# Decoding Multivariate Multiplicity Codes on Product Sets

Siddharth Bhandari, Prahladh Harsha, Mrinal Kumar, and Madhu Sudan

***Abstract*— The multiplicity Schwartz-Zippel lemma bounds the total multiplicity of zeroes of a multivariate polynomial on a product set. This lemma motivates the multiplicity codes of Kopparty, Saraf and Yekhanin [J. ACM, 2014], who showed how to use this lemma to construct high-rate locally-decodable codes. However, the algorithmic results about these codes crucially rely on the fact that the polynomials are evaluated on a vector space and not an arbitrary product set. In this work, we show how to decode multivariate multiplicity codes of large multiplicities in polynomial time over finite product sets (over fields of large characteristic and zero characteristic). Previously such decoding algorithms were not known even for a positive fraction of errors. In contrast, our work goes all the way to the distance of the code and in particular exceeds both the unique-decoding bound and the Johnson radius. For errors exceeding the Johnson radius, even combinatorial list-decodablity of these codes was not known. Our algorithm is an application of the classical polynomial method directly to the multivariate setting. In particular, we do not rely on a reduction from the multivariate to the univariate case as is typical of many of the existing results on decoding codes based on multivariate polynomials. However, a vanilla application of the polynomial method in the multivariate setting does not yield a polynomial upper bound on the list size. We obtain a polynomial bound on the list size by taking an alternative view of multivariate multiplicity codes. In this view, we glue all the partial derivatives of the same order together using a fresh set $\mathbf{z}$ of variables. We then apply the polynomial method by viewing this as a problem over the field $\mathbb{F}(\mathbf{z})$ of rational functions in $\mathbf{z}$.**

***Index Terms*— Coding theory, list-decoding, multiplicity codes.**

## I. Introduction

THE classical Schwartz-Zippel Lemma (due to Ore [2], Schwartz [3], Zippel [4] and DeMillo and Lipton [5]) states that if $\mathbb{F}$ is a field, and $f \in \mathbb{F}[x_1, x_2, \ldots, x_m]$ is a *non-zero* polynomial of degree $d$, and $S \subseteq \mathbb{F}$ is an arbitrary finite subset of $\mathbb{F}$, then the number of points on the grid[1] $S^m$ where $f$ is zero is upper bounded by $d|S|^{m-1}$. A higher order *multiplicity* version of this lemma (due to Dvir et al. [6]) states that the number of points on the grid $S^m$ where $f$ is zero with *multiplicity*[2] at least $s$ is upper bounded by $\frac{d|S|^{m-1}}{s}$.[3]

This innately basic statement about low degree polynomials has had innumerable applications in both theoretical computer science and discrete mathematics and has by now become a part of the standard toolkit when working with low degree polynomials [7], [8]. Despite this, the following natural algorithmic version of this problem remains open.

*Algorithmic SZ Question:* Let $\mathbb{F}$ be a field, and $S$, $d$, $m$ be as above. Design an efficient algorithm that takes as input an arbitrary function $P : S^m \to \mathbb{F}^{\binom{s+m-1}{m}}$ and finds a polynomial $f \in \mathbb{F}[x_1, x_2, \ldots, x_m]$ of degree at most $d$ (if one exists) such that the function $\mathsf{Enc}(f) : S^m \to \mathbb{F}^{\binom{s+m-1}{m}}$ defined as

$$\mathsf{Enc}(f)(\mathbf{a}) = \left( \frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a}) : \deg(\mathbf{x}^{\mathbf{e}}) < s \right)$$

differs from $P$ on less than $\frac{1}{2}\left(1 - \frac{d}{s|S|}\right)$ fraction of points on $S^m$.

The aforementioned multiplicity Schwartz-Zippel lemma (henceforth, referred to as the multiplicity SZ lemma for brevity) assures us that if there is a polynomial $f \in \mathbb{F}[x_1, x_2, \ldots, x_m]$ such that $\mathsf{Enc}(f)$ differs from $P$ on less than $\frac{1}{2}\left(1 - \frac{d}{s|S|}\right)$ fraction of points, then it must be unique! Thus, in some sense, the above question is essentially asking for an algorithmic version of the multiplicity SZ lemma.

Although a seemingly natural problem, especially given the ubiquitous presence of the SZ lemma in computer science, this question continues to remain open for even bivariate

Manuscript received 25 October 2022; revised 23 June 2023; accepted 1 August 2023. Date of publication 29 August 2023; date of current version 26 December 2023. The work of Siddharth Bhandari was supported in part by the Department of Atomic Energy, Government of India, under Project 12-R&D-TFR-5.01-0500; and in part by the Google Ph.D. Fellowship. The work of Prahladh Harsha was supported in part by the Department of Atomic Energy, Government of India, under Project 12-R&D-TFR-5.01-0500; and in part by the Swarnajayanti Fellowship and Google Research Award. The work of Mrinal Kumar was supported by the Department of Atomic Energy, Government of India, under Project 12-R&D-TFR-5.01-0500. The work of Madhu Sudan was supported in part by the Simons Investigator Award and in part by the NSF under Award CCF 1715187 and Award CCF 2152413. An earlier version of this paper was presented at the Proceedings of the 53rd ACM Symposium on Theory of Computing [DOI: 10.1145/3406325.3451027]. *(Corresponding author: Siddharth Bhandari.)*

Siddharth Bhandari was with the Tata Institute of Fundamental Research, Mumbai 400005, India, and also with the Simons Institute for the Theory of Computing, Berkeley, CA 94720 USA. He is now with the Toyota Technological Institute at Chicago, Chicago, IL 60637 USA (e-mail: siddharth@ttic.edu).

Prahladh Harsha is with the Tata Institute of Fundamental Research, Mumbai 400005, India (e-mail: prahladh@tifr.res.in).

Mrinal Kumar was with the Department of Computer Science and Engineering, IIT Bombay, Mumbai 400076, India. He is now with the Tata Institute of Fundamental Research, Mumbai 400005, India (e-mail: mrinal.kumar@tifr.res.in).

Madhu Sudan is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: madhu@cs.harvard.edu).

Communicated by A.-L. Horlemann-Trautmann, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2023.3306849

[1] We use "grids" and "product sets" interchangeably (see also Remark 1).

[2] This means that all the partial derivatives of $f$ of order at most $s-1$ are zero at this point. See Section III for a formal definition.

[3] This bound is only interesting when $|S| > d/s$ so that $\frac{d|S|^{m-1}}{s}$ is less than the trivial bound of $|S|^m$.

polynomials! In fact, even the $s = 1$ case, which corresponds to an algorithmic version of the classical SZ lemma (without multiplicities) was only very recently resolved in a beautiful work of Kim and Kopparty [9]. Unfortunately, their algorithm does not seem to extend to the case of $s > 1$, and they mention this as one of the open problems.

In this work, we make some progress towards answering the algorithmic SZ question. In particular, we design an efficient deterministic algorithm for this problem when the field $\mathbb{F}$ has characteristic zero or larger than the degree $d$, the dimension $m$ is an arbitrary constant and the multiplicity parameter $s$ is a sufficiently large constant. In fact, in this setting we prove a stronger result, which we now informally state (see Theorem 1 for a formal statement).

*Main Result:* Let $\varepsilon \in (0, 1)$ be an arbitrary constant, $m \in \mathbb{N}$ be a positive constant and $s$ be a large enough positive integer. Over fields $\mathbb{F}$ of characteristic zero or characteristic larger than $d$, there is a deterministic polynomial time algorithm that on input $P$ outputs all degree $d$ polynomials $f \in \mathbb{F}[x_1, x_2, \ldots, x_m]$ such that $\mathsf{Enc}(f)$ differs from the input function $P : S^m \to \mathbb{F}^{\binom{s+m-1}{m}}$ on less than $\left(1 - \frac{d}{s|S|} - \varepsilon\right)$ fraction of points on the grid $S^m$.

We note that the fraction of errors that can be tolerated in the above result is $1-\frac{d}{s|S|}-\varepsilon$, which is significantly larger than the error parameter in the algorithmic SZ question. In this regime, we no longer have the guarantee of a unique solution $f$ such that the function $\mathsf{Enc}(f)$ which is *close* to $P$. In fact, for this error regime, it is not even clear that the number of candidate solutions is polynomially bounded. The algorithm stated in the main result outputs all such candidate solutions, and in particular, shows that their number is polynomially bounded (for constant $m$). This fraction of errors is the best one can hope for since there are functions $P$ (for instance, the all zero's function) which have super-polynomially many polynomials of degree $d$ which are $\left(1 - \frac{d}{s|S|}\right)$-close to $P$. (see Section I).

In the language of error correcting codes, the algorithmic SZ question is the question of designing efficient unique-decoding algorithms for multivariate multiplicity codes over arbitrary product sets when the error is at most half the minimum distance. Our main result gives an efficient algorithm for list-decoding (which includes as a subcase the problem of unique-decoding) these codes from relative error $\delta - \varepsilon$, where $\delta := 1 - \frac{d}{s|S|}$ is the distance of the code, provided that the field has characteristic larger than $d$ or zero, $m$ is a constant and $s$ is large enough. In the next section, we define some of these notions, and state and discuss the results and the prior work in this language.

### A. Multiplicity Codes

Polynomial based error correcting codes, such as the Reed-Solomon codes and Reed-Muller codes, are a very important family of codes in coding theory both in theory and practice. Multiplicity codes are a natural generalization of Reed-Muller codes wherein at each evaluation point, one not only gives the evaluation of the polynomial $f$, but also all its derivatives up to a certain order.

Formally, let $\mathbb{F}$ be a field, $s$ a positive integer, $S \subset \mathbb{F}$ an arbitrary subset of the field $\mathbb{F}$, $d \leq s|S|$ the degree parameter and $m \geq 1$ the ambient dimension. The codewords of the $m$-variate order-$s$ multiplicity code of degree-$d$ polynomials over $\mathbb{F}$ on the grid $S^m$ is obtained by evaluating an $m$-variate polynomial of total degree at most $d$, along with all its derivatives of order less than $s$ at all points in the grid $S^m$. Thus, a codeword corresponding to the polynomial $f$ of total degree at most $d$ can be viewed as a function $\mathsf{Enc}_{s,S}(f)\colon S^m \to \mathbb{F}^{|E|}$ where $E := \{\mathbf{e} \in \mathbb{Z}^m_{\geq 0} \mid 0 \leq \|\mathbf{e}\|_1 < s\}$ and

$$\mathsf{Enc}_{s,S}(f)|_{\mathbf{a}} = \left(\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a}) : \mathbf{e} \in E\right)$$

where $\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}$ is the Hasse derivative of the polynomial $f$ with respect to $\mathbf{x}^{\mathbf{e}}$. The $s = 1$ version of these multiplicity codes corresponds to the classical Reed-Solomon codes (univariate case, $m = 1$) and Reed-Muller codes (multivariate setting, $m > 1$). The distance of these codes is $\delta := 1 - \frac{d}{s|S|}$, which follows from the multiplicity SZ Lemma mentioned earlier in the introduction.

Univariate multiplicity codes were first studied by Rosenbloom and Tsfasman [10] and Nielsen [11]. Multiplicity codes for general $m$ and $s$ were introduced by Kopparty et al. [12] in the context of local decoding. Subsequently, Kopparty [13] and Guruswami and Wang [14] independently proved that the univariate multiplicity codes over prime fields (or more generally over fields whose characteristic is larger than the degree of the underlying polynomials) achieve "list-decoding capacity". In the same work, Kopparty [13] proved that multivariate multiplicity codes are list decodable up to the Johnson radius if the underlying set $S$ has a nice algebraic structure (eg., $S = \mathbb{F}$).

We remark that in the case of univariate multiplicity codes (both Reed-Solomon and larger order multiplicity codes), the decoding algorithms work for all choices of the set $S \subset \mathbb{F}$. However, all decoding algorithms for the multivariate setting (both Reed-Muller and larger order multiplicity codes) work only when the underlying set $S$ has a nice algebraic structure (eg., $S = \mathbb{F}$) or when the degree $d$ is very small (cf, the Reed-Muller list-decoding algorithm of Sudan [15] and its multiplicity variant due to Guruswami and Sudan [16]). The only exception to this is the unique-decoding algorithm of Kim and Kopparty [9] of Reed-Muller codes over product sets.

### B. Our Results

Below we state and contrast our results on the problem of decoding multivariate multiplicity codes (over grids) from a $\delta - \varepsilon$ fraction of errors for any constant $\varepsilon \in (0, 1)$ where $\delta$ is the distance of the code. Our first result is as follows.

*Theorem 1 (List-Decoding of Multivariate Multiplicity Codes With Polynomial List Size):* For every $\varepsilon \in (0, 1)$ and integer $m$, for all $s \geq \frac{4}{\varepsilon} \cdot \binom{(20/\varepsilon)^m + m}{m}$, degree parameters $d$, fields $\mathbb{F}$ of size $q$ and characteristic larger than $d$, and any set $S \subseteq \mathbb{F}$ where $d < s|S|$, the following holds.

For an $m$-variate order-$s$ multiplicity code of degree-$d$ polynomials over $\mathbb{F}$ on the grid $S^m$, there is an efficient algorithm which when given a received word $P$, outputs all code words with agreement at least $(1 - \delta + \varepsilon)$ with $P$, where $\delta = 1-d/(s|S|)$ is the relative distance of this code. Moreover,

the algorithm requires $\mathrm{poly}(|S|^{m^2}, d^{m^2})$ operations over the field $\mathbb{F}$.

*Remark 1:* A general product set in $\mathbb{F}^m$ is of the form $S_1 \times S_2 \times \cdots S_m$, where each $S_i$ is a subset of $\mathbb{F}$. For ease of notation, we always work with product sets which are grids $S^m$ for some $S \subseteq \mathbb{F}$ even though all of our results hold for general product sets.

As indicated before, this is the best one can hope for with respect to polynomial time list-decoding algorithms for multiplicity codes since there are super-polynomially many codewords with minimum distance $\delta = 1 - d/(s|S|)$ (see Section I). Until recently, it was not known if multivariate multiplicity codes were list decodable beyond the Johnson radius (even for the case $S = \mathbb{F}$). For the case of grids $S^m$, where $S \subseteq \mathbb{F}$ is an arbitrary set, even unique-decoding algorithms were not known. We note that the above result does not yield a list-decoding algorithm for all multiplicities, but only for large enough multiplicities (based on the dimension $m$ and the error parameter $\varepsilon$).

Kopparty et al. [17] showed how to reduce the size of the list for univariate multiplicity codes from polynomial to constant (dependent only on the error parameter $\varepsilon$). We use similar ideas, albeit in the multivariate setting, to reduce the list size in Theorem 1 to constant (dependent only on the error parameter $\varepsilon$ and the dimension $m$).

*Theorem 2 (List-Decoding of Multivariate Multiplicity Codes With Constant List Size):* For every $\varepsilon \in (0,1)$ and integer $m$, for all $s \geq \frac{4}{\varepsilon} \cdot \binom{(20/\varepsilon)^m + m}{m}$, degree parameter $d$, fields $\mathbb{F}$ of size $q$ and characteristic larger than $d$, and any set $S \subseteq \mathbb{F}$ where $d < s|S|$, the following holds.

For an $m$-variate order-$s$ multiplicity code of degree-$d$ polynomials over $\mathbb{F}$ on the grid $S^m$, there is a randomized algorithm which requires $\mathrm{poly}\left(d^{m^2}, |S|^{m^2}, \exp\left(O\left(\frac{m^2}{\varepsilon}\log^3\frac{1}{\varepsilon}\right)\right)\right)$ operations over the field $\mathbb{F}$ and which when given a received word $P$, outputs all code words with agreement at least $(1 - \delta + \varepsilon)$ with $P$, where $\delta = 1 - d/(s|S|)$ is the relative distance of this code.

Moreover, the number of such codewords is at most $\exp\left(O\left(\frac{m^2}{\varepsilon}\log^2\frac{1}{\varepsilon}\right)\right)$.

*Remark 2:* We remark that by taking a slightly different view of the list-decoding algorithm in Theorem 1 and Theorem 2, the upper bound on the number of field operations needed in Theorem 1 and Theorem 2 can be improved to $\mathrm{poly}(|S|^m, d^m)$. We sketch this view in subsection IV-G and note the runtime analysis in Remark 5.

The above two results are a generalization of (and imply) the corresponding theorems for the univariate setting due to Kopparty [13] and Guruswami and Wang [14] and Kopparty et al. [17]. We remark that Kopparty et al. [17] in the recent improvement to their earlier work prove a similar list-decoding algorithm for multivariate multiplicity codes as Theorem 2 for the case when $S = \mathbb{F}$. Though their list-decoding algorithm does not extend to products sets, it has the added advantage that it is *local*.

As noted earlier the only previous algorithmic method for decoding polynomial-based codes over product sets was that of Kim and Kopparty [9]. We describe the ideas in our algorithm shortly (in Section II), but stress here that our approach is very different from that of Kim and Kopparty. Their work may be viewed as an algorithmic version of the inductive proof of the SZ lemma, and indeed recovers the SZ lemma as a consequence. Their work uses algorithmic aspects of algebraic decoding as a black box (to solve univariate cases). Our work, in contrast, only relies on the multiplicity SZ lemma as a black box. Instead, we open up the "algebraic decoding" black box and make significant changes there, thus adding to the toolkit available to deal with polynomial evaluations over product sets.

### C. Further Discussion and Open Problems

Our result falls short of completely resolving the algorithmic SZ question in two respects; though it works for all dimensions $m$ it only works when the multiplicity parameter $s$ is large enough and when the characteristic of the field is either zero or larger than the degree parameter. Making improvements on any of these fronts is an interesting open problem.

1) *All multiplicities:* The algorithms presented in this paper decode all the way up to distance if the multiplicity parameter $s$ is large enough. Recently, the work of [18] addressed the unique-decoding question when the multiplicity parameter $s$ is small. Specifically, they gave an algorithm to unique decode which runs in time $(sn)^{O(s+m)} \cdot \binom{s-1+m}{m}$. This is polynomial in the input size, i.e., $\left(n^m \cdot \binom{s-1+m}{m}\right)$, when $(sn)^{s+m}$ is polynomial in the input size: hence, for the case of $s = O(1)$ this is a truly polynomial time algorithm. Combining our algorithm with theirs, we get polynomial time unique-decoding algorithms for *all* settings of the multiplicity parameter $s$ when the dimension $m$ and the fractional distance $1 - \frac{d}{s|T|}$ are constant. However, for small multiplicities, the list-decoding problem is open. For $s = 1$, the result due to Kim and Kopparty [9] addresses the unique-decoding question, but the list-decoding question for product sets is open.
2) *Fields of small characteristic:* All known proofs of list-decoding multiplicity codes beyond the Johnson radius (both algorithmic and combinatorial) require the field to be of zero characteristic or large enough characteristic. The problem of list-decoding multiplicity codes over small characteristic beyond the Johnson radius is open even for the univariate setting. As pointed to us by Swastik Kopparty, this problem of list-decoding univariate multiplicity codes over fields of small characteristic beyond the Johnson radius is intimately related to list-decoding Reed-Solomon codes beyond the Johnson radius.

For a more detailed discussion of multiplicity codes and related open problems, we refer the reader to the excellent survey by Kopparty [19].

*1) Organization:* The rest of this paper is organized as follows. We begin with an overview of our proofs in Section II followed by some preliminaries (involving Hasse derivatives, their properties, multiplicity codes) in Section III. We then describe and analyze the list-decoding algorithm for multivariate multiplicity codes in Section IV, thus proving Theorem 1. In Section V, we then show how

to further reduce the list size to a constant, thus proving Theorem 2. In Section VI, we prove some properties of subspace restriction of multivariate multiplicity codes needed in Section V. In Section I, we show that there are super-polynomially many minimum-weight codewords, thus proving the tightness of Theorem 1 and 3 with respect to list-decoding radius.

## II. Proof Overview

In this section, we first describe some of the hurdles in extending the univariate algorithms of Kopparty [13] and Guruswami and Wang [14] to the multivariate setting, especially for product sets and then given a detailed overview of the proofs of Theorem 1 and Theorem 2.

### A. Background and Motivation for Our Algorithm

To explain our algorithm, it will be convenient to recall the general polynomial method framework underlying the list-decoding algorithms in the univariate setting due to Kopparty [13] and Guruswami and Wang [14]. Let $P : S \to \mathbb{F}^s$ be the received word and $1 \leq w \leq s$

- *Step 1: Algebraic Explanation.* Find a polynomial $Q(x, y_1, \ldots, y_w) \in \mathbb{F}[x, y_1, \ldots, y_w]$ of appropriate degree constraints that "explains" the received word $P$.
- *Step 2: $Q$ contains the close codewords.* Show that every low-degree polynomial $f$ whose encoding agrees with $P$ in more than $(1 - \delta + \varepsilon)$-fraction of points satisfies the following condition.

$$Q\left(x, f(x), \frac{\partial f}{\partial x}, \frac{\partial f}{\partial x^2}, \ldots, \frac{\partial f}{\partial x^{w-1}}\right) = 0.$$

- *Step 3: Reconstruction step.* Recover every polynomial $f$ that satisfies the above condition.

The main (and only) difference between the list-decoding algorithms of Kopparty [13] and Guruswami and Wang [14] is that Guruswami and Wang observe that it suffices to work with a polynomial $Q$ which is linear in the $y$-variables,[4] more precisely, $Q(x, y_1, \ldots, y_w)$ of the form $Q_0(x) + Q_1(x) \cdot y_1 + \cdots + Q_w(x) \cdot y_w$, while Kopparty allows for larger degrees in the $y$-variables. As a result, Kopparty performs the recovery step by solving a differential equation while Guruswami and Wang observe that due to the simple structure of $Q$, the solution can be obtained by solving a linear system of equations.

How is multivariate list-decoding performed? There are by now two standard approaches. Inspired by the Pellikaan-Wu [21] observation that Reed-Muller codes are a subcode of Reed-Solomon codes over an extension field, Kopparty performs a similar reduction of the multivariate multiplicity code to a univariate multiplicity code over an extension field. Another approach is to solve the multivariate case by solving the univariate subproblem on various lines in the space. However, both these approaches work only if the set $S = \mathbb{F}$ or has some special algebraic structure.

For our proof, we take an alternate approach and always work in the multivariate setting without resorting to a reduction to the univariate setting. As we shall see, our approach has some advantages over that of Kopparty [13], both in quantitative terms, since the algorithm can tolerate a larger number of errors, and in qualitative terms, since the underlying set of evaluation points does not have to be an algebraically nice subset of $\mathbb{F}^m$ as in [13]; evaluations on an arbitrary grid suffice for the algorithm to work.

To extend the univariate list-decoding algorithm outlined above to the multivariate setting, we adopt the following approach. We consider a new set of formal variables $\mathbf{z}$ and instead of directly working with the information about partial derivatives in the received word, we think of the partial derivatives of the same order as being *glued* together using monomials in $\mathbf{z}$. With this reorganized (and somewhat mysterious) view of the partial derivatives, we follow the outline of the univariate setting as described above. We find a linear polynomial $Q$ with coefficients from the field of fractions $\mathbb{F}(\mathbf{z})$ instead of just $\mathbb{F}$ in the interpolation step to explain the received word $P$. Thus, in this instance, the linear system in the interpolation step is over the field $\mathbb{F}(\mathbf{z})$. We then argue that $Q$ *contains* information about all the codewords that are close to the received word, and eventually *solve* $Q$ to recover all the codewords close to the received word. This might seem rather strange to begin with, but these ideas of gluing together the partial derivatives and working over the field $\mathbb{F}(\mathbf{z})$ immediately generalize the univariate list-decoding algorithm to the multivariate setting. Working with this field of fractions $\mathbb{F}(\mathbf{z})$ comes with its costs; it makes some of the steps costly and in particular, the recovery step far more elaborate than that in the Guruswami-Wang setting. However, this recovery step happens to be a special case of similar step in the recent work of Guo et al. [22] and we adapt their algorithm to our setting.

As a first attempt, a more standard way to generalize the algorithms of Kopparty [13] and Guruswami and Wang [14] to the multivariate setting would have been to work with the partial derivatives directly. And, while this approach seems alright for the interpolation step, it seems hard to work with when we try to solve the resulting equation to recover all the close enough codewords. In particular, it isn't even clear in this set up that the number of solutions of the algebraic explanation (and hence, the number of close enough codewords) is polynomially bounded. This mysterious step of gluing together derivatives of the same order in a reversible manner (in the sense that we can read off the individual derivatives from the glued term) gets around this problem, and makes it viable to prove a polynomial upper bound on the number of solutions, and eventually solve the equation to recover all the close enough codewords.

Given this background, we now give a more detailed outline of our algorithm below.

### B. Theorem 1: Multivariate List-Decoding Algorithm With Polynomially-Sized Lists

*1) Viewing the Encoding as a Formal Power Series:* Multiplicity codes are described by saying that the encoding of a polynomial $f \in \mathbb{F}[\mathbf{x}]$ consists of the evaluation of *all* partial derivatives of $f$ of order at most $s - 1$ at every point in the

[4]The authors attribute this observation to Vadhan [20] in [14].

appropriate evaluation set, e.g. the grid $S^m$. For our algorithm, we think of these partial derivatives of $f$ as being rearranged on the basis of the order of the derivatives as follows. We take a fresh set of formal variables $\mathbf{z}$ and define the following differential operators.

$$\Delta_i(f) := \sum_{\mathbf{e}:\|\mathbf{e}\|_1=i} \mathbf{z}^{\mathbf{e}} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}}$$

where $\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}$ denotes the Hasse derivative[5] of the polynomial $f$ with respect to $\mathbf{x}^{\mathbf{e}}$.

Let $\Delta(f)$ be an $s$ tuple of polynomials defined as follows.

$$\Delta(f) := (\Delta_0(f), \Delta_1(f), \ldots, \Delta_{s-1}(f)).$$

We view the encoding for $f$ as giving us the evaluation of the tuple $\Delta(f) \in \mathbb{F}[\mathbf{x}, \mathbf{z}]$ as $\mathbf{x}$ varies in $S^m$. Note that for every fixing of $\mathbf{x}$ to some $\mathbf{a} \in S^m$, $\Delta(f)(\mathbf{a})$ is in $\mathbb{F}[\mathbf{z}]^s$. Thus, the alphabet size is still large. Clearly, this is just a change of viewpoint, as we can go from the original encoding to this and back efficiently, and at this point it is unclear that this change of perspective would be useful.

*2) Finding an Equation Satisfied by All Close Enough Codewords:* Let $P$ be a received word. We view $P$ as a function $P : S^m \to \Sigma^s$, where $\Sigma = \mathbb{F}[\mathbf{z}]$, as discussed in the previous step. The goal of the decoding step is to find all the polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$, whose encoding is close enough to $P$.

As a first step towards this, we find a non-zero polynomial $Q(\mathbf{x}, \mathbf{y}) \in \mathbb{F}(\mathbf{z})[\mathbf{x}, \mathbf{y}]$ of the form

$$Q(\mathbf{x}, \mathbf{y}) = Q_1(\mathbf{x})y_1 + \cdots + Q_w(\mathbf{x})y_w,$$

which *explains* the received word $P$, i.e., for every $\mathbf{a} \in S^m$, $Q(\mathbf{a}, P(\mathbf{a})) = 0$, and $Q$ satisfies some appropriate degree constraints. Here $w \leq s$ is a parameter. For technical reasons, we also end up imposing some more constraints on $Q$ in terms of its partial derivatives, the details of which can be found in Section IV-C. Each of these constraints can be viewed as a homogeneous linear equation in the coefficients of $Q$ over the field $\mathbb{F}(\mathbf{z})$. We choose the degree of $Q$ to be large enough to ensure that this system has more variables than constraints, and therefore, has a non-zero solution.

This step is the interpolation step which shows up in any standard application of the polynomial method, and our set-up is a natural generalization of the set up in the list-decoding algorithm of Guruswami and Wang [14] for univariate multiplicity codes.

The key property of the polynomial $Q$ thus obtained is that for every degree $d$ polynomial $f \in \mathbb{F}[\mathbf{x}]$ whose encoding is close enough to $P$,

$$Q(\mathbf{x}, \Delta(f)) = Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

To see this, we note that from the upper bound on the degree of $Q$ and the fact that $f$ has degree at most $d$, the polynomial $Q(\mathbf{x}, \Delta(f)) \in \mathbb{F}(\mathbf{z})[\mathbf{x}]$ is of not too high degree in $\mathbf{x}$. Moreover, from the constraints imposed on $Q$ during interpolation, it follows that at every $\mathbf{a} \in S^m$ where the encoding of $f$ and $P$ agree, $Q(\mathbf{x}, \Delta(f))$ vanishes with high multiplicity. Thus, if the parameters are favorably set, it follows that $Q(\mathbf{x}, \Delta(f))$ has too many zeroes of high multiplicity on a grid, and hence by the multiplicity Schwartz-Zippel Lemma (see Lemma 1), $Q(\mathbf{x}, \Delta(f))$ must be identically zero.

We note that this is the only place in the proof where we use anything about the structure of the set of evaluation points, i.e., the set of evaluation points is a grid.

*3) Solving the Equation to Recover All Close Enough Codewords:* As the final step of our algorithm, we try to recover all polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ such that

$$Q(\mathbf{x}, \Delta(f)) = Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

$Q(\mathbf{x}, \Delta(f))$ can be viewed as a partial differential equation of order $w-1$ and degree one, and we construct all candidate solutions $f$ via the method of power series. We start by trying all possible choices of field elements for coefficients of monomials of degree at most $w-1$ in $f$, and iteratively recover the remaining coefficients of $f$ by reconstructing $f$ one homogeneous component at a time. Moreover, we observe that for each choice of the initial coefficients, there is a unique lift to a degree $d$ polynomial. Thus, the number of solutions is upper bounded by the number of initial choices, which is at most $|\mathbb{F}|^{\binom{w+m-1}{m}}$.

We note that this is one place where working with $\Delta_i(f)$ as opposed to having an equation in the individual partial derivatives of $f$ is of crucial help. Even though the equation $Q(\mathbf{x}, \Delta(f)) = 0$ is a partial differential equation of high order in $f$, the fact that these derivatives appear in a structured form via the operators $\Delta_i(f)$ helps us prove a polynomial upper bound on the number of such solutions and solve for $f$. Without this additional structure, it is unclear if one can prove a polynomial upper bound on the number of solutions of the corresponding equation.

This reconstruction step is a multivariate generalization of similar reconstruction steps in the list-decoding algorithms of Kopparty [13] and Guruswami and Wang [14] for univariate multiplicity codes. Interestingly, this is also a special case of a similar reconstruction procedure in the work of Guo et al. [22], where the polynomial $Q$ could potentially be of higher degree in $\mathbf{y}$ variables, and is given to us via an arithmetic circuit of small size and degree and the goal is to show that all (low degree) polynomials $f$, satisfying $Q(\mathbf{x}, \Delta(f)) \equiv 0$ have small circuits. In contrast, we are working with $Q$ which is linear in $\mathbf{y}$ and we have access to the coefficient representation of this polynomial, and construct the solutions $f$ in the monomial representation. As a consequence, the details of this step are much simpler here, when compared to that in [22].

In this step of our algorithm viewing the encoding in terms of the differential operators $\Delta_i()$ turns out to be useful. The iterative reconstruction outlined above crucially uses the fact that for any homogeneous polynomial $g \in \mathbb{F}[\mathbf{x}]$ of degree $r$, $\Delta_i(g)$ is a homogeneous polynomial in the $\mathbf{x}$ variables of degree exactly $r-i+1$. The other property that we use from $\Delta_i()$ is that given $\Delta_i(g)$ for any homogeneous polynomial $g$,

[5]Since we have both $\mathbf{x}$ and $\mathbf{z}$ variables, we use the notation $\frac{\partial f}{\partial x}$ to denote the Hasse derivative wrt variable $x$ to explicitly indicate which variable the derivative is being taken.

we can uniquely read off all the partial derivatives of order $i-1$ of $g$, and via a folklore observation of Euler, uniquely reconstruct the polynomial $g$ itself (see Lemma 6).

Finally, we note that the precise way of *gluing* together the partial derivatives of order $i$ in the definition of the operator $\Delta_i()$ is not absolutely crucial here, and as is evident in Lemma 6, many other candidates would have satisfied the necessary properties.

The details of this step are in Section IV-E, and essentially complete the proof of Theorem 1.

### C. Theorem 2: Reducing the List Size to a Constant

In Section V, we combine our proof of Theorem 1 with the techniques in the recent work of Kopparty et al. [17] to show that the list size in the decoding algorithm in Theorem 1 can be reduced to a constant.

The key to this step is the observation that since $Q(\mathbf{x},\mathbf{y})$ is linear in the $\mathbf{y}$ variables, the solutions $f$ of the equation $Q(\mathbf{x},\Delta(f))\equiv 0$ form an affine subspace of polynomials. The reconstruction algorithm in Section IV-E in fact gives us an affine subspace $V\subseteq\mathbb{F}[\mathbf{x}]$ of polynomials of degree at most $d$ which consists of all the solutions of $Q(\mathbf{x},\Delta(f))\equiv 0$.

This is precisely the setting in the work of Kopparty et al. [17] in the context of folded Reed-Solomon codes and univariate multiplicity codes, and we essentially apply their ideas off the shelf, and combine them with our proof of Theorem 1 to reduce the list size to a constant.

In general, this idea of solving $Q(\mathbf{x},\Delta(f))\equiv 0$ to recover a subspace, and then using the ideas in [17] to recover codewords in the subspace which are close to the received word has the added advantage that it can be applied over all fields. As an immediate consequence, we get an analog of Theorem 1 over infinite fields like rationals as well.

## III. Notation & Preliminaries

### A. Notation

We use the following notation.

- $\mathbb{F}$ is the field we work over, and we assume the characteristic of $\mathbb{F}$ to be either zero or larger than the degree parameter $d$ of the message space.
- We use bold letters to denote tuples of variables (i.e., $\mathbf{x}$, $\mathbf{z}$, $\mathbf{y}$ for $(x_1,\ldots,x_m)$, $(z_1,\ldots,z_m)$ and $(y_1,\ldots,y_w)$ respectively).
- We work with polynomials which are in general members of $\mathbb{F}(\mathbf{z})[\mathbf{x},\mathbf{y}]$. We denote monomials in $\mathbf{x}$ and $\mathbf{z}$ by $\mathbf{x}^{\mathbf{e}}$ $(=\prod_{i\in[m]}x_i^{e_i})$, $\mathbf{z}^{\mathbf{e}}$ $(=\prod_{i\in[m]}z_i^{e_i})$ respectively where $\mathbf{e}\in\mathbb{Z}^m_{\geq 0}$. The degree of the monomial is $\|\mathbf{e}\|_1=\sum_{i=1}^m e_i$.
- For $\mathbf{e},\mathbf{e}'\in\mathbb{Z}^m_{\geq 0}$ we say $\mathbf{e}'\leq\mathbf{e}$ iff for all $i\in[m]$ we have $e'_i\leq e_i$. Also, we use $\binom{\mathbf{e}}{\mathbf{e}'}$ to denote $\prod_{i\in[m]}\binom{e_i}{e'_i}$.
- For a natural number $n$, $[n]$ denotes the set $\{1,2,\ldots,n\}$.

### B. Hasse Derivatives

Throughout the paper we work with Hasse derivatives: we interchangeably use the term partial derivatives.

*Definition 1 (Hasse Derivative):* For a polynomial $f\in\mathbb{F}[\mathbf{x}]$ the Hasse derivative of type $\mathbf{e}$ is the coefficient of $\mathbf{z}^{\mathbf{e}}$ in the polynomial $f(\mathbf{x}+\mathbf{z})\in\mathbb{F}[\mathbf{x},\mathbf{z}]$. We denote this by $\frac{\partial f}{\partial\mathbf{x}^{\mathbf{e}}}$ or $\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}^{\mathbf{e}}}$

We state some basic properties of Hasse Derivatives below. Some of these are taken from [6, Proposition 4].

*Proposition 1 (Basic Properties of Hasse Derivatives):* Let $f,g\in\mathbb{F}[\mathbf{x}]$ and consider $\mathbf{e},\mathbf{e}'\in\mathbb{Z}^m_{\geq 0}$.

1) $\frac{\partial f}{\partial\mathbf{x}^{\mathbf{e}}}+\frac{\partial g}{\partial\mathbf{x}^{\mathbf{e}}}=\frac{\partial(f+g)}{\partial\mathbf{x}^{\mathbf{e}}}$.
2) If $f$ is a homogeneous polynomial of degree $d$ then $\frac{\partial f}{\partial\mathbf{x}^{\mathbf{e}}}$ is homogeneous polynomial of degree $d-\|\mathbf{e}\|_1$.
3) If $f=\mathbf{x}^{\mathbf{e}'}$ then $\frac{\partial f}{\partial\mathbf{x}^{\mathbf{e}}}=\binom{\mathbf{e}}{\mathbf{e}'}\mathbf{x}^{\mathbf{e}-\mathbf{e}'}$.
4) Hasse derivatives compose in the following manner:
$$\frac{\partial}{\partial\mathbf{x}^{\mathbf{e}}}\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}^{\mathbf{e}'}}=\binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}}\cdot\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}^{\mathbf{e}+\mathbf{e}'}}.$$
5) Product rule for Hasse derivatives:
$$\frac{\partial\left(\prod_{i\in[w]}f_i\right)}{\partial\mathbf{x}^{\mathbf{e}}}=\sum_{\mathbf{u}_1+\mathbf{u}_2+\ldots+\mathbf{u}_w=\mathbf{e}}\left(\prod_{i\in[w]}\frac{\partial f_i}{\partial\mathbf{x}^{\mathbf{u}_i}}\right).$$

*Proof:* Items 1 to 3 and 5 follow directly from Definition 1. For Item 4, observe that by linearity of Hasse derivatives we may assume WLOG that $f$ is a monomial, say $\mathbf{x}^{\tilde{\mathbf{e}}}$: in this case the claim follows from Item 3 and the fact that $\binom{\tilde{\mathbf{e}}}{\mathbf{e}}\cdot\binom{\tilde{\mathbf{e}}-\mathbf{e}}{\mathbf{e}'}=\binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}}\cdot\binom{\tilde{\mathbf{e}}}{\mathbf{e}+\mathbf{e}'}$. □

### C. Multiplicity Code

We now define the notion of multiplicity of a polynomial $f\in\mathbb{F}[\mathbf{x}]$ at a point $\mathbf{a}\in\mathbb{F}^m$. The multiplicity of $f$ at the origin is $\ell$ iff $\ell$ is the highest integer such that no monomial of total degree less than $\ell$ appears in the coefficient representation of $f$. We formalize this below using Hasse derivatives.

*Definition 2 (Multiplicity):* A polynomial $f\in\mathbb{F}[\mathbf{x}]$ is said to have multiplicity $\ell$ at a point $\mathbf{a}\in\mathbb{F}^m$, denoted by $\mathrm{mult}(f,\mathbf{a})$, iff $\ell$ is the largest integer such that for all $\mathbf{e}\in\mathbb{Z}^m_{\geq 0}$ with $\|\mathbf{e}\|_1<\ell$ we have $\frac{\partial f}{\partial\mathbf{x}^{\mathbf{e}}}(\mathbf{a})=0$. If no such $\ell$ exists then $\mathrm{mult}(f,\mathbf{a})=\infty$.

Dvir, Kopparty, Saraf and Sudan proved the following higher order multiplicity version of the classical Schwartz-Zippel lemma.

*Lemma 1 (multiplicity SZ Lemma [6, Lemma 2.7]):* Let $\mathbb{F}$ be any field and let $S$ be an arbitrary subset of $\mathbb{F}$. Then, for any non-zero $m$-variate polynomial $P$ of degree at most $d$,
$$\sum_{\mathbf{a}\in S^m}\mathrm{mult}(P,\mathbf{a})\leq d|S|^{m-1}.$$

The above lemma implies the classical SZ lemma, which states that two distinct $m$-variate polynomials of degree $d$ cannot agree everywhere on a grid $S^m$ for any set $S$ of size larger than $d$ trivially. This in particular tells us that the grid $S^m$ serves as hitting set for polynomials of degree at most $d$ provided $d<|S|$.

As mentioned before, a multiplicity code over a grid $S^m$ consists of evaluations of the message polynomial $f$ along with its derivatives of various orders (up to $s-1$), at the points of the grid.

*Definition 3 (multiplicity Code):* Let $s,m\in\mathbb{N}$, $d\in\mathbb{Z}_{\geq 0}$, $\mathbb{F}$ a field and $S\subset\mathbb{F}$ a non-empty finite subset. The $m$-variate

order-$s$ multiplicity code of degree-$d$ polynomials over $\mathbb{F}$ on the grid $S^m$ is defined as follows.

Let $E := \{\mathbf{e} \in \mathbb{Z}^m_{\geq 0} \mid 0 \leq \|\mathbf{e}\|_1 < s\}$. Note that $|E| = \binom{s+m-1}{m}$. The code is over alphabet $\mathbb{F}^E$ and has length $|S|^m$ (where the coordinates are indexed by elements of $S^m$).

The code is an $\mathbb{F}$-linear map from the space of degree $d$ polynomials in $\mathbb{F}[\mathbf{x}]$ to $\left(\mathbb{F}^E\right)^{S^m}$. The encoding of $f \in \mathbb{F}[\mathbf{x}]$ at a point $\mathbf{a} \in S^m$ is given by:

$$\mathsf{Enc}_{s,S}(f)|_{\mathbf{a}} = \left(\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a}) : \mathbf{e} \in E\right).$$

□

*Remark 3:*

- The distance of the code is exactly $\delta := 1 - \frac{d}{s|S|}$ and the rate of the of the code is $\frac{\binom{d+m}{m}}{\binom{s+m-1}{m} \cdot |S|^m}$.
- As mentioned in the introduction we can also view the encoding by clubbing partial derivatives of the same degree. Thus, the encoding of $f$ at a point $\mathbf{a}$ is $(\Delta_0(f)(\mathbf{a}), \Delta_1(f)(\mathbf{a}), \ldots, \Delta_{s-1}(f)(\mathbf{a})) \in \mathbb{F}[\mathbf{z}]^s$ where $\Delta_i(f)(\mathbf{a}) = \sum_{\mathbf{e}:\|\mathbf{e}\|_1=i} \mathbf{z}^{\mathbf{e}} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a})$.
- We think of $m$, $w$ and $s$ as constants, but $w$ much larger than $m$ and $s$ is much larger than $w$. The precise trade-offs will be alluded to when we need to set parameters in our proofs.

### D. Arithmetic Circuits and Related Notions

We now define the notion of arithmetic circuits and some properties that would be helpful for some of the technical statements later in this paper. For a detailed introduction to arithmetic circuits and the broad area of algebraic circuit complexity, we refer to the excellent survey of Shpilka and Yehudayoff [23].

*Definition 4 (Arithmetic Circuits):* An arithmetic circuit $C$ over a field $\mathbb{F}$ and an $n$-tuple $\mathbf{x}$ of variables is a directed acyclic graph where all the vertices of in-degree zero (referred to as leaf gates) are labelled by variables in $\mathbf{x}$ or elements of the field $\mathbb{F}$ and every other vertex is labelled by the symbol $(+)$ or the symbol $\times$ (and hence called a sum gate and a product gate respectively). The vertices of $C$ of out-degree zero are referred to as output gates.

An arithmetic circuit computes a polynomial in $\mathbb{F}[\mathbf{x}]$ in a natural inductive manner: a leaf gate computes the polynomial equal to its label, a sum gate $u$ computes the polynomial that is the sum of polynomials computed at all the gates $v_1, v_2, \ldots, v_t$ of $C$ such that $(v_t, u)$ is an edge in $C$ and a product gate $u$ computes the polynomial that is the product of polynomials computed at all the gates $v_1, v_2, \ldots, v_t$ of $C$ such that $(v_t, u)$ is an edge in $C$.

The size of $C$ is the number of edges in the underlying graph of $C$.

A very well studied family of polynomials in algebraic complexity is the determinant polynomial that we now formally define.

*Definition 5 (Determinant Polynomial):* Let $\mathbb{F}$ be any field and $n \in \mathbb{N}$ be a natural number. Then the determinant polynomial for $n \times n$ matrices over $\mathbb{F}$ that we denote by $\det_n$ is a polynomial of degree $n$ on $n^2$ variables $\{x_{i,j} : 1 \geq i, j \leq n\}$ and is equal to the determinant of the $n \times n$ matrix $X$ whose $(i, j)$ entry equals the variable $x_{i,j}$, i.e.,

$$\det_n := \sum_{\sigma \in S_n} (-1)^{\text{sign}(\sigma)} \prod_{i=1}^{n} x_{i,\sigma(i)},$$

where $S_n$ is the set of all permutations of $1, 2, \ldots, n$.

The following theorem gives an upper bound on the size of arithmetic circuits computing $\det_n$.

*Theorem 3 [24]:* Let $\mathbb{F}$ be any field. There exists a constant $c$ such that for all sufficiently large $n \in \mathbb{N}$, $\det_n$ can be computed by an arithmetic circuit of size at most $O(n^c)$ over $\mathbb{F}$.

Moreover, there is an algorithm that on input $n$, runs in time at most $O(n^c)$ and outputs the aforementioned circuit for $\det_n$.

We now define the notion of a hitting set that will be crucially used in our proof.

*Definition 6 (Hitting Set):* Let $\mathbb{F}$ be any field, $n$ be a natural number and $\mathcal{C}$ be a set of polynomials on $n$ variables with coefficients in $\mathbb{F}$. Then, a set of points $H \subseteq \mathbb{F}^n$ is said to be a hitting set for $\mathcal{C}$ if for every non-zero polynomial $P \in \mathcal{C}$, there is a point $\mathbf{a} \in H$ such that $P(\mathbf{a}) \neq 0$.

One example of a hitting set that is a direct consequence of Lemma 1 is the following. Let $\mathcal{C}$ be the set of all polynomials on $n$ variables and degree $d$. The any set $H$ of the form $S^n = S \times S \times \cdots \times S$ for any $S \subseteq \mathbb{F}$ of size greater than $d$ is a hitting set for $\mathcal{C}$.

### E. Computing on Polynomial Rings

In this section, we state a few basic results that show how to perform algebraic operations over polynomial rings.

The following lemma, proved via an easy application of polynomial interpolation, lets us construct the coefficient representation of a polynomial given an arithmetic circuit for it.

*Lemma 2:* Let $m \in \mathbb{N}$. There exists a deterministic algorithm that takes as input an arithmetic circuit $C$ of size $s$ that computes an $m$-variate polynomial $P \in \mathbb{F}[\mathbf{z}]$ of degree at most $d$ and outputs the coefficient vector of $P$ in at most $\text{poly}(d^m, s)$ field operations over $\mathbb{F}$

*Proof:* From Lemma 1, we know that no two degree $d$ polynomials can agree everywhere on a grid of size larger than $d$. So, we pick an arbitrary subset $S$ of $\mathbb{F}$ of size $d + 1$ and evaluate the circuit $C$ at all points on the grid $|S|^m$. This requires at most $\text{poly}(d^m, s)$ field operations. Now, given these evaluations, we set up a linear system in the coefficients of $P$ where for every $\mathbf{a}$ in the grid, we have a constraint of the form $P(\mathbf{a}) = C(\mathbf{a})$. We know that this system has a solution. Furthermore, from Lemma 1, we know that this system has a unique solution.

Solving this system gives us the coefficient vector of $P$ and requires at most $d^m$ additional field operations. □

The next lemma tells us how to perform linear algebra over the polynomial ring $\mathbb{F}[\mathbf{z}]$.

*Lemma 3 (Linear Algebra Over Polynomial Rings):* Let $A(\mathbf{z}) \in \mathbb{F}[\mathbf{z}]^{t' \times t}$ be a matrix such that each of its entries is a polynomial of degree at most $w$ in the variables $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ and $t' \leq t$. Then, there is a deterministic algorithm which takes as input the coefficient vectors of the

entries of $A$ and outputs a non-zero vector $\mathbf{u} \in \mathbb{F}[\mathbf{z}]^t$ in time $\mathrm{poly}(w^m, t^m)$ such that $A \cdot \mathbf{u} = \mathbf{0}$. Moreover, every entry in $\mathbf{u}$ is a polynomial of degree at most $tw$.

*Proof:* As a first step, we reduce this to the problem of solving a linear system of the form $A' \cdot \mathbf{u}' = \mathbf{b}$, where $A'$ and $\mathbf{b}$ have entries in $\mathbb{F}[\mathbf{z}]$ of degree at most $w$, and $A'$ is a square matrix of dimension at most $t'$, which is non-singular. At this point, we can just apply Cramer's rule to find a solution of this system.

Since $t' \leq t$, the rank $r$ of $A(\mathbf{z})$ over $\mathbb{F}(\mathbf{z})$ is at most $t'$. Thus, there is a square submatrix $A'(\mathbf{z})$ of $A$ such that $\det(A')$ is a non-zero polynomial of degree at most $wr \leq wt'$ in $\mathbb{F}[\mathbf{z}]$. For a hitting set $H_{wt',m}$ of polynomials of degree at most $wt'$ on $m$ variables over $\mathbb{F}$, we consider the set of matrices $\{A(\mathbf{c}) : \mathbf{c} \in H_{wt',m}\}$. From the guarantees of the hitting set, we know that there is a $\mathbf{c} \in H_{wt',m}$ such that $A'(\mathbf{c})$ is of rank equal to $r$. Let $\mathbf{c}_0 \in H_{wt',m}$ be such that the rank of $A(\mathbf{c}_0)$ over $\mathbb{F}$ is maximum among all matrices in the set $\{A(\mathbf{c}) : \mathbf{c} \in H_{wt',m}\}$. Moreover, let $A'(\mathbf{z})$ be a submatrix of $A(\mathbf{z})$ such that $\mathrm{rank}(A'(\mathbf{c}_0))$ equals $\mathrm{rank}(A(\mathbf{c}_0))$. From Lemma 1, there is an explicit hitting set $H_{wt',m}$ of size at most $(wt'+1)^m \leq (wt+1)^m$. Thus, we can find $A'(\mathbf{z})$ of rank equal to the rank of $A(\mathbf{z})$ with at most $\mathrm{poly}(w^m, t^m)$ field operations over $\mathbb{F}$. Without loss of generality, let us assume that $A'$ is the top left submatrix of $A$ of size $r$. Clearly, the $(r+1)$-st column of $A$ is linearly dependent on the first $r$ columns of $A$ over the field $\mathbb{F}(\mathbf{z})$. In other words, the linear system given by

$$A' \cdot \mathbf{u}' = \mathbf{b}$$

where $\mathbf{b} = (A_{1,r+1}, A_{2,r+1}, \ldots, A_{r,r+1})$, has a solution in $\mathbb{F}(\mathbf{z})$. Moreover, for every solution $\mathbf{u}'$ of this system, where $\mathbf{u}' = (u'_1, u'_2, \ldots, u'_r)$, the $t$ dimensional vector $(u'_1, u'_2, \ldots, u'_r, -1, 0, \ldots, 0)$ is in the kernel of $A(\mathbf{z})$. Also, since $A \cdot \mathbf{u} = \mathbf{0}$ is a homogeneous linear system, for any non-zero polynomial $P(\mathbf{z})$, $(P \cdot u'_1, P \cdot u'_2, \ldots, P \cdot u'_r, -P, 0, \ldots, 0)$ continues to be a non-zero vector in the kernel of $A(\mathbf{z})$.

Since $A'$ is non-singular, $\mathbf{u}' = (A')^{-1} \cdot \mathbf{b}$ is a solution to this system. Moreover, by Cramer's rule, $(A')^{-1} = \mathsf{adj}(A')/\det(A')$, where $\mathsf{adj}(A')$ is the adjugate matrix of $A'$ and $\det(A')$ is its determinant. Since, every entry of $\mathsf{adj}(A')$ is a polynomial in $\mathbb{F}[\mathbf{z}]$ of degree at most $tw$, we get a solution of the form $\mathbf{u}' = (p_1/\det(A'), p_2/\det(A'), \ldots, p_r/\det(A'))$ where each $p_i$ is a polynomial in $\mathbb{F}[\mathbf{z}]$ of degree at most $tw$. By getting rid of the denominators by scaling by $\det(A')$, we get that the non-zero $t$ dimensional vector $(p_1, p_2, \ldots, p_r, -\det(A'), 0, \ldots, 0)$ is in the kernel of $A(\mathbf{z})$.

Moreover, using the fact that the determinant polynomial has a polynomial size efficiently constructible circuit, and Lemma 2, we can output this vector, with each entry being a list of coefficients in $\mathbb{F}$ in time $\mathrm{poly}(w^m, t^m)$ via an efficient deterministic algorithm. □

## IV. List-Decoding the Multivariate Multiplicity Code

In this section, we prove Theorem 1. We follow the outline of the proof described in Section II. We start with the interpolation step.

### A. Viewing the Encoding as a Formal Power Series

The message space is the space of $m$-variate polynomials of degree at most $d$ over $\mathbb{F}$. In the standard encoding, we have access to evaluations of the polynomial and all its derivatives of order up to $s-1$ on all points on a grid $S^m \subseteq \mathbb{F}^m$.

For our proof, it will be helpful to group the derivatives of the same order together.

*Definition 7: Let $f \in \mathbb{F}[\mathbf{x}]$ be a polynomial. Then, for any $i \in \mathbb{Z}_{\geq 0}$, $\Delta_i(f)$ is defined as*

$$\Delta_i(f) := \sum_{\mathbf{e}:\|\mathbf{e}\|_1 = i} \mathbf{z}^{\mathbf{e}} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}}.$$

□

So, we have a distinct monomial in $\mathbf{z}$ attached to each of the derivatives. The precise form of the monomial in $\mathbf{z}$ is not important, and all that we will use is that these monomials are linearly independent over the underlying field, don't have very high degree and there aren't too many variables in $\mathbf{z}$.

Now, we think of the encoding of $f$ as giving us the evaluation of the tuple of polynomials $\Delta(f) = (\Delta_0(f(\mathbf{x})), \Delta_1(f(\mathbf{x})), \ldots, \Delta_{s-1}(\mathbf{x})) \in \mathbb{F}(\mathbf{z})[\mathbf{x}]^s$ as $\mathbf{x}$ takes values in $\mathbb{F}^m$.

Note that $\Delta_i(f)$ is a homogeneous polynomial of degree equal to $i$ in $\mathbf{z}$.

### B. The $\tau$ Operator

We will need to compute the Hasse derivative of $\Delta_i(f)$ with respect to $\mathbf{x}^{\mathbf{e}}$, i.e., $\frac{\partial \Delta_i(f)}{\partial \mathbf{x}^{\mathbf{e}}}$. From the definition of $\Delta_i(f)$, we have

$$\begin{aligned}\frac{\partial \Delta_i(f)}{\partial \mathbf{x}^{\mathbf{e}}} &= \sum_{\mathbf{e}':\|\mathbf{e}'\|_1 = i} \mathbf{z}^{\mathbf{e}'} \cdot \frac{\partial}{\partial \mathbf{x}^{\mathbf{e}}} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}'}} \\ &= \sum_{\mathbf{e}':\|\mathbf{e}'\|_1 = i} \mathbf{z}^{\mathbf{e}'} \cdot \binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}+\mathbf{e}'}} \\ &= \sum_{\mathbf{e}':\|\mathbf{e}'\|_1 = i} \mathbf{z}^{\mathbf{e}'} \cdot \binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}} \cdot \mathrm{coeff}_{\mathbf{z}^{\mathbf{e}+\mathbf{e}'}}(\Delta_{i+\|\mathbf{e}\|_1} f(\mathbf{x})).\end{aligned}$$

The key point to note is that the Hasse derivative of $\Delta_i(f)$ with respect to $\mathbf{x}^{\mathbf{e}}$ can be read off the coefficients of $\Delta_{i+\|\mathbf{e}\|_1}(f)$.

This motivates the following definition. Consider a tuple $P = (P_0, P_1, \ldots, P_{s-1})$, where for each $i$, $P_i$ is a homogeneous polynomial of degree $i$ in $\mathbb{F}[\mathbf{z}]$. For any $\mathbf{e} \in \mathbb{Z}^m_{\geq 0}$, and $i \leq s-1$ such that $i + \|\mathbf{e}\|_1 \leq s-1$, we define

$$\tau^{(i)}_{\mathbf{e}}(P) := \sum_{\mathbf{e}':\|\mathbf{e}'\|_1 = i} \mathbf{z}^{\mathbf{e}'} \cdot \binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}} \cdot \mathrm{coeff}_{\mathbf{z}^{\mathbf{e}+\mathbf{e}'}}(P_{i+\|\mathbf{e}\|_1}).$$

Thus, for

$$\Delta(f) = (\Delta_0(f(\mathbf{x})), \Delta_1(f(\mathbf{x})), \ldots, \Delta_{s-1}(\mathbf{x})),$$

we have

$$\tau^{(i)}_{\mathbf{e}}(\Delta(f)) = \frac{\partial \Delta_i(f)}{\partial \mathbf{x}^{\mathbf{e}}}.$$

### C. Interpolation Step

Let $P$ be the received word, Thus, we are given a collection of $s$-tuples of polynomials $P(\mathbf{a}) = (P_0(\mathbf{a}), P_1(\mathbf{a}), \ldots, P_{s-1}(\mathbf{a}))$ for every $\mathbf{a} \in S^m$, where each $P_i(\mathbf{a})$ is a homogeneous polynomial of degree $i$ in $\mathbf{z}$. From the earlier definition of $\tau$, given such a $P(\mathbf{a})$, we have $\tau_{\mathbf{e}}^{(i)}(P(\mathbf{a}))$ for every $i \leq w$ and $\mathbf{e}$ with $\|\mathbf{e}\|_1 \leq s-1-w$.

*Lemma 4:* Let $m$ and $s$ be constants. For every natural number $w \leq s-1-m$, and $D = 10|S|(s-w)/w^{1/m}$, there is a non-zero polynomial $Q(\mathbf{x}, \mathbf{y}) = Q_1(\mathbf{x})y_1 + \cdots + Q_w(\mathbf{x})y_w \in \mathbb{F}(\mathbf{z})[\mathbf{x}, \mathbf{y}]$ such that

- For every $i \in \{1, 2, \ldots, w\}$, the $\mathbf{x}$-degree of each $Q_i$ is at most $D$.
- For every $\mathbf{a} \in S^m$ and every $\mathbf{e} \in \mathbb{Z}^m_{\geq 0}$ such that $0 \leq \|\mathbf{e}\|_1 \leq s-1-w$, $\Delta_{\mathbf{e}}(Q)(\mathbf{a}) = 0$, where

$$\Delta_{\mathbf{e}}(Q)(\mathbf{a}) := \sum_{i=1}^{w} \sum_{\mathbf{e}' \leq \mathbf{e}} \frac{\partial Q_i(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}'}}(\mathbf{a}) \cdot \tau_{\mathbf{e}-\mathbf{e}'}^{(i-1)}(P(\mathbf{a})).$$

Here, $\mathbf{e}' \leq \mathbf{e}$ means that $\mathbf{e}$ dominates $\mathbf{e}'$ coordinate wise. Moreover, the coefficients of $Q$ are polynomials in $\mathbb{F}[\mathbf{z}]$ of degree at most $O(|S|^m s^{2m})$, and such a $Q$ can be deterministically constructed by using at most $\mathrm{poly}(|S|^{m^2}, s^{m^2}, d^m)$ operations over the field $\mathbb{F}$.

*Proof:* We start by showing the existence of a polynomial $Q$ with the appropriate degree constraints, followed by an analysis of the running time.

*1) Existence of $Q$:* We view the above constraints as a system of linear equations over the field $\mathbb{F}(\mathbf{z})$, where the variables are the coefficients of $Q$. The number of homogeneous linear constraints is $|S|^m \binom{s-w+m}{m}$ and the number of variables is $w\binom{D+m}{m}$.

By using the fact that $m$ is much smaller than $s$, and a crude approximation of the binomial coefficients, we have $|S|^m \binom{s-w+m}{m} \leq (2e|S|(s-w)/m)^m$ and $w\binom{D+m}{m} > w(D/m)^m$. Plugging in the value of $D$, we get $w(D/m)^m = (10|S|(s-w)/m)^m$, which is clearly greater than the number of constraints. (Notice that setting $D > 2e|S|(s-w)/w^{1/m}$ would have sufficed). Hence, there is a non-zero solution, where the coefficients of the polynomial are from the field $\mathbb{F}(\mathbf{z})$, i.e., are rational functions in $\mathbf{z}$.

Next we analyze the degree of these coefficients and show that we can recover such a $Q$ efficiently, with the appropriate degree bounds.

*2) The Running Time :* For the running time, we recall that each $\tau_{\mathbf{e}}^{(i)}(P)$ is a polynomial of degree at most $w-1$ in the $\mathbf{z}$ variables. As a consequence, observe that the linear system we have for the coefficients of $Q$ is of the form $A \cdot \mathbf{u} = 0$, where $A$ is a matrix with dimension at most $O(|S|^m(s-w)^m)$ over the ring $\mathbb{F}[\mathbf{z}]$, and every entry of $A$ is a polynomial in $\mathbb{F}[\mathbf{z}]$ of degree at most $w$. From Lemma 3, we get that we can find a non-zero solution in $\mathbb{F}[\mathbf{z}]$ using at most $\mathrm{poly}(|S|^{m^2}, s^{m^2})$ field operations over $\mathbb{F}$. Moreover, each of the coordinates of this output vector is a polynomial of degree at most $O(|S|^m(s-w)^m) \cdot w = O(|S|^m s^{2m})$ in $\mathbb{F}[\mathbf{z}]$. □

Going forward, we work with the polynomial $Q$ and the degree parameter $D$ as set in Lemma 4.

### D. Close Enough Codewords Satisfy the Equation

We now show that for every polynomial $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ whose encoding is close enough to the received word $P$, $f$ *satisfies* the equation $Q$ in some sense.

*Lemma 5:* If $f \in \mathbb{F}[\mathbf{x}]$ is a degree $d$ polynomial such that the number of $\mathbf{a} \in S^m$ which satisfy

$$P(\mathbf{a}) = \Delta(f)(\mathbf{a}),$$

is at least $T > (D+d) \cdot |S|^{m-1}/(s-w)$, then $Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f))$ is identically zero as a polynomial in $\mathbb{F}(\mathbf{z})[\mathbf{x}]$.

*Proof:* Define the polynomial $R \in \mathbb{F}(\mathbf{z})[\mathbf{x}]$ as follows

$$\begin{aligned} R(\mathbf{x}) &:= Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \\ &= \sum_{i=1}^{w} Q_i(\mathbf{x}) \cdot \Delta_{i-1}(f). \end{aligned}$$

$R$ is a polynomial in $\mathbf{x}$ of degree at most $D+d$ over the field $\mathbb{F}(\mathbf{z})$. Whenever $\mathbf{a}$ satisfies that $P(\mathbf{a}) = \Delta(f)(\mathbf{a})$, from the definitions of $\tau_{\mathbf{e}}^{(i)}$ and $\Delta_{\mathbf{e}}$, we have that for all $\mathbf{e}$ such that $0 \leq \|\mathbf{e}\|_1 \leq s-w-1$,

$$\begin{aligned} \frac{\partial R(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a}) &= \sum_{i=1}^{w} \sum_{\mathbf{e}' \leq \mathbf{e}} \frac{\partial Q_i(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}'}}(\mathbf{a}) \cdot \frac{\partial \Delta_{i-1}(f)}{\partial \mathbf{x}^{\mathbf{e}-\mathbf{e}'}}(\mathbf{a}) \\ &= \sum_{i=1}^{w} \sum_{\mathbf{e}' \leq \mathbf{e}} \frac{\partial Q_i(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}'}}(\mathbf{a}) \cdot \tau_{\mathbf{e}-\mathbf{e}'}^{(i-1)}(P(\mathbf{a})) \\ &= \Delta_{\mathbf{e}}(Q)(\mathbf{a}) \\ &= 0. \end{aligned}$$

Hence, at every point of agreement between $\Delta(f)$ and the received word $P$, $R(\mathbf{x})$ vanishes with multiplicity at least $s-w$. From Lemma 1, we know that if

$$T(s-w) > (D+d)|S|^{m-1},$$

then $R$ must be identically zero. □

Let us try to get a sense of the parameters here. The relative distance of this code is $\delta = 1 - \frac{d}{s|S|}$. Now, in $\frac{T}{|S|^m} > \frac{D+d}{|S|(s-w)}$, plugging in the value of $D$ from the earlier discussion gives us

$$\begin{aligned} \frac{T}{|S|^m} &> \frac{d}{|S|(s-w)} + \frac{10|S|(s-w)/w^{1/m}}{|S|(s-w)} \\ &= \frac{10}{w^{1/m}} + \left(\frac{s}{s-w}\right) \cdot \frac{d}{s|S|} \\ &= \frac{10}{w^{1/m}} + \left(\frac{w}{s-w}\right) \cdot \frac{d}{s|S|} + \frac{d}{s|S|}. \end{aligned}$$

In our final analysis for the proof of Theorem 1, we choose $w$ and $s$ large enough as a function of $\varepsilon$, so that this bound is of the form $\varepsilon + (1-\delta)$, which is precisely what is claimed in Theorem 1.

### E. Solving the Equation to Find Close Enough Codewords

All that remains now is to solve equations of the form $Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f))$ to recover $f$. This would be done via iteratively constructing $f$ one homogeneous component at a time. We will need the following easy observations.

*Lemma 6:* Let $\mathbb{F}$ be a field of characteristic zero or larger than $d$. Let $f \in \mathbb{F}[\mathbf{x}]$ be a polynomial of degree $d$, and for every $i \in \mathbb{Z}_{\geq 0}$, $\Delta_i$ be the differential form of order $i$ as defined in Definition 7. Then, the following are true.

- For each $i \in \mathbb{Z}_{\geq 0}$, $\Delta_i(f)$ is homogeneous in $\mathbf{z}$ and has degree $i$ in the $\mathbf{z}$ variables. Moreover, for any monomial $\mathbf{z}^{\mathbf{e}}$ of degree $i$, its coefficient in $\Delta_i(f)$ equals $\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}$.
- If $f$ is a homogeneous polynomial, then, for every $i \leq d$, $f$ can be uniquely recovered from all its partial derivatives of order $i$. As a consequence, for any homogeneous $f$, given the formal polynomial $\Delta_i(f)$, we can recover $f$.

*Proof:* The first item follows directly from the definition of $\Delta$ in Definition 7.

The second item follows from an immediate generalization of the following well known observation of Euler to Hasse derivatives. For any homogeneous polynomial $f$ of degree $d$,

$$d \cdot f = \sum_i x_i \cdot \frac{\partial f(\mathbf{x})}{\partial x_i}.$$

We also have that

$$\frac{\partial}{\partial \mathbf{x}^{\mathbf{e}}} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}'}} = \binom{\mathbf{e}+\mathbf{e}'}{\mathbf{e}} \cdot \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}+\mathbf{e}'}}.$$

Using this we can compute the first order Hasse derivatives of $\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}'}}$ for all $\|\mathbf{e}'\|_1 = i-1$ from $\Delta_i(f)$. So, for any $i$, given all Hasse derivatives of degree $i$, we can recover Hasse derivatives of degree $i-1$ (using Euler's formula), and so on, till we recover $f$. □

*Remark 4:* We remark that the second item in Lemma 6 is false for fields of small characteristic. For instance, if the characteristic is smaller than $d$, then even for a non-zero $f$, all its first order derivatives could be zero, and hence $f$ cannot be recovered from its first order derivatives.

The following lemma shows that under very mild conditions on $Q(\mathbf{x}, \mathbf{y})$, we can (efficiently) recover all polynomials $f$ of degree at most $d$ such that $Q(\mathbf{x}, \Delta_{<w}(f)) \equiv 0$. This will complete all the ingredients needed for the proof of Theorem 1.

*Lemma 7:* Let $\mathbb{F}$ be a finite field of characteristic larger than $d$ and let $Q(\mathbf{x}, \mathbf{y}) = Q_1(\mathbf{x})y_1 + \cdots + Q_w(\mathbf{x})y_w$ be any non-zero polynomial in $\mathbb{F}(\mathbf{z})[\mathbf{x}, \mathbf{y}]$ where, $\deg_{\mathbf{x}}(Q) \leq D + d$, $\deg_{\mathbf{z}}(Q) \leq \Gamma$ and $Q_i$ does not depend on $\mathbf{y}$. Then, there is a deterministic algorithm that outputs all polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ such that

$$Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

Moreover, the algorithm requires at most $\text{poly}\left(D^m, d^m, |\mathbb{F}|^{\binom{w+m}{w}}, \Gamma^m\right)$ arithmetic operations over the underlying field $\mathbb{F}$.

*Proof:* We will reconstruct $f$ iteratively, one homogeneous component at a time. This iterative process has to be started by fixing the homogeneous components of $f$ of degree at most $w$, and as will be evident from the discussion ahead, every fixing of this initial seed can be lifted to a unique $f$ of degree at most $d$ satisfying

$$Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

Before starting the reconstruction, we need to ensure appropriate non-degeneracy conditions which are typical in iterative reconstruction arguments of this kind.

*1) Preprocessing:* We know from the hypothesis of the lemma that $Q$ depends on at least one $y$ variable. Let $j$ be the largest index in $\{1, \ldots, w\}$ such that $Q$ depends on $y_j$, i.e., $Q_j$ is non-zero and $Q_i$ is identically zero for all $i > j$. For the ease of notation, we shall assume that $j = w$, thus, $Q_w$ is a non-zero polynomial. Recall that $f$ is a polynomial in $\mathbb{F}[\mathbf{x}]$ and each $\Delta_i(f)$ is a polynomial in $\mathbb{F}[\mathbf{x}, \mathbf{z}]$.

Since $Q_w(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is a non-zero polynomial, there is an $\mathbf{a} \in \mathbb{F}^m$ such that $Q_w(\mathbf{a}) \neq 0$.[6] Replacing the variable $x_i$ by $x'_i + a_i$ (i.e., translating the origin), we can ensure that in this translated coordinate system, $Q_w(\mathbf{x}' + \mathbf{a})$ is non-zero at the origin, i.e., when $\mathbf{x}'$ is set to $\mathbf{0}$. We work in this translated coordinate system for the ease of notation. Observe that every solution $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is bijectively mapped to a solution $\tilde{f}(\mathbf{x}') = f(\mathbf{x}' + \mathbf{a}) \in \mathbb{F}[\mathbf{x}']$ and given $\tilde{f}$, we can efficiently recover $f$. Also, note that $\Delta_i(f)(\mathbf{x}' + \mathbf{a}) = \Delta_i(f(\mathbf{x}' + \mathbf{a})) = \Delta_i(\tilde{f}(\mathbf{x}))$, i.e., taking derivatives and then setting $\mathbf{x} = \mathbf{x}' + \mathbf{a}$ is equivalent to first doing the translation $\mathbf{x} = \mathbf{x}' + \mathbf{a}$ and then taking derivatives. Let

$$Q'(\mathbf{x}') := Q(\mathbf{x}' + \mathbf{a}) = Q_1(\mathbf{x}' + \mathbf{a})y_1 + \cdots + Q_w(\mathbf{x}' + \mathbf{a})y_w,$$

and let $I = \langle x'_1, \ldots, x'_m \rangle$ be the ideal generated by $\{x'_1, \ldots, x'_m\}$.

*2) Iterative Reconstruction:* We are now ready to describe the iterative reconstruction of $\tilde{f}$.

- **Base Case:** We will try all possible values for the coefficients of monomials of degree at most $w$ in $\tilde{f}$ from the field $\mathbb{F}$. There are $|\mathbb{F}|^{\binom{w+m}{m}}$ possible choices. The next steps are going to uniquely lift each of these candidate solutions to a degree $d$ polynomial, so the number of solutions remains $|\mathbb{F}|^{\binom{w+m}{m}}$.
- **Induction Step:** We now assume that we have recovered $\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_t \in \mathbb{F}[\mathbf{x}']$ for some $t \geq w$, where $\tilde{f}_i$ is a homogeneous component of $\tilde{f}$ of degree $i$. The goal is to recover $\tilde{f}_{t+1}$, the $(t+1)$-st homogeneous component. Let $\tilde{f}_{\leq t} = \tilde{f}_0 + \tilde{f}_1 + \cdots + \tilde{f}_t$. Now, let us consider the equation $Q'(\mathbf{x}', \Delta_0(\tilde{f}), \Delta_1(\tilde{f}), \ldots, \Delta_{w-1}(\tilde{f})) = 0$ when we work modulo the ideal $I^{t-w+3}$. Clearly, the homogeneous components of $\tilde{f}$ of degree larger than $t+1$ do not contribute anything modulo $I^{t-w+3}$, and so we have,

$$Q'(\mathbf{x}', \Delta_0(\tilde{f}_{\leq t}), \Delta_1(\tilde{f}_{\leq t}), \ldots, \Delta_{w-1}(\tilde{f}_{\leq t} + \tilde{f}_{t+1})) = 0 \mod I^{t-w+3}.$$

Using the linearity of $\Delta_i$ and the fact that $Q'$ is linear in $\mathbf{y}$, we get

$$Q'(\mathbf{x}', \Delta_0(\tilde{f}_{\leq t}), \Delta_1(\tilde{f}_{\leq t}), \ldots, \Delta_{w-1}(\tilde{f}_{\leq t})) + Q_w(\mathbf{x}' + \mathbf{a}) \cdot \Delta_{w-1}(\tilde{f}_{t+1}) = 0 \mod I^{t-w+3}.$$

We know that the degree of $\Delta_{w-1}(\tilde{f}_{t+1})$ equals $t + 1 - (w-1) = t - w + 2$, and it is homogeneous in $\mathbf{x}'$. Also, we have ensured in the preprocessing phase that $Q_w(\mathbf{x}' + \mathbf{a}) \mod I = Q_w(\mathbf{a})$ is some non-zero

[6] This is assuming $\mathbb{F}$ is large enough, else we can find such an $\mathbf{a}$ in a large enough extension field of $\mathbb{F}$.

constant $\mathbb{F}$. Thus, this is a non-trivial linear equation in $\Delta_{w-1}(\tilde{f}_{t+1})$ and if we can use it to recover all the partial derivatives of $\tilde{f}_{t+1}$ of order $w-1$, we can then use Lemma 6 to recover $\tilde{f}_{t+1}$ itself. We elaborate on the details of this step of recovering the partial derivatives of $\tilde{f}_{t+1}$ from $\Delta_{w-1}(f_{t+1})$ next.

*3) Recovering Partial Derivatives of $\tilde{f}_{t+1}$ From $\Delta_{w-1}(\tilde{f}_{t+1})$:* Recall that since $\tilde{f}_{t+1}$ is a homogeneous polynomial in $\mathbb{F}[\mathbf{x}]$ of degree $t+1$, each of its partial derivatives of order $w-1$ is a homogeneous polynomial in $\mathbf{x}'$ of degree equal to $t+1-(w-1) = t-w+2$. Thus,

$$\Delta_{w-1}(\tilde{f}_{t+1}) := \sum_{\mathbf{e}:\|\mathbf{e}\|_1=w-1} \mathbf{z}^{\mathbf{e}} \cdot \frac{\partial \tilde{f}_{t+1}(\mathbf{x}')}{\partial \mathbf{x}'^{\mathbf{e}}}.$$

is a homogeneous polynomial in both $\mathbf{z}$ and $\mathbf{x}'$, with degree $w-1$ in $\mathbf{z}$ and degree $t-w+2$ in $\mathbf{x}'$. Our goal is to recover the coefficients of all monomials $\mathbf{z}^{\mathbf{e}}$ of degree $w-1$ in $\mathbf{z}$ when viewing $\Delta_{w-1}(\tilde{f}_{t+1})$ as a polynomial in $\mathbb{F}[\mathbf{x}][\mathbf{z}]$, and we have access to the expression

$$Q'(\mathbf{x}', \Delta_0(\tilde{f}_{\le t}), \Delta_1(\tilde{f}_{\le t}), \ldots, \Delta_{w-1}(\tilde{f}_{\le t}))$$

$$= -Q_w(\mathbf{a})\Delta_{w-1}(\tilde{f}_{t+1}) \mod I^{t-w+3}.$$

As a first step, observe that the polynomial $Q_w(\mathbf{a})\Delta_{w-1}(\tilde{f}_{t+1})$ has degree at most $\Gamma + w - 1$ in $\mathbf{z}$ and degree exactly $t-w+2$ in $\mathbf{x}'$. Moreover, since $Q_w(\mathbf{a}) \in \mathbb{F}[\mathbf{z}]$ is non-zero, the polynomials $\{Q_w(\mathbf{a})\mathbf{z}^{\mathbf{e}} : \deg(\mathbf{z}^{\mathbf{e}}) = w-1\}$ are linearly independent as polynomials of degree at most $\Gamma + (w-1)$ in $\mathbf{z}$ over the field $\mathbb{F}$. Therefore, for any hitting set $H \subseteq \mathbb{F}^m$ for $m$-variate polynomials of degree at most $\Gamma + (w-1)$, the evaluation vectors $\mathsf{Eval}_H(Q_w(\mathbf{a})\mathbf{z}^{\mathbf{e}})$ of these polynomials on $H$ are linearly independent over $\mathbb{F}$.[7] So, for every $\mathbf{x}'^{\mathbf{e}_0}$ of degree $w-1$, there exists an $\mathbb{F}$ linear combination of the polynomials $\{Q_w(\mathbf{a})\Delta_{w-1}(\tilde{f}_{t+1})\mathbf{b} : \mathbf{b} \in H\}$ which equals $\frac{\partial \tilde{f}_{t+1}(\mathbf{x}')}{\partial \mathbf{x}'^{\mathbf{e}}}$. Moreover, such a linear combination can be found (e.g. via Gaussian Elimination over the field $\mathbb{F}$) efficiently in the size of this linear system.

Thus, to recover the partial derivatives of order $w-1$ of $\tilde{f}_{t+1}$ given a monomial representation of $Q_w(\mathbf{a})\Delta_{w-1}(\tilde{f}_{t+1})$, we consider the hitting set $H$ of size $O(\Gamma + w)^m \le O(\Gamma \cdot w)^m$ for $m$-variate degree $\Gamma + (w-1)$ polynomials given by Lemma 1, compute the evaluation of the polynomials

$$Q_w(\mathbf{a}) \cdot \Delta_{w-1}(\tilde{f}_{t+1}) = \sum_{\mathbf{e}:\|\mathbf{e}\|_1=w-1} Q_w(\mathbf{a})\mathbf{z}^{\mathbf{e}} \cdot \frac{\partial \tilde{f}_{t+1}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}},$$

at every $\mathbf{b} \in H$, and take appropriate weighted linear combinations to recover each of the partial derivatives $\frac{\partial \tilde{f}_{t+1}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}}}$.

Since $Q'(\mathbf{x}', \Delta_0(\tilde{f}_{\le t}), \Delta_1(\tilde{f}_{\le t}), \ldots, \Delta_{w-1}(\tilde{f}_{\le t}))$ is a polynomial of degree at most $\Gamma + w$ in $\mathbf{z}$ and at most $D+d$ in $\mathbf{x}$, we can do the evaluations by writing the coefficient vector of this polynomial in time $\mathrm{poly}(D^m, d^m, \Gamma^m, w^m)$ and doing evaluations one monomial at a time.

[7]Here we are setting the $\mathbf{z}$ variables according to inputs in the hitting set $H$ and looking at the resulting vectors.

*4) The Running Time:* Observe that we can go from the original polynomial $Q$ to the polynomial $Q'$ by finding an appropriate $\mathbf{a}$ deterministically in time at most $(D+d)^m$ by just querying all points on a large enough grid in $\mathbb{F}^m$ (or a grid in an extension field of $\mathbb{F}$, if $\mathbb{F}$ isn't large enough). This follows from Lemma 1.

Once we have $Q'$, we reconstruct $f$ in $d$ iterations, so it suffices to estimate the cost of each iteration. As we just argued in the earlier part of the proof, every iteration just involves evaluating the polynomial $Q'(\mathbf{x}', \Delta_0(\tilde{f}_{\le t}), \Delta_1(\tilde{f}_{\le t}), \ldots, \Delta_{w-1}(\tilde{f}_{\le t}))$ at a hitting set $H$ of size at most $\mathrm{poly}(\Gamma^m, w^m)$ and solving about $w^m$ linear systems of the same size. The straightforward implementation of this takes no more than $\mathrm{poly}(D^m, d^m, \Gamma^m, w^m)$ field operations. □

As is evident from the proof of Lemma 7, the following more structured version of Lemma 7 is true.

*Lemma 8:* Let $\mathbb{F}$ be a field of characteristic zero or larger than $d$ and let $Q(\mathbf{x}, \mathbf{y}) = Q_1(\mathbf{x})y_1 + \cdots + Q_w(\mathbf{x})y_w$ be any non-zero polynomial in $\mathbb{F}(\mathbf{z})[\mathbf{x}, \mathbf{y}]$ where, $\deg_{\mathbf{x}}(Q) \le D+d$, $\deg_{\mathbf{z}}(Q) \le \Gamma$ and $Q_i$'s do not depend on $\mathbf{y}$. Then, there is a deterministic algorithm that outputs a linear space of polynomials in $\mathbb{F}[\mathbf{x}]$ of dimension at most $\binom{w+m}{m}$ over $\mathbb{F}$ which contains all polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ such that

$$Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

Moreover, the algorithm requires at most $\mathrm{poly}\,(D^m, d^m, \Gamma^m)$ arithmetic operations over the underlying field $\mathbb{F}$.

To bound the true running time of the algorithm in Lemma 8, we need to add a $\mathrm{poly}(\log \mathbb{F})$ factor in the upper bound on the field operations for finite fields and a polynomial factor in the bit complexity of the input over the field of rational numbers. While working over rationals, we might need a bit more care to solve the linear systems appearing in the proof of Lemma 7 efficiently, since the naive implementation of Gaussian Elimination might blow up the bit complexity of the numbers appearing at various intermediate stages.

### F. Putting Things Together

We are now ready to prove Theorem 1.

*Proof of Theorem 1:* Given the error parameter $\varepsilon$ and the number of variables $m$, we choose $s, w$ as follows.

- $w = \left(\frac{20}{\varepsilon}\right)^m$,
- $s = \frac{4}{\varepsilon} \cdot \binom{w+m}{m}$.

We note that for this choice of parameters, $\frac{w}{s-w} < \frac{\varepsilon}{2}$ and hence,

$$\frac{10}{w^{1/m}} + \frac{w}{s-w} < \varepsilon.$$

With this choice of parameters, we use Lemma 4 to construct a non-zero polynomial $Q$ which *explains* the received word $P$. Then, we use Lemma 7 to find all polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ such that

$$Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

We know, from Lemma 7 that the number of such solutions is upper bounded by $|\mathbb{F}|^{\binom{w+m}{m}}$ and from Lemma 5 that

every polynomial $f$ of degree at most $d$ in $\mathbb{F}[\mathbf{x}]$ such that $\mathrm{Dist}(\mathsf{Enc}(f), P)$ is at most $(1-\delta)-\varepsilon$, where $\delta = 1-d/(s|S|)$ satisfies the equation

$$Q(\mathbf{x}, \Delta_0(f), \Delta_1(f), \ldots, \Delta_{w-1}(f)) \equiv 0.$$

Thus, all such polynomials $f$ are included in the list of outputs. The running time of the algorithm follows from the running time guarantees in Lemma 4 and Lemma 7. □

### G. Another View of the Algorithm

We now discuss an alternative description of the decoding algorithm. In essence, this is just a rewording of the previous algorithm, but appears to have some qualitative advantages. For instance, the description itself seems simpler here as we don't need to introduce the $\mathbf{z}$ variables, but instead, end up working with a system of equations over the original field $\mathbb{F}$ itself. Moreover, the runtime analysis of the algorithm gives a slightly better bound of $\mathrm{poly}(|S|^m, d^m)$ on the number of field operations needed by the decoding algorithm as opposed to the bound of $\mathrm{poly}(|S|^{m^2}, d^{m^2})$ that is claimed in Theorem 1.

Given the received word $P : S^m \to \mathbb{F}^{\binom{s+m-1}{m}}$, we assume that the coordinates of $\mathbb{F}^{\binom{s+m-1}{m}}$ are indexed by $m$-variate monomials of degree at most $s-1$. Let $t = 10w^{m+1}$ and for each $i \in [s]$ and $j \in [t]$, let $\mathbf{a}_{i,j} \in \mathbb{F}^{\binom{i-2+m}{m}}$ be vectors such that for every $i$, the dimension of the space spanned by $\{\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \ldots, \mathbf{a}_{i,t}\}$ over $\mathbb{F}$ equals $\binom{i-2+m}{m}$. Again we think of the coordinates of $\mathbf{a}_{i,j}$ as being indexed by $m$-variate monomials of degree equal to $i-1$.

Now, from $P$, we construct $P_1, P_2, \ldots, P_t$ where each $P_j$ is a function $S^m$ to $\mathbb{F}^s$, such that for every $\mathbf{b} \in S^m$, the $i^{th}$ coordinate of $P_j(\mathbf{b})$ equals the weighted linear combination of the coordinates of $P(\mathbf{b})$ indexed by monomials of degree exactly $i-1$, with weights according to $\mathbf{a}_{i,j}$. In other words, the $i^{th}$ coordinate of $P_j(\mathbf{b})$ equals

$$\sum_{\mathbf{e} \in \mathbb{Z}^m_{\geq 0}, \|\mathbf{e}\|_1 = i-1} \mathbf{a}_{i,j}(\mathbf{e}) \cdot P(\mathbf{b})_{\mathbf{e}},$$

where $P(\mathbf{b})_{\mathbf{e}}$ is the coordinate of $P(\mathbf{b})$ indexed by $\mathbf{e}$. Now, for the interpolation step, for each $j \in [t]$, we find a polynomial $\tilde{Q}_j = \sum_{i=1}^{w} \tilde{Q}_{i,j}(\mathbf{x}) y_j$ of not too high degree which explains $P_j$ in the sense of Lemma 4. Note that each $\tilde{Q}_j$ is now a polynomial over the original field $\mathbb{F}$. An immediate instantiation of Lemma 5 for this setting shows that if $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ and $\mathsf{Enc}(f)$ and $P$ are close enough, then for every $j \in [t]$, $\tilde{Q}_j(\mathbf{x}, \Psi_j(f))$ must be identically zero, where $\Psi_j(f) = (\Psi_{j,1}(f), \ldots, \Psi_{j,w}(f))$ is defined as

$$\Psi_{j,i}(f) = \sum_{\mathbf{e} \in \mathbb{Z}^m_{\geq}, \|\mathbf{e}\|_1 = i-1} \mathbf{a}_{i,j}(\mathbf{e}) \cdot \frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}}.$$

Before going to the reconstruction step, we note that it might be the case that $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_t$ depend on different subsets of $\mathbf{y}$ variables. But since $t > w^{m+1}$, by averaging, it follows that there exist an $\ell \in [w]$ such that at least $w^m$ of the polynomials $\{\tilde{Q}_j : j \in [t]\}$ have the property that they depend on $y_\ell$ and do not depend on $y_{\ell'}$ for any $\ell' > \ell$. For the ease of notation, let us assume that $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_{t'}$ depend on $y_w$, where $t' = w^m$.

Now, to recover $f$, we solve the equations $\tilde{Q}_j(\mathbf{x}, \Psi_j(f)) \equiv 0$ for all $j \in [t']$. We solve for $f$ one homogeneous component as in the proof of Lemma 7. Assuming that we have recovered homogeneous components of degree at most $u$ of $f$, we can follow the proof of Lemma 7 to recover $\Psi_{j,w}(f_{u+1})$ for every $j \in [t']$, where $f_{u+1}$ is the homogeneous component of $f$ of degree $u+1$.[8] At this point, the choice of the vectors $\mathbf{a}_{i,j}$, the definition of $\Psi_{j,w}(f_{u+1})$ and the fact that $t' \geq w^m$, we get that we have sufficiently many linearly independent homogeneous linear equations in all the partial derivatives of $f_{u+1}$ of order $(w-1)$. Thus, we can solve this linear system to recover each of these partial derivatives and combine them according to Lemma 6 to obtain $f_{u+1}$, and proceed to the next step. Moreover, as in Lemma 7, if we start from the correct coefficients of $f$ in the base case of this process, each of the subsequent steps are unique.

Thus, instead of working with a single polynomial equation as in a standard application of the polynomial method, this algorithm proceeds via working simultaneously with many equations.

We now remark on the running time.

*Remark 5:* We note that in algorithm sketched above, the number of field operations needed is upper bounded by $\mathrm{poly}(|S|^m, d^m)$. This follows from the observation that in this algorithm we are essentially solving $w^m < d^m$ linear systems of size $\mathrm{poly}(|S|^m, d^m)$ over the underlying field $\mathbb{F}$ to recover all codewords close to the received word.

## V. Reducing the List Size to a Constant

In this section, we use the pruning algorithm due to Kopparty et al. [17] together with Lemma 8 to obtain a shorter list of correct polynomials, thereby improving the bound on the list size in Theorem 1 from a polynomial (in the input size) to an absolute constant depending only on the parameter $\varepsilon$ and dimension $m$. This would complete the proof of Theorem 2. The first step towards the goal of recovering codewords from a small linear space is the following theorem, which is a natural multivariate analog of [25, Theorem 17] in the work of Guruswami and Kopparty [25]. Our proof is essentially the same, apart from the fact that we are in the multivariate setting and hence have to work with Generalized Wronskians matrices.

*Theorem 4 (subspace restrictions):* Let $\mathbb{F}$ be a field of characteristic zero or larger than $d$. Let $\mu \geq w \in \mathbb{N}$ be parameters and let $W \subseteq \mathbb{F}[\mathbf{x}]$ be an $\mathbb{F}$-linear subspace of $m$-variate polynomials of degree at most $d$, such that dimension of $W$ is at most $w$. For any $\mathbf{a} \in \mathbb{F}^m$, let $H_{\mathbf{a}}$ be the $\mathbb{F}$-linear space of polynomials of degree at most $d$ which vanish with multiplicity at least $\mu$ at $\mathbf{a}$. Then, for every set $S \subseteq \mathbb{F}$, we have,

$$\sum_{\mathbf{a} \in S^m} \dim(H_{\mathbf{a}} \cap W) \leq \frac{dw|S|^{m-1}}{(\mu - w + 1)}.$$

We use this statement in our proof in this section, and prove it in Section VI.

[8] We might have to do an initial translation of coordinates as in the proof of Lemma 7.

### A. The Pruning Algorithm

The input to this algorithm is a received word $P$ and a linear subspace $W$ of polynomials of degree at most $d$ in $\mathbb{F}[\mathbf{x}]$ of dimension at most $w$. The goal is to output all polynomials in $f \in W$ such that $\mathsf{Enc}_{s,S}(f)$ agrees with $P$ on at least $\alpha = \delta+\varepsilon$ locations. The description of the algorithm has a parameter $r$, which we later set to an appropriate value.

*1) Algorithm A:*

1) Choose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r$ independently and uniformly at random from $S^m$.
2) If there is a unique polynomial $f \in W$ such that $\mathsf{Enc}_{s,S}(f)$ and $P$ agree on each of $\mathbf{a}_1, \ldots, \mathbf{a}_r$, then output $f$.

Clearly, the second step of the algorithm can be implemented efficiently via Gaussian Elimination.

The final pruning algorithm invokes Algorithm A multiple times and outputs the union of all the lists. In the rest of this section, we show that with high probability, this will output the list of all codewords close to the received word that are contained in the input linear space.

The algorithm and the analysis is precisely the same as that in the work of Kopparty et al. [17], apart from the fact that we invoke it for multivariate multiplicity codes, whereas in [17] it was designed for folded Reed Solomon Codes and univariate multiplicity codes. We briefly sketch some of the details in the rest of this section. For brevity, we again use $\mathsf{Enc}()$ for $\mathsf{Enc}_{s,S}()$. We also assume that the dimension $w$ of $W$ is less than the multiplicity parameter $s$ of the code.

*Lemma 9 (Analogous to [17, Lemma IV.5 (conference version)]):* For any polynomial $f \in W$ such that $\mathrm{Dist}(\mathsf{Enc}(f), P) < \alpha$, $f$ is output by Algorithm A with probability at least

$$(1-\alpha)^r - w\left(\frac{d}{|S|(s-w)}\right)^r.$$

Moreover, Algorithm A runs in polynomial time in the input size.

*Proof Sketch:* The proof of the lemma is precisely the same as that of [17, Lemma IV.5 (conference version)] except we use Theorem 4 as opposed to an analogous statement for folded Reed Solomon codes. □

We are now ready to prove Theorem 2.

*Proof of Theorem 2:* Following the proof of Theorem 1, given the error parameter $\varepsilon$ and the number of variables $m$, we choose $s, w$ as follows.

- $w = \left(\frac{20}{\varepsilon}\right)^m$,
- $s = \frac{4}{\varepsilon} \cdot \binom{w+m}{m}$.

We note that for this choice of parameters, $\frac{w}{s-w} < \frac{\varepsilon}{2}$ and hence,

$$\frac{10}{w^{1/m}} + \frac{w}{s-w} < \varepsilon,$$

as is needed to invoke Lemma 4. We now use Lemma 4 to construct the polynomial $Q$ which *explains* the received word $P$, and Lemma 8 to give us a subspace $W$ of polynomials in $\mathbb{F}[\mathbf{x}]$ of dimension at most $w = \binom{w+m}{m}$ over $\mathbb{F}$, that contains all polynomials $f \in \mathbb{F}[\mathbf{x}]$ of degree at most $d$ such that $\mathrm{Dist}(\mathsf{Enc}(f), P) < (\delta-\varepsilon)$, where $\delta = 1 - d/(s|S|)$ is the relative distance of the code. Let the parameter $r$ be set as

$$r = \frac{\log(2 \cdot \binom{w+m}{m})}{\log(1+\varepsilon/4)} \leq O\left(\frac{m^2 \log 1/\varepsilon}{\varepsilon}\right).$$

We now instantiate Lemma 9 with inputs being the received word $P$, the subspace $W$ of dimension at most $w' = \binom{w+m}{m}$ and the parameter $r$ as set above.

A single run of Algorithm A returns at most one polynomial $f$ in $W$ such that $\mathrm{Dist}(\mathsf{Enc}(f), P) < (\delta-\varepsilon)$. Moreover, every such $f$ is output with probability at least

$$\rho = (1-\delta+\varepsilon)^r - w'\left(\frac{d}{|S|(s-w')}\right)^r.$$

To simplify this, we note that from the choice of parameters

$$w'\left(\frac{d}{|S|(s-w')}\right)^r = \binom{w+m}{m}\left(\frac{s}{(s-w')} \cdot (1-\delta)\right)^r,$$

and plugging in the values of $s$ and $r$, we have

$$\begin{aligned}
&\leq \frac{1}{2} \cdot (1+\varepsilon/4)^r \left(\frac{1}{(1-\varepsilon/4)} \cdot (1-\delta)\right)^r \\
&\leq \frac{1}{2} \cdot \left(\frac{1+\varepsilon/4}{1-\varepsilon/4} \cdot (1-\delta)\right)^r \\
&\leq \frac{1}{2} \cdot (1-\delta+\varepsilon)^r,
\end{aligned}$$

where the last inequality follows from the fact that $\frac{1+\varepsilon/4}{1-\varepsilon/4} \cdot (1-\delta) \leq (1-\delta+\varepsilon)$, whenever $1+\delta-\varepsilon/2 > 0$, which is always true in our setting, since $\delta, \varepsilon \in (0,1)$. Thus, we get

$$\rho \geq \frac{1}{2}(1-\delta+\varepsilon)^r.$$

Hence, the number of polynomials in the space $W$ such that $\mathrm{Dist}(\mathsf{Enc}(f), P) < (\delta-\varepsilon)$ is at most $\frac{1}{\rho} = \frac{2}{(1-\delta+\varepsilon)^r}$.

It follows from a union bound that if we run Algorithm A about $O\left(\frac{1}{\rho} \cdot \log \frac{1}{\rho}\right)$ times with fresh randomness each time, and output every polynomial obtained, with high probability, we would have output all the polynomials $f$ in $W$ with $\mathrm{Dist}(\mathsf{Enc}(f), P) < (\delta-\varepsilon)$. Thus the number of runs of Algorithm A is

$$\begin{aligned}
O\left(\frac{1}{\rho} \cdot \log \frac{1}{\rho}\right) &= O\left(\frac{r \log(\frac{1}{1-\delta+\varepsilon})}{(1-\delta+\varepsilon)^r}\right) \\
&\leq \exp\left(O\left(\frac{m^2}{\varepsilon} \log^3 \frac{1}{\varepsilon}\right)\right).
\end{aligned}$$

The upper bound on the running time immediately follows from the running time guarantees in Lemma 4, Lemma 8 and the final pruning that happens in the process of recovering the relevant codewords from the subspace output by Lemma 8. □

## VI. Subspace Restrictions of Multivariate Multiplicity Codes

In this section, we prove Theorem 4. For the proof, we follow the outline of Guruswami and Kopparty [25] and essentially observe that (almost) everything works even for multivariate polynomials. The only difference is that instead of the Wronskian criterion for univariate polynomial, we need

to work with the following generalized Wronskian criterion for multivariate polynomials.

*Theorem 5 (Generalized Wronskian Criterion):* Let $f_1, f_2, \ldots, f_w \in \mathbb{F}[\mathbf{x}]$ be $m$-variate polynomials of maximum individual degree at most $d$. If the characteristic of $\mathbb{F}$ is zero or larger than $d$, then the following is true. $f_1, f_2, \ldots, f_w$ are linearly independent over $\mathbb{F}$ if and only if there exist monomials $\mathbf{x}^{\mathbf{e}_1}, \mathbf{x}^{\mathbf{e}_2}, \ldots, \mathbf{x}^{\mathbf{e}_w}$ such that for every $i \in [w]$, $\deg(\mathbf{x}^{\mathbf{e}_i}) \leq i-1$, and the $w \times w$ matrix $M_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}$ whose $(i,j)$ entry equals $\frac{\partial f_j}{\partial \mathbf{x}^{\mathbf{e}_i}}$ is full rank over the field $\mathbb{F}(\mathbf{x})$.

The classical Wronskian criterion (and its generalized counterpart) are typically proved for fields of characteristic zero and with the usual notion of partial derivatives (cf., Bostan and Dumas [26, Theorem 3]). These proofs extend to the above setting. For the sake of completeness, we provide an alternative proof of the above theorem in subsection II.

Equipped with this criterion, we are now ready to prove Theorem 4

*Proof of Theorem 4:* Let $f_1, f_2, \ldots, f_w \in W$ be linearly independent polynomials of degree at most $d$ which span $W$. Let $E$ be a subset of $\mu$-tuples of monomials defined as follows.

$$E := \{(\mathbf{x}^{\mathbf{e}_1}, \mathbf{x}^{\mathbf{e}_2}, \ldots, \mathbf{x}^{\mathbf{e}_\mu}) : \deg(\mathbf{x}^{\mathbf{e}_i}) \leq i-1\}.$$

For every $\psi = (\mathbf{x}^{\mathbf{e}_1}, \mathbf{x}^{\mathbf{e}_2}, \ldots, \mathbf{x}^{\mathbf{e}_\mu})$ in $E$, let $M_\psi \in \mathbb{F}[\mathbf{x}]^{\mu \times w}$ matrix defined as follows.

$$M_\psi := \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_1}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_1}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_1}} \\ \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_2}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_2}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_2}} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_\mu}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_\mu}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_\mu}} \end{bmatrix}.$$

And, let $\tilde{M}_\psi$ denote the $w \times w$ submatrix of $M_\psi$ by taking the first $w$ rows and columns, i.e.,

$$\tilde{M}_\psi := \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_1}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_1}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_1}} \\ \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_2}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_2}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_2}} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_1}{\partial \mathbf{x}^{\mathbf{e}_w}} & \frac{\partial f_2}{\partial \mathbf{x}^{\mathbf{e}_w}} & \cdots & \frac{\partial f_w}{\partial \mathbf{x}^{\mathbf{e}_w}} \end{bmatrix}.$$

From Theorem 5, we know that there exists $\psi_0$ in $E$ such that $\tilde{M}_{\psi_0}$ (and hence, $M_{\psi_0}$) is full rank over $\mathbb{F}(\mathbf{x})$. Let $L_{\psi_0}$ denote the determinant of $\tilde{M}_{\psi_0}$. Clearly, $L_{\psi_0}$ is a non-zero $m$-variate polynomial of degree at most $dw$. We note that for many choices of $\psi \in E$, the corresponding matrix $M_\psi$ could be of rank less than $w$. Perhaps somewhat surprisingly, all these matrices play a role in the proof. The proof essentially follows from the following claim.

*Claim 1: For every $\mathbf{a} \in \mathbb{F}^m$, the multiplicity of $L_{\psi_0}(\mathbf{x})$ at $\mathbf{a}$ is at least $(\mu - w + 1)\dim(H_\mathbf{a} \cap W)$.*

We first complete the proof of the theorem using the above claim and then prove the claim.

From Claim 1, we get

$$\sum_{\mathbf{a} \in S^m} (\mu - w + 1)\dim(H_\mathbf{a} \cap W) \leq \sum_{\mathbf{a} \in S^m} \text{mult}(L(\mathbf{x}), \mathbf{a}).$$

From the earlier discussion, $L_{\psi_0}$ is a non-zero polynomial of degree at most $dw$. Thus, by Lemma 1, the quantity $\sum_{\mathbf{a} \in S^m} \text{mult}(L(\mathbf{x}), \mathbf{a})$ is upper bounded by $dw|S|^{m-1}$, and this completes the proof of Theorem 4. □

We now prove the claim. For this, we need the following claim.

*Claim 2: For every $\psi \in E$, and for every $\mathbf{a} \in \mathbb{F}^m$,*

$$\text{rank}(M_\psi(\mathbf{a})) \leq w - \dim(H_\mathbf{a} \cap W).$$

*Proof of Claim 2:* We just follow the definition.

$$\begin{aligned}
&\dim(H_\mathbf{a} \cap W) \\
&= \dim(\{\mathbf{b} = (b_1, b_2, \ldots, b_w) \in \mathbb{F}^w : \\
&\text{mult}(\sum_{i=1}^{w} b_i f_i, \mathbf{a}) \geq \mu\}) \\
&= \dim(\{\mathbf{b} = (b_1, b_2, \ldots, b_w) \in \mathbb{F}^w : \\
&\forall \mathbf{x}^{\mathbf{e}} \text{ s.t } \deg(\mathbf{x}^{\mathbf{e}}) < \mu, \sum_{i=1}^{w} b_i \frac{\partial f_i}{\partial \mathbf{x}^{\mathbf{e}}}(\mathbf{a}) = 0\}) \\
&= \dim\left(\{\mathbf{b} = (b_1, b_2, \ldots, b_w) \in \mathbb{F}^w : \forall \psi \in E, (M_\psi(\mathbf{a}))\mathbf{b} = \mathbf{0}\}\right) \\
&\leq \min_{\psi \in E}(\dim(\text{Kernel}(M_\psi(\mathbf{a})))) \\
&\leq \min_{\psi \in E}(w - \text{rank}(M_\psi(\mathbf{a}))) \ .
\end{aligned}$$

□

*Proof of Claim 1:* To show the claim, we show that for every monomial $\mathbf{x}^{\mathbf{f}}$ of degree less than $(\mu - w + 1)\dim(H_\mathbf{a} \cap W)$, the Hasse derivative $\frac{\partial L_{\psi_0}}{\partial \mathbf{x}^{\mathbf{f}}}$ is zero at $\mathbf{a}$. Let $\psi_0 = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_w)$. Then, we have (using Proposition 1: Item 4 and 5).

$$\begin{aligned}
&\frac{\partial L_{\psi_0}}{\partial \mathbf{x}^{\mathbf{f}}}(\mathbf{a}) = \\
&\sum_{\mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_w = \mathbf{f}} \left( \prod_{j \in [w]} \binom{\mathbf{e}_j + \mathbf{u}_j}{\mathbf{u}_j} \right) \\
&\det(\tilde{M}_{(\mathbf{e}_1 + \mathbf{u}_1, \ldots, \mathbf{e}_w + \mathbf{u}_w)})(\mathbf{a}).
\end{aligned}$$

Now, we know that $\sum_j \|\mathbf{u}_j\|_1 < (\mu - w + 1)\dim(H_\mathbf{a} \cap W)$, so there are less than $\dim(H_\mathbf{a} \cap W)$ values of $j \in \{1, 2, \ldots, w\}$ such that $\|\mathbf{u}_j\|_1$ is more than $\mu - w$. Moreover, $\|\mathbf{u}_j\|_1 \leq \mu - w$ implies that $\|\mathbf{e}_j\|_1 + \|\mathbf{u}_j\|_1 \leq \mu - 1$. Thus, there is a $\psi \in E$, such that there are more than $w - \dim(H_\mathbf{a} \cap W)$ rows of the matrix $\tilde{M}_{(\mathbf{e}_1 + \mathbf{u}_1, \ldots, \mathbf{e}_w + \mathbf{u}_w)}(\mathbf{a})$ which are also rows in the matrix $M_\psi(\mathbf{a})$. But, from Claim 2, we know that for every $\psi \in E$, $M_\psi(\mathbf{a})$ has rank at most $w - \dim(H_\mathbf{a} \cap W)$. Thus, each of the matrices $\tilde{M}_{(\mathbf{e}_1 + \mathbf{u}_1, \ldots, \mathbf{e}_w + \mathbf{u}_w)}(\mathbf{a})$ in the summand above is rank deficient, and hence has determinant zero. □

## APPENDIX I
## EXPONENTIAL NUMBER OF CODEWORDS AT A DISTANCE $\delta$

Let $T \subseteq S$ be an arbitrary subset of size $d/s$. For a variable $x_1$, consider the polynomial $f(\mathbf{x}) = \prod_{b \in T}(x_1 - b)^{s-1}$. At every point $\mathbf{a} \in S^m$ such that $a_1 \in T$, $f(\mathbf{x})$ vanishes with multiplicity at least $s$. Moreover, the set $\{\mathbf{a} \in S^m : a_1 \in T\} \subseteq S^m$ is of size exactly $\frac{d}{s}|S|^{m-1}$. Thus, the encoding of

every polynomial in the set

$$\mathcal{M} = \left\{ \prod_{b \in T} (x_1 - b)^{s-1} : \deg(L(\mathbf{x})) = 1, T \subseteq S, |T| = d/s \right\}$$

under the $m$-variate multiplicity code, with multiplicity parameter $s$ agrees with the encoding of the polynomial 0 on at least $d/(qs)$ fraction of points, i.e., the relative distance between them is $(1-\delta)$, where $\delta$ is the distance of the code. Moreover, the set $\mathcal{M}$ is of size $\binom{q}{d/s}$, which is superpolynomially growing in $d$. In this sense, the error tolerance of the result in Theorem 1 is the best one could hope for (up to the $\varepsilon > 0$ term) if we are hoping for polynomial list size.

## Appendix II
## Generalized Wronskian Criterion

In this section, we give a proof of the generalized Wronskian criterion in the multivariate setting that works over fields of positive characteristic, and using the notion of Hasse derivatives.

We first state and prove a proposition which we will use to prove Theorem 5. Given a sequence $f_1, f_2, \ldots, f_w$ of $w$ $m$-variate polynomials of individual degree at most $d$ and a sequence $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_w$ of $w$ monomials, let $M_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w)$ be the $w \times w$ matrix whose $(i,j)$-th entry is $\frac{\partial f_j}{\partial \mathbf{x}^{\mathbf{e}_i}}$. Let $W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w) := \det\left(M_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w)\right)$: so, $W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w) \in \mathbb{F}[\mathbf{x}]$.

We say that $\mathbf{x}^{\mathbf{e}'} \leq \mathbf{x}^{\mathbf{e}}$ if $\mathbf{e}' \leq \mathbf{e}$, that is, for all $i \in [m]$: $e'_i \leq e_i$. Let $\lesssim$ be the graded-lexicographic order (deglex), which is an extension of the $\leq$ ordering: so, for distinct $\mathbf{e}$ and $\mathbf{e}'$, we have $\mathbf{x}^{\mathbf{e}'} \lesssim \mathbf{x}^{\mathbf{e}}$ iff $\|\mathbf{e}'\|_1 < \|\mathbf{e}\|_1$ or $\|\mathbf{e}'\|_1 = \|\mathbf{e}\|_1$ and $e'_i < e_i$ where $i$ is the first index where $e'_i \neq e_i$. Also, for a polynomial $f \in \mathbb{F}[\mathbf{x}]$, let $\widetilde{f}$ denote its monomial of minimum degree under $\lesssim$ if $f$ is non-zero and 0 otherwise. Thus, for every non-zero polynomial $f$ of the form $\sum_{\mathbf{e}} \alpha_{\mathbf{e}} \cdot \mathbf{x}^{\mathbf{e}}$ with $\alpha_{\mathbf{e}} \in \mathbb{F}$, $\widetilde{f}$ is $\mathbf{x}^{\mathbf{e}^*}$ where $\mathbf{x}^{\mathbf{e}^*}$ is the least monomial among the set of monomials $\{\mathbf{x}^{\mathbf{e}} : \alpha_{\mathbf{e}} \neq 0\}$. For a monomial, $\ell = \mathbf{x}^{\mathbf{e}}$ we denote $\|\mathbf{e}\|_1$ by $|\ell|$.

*Proposition 2:* 1) *(linear combinations) For a fixed $i$, let $f_i = \alpha_i f'_i + \sum_{j \neq i} \alpha_j f_j$ where $\alpha_j \in \mathbb{F}$. Then*

$$W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w) =$$

$$\alpha_i \cdot W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_{i-1}, f'_i, f_{i+1}, \ldots, f_w).$$

2) *(translation) Let $\mathbf{x} + \mathbf{1} = (x_1 + 1, x_2 + 1, \ldots, x_m + 1)$. Then*

$$(W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1(\mathbf{x}), \ldots, f_w(\mathbf{x})))(\mathbf{x} + \mathbf{1}) =$$

$$(W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1(\mathbf{x} + \mathbf{1}), \ldots, f_w(\mathbf{x} + \mathbf{1})))(\mathbf{x}).$$

3) *(minimum monomial) If $W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(\widetilde{f_1}, \ldots, \widetilde{f_w}) \neq 0$, then*

$$\widetilde{W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}}(f_1, \ldots, f_w) = W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(\widetilde{f_1}, \ldots, \widetilde{f_w}).$$

*Proof:* By linearity of Hasse derivatives we have

$$\frac{\partial f'_i}{\partial \mathbf{x}^{\mathbf{e}}} = \alpha_i \frac{\partial f_i}{\partial \mathbf{x}^{\mathbf{e}}} + \sum_{j \neq i} \alpha_j \frac{\partial f_j}{\partial \mathbf{x}^{\mathbf{e}}}.$$

Hence, $M_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_w)$ and $M_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(f_1, \ldots, f_{i-1}, f'_i, f_{i+1}, \ldots, f_w)$ are related by column elementary operations. Thus, their determinants are the same modulo a multiplicative factor of $\alpha_i$. This proves Item 1. The proof of Item 2 follows from the fact that for any $f \in \mathbb{F}[\mathbf{x}]$ we have $(\frac{\partial f}{\partial \mathbf{x}^{\mathbf{e}}})(\mathbf{x} + \mathbf{1}) = (\frac{\partial f(\mathbf{x}+\mathbf{1})}{\partial \mathbf{x}^{\mathbf{e}}})(\mathbf{x})$. Also, Item 3 follows directly by expanding out the determinant. □

Equipped with this proposition, we will now show that if $f_1, \ldots, f_w$ are linearly independent over $\mathbb{F}$, then there exist monomials $\mathbf{x}^{\mathbf{e}_1}, \ldots, \mathbf{x}^{\mathbf{e}_w}$ such that $W_{(\mathbf{x}^{\mathbf{e}_1},\ldots,\mathbf{x}^{\mathbf{e}_w})}(f_1, \ldots, f_w) \neq 0$ and $\deg(\mathbf{x}^{\mathbf{e}_i}) < i$.

*Proof of Theorem 5:* Using Proposition 2-Item 1 we can WLOG assume that each $f_i$ has a distinct minimum monomial. We can take an appropriate linear combination of the $f_i$s of the form $f_i \leftarrow f_i + \sum_{j \neq i} \alpha_j f_j$ (this preserves linear independence) to clear out a minimum monomial if it repeats. Hence, the minimal monomials $\widetilde{f_i}$s are all distinct. Further, by reordering if necessary we can assume that $\widetilde{f_i}$s are in increasing order according to $\lesssim$. Now, using Proposition 2-Item 3, we are left to show that there are $\mathbf{x}^{\mathbf{e}_1}, \ldots, \mathbf{x}^{\mathbf{e}_w}$ such that $\deg(\mathbf{x}^{\mathbf{e}_i}) < i$ and $W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(\widetilde{f_1}, \ldots, \widetilde{f_w}) \neq 0$. To show this first we massage the monomials in the following manner.

1) Set $t \leftarrow 0$ and for all $i \in [w]$ let $\ell_i^0 \leftarrow \widetilde{f_i}$.
2) While $(\exists i : |\ell_i^t| \geq i)$:
   a) For all $i$ let $g_i^{t+1} = \ell_i^t(\mathbf{x} + \mathbf{1})$.
   b) Take appropriate linear combinations of the form $g_i^{t+1} \leftarrow g_i^{t+1} - \sum_{j<i} \alpha_j \cdot g_j^{t+1}$ to ensure that all $\widetilde{g_i^{t+1}}$s are distinct.
   c) For all $i$ set $\ell_i^{t+1} \leftarrow \widetilde{g_i^{t+1}}$. Reorder to ensure that the $\ell_i^{t+1}$s are in increasing order wrt $\lesssim$.
   d) $t \leftarrow t + 1$.

We will now show that the while loop terminates in at most $w$ steps and at the end we have $|\ell_i^t| < i$ for all $i \in [w]$. Suppose we enter the while loop at a particular value of $t$. Let $i^*$ be the first index such that $|\ell_i^t| \geq i$. Observe that $g_{i^*}^{t+1}$ will include all monomials $\mathbf{x}^{\mathbf{e}'}$ such that $\mathbf{e}' \leq \mathbf{e}$ where $\ell_{i^*}^t = \mathbf{x}^{\mathbf{e}}$. This is because the characteristic of $\mathbb{F}$ is larger than the maximum individual degree. Hence, at time $t+1$ we will have $|\ell_j^{t+1}| < j$ for all $j \leq i^*$: for $j < i^*$ step 2$(b)$ does not increase the degree of $g_j^{t+1}$ and for $j = i^*$ the minimal monomial $\widetilde{g_{i^*}^{t+1}}$ will be of degree less than $i^*$ as $g_{i^*}^{t+1}$ includes a monomial of degree $i^* - 1$ which does not occur in any $g_j^{t+1}$ for $j < i^*$. Thus at termination, we have $|\ell_i^t| < i$ for all $i \in [w]$ and further the $\ell_i^t$s are all distinct monomials and in increasing order.

Also, by Proposition 2 we have that if $W_{\mathbf{e}_1,\ldots,\mathbf{e}_w}(\ell_1^{t+1}, \ldots, \ell_w^{t+1}) \neq 0$, then, $W_{\mathbf{e}_1,\ldots,\mathbf{e}_w}(\ell_1^t, \ldots, \ell_w^t) \neq 0$. At termination set $\ell_i = \ell_i^t$. Hence, we are left to show that there are $\mathbf{x}^{\mathbf{e}_1}, \ldots, \mathbf{x}^{\mathbf{e}_w}$ such that $\deg(\mathbf{x}^{\mathbf{e}_i}) < i$ and $W_{(\mathbf{e}_1,\ldots,\mathbf{e}_w)}(\ell_1, \ldots, \ell_w) \neq 0$. Towards this end observe that the matrix $M_{(\ell_1,\ldots,\ell_w)}(\ell_1, \ldots, \ell_w)$ is upper triangular with all the diagonal entries as 1. For contradiction suppose that $i > j$ and $\frac{\partial \ell_i}{\partial \ell_j} \neq 0$: then $\ell_j > \ell_i$ which is a contradiction.

Hence, $W_{(\ell_1,\ldots,\ell_w)}(\ell_1,\ldots,\ell_w) = 1$. Thus, letting $\mathbf{x}^{\mathbf{e}_i} = \ell_i$ for all $i \in [w]$ gives us the requisite monomials $\mathbf{x}^{\mathbf{e}_i}$.

The other direction that if there are monomials $\mathbf{x}^{\mathbf{e}_1},\ldots,\mathbf{x}^{\mathbf{e}_w}$ such that $W_{(\mathbf{x}^{\mathbf{e}_1},\ldots,\mathbf{x}^{\mathbf{e}_w})}(f_1,\ldots,f_w) \neq 0$ then $f_1,\ldots,f_w$ are linearly independent, is simpler. Suppose the $f_i$s are linearly dependent and in particular, $\sum_i \alpha_i f_i$ be a non-trivial linear combination which is zero. Due to linearity of Hasse derivatives we have $(\alpha_1,\ldots,\alpha_w) \in \ker\left(M_{(\mathbf{x}^{\mathbf{e}_1},\ldots,\mathbf{x}^{\mathbf{e}_w})}(f_1,\ldots,f_w)\right)$. This completes the proof of Theorem 5. □

## ACKNOWLEDGMENT

The authors would like to thank Swastik Kopparty for many insightful discussions on multiplicity codes and on the results in [12] and [13].

## REFERENCES

[1] S. Bhandari, P. Harsha, M. Kumar, and M. Sudan, "Decoding multivariate multiplicity codes on product sets," in *Proc. 53rd Annu. ACM SIGACT Symp. Theory Comput.*, S. Khuller and V. V. Williams, Eds., Jun. 2021, pp. 1489–1501.

[2] Ø. Ore, "Über höhere kongruenzen (German) [about higher congruences]," *Norsk Mat. Forenings Skrifter*, vol. 1, no. 7, p. 15, 1922.

[3] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, Oct. 1980.

[4] R. Zippel, "Probabilistic algorithms for sparse polynomials," in *Proc. Int. Symp. Symbolic Algebr. Comput. (EUROSAM)*, in Lecture Notes in Computer Science, vol. 72, E. W. Ng, Ed. Berlin, Germany: Springer, 1979, pp. 216–226, doi: 10.1007/3-540-09519-5_73.

[5] R. A. Demillo and R. J. Lipton, "A probabilistic remark on algebraic program testing," *Inf. Process. Lett.*, vol. 7, no. 4, pp. 193–195, Jun. 1978.

[6] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan, "Extensions to the method of multiplicities, with applications to Kakeya sets and mergers," *SIAM J. Comput.*, vol. 42, no. 6, pp. 2305–2328, Jan. 2013.

[7] S. Saraf, "The method of multiplicities," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2011. [Online]. Available: http://hdl.handle.net/1721.1/68494

[8] L. Guth, *Polynomial Methods in Combinatorics* (University Lecture Series), vol. 64. Providence, RI, USA: American Mathematical Society, 2016. [Online]. Available: https://bookstore.ams.org/ulect-64/

[9] J. Y. Kim and S. Kopparty, "Decoding Reed–Müller codes over product sets," *Theory Comput.*, vol. 13, no. 1, pp. 1–38, 2017.

[10] M. Y. Rosenbloom and M. A. Tsfasman, "Codes for the $m$-metric," *Problemy Peredachi Informatsii*, vol. 33, no. 1, pp. 55–63, 1997.

[11] R. R. Nielsen, "List decoding of linear block codes," Ph.D. dissertation, Dept. Math., Tech. Univ. Denmark, Lyngby, Denmark, 2001. [Online]. Available: https://orbit.dtu.dk/en/publications/list-decoding-of-linear-block-codes

[12] S. Kopparty, S. Saraf, and S. Yekhanin, "High-rate codes with sublinear-time decoding," *J. ACM*, vol. 61, no. 5, pp. 28:1–28:20, 2014.

[13] S. Kopparty, "List-decoding multiplicity codes," *Theory Comput.*, vol. 11, pp. 149–182, May 2015.

[14] V. Guruswami and C. Wang, "Linear-algebraic list decoding for variants of Reed–Solomon codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3257–3268, Jun. 2013.

[15] M. Sudan, "Decoding of Reed Solomon codes beyond the error-correction bound," *J. Complex.*, vol. 13, no. 1, pp. 180–193, Mar. 1997.

[16] V. Guruswami and M. Sudan, "Improved decoding of Reed–Solomon and algebraic-geometry codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1757–1767, Sep. 1999.

[17] S. Kopparty, N. Ron-Zewi, S. Saraf, and M. Wootters, "Improved decoding of folded Reed–Solomon and multiplicity codes," in *Proc. IEEE 59th Annu. Symp. Found. Comput. Sci. (FOCS)*, M. Thorup, Ed., Oct. 2018, pp. 212–223.

[18] S. Bhandari, P. Harsha, M. Kumar, and A. Shankar, "Algorithmizing the multiplicity Schwartz–Zippel lemma," in *Proc.34th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, N. Bansal, Ed., 2023, pp. 2816–2835.

[19] S. Kopparty, "Some remarks on multiplicity codes," in *Discrete Geometry and Algebraic Combinatorics* (Contemporary Mathematics), vol. 625, A. Barg and O. R. Musin, Eds. Providence, RI, USA: AMS, 2014, pp. 155–176.

[20] S. P. Vadhan, "Pseudorandomness," *Found. Trends Theor. Comput. Sci.*, vol. 7, nos. 1–3, pp. 1–336, 2012.

[21] R. Pellikaan and X.-W. Wu, "List decoding of q-ary Reed–üller codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 679–682, Apr. 2004.

[22] Z. Guo, M. Kumar, R. Saptharishi, and N. Solomon, "Derandomization from algebraic hardness," *SIAM J. Comput.*, vol. 51, no. 2, pp. 315–335, Apr. 2022.

[23] A. Shpilka and A. Yehudayoff, "Arithmetic circuits: A survey of recent results and open questions," *Found. Trends Theor. Comput. Sci.*, vol. 5, nos. 3–4, pp. 207–388, 2009. [Online]. Available: https://www.cs.tau.ac.il/~shpilka/publications/SY10.pdf

[24] M. Mahajan and V. Vinay, "Determinant: Combinatorics, algorithms, and complexity," *Chicago J. Theor. Comput. Sci.*, vol. 1997, no. 5, pp. 1–28, 1997.

[25] V. Guruswami and S. Kopparty, "Explicit subspace designs," *Combinatorica*, vol. 36, no. 2, pp. 161–185, Apr. 2016.

[26] A. Bostan and P. Dumas, "Wronskians and linear independence," *Amer. Math. Monthly*, vol. 117, no. 8, pp. 722–727, 2010.

[27] R. Lidl and H. Niederreiter, *Finite Fields* (Encyclopedia of Mathematics and its Applications), vol. 2, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.

**Siddharth Bhandari** received the M.S. degree in computer science from the Chennai Mathematical Institute in 2017 and the Ph.D. degree from the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai. He is currently with the Toyota Technological Institute at Chicago. His research interests include coding theory, combinatorics, sampling algorithms, zero-error information theory, and causal inference.

**Prahladh Harsha** received the bachelor's degree in computer science and engineering from IIT Madras in 1998 and the master's and Ph.D. degrees in computer science from MIT in 2000 and 2004, respectively. He is currently a Theoretical Computer Scientist with the Tata Institute of Fundamental Research (TIFR). Prior to joining TIFR in 2010, he was at Microsoft Research, Silicon Valley, and the Toyota Technological Institute at Chicago. His research interests include computational complexity, algebraic coding theory, probabilistically checkable proofs, and hardness of approximation.

**Mrinal Kumar** received the Ph.D. degree from Rutgers University, under the supervision of Swastik Kopparty and Shubhangi Saraf. He is currently a Reader with the School of Technology and Computer Science, Tata Institute of Fundamental Research (TIFR), Mumbai. His research interests include algebra and computation, algebraic complexity theory, and error correcting codes.

**Madhu Sudan** received the bachelor's degree from IIT Delhi in 1987 and the Ph.D. degree from UC Berkeley in 1992. From 1992 to 2015, he was with IBM Research, MIT, and Microsoft Research. Since 2015, he has been a Gordon McKay Professor with the School of Engineering and Applied Sciences, Harvard University. His research interests include mathematical studies of communication and computation. Specifically, his research focuses on concepts of reliability and mechanisms that are, or can be, used by computers to interact reliably with each other. His research draws on tools from computational complexity, which studies efficiency of computation, and many areas of mathematics, including algebra and probability theory. He is a fellow of ACM and AMS. He is a member of the American Academy of Arts and Sciences and the National Academy of Sciences. He was a recipient of the Nevanlinna Prize, the Infosys Foundation Prize in Mathematical Sciences, and the IEEE Hamming Medal.