

# High Rate Multivariate Polynomial Evaluation Codes

Swastik Kopparty \*

Mrinal Kumar <sup>†</sup>

Harry Sha <sup>‡</sup>

## Abstract

The classical Reed-Muller codes over a finite field  $\mathbb{F}_q$  are based on evaluations of  $m$ -variate polynomials of degree at most  $d$  over a product set  $U^m$ , for some  $d < |U|$ . Because of their good distance properties, as well as the ubiquity and expressive power of polynomials, these codes have played an influential role in coding theory and complexity theory. This is especially so in the setting of  $U$  being  $\mathbb{F}_q$  where they possess deep locality properties. However, these Reed-Muller codes have a significant limitation in terms of the rate achievable — the rate cannot be more than  $\frac{1}{m!} = \exp(-m \log m)$ .

In this work, we give the first constructions of multivariate polynomial evaluation codes which overcome the rate limitation — concretely, we give explicit evaluation domains  $S \subseteq \mathbb{F}_q^m$  on which evaluating  $m$ -variate polynomials of degree at most  $d$  gives a good code. For  $m = O(1)$ , these new codes have relative distance  $\Omega(1)$  and rate  $1 - \varepsilon$  for any  $\varepsilon > 0$ . In fact, we give two quite different constructions, and for both we develop efficient decoding algorithms for these codes that can decode from half the minimum distance.

The first of these codes is based on evaluating multivariate polynomials on simplex-like sets. The distance of this code is proved via a generalized Schwartz-Zippel lemma on the probability of non-zeros when evaluating polynomials on sparser subsets of  $U^m$  — the final bound only depends on the “shape” of the set, and recovers the Schwartz-Zippel bound for the case of the full  $U^m$ , while still being  $\Omega(1)$  for much sparser simplex-like subsets of  $U^m$ .

The second of these codes is more algebraic and, surprisingly (to us), has some strong locality properties. It is based on evaluating multivariate polynomials at the intersection points of hyperplanes in general position. It turns out that these evaluation points have many large subsets of collinear points. These subsets form the basis of a simple local characterization, and using some deeper algebraic tools generalizing ideas from Polischuk-Spielman [PS94], Raz-Safra [RS97] and Ben-Sasson-Sudan [BSS06], we show that this gives a local test for these codes. Interestingly, the set of evaluation points for these locally testable multivariate polynomial evaluation codes can be as small as  $O(d^m)$ , and need not occupy a constant or even noticeable fraction of the full space  $\mathbb{F}_q^m$ .

---

\*Department of Mathematics and Department of Computer Science, University of Toronto, Toronto, Canada. Email: [swastik.kopparty@utoronto.ca](mailto:swastik.kopparty@utoronto.ca). Research supported by an NSERC grant.

<sup>†</sup>Tata Institute of Fundamental Research, Mumbai, India. Email: [mrinal@tifr.res.in](mailto:mrinal@tifr.res.in). Research supported by the Department of Atomic Energy, Government of India, under project number RTI400112 and in part by a Google Research grant and SERB grants.

<sup>‡</sup>Department of Computer Science, University of Toronto, Toronto, Canada. Email: [shaharry@cs.toronto.edu](mailto:shaharry@cs.toronto.edu).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Our results . . . . .	4
1.1.1	CAP codes . . . . .	6
1.1.2	GAP codes . . . . .	8
1.1.3	Comparison to Reed-Muller codes . . . . .	10
1.2	Other Related Work . . . . .	10
<b>2</b>	<b>Notation and preliminaries</b>	<b>12</b>
<b>3</b>	<b>Explicit High Rate Polynomial Evaluation Codes</b>	<b>15</b>
3.1	A Geometric Construction . . . . .	16
3.2	A Construction Based on Simplices . . . . .	18
<b>4</b>	<b>Zero Patterns of Polynomials</b>	<b>20</b>
4.1	The Bivariate Case . . . . .	20
4.2	The General Case . . . . .	22
4.3	Shifting . . . . .	23
4.4	d-robustness . . . . .	26
4.5	The step evaluation set . . . . .	29
<b>5</b>	<b>Unique Decoding of CAP Codes</b>	<b>30</b>
5.1	Uneven GMD . . . . .	30
5.2	Decoding in two-dimensions . . . . .	33
5.3	Decoding on the $m$ -dimensional simplex . . . . .	35
<b>6</b>	<b>Unique Decoding of GAP codes</b>	<b>37</b>
6.1	Decoding in two dimensions . . . . .	39
6.2	Decoding in higher dimensions . . . . .	42
<b>7</b>	<b>Local testability of GAP codes</b>	<b>45</b>
7.1	Local characterization . . . . .	47
7.2	Soundness of the plane-point test . . . . .	50
7.2.1	Robust local characterization . . . . .	50
7.2.2	Analysis of the plane-point test . . . . .	52
7.3	Soundness of the line-point test . . . . .	55
7.3.1	Proof of Theorem 7.3 in two dimensions . . . . .	57
7.3.2	Proof of Theorem 7.3 in higher dimensions . . . . .	58
7.4	Robust local characterization of bivariate polynomials using restrictions to lines . . . . .	59

7.5	The divisibility lemma . . . . .	61
7.5.1	Proof of Lemma 7.12 in 2 dimensions . . . . .	62
7.5.2	Proof of Lemma 7.12 for general $m$ . . . . .	66
<b>A</b>	<b><math>d</math>-robustness in 2 variables</b>	<b>74</b>
A.1	High robustness regime . . . . .	74
A.2	Low robustness regime . . . . .	77
<b>B</b>	<b>Robust local characterization in <math>m</math> dimensions using error-locator polynomials</b>	<b>78</b>
<b>C</b>	<b>Intersection points of hyperplanes in general position are an interpolating set</b>	<b>82</b>

# 1 Introduction

Polynomial evaluation codes, such as the Reed-Solomon and the Reed-Muller codes, have played a central and outsized role in the classical theory of error-correcting codes. In recent decades, they have also been key to fundamental advances in complexity theory, pseudorandomness, cryptography and extremal combinatorics, for example in PCPs, interactive proofs, randomness extractors, pseudorandom generators and algebraic complexity theory. Furthermore, the pursuit of algorithms for and understanding of these codes has generated deep ideas in algebraic algorithms, property testing, and pseudorandom constructions.

A polynomial evaluation code is an error-correcting code obtained by evaluating  $m$ -variate polynomials of degree<sup>1</sup> at most  $d$  with coefficients in a finite field  $\mathbb{F}$  at some subset  $S$  of  $\mathbb{F}^m$ .

$$\mathcal{C} = \{(P(\mathbf{a}))_{\mathbf{a} \in S} \mid P(X_1, \dots, X_m) \in \mathbb{F}[X_1, \dots, X_m], \deg(P) \leq d\} \subseteq \mathbb{F}_q^S.$$

The rate of this code (which is the ratio of the dimension of the code as a linear space to the length of the codewords) is easily calculated to be  $R = \frac{\binom{d+m}{m}}{|S|}$ . The relative distance of this code (which is the largest fraction of coordinates on which any two codewords are guaranteed to be different) equals the largest  $\delta$  such that any polynomial of degree at most  $d$  has at least  $\delta|S|$  non-zero evaluations on  $S$ , and is more complicated to determine. We are interested in families of codes, with their length  $|S|$  going to infinity, and the rate  $R$  and the relative distance  $\delta$  are both  $\Omega(1)$ .

The most basic and well-known polynomial evaluation code is the  $m = 1$  case: the Reed-Solomon code. Regardless of the choice of  $S \subseteq \mathbb{F}$  of size  $n$ , this code has relative distance  $\delta = 1 - \frac{d}{n}$ : since distinct polynomials of degree  $d$  can agree on at most  $d$  points. Thus, Reed-Solomon codes achieve a rate-distance tradeoff of  $R + \delta = 1$ , which turns out to be optimal among all codes.

---

<sup>1</sup>Throughout this paper, degree refers to the total degree of a polynomial. Thus the degree of  $X^3Y^7 + X^6Y^6$  equals 12.

For the multivariate case,  $m \geq 2$ , the standard code is called the Reed-Muller code, and is based on choosing the evaluation domain  $S$  to be a product set  $U^m$ , where<sup>2</sup>  $|U| > d$ . In this case, the Schwartz-Zippel lemma provides the necessary bound on the minimum distance: it says that this code has relative distance at least  $1 - \frac{d}{|U|}$ . This gives a worse rate-distance tradeoff of

$$R = \frac{(1 - \delta)^m}{m!}.$$

For constant  $m$ , this gives constant rate and constant relative distance, albeit with  $R < 1/2$  unless  $m = 1$ .

Reed-Muller codes with  $U$  equal to the whole finite field  $\mathbb{F}$  are especially interesting because of their amazing locality properties. Indeed,  $\mathbb{F}^m$  contains within it  $m - 1$  and lower dimensional spaces, and  $m$ -variate polynomials of degree  $d$  stay degree  $d$  when restricted to these lower dimensional spaces. This is the basis for a natural local characterization of codewords – and leads to the extremely useful local testability and local decodability of these codes.

This rate limitation of multivariate polynomial codes is what our paper addresses – we find evaluation domains for multivariate polynomials of total degree  $d$  so that the resulting code has rate close to 1 while still having constant relative distance. This is a derandomization question – a random such evaluation domain of the right size has this property with high probability.

We give two kinds of explicit such evaluation domains, along with polynomial time decoding algorithms for the corresponding codes. The first code is more combinatorial, and its analysis and decoding involves a generalized Schwartz-Zippel lemma bounding the probability that a polynomial evaluates to zero on a general “shape”. The second kind is more geometric, based on intersections of hyperplanes in general position, and its analysis and decoding comes from the presence of the many intersecting lines contained within the evaluation domain. The rich geometric structure that comes from these lines, to our surprise, leads to the *local testability* of these codes, something that we did not expect was possible from such a drastic sparsification of the Reed-Muller code.

## 1.1 Our results

We give explicit multivariate polynomial evaluation codes with rate close to 1 and constant relative distance for the setting of  $m = O(1)$ .

**Theorem A (Informal):** *For  $m = O(1)$  and any  $\varepsilon > 0$ , there are explicit polynomial evaluation codes of  $m$ -variate polynomials that have rate  $R \geq 1 - \varepsilon$  and relative distance  $\Omega(1)$  (see [Theorems 3.5 and 3.12](#)).*

---

<sup>2</sup>Reed-Muller codes also include the code of degree  $d$  polynomials evaluated on  $U^m$  with the degree  $d \geq |U|$  (and individual degree  $< |U|$ ), but these codes cannot achieve constant rate and relative distance.

We give three proofs of this theorem.

The fastest (but least informative) proof is obtained via a general result, proved using an application of the polynomial method, showing that *interpolating sets* for  $m$ -variate polynomials of degree  $(1 + \Omega(1))d$  are themselves automatically good evaluation domains for polynomials of degree  $d$ ; any polynomial of degree  $d$  has many non-zero evaluations on such a set. This construction gives the tradeoff:

$$R^{1/m} + \delta^{1/m} = 1, \quad (1.1)$$

which for constant  $m$  gives us the statement of Theorem A.

Instantiating this with two known constructions of interpolating sets from the literature<sup>3</sup> (and elegantly presented in the paper of Bläser and Pandey [BP20]), we get the two high rate multivariate polynomial evaluation codes which this paper is about. We give a small taste of these codes now and in a later subsection we describe them in full generality.

- **CAP (Combinatorial Arrays for Polynomials) codes:** These codes are obtained by evaluating multivariate polynomials on simplex-like sets.

A simple example of such a set for the case  $m = 3$  is given below. Assume  $\mathbb{F}$  has characteristic at least  $1.1d$ . Then taking  $S \subseteq \mathbb{F}^3$  to be the “simplex” given by:

$$S = \{(i, j, k) : 0 \leq i, j, k \leq 1.1d, i + j + k \leq 1.1d\},$$

the code obtained by evaluating 3-variate polynomials of degree at most  $d$  on  $S$  turns out to have rate  $R = \left(\frac{1}{1.1}\right)^3$  and relative distance  $\delta = \left(\frac{0.1}{1.1}\right)^3$ .

This example seems somewhat related to the Reed-Muller code with evaluation domain  $[1.1d]^3$ . Indeed it is very closely related, and our most general CAP codes come from a unified viewpoint which captures both of these and many more kinds of evaluation domains.

- **GAP (Geometric Arrays for Polynomials) codes:** These codes are obtained by evaluating multivariate polynomials on intersection points of hyperplanes in general position.

A simple example of such a set for the case  $m = 3$  is given below. Let  $U \subseteq \mathbb{F}$  with  $|U| = 1.1d$ . Then we take the set  $S \subseteq \mathbb{F}^3$  by evaluating the elementary symmetric polynomials in 3 variables on distinct 3-tuples of  $U$ :

$$S = \{(a + b + c, ab + bc + ca, abc) : a, b, c \in U, \text{ distinct}\},$$

The code obtained by evaluating 3-variate polynomials of degree at most  $d$  on  $S$  turns out to have rate  $R = \left(\frac{1}{1.1}\right)^3$  and relative distance  $\delta = \left(\frac{0.1}{1.1}\right)^3$ .

---

<sup>3</sup>See [GS00] for a detailed history of interpolating sets for multivariate polynomials.

It may not be clear why the example above has anything to do with hyperplanes in general position, but it does, and it in fact contains many large collinear subsets of points. This leads to the (very surprising to us) local testability of these codes (Theorem C below).

Our second and third proofs of Theorem A are by a direct analysis of the distance of CAP and GAP codes. This yields much more information, and in particular, allows us to prove our next main result: that there are efficient unique decoding algorithms for both CAP codes and GAP codes up to half their minimum distance.

**Theorem B (Informal):** *There are polynomial time algorithms for unique decoding CAP codes and GAP codes from a fraction of errors that equals half their minimum distance (see [Theorems 5.7](#) and [6.12](#)).*

A unique decoding algorithm that decodes a code unto half its minimum distance is effectively providing an “algorithmic proof” of the minimum distance of the code. To prove Theorem B, we develop alternate and deeper understandings of these codes, and then identify and algorithmize the core combinatorial and geometric phenomena that give CAP codes and GAP codes their distance.

### 1.1.1 CAP codes

The CAP code described in the example above is based on evaluating polynomials at a small subset of the product set  $[1.1d]^3$ , thus improving upon the standard Reed-Muller codes, which require the whole product set.

In [Section 4](#), we develop a second, more transparent, proof that CAP codes have distance, based on a new generalization of the Schwartz-Zippel Lemma. This generalization comes from an observation that the standard proof of the Schwartz-Zippel lemma not only bounds the number of zeros in the grid, also gives information about the shape in which the zeros may appear. This information is quite powerful and underlies the ability of CAP codes to achieve rate close to 1.

To state this generalized Schwartz-Zippel Lemma, we introduce some notation to specify “shapes”. Let  $t$  be a natural number. Let  $J$  be a subset of  $[t]^m$  which is downward closed under the coordinatewise order  $\preceq$ : namely, if  $\mathbf{j}, \mathbf{j}' \in [t]^m$  with  $\mathbf{j}' \preceq \mathbf{j}$  (which means we have coordinatewise inequality), then  $\mathbf{j} \in J$  implies  $\mathbf{j}' \in J$ .

We say that a downward-closed set  $J \subseteq [t]^m$  has  $d$ -robustness at least  $B$  if, whenever we delete  $d$  axis parallel  $(m - 1)$ -dimensional hyperplanes from  $J$ , at least  $B$  points of  $J$  remain.

We can now state our generalized Schwartz-Zippel lemma.

**Generalized Schwartz-Zippel Lemma** *Let  $J$  be a subset of  $[t]^m$  with  $d$ -robustness at least  $B$ .*

Let  $U \subseteq \mathbb{F}$  with  $|U| = t$ , and let  $U = \{u_1, \dots, u_t\}$ . Define  $U[J]$  to be the set  $S \subseteq U^m$  given by:

$$S = \{(u_{j_1}, u_{j_2}, \dots, u_{j_m}) \mid \mathbf{j} \in J\}.$$

Let  $P(X_1, \dots, X_m) \in \mathbb{F}[X_1, \dots, X_m]$  be a nonzero polynomial of degree at most  $d$ . Then:

$$|\{\mathbf{a} \in S \mid P(\mathbf{a}) \neq 0\}| \geq B.$$

Thus if we find a subset  $J$  with size  $n$  and  $d$ -robustness  $\geq \delta n$ , the polynomial evaluation code with evaluation set  $U[J]$  will have relative distance  $\delta$  (and rate  $\frac{\binom{d+m}{m}}{n}$ ).

When we take  $J = [t]^m$ , then we recover the standard Schwartz-Zippel lemma. When we take  $J = \{\mathbf{j} \in [t]^m \mid \sum_i j_i \leq t\}$ , we recover the bound for simplex-like CAP codes described earlier (but in much greater generality – since we do not need the regular-spacedness of the set  $U$ ).

Other  $J$  can be taken, and they can give other rate vs distance tradeoffs. For given constants  $R, \delta \in [0, 1]$  and  $m = O(1)$ , the question of whether there is a downward-closed  $J \subseteq [t]^m$  (for some large  $t$ ) of a given size  $n = (1 + o(1)) \cdot \frac{1}{R} \cdot \frac{d^m}{m!}$  with  $d$ -robustness  $\delta n$  is an interesting combinatorial question. Interestingly, there are choices of  $J$  which go beyond (even in the asymptotic setting) both the full product set and the simplex constructions, even though both are “locally optimal”.

The proof of the Generalized Schwartz-Zippel Lemma is by induction on a stronger statement – this statement not only bounds the size of, but also gives structural information about, the “shape” of the zero set of any multivariate polynomial. This refined information also turns out to be useful for our decoding algorithms, which we discuss next.<sup>4</sup>

**Decoding:** Based on the ideas above, we show that CAP codes can be decoded in polynomial time from half their minimum distance. We do this by developing a suitable generalization of a peeling-based algorithm of Kim and Kopparty [KK17] for decoding Reed-Muller codes on a product set. Suppose we are trying to decode the bivariate Reed-Muller code over a grid  $U^2$ . At a high level, [KK17] write  $f(X, Y) = \sum_{i=0}^d f_i(X)Y^i$ . They then decode the univariate polynomials  $f(x, Y)$  on each column. From these decodings, they have estimates of  $f_d(x)$  for each  $x \in U$  and use these to decode for  $f_d$  using the GMD algorithm – the classical algorithm for decoding concatenated codes that uses “soft” information. Peeling off the polynomial  $f_d(X)Y^d$  from  $f(X, Y)$ , and iteratively applying this argument again gives the final algorithm. The main challenge for adapting this algorithm to CAP codes is that each column has a variable length (and thus, each univariate polynomial  $f(x, Y)$  is evaluated on a different number of points). We address this by

---

<sup>4</sup>It was pointed to us by S Venkitesh [?] that this generalized Schwartz-Zippel lemma can also be shown as a consequence of a classic theorem of Macaulay [?] using a slight adaptation of the techniques in [?].

introducing a variant of the GMD algorithm that allows for varying inner code lengths and inner code distances. This uses a recent new combinatorial analysis of the GMD algorithm by [BHS23].

### 1.1.2 GAP codes

GAP codes are based on intersections of hyperplanes in general position. Suppose we take a collection  $\mathcal{H}$  of  $t$  such hyperplanes in  $\mathbb{F}^m$ . Let  $\mathcal{S}$  be the set of all  $m$ -wise intersections of hyperplanes from  $\mathcal{H}$ . This set  $\mathcal{S}$ , of size  $\binom{t}{m}$ , is the set of evaluation points for GAP codes<sup>5</sup>.

Combinatorially these sets look quite complex, but there is rich structure sitting inside them. Indeed, if we take the intersection of some  $(m - 1)$  hyperplanes from  $\mathcal{H}$ , we get a line  $L$  in  $\mathbb{F}^m$ . It is easy to see that this line contains  $t - (m - 1)$  points of  $\mathcal{S}$ . If  $t - (m - 1)$  is at least  $(1 + \Omega(1))$  times  $d$ , then polynomials of degree  $d$  evaluated on  $\mathcal{S}$  will look like Reed-Solomon codes of constant relative distance when restricted  $\mathcal{S} \cap L$ .

**Decoding:** We can use these observations to get a decoding algorithm for GAP codes up to half the minimum distance. The algorithm is based on decoding along each line  $\mathcal{S} \cap L$  to the nearest Reed-Solomon codeword. How do we stitch together all these Reed-Solomon codewords? Miraculously, viewing things the right way, the  $m$ -variate GAP code turns out to be a *concatenated code*, where the outer codes are Reed-Solomon codes over the function field  $\mathbb{F}(X_1, \dots, X_{m-1})$ , and the inner codes are  $(m - 1)$ -variate GAP codes. From this point of view, the way to stitch together all these Reed-Solomon codewords is to simply use the classical GMD decoder of Forney for concatenated codes!

Compared to the big struggle that happened within [KK17] for decoding Reed-Muller codes, which involved slowly correcting errors in tiny steps via a back and forth between the inner Reed-Solomon decoder and the GMD decoder, and a refined version of that struggle that we just did for decoding CAP codes, this clean decoder for GAP codes is a source of great joy (and relief).

**Local Testing:** These local Reed-Solomon codes present in GAP codes also form the basis for a local characterization – we show in Theorem 7.5 that a function defined on  $\mathcal{S}$  is a codeword of the GAP code if and only if its restriction to any line  $L$  gives a codeword of the Reed-Solomon code of degree  $d$  univariate polynomials.

Using (and generalizing) the ideas of Polischuk-Spielman [PS94] and Raz-Safra [RS97], we upgrade this local characterization to a *robust* local characterization (involving planes), and thereby show local testability of these codes.

---

<sup>5</sup>The example GAP code presented above was just an instantiation of this framework with the “moment-curve construction” of hyperplanes in general position.



**Theorem C (Informal):** For  $m = O(1)$ , GAP codes of length  $n$  with constant relative distance are locally testable with  $O(n^{2/m})$  queries. In particular, for any  $\varepsilon > 0$ , there exist GAP codes of length  $n$  and rate  $(1 - \varepsilon)$  that are locally testable with  $O(n^{2/m})$  queries (see [Theorem 7.2](#)).

The above result is proved using a version of the plane-vs-point test for standard Reed-Muller codes of Raz-Safra [[RS97](#)] (see also its abstraction to tensor codes by Ben-Sasson-Sudan [[BSS06](#)] and Viderman [[Vid15](#)]). The rich geometric structure within GAP codes turns out to be sufficient to implement this proof strategy, despite the drastically smaller number of planes available.

In fact, the proof turns out to be substantially *easier* than both the proofs of Raz-Safra and Ben-Sasson-Sudan. The underlying reason is the following fact ([Lemma 7.6](#)) about gluing polynomials defined on hyperplanes *in general position*: if we are given  $(m - 1)$ -dimensional hyperplanes  $H_1, \dots, H_\ell$  in  $\mathbb{F}^m$  in general position for some  $\ell \geq d$ , and degree  $d$  polynomials  $P_i : H_i \rightarrow \mathbb{F}$  defined on them which are mutually consistent on the intersections  $H_i \cap H_j$ , then there is a unique degree  $d$  polynomial  $P(X_1, \dots, X_m)$  consistent with all of them; namely  $P|_{H_i} = P_i$  for all  $i$ . Without the general position assumption, this no longer holds (for example, because of parallel hyperplanes), and this complicates the analysis of the hyperplane consistency graph in [[RS97](#)] and [[BSS06](#)]. We think this explains why GAP codes admit a much simpler analysis, and also why the analysis directly works in the low-distance/high-rate regime<sup>6</sup>.

We also analyze the natural line-vs-point test for GAP codes, but this ends up being quantitatively weaker – it only gives nontrivial testability when the rate is at most  $(1 - \varepsilon)2^{-m}$  (which is still higher than the rate achievable by the classical Reed-Muller code). Thus, unlike for the plane-vs-point test, our analysis for the line-vs-point test does not work all the way up to rate approaching 1.

Along the way, we prove an independently interesting statement ([Lemma 7.12](#)) about a local-to-global phenomenon for divisibility. If we have two polynomials  $A(X_1, \dots, X_m)$  and  $B(X_1, \dots, X_m)$  of degree at most  $d$  such that for at least  $3d$  hyperplanes  $H$  in general position, the  $(m - 1)$ -variate polynomial  $A|_H$  divides the  $(m - 1)$ -variate polynomial  $B|_H$ , then in fact  $A$  divides  $B$ . Such statements with a  $O(d^2)$  bound on the number of hyperplanes needed were implicitly proved in a work of Forbes [[For15](#)] and a work of Harsha, Kumar, Saptharishi and Sudan [[HKSS23](#)]. Our work is the first to get an  $O(d)$  bound for this setting. For  $m = 2$  our proof is based on an adaptation of an argument of Polischuk-Spielman, using Bezout’s theorem and a basic claim about intersection multiplicities of two curves and with a common tangent. For larger  $m$  it involves much more non-trivial algebraic machinery involving intersection multiplicities of varieties, but the final statement is still quite clean.

Finally, we remark that GAP codes can be viewed as Tanner codes (or HDX-Tanner codes)

---

<sup>6</sup>For the Ben-Sasson-Sudan setting of tensor codes, the original proof only worked when the base code had large distance. Testability when the base code has low distance was only later discovered by Viderman [[Vid15](#)] using some further clever ideas.

on the complete graph (for  $m = 2$ ) or the complete  $m$ -dimensional complex (for larger  $m$ ) on  $t$  vertices, where the inner codes are Reed-Solomon codes. That these codes have high rate is a surprise; it goes beyond what pure dimension counting implies. That these codes are in fact multivariate polynomial evaluation codes is another surprise. This viewpoint is related to the recent constructions of Dinur, Liu and Zhang [DLZ23].

### 1.1.3 Comparison to Reed-Muller codes

How do CAP codes and GAP codes compare to Reed-Muller codes?

At a very high level, CAP codes take the principles underlying Reed-Muller codes and squeeze everything that could possibly be squeezed out of it. The proof of distance of CAP codes generalizes the Schwartz-Zippel lemma, and our decoding algorithms for CAP codes generalize the decoding algorithms for Reed-Muller codes from [KK17]. We feel that our generalized Schwartz-Zippel lemma as well as our decoding algorithm for CAP codes sheds light on the true content of the Schwartz-Zippel lemma and the reason for the efficient decodability of Reed-Muller codes, while simultaneously highlighting the route to modifying Reed-Muller codes to achieve higher rate.

On the other hand, GAP codes look nothing like Reed-Muller codes, and in many ways they are much cleaner and more natural than Reed-Muller codes. The proof of distance is much simpler, and in fact the decoding algorithm just follows directly from viewing it as a concatenated code in the right way. Even more interestingly, GAP codes are locally testable (in part because of the many lower dimension subspaces present in the evaluation domain), while Reed-Muller codes are not unless the evaluation domain is all of  $\mathbb{F}_q^m$ .

Reed-Muller codes on general product sets  $U^m$  also have many local constraints, with  $m$  axis-parallel lines through each evaluation point. The local constraint says that the restriction to these lines should be degree at most  $d$ . But the key difference with GAP codes is that these local constraints do not characterize Reed-Muller codes – they only characterize evaluation codes of multivariate polynomials of *individual degree* at most  $d$  – and this is a significantly bigger code than the Reed-Muller code which has a total degree condition.

Finally we remark that since the evaluation domains for CAP and GAP codes do not essentially depend on the degree  $d$ , CAP and GAP codes automatically are *multiplication codes* [Mei13] – the pointwise product of codewords is a codeword of another good code – one of the key properties of Reed-Muller codes which is important for applications in complexity theory.

## 1.2 Other Related Work

There have been many influential works studying the problem of finding small evaluation sets for  $m$ -variate degree- $d$  polynomials which preserve their good distance. Our results are focused on  $m$  constant and  $d$  big (where we get codes rate nearly 1), while the existing literature addresses large

$m$ . The known results give good-distance polynomial evaluation codes with inverse polynomial rate when  $m$  and  $d$  are both large, and rate nearly 1 when  $d$  is constant and  $m$  is large.

Recall the setup. Let  $k = \binom{m+d}{d}$ , which is the dimension of the space of  $m$ -variate polynomials of degree at most  $d$ . We are seeking evaluation domains where such polynomials still have constant fraction distance on this domain, with the size of the evaluation domain being as small a function of  $k$  as possible.

There is a long line of work (Chen-Kao [CK97], Lewin-Vadhan [LV98], Agrawal-Biswas [AB03], Klivans-Spielman [KS01]) studying pseudorandom generators against low degree polynomials over large fields using few random bits. The fundamental (and very general) result of Klivans and Spielman on fooling sparse polynomials from this line of work yields explicit evaluation sets for  $m$ -variate degree- $d$  polynomials of size  $\binom{m+d}{d}^{O(1)} = k^{O(1)}$ . When  $m = d$  for example, this gives a multivariate polynomial evaluation code with block-length  $k^{O(1)}$ , while the evaluation domains coming from the Schwartz-Zippel lemma would have to have size  $d^m = k^{\Omega(\log k)}$ . These results require the field size to be very large (at least  $\binom{m+d}{d}$ ).

Another line of work exploiting deeper algebraic geometric properties of polynomials over large fields gives interesting evaluation sets over fields of moderately large size. These works by Bogdanov [Bog05], Derksen-Viola [DV22] and Dwivedi-Guo-Volk [DGV24] lead to explicit evaluation sets of size  $k^{O(1)}$  when the field size is  $\text{poly}(d, n)$  (this can also work over slightly smaller fields, giving slightly larger explicit evaluation sets).

In the other extreme setting of parameters, when  $d = O(1)$ ,  $m$  is large and the field is small (e.g.  $\mathbb{F}_2$ ), there is another approach to this problem<sup>7</sup> using sumsets of epsilon-biased sets. Here the result of Dvir-Shpilka [DS11] and Viola [Vio08] (using a key construction of Bogdanov-Viola [BV10], see also Lovett [Lov08]) give explicit evaluation sets of size  $(1 + \epsilon)k$ . There is also an elegant decoding algorithm for these codes, given by Dvir and Shpilka [DS11].

We also mention a classical coding theory approach to constructing evaluation sets for the setting  $m = 2$ , based on Algebraic Geometry codes. If we take the  $\mathbb{F}$ -points of an irreducible degree  $d + 1$  curve in the plane, the number of points where a bivariate polynomial  $Q(X_1, X_2)$  of degree at most  $d$  can vanish on  $C$  is at most  $d \cdot (d + 1)$  (by Bezout's theorem). Thus, if we take  $n = (1 + \Omega(1))d^2$  points on  $C$  as our evaluation domain, we guarantee that polynomials of degree  $d$  have positive fraction distance on this domain. Note, however, that this argument cannot achieve rates  $\geq 1/2$ , and is thus comparable to the Schwartz-Zippel approach for  $m = 2$ .

A very interesting recent work by Dinur, Liu and Zhang [DLZ23] gave new constructions of constant query locally testable codes of subconstant rate using the explicit description of high dimensional expanders using finite field geometry, and with Reed-Solomon "inner codes" in a

---

<sup>7</sup>To be precise, here we look for evaluation domains in  $\mathbb{F}_q^m$  for polynomials of total degree  $d$  and *individual degree* at most  $q - 1$ : the individual degree constraint is natural because points in  $\mathbb{F}_q^m$  cannot witness the nonzeroness of the polynomials  $X_i^q - X_i$ . For this setting,  $k$  is taken to be the dimension of this space of polynomials, and may be smaller than  $\binom{m+d}{d}$ .

Tanner code like construction. Our GAP codes can be viewed through a similar lens, but with the complete  $m$ -dimensional complex in place of the HDX.

Another interesting recent line of work is by Bafna, Srinivasan and Sudan [BSS20] and Amireddy, Srinivasan and Sudan [ASS23], who studied the local properties of multivariate polynomial codes evaluated on product sets. These gave the first examples of nontrivial locality for multivariate polynomial evaluation codes which are not evaluated on the full  $\mathbb{F}^m$ . Our GAP codes give another such example (but only for local testability).

We close by mentioning two nice open problems.

1. Are there explicit polynomial evaluation codes of rate and relative distance both  $\Omega(1)$  when  $m$  and  $d$  are both large?
2. By a small variant of the argument of Saraf-Yekhanin [SY11], a random choice of evaluation points in  $\mathbb{F}^m$  for the space of polynomials of degree at most  $d$  will achieve near-optimal tradeoff of  $R + \delta = 1 - o(1)$ , provided  $|\mathbb{F}| > \omega(d)$ . Thus, there are better multivariate polynomial evaluation codes out there, even for  $m = 2$ . We think finding an explicit one of them for which there is efficient decoding (and possibly locally decoding and locally testing) is a very interesting problem for future work.

**Organization:** The rest of this paper is organized as follows. We set up the necessary notation and prelims in Section 2. We describe our construction of high rate codes in Section 3, followed by a more detailed discussion on the ideas underlying the construction and properties of CAP codes in Section 4. We describe the decoding algorithms for CAP codes in Section 5 and GAP codes in Section 6. We prove the local testability of GAP codes in Section 7.

## 2 Notation and preliminaries

### Notation

Throughout the paper, we use  $\mathbb{F}$  to denote a field. We use  $X, Y, Z$  etc in capitals to denote formal variables, and  $x, y, z$  etc in small to denote field elements in  $\mathbb{F}$ . We use bold for vectors and subscripts to denote the entries. For example,  $\mathbf{x} \in \mathbb{N}^m$ , and the entries of  $\mathbf{x}$  are  $x_1, \dots, x_m$ . Let  $\mathbb{N}_{<k} = \{0, 1, \dots, k-1\}$ . The arity of bold letters will be clear from the context. For a polynomial  $f \in \mathbb{F}[X_1, \dots, X_m]$ ,  $Z(f), N(f)$  be the set of zeros and non-zeros of  $f$  in  $\mathbb{F}$ .

For any alphabet  $\Sigma$ , natural number  $n \in \mathbb{N}$ , and strings  $\mathbf{a}, \mathbf{b} \in \Sigma^n$ ,  $\Delta(\mathbf{a}, \mathbf{b})$  denotes the Hamming distance between  $\mathbf{a}$  and  $\mathbf{b}$ , i.e. the number of coordinates on which  $\mathbf{a}$  and  $\mathbf{b}$  disagree with each other, and  $\delta(\mathbf{a}, \mathbf{b})$  denotes their fractional Hamming distance, i.e.  $\delta(\mathbf{a}, \mathbf{b}) = \Delta(\mathbf{a}, \mathbf{b})/n$ .

## Error-correcting codes

The following are standard definitions in coding theory. A set  $C \subset \Sigma^n$  is called an error-correcting code of distance at least  $d$  if for any  $c, c' \in C$ ,  $\Delta(c, c') \geq d$ . The relative distance,  $d/n$ , typically denoted as  $\delta(C)$  or simply  $\delta$  where  $C$  is clear from the context. The dimension of the code is  $\log_{|\Sigma|}(C)$ . The rate of the code, typically  $R$ , is equal to  $\log_{|\Sigma|}(C)/N$ .

## Codes based on polynomial evaluation

This paper is focused on error-correcting codes obtained from polynomial evaluations. The message space is the space of  $m$ -variate polynomials of total degree at most  $d$ , and a polynomial,  $f$ , is encoded by recording the evaluations of  $f$  on a subset  $S \subset \mathbb{F}^m$ . We think of the list of evaluations  $(f(\mathbf{x}))_{\mathbf{x} \in S}$  as a function from  $S$  to  $\mathbb{F}$ . More formally, we have the following definition.

**Definition 2.1** (Polynomial Evaluation Code). *The  $m$ -variate Polynomial Evaluation Code of degree  $d$  on a set  $S$ , denote  $E_{m,d,S}$  is the set of evaluations of (total) degree  $d$ ,  $m$ -variate polynomials on the set  $S$ . That is*

$$E_{m,d,S} = \{f : S \rightarrow \mathbb{F} : f \in \mathbb{F}[X_1, \dots, X_m], \deg(f) \leq d\}$$

◇

Important examples of polynomial evaluation codes include Reed-Solomon codes, where  $m = 1$ , and Reed-Muller Codes, where  $S$  is a  $m$ -dimensional grid.

**Definition 2.2** (Reed-Solomon Code). *For any  $d, n \in \mathbb{N}$  with  $d < n$ , let  $\text{RS}_{d,n} = E_{1,d,U}$  where  $U$  is some subset of  $\mathbb{F}$  of size  $n$ .*

◇

**Definition 2.3** (Reed-Muller Code). *For any  $d, n, m \in \mathbb{N}$  with  $d < n$ ,  $m \geq 1$ , let  $\text{RM}(m, d, n) = E_{m,d,U^m}$ , where  $U$  is some subset of  $\mathbb{F}$  of size  $n$ .*

◇

Let  $C$  be a polynomial evaluation code where the evaluation set is  $S$ . Define an error and erasure pattern to be a function  $e : S \rightarrow (\mathbb{F} \cup \{?\})^N$ . When added to a codeword, the symbol  $?$  corresponds to erasures, and non-zero entries of  $e$  correspond to errors. Define the weight of an error and erasure pattern to be  $\text{wt}(e) = 2|\{x \in S : e(x) \neq 0, e(x) \neq ?\}| + |\{x \in S : e(x) = ?\}|$ ; i.e., the weight of an error and erasure pattern is twice the number of errors plus the number of erasures. We call  $e$  an error pattern when there are no erasures.

Let  $A$  be an algorithm and  $C$  be a code. We say that  $A$  decodes  $C$  from error (and erasure) patterns of weight at most  $D$  if for any error (and erasure) pattern  $e$  of weight at most  $D$ , and any codeword  $c \in C$ ,  $A$  returns  $c$  on input  $c + e$ .

## Hitting sets and interpolating sets

**Definition 2.4** (Hitting sets). *A subset  $S \subseteq \mathbb{F}^m$  is called a hitting set for  $m$ -variate polynomials of degree  $d$  if every non-zero polynomial  $f$  of degree at most  $d$  in  $\mathbb{F}[X_1, \dots, X_m]$ , there is some point  $\mathbf{x} \in S$  such that*

$f(\mathbf{x}) \neq 0$ . ◇

**Definition 2.5** (Interpolating sets). A set  $S \subseteq F^m$  is an interpolating set for  $m$ -variate polynomials of degree  $d$  if for all functions  $f : S \rightarrow \mathbb{F}$ , there is a unique polynomial,  $g \in \mathbb{F}[X_1, \dots, X_m]$  of degree at most  $d$  such that for all  $a \in S$ ,  $f(a) = g(a)$ . ◇

## Concatenated Codes and GMD decoding

A key step in our decoding algorithms is to express CAP and CAP codes as concatenated codes. This allows us to leverage known algorithms for decoding concatenated codes such GMD algorithm of Forney [For66].

**Definition 2.6.** Let  $C_{out} : \Sigma_1^K \rightarrow \Sigma_2^N$  be a code with block length  $N$  over an alphabet  $\Sigma_2$  of size  $Q$ . Let  $C_{in} = (C_1, C_2, \dots, C_N)$  be an  $N$ -tuple of codes where for each  $i$ , the code  $C_i$  is of the form  $C_i : \Sigma_2 \rightarrow \Sigma_3^n$ , i.e. the message space is identified with the alphabet  $\Sigma_2$  of  $C_{out}$  and  $C_i$  maps each such alphabet to a vector of length  $n$  over a new alphabet  $\Sigma_3$ . The concatenation of  $C_{out}$  with  $C_{in}$ , denoted as  $C_{in} \circ C_{out}$  is a code that maps messages in the space  $\Sigma_1^K$  to vectors of length  $(Nn)$  over the alphabet  $\Sigma_3$ , and the encoding maps a message  $\mathbf{m} \in \Sigma_1^K$  to the codeword

$$C_{in} \circ C_{out}(\mathbf{m}) = (C_1(C_{out}(\mathbf{m})_1), C_2(C_{out}(\mathbf{m})_2), \dots, C_N(C_{out}(\mathbf{m})_N)),$$

where,

$$C_{out}(\mathbf{m}) = (C_{out}(\mathbf{m})_1, C_{out}(\mathbf{m})_2, \dots, C_{out}(\mathbf{m})_N) \in \Sigma_2^N.$$

◇

We now state a result of Forney for decoding concatenated codes up to half their minimum distance that will be crucially used in our proofs.

**Theorem 2.7** (Forney [For66]). Let  $C_{out}, C_{in} = (C_1, C_2, \dots, C_N)$  be codes and  $C_{in} \circ C_{out}$  be their concatenation as in Definition 2.6. Suppose each inner code  $C_i$  has distance  $d$ , and  $C_{out}$  has distance  $D$ . Suppose there exists algorithms  $A_1, \dots, A_N, A_{out}$  such that

- For every  $i$ ,  $A_i$  decodes  $C_i$  from error patterns of weight less than  $d$ .
- $A_{out}$  decodes  $C_{out}$  from errors and erasures patterns of weight less than  $D$

Then, there is a deterministic algorithm for (unique) decoding of the concatenated code  $C_{in} \circ C_{out}$  from error patterns of weight less than  $dD$ . Furthermore, the algorithm makes at most one function call to each  $A_i$ , at most  $O(N)$  function calls to the algorithm  $A_{out}$ , and performs at most  $\text{poly}(N, n, \log(|\Sigma_1| \cdot |\Sigma_2| \cdot |\Sigma_3|))$  additional computation.

### 3 Explicit High Rate Polynomial Evaluation Codes

The following technical lemma gives a general connection between robust hitting sets for polynomials of degree  $d$  and hitting sets for polynomials of slightly higher degree. The lemma will give us a clean and direct way of analyzing our constructions of high-rate polynomial evaluation codes.

**Lemma 3.1.** *Let  $m, d, D$  be natural numbers with  $D > d$  and  $\mathbb{F}$  be any field. Let  $S$  be a subset of  $\mathbb{F}^m$  that is a hitting set for polynomials of degree  $D$ . Then, for any non-zero  $m$ -variate polynomial  $f$  of degree at most  $d$ , the number of non-zeroes of  $f$  in the set  $S$  is at least  $\binom{D-d+m}{m}$ .*

*Proof.* The proof of the lemma is via an instantiation of the polynomial method.

Let  $f$  be an arbitrary non-zero polynomial on  $m$  variables and degree  $d$  and  $T_f \subseteq S$  denote the set of non-zeroes of  $f$  in the set  $S$ . Suppose  $|T_f| < \binom{D-d+m}{m}$ . The next general interpolation claim exploits this assumption.

**Claim 3.2.** *If  $|T_f| < \binom{D-d+m}{m}$ , then there is a non-zero polynomial  $g$  on  $m$  variables and degree at most  $(D - d)$  that vanishes everywhere on  $T_f$ .*

*Proof of Claim 3.2.* For the proof of the claim, we think of the coefficients of  $g$  as formal variables and consider the system of homogeneous linear equations obtained by imposing the constraints that  $g$  vanishes at every point in  $T_f$ . The number of variables in this linear system equals the number of monomials of degree at most  $(D - d)$  on  $m$  variables, i.e.,  $\binom{D-d+m}{m}$  and the number of homogeneous linear constraints equals the size of  $T_f$ , which by our assumption is strictly less than  $\binom{D-d+m}{m}$ . Thus, the number of variables in this homogeneous linear system is strictly more than the number of constraints imposed. Hence, the system has a non-zero solution.  $\square$

With the  $g$  given by the above claim in hand, consider the polynomial  $h = f \times g$ . Clearly, the degree of  $h$  is at most  $D$ , and  $h$  is non-zero since  $g$  and  $f$  are non-zero. Moreover, since  $f$  vanishes on  $S \setminus T_f$  (by definition of  $T_f$ ) and  $g$  vanishes on  $T_f$  (by construction of  $g$ ), we get that  $h$  vanishes on all points in the set  $S$ , which contradicts the fact that  $S$  is a hitting set for  $m$ -variate polynomials of degree at most  $D$ . Thus, the size of the set  $T_f$  must be at least  $\binom{D-d+m}{m}$ .  $\square$

We now discuss the construction of two sets of evaluation points that give us polynomial evaluation codes of high rate and constant relative distance (when the number of variables is constant). From Lemma 3.1, we get that one way of doing this is to construct hitting sets of small size. Two such constructions of hitting sets were given by Bläser & Pandey [BP20]. In fact, their hitting sets are optimal, in the sense that their hitting set for  $m$ -variate degree  $D$  polynomials has size equal to  $\binom{m+D}{D}$  (in other words, these hitting sets are interpolating sets). In the next section, we discuss these constructions and invoke them with Lemma 3.1 to get our codes.



### 3.1 A Geometric Construction

To describe this construction, we rely on the notion of hyperplanes in general position that we now define. Recall that a hyperplane  $H$  in  $\mathbb{F}^m$  is just an affine subspace of co-dimension one. In other words, there is a non-constant linear polynomial  $L(X_1, X_2, \dots, X_m) := \alpha_0 + \sum_{i=1}^m \alpha_i X_i$ , where each  $\alpha_i$  is a field element such that

$$H = \{(b_1, \dots, b_m) \in \mathbb{F}^m : L(b_1, \dots, b_m) = 0\}.$$

We say that a set  $\mathcal{H} = \{H_1, H_2, \dots, H_t\}$  of hyperplanes in  $\mathbb{F}^m$  are in *general position* if for every subset of hyperplanes in  $\mathcal{H}$  of size  $m$  intersect in one point, and no subset of hyperplanes in  $\mathcal{H}$  of size  $m + 1$  intersect.

Given any collection of  $D + m$  hyperplanes in general position in  $\mathbb{F}^m$ , Bläser & Pandey gave a construction of hitting sets of optimal size for  $m$ -variate degree  $D$  polynomials. It is not hard to show that this set of points is, in fact, an interpolating set (and not just a hitting set) for degree  $D$  polynomials on  $m$  variables. This stronger property turns out to be useful for our proofs in this paper.

**Theorem 3.3** (Bläser & Pandey [BP20]). *Let  $m, D$  be natural numbers,  $\mathbb{F}$  be a field, and let  $\mathcal{H}$  be a set of size  $(m + D)$  hyperplanes in general position in  $\mathbb{F}^m$ . Let  $T \subseteq \mathbb{F}^m$  be the set of all points obtained by intersecting some  $m$  elements of  $\mathcal{H}$ . Then,  $T$  is of size  $\binom{m+D}{m}$  and  $T$  is an interpolating set for  $m$ -variate polynomials of degree  $D$  over  $\mathbb{F}$ .*

We give a quick proof in [Appendix C](#), which is slightly simpler than the already simple proof in [BP20].

We refer to codes obtained from hitting sets obtained from [Theorem 3.3](#) as Geometric Arrays for Polynomials (GAP).

**Definition 3.4** (GAP). *Let  $t > 0$ , and  $\mathbb{F}$  be a finite field of size at least  $t$ . Define  $\text{GAP}_{m,d,t} = E_{m,d,T}$ , where  $T$  is the set of  $m$ -wise intersections of some  $t$  hyperplanes in general position. I.e.,  $\text{GAP}_{m,d,t}$  is the set of evaluations of  $m$ -variate polynomials of total degree at most  $d$  on  $T$ . To specify a specific set of  $t$  hyperplanes,  $\mathcal{H}$ , we write  $\text{GAP}_{m,d,t}^{\mathcal{H}}$*   $\diamond$

We now combine [Theorem 3.3](#) and [Lemma 3.1](#) to obtain the following theorem.

**Theorem 3.5.** *Let  $m, d, t \in \mathbb{N}$  with  $t > m + d$ . Then  $\text{GAP}_{m,d,t}$  has distance at least  $\binom{t-d}{m}$ . Furthermore, for any  $\varepsilon > 0$ , if  $t = m + d + \varepsilon d - 1$ , then rate of  $\text{GAP}_{m,d,t}$  is least  $\left(\frac{1}{1+\varepsilon}\right)^m$  and relative distance is at least  $\left(\frac{\varepsilon}{1+\varepsilon}\right)^m$ .*

*Proof.* Let  $C = \text{GAP}_{m,d,t}$ . Let  $\mathcal{H}$  be a set of  $t$  hyperplanes in general position, and let  $T$  be the set of their  $m$ -wise intersections.



By [Theorem 3.3](#),  $T$  is a hitting set for polynomials of degree  $t - m$ . Let  $f$  be any  $m$ -variate polynomial of degree at most  $d$ . By [Lemma 3.1](#), the number of non-zeros of  $f$  in  $T$  is at least  $\binom{t-m-d+m}{m} = \binom{t-d}{m}$ . Therefore, since  $C$  is linear,  $C$  has distance at least  $\binom{t-d}{m}$ .

Now let  $t = m + d + \varepsilon d$ . We will now compute the rate,  $R$ , and relative distance,  $\delta$ . Recall that  $C$  has dimension  $\binom{m+d}{m}$ , and block length  $\binom{t}{m}$ . Thus,

$$\begin{aligned} R &= \frac{\binom{m+d}{m}}{\binom{m+d+\varepsilon d}{m}} \\ &= \frac{(m+d)!}{m!d!} \cdot \frac{m!(d+\varepsilon d)!}{(m+d+\varepsilon d)!} \\ &= \prod_{i=0}^{m-1} \frac{(m+d-i)}{(m+d+\varepsilon d-i)}. \end{aligned}$$

Now, using the observation that for all  $i \leq m$ ,  $\frac{(m+d-i)}{(m+d+\varepsilon d-i)}$  is at least  $1/(1+\varepsilon)$ , we get that  $R \geq \left(\frac{1}{1+\varepsilon}\right)^m$ . Then,

$$\begin{aligned} \delta &= \frac{\binom{m+\varepsilon d}{m}}{\binom{m+d+\varepsilon d}{m}} \\ &= \frac{(m+\varepsilon d)!}{m!(\varepsilon d)!} \cdot \frac{m!(d+\varepsilon d)!}{(m+d+\varepsilon d)!} \\ &= \prod_{i=0}^{m-1} \frac{(m+\varepsilon d-i)}{(m+d+\varepsilon d-i)}. \end{aligned}$$

We now observe that for all  $i \leq m$ ,  $\frac{(m+\varepsilon d-i)}{(m+d+\varepsilon d-i)}$  is at least  $\varepsilon/(1+\varepsilon)$ . Thus,  $\delta \geq \left(\frac{\varepsilon}{1+\varepsilon}\right)^m$ , as required.  $\square$

**Remark 3.6.** The distance and rate trade-off for GAP codes is  $R^{1/m} + \delta^{1/m} \geq 1$ .  $\diamond$

In comparison, the trade-off obtained by the standard Reed-Muller code,  $\text{RM}_{m,d,\ell}$  is  $R = (1 - \delta)^m / m!$ . Note that the rate of such codes are always bounded by  $1/m!$ . By contrast, GAP codes can have rate arbitrarily close to 1, while having constant relative distance.

This construction can be instantiated using any explicit collection of hyperplanes in general position<sup>8</sup>. For example, here is a classical construction of hyperplanes in general position coming from Vandermonde matrices, which works over any sufficiently large field.

**Observation 3.7** (The Vandermonde hyperplane family in general position, see eg. Bläser & Pandey [BP20]). Let  $t, m$  be natural numbers with  $t \geq m$  and let  $\mathbb{F}$  be a field of size at least  $t$ . Let  $\alpha_1, \alpha_2, \dots, \alpha_t$  be  $t$  distinct elements from the field  $\mathbb{F}$ .

---

<sup>8</sup>which are equivalent to MDS codes.

For  $i \in [t]$ , let  $H_i \subseteq \mathbb{F}^m$  be the hyperplane given by:

$$H_i := \{(b_1, \dots, b_m) \in \mathbb{F}^m : L_{\alpha_i}(b_1, \dots, b_m) = 0\},$$

where for  $\alpha \in \mathbb{F}$ ,  $L_\alpha(X_1, \dots, X_m)$  is the linear function given by:

$$L_\alpha(\mathbf{X}) := \alpha^m - \alpha^{m-1}X_1 + \alpha^{m-2}X_2 + \dots + (-1)^{m-1}\alpha X_{m-1} + (-1)^m X_m.$$

Then, the hyperplanes  $H_1, H_2, \dots, H_t$  are in general position in  $\mathbb{F}^m$ .

The nonstandard choice of signs in the above construction is to enable the following elegant description of the evaluation points of the associated GAP code. Viewing  $\mathbb{F}^m$  as the space of monic univariate polynomials of degree  $m$  (by associating the point  $(b_1, \dots, b_m)$  with the polynomial  $T^m - b_1 T^{m-1} + \dots + (-1)^m b_m \in \mathbb{F}[T]$ ), we get that the hyperplane  $L_\alpha$  is simply the set of monic polynomials that vanish at  $\alpha$ . Then the intersection of  $m$  such hyperplanes  $\{L_\alpha : \alpha \in J\}$  (where  $|J| = m$ ) is simply the unique polynomial of degree  $m$  whose set of roots equals  $J$  – its coefficients are thus the (signed) elementary symmetric polynomials evaluated at the elements of  $J$ . Instantiating this, we get the following very clean description of the evaluation domain for a GAP code.

**Definition 3.8** (Vandermonde GAP code evaluation domain). *Let  $A \subseteq \mathbb{F}$  be a set of size  $t$ . Define the evaluation domain:*

$$S_{A,m} = \{(e_1(\mathbf{a}), e_2(\mathbf{a}), \dots, e_m(\mathbf{a})) \mid \mathbf{a} = (a_1, \dots, a_m), \text{ the } a_i \text{ are pairwise distinct elements of } A\}.$$

◇

Note that  $|S_{A,m}| = \binom{t}{m}$ . By the previous discussion, evaluating polynomials of degree  $d$  on  $S_{A,m}$  yields a code of distance  $\binom{t-d}{m}$ .

### 3.2 A Construction Based on Simplices

We now give another construction of a polynomial evaluation code with parameters matching those of the construction in [Theorem 3.5](#). The overall framework of analysis will also remain the same and rely on the use of [Lemma 3.1](#). The main difference is that we instantiate this framework using an alternative construction of hitting sets for polynomials based on taking the hitting set to be all points in an appropriate simplex. Let

$$\Lambda_{m,D} = \{(x_1, \dots, x_m) \in \mathbb{N}^m : x_i \in \mathbb{N} \text{ and } x_1 + x_2 + \dots + x_m < D\}$$

be the  $m$ -dimensional simplex of side length  $D$ . This construction of hitting sets is also described in the work of Bläser & Pandey [[BP20](#)]. We start with a statement of their result.

**Theorem 3.9** (Bläser & Pandey [BP20]). *Let  $m, D$  be natural numbers and  $\mathbb{F}$  be a field of characteristic greater than  $D$  or zero. Then,  $\Lambda_{m,D+1}$  is of size  $\binom{m+D}{m}$  and is a hitting set for  $m$ -variate polynomials of degree  $D$  over the field  $\mathbb{F}$ .*

Since the characteristic of  $\mathbb{F}$  is either zero or larger than  $D$ , we can view the set  $\{0, 1, \dots, D\}$  as a subset of  $\mathbb{F}$  in a natural sense. Thus, the set  $\Lambda_{m,D+1}$  as constructed above is indeed a subset of  $\mathbb{F}^m$ .

Instantiating this hitting/interpolating set with Lemma 3.1, we immediately get another family of multivariate polynomial evaluation codes with high rate. In fact, investigating the combination of the proof of Theorem 3.9 from [BP20] and Lemma 3.1 together, we develop a new “Schwartz-Zippel lemma theory”, which is able to analyze the distance of polynomial codes evaluated over a wide family of combinatorial shapes – simultaneously generalizing the original Schwartz-Zippel Lemma for general product sets and the distance bound that we just saw for the above rigid  $\{0, 1, \dots, D\}$  simplex. See the statement in subsubsection 1.1.1.

For now, we will simply state a special case: combinatorial simplices.

**Definition 3.10** (Combinatorial Simplex). *Let  $A \subseteq \mathbb{F}$  with  $|A| = \ell$ . Let  $A = \{a_0, a_1, \dots, a_{\ell-1}\}$  be an arbitrary ordering of  $A$ . Define:*

$$\Lambda_{m,A} = \{(a_{x_1}, a_{x_2}, \dots, a_{x_m}) \mid x_i \in \mathbb{N} \text{ with } \sum_i x_i < \ell\}.$$

◇

In Section 4 we show that combinatorial simplices are good evaluation domains for polynomials. We refer to the codes obtained from combinatorial simplices as Combinatorial Arrays for Polynomials (CAP) codes.

**Definition 3.11** (CAP). *Define  $\text{CAP}_{m,d,\ell} = E_{m,d,S}$ , where  $S = \Lambda_{m,A}$  for some set  $A$  of size  $\ell$ . That is,  $\text{CAP}_{m,d,\ell}$  is the set of evaluations of  $m$ -variate polynomials of degree at most  $d$  on the combinatorial simplex of  $A$ , where  $|A| = \ell$ .*

◇

We again have that these codes have good rate and distance, with rate able to approach 1 when  $m = O(1)$ .

**Theorem 3.12.** *Let  $m, d, \ell \in \mathbb{N}$  with  $\ell > d$ . Then  $\text{CAP}_{m,d,\ell}$  has distance at least  $\binom{\ell-d+m-1}{m}$ . Furthermore, for any  $\varepsilon > 0$ , if  $\ell = d + \varepsilon d$ , then rate of  $\text{CAP}_{m,d,\ell}$  is least  $\left(\frac{1}{1+\varepsilon}\right)^m$  and relative distance is at least  $\left(\frac{\varepsilon}{1+\varepsilon}\right)^m$ .*

The proof of the distance and the calculations of rate and relative distance are the same as that of the proof of Theorem 3.5.

## 4 Zero Patterns of Polynomials

In this section we develop an alternate approach to the distance of CAP codes by understanding the zero patterns of polynomials. Though this method requires a bit more effort, the technique we develop leads to a generalized Schwartz-Zippel lemma, and is able to prove the distance properties of a wider variety of evaluation sets. It will also be the foundation for our efficient decoding algorithms for CAP codes.

As a particular application of this more general theory, we give an example in [Section 4.5](#) of an evaluation set in 2-dimensions obtaining better rate versus distance trade-offs than both CAP/GAP codes and traditional RM codes for some settings of rate and distance.

Since polynomial evaluation codes are linear, finding the distance of a polynomial evaluation code is equivalent to finding lower bounds on the minimum number of non-zeros in  $S$  of any non-zero polynomial of total degree  $d$ . Alternatively, one can find upper bounds on the number of zeros in a set  $S$ . In the univariate case, the degree mantra states that univariate polynomials of degree at most  $d$  have at most  $d$  zeros. When the evaluation set is the grid, we have the Schwartz-Zippel lemma.

Thus, in order to prove the distance properties of polynomial evaluation codes on more general sets, we will study the zero patterns of polynomials and prove a generalization of the Schwartz-Zippel lemma. We state the The Schwartz-Zippel Lemma below for reference.

**Lemma 4.1** (Schwartz-Zippel). *Let  $f \in \mathbb{F}[X_1, \dots, X_m]$  be any non-zero polynomial of total degree at most  $d$ , let  $S \subset \mathbb{F}$ . Then  $|Z(f) \cap S^m| \leq d|S|^{m-1}$*

Throughout this section, we identify  $\mathbb{F} = \mathbb{F}_q$  and  $\mathbb{N}$  in the following way. Suppose  $\{\alpha_0, \dots, \alpha_{q-1}\}$  are the elements of  $\mathbb{F}$ , think of  $k \in \mathbb{N}$  as  $\alpha_k$ . Note that the mapping of  $\alpha_i$  to elements of  $\mathbb{F}$  is irrelevant since we will only be working on the indices. We will always assume  $\mathbb{F}$  is large enough so that when we write  $\mathbb{N}_{<\ell}$ , there are indeed at least  $\ell$  distinct elements in  $\mathbb{F}$ . Furthermore, we will write  $\mathbb{N} = \mathbb{N}_{<q} = \mathbb{F}$ .

### 4.1 The Bivariate Case

We start by considering the bivariate case. The standard proof of the Schwartz-Zippel Lemma proceeds by induction on  $m$ . We will present the first inductive step (proving the bivariate case from the univariate case), and observe that it not only gives bounds on the number of zeros but also provides restrictions on their locations in the grid.

Let  $d_Y$  be the  $Y$  degree of  $f$  (i.e., the maximum power of  $Y$  in any monomial with non-zero coefficient). Then we can write  $f(X, Y) = \sum_{i=0}^{d_Y} f_i(X)Y^i$ , where  $f_i$  is a univariate polynomial in  $X$  of degree at most  $d - i$ . Then, consider the number of zeros in each column. I.e., for each  $x$ , how many zeros are there in  $\{(x, y) : y \in S\}$ ? Let  $x \in S$ , there are two cases.

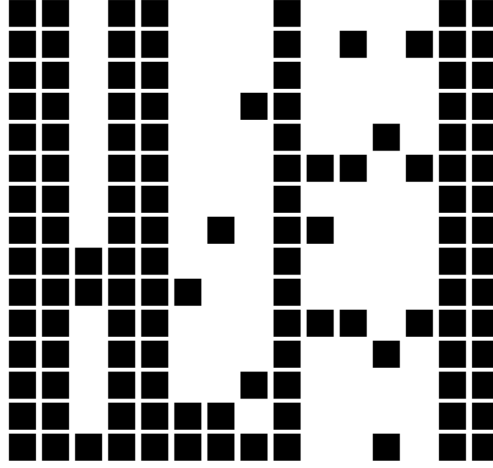


Figure 1: Possible locations of zeros of a bivariate polynomial,  $f(X, Y)$  of degree 10,  $Y$ -degree 3, evaluated on  $\mathbb{N}_{<15} \times \mathbb{N}_{<15}$ . Zeros are denoted with black squares.

1. If  $f_{d_Y}(x) \neq 0$ , then the univariate polynomials  $f(x, Y)$  is non-zero (since the coefficient for  $Y^{d_Y}$  is non-zero). Furthermore,  $f(x, Y)$  has degree at most  $d_Y$ , so there are at most  $d_Y$  zeros in the  $x$ th column.
2.  $f_{d_Y}(x) = 0$ , then  $f(x, Y)$  may be identically zero, so the entire  $x$ th with the column could be zero.

The proof of the Schwartz-Zippel Lemma then proceeds to bound the number of zeros as follows. Since  $f_{d_Y}$  has degree at most  $d - d_Y$ , case 2 happens for at most  $d - d_Y$  values of  $x$ , and thus the total number of zeros is at most

$$|S|(d - d_Y) + d_Y(|S| - d + d_Y) = d|S| + d_Y(d_Y - d) \leq d|S|.$$

Upon close inspection, the proof not only bounds the number of zeros but also proves that the zeros can only occur in a certain pattern. In particular, there exists  $a \in \{0, 1, 2, \dots, d\}$  (which is the  $Y$  degree of  $f$ ), such that at most  $d - a$  columns are entirely zero, and the rest of the column have at most  $a$  zeros. Figure 1 shows an example.

To demonstrate the power of this structure, consider the following set.

**Example 4.2** (Triangular Evaluation Set).  $T = \Lambda_{2, \frac{3}{2}d} = \{(x, y) \in \mathbb{N} \times \mathbb{N} : x + y < \frac{3}{2}d\}$ .  $\diamond$

Let  $f \in \mathbb{F}[X, Y]$  be a polynomial of total degree at most  $d$ , how many points in  $T$  can  $f$  vanish? Note that  $T$  is contained in the grid  $(\mathbb{N}_{<\frac{3}{2}d})^2$ . Applying the Schwartz-Zippel Lemma directly to this grid, we get that there are at most  $\frac{3}{2}d^2$  zeros. By a stars and bars argument,  $|T| = \binom{\frac{3}{2}d}{2} \approx \frac{9}{8}d^2 \leq$



Figure 2: A possible zero pattern (same as Figure 1) overlaid with the triangle evaluation set from Example 4.2 in orange.

$\frac{3}{2}d^2$ . Thus, the Schwartz-Zippel lemma does not rule out the possibility that  $f$  vanishes entirely on  $T$ .

Now, let's consider the structure of the zeros. From the previous discussion, we know that there exists some  $a \in \{0, 1, \dots, d\}$  such that  $f$  vanishes on at most  $d - a$  of the columns, and has at most  $a$  zeros on the remaining columns. Let  $C \subset S$  be the set of columns on which  $f$  vanishes. Then the smallest value of  $x$  for which the  $x$ th column does not vanish is at most  $d - a$  (in the case that  $C = \{0, 1, \dots, d - a - 1\}$ ). However, this column has at least  $\frac{3}{2}d - (d - a) = d/2 + a$  points in  $T$ . Since  $f$  can vanish on at most  $a$  of these points,  $f$  must be non-zero on at least  $d/2$  points in this column. Extending this argument to the remaining columns,  $f$  must be non-zero on at least  $(d/2) + (d/2 - 1) + (d/2 - 2) + \dots + 1 \approx d^2/8$  points in  $T$ . This example is illustrated in Figure 2.

Slightly more generally, if we fix any  $\varepsilon > 0$ , and take  $T = \{(x, y) \in S : x + y \leq (1 + \varepsilon)d\}$ , we have that any degree  $d$  polynomial is non-zero on at least  $\varepsilon^2 d^2/2$  points. As a polynomial evaluation code,  $E_{2,d,T}$  has rate  $\frac{1}{(1+\varepsilon)^2}$ , and relative distance  $\frac{\varepsilon^2}{(1+\varepsilon)^2}$ . Thus, taking  $\varepsilon$  to be small, we can get codes with rate arbitrarily close to 1, with constant distance. Notice that this recovers Theorem 3.12 for  $m = 2$ .

In the next section, we formalize the idea of zero patterns and generalize it to  $m$  variables.

## 4.2 The General Case

**Slicing.** To facilitate working over high-dimensional grids, we will introduce some notation for slices. Let  $S \subset \mathbb{N}^m$ . For  $\mathbf{x} \in \mathbb{N}^{m-1}$ , and  $i \in J$ , define  $S_{\mathbf{x},i} = \{a \in \mathbb{N} : (x_1, \dots, x_{i-1}, a, x_i, \dots, x_{m-1}) \in S\}$ ,

and call  $S_{\mathbf{x},i}$  the slice of  $S$  at  $\mathbf{x}$  in the  $i$  direction. Note that to define a subset of  $\mathbb{N}^m$ , it suffices to specify all the slices in the  $i$  direction for some  $i \in [m]$ .

We define a set of zero patterns,  $\mathcal{Z}_{m,d}$ .

**Definition 4.3** ( $\mathcal{Z}_{m,d}$ ).  $\mathcal{Z}_{m,d}$  is a subset of  $\wp(\mathbb{N}^m)$  and is defined recursively as follows.  $\mathcal{Z}_{0,d} = \{\emptyset\}$ , and  $\forall E \subset \mathbb{N}^m$ ,  $E \in \mathcal{Z}_{m,d}$  if and only if there exists  $a \in \{0, 1, \dots, d\}$ , and  $D \in \mathcal{Z}_{m-1,d-a}$  such that for all  $\mathbf{x} \in \mathbb{N}^{d-1}$ , if  $\mathbf{x} \in D$ , then  $E_{\mathbf{x},m} = \mathbb{N}$ , and if  $\mathbf{x} \notin D$ ,  $|E_{\mathbf{x},m}| = a$ <sup>1</sup>.  $\diamond$

The following lemma shows that for every non-zero polynomial  $f$  on  $m$  variables with total degree at most  $d$ , the zeros of  $f$  are contained in some set in  $\mathcal{Z}_{m,d}$ .

**Lemma 4.4** (Zero Patterns). *Let  $f$  be a non-zero polynomial on  $m$  variables with total degree at most  $d$ . Let  $Z(f)$  be the set of zeros of  $f$  on  $\mathbb{N}^m$ . Then there is some  $E \in \mathcal{Z}_{m,d}$  such that  $Z(f) \subset E$ .*

*Proof.* By induction on  $m$ .

**Base case.** For  $m = 0$ ,  $f$  is a polynomial on zero variables and thus a constant. Since  $f$  is non-zero, it has no zeros; hence,  $Z(f) = \emptyset$ .

**Inductive step.** Assume the claim is true for  $m - 1$ . Let  $f \in \mathbb{F}[X_1, \dots, X_m]$  be a non-zero polynomial with total degree at most  $d$ . Let  $d_m \in \{0, 1, \dots, d\}$  be the  $X_m$  degree of  $f$ , and write  $f = \sum_{i=0}^{d_m} f_i(X_1, \dots, X_{m-1})X_m^i$ . Note that  $f_{d_m}$  is a non-zero polynomial on  $m - 1$  variables of total degree at most  $d - d_m$ . Thus, the induction hypothesis applies, and there exists some  $D \in \mathcal{Z}_{m-1,d-d_m}$  such that  $Z(f_{d_m}) \subset D$ .

We'll construct  $E$  by its  $m$  directional slices as follows. For  $\mathbf{x} \in D$ , we take everything - i.e.,  $E_{\mathbf{x},m} = \mathbb{N}$ . For  $\mathbf{x} \notin D$ ,  $f(\mathbf{x}, X_m)$  is a non-zero univariate polynomial of degree  $d_m$ , and thus has at most  $d_m$  zeros. Thus,  $|Z(f)_{\mathbf{x},m}| \leq d_m$ . So for each  $\mathbf{x} \notin D$ , we set  $E_{\mathbf{x},m}$  to some subset containing  $Z(f)_{\mathbf{x},m}$  of size exactly  $d_m$ .

Then,  $E \in \mathcal{Z}_{m,d}$  and  $E$  contains  $Z(f)$ , so we're done.  $\square$

### 4.3 Shifting

To get a better handle on  $\mathcal{Z}_{m,d}$ , we will define a shifting operation and find that, after shifting, elements of  $\mathcal{Z}_{m,d}$  are very easy to describe.

Let's again think back to the bivariate case. Suppose  $f \in \mathbb{F}[X, Y]$  is any polynomial of total degree at most  $d$  and  $Y$  degree  $d_Y$ . We earlier found that there are at most  $d - d_Y$  columns where  $f$  was completely zero, and the remaining columns had at most  $d_Y$  zeros each. Let  $C \subset \mathbb{N}$  be the subset of entirely zero columns. Consider an operation that first pushes zeros in each column as far down as possible. This has no effect for columns that were already completely zeros, but now, for each other column, the zeros are now contained within the first  $d_Y$  rows. Next, for each row, push all the zeros as far left as possible. Since there are now at most  $d - d_Y$  columns that have

<sup>1</sup>Note that we could have had replaced this condition with  $|E_{\mathbf{x},m}| \leq a$ , but having  $|E_{\mathbf{x},m}|$  exactly equal to  $a$  will make some later proofs slightly cleaner.

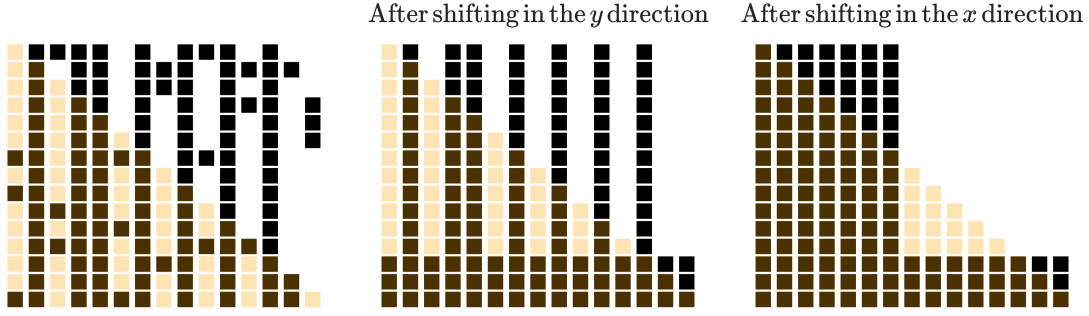


Figure 3: Example of shifting the zeros from Figure 2

non-zero values above the  $d_Y$ th row, we have that all of the zeros are contained in the first  $d_Y$  rows and  $d - d_Y$  columns. This shifting procedure is shown in Figure 3.

We now formally define the  $i$ th shifting operator,  $\sigma_i(Z)$ , that pushes points down in the  $i$  direction. Define  $\sigma_i(Z)$  by its slices in the  $i$  direction. For any  $\mathbf{x} \in \mathbb{N}^{m-1}$ ,

$$(\sigma_i(Z))_{\mathbf{x},i} = \mathbb{N}_{<|Z_{\mathbf{x},i}|}.$$

That is, we replace every slice in the  $i$  direction with an initial segment of the same size as the original slice.

We now prove some basic facts about  $\sigma_i$ .

**Lemma 4.5** (Properties of  $\sigma_i$ ).

- (I)  $\sigma_i(Z)$  can be obtained by decreasing the  $i$  coordinate of zero or more elements of  $Z$ .
- (II) For any  $i \in [m]$ .  $\sigma_i(\mathbb{N}^m) = \mathbb{N}^m$
- (III) Let  $i, j \in [m]$  with  $i \neq j$ , and  $A, B \subset \mathbb{N}^m$ , and suppose  $A$  and  $B$  are disjoint in a coordinate  $i$ , that is  $\{a_i : \mathbf{a} \in A\} \cap \{b_i : \mathbf{b} \in B\} = \emptyset$ , then if  $j \neq i$ ,  $\sigma_j(A \cup B) = \sigma_j(A) \cup \sigma_j(B)$ .
- (IV) Let  $j \in [m-1]$ , and  $A \subset \mathbb{N}^{m-1}$  and  $C \subset \mathbb{N}$ . Then  $\sigma_j(A \times C) = \sigma_j(A) \times C$ .

*Proof.*

- (I) For any  $\mathbf{x} \in \mathbb{N}^{m-1}$ ,  $(\sigma_i(Z))_{\mathbf{x},i}$  is the initial segment of  $\mathbb{N}$  of size  $|Z_{\mathbf{x},i}|$ . Let  $z_0, z_1, \dots$  be the elements of  $Z_{\mathbf{x},i}$  in ascending order. Then  $f : Z_{\mathbf{x},i} \rightarrow \sigma_i(Z)_{\mathbf{x},i}$  defined by  $f(z_j) = j$  is always such that  $f(x) \leq x$ .
- (II) For every  $\mathbf{x} \in \mathbb{N}^{m-1}$ ,  $(\mathbb{N}^m)_{\mathbf{x},i} = \mathbb{N}$ . Thus,  $(\sigma_i(\mathbb{N}^m))_{\mathbf{x},i} = \mathbb{N}$ , and  $\sigma_i(\mathbb{N}^m) = \mathbb{N}^m$ .
- (III) Let  $\mathbf{x} \in \mathbb{N}^{m-1}$ . Since  $A$  and  $B$  are disjoint in some coordinate  $i$ , at most one of  $A_{\mathbf{x},j}$  and  $B_{\mathbf{x},j}$  can be non-empty (otherwise, there would be elements in  $A$  and  $B$  that both have  $i$  coordinate  $x_i$ ). Thus,  $A_{\mathbf{x},j} \cup B_{\mathbf{x},j}$  is either  $A_{\mathbf{x},j}$  or  $B_{\mathbf{x},j}$ , and consequently,  $|A_{\mathbf{x},j} \cup B_{\mathbf{x},j}| =$



$\max(|A_{\mathbf{x},j}|, |B_{\mathbf{x},j}|)$ . We now show  $\sigma_j(A \cup B)$  and  $\sigma_j(A) \cup \sigma_j(B)$  are equal by comparing their  $j$ -directional slices.

$$\begin{aligned} (\sigma_j(A) \cup \sigma_j(B))_{\mathbf{x},j} &= \mathbb{N}_{<|A_{\mathbf{x},j}|} \cup \mathbb{N}_{<|B_{\mathbf{x},j}|} \\ &= \mathbb{N}_{<\max(|A_{\mathbf{x},j}|, |B_{\mathbf{x},j}|)} \\ &= \mathbb{N}_{<|A_{\mathbf{x},j} \cup B_{\mathbf{x},j}|} \\ &= (\sigma_j(A \cup B))_{\mathbf{x},j} \end{aligned}$$

(IV) Let  $c \in \mathbb{N}$ , and  $\mathbf{x} \in \mathbb{N}^{m-2}$ . Note that the slice at  $\mathbf{x}$  in the  $j$ th direction for  $A$  is the same as the slice at  $[x_1, \dots, x_{m-2}, c]$  in the  $j$ th direction for  $A \times \{c\}$  since all we changed was to set the last coordinate to  $c$  for each element of  $A$ . Thus,  $\sigma_j(A \times \{c\}) = \sigma_j(A) \times \{c\}$ . Then note that  $A \times C = \bigcup_{c \in C} A \times \{c\}$ . Each of these sets in the union is disjoint last coordinate, so by [Lemma 4.5\(III\)](#),

$$\sigma_j(A \times C) = \bigcup_{c \in C} \sigma_j(A \times \{c\}) = \bigcup_{c \in C} \sigma_j(A) \times \{c\} = \sigma_j(A) \times C$$

□

Let  $\mathbf{d} \in \mathbb{N}^m$ , and define  $L(\mathbf{d}) = \{\mathbf{x} \in \mathbb{N}^m : x_1 < d_1 \vee x_2 < d_2 \vee \dots \vee x_m < d_m\}$ . We use  $L$  to denote this set since, in the two-dimensional case, it has an “L” shape. We now claim that every element of  $\mathcal{Z}_{m,d}$  looks like  $L(\mathbf{d})$  for some  $\mathbf{d}$  after shifting in every coordinate.

**Lemma 4.6.** *Let  $\sigma_{1,\dots,m} = \sigma_1 \circ \sigma_2 \circ \dots \circ \sigma_m$ . For every  $E \in \mathcal{Z}_{m,d}$ ,  $\sigma_{1,\dots,m}(E) = L(\mathbf{d})$  for some  $\mathbf{d} \in \{0, 1, \dots, d\}^m$  with  $d_1 + d_2 + \dots + d_m = d$ .*

*Proof.* By induction on  $m$ . To make the base case work nicely, define the empty concatenation of operators to be the identity, and define  $L(\emptyset) = \emptyset$

**Base case.** For  $m = 0$ ,  $E = \emptyset = L(\emptyset)$

**Inductive Step.** Suppose the claim is true for  $m - 1$ , and let  $E \in \mathcal{Z}_{m,d}$ . Then by the recursive definition of  $\mathcal{Z}$ , there exists some  $d_m \leq d$ , and  $D \in \mathcal{Z}_{m-1,d-d_m}$  such that  $E_{\mathbf{x},m} = \mathbb{N}$  for  $\mathbf{x} \in D$ , and  $|E_{\mathbf{x},m}| = d_m$  for  $\mathbf{x} \notin D$ . Then,

$$\sigma_m(E) = D \times \mathbb{N} \cup \overline{D} \times \mathbb{N}_{<d_m}$$

Since  $D$  and  $\overline{D}$  partition  $\mathbb{N}^{m-1}$ , we have that every  $\mathbf{x} \in \mathbb{N}^m$  with  $x_m < d_m$  is in  $\sigma_m(E)$ . Thus, we

can rewrite the set as  $D \times \mathbb{N}_{\geq d_m} \cup \mathbb{N}^{m-1} \times \mathbb{N}_{< d_m}$ . Plugging this in, we get

$$\begin{aligned}
\sigma_{1,\dots,m}(E) &= \sigma_{1,\dots,m-1}(D \times \mathbb{N}_{\geq d_m} \cup \mathbb{N}^{m-1} \times \mathbb{N}_{< d_m}) \\
&= \sigma_{1,\dots,m-1}(D \times \mathbb{N}_{\geq d_m}) \cup \sigma_{1,\dots,m-1}(\mathbb{N}^{m-1} \times \mathbb{N}_{< d_m}) \\
&= \sigma_{1,\dots,m-1}(D) \times \mathbb{N}_{\geq d_m} \cup \sigma_{1,\dots,m-1}(\mathbb{N}^{m-1}) \times \mathbb{N}_{< d_m} \\
&= \sigma_{1,\dots,m-1}(D) \times \mathbb{N}_{\geq d_m} \cup \mathbb{N}^{m-1} \times \mathbb{N}_{< d_m}.
\end{aligned}$$

The second line follows from the fact that  $D \times \mathbb{N}_{\geq d_m}$ , and  $\mathbb{N}^{m-1} \times \mathbb{N}_{< d_m}$  are disjoint in the  $m$  coordinate, and  $\sigma_i$  for  $i < m$  does not change the  $m$  coordinate, and applying [Lemma 4.5\(III\)](#) repeatedly. The third line follows from [Lemma 4.5\(IV\)](#), and the fourth from [Lemma 4.5\(II\)](#).

From the inductive hypothesis, there exists  $\mathbf{d} \in \{0, 1, \dots, d\}^{m-1}$  such that  $d_1 + d_2 + \dots + d_{m-1} = d - d_m$ , and  $\sigma_{1,\dots,m-1}(D) = L(d_1, \dots, d_{m-1})$ . Substituting this back, we finally get

$$\sigma_{1,\dots,m}(E) = L(d_1, \dots, d_{m-1}) \times \mathbb{N}_{\geq d_m} \cup \mathbb{N}^{m-1} \times \mathbb{N}_{< d_m} = L(d_1, \dots, d_m).$$

Also note that  $\sum_{i=1}^m d_i = d_m + \sum_{i=1}^{m-1} d_i = d_m + d - d_m = d$ , and this completes the proof.  $\square$

#### 4.4 d-robustness

Since we want to prove upper bounds on the number of zeros in  $S$ , we require that  $S$  satisfies some property to ensure that the shifting operation does not move zeros out of the set. Shifting can be viewed as decreasing the values of certain coordinates ([Lemma 4.5\(I\)](#)), so a very natural definition is the following.

**Definition 4.7** (Downward Closed).  $S \subset \mathbb{N}^m$  is downward closed if for all  $\mathbf{x}, \mathbf{y} \in \mathbb{N}^m$ , with  $x_1 \leq y_1, x_2 \leq y_2, \dots, x_m \leq y_m$ , if  $\mathbf{y} \in S$ , then  $\mathbf{x} \in S$ . In other words, if  $\mathbf{x}$  is at most  $\mathbf{y}$  in every coordinate, and  $\mathbf{y} \in S$ , then  $\mathbf{x} \in S$ .  $\diamond$

Given, this definition, we see that shifting cannot decrease the size of the intersection with a downward closed set.

**Lemma 4.8.** Let  $S, Z \subset \mathbb{N}^m$  such that  $S$  is downward closed. Then for any  $i \in [m]$ ,

$$|S \cap Z| \leq |S \cap \sigma_i(Z)|$$

Our next lemma bounds the number of zeros of an  $m$ -variate polynomial can have on a downward closed set.

**Lemma 4.9.** Let  $f \in \mathbb{F}[X_1, \dots, X_m]$  be of total degree  $d$ , and  $S$  be a downward closed set. Then there exists  $\mathbf{d} \in \{0, 1, \dots, d\}^m$  such that  $\sum_{i=1}^m d_i = d$ , and

$$|Z(f) \cap S| \leq |L(\mathbf{d}) \cap S|$$

*Proof.* By Lemma 4.4,  $Z(f) \subset E$  for some  $E \in \mathcal{Z}_{m,d}$ . By Lemma 4.6, we have that there is some  $\mathbf{d} \in \{0, 1, \dots, d\}^m$ , with  $\sum_{i=1}^m d_i = d$ , such that  $\sigma_{1,\dots,m}(E) = L(\mathbf{d})$ . Then, we have

$$\begin{aligned} |Z(f) \cap S| &\leq |E \cap S| && (Z(f) \subset E) \\ &\leq |\sigma_{1,\dots,m}(E) \cap S| && (\text{Lemma 4.8}) \\ &= |L(\mathbf{d}) \cap S| \end{aligned}$$

□

We also state the corresponding bound for non-zeros, a direct result of the previous lemma. For  $\mathbf{d} \in \{0, 1, \dots, d\}^m$ , define  $H(\mathbf{d}) = \overline{L(\mathbf{d})}$ . I.e.,  $H(\mathbf{d})$  is the set of points  $\mathbf{x}$  in which every coordinate  $x_i$  is at least  $d_i$ .

**Lemma 4.10.** *Let  $f \in \mathbb{F}[X_1, \dots, X_m]$  be of total degree  $d$ , and  $S$  be a downward closed set. Then, there exists  $\mathbf{d} \in \{0, 1, \dots, d\}^m$  such that  $\sum_{i=1}^m d_i = d$ , and*

$$|N(f) \cap S| \geq |H(\mathbf{d}) \cap S|$$

Now, we define a combinatorial property of a set  $S$  that guarantees that any  $m$ -variate polynomial of total degree at most  $d$  will have at least some number of non-zeros on  $S$ .

**Definition 4.11** ( $d$ -robustness of a downward closed set). *Let  $S \subset \mathbb{N}^m$  be downward closed. Define the  $d$ -robustness of  $S$ , as follows.*

$$\Pi_d(S) = \min \left\{ |H(\mathbf{d}) \cap S| : \mathbf{d} \in \mathbb{N}^m, \sum_{i=1}^m d_i = d. \right\}$$

The relative  $d$ -robustness is then  $\pi_d(S) = \Pi_d(S) / |S|$ . ◇

In words,  $\Pi_d(S) \geq B$  if for any choice of  $\mathbf{d}$  such that  $\sum_{i=1}^m d_i = d$ , at least  $B$  elements  $\mathbf{x}$  in  $S$  are such that  $x_i \geq d_i$  for all  $i \in [m]$ .

Here are several useful facts about  $d$ -robustness

**Remark 4.12.** *For any downward closed  $S$ .  $\Pi_d(S) = |S| - \max \{|L(\mathbf{d}) \cap S| : \mathbf{d} \in \mathbb{N}^m, \sum_{i=1}^m d_i = d\}$ . Additionally, the following conditions are equivalent*

- $\pi_d(S) \geq \delta$
- For every  $\mathbf{d}$  with  $\sum_{i=1}^m d_i = d$ ,  $\Pr_{\mathbf{x} \sim S}[\mathbf{x} \in H(\mathbf{d})] \geq \delta$
- For every  $\mathbf{d}$  with  $\sum_{i=1}^m d_i = d$ ,  $\Pr_{\mathbf{x} \sim S}[\mathbf{x} \in L(\mathbf{d})] \leq 1 - \delta$

◇

Following the definition of  $d$ -robustness, and [Lemma 4.10](#), we have the following bounds on the number of zeros/non-zeros in sets with  $d$ -robustness  $\delta n$ .

**Lemma 4.13** (Zeros in  $d$  robust sets). *Let  $f \in \mathbb{F}[X_1, \dots, X_m]$  be a polynomial of total degree  $d$ . Let  $S$  be downward closed of size  $n$  such that  $\Pi_d(S) \geq \delta n$ . Then, we get the following equivalent results:*

1.  $|N(f) \cap S| \geq \delta n$ .
2.  $\Pr_{\mathbf{x} \sim S}[\mathbf{x} \in N(f)] \geq \delta$ .
3.  $|Z(f) \cap S| \leq (1 - \delta)|S|$ .
4.  $\Pr_{\mathbf{x} \sim S}[\mathbf{x} \in Z(f)] \leq 1 - \delta$ .

A direct consequence of [Lemma 4.13](#) (versions 1, 2) is that polynomial evaluation codes on downward closed sets have distance.

**Corollary 4.14** (Distance of polynomial evaluation codes). *Let  $m, d \in \mathbb{N}$  with  $m \geq 1$ , and let  $S \subset \mathbb{N}^m$ . If  $\pi_d(S) \geq \delta$ , then  $E_{m,d,S}$  has relative distance at least  $\delta$ .*

Thus, to compute the distance of a polynomial evaluation code where the evaluation set is downward closed, we just need to compute the  $d$ -robustness of the set. A direct consequence of [Lemma 4.13](#) (versions 3, 4) is a generalization of the Schwartz-Zippel Lemma.

**Corollary 4.15** (Schwartz-Zippel for downward closed sets). *Let  $m, d \in \mathbb{N}$  with  $m \geq 1$ , and let  $S \subset \mathbb{N}^m$ . Let  $f \in [X_1, \dots, X_m]$  be a polynomial of total degree  $d$ , and suppose  $\pi_d(S) \geq \delta$ . Then  $\Pr_{\mathbf{x} \sim S}[f(\mathbf{x}) = 0] \leq 1 - \delta$ , equivalently,  $|Z(f) \cap S| \leq (1 - \delta)|S|$ .*

To see that [Corollary 4.15](#) is indeed a generalization of the Schwartz-Zippel lemma, calculate  $d$ -robustness of the grid of side length  $\ell$ ,  $(\mathbb{N}_{<\ell})^m$ . Let  $\mathbf{d} \in \{0, 1, \dots, d\}^m$  such that  $\sum_{i=1}^m d_i = d$ . We have

$$\Pr[\mathbf{x} \in L(\mathbf{d})] = \Pr[\exists i \in [m]. (x_i < d_i)] \leq \sum_{i=1}^m \Pr[x_i < d_i] = \sum_{i=1}^m d_i / \ell = d / \ell.$$

Thus,  $(\mathbb{N}_{<\ell})^d$  has relative  $d$ -robustness  $1 - d/\ell$ , and we have  $\Pr[f(x) = 0] \leq d/\ell$ , which is exactly what is given by Schwartz-Zippel.

As an alternate proof that CAP codes have distance ([Theorem 3.12](#)), we can calculate the  $d$ -robustness of simplex,  $\Lambda_{m,\ell}$ . Let  $\mathbf{d} \in \{0, 1, \dots, d-1\}^m$  be such that  $\sum_{i=1}^m d_i = d$ . Then

$$\Lambda_{m,\ell} \cap H(\mathbf{d}) = \{\mathbf{x} \in \mathbb{N}^m : x_i \geq d_i, \sum_{i=1}^m x_i < \ell\}$$

Then,  $f(\mathbf{x}) = \mathbf{x} - \mathbf{d}$  is a bijection from  $\Lambda_{m,\ell} \cap H(\mathbf{d})$  to  $\Lambda_{m,\ell-d}$ . Thus, the  $d$ -robustness of  $\Lambda_{m,\ell}$  is

$$\binom{m + \ell - d - 1}{m}.$$

Combining this with [Lemma 4.13](#), we get that  $\text{CAP}_{m,d,\ell}$  has distance at least  $\binom{m+\ell-d-1}{m}$ , recovering [Theorem 3.12](#). Note that we get the same result for  $\Lambda_{m,A}$ , where  $A \subset \mathbb{F}$  has size  $\ell$ , by identifying the elements of  $A$  with  $\{0, 1, \dots, \ell - 1\}$  using any ordering of the elements in  $A$ .

## 4.5 The step evaluation set

We can also use [Lemma 4.13](#) to prove that other evaluation sets provide codes with good distance. For example, in this section, we show an evaluation set beating both the grid and the simplex for certain values of  $\delta$ . The evaluation set is  $\text{Step}(\ell) = \{(x, y) \in \mathbb{N}_{<\ell} \times \mathbb{N}_{<\ell} : x < \ell/2 \vee y < \ell/2\}$ .

**Claim 4.16** (Step Evaluation Set).  $E_{2,d,\text{Step}(\ell)}$  has relative distance  $\frac{2}{3}(1 - d/\ell)$ , and rate at least  $\frac{2d^2}{3\ell^2}$ .

*Proof.* First we calculate the rate. Note that  $|\text{Step}(\ell)| = \frac{3}{4}\ell^2$ , so the rate of the code is  $\frac{\binom{d+2}{2}}{\frac{3}{4}\ell^2} \geq \frac{2d^2}{3\ell^2}$ . To calculate the distance, we find the  $d$ -robustness of  $\text{Step}(\ell)$ , and apply [Corollary 4.14](#). To find the  $\pi_d(\text{Step}(\ell))$ , we study the size of  $\text{Step}(\ell) \cap L(a, d - a)$  for each  $a \in \{0, 1, \dots, d\}$ . WLOG, suppose  $a \geq d - a$ . Consider two cases.

- $a \leq \ell/2$ . Since  $a \geq d - a$ , we also have  $d - \ell/2 \leq a$ . Then,  $|\text{Step}(\ell) \cap L(a, d - a)| = a\ell + (d - a)\ell - a(d - a)$ . Since this is a quadratic expression in  $a$  with a positive coefficient, for  $a \in [d - \ell/2, \ell/2]$ , this function is maximized at  $a = \ell/2$  (or equally  $a = d - \ell/2$ ), and we have  $|\text{Step}(\ell) \cap L(a, d - a)| \leq d\ell/2 + \ell^2/4$ .
- $a \geq \ell/2$ . In this case,  $|\text{Step}(\ell) \cap L(a, d - a)| = \ell a - (a - \ell/2)\ell/2 + (d - a)(\ell - a)$ . This function is another quadratic in  $a$  with a positive coefficient of  $a^2$ . Therefore, it is again maximized at the boundaries  $a = \ell/2$ , and  $a = d$ , and  $|\text{Step}(\ell) \cap L(a, d - a)| \leq d\ell/2 + \ell^2/4$ .

Rearranging, we find that  $\text{Step}(\ell)$  has relative  $d$ -robustness equal to  $\frac{2}{3}(1 - d/\ell)$ , as required.  $\square$

**Remark 4.17.** The rate distance trade-off of  $\delta = 2/3 - \sqrt{2R/3}$ .  $\diamond$

[Figure 4](#) compares the rate versus distance trade-offs for the simplex, grid, and step. In particular, we see that for  $\delta \in [0.078, 0.21]$ , the step construction beats both the simplex and the grid. <sup>9</sup>

The following is a natural question. For a fixed  $\delta$ , what is the smallest set  $S$  with relative  $d$ -robustness equal to  $\delta$ ? We explore this question in [Appendix A](#), where we prove several lower bounds in two dimensions. In particular, we show that for  $\delta > 1/2$ , the grid is optimal, and for  $\delta \rightarrow 0$ , the simplex has the right asymptotic behavior.

<sup>9</sup>The exact range for when the step construction beats the grid is  $\delta \in [\frac{1}{3}(3 - \sqrt{6} - \sqrt{5 - 2\sqrt{6}}), \frac{1}{2} - \frac{1}{6}\sqrt{3}]$ .

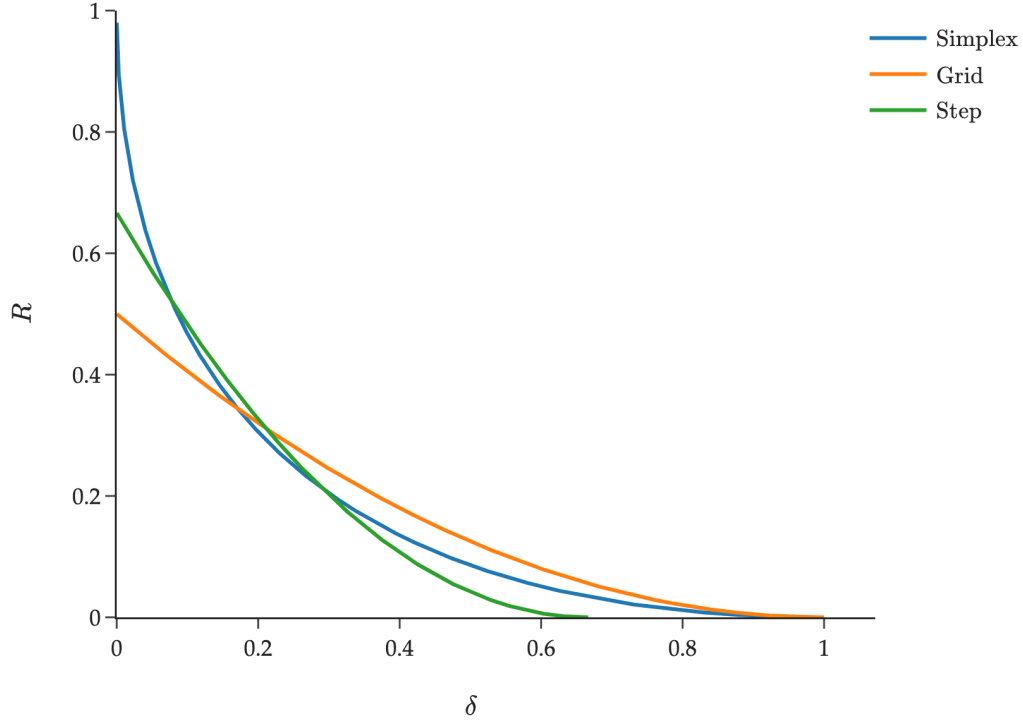


Figure 4: Comparing rate vs. distance trade-offs for the grid, simplex and step evaluation sets

## 5 Unique Decoding of CAP Codes

In this section, we give an efficient decoding algorithm for CAP codes. The algorithm is based on the decoding algorithm for Reed Muller codes given by Kim and Kopparty [KK17].

### 5.1 Uneven GMD

In this section we describe a variant of the GMD decoding algorithm of Forney [For66] that allows for inner codes of varying distances. This variant of GMD decoding, Algorithm 1, will be crucial in our decoding algorithm for the simplex evaluation set. As it turns out, the only change we need to make to the standard (deterministic) GMD algorithm is the setting of the weights. For a reference on the GMD algorithm, see section 14.3 of [GRS12].

**Theorem 5.1** (Uneven GMD Decoding). *Let  $C_{out}$  code with block length  $N$  and distance  $D$ . For  $i \in [N]$ , let  $C_i$  be a code with distance  $d_i$ . Let  $C = (C_1, \dots, C_N) \circ C_{out}$  be the concatenated code as in Definition 2.6. Then,  $C$  has minimum distance at least  $\min_{S \subseteq [N]: |S|=D} \sum_{i \in S} d_i$ , i.e. the sum of the  $D$  smallest inner distances.*

Furthermore, suppose there exist decoding algorithms  $A_{out}, A_1, A_2, \dots, A_N$  with corresponding time complexities  $T_{out}, T_1, \dots, T_N$ , such that  $A_{out}$  decodes  $C_{out}$  from error and erasure patterns of weight less than

---

**Algorithm 1:**  $\text{GMD}(r, A_{\text{out}}, A_1, \dots, A_N)$ 

---

```
1 Write  $r$  as  $N$  blocks  $r_1, \dots, r_N$  corresponding to  $N$  inner codewords
2 for  $i \in [N]$  do
3    $y_i = A_i(r_i)$ 
4    $s(i) \leftarrow$  number of erasures in  $r_i$ .
5    $x(i) \leftarrow$  number of non-erasure coordinates in which  $y_i$  and  $r_i$  differ
6    $w(i) \leftarrow d_i - 2x(i) - s(i)$ 
7 for  $i \in [N]$  do
8    $t \leftarrow w(i)$ 
9   for  $j \in [N]$  do
10    if  $w(j) < t$  then
11       $z_j \leftarrow ?$ 
12    else
13       $z_j \leftarrow r_j$ .
14   $c \leftarrow A_{\text{out}}(\mathbf{z})$ 
15  if  $\text{wt}(C(c^{(i)}) - r) < dD$  then
16    return  $c$ 
17 return  $\perp$ 
```

---

$D$ , and for each  $i \in [N]$ ,  $A_i$  decodes  $C_i$  from error (and erasure) patterns of weight less than  $d_i$ .

Then [Algorithm 1](#) instantiated with the  $A_{\text{out}}, A_1, \dots, A_N$  decodes the concatenated code  $C$  from error (and erasure) patterns of weight less than  $\min_{S \subseteq [N]: |S|=D} \sum_{i \in S} d_i$ . Furthermore, the time complexity of the algorithm is  $\text{poly}(T_{\text{out}}, T_1, \dots, T_N, N, n)$ .

*Proof.* The claim about the distance of the concatenated code is simple. WLOG suppose  $d_1 \leq d_2 \leq \dots \leq d_N$  so that  $\min_{S \subseteq [N]: |S|=D} \sum_{i \in S} d_i = \sum_{i=1}^D d_i$ . Let  $m, m'$  be distinct messages with  $m \neq m'$ . Let  $X \subseteq [N]$  be the set of indices where the encodings of  $m$  and  $m'$  differ in the outer code. Since the distance of the outer code is at least  $D$ ,  $|X| \geq D$ . For each  $i \in X$ , the inner encoding of  $i$ th symbol of the outer codeword must differ in at least  $d_i$  coordinates. Thus, the total number of coordinates in which the codewords differ is at least  $\sum_{i \in X} d_i$ , which is at least  $\sum_{i=1}^D d_i$ .

We now show the correctness of [Algorithm 1](#). The proof follows the combinatorial proof of Forney's algorithm given in [\[BHKS23\]](#). Let  $e$  be an error pattern of weight at most  $\sum_{i=1}^D d_i$ . Let  $E$  be the total number of errors and  $S$  be the total number of erasures. Recall  $\text{wt}(e) = 2E + S$ . Let  $e(i)$  be the number of errors in the  $i$ th block, and  $s(i)$  be the number of erasures in the  $i$ th block.

Consider the situation after decoding the inner codes. Let  $x(i)$  be the Hamming distance between the closest (inner) codeword and the received word in the  $i$ th block not counting erasures. Define  $w(i) = d_i - 2x(i) - s(i)$  to be the weight of each block. Intuitively, the  $w(i)$  captures the 'confidence' in the  $i$ th block. We'll now show that there exists a threshold,  $t$ , such that setting blocks with weight less than  $t$  to erasures allows the outer decoder to decode correctly.

Let  $A \subseteq [N]$  be the set of correctly decoded blocks. Let  $B \subseteq [N]$  be the set of incorrectly decoded blocks. Note that these are unknown to the algorithm and are purely for analysis. Let  $A_{\geq t} = \{i \in A : w(i) \geq t\}$ . Define  $B_{\geq t}$  similarly. The number of errors remaining after erasing blocks with weight  $< t$  is  $|B_{\geq t}|$ . The number of blocks we set to erasures is  $N - |B_{\geq t}| - |A_{\geq t}|$ . Thus, the outer decoder can decode so long as

$$2|B_{\geq t}| + N - |B_{\geq t}| - |A_{\geq t}| < D.$$

I.e.,

$$|B_{\geq t}| + N - D < |A_{\geq t}|.$$

**Claim 5.2.** *There exists a threshold  $t \in \{w(i) : i \in [N]\}$  such that the above holds.*

If the claim holds, we can decode by trying each weight as a threshold. The rest of this proof will show that the claim is true.

By contradiction, suppose the claim is false, i.e.

$$|B_{\geq t}| + N - D \geq |A_{\geq t}|, \tag{5.3}$$

for every choice in  $t \in \{w(i) : i \in [N]\}$ . The contradiction we will eventually obtain is that the weight of  $e$ , i.e.,  $2E + S$ , is too large.

As a first step, we bound the size of  $|B|$ . Since the inner codes are decodable as long as  $2e(i) + s(i) < d_i$ , we have

$$\sum_{i \in [B]} 2e(i) + s(i) \geq \sum_{i \in B} d_i \geq \sum_{i=1}^{|B|} d_i,$$

where the second inequality holds because  $d_1 \leq d_2 \leq \dots \leq d_N$ . Thus, since we have  $2E + S < \sum_{i=1}^D d_i$ ,  $|B| < D$ . In particular, let  $|B| = D - u$  for some positive integer  $u$ , and  $|A| = N - D + u$ .

Let  $a_1 \geq a_2 \geq \dots \geq a_{N-D+u}$  be an enumeration of  $A$ , and. Let  $b_1 \geq b_2 \geq \dots \geq b_{D-u}$  be an enumeration of  $B$ .

We show that for every  $r \in [u]$ ,  $w(a_{N-D+r}) \leq w(b_r)$ . By contradiction, suppose  $w(a_{N-D+r}) > w(b_r)$ . Then consider setting  $t = w(a_{N-D+r})$ , we have  $|A_{\geq t}| = N - D + r$ , and  $|B_{\geq t}| < r$ , which contradicts Equation (5.3).

We now use this relationship to bound the weight of the errors in blocks  $a_{N-D+r}$  and  $b_r$ . For simplicity, let's abbreviate  $a = a_{N-D+r}$ , and  $b = b_r$ . Substituting the definition of the weights, we



get

$$\begin{aligned} d_a - 2x(a) - s(a) &\leq d_b - 2x(b) - s(b) \iff \\ d_a - d_b &\leq 2x(a) + s(a) - 2x(b) - s(b) \end{aligned}$$

By the triangle inequality on block  $b$ , we have that  $x(b) \geq d_b - s(b) - e(b)$ . Additionally, since  $a \in A$ , the  $a$ th block was correctly decoded, thus  $x(a) = e(a)$ . Substituting these facts into the above and rearranging, we get

$$d_a + d_b \leq 2e(a) + s(a) + 2e(b) + s(b)$$

For the remaining blocks in  $i \in B$ , we know that the weight of the errors in block  $i$  is at least  $d_i$  since otherwise, the inner codeword would have been correctly decoded. Summing over the weights of the errors in each of these blocks, we have

$$\begin{aligned} 2E + S &\geq \sum_{r \in [u]} d_{a_{N-D+r}} + d_{b_r} + \sum_{i=u+1}^{D-u} 2e(b_i) + s(b_i) \\ &\geq \sum_{r \in [u]} d_{a_{N-D+r}} + d_{b_r} + \sum_{i=u+1}^{D-u} d_{b_i}. \end{aligned}$$

There are  $2u + D - 2u = D$  unique inner distances in the sum, which is at least the sum of the  $D$  smallest inner distances,  $\sum_{i=1}^D d_i$ , which is a contradiction since  $2E + S$  was supposed to be less than that quantity.  $\square$

**Remark 5.4.** *Note that if the inner code can only be decoded from errors, then the theorem is still true if there are only errors in the received word (and no erasures). However, since we are setting certain blocks to erasures for the outer decoding, the outer decoder must be able to correct errors and erasures even if there are no erasures in the received word.*  $\diamond$

## 5.2 Decoding in two-dimensions

Let  $f \in \mathbb{F}[X, Y]$  be any bivariate polynomial of total degree at most  $d$ . We can write it as

$$f(X, Y) = \sum_{i=0}^d c_i(X) Y^i,$$

where  $c_i$ , a univariate polynomial in  $X$ , is the coefficient of  $Y^i$ . Let  $G = \mathbb{N}_{<\ell} \times \mathbb{N}_{<\ell}$  be the two dimensional  $\ell \times \ell$  grid. View the grid as having  $\ell$  columns, each of size  $\ell$ . Let  $f : G \rightarrow \mathbb{F}$  be a codeword and  $e : G \rightarrow \mathbb{F}$  be an error pattern. The received word is  $r = f + e$ . Kim and Kopparty

combine the following observations to decode polynomial evaluation codes on the grid. Firstly, the  $x$ th column contains  $\ell$  evaluations of the univariate polynomial  $f(x, Y)$ , and thus, we can decode each column using standard RS decoding. Secondly, for any fixed  $i$  and  $x$ , the coefficient of  $Y^i$  in  $f(x, Y)$  is the evaluation of  $c_i$  at  $x$ . Thus, the decodings of each of the  $\ell$  columns combine to give  $\ell$  evaluations of  $c_i$ , and we can decode for  $c_i$  using standard RS decoding. Their algorithm is iterative, where the first iteration uses this procedure to find  $c_d$ , the second finds  $c_{d-1}$ , and so on.

We first present an alternate analysis. In particular, to retrieve the coefficient  $c_d$ , we view the codeword  $f$  as the encoding of  $c_d$  in some concatenated code and directly apply [Algorithm 1](#). The decoding algorithm is formally defined in [Algorithm 2](#). In [Algorithm 2](#), and in the rest of the paper, let  $\text{Coeff}_{Y^k}$  be a function that maps a polynomial to its  $Y^k$  coefficient, and let  $\text{RSDecode}_{d,n}$  be the decoding algorithm for  $\text{RS}_{d,n}$  correcting error and erasure patterns of weight less than  $n - d$ . Note that  $\text{RSDecode}$  is the standard Berlekamp-Welch algorithm [[WB86](#)].

---

**Algorithm 2:**  $\text{DecodeBivariate}_{d,\ell}(r, k)$

---

```

1 if  $k < 0$  then
2   | return []
3 else
4   |  $A_{\text{out}} \leftarrow \text{RSDecode}_{d-k,\ell}$ 
5   |  $A_x \leftarrow \text{Coeff}_{Y^k} \circ \text{RSDecode}_{k,\ell-x}$  for each  $x \in \{0, 1, \dots, \ell - 1\}$ 
6   |  $c_k \leftarrow \text{GMD}(A_{\text{out}}, A_0, \dots, A_{\ell-1}, r)$ 
7   | return  $\text{DecodeBivariate}_{d,\ell}(r - c_k Y^k, k - 1) + [c_k]$ 

```

---

Recall that the distance of  $\text{CAP}_{m,d,\ell}$  is  $\binom{\ell-d+m-1}{m} = |\Lambda_{m,\ell-d}|$ .

**Theorem 5.5.** *Let  $d, \ell \in \mathbb{N}$  with  $\ell > d$ . Let  $f \in \text{CAP}_{2,d,\ell}$ , be any codeword with  $Y$ -degree at most  $k$  (and total degree at most  $d$ ),  $e$  be an error pattern with weight less than  $\binom{\ell-d+1}{2}$ , and let  $r = f + e$ . Write  $f = \sum_{i=0}^k c_i(X)Y^i$ , Then  $\text{DecodeBivariate}_{d,\ell}(r, k)$  ([Algorithm 2](#)) returns  $f$  as a list of coefficients  $[c_0, \dots, c_k]$ .*

*Proof.* By induction. Suppose the claim is true for  $k - 1$ . We'll show the claim for  $k$ .

Let  $f$  be a polynomial with  $Y$ -degree at most  $k$ . Thus,  $f = \sum_{i=0}^k c_i(X)Y^i$ , where  $c_i$  is a univariate polynomial in  $X$  of degree at most  $d - i$ .

We claim that the evaluations of  $f$  on  $S$  is in one-to-one correspondence with the encoding of  $c_k$  under the following concatenated code.

Let  $C_{\text{out}}$  be the code that encodes  $c_k$  by evaluating it on  $\mathbb{N}_{<\ell}$ . For each  $x \in \{0, \dots, \ell - 1\}$ , define the inner code  $C_x$  as the code that encodes elements of  $\alpha \in \mathbb{F}$  in the following way. For any  $\alpha \in \mathbb{F}$ , map it to the univariate polynomial  $\alpha Y^k + \sum_{i=0}^{k-1} c_i(x)Y^i$ , and evaluate this polynomial on  $\mathbb{N}_{<\ell-x}$ . Let  $C = (C_0, \dots, C_{\ell-1}) \circ C_{\text{out}}$ .

Let  $(x, y) \in \Lambda_{2,\ell}$ . Then  $f(x, y)$  is the  $y$ th element of the  $x$ th block in the encoding of  $c_k$  under  $C$ . Indeed, the  $x$ th block of the outer code is  $c_k(x)$ , and the  $y$ th element of the encoding of  $c_k(x)$  under the inner code is  $c_k(x)y^k + \sum_{i=0}^{k-1} c_i(x)y^i = f(x, y)$ . Furthermore, it is not hard to check that if  $(x, y) \in \Lambda_{2,\ell}$ , then  $C_x$  indeed contains the evaluation at  $y$ : In  $C_x$ , the polynomial is evaluated on  $\mathbb{N}_{<\ell-x}$ , and  $y < \ell - x$  since  $(x, y) \in \Lambda_{2,\ell}$ . This establishes the one-to-one correspondence.

We now show that the concatenated code can be decoded to the desired distance using [Algorithm 1](#).  $C_{out} = \text{RS}_{d-k,\ell}$ , can thus be decoded from errors and erasures up to weight  $\ell - d + k$  using  $\text{RSDecode}_{d-k,\ell}$ . For each  $x \in \mathbb{N}_{<\ell}$ ,  $C_x$  is a subset of  $\text{RS}_{k,\ell-x}$ , and thus one can decode for the polynomial  $\alpha Y^k + \sum_{i=0}^{k-1} c_i(x)Y^i$  using  $\text{RSDecode}_{k,\ell-x}$  for error and erasure patterns with weight less than  $\max(\ell - x - k, 0)$ . One can then extract  $\alpha$  as the coefficient of  $Y^k$ .

The distance of the concatenated code is then the sum of the  $\ell - d + k$  smallest inner distances, which is the distances of the last  $\ell - d + k$  inner codes. Note that the distance of  $C_x$  is 0 for each  $x \in \{\ell - k, \ell - k + 1, \dots, \ell - 1\}$  (the last  $k$  codes). For  $x = \ell - k - i$ , the distance is  $i$ . Thus, the sum of the  $\ell - d + k$  smallest inner distances is just  $1 + 2 + \dots + \ell - d = \binom{\ell-d+1}{2}$ . Since the weight of  $e$  is less than  $\binom{\ell-d+1}{2}$ , the GMD algorithm [Algorithm 1](#) correctly decodes  $c_k$ .

Then,  $r - c_k Y^k = (f - c_k Y^k) + e$ . This is now a received word with  $Y$ -degree at most  $k - 1$ . By the inductive hypothesis, the recursive call correctly returns  $[c_0, \dots, c_{k-1}]$ , and the final return value of the function is the full list of coefficients  $[c_0, c_1, \dots, c_k]$  as required.  $\square$

**Corollary 5.6.**  *$\text{CAP}_{2,d,\ell}$  can be decoded from error patterns of weight less than the minimum distance.*

*Proof.* For any  $f \in \text{CAP}_{2,d,\ell}$ , the total degree of  $f$  is at most  $d$ . Therefore, the  $Y$ -degree of  $f$  is at most  $d$ , and by [Theorem 5.5](#),  $\text{DecodeBivariate}_{d,\ell}(r, d)$  correctly decodes so long as the weight of the error pattern is less than the minimum distance.  $\square$

### 5.3 Decoding on the $m$ -dimensional simplex

We now generalize to  $m$  dimensions. For brevity, for any  $\mathbf{v} \in \mathbb{N}^{m-1}$ , let  $\mathbf{X}^{\mathbf{v}} = X_1^{e_1} X_2^{e_2} \dots X_{m-1}^{e_{m-1}}$ , and write  $\deg(\mathbf{v}) = v_1 + v_2 + \dots + v_{m-1}$ . Write an  $m$ -variate polynomial  $f$  as an element of  $(\mathbb{F}[X_m])[X_1, \dots, X_{m-1}]$ . In particular, write

$$f = \sum_{\mathbf{v}: \deg(\mathbf{v}) \leq d} c_{\mathbf{v}}(X_m) \mathbf{X}^{\mathbf{v}},$$

where  $c_{\mathbf{v}}$  is a univariate polynomial of degree at most  $d - \deg(\mathbf{v})$ . Let  $\prec$  be an ordering over  $\mathbb{N}^{m-1}$ , corresponding to the graded lexicographical ordering of monomials. That is,  $\mathbf{v} \prec \mathbf{v}'$  iff either  $\deg(\mathbf{v}) < \deg(\mathbf{v}')$  or  $\deg(\mathbf{v}) = \deg(\mathbf{v}')$  and  $\mathbf{v}$  comes before  $\mathbf{v}'$  in lexicographical order. For example,

$$[0, 0] \prec [0, 1] \prec [1, 0] \prec [0, 2] \prec [1, 1] \prec [2, 0].$$

Notice that  $\prec$  is a total order. Our algorithm will recover  $c_v$  in descending order with respect to  $\prec$ . Define the function  $\text{prev} : \mathbb{N}^{m-1} \rightarrow \mathbb{N}^{m-1} \cup \{\perp\}$ , where  $\text{prev}(\mathbf{v})$  returns the direct predecessor of  $\mathbf{v}$  with respect to the order  $\prec$ . I.e.  $\text{prev}(\mathbf{v}) = \max(\{\mathbf{v}' : \mathbf{v}' \prec \mathbf{v}\})$ . As an edge case, define  $\text{prev}([0, 0, \dots, 0]) = \perp$  to specify an invalid monomial. The algorithm for decoding on the  $m$ -variable simplex is described in [Algorithm 3](#).

---

**Algorithm 3:**  $\text{DecodeSimplex}_{m,d,\ell}(r, \mathbf{v})$

---

```

1 if  $\mathbf{v} = \perp$  then
2   return []
3 else
4    $k \leftarrow \deg(\mathbf{v})$ 
5    $A_{\text{out}} \leftarrow \text{RSDecode}(d - k, \ell)$ 
6   for  $i \in \{0, \dots, \ell - 1\}$  do
7      $A_i \leftarrow \text{Coeff}_{\mathbf{X}^{\mathbf{v}}} \circ \text{DecodeSimplex}_{m-1,k,\ell-i}(\cdot, [k, 0, \dots, 0])$ 
8    $c_v \leftarrow \text{GMD}(A_{\text{out}}, A_0, \dots, A_{\ell-1}, r)$ 
9    $\mathbf{v}' = \text{prev}(\mathbf{v})$ 
10  return  $\text{DecodeSimplex}_{m,d,\ell}(r - c_v \mathbf{X}^{\mathbf{v}}, \mathbf{v}') + [c_v]$ 

```

---

**Theorem 5.7.** Let  $m, d, \ell \in \mathbb{N}$ , such that  $\ell > d$ , and  $m \geq 1$ . Let  $f \in \text{CAP}_{m,d,\ell}$ , and write  $f = \sum_{\mathbf{u}: \deg(\mathbf{u}) \leq d} c_{\mathbf{u}}(X_m) \mathbf{X}^{\mathbf{u}}$ . Let  $\mathbf{v}$  be an exponent vector of degree at most  $d$  and suppose for all  $\mathbf{w}$  with  $\mathbf{v} \prec \mathbf{w}$ ,  $c_{\mathbf{w}} = 0$ . Let  $e$  be an error pattern with weight less than  $|\Lambda_{m,\ell-d}|$ , and let  $r = f + e$ . Then  $\text{DecodeSimplex}_{m,d,\ell}(r, \mathbf{v})$  ([Algorithm 3](#)) returns  $f$  as a list of coefficients  $(c_{\mathbf{v}})_{\mathbf{v}: \deg(\mathbf{v}) \leq d}$ .

*Proof.* Let  $\ell, d \in \mathbb{N}$  with  $\ell > d$ . For notational simplicity, suppose the code is the CAP code on the standard simplex of side length  $\ell$ ,  $\Lambda_{m,\ell}$ . The proof works exactly the same on  $\Lambda_{m,A}$  for any  $A \subset F$  with  $|A| = \ell$  by replacing elements of  $A^m$  with the vector of their indices in  $\{0, 1, \dots, \ell - 1\}$ .

By induction on  $m$ .

For the base case  $m = 1$ , observe that  $E_{m,d,\Lambda_{m,\ell}}$  is a Reed-Solomon code, so  $\text{DecodeSimplex}_{1,d,\ell}$  is  $\text{RSDecode}_{d,\ell}$ .

Now, suppose the theorem is true for  $\text{DecodeSimplex}_{m-1,d',\ell'}$  for any  $d'$  and  $\ell'$  with  $\ell' > d'$ . We'll show the theorem for  $\text{DecodeSimplex}_{m,d,\ell}$ . The proof is similar to that of [Theorem 5.5](#). By induction on  $\mathbf{v}$  with respect to  $\prec$ .

For the inductive hypothesis, suppose that for any polynomial  $g$  where the maximum monomial with a non-zero coefficient is  $\mathbf{X}^{\text{prev}(\mathbf{v})}$ , and error pattern  $e$  with weight less than  $|\Lambda_{m,\ell-d}|$ ,  $\text{DecodeSimplex}_{m,d,\ell}(g + e, \text{prev}(\mathbf{v}))$  returns  $g$ .

Let  $f$  be any polynomial where the maximum monomial with a non-zero coefficient is  $\mathbf{X}^{\mathbf{v}}$ , and let  $e$  be an error pattern of weight less than  $|\Lambda_{m,\ell-d}|$ , we'll show  $\text{DecodeSimplex}_{m,d,\ell}(f + e, \text{prev}(\mathbf{v}))$  returns  $f$ . Let  $k = \deg(\mathbf{v})$ . We first show that  $f$  is also the encoding of  $c_v$  under the following concatenated code.

Let  $C_{out}$  be the code that encodes the univariate polynomial  $c_v$  by evaluating it on  $\mathbb{N}_{<\ell}$ . For  $i \in \mathbb{N}_{<\ell}$ , define the inner code  $C_i$  to be the code that encodes symbols  $\alpha \in \mathbb{F}$  as a  $m-1$  variate polynomial

$$f_\alpha(X_1, \dots, X_{m-1}) = \alpha \mathbf{X}^v + \sum_{\mathbf{u}: \mathbf{u} \prec v} c_u(x_m) \mathbf{X}^u,$$

and evaluates it on  $\Lambda_{m-1, \ell-i}$ . Consider the code  $C = (C_0, C_1, \dots, C_{\ell-1}) \circ C_{out}$ .

We claim that the evaluation of  $f$  on  $\Lambda_{m, \ell}$  is the encoding of  $c_v$  under  $C$ . Let  $(x_1, \dots, x_{m-1}, i) \in \Lambda_{m, \ell}$ . Then, this corresponds to the element indexed by  $(x_1, \dots, x_{m-1})$  in the  $i$ th block. Since the  $i$ th element in the outer encoding of  $c_v$  is  $c_v(i)$ , and the coordinate of the  $i$ th inner code indexed by  $(x_1, \dots, x_{m-1})$  is the evaluation  $f_{c_i}(x_1, \dots, x_{m-1}) = f(x_1, \dots, x_m)$ . Since  $x_1 + x_2 + \dots + x_{m-1} + i < \ell$ , we have  $x_1 + x_2 + \dots + x_{m-1} < \ell - i$ , so  $(x_1, \dots, x_{m-1}) \in \Lambda_{m-1, \ell-i}$  and hence is indeed an evaluation point in  $C_i$ .

The degree of  $c_v$  is at most  $d-k$ , thus  $C_{out} = RS(d-k, \ell)$ , and can be decoded from error and erasure patterns with weight less than  $\ell - d + k$ . For each  $i \in \mathbb{N}_{<\ell}$ ,  $C_i$  is a subcode of  $E_{m-1, k, \Lambda_{m-1, \ell-i}}$  since  $f_\alpha$  has degree at most  $\deg(v) = k$ . Thus, for any  $\alpha$ ,  $f_\alpha$  can be decoded from error and erasure patterns with weight less than  $|\Lambda_{m-1, \ell-i-k}|$  using the (outer) inductive hypothesis. To retrieve  $\alpha$ , we just extract the coefficient of  $\mathbf{X}^v$ .

Instantiating the GMD decoder (Algorithm 1) with these outer and inner decoders, we get that we can decode so long as the number of errors is less than the sum of the  $\ell - d + k$  smallest inner distances. Note that for  $x \in \{\ell - k, \ell - k + 1, \dots, \ell - 1\}$ , the last  $k$  codes, the distance is zero. For  $x = \ell - k - i$ , the distance is  $|\Lambda_{m-1, i}|$ . Thus, the sum of the  $\ell - d + k$  smallest inner distance is  $\sum_{i=1}^{\ell-d} |\Lambda_{m-1, i}| = |\Lambda_{m, \ell-d}|$ . Since the weight of  $e$  is less than  $|\Lambda_{m, \ell-d}|$ , GMD instantiated with these outer and inner decoders correctly decodes  $c_v$ .

Then,  $r - c_v Y^k = (f - c_v Y^k) + e$ . Notice that this is now a received word where the maximum monomial present with respect to  $\prec$  is at most  $\text{prev}(v)$ . By the inductive hypothesis, the recursive call correctly returns  $[c_u]_{\mathbf{u} \prec v}$ , and the final return value of the function is the full list of coefficients.  $\square$

**Corollary 5.8.**  $CAP_{m, d, \ell}$  can be decoded from error patterns of weight less than the minimum distance.

*Proof.* Note that for any  $f \in CAP_{m, d, \ell}$ , the total degree is at most  $d$ . Viewing  $f \in (\mathbb{F}[X_m])[X_1, \dots, X_{m-1}]$ , the greatest monomial with respect to  $\prec$  is  $X_1^d$ , corresponding to the exponent vector  $[d, 0, 0, \dots, 0]$ . Thus, by Theorem 5.7, DecodeSimplex with the  $v$  parameter as  $[d, 0, \dots, 0]$  decodes any received word so long as the weight of the error patten is less than the minimum distance.  $\square$

## 6 Unique Decoding of GAP codes

This section gives efficient algorithms for decoding GAP codes Theorem 3.5 up to half their minimum distance. At a high level, for our decoding algorithms, we essentially view these codes as

a specific instance of decoding concatenated codes where the outer code is a Reed-Solomon code, and the inner codes are GAP codes in a smaller dimension, up to some (crucial) technical differences. We massage this point of view sufficiently to be in a position to invoke the framework of generalized minimum distance (GMD) decoding for concatenated codes of Forney [For66]. While this appears to be a natural approach, we note that apriori, it is quite unclear whether this strategy can decode up to half the minimum distance of the code. For instance, we do not know if this strategy can give a simple unique decoding algorithm for decoding polynomial evaluation codes on arbitrary product sets (a problem that was solved only relatively recently by Kim and Kopparty [KK17]) or our second construction of such codes in Theorem 3.12, both of which, though relying on GMD decoding seem to require further ideas and a more complex algorithm overall. Perhaps this innate simplicity of the decoding algorithm is further evidence that the codes in Theorem 3.5 are quite natural and merit further investigation.

In this section, we work with concatenated codes where the outer code is a Reed-Solomon code, and the inner codes are all polynomial evaluation codes with the same block length and minimum distance, with potentially different evaluation points. The Reed-Solomon codes in the outer codes here are in a slightly unusual setting, and we formally define them now.

**Definition 6.1** (Reed-Solomon codes over a polynomial ring). *Let  $\mathbb{F}$  be a finite field,  $\mathbf{X} = (X_1, \dots, X_{m-1})$  be an  $(m-1)$ -tuple of variables,  $S \subseteq \mathbb{F}[\mathbf{X}]$  be a set of  $t$  affine linear forms and  $d \in \mathbb{N}$  be a natural number. Then, Reed-Solomon codes  $\mathcal{RS}(m, t, d, S)$  are defined as follows.*

*The message space of the code is identified with the set of  $m$ -variate polynomials  $f(\mathbf{X}, Y)$  of total degree at most  $d$ . To encode any such message  $f(\mathbf{X}, Y)$ , we view it as a degree  $d$  univariate in  $Y$  with coefficients in the field  $\mathbb{F}(\mathbf{X})$  and the codeword associated to  $f$  equals  $(f(\mathbf{X}, L_1(\mathbf{X})), \dots, f(\mathbf{X}, L_t(\mathbf{X})))$ , i.e. the evaluation of  $f(\mathbf{X}, Y) \in (\mathbb{F}(\mathbf{X}))[Y]$  on the  $t$  inputs  $L_1(\mathbf{X}), \dots, L_t(\mathbf{X}) \in S$ .  $\diamond$*

The following observation summarises some of the basic properties of this code.

**Observation 6.2.** *Let  $\mathcal{RS}(m, t, d, S)$  be the codes defined in Definition 6.1. Then, the following are true.*

- *The distance of the code is  $(t - d)$ .*
- *For any message polynomial  $f(\mathbf{X}, Y)$  of total degree at most  $d$ , every coordinate of its encoding  $(f(\mathbf{X}, L_1(\mathbf{X})), \dots, f(\mathbf{X}, L_t(\mathbf{X})))$  is an  $(m-1)$ -variate polynomial of degree  $d$  in  $\mathbb{F}[\mathbf{X}]$ . Thus, the alphabet of the code can be naturally identified with  $\mathbb{F}^{\binom{d+m-1}{m-1}}$ .*

Reed-Solomon codes are very well studied from the point of view of decoding algorithms and, in particular, are known to be efficiently decodable up to half their minimum distance via the well-known decoding algorithm of Berlekamp & Welch [WB86]. However, typically, we study Reed-Solomon codes where the underlying alphabet is a finite field, in contrast to the Reed-Solomon codes in Definition 6.1 where the underlying alphabet is the set of  $(m-1)$ -variate polynomials of degree  $d$  over the field  $\mathbb{F}$ . The standard Berlekamp-Welch algorithm can be naturally extended to

this setting with two additional technical ingredients. The first ingredient is an efficient algorithm that, when given a rank deficient  $N \times N$  matrix  $M$  whose entries are  $m$ -variate polynomials in  $\mathbb{F}[\mathbf{X}]$  of degree at most  $D$  (for some parameter  $D$ ) outputs a non-zero vector with entries in  $\mathbb{F}[\mathbf{X}]$  that is in the kernel of  $M$ . Lemma 3 in [BHK23] addresses precisely this problem and shows that there is a deterministic algorithm for this problem that runs in time  $\text{poly}(N^m, D^m)$ . Moreover, the degree of every coordinate of the output vector of the algorithm is at most  $ND$ . The second technical ingredient is an efficient algorithm that takes any two  $m$ -variate polynomials  $F$  and  $G$  of degree at most  $D$  as input and outputs the quotient  $G/F$  if  $F$  divides  $G$  and a bot ( $\perp$ ) otherwise. One of the ways of doing this is to fix any valid monomial ordering (e.g., the graded lexicographic order) and just do a standard long division. This gives a deterministic algorithm for this problem with time complexity  $\text{poly}(D^m)$ .

With these two ingredients in place, the Berlekamp-Welch decoder for Reed-Solomon codes naturally extends to the codes in Definition 6.1. For our applications here, we need to decode Reed-Solomon codes from errors and erasures, which the Berlekamp-Welch algorithm handles without any further issues.

**Theorem 6.3** (Berlekamp-Welch [WB86]). *Let  $\mathbb{F}$  be any finite field. There is a deterministic algorithm, that for all  $d, m, t \in \mathbb{N}$  and subsets  $S \subseteq \mathbb{F}[\mathbf{X}]$  of size  $t$  of affine linear forms, decodes the code  $\mathcal{RS}(m, t, d, S)$  in time  $\text{poly}(t^m, d^m, \log |\mathbb{F}|)$  from errors and erasures with weight less than the distance of the code.*

We start with the bivariate case and then rely on induction on the number of variables to lift these ideas to the multivariate case.

## 6.1 Decoding in two dimensions

Before moving on to the decoding algorithm, we first view  $\text{GAP}_{m,d,t}$  as a concatenated code. We begin with the  $m = 2$  case, where the details are cleaner, and which will also serve as the base case for our induction based argument.

We have the following set up. Let  $\mathcal{H} = \{\ell_1, \ell_2, \dots, \ell_t\}$  be a set of  $t$  lines in two dimensions that are in general position, i.e., every pair of these lines intersect at a point, and no three of them intersect anywhere. For every  $i$ , let  $\ell_i$  be parameterized as

$$\ell_i := \{(x, a_i x + b_i) : x \in \mathbb{F}\},$$

where  $a_i, b_i$  elements of  $\mathbb{F}$ . Let  $T$  be the set of pairwise intersections of lines in  $\mathcal{H}$ . We will decode  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ . Recall that  $\text{GAP}_{2,d,t}^{\mathcal{H}}$  has distance  $\binom{t-d}{2}$ .

**An intermediate concatenated code:** For describing our decoding algorithm, we will rely on another family of codes that are closely related to  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ . These intermediate codes are denoted by  $\mathcal{I}_2^{\mathcal{H}}$  and are defined as follows.



**Definition 6.4.** The message space of  $\mathcal{I}_2^{\mathcal{H}}$  is the space of bivariate polynomials of total degree at most  $d$  in  $\mathbb{F}[X, Y]$ . The block length of the code is  $t(t-1)$  and the alphabet is  $\mathbb{F}$ . We index the coordinates of the codewords by pairs  $(i, j)$ , where  $i, j \in \{1, 2, \dots, t\}$ ,  $i \neq j$ , and the  $(i, j)$ th coordinate of the codeword corresponding to a message polynomial  $f$  is the evaluation of  $f$  at the intersection of the lines  $\ell_i$  and  $\ell_j$ .  $\diamond$

From the above definition, it follows that the main difference between the encoding of  $f$  under the codes  $\mathcal{I}_2^{\mathcal{H}}$  and  $\text{GAP}_{2,d,t}^{\mathcal{H}}$  is that in the former, the evaluation of  $f$  on the intersection point of any two distinct lines  $\ell_i$  and  $\ell_j$  appears on two separate coordinates, namely,  $(i, j)$  and  $(j, i)$ , whereas in the latter, this evaluation appears exactly once, at a coordinate indexed by the set  $\{i, j\}$ . Moreover, given the encoding of any message  $f$  under one of these codes, we can efficiently compute its encoding as a codeword of the other code.

The following lemma summarises some more interesting properties of this new code  $\mathcal{I}_2^{\mathcal{H}}$  that will be useful for our algorithm. Again, we follow the notation set up earlier in this section.

**Lemma 6.5.** Let  $C_{\text{out}}$  be the code  $\mathcal{RS}(2, t, d, S)$ , where  $S$  is the set  $\{a_i X + b_i : i \in \{1, 2, \dots, t\}\}$  of  $t$  distinct affine forms.

For  $i \in \{1, 2, \dots, t\}$ , let  $C_i$  be the Reed-Solomon code of dimension  $(d+1)$  and block length  $(t-1)$  with the evaluation points being the set  $S_i = \{\alpha_j : (\alpha_j, a_i \alpha_j + b_i) \in \ell_i \cap \ell_j \text{ for some } j \neq i\}$ .

Then, the following are true.

- $\mathcal{I}_2^{\mathcal{H}}$  can be viewed as the code obtained by the concatenation of a code  $C_{\text{out}}$  and  $C_{\text{in}} = (C_1, \dots, C_t)$ .
- $\delta(\mathcal{I}_2^{\mathcal{H}}) = \delta(\text{GAP}_{2,d,t}^{\mathcal{H}}) = \delta(C_{\text{out}}) \cdot \delta(C_i)$ , for every  $i \in \{1, 2, \dots, t\}$ .

We now sketch the proof of the lemma.

*Proof Sketch.* The first item immediately follows from the definitions of the codes involved.

For the second item, we note that  $\mathcal{I}_2^{\mathcal{H}}$ , as defined, is a linear code, and hence its distance equals the minimum of the weights of non-zero codewords in it. As discussed earlier in this section, there is a bijection between  $\mathcal{I}_2^{\mathcal{H}}$  and  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ , where every codeword of the former is obtained by repeating twice every symbol in a codeword of the latter (up to an appropriate indexing of coordinates). Thus, the Hamming weight of any codeword of minimum weight in  $\mathcal{I}_2^{\mathcal{H}}$  is twice that of  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ , and hence equals  $2 \cdot \binom{t-d}{2} = (t-d)(t-d-1)$ . Since the block length of  $\mathcal{I}_2^{\mathcal{H}}$  is also twice the block length of  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ , we get that their fractional distances must be the same. The last equality immediately follows from the first item of the lemma and by observing that the distance of  $C_{\text{out}}$ , which is really a Reed-Solomon code over the field  $\mathbb{F}(X)$  equals  $(t-d)$ . The distance of every inner code  $C_i$  equals  $(t-1-d)$ , and hence their product equals  $(t-d)(t-d-1)$ .  $\square$

From Lemma 6.5, we have that  $\mathcal{I}_2^{\mathcal{H}}$  is a concatenated code. Thus, from Theorem 2.7 and Theorem 6.3, we have the following corollary.

**Corollary 6.6.** There is a deterministic algorithm, denoted by `DecodeIntermediateBivariate`, that decodes the code  $\mathcal{I}_2^{\mathcal{H}}$  from error patterns of weight less than the minimum distance in time  $\text{poly}(t, d, \log |\mathbb{F}|)$ .



**The decoding algorithm :** We now describe our decoding algorithm for two dimensional GAP codes.

---

**Algorithm 4:** DecodeBivariateGeometric  $(t, d, \mathcal{H}, \mathbf{r})$

---

**Input:**  $t, d \in \mathbb{N}$  with  $t \geq d$ , a set  $\mathcal{H} = \{\ell_1, \dots, \ell_t\}$  of lines in  $\mathbb{F}^2$  in general position and received word  $\mathbf{r} \in \mathbb{F}^{\binom{t}{2}}$

**Output:** The unique polynomial  $f \in \mathbb{F}[X, Y]$  of degree at most  $d$  such that  $\Delta(\text{GAP}_{2,d,t}^{\mathcal{H}}(f), \mathbf{r}) < \binom{t-d}{2}/2$  or,  $\perp$  if no such polynomial exists.

```

1 for  $i \in \{1, \dots, t\}$  do
2   for  $j \in \{1, \dots, t\} \setminus \{i\}$  do
3      $\tilde{\mathbf{r}}(i, j) \leftarrow \mathbf{r}(\{i, j\})$ 
4 Output DecodeIntermediateBivariate( $t, d, \mathcal{H}, \tilde{\mathbf{r}}$ ) (Corollary 6.6).
```

---

Given the discussion leading up to this point, the analysis of [algorithm 4](#) follows almost immediately and is summarised in the following theorem.

**Theorem 6.7.** *Let  $t, d \in \mathbb{N}$  with  $t \geq d$ , and  $\mathcal{H} = \{\ell_1, \dots, \ell_t\}$  be a set of  $t$  lines in  $\mathbb{F} \times \mathbb{F}$  in general position. Then [algorithm 4](#) decodes  $\text{GAP}_{2,d,t}^{\mathcal{H}}$  from error patterns with weight less than the minimum distance. I.e., when the number of errors is less than half the minimum distance. Moreover, the time complexity of [algorithm 4](#) is at most  $\text{poly}(t, d, |\mathbb{F}|)$ .*

*Proof.* The bound on the time complexity follows immediately from the observation that the construction of the word  $\tilde{\mathbf{r}}$  from the received word  $\mathbf{r}$  can be done in  $O(t^2)$  time and the bound on the time complexity of the unique decoding algorithm for code  $\mathcal{I}_2^{\mathcal{H}}$  in [Corollary 6.6](#).

We now argue the correctness of the algorithm. Given the received word  $\mathbf{r}$ , the algorithm first constructs the received word  $\tilde{\mathbf{r}}$  for the intermediate code  $\mathcal{I}_2^{\mathcal{H}}$  and then invokes the unique decoding algorithm for the code  $\mathcal{I}_2^{\mathcal{H}}$  given in [Corollary 6.6](#) on the received word  $\tilde{\mathbf{r}}$ . Thus, the following claim, together with the correctness of the unique decoding algorithm for  $\mathcal{I}_2^{\mathcal{H}}$  in [Corollary 6.6](#) implies the correctness of [algorithm 4](#).

**Claim 6.8.** *The received word  $\mathbf{r}$  is within half the minimum distance of the code  $\text{GAP}_{2,d,t}^{\mathcal{H}}$  if and only if the word  $\tilde{\mathbf{r}}$  is within half the minimum distance of the code  $\mathcal{I}_2^{\mathcal{H}}$ .*

*Proof of claim.* Let  $\mathbf{c}$  be the codeword of the code  $\text{GAP}_{2,d,t}^{\mathcal{H}}$  closest to the received word  $\mathbf{r}$  and let  $\tilde{\mathbf{c}} \in \mathbb{F}^{t(t-1)}$  be defined as

$$\forall i \neq j \in \{1, \dots, t\}, \quad \tilde{\mathbf{c}}(i, j) = \tilde{\mathbf{c}}(j, i) = \mathbf{c}(\{i, j\}).$$

From the definition of the code  $\mathcal{I}_2^{\mathcal{H}}$ , it follows that  $\tilde{\mathbf{c}}$  is a codeword in  $\mathcal{I}_2^{\mathcal{H}}$ . In fact,  $\mathbf{c}$  and  $\tilde{\mathbf{c}}$  are encodings of the same message polynomial. Moreover, from the definitions of  $\tilde{\mathbf{r}}$  and  $\tilde{\mathbf{c}}$ , it follows that  $\Delta(\tilde{\mathbf{c}}, \tilde{\mathbf{r}}) = 2 \cdot \Delta(\mathbf{c}, \mathbf{r})$ . Using this bound and the fact that the block length of  $\mathcal{I}_2^{\mathcal{H}}$  is twice that

of  $\text{GAP}_{2,d,t}^{\mathcal{H}}$ , we get that the fractional Hamming distances,  $\delta(\tilde{\mathbf{c}}, \tilde{\mathbf{r}})$  and  $\delta(\mathbf{c}, \mathbf{r})$  must be equal to each other. The claim now follows immediately from [Lemma 6.5](#), which shows that the fractional minimum distances of these codes are equal.  $\square$

$\square$

## 6.2 Decoding in higher dimensions

In this section, we extend the ideas in the previous section from the bivariate to the multivariate case. Our technical argument is based on an induction on the number of variables. We start by setting up the notation. We have a set  $\mathcal{H}$  of hyperplanes  $\mathcal{H} = \{H_1, H_2, \dots, H_t\}$  in  $\mathbb{F}^m$  that are in general position, i.e., every subset of size  $m$  of these hyperplanes intersect at a point, and no  $(m+1)$  of them intersect anywhere. Without loss of generality (up to an invertible change of basis), we can assume that the hyperplanes can be expressed as

$$H_i = \left\{ (\mathbf{a}, L_i(\mathbf{a})) : \mathbf{a} \in \mathbb{F}^{m-1} \right\},$$

where  $L_i(\mathbf{X}) \in \mathbb{F}[\mathbf{X}]$  is an affine linear form on variables  $X_1, X_2, \dots, X_{m-1}$ .

The set of evaluation points is the set of  $m$ -wise intersection of these hyperplanes, and hence the corresponding polynomial evaluation code  $\text{GAP}_{m,d,t}^{\mathcal{H}}$  has block length  $\binom{t}{m}$ , and from [Theorem 3.5](#), we have that its distance is  $\binom{t-d}{m}$ . This section aims to efficiently decode these codes from less than  $\binom{t-d}{m}/2$  errors. As in the case of two dimensions, once again, we relate this code to an intermediate code  $\mathcal{I}_m^{\mathcal{H}}$  of equal fractional distance and observe that  $\mathcal{I}_m^{\mathcal{H}}$  is obtained by concatenating a Reed-Solomon code with inner codes that are all instances of  $\text{GAP}_{m-1,d,t-1}$ . These inner codes can be efficiently decoded from half their minimum distance via induction, and we can invoke [Theorem 6.3](#) and [Theorem 2.7](#) to get our decoding algorithm for  $\text{GAP}_{m,d,t}^{\mathcal{H}}$ .

**An intermediate concatenated code:** We start by defining the intermediate code. Throughout this section,  $\mathbf{X}$  denotes the  $(m-1)$ -tuple  $(X_1, X_2, \dots, X_{m-1})$  of variables.

**Definition 6.9.** *The message space of  $\mathcal{I}_m^{\mathcal{H}}$  is the space of  $m$ -variate polynomials of total degree at most  $d$  in  $\mathbb{F}[\mathbf{X}, Y]$ . The block length of the code is  $m \cdot \binom{t}{m}$  and the alphabet is  $\mathbb{F}$ . We index the coordinates of a codeword of this code by pairs  $(i, \mathbf{j})$ , where  $i \in \{1, 2, \dots, t\}$  and  $\mathbf{j}$  is a subset of size  $(m-1)$  of  $\{1, 2, \dots, t\} \setminus \{i\}$ . The  $(i, \mathbf{j})$ th coordinate of the codeword corresponding to a message polynomial  $f$  is the evaluation of  $f$  at the intersection of  $m$  hyperplanes  $\{H_i\} \cup \{H_\ell : \ell \in \mathbf{j}\}$ .  $\diamond$*

Once again, it follows that code words of  $\mathcal{I}_m^{\mathcal{H}}$  and  $\text{GAP}_{m,d,t}^{\mathcal{H}}$  are in a natural bijection, and one can go from a codeword of the latter to the former by repeating every coordinate exactly  $m$  times. The following analog of [Lemma 6.5](#) follows almost immediately from the discussion so far.

**Lemma 6.10.** Let  $C_{out}$  be the code  $RS(m, t, d, S)$  from [Definition 6.1](#) where  $S$  is the set  $\{L_1(\mathbf{X}), \dots, L_t(\mathbf{X})\}$  of  $t$ -distinct affine forms and  $H_1, H_2, \dots, H_t$  are hyperplanes in general position such that for every  $i$ ,  $H_i = \{(\mathbf{a}, L_i(\mathbf{a})) : \mathbf{a} \in \mathbb{F}^{m-1}\}$ .

For  $i \in \{1, 2, \dots, t\}$ , let  $C_i = \text{GAP}_{m-1, d, t-1}^{\mathcal{H}_i}$ , where  $\mathcal{H}_i = \{H_j \cap H_i : j \in \{1, 2, \dots, t\} \setminus \{i\}\}$ . Then, the following are true.

- $\mathcal{I}_m^{\mathcal{H}}$  is the code obtained by the concatenation of  $C_{out}$  and  $C_{in} = (C_1, \dots, C_t)$ .
- $\delta(\mathcal{I}_m^{\mathcal{H}}) = \delta(\text{GAP}_{m, d, t}^{\mathcal{H}}) = \delta(C_{out}) \cdot \delta(C_i)$ , for every  $i \in \{1, 2, \dots, t\}$ .

From [Lemma 6.10](#), we note that to invoke [Theorem 2.7](#) to decode  $\mathcal{I}_m^{\mathcal{H}}$ , we need the algorithm from [Theorem 6.3](#) and a unique decoding algorithm for  $\text{GAP}_{m-1, d, t-1}$ . For the latter, we note that if  $\mathcal{H} = \{H_1, H_2, \dots, H_t\}$  is a set of hyperplanes in  $\mathbb{F}^m$  in general position, then for every  $i \in \{1, 2, \dots, t\}$ ,  $\mathcal{H}_i = \{H_j \cap H_i : j \neq i\}$  are hyperplanes in  $\mathbb{F}^{m-1}$  in general position. Thus, the code  $C_i$  is indeed a GAP code for polynomial in one fewer variable and by the induction hypothesis, such GAP codes are efficiently decodable up to half their minimum distance. The above discussion, together with [Theorem 2.7](#), immediately implies the following.

**Lemma 6.11.** There is an algorithm, denoted by `DecodeIntermediateMultivariate`, that decodes  $\mathcal{I}_m^{\mathcal{H}}$  from error patterns of weight less than the minimum distance. I.e., when the number of errors is less than half the minimum distance. Moreover, the algorithm is deterministic and on every input, makes at most  $t$  function calls to a unique decoder for the polynomial evaluation code  $\text{GAP}_{m-1, d, t-1}$  ([algorithm 5](#)), at most  $t$  function calls to the Berlekamp-Welch decoder for Reed-Solomon codes in [Theorem 6.3](#), and performs at most  $\text{poly}(t^m, \log |\mathbb{F}|)$  additional computation.

**The decoding algorithm :** We now describe our decoding algorithm for  $\text{GAP}_{m, d, t}^{\mathcal{H}}$ .

---

**Algorithm 5:** `DecodeMultivariateGeometric` ( $m, t, d, \mathcal{H}, \mathbf{r}$ )

---

**Input:**  $m, t, d \in \mathbb{N}$  with  $t \geq d$ , a set  $\mathcal{H} = \{H_1, \dots, H_t\}$  of hyperplanes in  $\mathbb{F}^m$  in general position and received word  $\mathbf{r} \in \mathbb{F}^{\binom{t}{m}}$

**Output:** The unique  $m$ -variate polynomial  $f \in \mathbb{F}[\mathbf{X}, Y]$  of degree at most  $d$  such that  $\Delta(\text{GAP}_{m, d, t}^{\mathcal{H}}(f), \mathbf{r}) < \binom{t-d}{m}/2$  or,  $\perp$  if no such polynomial exists.

```

1 if  $m = 2$  then
2   | Output DecodeBivariateGeometric ( $t, d, \mathcal{H}, \mathbf{r}$ ) (algorithm 4).
3 else
4   | for  $i \in \{1, \dots, t\}$  do
5     |   for  $\mathbf{j} \subseteq \{1, \dots, t\} \setminus \{i\}, |\mathbf{j}| = m - 1$  do
6       |     |  $\tilde{\mathbf{r}}(i, \mathbf{j}) \leftarrow \mathbf{r}(\{i\} \cup \mathbf{j})$ 
7   | Output DecodeIntermediateMultivariate ( $m, d, t, \mathcal{H}, \tilde{\mathbf{r}}$ ) (Lemma 6.11).
```

---

Note that the unique decoding algorithm for  $\mathcal{I}_m^{\mathcal{H}}$  in [Lemma 6.11](#) makes function calls to `algo-`

algorithm 5, but for the number of variables being  $(m - 1)$ , which is handled recursively. The following theorem sums up the analysis of algorithm 5.

**Theorem 6.12.** *Let  $m, t, d \in \mathbb{N}$  with  $t \geq d$ , and  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of  $t$  hyperplanes in  $\mathbb{F}^m$  in general position. Then DecodeMultivariateGeometric decodes  $\text{GAP}_{m,d,t}^{\mathcal{H}}$  from error patterns of weight less than the minimum distance. I.e., when the number of errors is less than half the minimum distance. Moreover, the time complexity of algorithm 5 is at most  $\text{poly}(t^m, \log |\mathbb{F}|)$ .*

*Proof.* The time complexity of the algorithm can be upper bounded by the time taken to generate the word  $\tilde{\mathbf{r}}$  and then the time taken in the function call to DecodeIntermediateMultivariate( $m, d, t, \mathcal{H}, \tilde{\mathbf{r}}$ ). By Lemma 6.11, on each such function call, the algorithm DecodeIntermediateMultivariate makes  $t$  calls to algorithm 5, but now for the number of variables being  $(m - 1)$  and the number of hyperplanes being  $(t - 1)$  and  $t$  calls to the Reed-Solomon decoder given in Theorem 6.3 which takes time  $\text{poly}(t^m, d^m, \log |\mathbb{F}|) \leq \text{poly}(t^m, \log |\mathbb{F}|)$ , in addition to some  $\text{poly}(t^m, \log |\mathbb{F}|)$  additional computation. Thus, the time complexity  $T(m, t, d)$  of algorithm 5 can be upper bounded by the following recursion.

$$\begin{aligned} T(2, d, t) &\leq \text{poly}(t, d, \log |\mathbb{F}|) \\ \forall m > 2, T(m, d, t) &\leq t \cdot T(m - 1, d, t - 1) + \text{poly}(t^m, \log |\mathbb{F}|), \end{aligned}$$

which implies that  $T(m, d, t)$  is at most  $\text{poly}(t^m, \log |\mathbb{F}|)$ .

We now argue the correctness of the algorithm. This follows the outline of the proof of correctness of algorithm 4 in Theorem 6.7. Given the received word  $\mathbf{r}$ , the algorithm first constructs the received word  $\tilde{\mathbf{r}}$  for the intermediate code  $\mathcal{I}_m^{\mathcal{H}}$  and then invokes the unique decoding algorithm for the code  $\mathcal{I}_m^{\mathcal{H}}$  given in Corollary 6.6 on the received word  $\tilde{\mathbf{r}}$ . Therefore, together with the correctness of the unique decoding algorithm for  $\mathcal{I}_m^{\mathcal{H}}$  in Lemma 6.11, the following claim implies the correctness of algorithm 5.

**Claim 6.13.** *The received word  $\mathbf{r}$  is within half the minimum distance of the code  $\text{GAP}_{m,d,t}^{\mathcal{H}}$  if and only if the word  $\tilde{\mathbf{r}}$  is within half the minimum distance of the code  $\mathcal{I}_m^{\mathcal{H}}$ .*

*Proof of claim.* Let  $\mathbf{c}$  be the codeword of the code  $\text{GAP}_{m,d,t}^{\mathcal{H}}$  closest to the received word  $\mathbf{r}$  and let  $\tilde{\mathbf{c}}$  be defined as

$$\forall i \in \{1, \dots, t\}, \mathbf{j} \subseteq \{1, \dots, t\} \setminus \{i\}, |\mathbf{j}| = m - 1, \quad \tilde{\mathbf{c}}(i, \mathbf{j}) = \mathbf{c}(\{i\} \cup \mathbf{j}).$$

From the definition of the code  $\mathcal{I}_m^{\mathcal{H}}$ , it follows that  $\tilde{\mathbf{c}}$  is a codeword in  $\mathcal{I}_m^{\mathcal{H}}$ . In fact,  $\mathbf{c}$  and  $\tilde{\mathbf{c}}$  are encodings of the same message polynomial. Moreover, from the definitions of  $\tilde{\mathbf{r}}$  and  $\tilde{\mathbf{c}}$ , it follows that  $\Delta(\tilde{\mathbf{c}}, \tilde{\mathbf{r}}) = m \cdot \Delta(\mathbf{c}, \mathbf{r})$ . The claim now follows from this bound, the fact that the block length of  $\mathcal{I}_m^{\mathcal{H}}$  is also  $m$  times the block length of  $\text{GAP}_{m,d,t}^{\mathcal{H}}$ , and the fact that the fractional minimum distances of these codes are equal to each other as shown in Lemma 6.10.  $\square$

□

**Remark 6.14.**  $\text{GAP}_{m,d,t}$  can also be efficiently decoded from error *and erasure* patterns of weight at most the minimum distance. ◇

This follows from the fact that the GMD algorithm can handle erasures as long as each inner decoder can handle erasures. We will elaborate more on this in the next section.

## 7 Local testability of GAP codes

Next, we discuss the local testability of GAP codes. We analyze two very natural local tests for GAP codes, the line-point test, and the plane-point test, and show their local testability via these tests.

First, we fix some notations and define our tests. Throughout this section, let  $\mathbb{F}$  be a field,  $m, d, t \in \mathbb{N}$  be such that  $t \geq d + m$ , and let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of hyperplanes in general position in  $\mathbb{F}^m$ . Let  $\mathcal{S}, \mathcal{L}, \mathcal{P}$  denote the set of all  $m$ ,  $(m - 1)$ , and  $(m - 2)$ -wise intersections of the hyperplanes in  $\mathcal{H}$  respectively. Note that  $\mathcal{S}, \mathcal{L}, \mathcal{P}$  correspond to points, lines, and planes in  $\mathbb{F}^m$  respectively, and furthermore,  $\mathcal{S}$  is the set of evaluation points of  $\text{GAP}_{m,d,t}^{\mathcal{H}}$ .

**Definition 7.1** (The line-point test). *Given query access to a received word  $f : \mathcal{S} \rightarrow \mathbb{F}$ , the line-point low-degree test is as follows.*

- pick a random line  $\ell$  from the set  $\mathcal{L}$  and a random point  $x$  on  $\ell$
- let  $g_\ell$  be a degree  $d$  univariate polynomial that is closest to the restriction of  $f$  on  $\ell \cap \mathcal{S}$
- accept if  $g_\ell(x) = f(x)$ , reject otherwise

◇

The plane-point test is similar to the definition above, except that instead of picking a random line, we pick a random plane  $P$  from  $\mathcal{P}$  and a random point on this plane. We accept if and only if the degree  $d$  bivariate polynomial closest to the restriction of  $f$  on  $P$  agrees with  $f$  on the chosen point.

Both of these tests have perfect completeness. Namely, if  $f = P|_{\mathcal{S}}$  for some polynomial  $P(X_1, \dots, X_m)$  of degree at most  $d$ , then the tests accept with probability 1. Thus, we are interested in the soundness of the tests.

The main result of this section is an analysis of the soundness of these tests. Specifically, we show that the plane-point test is a good local test for GAP codes with constant  $m$  and constant rate  $R < 1$ .

**Theorem 7.2.** *For all  $m \in \mathbb{N}$ ,  $R < 1$ , there exists a constant  $K_{m,R}$  such that for any  $m$ -variate GAP code  $C$  of rate  $\leq R$ , and any received word  $f$ , if the plane-point test on  $f$  has rejection probability  $p$ , then*

$$\delta_C(f) \leq K_{m,R} \cdot p.$$

For the line-point test, we are only able to show soundness at rates below some  $R_m = \exp(-m)$ .

**Theorem 7.3.** *For all  $m \in \mathbb{N}$ , there exists constants  $R_m, K_m$  such that for any  $m$ -variate GAP code  $C$  of rate  $\leq R_m$ , and any received word  $f$ , if the line-point test on  $f$  has rejection probability  $p$ , then*

$$\delta_C(f) \leq K_m \cdot p.$$

For fixed  $m$ , if the GAP code has block length  $N$ , then the query complexity of the line-point test is  $O(N^{1/m})$ , and the query complexity of the plane-point test is  $O(N^{2/m})$ . Then, a simple corollary of Theorem 7.2 is the local testability, with polynomially small query complexity, of high rate GAP codes. The first time such parameters were achieved was by Viderman [Vid15], although today constant rate locally testable codes with constant query complexity are known through the celebrated results of [DEL<sup>+</sup>22, PK22].

**Theorem 7.4.** *For any  $\lambda, R \in (0, 1)$ , there are GAP codes of arbitrarily large codeword length  $N$ , with rate  $R$  and distance  $\delta_{R,\lambda}$ , that can be locally tested with  $O(N^\lambda)$  queries.*

This follows immediately from Theorem 7.2 by taking  $m$  big enough so that  $\frac{2}{m} < \lambda$ .

## Additional notation

Before proceeding further, we set up some additional notation and conventions for this section that we repeatedly appeal to. In many of the technical statements in this section, we work with an appropriate collection of hyperplanes  $\mathcal{H}$  in  $\mathbb{F}^m$  in general position and study properties of functions on the point set  $\mathcal{S}$  obtained by taking  $m$ -wise intersection of these hyperplanes. In this setting, unless otherwise stated, all functions that we study should be thought of as functions on the set  $\mathcal{S}$  or its intersection with the appropriate domain at hand. For instance, when we say that for an  $H \in \mathcal{H}$ , let  $f_H$  be the restriction of  $f$  to the hyperplane  $H$ , we mean that  $f_H$  is a function of the form  $f_H : \mathcal{S} \cap H \rightarrow \mathbb{F}$ . We also follow this convention of denoting by  $f_H$  the restriction of a global function  $f$  to a hyperplane (or, more generally, an affine subspace)  $H$ . We also say that two functions  $f, g$  are consistent with each other if they take the same values on all points at the intersection of their domains. For example, given a function  $f$  on a hyperplane  $H$  and a function  $g$  on a hyperplane  $H'$ , we say that  $f$  and  $g$  are consistent if they agree everywhere on  $H \cap H' \cap \mathcal{S}$ . Otherwise, they are said to be inconsistent.

Throughout this section, the field  $\mathbb{F}$  is assumed to be large enough, essentially so that  $\mathbb{F}^m$

contains  $t$  hyperplanes in general position. From [Observation 3.7](#), we know that  $|\mathbb{F}| \geq t$  suffices for this.

In terms of the asymptotic behavior of the parameters involved in the results in this section, we should think of the number of variables  $m$  as an arbitrary constant, and parameters  $d, t$  and the field size  $|\mathbb{F}|$  to be growing asymptotically. We also note that in our proofs, we haven't tried to achieve the best possible quantitative bounds on various parameters involved. We anticipate that these bounds can be tightened further using a more careful analysis.

## Organization

In [Section 7.1](#), we prove a local characterization lemma which states that the property of being degree  $d$  on restrictions to each hyperplane is a characterization of degree  $d$  polynomials. More precisely, if there are  $(m - 1)$ -variate degree  $d$  polynomials  $g_H$  for each  $H \in \mathcal{H}$ , such that each pair  $(g_H, g_{H'})$  are consistent on their intersection, then there is a  $m$ -variate degree  $d$  polynomial  $f$  such that  $f$  is consistent with  $g_H$  for all  $H \in \mathcal{H}$ .

We then prove the soundness of the plane-point test in [Section 7.2](#) and that of the line-point test in [Section 7.3](#), [Section 7.4](#) and [Section 7.5](#). [Section 7.5](#) contains a general divisibility lemma ([Lemma 7.12](#)) relating divisibility on hyperplane restrictions with global divisibility, a statement of independent interest.

The proofs of soundness of both tests use a robust version of the local characterization lemma, which states that if almost all (though not necessarily all) of the hyperplane polynomials are consistent with each other, then they can be stitched together into a global degree  $d$  polynomial that is consistent with many of them. We have two alternate versions of the robust local characterization, one with better parameters that works for  $m \geq 3$  ([Section 7.2](#)). one with worse parameters that works for  $m \geq 2$  ([Section 7.4](#)).

The first robust local characterization lemma is based on the Raz-Safra [[RS97](#)] and Ben-Sasson-Sudan [[BSS06](#)] hyperplane consistency graph method. The second one is more algebraic, and uses a refined version of the Polischuk-Spielman method [[PS94](#)] based on Bezout's theorem, paying careful attention to intersection multiplicities.

### 7.1 Local characterization

If  $f$  is a  $m$ -variate degree  $d$  polynomial, the restrictions of  $f$  to lines are univariate polynomials of degree at most  $d$ . In this section, we prove something of a converse to that statement. Formally, we show the following.

**Theorem 7.5.** *Let  $m, d, t \in \mathbb{N}$  be such that  $t \geq d + m$ ,  $\mathbb{F}$  be a large enough field, and let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of hyperplanes in general position in  $\mathbb{F}^m$ . Let  $\mathcal{S} \subseteq \mathbb{F}^m$  denote the set of all  $m$ -wise intersection points of the hyperplanes in  $\mathcal{H}$  and let  $\mathcal{L}$  be the set of lines determined by the  $(m - 1)$ -wise*



intersection of hyperplanes in  $\mathcal{H}$ . For every line  $\ell$  in  $\mathcal{L}$ , let  $g_\ell$  be a univariate polynomial of degree at most  $d$  defined on  $\ell$ .

If for all distinct pairs of lines  $\ell_1, \ell_2 \in \mathcal{L}$ , the line polynomials  $g_{\ell_1}$  and  $g_{\ell_2}$  agree on the point  $\ell_1 \cap \ell_2 \in \mathcal{S}$ , then there is an  $m$ -variate degree  $d$  polynomial  $f$  such that for every line  $\ell \in \mathcal{L}$ , the restriction of  $f$  to the line  $\ell$  equals  $g_\ell$ .

This statement talks about deducing the  $m$ -dimensional structure from 1-dimensional restrictions. The proof will recursively use the following lemma, which deduces  $m$ -dimensional structure from  $(m - 1)$ -dimensional restrictions.

**Lemma 7.6.** *Let  $m, d, t \in \mathbb{N}$  be such that  $t \geq d + m$  and  $\mathbb{F}$  be a large enough field. Let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of hyperplanes in general position in  $\mathbb{F}^m$  and let  $\mathcal{S} \subseteq \mathbb{F}^m$  denote the set of all  $m$ -wise intersection points of the hyperplanes in  $\mathcal{H}$ . For every hyperplane  $H$  in  $\mathcal{H}$ , let  $g_H : H \rightarrow \mathbb{F}$  be a degree  $d$  polynomial on  $(m - 1)$  variables.*

*If for all distinct pairs of hyperplanes  $H, H' \in \mathcal{H}$ , the polynomials  $g_H$  and  $g_{H'}$  agree on  $H \cap H' \subseteq \mathcal{S}$ , then there is an  $m$ -variate degree  $d$  polynomial  $f : \mathcal{S} \rightarrow \mathbb{F}$  such that for every hyperplane  $H \in \mathcal{H}$ , the restriction of  $f$  to  $H$  equals  $g_H$ .*

*Proof.* For every  $i \in \{1, 2, \dots, t\}$ , let  $\mathcal{H}_i$  be the set  $\{H_j \cap H_i : j \neq i\}$  of subspaces of  $\mathbb{F}^m$  of co-dimension 2. However, we will instead think of  $\mathcal{H}_i$  as a set of  $(t - 1)$  hyperplanes in the  $(m - 1)$  dimensional ambient space  $\mathbb{F}^{m-1}$  (identified with  $H_i$ ). Moreover, since  $\mathcal{H}$  is a set of hyperplanes in  $\mathbb{F}^m$  in general position, we have that for every  $i$ ,  $\mathcal{H}_i$  is a set of hyperplanes in general position in  $\mathbb{F}^{m-1}$ . Let  $\mathcal{S}_i$  be the points given by  $(m - 1)$ -wise intersection of hyperplanes in  $\mathcal{H}_i$ . Note that  $\mathcal{S}_i$  is precisely the subset of points of  $\mathcal{S}$  that are contained in the hyperplane  $H_i$ .

Let  $v : \mathcal{S} \rightarrow \mathbb{F}$  be a function defined as follows. For any  $\mathbf{a} \in \mathcal{S}$ , let  $\{H_{i_1}, H_{i_2}, \dots, H_{i_m}\} \subseteq \mathcal{H}$  be the hyperplanes that intersect at  $\mathbf{a}$ . Then,  $v(\mathbf{a})$  is defined to be equal to  $g_{H_{i_j}}(\mathbf{a})$  for  $j \in \{1, 2, \dots, m\}$ . Since the hyperplane polynomials are mutually consistent with each other, the function  $v$  is well-defined. We now define the global degree  $d$  polynomial  $f$  using the function  $v$ .

**Defining the global function:** Let  $\mathcal{H}'$  be an arbitrary subset of  $\mathcal{H}$  of size exactly  $(d + m)$ , and let  $\mathcal{S}'$  be the set of points obtained by taking  $m$ -wise intersection of hyperplanes in  $\mathcal{H}'$ . Clearly,  $\mathcal{S}'$  is a subset of  $\mathcal{S}$ . Let  $f$  be the *unique* degree  $d$  polynomial on  $m$ -variables such that  $f(\mathbf{a}) = v(\mathbf{a})$  for every  $\mathbf{a} \in \mathcal{S}'$ . The existence and uniqueness of  $f$  again follows from the fact that the set  $\mathcal{S}'$  is an interpolating set for degree  $d$  polynomials on  $m$  variables ([Theorem 3.3](#)).

We now argue that for every  $i$ , the restriction of the global function  $f$  to any hyperplane  $H_i \in \mathcal{H}$  equals the function  $g_{H_i}$ . We do this in two steps.

**Case I -  $H_i \in \mathcal{H}'$ :** Let  $\mathcal{S}'_i$  of points obtained by taking the intersection of any  $(m - 1)$  hyperplanes in  $\mathcal{H}' \setminus \{H_i\}$  together with  $H_i$ , and let  $f_{H_i}$  be the degree  $d$  polynomial (on  $(m - 1)$ -variables) obtained by restricting  $f$  to  $H_i$ . Note that  $\mathcal{S}'_i$  is a subset of  $\mathcal{S}_i \cap \mathcal{S}'$  and is an interpolating set for



degree  $d$  polynomials on the space  $H_i$  since it is the set of  $(m - 1)$ -wise intersections of  $(d + m - 1)$  hyperplanes in general position in this  $(m - 1)$ -dimensional ambient space ([Theorem 3.3](#)). From the definition of  $f$ , and the properties of  $f_{H_i}$  discussed earlier, we know that for every point  $\mathbf{a} \in \mathcal{S}'_i$ ,  $f(\mathbf{a}) = f_{H_i}(\mathbf{a}) = g_{H_i}(\mathbf{a}) = v(\mathbf{a})$ . Note that we use the fact that  $\mathcal{S}'_i$  is a subset of  $\mathcal{S}'$  for this. But,  $f_{H_i}$  and  $g_{H_i}$  are both degree  $d$  functions on the space  $H_i$  that agree everywhere on the set  $\mathcal{S}'_i$  which is an interpolating set for such functions. Thus, they must be equal to each other.

**Case II -  $H_i \notin \mathcal{H}'$  :** Now, the set  $\mathcal{S}'_i$  of points obtained by taking the intersection any  $(m - 1)$  hyperplanes in  $\mathcal{H}'$  together with  $H_i$ .  $f_{H_i}$  is the restriction of  $f$  to  $H_i$  as defined for Case I. Let  $\mathbf{a} \in \mathcal{S}'_i$  be any point and let  $H_j \in \mathcal{H}'$  be such that  $\mathbf{a} \in H_j$ . So, we have  $g_{H_i}(\mathbf{a}) = g_{H_j}(\mathbf{a}) = f_{H_j}(\mathbf{a})$ , since  $g_{H_i}$  and  $g_{H_j}$  are consistent with each other and  $g_{H_j} = f_{H_j}$  from the case above. But then,  $f_{H_j}$  and  $f_{H_i}$  are both restrictions of the same global function  $f$ , and  $\mathbf{a}$  is a common point in the domain of these functions. So,  $f_{H_i}(\mathbf{a}) = f_{H_j}(\mathbf{a})$ . Therefore, we get that  $g_{H_i}(\mathbf{a}) = g_{H_j}(\mathbf{a}) = f_{H_i}(\mathbf{a})$ . Hence, we have two  $(m - 1)$ -variate degree  $d$  polynomials on  $g_{H_i}$  and  $f_{H_i}$  on  $H_i$  that are equal to each other on the set  $\mathcal{S}'_i$ , which is again an interpolating set for such polynomials. Thus, they must be equal to each other.  $\square$

We can now complete the proof of the local characterization from 1-dimensional restrictions.

*Proof of [Theorem 7.5](#).* The proof of the theorem is via induction on  $m$ , with the  $m = 2$  case of [Lemma 7.6](#) being the base case. So, we now assume that  $m > 2$  and the theorem is true in  $(m - 1)$  dimensions and prove it for  $m$  dimensions. The overall structure of the proof follows that of the proof of [Lemma 7.6](#).

For every  $i \in \{1, 2, \dots, t\}$ , let  $\mathcal{H}_i$  be the set  $\{H_j \cap H_i : j \neq i\}$  of subspaces of  $F^m$  of co-dimension 2, which as before, we think of  $\mathcal{H}_i$  as a set of  $(t - 1)$  hyperplanes in the  $(m - 1)$  dimensional ambient space  $\mathbb{F}^{m-1}$  (identified with  $H_i$ ). Moreover, since  $\mathcal{H}$  is a set of hyperplanes in  $\mathbb{F}^m$  in general position, we have that for every  $i$ ,  $\mathcal{H}_i$  is a set of hyperplanes in general position in  $\mathbb{F}^{m-1}$ . Let  $\mathcal{S}_i$  be the points given by  $(m - 1)$ -wise intersection of hyperplanes in  $\mathcal{H}_i$ ,  $\mathcal{L}_i$  be the lines given by  $(m - 2)$ -wise intersection of hyperplanes in  $\mathcal{H}_i$ . Note that  $\mathcal{S}_i$  is precisely the subset of points of  $\mathcal{S}$  that are contained in the hyperplane  $H_i$  and  $\mathcal{L}_i$  is precisely the subset of lines of  $\mathcal{L}$  that are contained in the hyperplane  $H_i$ . Now, clearly, the set of line polynomials for lines in  $\mathcal{L}_i$  also satisfies the hypothesis of the theorem, namely that any of two of these polynomials are consistent. But now, we are in  $(m - 1)$  dimensions. Hence, by the induction hypothesis, we get that there is a degree  $d$  polynomial  $f_{H_i}$  such that for every line  $\ell \in \mathcal{L}_i$ , the restriction of  $f_{H_i}$  on  $\ell$  equals the polynomial  $g_\ell$ .

Moreover, since all the line polynomials are consistent with each other and the hyperplane polynomials  $f_{H_i}$  are consistent with all line polynomials on the respective hyperplane, we get that

the hyperplane polynomials  $f_{H_i}$  obtained above must also be consistent with each other. Thus, we are in the setting of [Lemma 7.6](#), and get that there is an  $m$ -variate degree  $d$  polynomial  $f$  that is consistent with all the hyperplane polynomials  $f_{H_i}$ . Since these hyperplane polynomials are consistent with the line polynomials, we get that  $f$  is consistent with all the line polynomials, thereby completing the proof of the theorem.  $\square$

## 7.2 Soundness of the plane-point test

In this subsection, we prove [Theorem 7.2](#).

We begin by proving a robust analogue of the local characterization lemma, [Lemma 7.6](#), for  $m \geq 3$ . Here, in the hypothesis, we are only guaranteed that the low-degree functions on *most* (as opposed to all) pairs of hyperplanes are consistent with each other. The conclusion again is in the spirit of that in [Lemma 7.6](#) and guarantees a single low degree polynomial *explaining* most of these hyperplane polynomials.

### 7.2.1 Robust local characterization

**Lemma 7.7** (Robust Local Characterization). *Let  $m \in \mathbb{N}_{\geq 3}$ , and  $\rho, \beta, c \in (0, 1)$  be constants. Let  $t$  be a growing parameter and  $d = \rho t$ . Let  $H_1, \dots, H_t \subset \mathbb{F}_q^m$  be hyperplanes in general position, and let  $g_1, \dots, g_t$  be corresponding  $(m-1)$ -variate polynomials of degree at most  $d$ .*

*Suppose at least  $1 - c$  fraction of pairs of the  $g_i$  are consistent, and*

$$1 - \rho > \sqrt{2c} + \beta. \tag{7.8}$$

*Then, there exists some  $m$ -variate polynomial  $h$  such that  $h|_{H_i} \equiv g_i$  for at least a  $1 - c/\beta$  fraction choices of  $i \in [t]$ .*

*Proof.* The proof closely follows [\[BSS06\]](#). Define an inconsistency graph  $G$  as follows. The vertices are  $[t]$ , and there is an edge between  $i$  and  $j$  if and only if  $g_i$  and  $g_j$  are inconsistent (they are not equal to each other on  $H_i \cap H_j$ ). The goal is to find a large independent set in  $G$  (corresponding to a large set of pairwise consistent polynomials) and apply [Lemma 7.6](#). We first show that  $G$  has the following property.

**Claim 7.9.** *If there is an edge between  $i$  and  $j$ , then either the degree of  $i$  or the degree of  $j$  is at least  $\frac{1}{2}(1 - \rho)t(1 - o(1))$ .*

Suppose  $i$  and  $j$  are inconsistent, then  $g_i$  and  $g_j$  restricted to  $H_i \cap H_j$  are distinct codewords of  $\text{GAP}_{m-2,d,t-2}$ , and hence disagree on at least  $\Delta$  points, where  $\Delta$  is the distance of  $\text{GAP}_{m-2,d,t-2}$ . Let  $S$  be the subset of points where  $g_i$  and  $g_j$  differ. Recall  $m$ -wise intersections of hyperplanes specify the points, and thus, we can think of  $S$  as being a subset of  $\binom{[t] \setminus \{i,j\}}{m-2}$  (since points in  $S$  must lie on

$H_i$  and  $H_j$ ). Let  $T$  be the set of indices of the hyperplanes that contain some disagreement between  $g_i$  and  $g_j$ . For any  $k \in T$ ,  $g_k$  must be inconsistent with at least one of  $g_i$  and  $g_j$ . We now bound the size of  $T$ .

$$\begin{aligned}
|S| &\leq \binom{|T|}{m-2} \\
\implies \Delta &\leq \binom{|T|}{m-2} \\
\implies (1-\rho)^{m-2} \binom{t-2}{m-2} &\leq \binom{|T|}{m-2} \\
\implies (1-\rho)^{m-2} &\leq \frac{|T|!}{(|T|-m+2)!} \cdot \frac{(t-2-m+2)!}{(t-2)!} \\
\implies (1-\rho)^{m-2} &\leq \prod_{i=0}^{m-3} \frac{|T|-i}{t-2-i} \\
\implies (1-\rho)^{m-2} &\leq \left( \frac{|T|}{t-2} \right)^{m-2} \\
\implies (1-\rho)(t-2) &\leq |T|.
\end{aligned}$$

The claim then follows from applying the Pigeonhole Principle.

Thus, the vertices with degree  $< \frac{1}{2}(1-\rho)t(1-o(1))$  form an independent set,  $I$ . Let  $z$  be the fraction of vertices of degree at least  $\frac{1}{2}(1-\rho)t(1-o(1))$ . Counting degrees, we have that  $\frac{zt^2}{2}(1-\rho)(1-o(1)) \leq 2c\binom{t}{2}$ , which implies  $z \leq \frac{2c}{1-\rho} + o(1)$ .

To apply the local characterization lemma (Lemma 7.6), we need  $(1-z)t \geq d+m$ , which is true when

$$\begin{aligned}
1-z &\geq \rho + m/t \\
\iff 1-\rho &\geq \frac{2c}{1-\rho} + o(1) \\
\iff (1-\rho)^2 &\geq 2c + o(1) \\
\iff 1-\rho &\geq \sqrt{2c} + o(1)
\end{aligned}$$

which is satisfied by the assumption of the theorem.

Applying Lemma 7.6, we have that there exists a  $m$ -variate polynomial  $h$  consistent with each of the polynomials  $g_i$  for  $i \in I$ , which is a set of fractional size at least  $1-z$ , we'll show that it must be consistent with an even larger set of polynomials.

Let  $i \in [t]$ , and suppose  $g_i$  is consistent with at least  $d+1$  polynomials in  $I$ . This implies that  $g_i|_{H_j} \equiv h|_{H_i \cap H_j}$  for at least  $d+1$  values of  $j$ , which implies  $h|_{H_i} \equiv g_i$  since both of these polynomials have degree at most  $d$ . Hence,  $h$  is consistent with  $g_i$  as well. For any  $i$ , let  $X_i$  be the number of polynomials that are inconsistent with  $i$ , we have  $\mathbb{E}_{i \in [t]}[X_i] = ct$ . Note that  $t - X_i$  is the number

of polynomials consistent with  $i$ . If  $t - X_i \geq zt + d + 1$ , then  $g_i$  is consistent with at least  $d + 1$  polynomials in  $I$ , hence also consistent with  $h$ . We'll show that this is true for many  $i$ . By Markov's inequality,

$$\begin{aligned} \Pr(X_i > (1 - z)t - d - 1) &\leq \frac{ct}{(1 - z)t - d - 1} \\ &\leq \frac{c}{1 - z - \rho - 1/t} \\ &= \frac{c}{1 - z - \rho - 1/t} \\ &\leq \frac{c}{\beta}. \end{aligned}$$

Note that we get the last inequality since  $z \leq \frac{2c}{1-\rho} \leq \frac{2c}{\sqrt{2c}} = \sqrt{2c}$ , and  $1 - \rho > \sqrt{2c} + \beta$ . Thus, at least a  $1 - c/\beta$  fraction of the  $g_i$ s is consistent with  $h$  as required.  $\square$

### 7.2.2 Analysis of the plane-point test

We now analyze the plane-point test and prove the following lemma.

**Lemma 7.10.** *Let  $m \in \mathbb{N}_{\geq 2}$ , and  $\rho, p, \beta \in (0, 1)$  be constants. Let  $t$  be a sufficiently large growing parameter, and let  $d/t \leq \rho$ . Define the sequences  $(Q_{m,\rho})_{m \geq 3}$  and  $(P_{m,\rho})_{m \geq 3}$  as follows.*

$$P_{m,\rho} = \begin{cases} 1 & m = 2 \\ \frac{(1-\rho-\beta)^2(1-\rho)}{4} \cdot P_{m-1,\rho} & m \geq 3 \end{cases}$$

$$Q_{m,\rho} = \begin{cases} 1 & m = 2 \\ \frac{3}{P_{m-1,\rho}(1-\rho)^{m-2}\beta} & m \geq 3 \end{cases}$$

Let  $C = \text{GAP}_{m,d,t}$ , and let  $f$  be a received word. Suppose the plane-point test accepts  $f$  with at least probability at least  $1 - p$ . Then, if  $p \leq P_{m,\rho}$ , then

$$\delta_C(f) \leq Q_{m,\rho}p$$

*Proof.* The proof of the lemma is by induction on  $m$ .

For the base case,  $m = 2$ . We query all the points since the entire code lies on a two-dimensional plane. Let  $g$  be the closest bivariate polynomial to  $f$ , then the acceptance probability is  $\mathbb{E}_x[\mathbb{1}_{g(x)=f(x)}] = \alpha(g, f)$ . Thus, we have  $1 - p \geq \alpha(g, f)$ , so  $\delta_C(f) \leq p$  as required.

Let  $m \geq 3$  for the inductive step, and suppose the planes test works for the  $(m - 1)$ -variate GAP code. Let  $f_i = f|_{H_i}$ . We give a brief proof overview below.

1. Using the inductive hypothesis, find  $(m - 1)$ -variate polynomials  $g_1, g_2, \dots, g_t$  such that  $g_i$  is

close to  $f_i$ .

2. Then, we will use [Lemma 7.7](#) (Robust Characterization Lemma) on  $g_1, \dots, g_t$  to extract an  $m$ -variate polynomial  $h$  consistent with most of the  $g_i$ . To do so, we must show that many pairs  $g_i$  and  $g_j$  are consistent.
3. Finally, we will bound the distance between  $h$  and  $f$ .

**Finding the  $g_i$ .** Let  $p_i$  be the rejection probability given that the selected plane lies on  $H_i$ . Call  $i$  *good* if  $p_i \leq P_{m-1,\rho}$ , that is  $p_i$  satisfies the conditions of the  $(m-1)$ -variate version of the theorem. A simple calculation shows that the average rejection probability on each hyperplane is the overall rejection probability, i.e.  $\mathbb{E}[p_i] = p$ . Applying Markov's Inequality, we have

$$\Pr[i \text{ is bad}] \leq \Pr[p_i > P_{m-1,\rho}] \leq \frac{p}{P_{m-1,\rho}}.$$

Let  $\gamma = \frac{p}{P_{m-1,\rho}}$ . Then, at least a  $1 - \gamma$  fraction of the  $i$ s are good. For each  $i \in [t]$ , let  $g_i$  be the  $(m-1)$ -variate polynomial obtained by applying the  $(m-1)$ -variate version of the theorem if  $i$  is good and the 0 polynomial otherwise.

Note that if  $i$  is bad,  $g_i$  might be far from  $f_i$ . However, since most of the indices  $i$  are good, we will show that the *average* agreement between  $g_i$  and  $f_i$  is large. To simplify the notation a bit, let  $\alpha_i = \alpha(g_i, f_i)$ , and  $\eta_i = 1 - Q_{m-1,\rho}p_i$ .

$$\begin{aligned} \mathbb{E}_i[\alpha(g_i, f_i)] &= \mathbb{E}_i[\alpha_i] \\ &= \mathbb{E}_i[\alpha_i - \eta_i + \eta_i] \\ &= \mathbb{E}_i[\eta_i] + \mathbb{E}[\alpha_i - \eta_i] \\ &= 1 - Q_{m-1,\rho}p + \mathbb{E}[\alpha_i - \eta_i]. \end{aligned}$$

Focusing on the  $\mathbb{E}[\alpha_i - \eta_i]$ , we have that for each good  $i$ ,  $\alpha_i \geq \eta_i$  by the distance guarantee of the  $(m-1)$ -variate version of the theorem, and for each  $i$  that is not good (of which there are most a  $\gamma$  fraction),  $\alpha_i - \eta_i \geq Q_{m-1,\rho}p_i - 1 \geq Q_{m-1,\rho}P_{m-1,\rho} - 1$ , where we get the first inequality since the agreement is always at least zero, and the second inequality because  $i$  is not good. Thus, we have

$$\mathbb{E}_i[\alpha(g_i, f_i)] \geq 1 - Q_{m-1,\rho}p + \gamma(Q_{m-1,\rho}P_{m-1,\rho} - 1) = 1 - \gamma.$$

**Many pairs are consistent.** Now that we know the average agreement between the  $g_i$  and  $f_i$  is large, we will show that many pairs,  $(g_i, g_j)$ , are consistent. Towards this goal, we first show that

the average agreement of  $g_i$  and  $g_j$  is high.

$$\begin{aligned}
\mathbb{E}_{i,j}[\alpha(g_i, g_j)] &= \mathbb{E}_x \mathbb{E}_{H_i, H_j \ni x} [\mathbb{1}_{g_i(x)=g_j(x)}] \\
&\geq \mathbb{E}_x \mathbb{E}_{H_i, H_j \ni x} [\mathbb{1}_{g_i(x)=f(x)} \mathbb{1}_{g_j(x)=f(x)}] \\
&= \mathbb{E}_x \left[ \mathbb{E}_{i: x \in H_i} [\mathbb{1}_{g_i(x)=f(x)}] \right]^2 \\
&\geq \left[ \mathbb{E}_x \mathbb{E}_{i: x \in H_i} [\mathbb{1}_{g_i(x)=f(x)}] \right]^2 \\
&= [\mathbb{E}_i [\alpha(g_i, f_i)]]^2 \\
&\geq (1 - \gamma)^2 \\
&> 1 - 2\gamma.
\end{aligned}$$

Let  $c$  be the fraction of inconsistent pairs. If a pair is consistent, they have agreement 1. On the other hand, if a pair is not consistent, they have agreement at most  $1 - \delta_{m-2}$ , which is  $1 - (1 - \rho)^{m-2}$ . Thus,

$$\begin{aligned}
1 - 2\gamma &< 1 - c + c(1 - (1 - \rho)^{m-2}) \\
\implies c &< \frac{2\gamma}{(1 - \rho)^{m-2}}.
\end{aligned}$$

**Applying Lemma 7.7.** To apply Lemma 7.7, we need to verify condition 7.8 holds.

In particular, we need

$$\begin{aligned}
(1 - \rho - \beta)^2 &\geq 2c \\
\iff \frac{(1 - \rho - \beta)^2}{2} &\geq \frac{2p}{P_{m-1,\rho}(1 - \rho)^{m-2}} \\
\iff \frac{(1 - \rho - \beta)^2(1 - \rho)^{m-2}}{4} \cdot P_{m-1,\rho} &\geq p \\
\iff P_{m,\rho} &\geq p.
\end{aligned}$$

Applying Lemma 7.7, we obtain a  $m$ -variate polynomial  $h$  consistent with at least a  $1 - c/\beta$  fraction of the  $g_i$ . Finally, we show that  $h$  is not too far from  $f$ .

**Agreement between  $h$  and  $f$ .** Let  $h_i$  be the restriction of  $h$  to  $H_i$ . Let  $X_i = \alpha(h_i, f_i)$ , and  $Y_i = \alpha(f_i, g_i)$ .

$$\alpha(h, f) = \mathbb{E}[X_i] = \mathbb{E}[X_i + Y_i - Y_i] > 1 - \gamma + \mathbb{E}[X_i - Y_i] \geq 1 - \gamma - c/\beta.$$

where we get the last inequality since  $g_i \neq h_i$  on at most a  $c/\beta$  fraction of  $i \in [t]$ , and the different of two number each in  $(0, 1)$  is at least  $-1$ .

Substituting bound on  $c$ ,

$$\begin{aligned}\alpha(h, f) &\geq 1 - \gamma - \frac{2\gamma}{(1-\rho)^{m-2}\beta} \\ &= 1 - p \cdot \frac{1}{P_{m-1,\rho}} \left(1 + \frac{2}{(1-\rho)^{m-2}\beta}\right) \\ &> 1 - p \left(\frac{3}{P_{m-1,\rho}(1-\rho)^{m-2}\beta}\right) \\ &= 1 - pQ_{m,\rho}\end{aligned}$$

Therefore,  $\delta_C(f) \leq pQ_{m,\rho}$ , as required.

**Bounds on  $P_{m,\rho}$ ,  $Q_{m,\rho}$ .**

$$P_{m,\rho} = \prod_{i=0}^{m-3} \left( \frac{(1-\rho-\beta)^2(1-\rho)^{m-2-i}}{4} \right) \geq \frac{(1-\rho-\beta)^{2m-2}(1-\rho)^{m^2}}{4^{m-2}}$$

Then,

$$Q_{m,\rho} = \frac{3}{P_{m-1,\rho}(1-\rho)^{m-2}\beta} \leq \frac{3 \cdot 4^{m-3}}{\beta(1-\rho-\beta)^{2m-4}(1-\rho)^{(m-1)^2+m-2}}$$

Note that the bounds on  $P_{m,\rho}$  and  $Q_{m,\rho}$  are exponential in  $m$ ; however, since  $m, \rho$  and  $\beta$  and constants,  $P_{m,\rho}$  and  $Q_{m,\rho}$  are also constants.

□

To prove [Theorem 7.2](#), apply [Lemma 7.10](#) with  $\rho = R^{1/m}$ ,  $\beta = \frac{1}{2}(1-\rho)$ , and  $C_{m,R} = \max\{1/P_{m,\rho}, Q_{m,\rho}\}$ .

### 7.3 Soundness of the line-point test

In this subsection, we prove [Theorem 7.3](#) on the soundness of the line-point test. The proof will use the soundness of the plane-point test, along with a robust local characterization of bivariate polynomials based on restrictions to lines. We now state this robust local characterization statement.

**Theorem 7.11.** *There exists a constant  $\varepsilon_0 \in (0, 1)$  such that for all  $\varepsilon \in [0, 1)$ , with  $\varepsilon < \varepsilon_0$ ,  $d, t \in \mathbb{N}$  with  $t > \max\{\frac{d+2}{1-\sqrt{\varepsilon}}, \frac{2d}{1-3\sqrt{\varepsilon}}\}$ , and sufficiently large field  $\mathbb{F}$ , the following is true.*

*Let  $\mathcal{H} = \{\ell_1, \dots, \ell_t\}$  be a set of lines in general position in  $\mathbb{F}^2$ . Let  $\mathcal{S} \subseteq \mathbb{F}^m$  denote the set of all pairwise intersection points of the lines in  $\mathcal{H}$ . For every line  $\ell$  in  $\mathcal{H}$ , let  $g_\ell : \ell \rightarrow \mathbb{F}$  be a degree  $d$  univariate polynomial.*

If the polynomials  $g_\ell$  and  $g_{\ell'}$  agree with each other on  $\ell \cap \ell'$  for at least  $(1 - \varepsilon)\binom{t}{2}$  pairs  $\{\ell, \ell'\} \subseteq \mathcal{H}$ , then there is a bivariate degree  $d$  polynomial  $f : \mathcal{S} \rightarrow \mathbb{F}$  and a set  $\tilde{\mathcal{H}} \subseteq \mathcal{H}$  of size at least  $(1 - 2\varepsilon)t$  such that for every  $\ell \in \tilde{\mathcal{H}}$ , the restriction of  $f$  on  $\ell$  equals  $g_\ell$ .

This can be viewed as a robust analogue of [Lemma 7.6](#) for the case  $m = 2$ . We already proved such a robust analogue for  $m > 2$  in [Lemma 7.7](#) during the analysis of the plane-point test. The proof appears in [Section 7.4](#).

The key ingredient of this robust local characterization is a general divisibility lemma, given below.

**Lemma 7.12.** *Let  $\mathbb{F}$  be any field,  $e, d, t, m$  be natural numbers, and  $\mathcal{H}$  be a set of  $t$  hyperplanes in  $\mathbb{F}^m$  in general position.  $E(X_1, \dots, X_m)$  and  $P(X_1, \dots, X_m)$  are polynomials in  $\mathbb{F}[X_1, \dots, X_m]$  of degree equal to  $e$  and  $d$  respectively. For any hyperplane  $H \subseteq \mathbb{F}^m$ , let  $E_H(Z_1, \dots, Z_{m-1})$  and  $P_H(Z_1, \dots, Z_{m-1})$  respectively be the  $m - 1$ -variate polynomials in  $\mathbb{F}[Z_1, \dots, Z_{m-1}]$  obtained by the restriction of  $E$  and  $P$  to the hyperplane  $H$ .*

*If  $t > 2d + e$  and for all hyperplanes  $H \in \mathcal{H}$ , there exists a polynomial  $Q_H(Z_1, \dots, Z_{m-1})$  such that*

$$P_H(Z_1, \dots, Z_{m-1}) = E_H(Z_1, \dots, Z_{m-1}) \cdot Q_H(Z_1, \dots, Z_{m-1}),$$

*then, there exists a polynomial  $Q(X_1, \dots, X_m) \in \mathbb{F}[X_1, \dots, X_m]$  such that*

$$P(X_1, \dots, X_m) = E(X_1, \dots, X_m) \cdot Q(X_1, \dots, X_m).$$

The proof appears in [Section 7.5](#). For the robust local characterization of bivariate polynomials with lines, [Theorem 7.11](#), we only need the  $m = 2$  case of this lemma. Nevertheless we state and prove the result for general  $m$  because it seems to be a basic statement of independent interest.

In terms of the quantitative parameters, the following simple example shows that at least for  $m = 2$ , the bounds on  $t$  in the lemma are essentially tight.

**An almost tight example:** Let  $E(X_1, X_2)$  be the polynomial  $X_1$  of degree 1 and let  $P(X_1, X_2)$  be the polynomial  $\prod_{i=1}^d (X_2 - i^2)$  of degree  $d$ . Let us now consider the lines  $\ell_i$  defined by solutions to the linear equation  $X_2 = i \cdot X_1 + i^2$  as  $i$  varies in the set  $\{-1, 1, -2, 2, \dots, -d, d\}$ . Clearly, for each of these lines, the restriction of  $P$  on the line is divisible by the restriction of  $E$  on the line. However,  $P$  is not divisible by  $E$  in the ring  $\mathbb{F}[X_1, X_2]$ .

[Lemma 7.12](#) says that if the number of such lines exceeds  $2d + 1$ , then divisibility of restrictions guarantees divisibility as bivariate polynomials and this example shows that the parameter  $t$  in the statement of [Lemma 7.12](#) must be at least  $2d$  for this to happen.



### 7.3.1 Proof of Theorem 7.3 in two dimensions

We recall the  $m = 2$  case of Theorem 7.3. Throughout this section, we denote the fractional agreement between two functions  $f, g$  by  $\alpha(f, g)$ .

**Theorem 7.13.** *There exists a constant  $\varepsilon_0 \in (0, 1)$  such that for all  $\varepsilon \in [0, 1)$  with  $\varepsilon < \varepsilon_0$ , natural numbers  $t, d$  with  $t > \max\{\frac{d+2}{1-\sqrt{\varepsilon}}, \frac{2d}{1-3\sqrt{\varepsilon}}\}$ , and sufficiently large field  $\mathbb{F}$ , the following is true.*

*Let  $\mathcal{H} = \{\ell_1, \dots, \ell_t\}$  be a set of lines in general position in  $\mathbb{F}^2$ . Let  $\mathcal{S} \subseteq \mathbb{F}^2$  denote the set of all pairwise intersection points of the lines in  $\mathcal{H}$ . If  $f : \mathcal{S} \rightarrow \mathbb{F}$  is a function for which the line-point test on  $f$  accepts with probability at least  $(1 - \varepsilon)$ , then there exists a degree  $d$  polynomial  $h \in \mathbb{F}[X, Y]$  such that  $f$  and  $h$  agree on at least  $(1 - 5\varepsilon - o(1))$  fraction of points in  $\mathcal{S}$ .*

*Proof.* For every line  $\ell \in \mathcal{H}$ , let  $g_\ell$  be the degree  $d$  polynomial that is closest to the restriction of  $f$  on  $\ell$  (henceforth denoted by  $f_\ell$  for brevity). Let  $\alpha(f_\ell, g_\ell)$  denote the fractional agreement between  $f_\ell$  and  $g_\ell$ . Since the line-point test passes with probability  $(1 - \varepsilon)$ , we have that

$$\mathbb{E}_\ell[\alpha(f_\ell, g_\ell)] = \mathbb{E}_{\ell, x \in \ell}[\mathbb{1}_{g_\ell(x)=f(x)}] \geq (1 - \varepsilon).$$

From the above statement, we would like to conclude that for most pairs of lines  $\ell, \ell'$ , the polynomials  $g_\ell$  and  $g_{\ell'}$  are consistent with each other, i.e. they agree on  $\ell \cap \ell'$ . To this end, we have the following sequence of inequalities.

$$\begin{aligned} \mathbb{E}_{\ell, \ell'}[\mathbb{1}_{g_\ell(\ell \cap \ell')=g_{\ell'}(\ell \cap \ell')}] &= \mathbb{E}_x \mathbb{E}_{\ell, \ell' \ni x}[\mathbb{1}_{g_\ell(x)=g_{\ell'}(x)}] \\ &\geq \mathbb{E}_x \mathbb{E}_{\ell, \ell' \ni x}[\mathbb{1}_{g_\ell(x)=f(x)} \cdot \mathbb{1}_{g_{\ell'}(x)=f(x)}] \\ &= \mathbb{E}_x \left( \mathbb{E}_{\ell \ni x}[\mathbb{1}_{g_\ell(x)=f(x)}] \right)^2 \\ &\geq \left( \mathbb{E}_x \mathbb{E}_{\ell \ni x}[\mathbb{1}_{g_\ell(x)=f(x)}] \right)^2 \\ &\geq (1 - \varepsilon)^2 \\ &\geq 1 - 2\varepsilon. \end{aligned}$$

Here,  $x$  ranges over the points in  $\mathcal{S}$  and the lines  $\ell, \ell'$  range over lines in the set  $\mathcal{L}$ , and the third inequality relies on an application of the Cauchy-Schwartz inequality.

Thus, we have that for at least  $(1 - 2\varepsilon - 1/t)$  fraction of all pairs of distinct lines  $\ell, \ell'$  in  $\mathcal{L}$  are consistent with each other. Therefore, from the robust local characterization theorem (Theorem 7.11) to this setting (we note that the inequalities needed in the hypothesis are satisfied for our choice of parameters here), we get that there is bivariate polynomial  $h$  of degree at most  $d$

such that for at least  $(1 - 4\varepsilon - 2/t)t$  lines  $\ell$  in  $\mathcal{H}$ , we have that

$$h_\ell = g_\ell.$$

Thus, on lines  $\ell$  where  $h_\ell = g_\ell$ , we have that  $\alpha(h_\ell, f_\ell) = \alpha(g_\ell, f_\ell)$  and on the remaining  $(4\varepsilon + 2/t)$  fraction of lines, we have that  $\alpha(h_\ell, f_\ell) \geq 0$  (and  $\alpha(g_\ell, f_\ell)$  could be as large as 1). Thus, the overall fractional agreement between  $f$  and  $h$  is at least  $1 - \varepsilon - 4\varepsilon - 2/t = (1 - 5\varepsilon - o(1))$ .  $\square$

An easy corollary of the above theorem, whose form will be useful for us in the higher dimensional analysis of the test is stated below. The proof follows from the fact that if the parameter  $\varepsilon$  is chosen to be sufficiently compared to the parameter  $\varepsilon_0$  in [Theorem 7.13](#), then the corollary below does not guarantee any non-trivial agreement between  $f$  and  $h$ , else, it guarantees an agreement that is only weaker than that in [Theorem 7.13](#).

**Corollary 7.14.** *There is a constant  $\varepsilon_0 \in (0, 1)$  such that for every  $\varepsilon \in [0, 1)$ , natural numbers  $t, d$  with  $t > \max\{\frac{d+2}{1-\sqrt{\varepsilon}}, \frac{2d}{1-3\sqrt{\varepsilon}}\}$ , and sufficiently large field  $\mathbb{F}$ , the following is true.*

*Let  $\mathcal{H} = \{\ell_1, \dots, \ell_t\}$  be a set of lines in general position in  $\mathbb{F}^2$ . Let  $\mathcal{S} \subseteq \mathbb{F}^2$  denote the set of all pairwise intersection points of the lines in  $\mathcal{H}$ . If  $f : \mathcal{S} \rightarrow \mathbb{F}$  is a function for which the line-point test on  $f$  accepts with probability at least  $(1 - \varepsilon)$ , then there exists a degree  $d$  polynomial  $h \in \mathbb{F}[X, Y]$  such that  $f$  and  $h$  agree on at least  $(1 - \frac{5\varepsilon + o(1)}{\varepsilon_0})$  fraction of points in  $\mathcal{S}$ .*

### 7.3.2 Proof of [Theorem 7.3](#) in higher dimensions

We now give an analysis of the line-point test for  $m \geq 3$ . We do this in two steps: we use [Lemma 7.15](#) to show that if the line-point test passes for a function  $f$  with high probability, the plane-point test also passes with high probability; then we use the soundness of the plane-point test, [Theorem 7.2](#), to deduce that  $f$  is close to low degree.

**Lemma 7.15.** *Let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of hyperplanes in general position in  $\mathbb{F}^m$ . Let  $\mathcal{S} \subseteq \mathbb{F}^m$  denote the set of all  $m$ -wise intersection points of the hyperplanes in  $\mathcal{H}$  and  $f : \mathcal{S} \rightarrow \mathbb{F}$  be a function. If the line-point test on function  $f$  accepts with probability at least  $(1 - \varepsilon)$ , then the plane-point test on  $f$  accepts with probability at least  $(1 - \varepsilon/\varepsilon_0 - o(1))$ .*

*Proof.* Since the hyperplanes in  $\mathcal{H}$  are in general position, we have that the number of lines on every plane is the same, and every line is on the same number of planes. Thus, the distribution on  $\mathcal{L}$  obtained by sampling a uniformly random plane  $P$  from  $\mathcal{P}$  and then sampling a uniformly random line  $\ell$  from the set of lines contained in  $P$  is uniform. Following the notation in [Theorem 7.13](#), we denote  $g_\ell$  to be the degree  $d$  univariate that is closest to the restriction  $f_\ell$  of  $f$  on the line  $\ell$ , and

$\alpha(\cdot)$  denotes the fractional agreement of two functions. Therefore, we have

$$\mathbb{E}_{\ell \in \mathcal{L}} [\alpha(f_\ell, g_\ell)] = \mathbb{E}_{P \in \mathcal{P}} \left( \mathbb{E}_{\ell \in \mathcal{L}, \ell \subseteq P} [\alpha(f_\ell, g_\ell)] \right),$$

and these quantities are at least  $(1 - \varepsilon)$  since the line-point test passes with probability at least  $(1 - \varepsilon)$ .

We now note that for any plane  $P$ , the quantity  $(\mathbb{E}_{\ell \in \mathcal{L}, \ell \subseteq P} [\alpha(f_\ell, g_\ell)])$  is precisely the probability that the line-point test passes when invoked on the restriction  $f_P$  of  $f$  on the plane  $P$ . Let this probability be denoted by  $\varepsilon_P$ . From [Corollary 7.14](#), we have that for a bivariate function (namely,  $f_P$  here), if the line-point test passes with probability  $\varepsilon_P$ , then there is a degree  $d$  polynomial  $h_P$  such that  $\alpha(h_P, f_P)$  is at least  $1 - 5\varepsilon_P/\varepsilon_0 - o(1)$ , for some fixed constant  $\varepsilon_0 \in (0, 1)$ . Thus, we have that

$$\mathbb{E}_{P \in \mathcal{P}} \alpha(h_P, f_P) \geq \mathbb{E}_{P \in \mathcal{P}} (1 - 5\varepsilon_P/\varepsilon_0 - o(1)).$$

Now, using the fact that  $\mathbb{E}_P(1 - \varepsilon_P) = \mathbb{E}_{\ell \in \mathcal{L}} [\alpha(f_\ell, g_\ell)] \geq (1 - \varepsilon)$ , we get that

$$\mathbb{E}_{P \in \mathcal{P}} \alpha(h_P, f_P) \geq 1 - 5\varepsilon/\varepsilon_0 - o(1).$$

Now, for every plane  $P$ , if  $h_P$  is replaced by the degree  $d$  bivariate closest to  $f_P$ , the quantity on the left-hand side can only increase. Thus, we have that the plane-point test on the function  $f$  passes with probability at least  $(1 - 5\varepsilon/\varepsilon_0 - o(1))$ , as we wanted to show.  $\square$

## 7.4 Robust local characterization of bivariate polynomials using restrictions to lines

In this section, we prove [Theorem 7.11](#).

*Proof of Theorem 7.11.* Let  $\mathcal{S}'$  be the set of points  $\mathbf{a} \in \mathcal{S}$  such that the two lines  $\ell, \ell'$  in  $\mathcal{H}$  that pass through  $\mathbf{a}$  agree with each other on  $\mathbf{a}$ , i.e.,  $g_\ell(\mathbf{a}) = g_{\ell'}(\mathbf{a})$ . From the hypothesis of the theorem, we have that  $|\mathcal{S}'| \geq (1 - \varepsilon) \cdot |\mathcal{S}| = (1 - \varepsilon) \cdot \binom{t}{2}$ .

**An error locator polynomial:** Our first observation is that there is a non-zero bivariate polynomial  $E(x, y)$  of degree at most  $\sqrt{\varepsilon}t$  that vanishes on all points of the set  $\mathcal{S} \setminus \mathcal{S}'$ . To see this, we think of the coefficients of  $E$  as formal variables of a polynomial of degree  $\sqrt{\varepsilon}t$  and set up a system of homogeneous linear constraints asserting that  $E$  vanishes on all points in  $\mathcal{S} \setminus \mathcal{S}'$ . The number of constraints is at most  $\varepsilon \binom{t}{2}$ , whereas the number of variables is  $\binom{\sqrt{\varepsilon}t+2}{2}$ , which exceeds the number of constraints. Thus, there is a non-zero solution. Such a polynomial  $E$  plays the role of an *error locator* for the rest of the proof. Let  $e \leq \sqrt{\varepsilon}t$  denote the degree of  $E$ .

**A global low degree polynomial:** For every  $\ell \in \mathcal{H}$ , let  $P_\ell(z)$  be the polynomial

$$P_\ell(Z) := g_\ell(Z) \cdot E_\ell(Z),$$

where  $E_\ell$  is the restriction of  $E(X, Y)$  to the line  $\ell$ . Clearly,

$$\deg(P_\ell) = \deg(g_\ell) + \deg(E_\ell) \leq d + \deg(E_\ell).$$

We claim that for every pair of distinct lines  $\ell, \ell' \in \mathcal{H}$ , the polynomials  $P_\ell$  and  $P_{\ell'}$  are consistent with each other, i.e. they agree on the point  $\mathbf{a}$  where  $\{\mathbf{a}\} = \ell \cap \ell'$ . To argue this, we consider two cases based on whether or not  $\mathbf{a}$  is in  $\mathcal{S}'$ . If  $\mathbf{a} \in \mathcal{S}'$ , then we have

$$P_\ell(\mathbf{a}) = g_\ell(\mathbf{a}) \cdot E_\ell(\mathbf{a}) = g_{\ell'}(\mathbf{a}) \cdot E_{\ell'}(\mathbf{a}) = P_{\ell'}(\mathbf{a}).$$

Here the second equality follows from the fact that since  $\mathbf{a}$  is in  $\mathcal{S}'$ , and  $\ell$  and  $\ell'$  are incident to it, the corresponding line polynomials  $g_\ell$  and  $g_{\ell'}$  must agree at  $\mathbf{a}$ , and the third equality follows from the fact that by definition,  $E_\ell$  and  $E_{\ell'}$  are restrictions of  $E$  and hence must agree at  $\ell \cap \ell'$ .

If  $\mathbf{a} \notin \mathcal{S}'$ , then by the definition of  $E$ , we get that

$$P_\ell(\mathbf{a}) = P_{\ell'}(\mathbf{a}) = g_\ell(\mathbf{a}) \cdot E_\ell(\mathbf{a}) = g_{\ell'}(\mathbf{a}) \cdot E_{\ell'}(\mathbf{a}) = 0.$$

Applying the local characterization lemma ([Lemma 7.6](#)) to this collection  $\{P_\ell : \ell \in \mathcal{H}\}$  of line polynomials of degree  $d + e$ , we get that if  $t \geq d + e + 2$  (which holds since  $t > \frac{d+2}{1-2\sqrt{\varepsilon}}$  from the hypothesis of the theorem), then there is a global polynomial  $P(X, Y)$  of degree at most  $(d + e)$  such that for every line  $\ell \in \mathcal{H}$ , the restriction of  $P$  on  $\ell$  equals  $P_\ell$  (hence justifying the notation).

Now, from the above conclusion and the definition of  $P_\ell$ , we get that for every  $\ell \in \mathcal{H}$ , the restriction  $E_\ell$  of the error locator polynomial  $E$  on  $\ell$  divides the restriction  $P_\ell$  of  $P$  on the line  $\ell$ . Since  $t > \frac{2d}{1-3\sqrt{\varepsilon}}$  and  $e \leq \sqrt{\varepsilon}t$ , we get that  $t > 2d + 3e = 2(d + e) + e \geq 2 \max_\ell \{\deg(P)\} + \deg(E)$ . Thus, from the divisibility lemma ([Lemma 7.12](#)), we have that the bivariate polynomial  $E(X, Y)$  divides the bivariate polynomial  $P(X, Y)$ .

Let  $f(X, Y)$  be the resulting quotient  $P(X, Y)/E(X, Y)$ . Clearly,  $\deg(f)$  is equal to  $\deg(P) - \deg(E)$ , which is at most  $(d + e) - e = d$ . To prove the theorem, it now suffices to show that the restriction  $f_\ell$  of  $f$  on a line  $\ell$  equals  $g_\ell$  for most lines  $\ell$ .

**Agreement with many lines polynomials :** From the definition of  $f$ , we get that for every  $\ell \in \mathcal{H}$ ,

$$P_\ell(Z) = f_\ell(Z) \cdot E_\ell(Z),$$

but from the definition of  $P_\ell$ , we have that

$$P_\ell(Z) = g_\ell(Z) \cdot E_\ell(Z).$$

Thus, if  $E_\ell(Z)$  is not identically zero, it must be the case that  $f_\ell(Z)$  equals  $g_\ell(Z)$ . We now recall [Claim 7.21](#) from the proof of [Lemma 7.16](#) which shows that the number of lines  $\ell$  where  $E(X, Y)$  can vanish identically is at most  $\deg(E) = e \leq \sqrt{\varepsilon}t$ . Let  $\mathcal{B}$  denote the set of all such lines. Thus, on all lines  $\ell \in \mathcal{H} \setminus \mathcal{B}$ , we get that  $f_\ell$  must equal  $g_\ell$ . Thus, the number of such *good* lines is at least  $(1 - \sqrt{\varepsilon})t$ . To complete the proof, we now argue that the fraction of such good lines is even higher.

**An improved bound on good lines:** For a line  $\ell \in \mathcal{H}$ , let  $\mathcal{I}_\ell \subseteq \mathcal{H}$  be the set of lines  $\ell'$  such that  $g_\ell$  and  $g_{\ell'}$  do not agree on  $\ell \cap \ell'$ . Thus,  $\mathcal{I}_\ell$  is the set of lines that are *inconsistent* with  $\ell$ . We say that the lines outside  $\mathcal{I}_\ell$  are *consistent* with  $\ell$ . From the hypothesis of the theorem

$$\mathbb{E}_{\ell \in \mathcal{H}} [|\mathcal{I}_\ell|] \leq \varepsilon t.$$

Thus, by Markov's inequality, we get that

$$\Pr_{\ell \in \mathcal{H}} [|\mathcal{I}_\ell| \geq ((1 - \sqrt{\varepsilon})t - d)] \leq \frac{\varepsilon t}{(1 - \sqrt{\varepsilon})t - d} \leq 2\varepsilon,$$

where the last inequality follows from the fact that  $t > \frac{2d}{1-3\sqrt{\varepsilon}} \geq \frac{2d}{1-2\sqrt{\varepsilon}}$ . Thus, we have that for  $(1 - 2\varepsilon)t$  of the lines  $\ell$ , the set  $\mathcal{I}_\ell$  of lines that are inconsistent with  $\ell$  is of size less than  $((1 - \sqrt{\varepsilon})t - d)$ . Let  $\ell$  be one such line. Then,  $\ell$  is consistent with at least  $(\varepsilon\sqrt{t} + d + 1)$  lines in  $\mathcal{H}$ . Among these consistent lines, at most  $|\mathcal{B}| \leq \sqrt{\varepsilon}t$  of these can be in the set  $\mathcal{B}$ . Thus, there are at least  $(d + 1)$  lines  $\ell'$  such that  $\ell$  and  $\ell'$  are consistent with each other, i.e.  $g_\ell$  and  $g_{\ell'}$  agree on  $\ell \cap \ell'$ , and  $f_{\ell'} = g_{\ell'}$ . So, there are at least  $(d + 1)$  distinct points on  $\ell$ , namely its distinct intersection points  $\mathbf{a}$  with  $(d + 1)$  lines  $\ell' \notin \mathcal{B}$  that  $\ell$  is consistent with, such that

$$f_\ell(\mathbf{a}) = f_{\ell'}(\mathbf{a}) = g_\ell(\mathbf{a}) = g_{\ell'}(\mathbf{a}) = f(\mathbf{a}).$$

Thus,  $f_\ell$  and  $g_\ell$  must be equal to each other.

Therefore, the set of lines  $\ell$  on which the restriction of  $f$  equals  $g_\ell$  is of size at least  $(1 - 2\varepsilon)t$ . This completes the proof of the theorem.  $\square$

## 7.5 The divisibility lemma

In this subsection, we prove the divisibility lemma, [Lemma 7.12](#).

### 7.5.1 Proof of Lemma 7.12 in 2 dimensions

We now prove the technically easier  $m = 2$  case of Lemma 7.12, which we state below as a lemma.

**Lemma 7.16.** *Let  $\mathbb{F}$  be any field,  $e, d, t$  be natural numbers, and  $\mathcal{H}$  be a set of  $t$  lines in  $\mathbb{F}^2$  in general position.  $E(X, Y)$  and  $P(X, Y)$  are polynomials in  $\mathbb{F}[X, Y]$  of degree equal to  $e$  and  $d$  respectively. For any line  $\ell \subseteq \mathbb{F}^2$ , let  $E_\ell(Z)$  and  $P_\ell(Z)$  respectively be the univariate polynomials in  $\mathbb{F}[z]$  obtained by the restriction of  $E$  and  $P$  to the line  $\ell$ .*

*If  $t > 2d + e$  and for all lines  $\ell \in \mathcal{H}$ , there exists a polynomial  $Q_\ell(Z)$  such that*

$$P_\ell(Z) = E_\ell(Z) \cdot Q_\ell(Z),$$

*then, there exists a polynomial  $Q(X, Y) \in \mathbb{F}[X, Y]$  such that*

$$P(X, Y) = E(X, Y) \cdot Q(X, Y).$$

### Bezout's theorem related preliminaries

To prove the lemma, we first take a detour and discuss Bezout's theorem and some of its applications, which turn out to be crucial for the rest of our analysis. We start with a discussion of the bivariate case first before moving on to high dimensions.

For the rest of this section, we assume  $\mathbb{F}$  is an algebraically closed field. We begin by defining the intersection multiplicity of two bivariate polynomials at a point  $(\alpha, \beta) \in \mathbb{F}^2$ . The definition is in terms of the local ring of  $\mathbb{F}^2$  at that point, which we now define.

**Definition 7.17** (Local ring at a point in  $\mathbb{F}^2$ ). *Let  $p = (\alpha, \beta) \in \mathbb{F}^2$ . The local ring of  $p$ , denoted by  $\mathcal{O}_p$  is defined as:*

$$\mathcal{O}_p = \left\{ \frac{A(x, y)}{S(X, Y)} \mid A(X, Y), S(X, Y) \in \mathbb{F}[X, Y], S(\alpha, \beta) \neq 0 \right\},$$

*with ring operations being the usual operations on rational functions in  $X, Y$ .* ◇

We now define the notion of intersection multiplicity of curves at a point.

**Definition 7.18** (Intersection Multiplicity of curves in  $\mathbb{F}^2$  [Ful08, Section 3.3]). *Let  $A(X, Y), B(X, Y) \in \mathbb{F}[X, Y]$  be relatively prime polynomials. Let  $p \in \mathbb{F}^2$  be a point such that  $F(p) = G(p) = 0$ .*

*Then the intersection multiplicity of  $A$  and  $B$  at  $p$ , denoted  $\mathbf{IMult}(A, B; p)$  is defined by:*

$$\mathbf{IMult}(A, B; p) = \dim_{\mathbb{F}} (\mathcal{O}_p / I),$$

*where  $\mathcal{O}_p$  is the local ring of the point  $p \in \mathbb{F}^2$ , and  $I$  is the ideal of  $\mathcal{O}_p$  generated by  $A$  and  $B$ .* ◇

To help get some intuition for this notion, we mention its relation with the more familiar notion of the vanishing multiplicity of a polynomial at a point. First, if  $F$  vanishes with multiplicity  $m_1$  at

$p$  and  $G$  vanishes with multiplicity  $m_2$  at  $p$ , then the intersection multiplicity of  $F$  and  $G$  at  $p$  is at least  $m_1 \cdot m_2$ . It could be more; it will be strictly greater than  $m_1 \cdot m_2$  if (and only if) the curves cut out by  $F$  and  $G$  have a common tangent at  $p$ .

We are now ready to state Bezout's theorem for curves in a two-dimensional space.

**Theorem 7.19** (Bezout's theorem for curves in 2-dimensional space [Ful08, Section 5.3]). *Let  $A(X, Y)$ ,  $B(X, Y)$  in  $\mathbb{F}[X, Y]$  be relatively prime polynomials of degrees  $d, e$  respectively. Let  $p_1, \dots, p_s \in \mathbb{F}^2$  be common zeroes of  $F$  and  $G$ .*

*Then*

$$\sum_{i=1}^s \mathbf{IMult}(A, B; p_i) \leq d \cdot e.$$

We can now prove the following technical lemma, which we will crucially use later in proving the testability of 2-dimensional GAP codes.

**Lemma 7.20.** *Let  $\mathbb{F}$  be an algebraically closed field and  $U(X, Y), V(X, Y)$  be non-zero polynomials. Let  $a, b, \alpha \in \mathbb{F}$  and  $k \in \mathbb{N}$  be such that the univariate polynomials  $U(Z, aZ + b), V(Z, aZ + b)$  vanish with multiplicity at least  $k$  at  $\alpha$ .*

*Then, the intersection multiplicity of the curves  $U(X, Y) = 0$  and  $V(X, Y) = 0$  at the point  $(\alpha, a\alpha + b)$  is at least  $k$ .*

*Proof.* Let  $p = (\alpha, a\alpha + b) \in \mathbb{F}^2$ .

Now since  $U(Z, aZ + b)$  vanishes with multiplicity at least  $k$  at  $\alpha$ . Then  $(Z - \alpha)^k$  divides  $U(Z, aZ + b)$ , and this means that  $U(X, Y)$  can be written in the form:

$$U(X, Y) = (Y - aX - b) \cdot A(X, Y) + (X - \alpha)^k \cdot B(X, Y)$$

for some  $A(X, Y), B(X, Y) \in \mathbb{F}[X, Y]$ . Similarly  $V(Z, aZ + b)$  vanishes with multiplicity at least  $k$  at  $\alpha$ , and thus  $V(X, Y)$  is of the form:

$$V(X, Y) = (Y - aX - b) \cdot A'(X, Y) + (X - \alpha)^k \cdot B'(X, Y)$$

for some  $A'(X, Y), B'(X, Y) \in \mathbb{F}[X, Y]$ .

Thus, the ideal  $I = \langle U, V \rangle$  of  $\mathcal{O}_p$  is contained in the ideal  $J = \langle (Y - aX - b), (X - \alpha)^k \rangle$  of  $\mathcal{O}_p$ . This means that

$$\mathbf{IMult}(U, V; p) = \dim_{\mathbb{F}}(\mathcal{O}_p / I) \geq \dim_{\mathbb{F}}(\mathcal{O}_p / J).$$

Thus, to complete the proof, it suffices to show that  $\dim_{\mathbb{F}}(\mathcal{O}_p / J) \geq k$ . To see this<sup>10</sup>, we will show that the elements  $1, (X - \alpha), \dots, (X - \alpha)^{k-1}$  are  $\mathbb{F}$ -linearly independent in the  $\mathbb{F}$ -vector space  $\mathcal{O}_p / J$ .

<sup>10</sup>An alternative proof of this fact is as follows.  $\dim_{\mathbb{F}}(\mathcal{O}_p / J)$  equals the intersection multiplicity of the polynomials  $Y - aX - b$  and  $(X - \alpha)^k$  at the point  $(\alpha, a\alpha + b)$ , and this should be  $k$  (by Bezout's theorem) because it is the intersection multiplicity at the unique point of intersection (in the full projective plane) of two relatively prime polynomials, one of degree  $k$  and one of degree 1.

We do this via contradiction. Suppose these elements are linearly dependent over  $\mathbb{F}$  in  $\mathcal{O}_p/J$ . Then there exist  $a_0, \dots, a_{k-1} \in \mathbb{F}$ , not all 0, such that

$$\sum_{i=0}^{k-1} a_i (X - \alpha)^i \in J$$

in the ring  $\mathcal{O}_z$ . Let  $j$  be the smallest such that  $a_j \neq 0$ . This means (recalling the definition of the local ring) that there exists some  $S(X, Y) \in \mathbb{F}[X, Y]$  with  $S(\alpha, a\alpha + b) \neq 0$ , and some  $R(X, Y), R'(X, Y) \in \mathbb{F}[X, Y]$  such that we have the following equation in  $\mathbb{F}[X, Y]$ .

$$S(X, Y) \cdot (X - \alpha)^j \cdot \left( \sum_{i=j}^{k-1} a_i (X - \alpha)^{i-j} \right) = R(X, Y) \cdot (Y - aX - b) + R'(X, Y) \cdot (X - \alpha)^k.$$

Setting  $Y = aX + b$  in this equation, we get

$$S(X, aX + b) \cdot (X - \alpha)^j \cdot \left( \sum_{i=j}^{k-1} a_i (X - \alpha)^{i-j} \right) = R'(X, aX + b) \cdot (X - \alpha)^k.$$

Since both sides are divisible by  $(X - \alpha)^j$ , we can cancel it from both sides, and we get

$$S(X, aX + b) \cdot \left( a_j + \sum_{i=j+1}^{k-1} a_i (X - \alpha)^{i-j} \right) = R'(X, aX + b) \cdot (X - \alpha)^{k-j}.$$

Finally, substituting  $X = \alpha$  into this equation and using the fact that  $S(\alpha, a\alpha + b) \neq 0$ , we get that  $a_j = 0$ , a contradiction.

This gives us our desired lower bound on the intersection multiplicity of  $U(X, Y)$  and  $V(X, Y)$  at the point  $(\alpha, a\alpha + b)$ .  $\square$

## Proof of Lemma 7.16

We now proceed with the proof of Lemma 7.16.

*Proof of Lemma 7.16.* Let  $E(X, Y) = G(X, Y) \cdot E'(X, Y)$  and  $P(X, Y) = G(X, Y) \cdot P'(X, Y)$ , for polynomials  $E', P', G$  where  $G$  is the GCD of  $E$  and  $P$ . Therefore,  $E'$  and  $P'$  are relatively prime. From the hypothesis of the lemma, we get that for every line  $\ell \in \mathcal{H}$ , we have that there exists a polynomial  $Q_\ell(Z)$  such that

$$P'_\ell(Z) \cdot G_\ell(Z) = E'_\ell(Z) \cdot G_\ell(Z) \cdot Q_\ell(Z),$$

where  $E'_\ell, P'_\ell, G_\ell$  denote the restrictions of  $E', P'$  and  $G$  respectively to the line  $\ell$ . If the degree of  $E'$  is zero, i.e. it is a constant, then clearly,  $E(X, Y)$  must divide  $P(X, Y)$  with quotient being a constant



multiple of  $P'(X, Y)$  and we are done. So, for the rest of this proof, we assume that the degree of  $E'$  is at least one. Let  $e'$  denote the degree of  $E'(X, Y)$  and  $d'$  denote the degree of  $P'(X, Y)$ .

We say that a line  $\ell \in \mathcal{H}$  is *bad* if either  $G_\ell$  is identically zero, or the degree of  $E'_\ell$  is strictly less than the degree of  $E'(X, Y)$ . Let  $\mathcal{H}'$  be the subset of  $\mathcal{H}$  are *not* bad. This is the set of lines of interest to us for the rest of the proof, and the following claim, whose proof we defer to the end, gives a lower bound on the number of such lines.

**Claim 7.21.**

$$|\mathcal{H}'| \geq |\mathcal{H}| - e.$$

For any such line  $\ell \in \mathcal{H}'$ , we have that the univariate polynomial  $E'_\ell$  has degree equal to  $e'$  and divides the univariate polynomial  $P'_\ell$ . In this case, each of the zeroes  $\alpha$  of  $E'_\ell(Z)$  is also a zero of  $P'_\ell$ , and in particular, this means that the point  $(\alpha, m\alpha + b)$  is a common zero of the bivariables  $E'(X, Y)$  and  $P'(X, Y)$ . Here  $m, b \in \mathbb{F}$  are such that  $\ell := \{(a, am + b) : a \in \mathbb{F}\}$ . Note that the zero  $\alpha$  might not be in  $\mathbb{F}$  but certainly lives the algebraic closure  $\overline{\mathbb{F}}$  of  $\mathbb{F}$ .

The idea for the rest of the proof is that if the zeroes of  $E'_\ell$  happened to be all distinct, then, for each line  $\ell$  in  $\mathcal{H}'$ , we get  $e'$  distinct common zeroes of  $E'(X, Y)$  and  $P'(X, Y)$  in  $\overline{\mathbb{F}}^2$ . Moreover, we know that these polynomials do not have a common factor. Hence, from Bezout's theorem (Theorem 7.19), we get that

$$\frac{1}{2} \cdot e' \cdot |\mathcal{H}'| \leq \deg(P') \cdot \deg(E') = d' \cdot e',$$

where the factor  $1/2$  is to account for the fact that each common zero  $(\alpha, m\alpha + b)$  can be a point of intersection of two lines in  $\mathcal{H}'$ , and be counted twice. Moreover, since  $\mathcal{H}'$  is a set of lines in general position, no point  $(\alpha, m\alpha + b)$  can lie on three or more lines in  $\mathcal{H}'$ . This gives an upper bound of  $2d' \leq 2d$  on  $|\mathcal{H}'|$  and of  $(2d + e)$  on  $|\mathcal{H}|$  using Claim 7.21.

The technical issue with this argument is that the polynomial  $E'_\ell$  might not have distinct zeroes but have zeroes  $\alpha$  with high multiplicity, and this weakens the counting argument above and only yields a bound of  $d'e' + e \leq de + e$  on  $|\mathcal{H}|$ . To obtain a better bound, we note from Lemma 7.20 that a zero of *multiplicity*  $k$  of  $E'_\ell$  (and hence  $P'_\ell$ ) also yields a common point on the curves of  $E'(X, Y) = 0$  and  $P'(X, Y) = 0$  of intersection multiplicity  $k$ . Now, using Theorem 7.19 and the counting argument above, we get the lemma.  $\square$

Next, we prove Claim 7.21.

*Proof of Claim 7.21.* Let a line  $\ell$  be parameterized by its slope  $m$  and intercept  $b$  such that  $\ell := \{(a, am + b) : a \in \mathbb{F}\}$ . We will show that if  $\ell$  is *bad*, then its slope  $m$  must be a zero of a non-zero univariate polynomial of low degree. Thus, there are only a few choices of the slope  $m$ . Moreover,

since the lines in  $\mathcal{H}$  are in general position, there is at most one line of slope  $m$  for any  $m \in \mathbb{F}$ . Thus, the number of bad lines must be small.

Let  $e' \geq 1$  be the degree of  $E'(X, Y)$ , and let  $U_i(X, Y)$  denote the homogeneous component of  $E'(X, Y)$  of degree equal to  $i$ . Thus,

$$E'(X, Y) = \sum_{i=0}^{e'} U_i(X, Y).$$

Let  $U_{e'}(X, Y) := \sum_{j=0}^{e'} u_{e',j} Y^j X^{e'-j}$  for constants  $u_{e',j}$ , not all of which are zero. Thus, the restriction of  $E'$  on  $\ell := \{(a, am + b) : a \in \mathbb{F}\}$  is given by

$$E'_\ell(Z) = E'(Z, mZ + b) = \sum_{i=0}^{e'} U_i(Z, mZ + b),$$

and the coefficient of  $Z^{e'}$  in  $E'_\ell(Z)$  equals  $U_{e'}(1, m) = \sum_{j=0}^{e'} u_{e',j} m^j$ . In other words, the degree of  $E'_\ell(Z)$  is less than  $e'$  if and only if the slope  $m$  of  $\ell'$  is a zero of the univariate polynomial  $(\sum_{j=0}^{e'} u_{e',j} \eta^j)$  ( $\eta$  is a formal variable here). Now,  $U_{e'}(X, Y)$  is non-zero, we have that  $(\sum_{j=0}^{e'} u_{e',j} \eta^j)$  is a non-zero polynomial of degree  $e'$ , there are at most  $e'$  such choices of  $m$ .

A very similar argument implies also that the number of lines  $\ell$  on which  $G(X, Y)$  vanishes identically, i.e.  $G_\ell$  is identically zero is also bounded by  $\deg(G)$ .

Thus, the slope  $m$  of any bad line must come from a set of size at most  $\deg(E') + \deg(G) = \deg(E) = e$ , and since there are no parallel lines in  $\mathcal{H}$ , the number of bad lines is at most  $e$ .  $\square$

### 7.5.2 Proof of Lemma 7.12 for general $m$

Our proof of this lemma follows the same high-level strategy as for the 2-dimensional case. However, because we need to deal with a higher-dimensional version of Bezout's theorem and higher-dimensional intersection multiplicities, we end up needing some slightly heavier commutative algebra. For algebra and algebraic geometry background, we refer the reader to [Eis95, Har77]. We only define the algebraic concepts about which we need to reason.

#### Bezout's theorem related preliminaries

We now give a quick introduction to the statement of Bezout's theorem that we need.

Let  $A(X_1, \dots, X_m)$  and  $B(X_1, \dots, X_m)$  be relatively prime polynomials; they define codimension 1 varieties  $V(A)$  and  $V(B)$  in  $\mathbb{F}^m$ . We want to study their intersection. Since  $A$  and  $B$  are relatively prime, their intersection will be codimension 2.

Informally, Bezout's theorem for this situation is the following. Let  $W_1, \dots, W_s \subseteq \mathbb{F}_q^m$  be the irreducible components of the intersection of  $V(A)$  and  $V(B)$  (they are all codimension 2 by [Har77,

Theorem I.7.1]). For this setup, we can define the intersection multiplicity  $\mathbf{IMult}(A, B; W_i)$  of  $A$  and  $B$  at each variety  $W_i$ . Bezout's theorem says that

$$\sum_i \mathbf{IMult}(A, B; W_i) \cdot \deg(W_i) \leq \deg(A) \cdot \deg(B),$$

where  $\deg(W_i)$  is the degree of the variety  $W_i$ . In fact, there is equality in this, provided that  $A$  and  $B$  do not have a dimension  $m - 2$  component contained in the “hyperplane at infinity”.

To state Bezout's theorem formally and for our proof, we will need to understand something about intersection multiplicities. We start with the definition.

**Definition 7.22** (Intersection multiplicity in  $m$  dimensions [Har77, Section I.7]).<sup>11</sup> Let  $A(X_1, \dots, X_m), B(X_1, \dots, X_m) \in \mathbb{F}[X_1, \dots, X_m]$  be relatively prime polynomials, and let  $V(A), V(B) \subseteq \mathbb{F}^m$  be their zero sets. Let  $W$  be an irreducible variety in  $\mathbb{F}^m$  of dimension 2. Let  $\mathfrak{p}$  be the prime ideal of  $W$  in  $\mathbb{F}[X_1, \dots, X_m]$ .

Then we define the intersection multiplicity of  $A$  and  $B$  at  $W$ , denoted by  $\mathbf{IMult}(A, B; W)$  as

$$\mathbf{IMult}(A, B; W) = \text{length}_{\mathcal{O}_{\mathfrak{p}}}(\mathcal{O}_{\mathfrak{p}}/I),$$

where  $\mathcal{O}_{\mathfrak{p}}$  is the local ring of  $\mathfrak{p}$  and  $I$  is the ideal of  $\mathcal{O}_{\mathfrak{p}}$  generated by  $A$  and  $B$ . ◇

Note that if  $A$  and  $B$  do not both vanish on  $W$ , then  $\mathbf{IMult}(A, B; W) = 0$ .

**Intersections with a hyperplane:** One property of  $\mathbf{IMult}$  that we will need concerns the case where  $B$  is degree 1 (i.e.,  $V(B)$  is a hyperplane  $H$ ). In this case, for any irreducible polynomial  $U(Z_1, \dots, Z_{m-1})$  viewed as a polynomial on  $H$ , if we let  $W$  denote the  $m - 2$  dimensional irreducible variety  $V(U) \subseteq H$ , then we have:

$$\mathbf{IMult}(A, B; W) = \text{the highest power of } U \text{ that divides } A_H.$$

Abusing notation, when  $B$  is degree 1 and thus  $V(B)$  is a hyperplane  $H$ , we will sometimes refer to  $\mathbf{IMult}(A, B; W)$  as  $\mathbf{IMult}(A, H; W)$ .

We will also need the property that  $\deg(W)$  in this case equals the degree of the polynomial  $U(Z_1, \dots, Z_{m-1})$ .

---

<sup>11</sup>There are some minor differences between what we write here and the definition in [Har77].

1. We only talk about affine varieties.
2. We only talk about the intersection of codimension 1 varieties, while in [Har77], one of the varieties is allowed to be of arbitrary codimension.
3. We do not require  $A$  and  $B$  to be irreducible. Since we are in the codimension 1 setting,  $A$  and  $B$  can be factored into irreducible factors, and everything generalizes since the definition of intersection multiplicity satisfies the bilinearity relations  $\mathbf{IMult}(A_1 A_2, B; W) = \mathbf{IMult}(A_1, B; W) + \mathbf{IMult}(A_2, B; W)$  and  $\mathbf{IMult}(A, B_1 B_2; W) = \mathbf{IMult}(A, B_1; W) + \mathbf{IMult}(A, B_2; W)$  whenever the  $A$ 's and  $B$ 's are relatively prime.

With the definition of **IMult** in hand, we can now state Bezout's theorem in the  $m$ -dimensional setting.

**Theorem 7.23** (Bezout's theorem in  $m$ -dimensions (for 2 hypersurfaces) [Har77, Theorem I.7.7]). *Let  $A(X_1, \dots, X_m), B(X_1, \dots, X_m) \in \mathbb{F}[X_1, \dots, X_m]$  be relatively prime polynomials, and let  $V(A), V(B) \subseteq \mathbb{F}^m$  be their zero sets. Let  $W_1, \dots, W_s$  be the irreducible components of  $V(A) \cap V(B)$ . Then,*

$$\sum_{j=1}^s \mathbf{IMult}(A, B; W_j) \leq \deg(A) \cdot \deg(B).$$

*Furthermore, if the closures  $V(A)$  and  $V(B)$  in projective space  $\mathbb{P}^m(\mathbb{F})$  do not contain any common dimension  $m - 2$  components, then the inequality above is, in fact, an equality.*

### Proof of Lemma 7.12

We can now prove the divisibility lemma for general  $m$ , Lemma 7.12.

*Proof of Lemma 7.12.* The high-level outline of the lemma is similar to that of the two-dimensional case. We begin with a high-level sketch. First, we may assume that  $E$  and  $P$  are relatively prime; otherwise we can divide out any common factors and reduce to the relatively prime case. Now, suppose  $e \geq 1$  – otherwise  $e = 0$ , and the conclusion holds trivially.

Our plan is to apply Bezout's theorem to  $E$  and  $P$ . Each  $H$  for which  $E_H$  divides  $P_H$  will account for some irreducible components  $W_i$  in the intersection of  $V(E)$  and  $V(P)$  which have a large total of the product of their degree and intersection multiplicity. This gives a lower bound, via Bezout's theorem, on  $\deg(E) \cdot \deg(P)$ , for which we have an apriori upper bound. Resolving this yields the result.

As in the  $m = 2$  case, we need a lemma about the intersection multiplicity of hypersurfaces that share a common tangent hyperplane. We prove this lemma at the end of this section.

**Lemma 7.24.** *Suppose  $E, P$  are relatively prime polynomials in  $\mathbb{F}[X_1, \dots, X_m]$  and  $H$  is a hyperplane in  $\mathbb{F}^m$ . Let  $W$  be an irreducible variety of dimension  $m - 2$  in  $H$ , and suppose both  $\mathbf{IMult}(E, H; W)$  and  $\mathbf{IMult}(P, H; W)$  are at least  $k$ . Then,*

$$\mathbf{IMult}(E, P; W) \geq k.$$

Now, we proceed with the formal proof. Let

$$\mathcal{H}' = \{H \in \mathcal{H} \mid \deg(E_H) = e\}.$$

Note that there can be very few  $H \in \mathcal{H}$  which do not lie in  $\mathcal{H}'$ . This is a natural generalization of Claim 7.21. Indeed, suppose  $E(Z_1, Z_2, \dots, Z_{m-1}, \langle \mathbf{a}, \mathbf{Z} \rangle + b)$  has degree strictly less than  $e$ . Then

letting  $E^*$  be the highest degree homogeneous part of  $E$ , we have that  $E^*(Z_1, \dots, Z_{m-1}, \langle \mathbf{a}, \mathbf{Z} \rangle) = 0$ , and thus  $E^*(Z_1, \dots, Z_{m-1}, Y)$  is divisible by  $(Y - \langle \mathbf{a}, \mathbf{Z} \rangle)$ . Since the  $H$  in  $\mathcal{H}$  are in general position and thus correspond to different  $\mathbf{a}$ s, we get that each  $H \in \mathcal{H} \setminus \mathcal{H}'$  yields a distinct factor of  $E^*$ , and so there can be at most  $e$  such  $H$ . Thus, we showed,

$$|\mathcal{H}'| \geq |\mathcal{H}| - e > 2d.$$

Now let  $H \in \mathcal{H}'$ , and let us consider the polynomials  $E_H$  and  $P_H$ . Let  $\mathcal{W}_H$  be the set of irreducible components of  $V(E) \cap H$ . Since  $E_H$  divides  $P_H$ , then for every irreducible factor  $U(Z_1, \dots, Z_{m-1})$  of  $E_H$ , the highest power of  $U$  that divides  $P_H$  is at least the highest power of  $U$  that divides  $E_H$ . Thus  $W \in \mathcal{W}_H$ , we have,

$$\mathbf{IMult}(P, H; W) \geq \mathbf{IMult}(E, H; W).$$

Now, from [Lemma 7.24](#), we get that for every  $H \in \mathcal{H}'$  and every  $W \in \mathcal{W}_H$ ,

$$\mathbf{IMult}(E, P; W) \geq \mathbf{IMult}(E, H; W).$$

Thus we have,

$$\begin{aligned} & \sum_{W \in \mathcal{W}_H} \mathbf{IMult}(E, P; W) \deg(W) \\ & \geq \sum_{W \in \mathcal{W}_H} \mathbf{IMult}(E, H; W) \deg(W) \\ & = \deg(E_H) \\ & = e. \end{aligned}$$

In words, every  $H \in \mathcal{H}'$  accounts for many of the intersection multiplicity of  $E$  and  $P$ . Now consider the expression  $\Delta$  defined below.

$$\begin{aligned} \Delta &:= \sum_{H \in \mathcal{H}'} \sum_{W \in \mathcal{W}_H} \mathbf{IMult}(E, P; W) \deg(W) \\ &\geq |\mathcal{H}'| \cdot e \\ &> 2de, \end{aligned}$$

where the inequality is strict because of our assumption that  $e \geq 1$ .

On the other hand,  $\Delta$  can be bounded from above by Bezout's theorem. Let  $\mathcal{F}$  be the set of all irreducible components of  $V(E) \cap V(P)$ . (Note that they are all  $m - 2$  dimensional by [\[Har77](#),

Theorem I.7.1].). From [Theorem 7.23](#), we have

$$\begin{aligned}
\Delta &= \sum_{H \in \mathcal{H}'} \sum_{W \in \mathcal{W}_H} \mathbf{IMult}(E, P; W) \deg(W) \\
&= \sum_{H \in \mathcal{H}'} \sum_{W \in \mathcal{F}} 1_{W \subseteq H} \mathbf{IMult}(E, P; W) \deg(W) \\
&= \sum_{W \in \mathcal{F}} \sum_{H \in \mathcal{H}'} 1_{W \subseteq H} \mathbf{IMult}(E, P; W) \deg(W) \\
&= \sum_{W \in \mathcal{F}} \mathbf{IMult}(E, P; W) \deg(W) \left( \sum_{H \in \mathcal{H}'} 1_{W \subseteq H} \right) \\
&\leq 2 \sum_{W \in \mathcal{F}} \mathbf{IMult}(E, P; W) \deg(W) \\
&\leq 2de,
\end{aligned}$$

where the last line is Bezout's theorem, and the penultimate inequality is because any  $(m-2)$ -dimensional variety  $W$  can be contained in at most 2 of the hyperplanes  $H \in \mathcal{H}'$  (since they are in general position – and this can only happen if  $W$  is the  $(m-2)$ -dimensional intersection of two hyperplanes in  $\mathcal{H}'$ ).

This gives us a contradiction, and the result follows.  $\square$

#### Proof of [Lemma 7.24](#)

We now prove [Lemma 7.24](#).

*Proof of [Lemma 7.24](#).* Let us rename the  $m$  variables  $X_1, \dots, X_m$  to  $Z_1, \dots, Z_{m-1}, Y$ .

By an affine change of variables, we may assume that the hyperplane  $H$  is defined by  $Y = 0$ , and  $W$ , the irreducible  $m-2$ -dimensional subvariety of  $H$ , is defined by the two equations  $Y = 0, U(Z_1, \dots, Z_{m-1}) = 0$ , where  $U$  is an irreducible polynomial. (This uses the fact that every irreducible subvariety that is codimension 1 in an affine space can be defined by a single irreducible polynomial.) Thus the prime ideal of  $\mathbb{F}[Z_1, \dots, Z_{m-1}, Y]$  associated with  $W$ ,  $\mathfrak{p}$ , is generated by  $y$  and  $U(Z_1, \dots, Z_{m-1})$ .

After the variable change, the polynomial  $E_H(Z_1, \dots, Z_{m-1})$  becomes  $E(Z_1, \dots, Z_{m-1}, 0)$  and  $P_H(Z_1, \dots, Z_{m-1})$  becomes  $P(Z_1, \dots, Z_{m-1}, 0)$ . Thus  $E(Z_1, \dots, Z_{m-1}, Y)$  and  $P(Z_1, \dots, Z_{m-1}, Y)$  are of the form:

$$E(Z_1, \dots, Z_{m-1}, Y) = E_H(Z_1, \dots, Z_{m-1}) + Y \cdot C_E(Z_1, \dots, Z_{m-1}, Y),$$

$$P(Z_1, \dots, Z_{m-1}, Y) = P_H(Z_1, \dots, Z_{m-1}) + Y \cdot C_P(Z_1, \dots, Z_{m-1}, Y),$$

for some polynomials  $C_E, C_P$ .

Our hypotheses  $\mathbf{IMult}(E, H; W) \geq k$  and  $\mathbf{IMult}(P, H; W) \geq k$  tell us that  $E_H(Z_1, \dots, Z_{m-1})$  and  $P_H(Z_1, \dots, Z_{m-1})$  are both divisible by  $U(Z_1, \dots, Z_{m-1})^k$ .

Thus, the polynomials  $E(Z_1, \dots, Z_{m-1}, Y)$  and  $P(Z_1, \dots, Z_{m-1}, Y)$  both lie in the ideal  $J$  of  $\mathbb{F}[Z_1, \dots, Z_{m-1}, Y]$  generated by  $U(Z_1, \dots, Z_{m-1})^k$  and  $Y$ .

Now we consider  $\mathbf{IMult}(E, P; W)$ . Let  $\mathfrak{p}$  be the prime ideal of  $W$  in  $\mathbb{F}[Z_1, Z_2, \dots, Z_{m-1}, Y]$ . By definition:

$$\mathbf{IMult}(E, P; W) = \text{length}_{\mathcal{O}_{\mathfrak{p}}}(\mathcal{O}_{\mathfrak{p}}/I),$$

where  $I$  is the ideal of  $\mathcal{O}_W$  generated by  $E$  and  $P$ . By our previous observation, if we let  $J$  be the ideal of  $\mathcal{O}_{\mathfrak{p}}$  generated by  $U(Z_1, \dots, Z_{m-1})^k$  and  $Y$ , then  $I \subseteq J$ . Thus

$$\mathbf{IMult}(E, P; W) \geq \text{length}_{\mathcal{O}_{\mathfrak{p}}}(\mathcal{O}_{\mathfrak{p}}/J).$$

This latter expression simply equals  $\mathbf{IMult}(Y, U(Z_1, \dots, Z_{m-1})^k; W)$ , which, by the general comments about intersection multiplicities of polynomials with hyperplanes, equals  $k$ . Alternatively,  $\text{length}_{\mathcal{O}_{\mathfrak{p}}}(\mathcal{O}_{\mathfrak{p}}/J)$  can be seen to be at least  $k$  because of the length  $k$  sequence of  $\mathcal{O}_{\mathfrak{p}}$  modules

$$\{0\} \subseteq \mathcal{O}_{\mathfrak{p}}/\langle Y, U \rangle \subseteq \dots \subseteq \mathcal{O}_{\mathfrak{p}}/\langle Y, U^i \rangle \subseteq \dots \subseteq \mathcal{O}_{\mathfrak{p}}/\langle Y, U^k \rangle = \mathcal{O}_{\mathfrak{p}}/J.$$

This completes the proof. □

## References

- [AB03] Manindra Agrawal and Somenath Biswas. **Primality and identity testing via Chinese remaindering**. *J. ACM*, 50(4):429–443, jul 2003.
- [ASS23] Prashanth Amireddy, Srikanth Srinivasan, and Madhu Sudan. **Low-Degree Testing over Grids**. In *Approximation, Randomization, and Combinatorial Optimization*, Leibniz International Proceedings in Informatics, LIPIcs, 9 2023.
- [BHKS23] Siddharth Bhandari, Prahladh Harsha, Mrinal Kumar, and Ashutosh Shankar. **Algorithmizing the Multiplicity Schwartz-Zippel Lemma**. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2816–2835. Society for Industrial and Applied Mathematics, 2023.
- [Bog05] Andrej Bogdanov. **Pseudorandom generators for low degree polynomials**. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, page 21–30, New York, NY, USA, 2005. Association for Computing Machinery.
- [BP20] Markus Bläser and Anurag Pandey. **Polynomial Identity Testing for Low Degree Polynomials with Optimal Randomness**. In *Approximation, Randomization, and Combinatorial*

- Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*, volume 176 of *LIPIcs*, pages 8:1–8:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.
- [BSS06] Eli Ben-Sasson and Madhu Sudan. **Robust Locally Testable Codes and Products of Codes**. *Random Structures & Algorithms*, 28(4):387–402, 2006.
  - [BSS20] Mitali Bafna, Srikanth Srinivasan, and Madhu Sudan. **Local decoding and testing of polynomials over grids**. *Random Structures and Algorithms*, 57, 06 2020.
  - [BV10] Andrej Bogdanov and Emanuele Viola. **Pseudorandom Bits for Polynomials**. *SIAM J. Comput.*, 39(6):2464–2486, 2010.
  - [CK97] Zhi-Zhong Chen and Ming-Yang Kao. **Reducing randomness via irrational numbers**. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, STOC '97*, page 200–209, 1997.
  - [DEL<sup>+</sup>22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. **Locally testable codes with constant rate, distance, and locality**. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2022*, page 357–374, New York, NY, USA, 2022.
  - [DGV24] Ashish Dwivedi, Zeyu Guo, and Ben Lee Volk. **Optimal Pseudorandom Generators for Low-Degree Polynomials Over Moderately Large Fields**, 2024. Pre-print available at [arXiv:2402.11915](https://arxiv.org/abs/2402.11915).
  - [DLZ23] Irit Dinur, Siqi Liu, and Rachel Yun Zhang. New codes on high dimensional expanders. *arXiv preprint arXiv:2308.15563*, 2023.
  - [DS11] Zeev Dvir and Amir Shpilka. **Noisy Interpolating Sets for Low-Degree Polynomials**. *Theory of Computing*, 7(1):1–18, 2011.
  - [DV22] Harm Derksen and Emanuele Viola. **Fooling polynomials using invariant theory**. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 399–406, 2022.
  - [Eis95] David Eisenbud. *Commutative Algebra with a View Toward Algebraic Geometry*. Springer New York, NY, 1995.
  - [For66] G. Forney. **Generalized Minimum Distance Decoding**. *IEEE Transactions on Information Theory*, 12(2):125–131, 1966.



- [For15] Michael A. Forbes. **Deterministic Divisibility Testing via Shifted Partial Derivatives**. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 451–465, 2015.
- [Ful08] W. Fulton. *Algebraic Curves: An Introduction to Algebraic Geometry*. 2008.
- [GRS12] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. *Draft available at <http://www.cse.buffalo.edu/atri/courses/coding-theory/book>*, 2(1), 2012.
- [GS00] Mariano Gasca and Thomas Sauer. **Polynomial interpolation in several variables**. *Advances in Computational Mathematics*, 12(4):377–410, 2000.
- [Har77] Robin Hartshorne. *Algebraic Geometry*. Springer New York, NY, 1977.
- [HKSS23] Prahladh Harsha, Mrinal Kumar, Ramprasad Saptharishi, and Madhu Sudan. An Improved Line-Point Low-Degree Test, 2023. Pre-print available at [arXiv:2311.12752](https://arxiv.org/abs/2311.12752).
- [KK17] John Y. Kim and Swastik Kopparty. **Decoding Reed-Muller Codes over Product Sets**. *Theory Comput.*, 13(1):1–38, 2017. (Preliminary version in *31st IEEE Conference on Computational Complexity*, 2016). Pre-print available at [arXiv:1511.07488](https://arxiv.org/abs/1511.07488).
- [KS01] Adam R. Klivans and Daniel Spielman. **Randomness efficient identity testing of multivariate polynomials**. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing, STOC '01*, page 216–223, 2001.
- [Lov08] Shachar Lovett. **Unconditional pseudorandom generators for low degree polynomials**. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing, STOC '08*, page 557–562, New York, NY, USA, 2008. Association for Computing Machinery.
- [LV98] Daniel Lewin and Salil Vadhan. **Checking polynomial identities over any field: towards a derandomization?** In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98*, page 438–447, 1998.
- [Mei13] Or Meir. **Combinatorial PCPs with Short Proofs**. *Electron. Colloquium Comput. Complex.*, TR13-134, 2013. Pre-print available at [arXiv:TR13-134](https://arxiv.org/abs/1301.1111).
- [PK22] Pavel Panteleev and Gleb Kalachev. **Asymptotically good Quantum and locally testable classical LDPC codes**. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2022*, page 375–388, 2022.
- [PS94] Alexander Polishchuk and Daniel A. Spielman. **Nearly-linear size holographic proofs**. In *Symposium on the Theory of Computing*, 1994.

- [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 Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 475–484, 1997.
- [SY11] Shubhangi Saraf and Sergey Yekhanin. **Noisy Interpolation of Sparse Polynomials, and Applications**. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, USA, June 8-10, 2011*, pages 86–92. IEEE Computer Society, 2011.
- [Vid15] Michael Viderman. **A combination of testability and decodability by tensor products**. *Random Struct. Algorithms*, 46(3):572–598, 2015.
- [Vio08] Emanuele Viola. **The Sum of  $d$  Small-Bias Generators Fools Polynomials of Degree  $d$** . In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 124–127, 2008.
- [WB86] Lloyd R. Welch and Elwyn R. Berlekamp. Error correction for algebraic block codes, December 1986. U.S. Patent Number 4,633,470.

## A $d$ -robustness in 2 variables

In this section, we investigate the problem of finding the smallest set with a given  $d$ -robustness. This corresponds to the best  $R$  vs  $\delta$  tradeoff that can be obtained from our generalized Schwartz-Zippel lemma for shapes. Our main results are that in the regime of large  $\delta$  (close to 1) and small  $\delta$  (close to 0), the grid and the simplex are essentially optimal. As we already saw, for intermediate  $\delta$ , there are shapes that outperform both of these.

We first show the grid is optimal for relative  $d$ -robustness greater than  $1/2$ .

### A.1 High robustness regime

First, we show that any set with relative  $d$ -robustness greater than  $1/2$  must contain the point  $d, d$ .

**Lemma A.1.** *Let  $\delta > 1/2$ . Suppose  $S \subset \mathbb{N}^2$  is such that  $\pi_d(S) \geq \delta$ . Then  $S$  contains  $(d, d)$ .*

*Proof.* By contradiction, suppose  $(d, d) \notin S$ . Define the following partition of  $S$ .  $S_x = S \cap H(d, 0)$ ,  $S_y = S \cap H(0, d)$ , and  $S_\perp = \{(x, y) : x < d \wedge y < d\}$ . In words,  $S_x$  is the subset of  $S$  with  $x$  coordinate at least  $d$ ,  $S_y$  is the subset of  $S$  with  $y$  coordinate at least  $d$ , and  $S_\perp$  is the subset where both  $x$  and  $y$  coordinates are less than  $d$ . Since  $S$  is downward closed, there are no elements of  $S$  with both  $x$  and  $y$  coordinates at least  $d$ , and thus  $(S_x, S_y, S_\perp)$  is indeed a partition of  $S$ .

Since  $\pi_d(S) \geq \delta$   $d$ -robust, we have  $|S_x| \geq \delta|S|$ , and  $|S_y| \geq \delta|S|$ . Thus, we have  $|S_x| + |S_y| \geq 2\delta|S|$  which implies  $|S| \geq 2\delta|S|$ , so  $\delta \leq 1/2$ , which is a contradiction.  $\square$

We now prove the following theorem, which implies that the grid is optimal for  $\delta \geq 1/2$ .

**Theorem A.2.** Let  $S \in \mathbb{N}^2$  with  $\pi_d(S) > 1/2$ . Then

$$|S| \geq \left( \sqrt{\delta|S| + \frac{d^2}{4}} + \frac{d}{2} \right)^2.$$

Furthermore, setting  $R = \frac{d^2}{2|S|}$ , we have  $R \leq (1 - \delta)^2/2$ .

*Proof.* Let  $V_x = S \cap H(d, 0)$ . Let  $V_y = S \cap H(0, d)$ . Let  $U = V_x \cap V_y$ . By hypothesis,  $|V_x|, |V_y| \geq \delta|S|$ . Let  $c = \frac{|V_x| + |V_y|}{2}$ , and note that  $c \geq \delta|S|$ .

We have:

$$\begin{aligned} |S| &= d^2 + |V_x| + |V_y| - |V_x \cap V_y| \\ &= d^2 + 2c - |U| \end{aligned}$$

We now get an upper bound on  $|U|$  in terms of  $c$ . Let  $a_x$  be the largest  $a$  such that  $(d + a, d) \in S$ . Let  $a_y$  be the largest  $a$  such that  $(d, d + a) \in S$ . Then  $|U| \leq a_x a_y$ . Furthermore, we have  $|V_x| - |U| \geq da_x$  and  $|V_y| - |U| \geq da_y$ .

Thus

$$|U| \leq \frac{(|V_x| - |U|)(|V_y| - |U|)}{d^2}.$$

By the AM-GM inequality,

$$|U| \leq \frac{(c - |U|)^2}{d^2}.$$

Simplifying,

$$|U|^2 - (2c + d^2)|U| + c^2 \geq 0.$$

Solving, we get:

$$\begin{aligned} |U| &\leq c + \frac{d^2}{2} - \sqrt{\left(c + \frac{d^2}{2}\right)^2 - c^2} \\ &= c + \frac{d^2}{2} - \sqrt{\frac{d^2}{2} \left(2c + \frac{d^2}{2}\right)} \\ &= c + \frac{d^2}{2} - d\sqrt{c + \frac{d^2}{4}}. \end{aligned}$$

Thus

$$\begin{aligned}
|S| &\geq d^2 + 2c - \left( c + \frac{d^2}{2} - d\sqrt{c + \frac{d^2}{4}} \right) \\
&= \frac{d^2}{2} + c + d\sqrt{c + \frac{d^2}{4}} \\
&= \left( \sqrt{c + \frac{d^2}{4}} + \frac{d}{2} \right)^2 \\
&\geq \left( \sqrt{\delta|S| + \frac{d^2}{4}} + \frac{d}{2} \right)^2,
\end{aligned}$$

as claimed. Then, substituting  $|S| = \frac{d^2}{2R}$ , and dividing through by  $d^2$ , we get

$$\frac{1}{2R} \geq \left( \sqrt{\frac{\delta}{2R} + \frac{1}{4}} + 1/2 \right)^2.$$

Expanding and multiplying through by  $2R$ , we get

$$1 \geq \delta + R + \sqrt{2\delta R + R^2}.$$

Removing the square root, we get

$$(1 - \delta - R)^2 - 2\delta R - R^2 \geq 0.$$

Expanding and simplifying, we get

$$\delta^2 - 2R - 2\delta + 1 \geq 0.$$

Finally, rearranging for  $R$ , we get

$$R \leq (1 - \delta)^2/2,$$

as claimed. □

Let  $S$  be the grid of side length  $\ell$ . Recall from the discussion in [Section 4.4](#) that  $S$  has relative  $d$ -robustness  $1 - d/\ell$ . Then the RHS of the inequality is  $\frac{d^2}{2\ell^2} = \frac{d^2}{2|S|} = R$ . Hence, the grid is the smallest set with relative  $d$ -robustness  $\delta$  for  $\delta > 1/2$ . Next, we study the low-distance regime.

## A.2 Low robustness regime

**Theorem A.3.** Let  $\delta > 0$ , and  $S \subset \mathbb{N}^2$  be such that  $\pi_d(S) \geq \delta$ . Let  $N = |S|$ , and  $R = \frac{d^2}{2N}$ . Then we have

$$N - \frac{d^2}{2} \geq \sqrt{2\delta N}d,$$

and

$$R \leq 1 - 2\sqrt{\delta^2 + \delta} + 2\delta$$

*Proof.* Let  $U = \Lambda_{2,d}$  be the two-dimensional simplex of side length  $d$ . Since  $\pi_d(S) \geq \delta > 0$ ,  $U \subset S$  (otherwise  $S$  is contained within the first  $d$  column or the first  $d$  rows). Let  $V = \overline{U} \cap S$ . We will find several subsets  $V_i$  of  $V$  that are mutually disjoint. Taking the sum of sizes of these  $V_i$  will yield our lower bound.

For convenience, define  $X_{<a} = \{(x, y) \in \mathbb{N}^2 : x < a\}$ . Similarly define  $X_{\geq a}, Y_{<a}$ .

Let  $S_1 = V \cap H(0, d)$ . By hypothesis,  $|S_1| \geq \delta N$ . Let  $b_1$  be the smallest value for which  $S_1 \subset X_{<b_1}$ . I.e.,  $b_1$  is the number of columns in  $S_1$ . Then, since  $S$  is downward closed, we have that the set  $X_{<b_1} \cap Y_{<d} \subset S$ . Define  $T_1 = X_{<b_1} \cap Y_{<d} \cap \overline{U}$ , and note that it is disjoint from  $S_1$ . A simple counting argument shows that  $|T_1| = \binom{b_1}{2}$ . We take  $V_1 = S_1 \cup T_1$ , and note that  $|V_1| \geq \delta N + \binom{b_1}{2}$ .

Now we consider the part of  $V$  with  $x$  coordinate at least  $b_1$ .

Let  $S_2 = V \cap H(b_1, d - b_1)$ . By hypothesis, we have  $|S_2| \geq \delta N$ . Suppose  $S_2$  has  $b_1 + b_2$  columns, i.e.  $b_2$  is the smallest number such that  $S_2 \subset X_{<b_1+b_2}$ . By the same logic, we have that  $T_2 = X_{\geq b_1} \cap X_{<b_2} \cap Y_{<d-b_1} \cap \overline{U}$  is a subset of  $V$  that is disjoint from  $S_2$ . We take  $V_2 = S_2 \cup T_2$ , and note that  $|V_2| \geq \delta N + \binom{b_2}{2}$ . Also, note that  $V_2$  is disjoint from  $V_1$  since their projections on the  $X$  axis are disjoint.

We repeat the process of taking

$$S_k = V \cap H\left(\sum_{i=1}^k b_i, d - \sum_{i=1}^{k-1} b_i\right),$$

and

$$T_k = X_{\geq \sum_{i=1}^{k-1} b_i} \cap X_{<\sum_{i=1}^k b_i} \cap Y_{<d-\sum_{i=1}^{k-1} b_i} \cap \overline{U},$$

until  $\sum_{i=1}^k b_i \geq d$ .

Let  $\hat{b}_k = d - \sum_{i=1}^{k-1} b_i$ , it is not hard to see that  $|T_k| \geq \binom{\hat{b}_k}{2}$ . Replace  $b_k$  with  $\hat{b}_k$ , so that  $\sum_{i=1}^k b_i = d$ , and  $T_i \geq \binom{b_i}{2}$  for each  $i \in [k]$ . Thus,  $|V_i| \geq \delta N + \binom{b_i}{2}$  for each  $i \in [k]$ . We now use these to bound the size of  $S$ . We have

$$\begin{aligned}
N &\geq |U| + k\delta N + \sum_{i=1}^k \binom{b_i}{2} \\
&= \frac{d(d-1)}{2} + k\delta N + \sum_{i=1}^k \frac{b_i(b_i-1)}{2} \\
&= \frac{d^2}{2} + k\delta N + \sum_{i=1}^k \frac{b_i^2}{2} - d \\
&\geq \frac{d^2}{2} + k\delta N + \frac{d^2}{2k} - d \\
&\geq \frac{d^2}{2} + d\sqrt{2\delta N} - d.
\end{aligned}$$

The third line follows from the fact that  $\sum_{i=1}^k b_i = d$ , the fourth line follows from the Cauchy-Schwartz inequality, and the fifth line follows from the AM-GM inequality. Substituting  $N = \frac{d^2}{2R}$ , and rearranging, we get that

$$1 \geq 2\sqrt{R\delta} + R - O(1/d)$$

Ignoring the  $O(1/d)$  term and solving for  $R$ , we get that  $R \leq 1 - 2\sqrt{\delta^2 + \delta} + 2\delta$ .

□

Recall that the simplex evaluation set obtains a trade-off  $R = (1 - \sqrt{\delta})^2 = 1 - 2\sqrt{\delta} + \delta$ . Taking the Taylor expansion of the expression from [Theorem A.3](#), get  $R \leq 1 - 2\sqrt{\delta} + 2\delta + O(\delta^{3/2})$ . Thus, the simplex has the optimal asymptotic behaviour as  $\delta$  approaches 0, namely,  $R = 1 - 2\sqrt{\delta} + O(\delta)$ . [Figure 5](#) compares the trade-off obtained by the simplex to the upper bound in [Theorem A.3](#).

## B Robust local characterization in $m$ dimensions using error-locator polynomials

In this section we give a proof of a quantitatively weaker robust local characterization (which is completely subsumed by our results in [Section 7](#)), using an interpolated error-locator polynomial in the spirit of our [Theorem 7.11](#). This uses our divisibility lemma for general  $m$ . We only include this because we find the proof method interesting.

The quantitatively weak aspect is that we cannot take  $\frac{d}{t}$  close to 1 in the statement, and thus it does not apply to high rate GAP codes<sup>12</sup>.

---

<sup>12</sup>The source of this weakness is the factor 2 in the divisibility lemma – however this factor 2 is tight for the divisibility lemma.

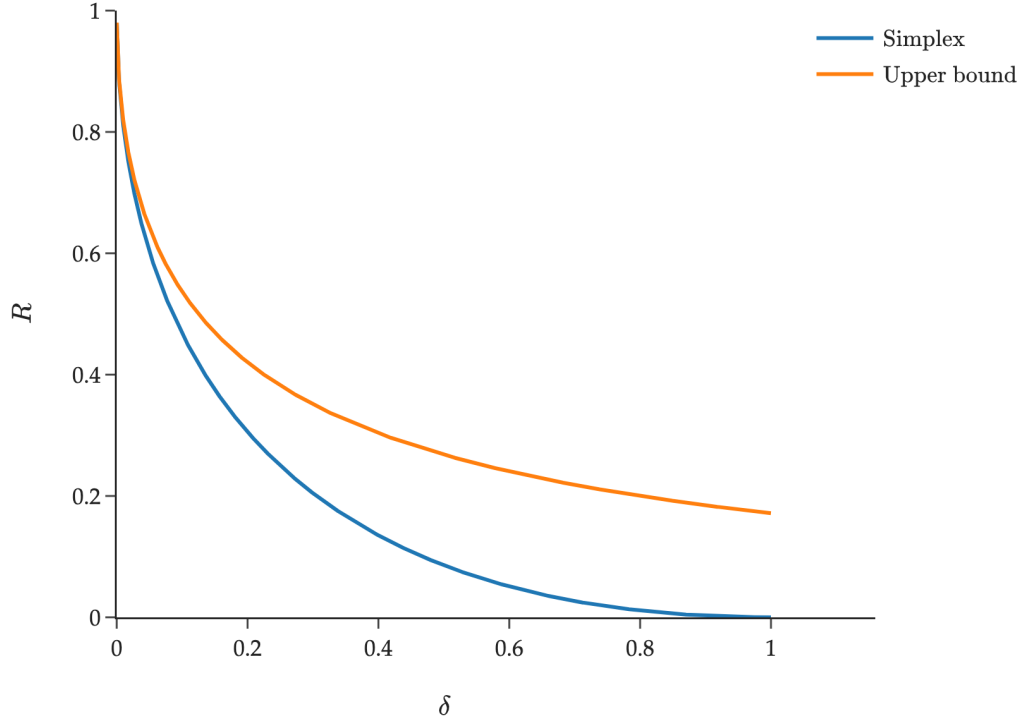


Figure 5: Comparing rate vs. robustness trade-off of the simplex construction to the upper bound

**Theorem B.1.** For all  $m, d, t \in \mathbb{N}$  and  $\varepsilon \in [0, 1)$  with  $t \geq 12(d + m)$ ,  $\varepsilon < \frac{1}{100m^2}$ , and for all large enough field  $\mathbb{F}$ , the following is true.

Let  $\mathcal{H} = \{H_1, \dots, H_t\}$  be a set of hyperplanes in general position in  $\mathbb{F}^m$ . Let  $\mathcal{S} \subseteq \mathbb{F}^m$  denote the set of all  $m$ -wise intersection points of the hyperplanes in  $\mathcal{H}$ . For every hyperplane  $H$  in  $\mathcal{H}$ , let  $g_H : H \rightarrow \mathbb{F}$  be a degree  $d$  polynomial on  $(m - 1)$  variables.

If the polynomials  $g_H$  and  $g_{H'}$  agree with each other on  $H \cap H'$  for at least  $(1 - \varepsilon)\binom{t}{2}$  pairs  $\{H, H'\} \subseteq \mathcal{H}$ , then there is an  $m$ -variate degree  $d$  polynomial  $f : \mathcal{S} \rightarrow \mathbb{F}$  and a set  $\tilde{\mathcal{H}} \subseteq \mathcal{H}$  of size at least  $(1 - 2\varepsilon)t$  such that for every  $H \in \tilde{\mathcal{H}}$ , the restriction of  $f$  on  $H$  equals  $g_H$ .

*Proof of Theorem B.1.* Let  $\mathcal{U}$  be the set of all subspaces of co-dimension two obtained by taking pairwise intersections of distinct hyperplanes in  $\mathcal{H}$ , and  $\mathcal{U}' \subseteq \mathcal{U}$  be the subset of co-dimension two subspaces  $A$  such that if  $A = H \cap H'$  for  $H, H' \subseteq \mathcal{H}$ , then  $g_H$  and  $g_{H'}$  agree on  $A$ , i.e.  $g_H$  and  $g_{H'}$  are consistent with each other. From the hypothesis of the theorem, we have that  $|\mathcal{U}'| \geq (1 - \varepsilon) \cdot |\mathcal{U}| = (1 - \varepsilon) \cdot \binom{t}{2}$ .

**An error locator polynomial:** We show that there is a non-zero  $m$ -variate polynomial  $E(\mathbf{X})$  of degree at most  $\sqrt{\frac{\varepsilon}{2}}tm$  that vanishes identically on all subspaces in the set  $\mathcal{U} \setminus \mathcal{U}'$ . This is again via an interpolation argument. We think of the coefficients of  $E$  as formal variables of a polynomial

of degree  $\sqrt{\frac{\varepsilon}{2}}tm$  on  $m$  variables. Now, for any linear subspace  $A$  of co-dimension two, we impose the constraint that the restriction of  $E$  on  $A$ , denoted by  $E^A$  vanishes identically.<sup>13</sup> Note that up to an invertible change of basis,  $E^A$  is a polynomial of degree at most  $\deg(E)$  on  $(m-2)$  variables. Thus, imposing the constraint that  $E_A$  is identically zero corresponds to  $\binom{\deg(E)+m-2}{m-2}$  homogeneous linear constraints on the coefficients of  $E(\mathbf{X})$ . Thus, we have a system of  $\left(\varepsilon \cdot \binom{t}{2} \cdot \binom{\deg(E)+m-2}{m-2}\right)$  homogeneous linear constraints on  $\binom{\deg(E)+m-2}{m}$  many variables. Thus, if

$$\binom{\deg(E)+m}{m} > \varepsilon \binom{t}{2} \binom{\deg(E)+m-2}{m-2},$$

The system has a non-zero solution. For  $\deg(E) = \sqrt{\frac{\varepsilon}{2}}tm$ , it can be verified that this inequality holds. Let  $E$  be one such non-zero polynomial, and  $e \leq \sqrt{\frac{\varepsilon}{2}}tm$  denote its degree.

**A global low degree polynomial:** For every  $H \in \mathcal{H}$ , let  $P_H(\mathbf{Z})$  be the polynomial

$$P_H(\mathbf{Z}) := g_H(\mathbf{Z}) \cdot E_H(\mathbf{Z}),$$

where  $E_H$  is the restriction of  $E$  to the hyperplane  $H$ . Clearly,

$$\deg(P_H) = \deg(g_H) + \deg(E_H) \leq d + \deg(E_H).$$

We claim that for every pair of distinct hyperplanes  $H, H' \in \mathcal{H}$ , the polynomials  $P_H$  and  $P_{H'}$  are consistent with each other, i.e. they agree on the subspace  $H \cap H'$ . To argue this, we consider two cases based on whether or not  $A = H \cap H'$  is in  $\mathcal{U}'$ . If  $H \cap H' \in \mathcal{U}'$ , then we have that  $g_H$  and  $g_{H'}$  are consistent, and hence

$$P_H^A = g_H^A \cdot E_H^A = g_{H'}^A \cdot E_{H'}^A = P_{H'}^A.$$

If  $A \notin \mathcal{U}'$ , then  $E_H^A$  is identically zero from the definition of  $E$ , and hence

$$P_H^A = P_{H'}^A \equiv 0.$$

Now, applying the local characterization theorem (Theorem 7.5) to this collection  $\{P_H : H \in \mathcal{H}\}$  of hyperplane polynomials of degree  $d + e$ , we get that if  $t \geq d + e + m$  (which holds from the lower bound on  $t$  in the hypothesis of the theorem), then there is a global  $m$ -variate polynomial  $P$  of degree at most  $(d + e)$  such that for every hyperplane  $H \in \mathcal{H}$ , the restriction of  $P$  on  $H$  equals  $P_H$ .

This implies that for every  $H \in \mathcal{H}$ , the restriction  $E_H$  of  $E$  divides the restriction  $P_H$  of  $P$  on

---

<sup>13</sup>Throughout this proof, we will abuse notation slightly and denote the restriction of polynomials such as  $E$  on any subspace  $A$  of co-dimension 2 by  $E^A$ .



the  $H$ . Since  $t > 2(d + e) + e \geq 2 \max_H \{\deg(P)\} + \deg(E)$ , from [Lemma 7.12](#), we get that the  $m$ -variate polynomial  $E$  divides the  $m$ -variate polynomial  $P$ .

Let  $f(\mathbf{X})$  be the resulting quotient  $P(\mathbf{X})/E(\mathbf{X})$ . Clearly,  $\deg(f)$  is equal to  $\deg(P) - \deg(E)$ , which is at most  $(d + e) - e = d$ . To prove the theorem, it now suffices to show that the restriction  $f_H$  of  $f$  on a hyperplane  $H$  equals the hyperplane polynomial  $g_H$  for many hyperplanes  $H$  in  $\mathcal{H}$ . Once again, we first prove a weaker bound for this and then amplify it to complete the proof of the theorem.

**Agreement with many hyperplane polynomials :** As in the proof of [Theorem 7.11](#), we again notice that if  $E_H$  is not identically zero for a hyperplane  $H \in \mathcal{H}$ , then  $P_H/E_H$  is a polynomial and equals  $P_H/E_H = g_H = f_H$ . We now observe that there are at most  $\deg(E)$  hyperplanes  $H$  in  $\mathcal{H}$  such that  $E_H$  is identically zero. To see this, note that the polynomial  $E_H$  can be obtained from the polynomial  $E$  by substituting one of the variables of  $E$  (w.l.o.g denoted by  $X_m$  here) by an affine linear form  $L$  in the variables  $X_1, X_2, \dots, X_{m-1}$ . Thus,  $E_H$  is identically zero if and only if  $E$  is divisible by  $(X_m - L)$ . So, every hyperplane  $H$  such that  $E_H$  is identically zero yields a linear factor of  $E$ , and all these factors are distinct (since the hyperplanes are distinct and, in fact, in general position). Thus,  $E$  is divisible by the product of all such linear factors, and hence the number of such factors, and therefore, such hyperplanes cannot exceed  $\deg(E)$ . Thus,  $f_H$  equals  $g_H$  for at least  $(t - e)$  of the hyperplanes. Now, we amplify this bound further to complete the proof of the theorem.

**An improved bound :** This final step also follows the outline of this step for the two-dimensional case. For  $H \in \mathcal{H}$ , let  $\mathcal{I}_H \subseteq \mathcal{H}$  be the set of hyperplanes  $H'$  such that  $g_H^{H \cap H'} \neq g_{H'}^{H \cap H'}$ . From the hypothesis of the theorem

$$\mathbb{E}_{H \in \mathcal{H}} [|\mathcal{I}_H|] \leq \varepsilon t.$$

Thus, by Markov's inequality, we get that

$$\Pr_{H \in \mathcal{H}} [|\mathcal{I}_H| \geq ((1 - \sqrt{\varepsilon/2m})t - d - m)] \leq \frac{\varepsilon t}{(1 - \sqrt{\varepsilon/2m})t - d - m} \leq 2\varepsilon,$$

where the last inequality follows from the bounds on  $t$  and  $\varepsilon$  in the theorem. Thus, we have that for  $(1 - 2\varepsilon)t$  of hyperplanes  $H$ , the set  $\mathcal{I}_H$  is of size less than  $((1 - \sqrt{\varepsilon/2m})t - d - m)$ . Any such  $H$  is consistent with at least  $d + m$  of the hyperplanes  $H'$  such that  $E_{H'}$  does not vanish, and  $f_{H'}$  equals  $g_{H'}$ , which in-turn equals  $g_H$ . Let  $\mathcal{A}$  be this set and let  $\mathcal{A}_H = \{H' \cap H : H' \in \mathcal{A}\}$ . Now, if we focus on the  $(m - 1)$ -dimensional linear space  $H$ , then  $\mathcal{A}_H$  is a set of  $(d + m)$  co-dimension one subspaces within  $H$ , which are also in general position. Moreover,  $f_H$  and  $g_H$  agree on each of these subspaces. Thus, they agree on the set of points corresponding to the  $(m - 1)$ -

wise intersection of subspaces in  $\mathcal{A}_H$ . Thus, [Theorem 3.3](#), we get that  $g_H$  and  $f_H$  must be equal to each other.

Therefore, for at least  $(1 - 2\varepsilon)t$  hyperplanes  $H \in \mathcal{H}$ ,  $f_H$  equals  $g_H$ . This completes the proof of the theorem.  $\square$

## C Intersection points of hyperplanes in general position are an interpolating set

We give a quick proof of Bläser and Pandey's [Theorem 3.3](#).

*Proof.* For each  $x \in T$ , we will show that there is a polynomial  $P_x$  in  $\mathbb{F}[X_1, \dots, X_m]$  of degree at most  $d$  which is nonzero at  $x$  and zero at  $T \setminus \{x\}$ . This immediately implies that every function from  $T$  to  $\mathbb{F}$  can be interpolated by a polynomial of degree at most  $d$  – and since  $|T|$  equals the dimension of the space of all polynomials of degree at most  $d$ , we get that  $T$  is an interpolating set.

It remains to show the existence of  $P_x$ . Let  $H_1, \dots, H_m \in \mathcal{H}$  be the  $m$  hyperplanes whose intersection is  $x$ . By the general position assumption about  $\mathcal{H}$ , we have that  $x$  does not lie on any other hyperplane in  $\mathcal{H}$ . Define

$$P_x(X_1, \dots, X_m) = \prod_{H \in \mathcal{H} \setminus \{H_1, \dots, H_m\}} H(X_1, \dots, X_m).$$

(Here, abusing notation, we used  $H(X_1, \dots, X_m)$  for a degree 1 polynomial whose zero set equals the hyperplane  $H$ ). It is easy to see that  $P_x$  does not vanish on  $x$  and vanishes on other points of  $T$ , and that the degree of  $P_x$  is at most  $D$ . This completes the proof.  $\square$