Proximity Gaps for Reed–Solomon Codes

ELI BEN-SASSON, StarkWare Industries Ltd., Israel
DAN CARMON, StarkWare Industries Ltd., Israel
YUVAL ISHAI*, Technion, Israel
SWASTIK KOPPARTY†, Departments of Mathematics and of Computer Science, University of Toronto, Canada
SHUBHANGI SARAF‡, Departments of Mathematics and of Computer Science, University of Toronto, Canada

A collection of sets displays a *proximity gap* with respect to some property if for every set in the collection, either (i) all members are δ-close to the property in relative Hamming distance or (ii) only a tiny fraction of members are δ-close to the property. In particular, no set in the collection has roughly half of its members δ-close to the property and the others δ-far from it.

We show that the collection of affine spaces displays a proximity gap with respect to Reed–Solomon (RS) codes, even over small fields, of size polynomial in the dimension of the code, and the gap applies to any δ smaller than the Johnson/Guruswami–Sudan list-decoding bound of the RS code. We also show near-optimal gap results, over fields of (at least) *linear* size in the RS code dimension, for δ smaller than the unique decoding radius. Concretely, if δ is smaller than half the minimal distance of an RS code $V \subset \mathbb{F}_q^n$, every affine space is either entirely δ-close to the code, or alternatively at most an (n/q)-fraction of it is δ-close to the code. Finally, we discuss several applications of our proximity gap results to distributed storage, multi-party cryptographic protocols, and concretely efficient proof systems.

We prove the proximity gap results by analyzing the execution of classical algebraic decoding algorithms for Reed–Solomon codes (due to Berlekamp–Welch and Guruswami–Sudan) on a *formal element* of an affine space. This involves working with Reed–Solomon codes whose base field is an (infinite) rational function field. Our proofs are obtained by developing an extension (to function fields) of a strategy of Arora and Sudan for analyzing low-degree tests.

CCS Concepts: • **Theory of computation** → **Interactive proof systems**; **Error-correcting codes**; Communication complexity; • **Computing methodologies** → **Linear algebra algorithms**; *Algebraic algorithms*; • **Mathematics of computing** → Coding theory.

Additional Key Words and Phrases: Reed-Solomon Codes, Proximity Gaps, Interactive Oracle Proofs (IOP), Interactive Oracle Proofs of Proximity (IOPP), Verifiable Secret Sharing

*Supported by ERC Project NTSC (742754), NSF-BSF grant 2015782, BSF grant 2018393, and a grant from the Ministry of Science and Technology, Israel and Department of Science and Technology, Government of India. Work done in part while participating in the Simons Institute program on Proofs, Consensus, and Decentralizing Society.
†Research supported in part by NSF grants CCF-1540634 and CCF-1814409 and BSF grant 2014359. Work done in part while attending a workshop at the Simons Institute program on Proofs, Consensus, and Decentralizing Society.
‡Research supported in part by NSF grants CCF-1540634 and CCF-1909683, BSF grant 2014359, a Sloan research fellowship and the Simons Collaboration on Algorithms and Geometry

Authors' addresses: Eli Ben-Sasson, eli@starkware.co, StarkWare Industries Ltd., Netanya, Israel; Dan Carmon, dancar@starkware.co, StarkWare Industries Ltd., Netanya, Israel; Yuval Ishai, yuvali@cs.technion.ac.il, Technion, Haifa, Israel; Swastik Kopparty, swastik.kopparty@gmail.com, Departments of Mathematics and of Computer Science, University of Toronto, Toronto, Ontario, Canada; Shubhangi Saraf, shubhangi.saraf@gmail.com, Departments of Mathematics and of Computer Science, University of Toronto, Toronto, Ontario, Canada.

0004-5411/2023/8-ART
https://doi.org/10.1145/3614423

1 INTRODUCTION

A variety of protocols, arising in the contexts of interactive proofs, distributed storage and cryptography, give rise to the following problem regarding proximity to a linear code $V \subset \mathbb{F}_q^n$ over a finite field $\mathbb{F}_q$ of minimal relative distance δ_V. These myriad protocols assume oracle access to a batch of vectors $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^n$ and their soundness requires that each and every vector u_i be close to V in relative Hamming distance. Furthermore, soundness deteriorates as a function of the largest distance between some vector u_i and the code V. Thus, we seek protocols that minimize the number of queries to the entries of the vectors in $\mathbf{u}$, while maximizing the probability of recognizing when some vector u_i is significantly far from V.

The linearity of V suggests a natural approach, first explored by Rothblum, Vadhan and Wigderson [35]: sample a uniformly random vector u' in the span of $\mathbf{u}$ (denoted $\mathrm{span}(\mathbf{u})$) and view the distance between u' and V, denoted $\Delta(u', V)$, as a proxy for the maximal distance between some member of $\mathbf{u}$ and V. To argue soundness, we would like to show that if even a single u_i is δ-far from (all members of) V, then a randomly chosen u' is also far from V. Indeed, the paper [35] that suggested this approach also showed for any V, that whenever a single u_i is δ-far from V, then nearly all samples u' are at least $\delta/2$-far from V. Here and henceforth, we use Δ to denote relative Hamming distance and say "u is δ-close to V", denoted $\Delta(u, V) \leq \delta$, when $\Delta(u, v) \leq \delta$ for some $v \in V$; otherwise we say "u is δ-far from V" (denoted $\Delta(u, V) > \delta$).

Note that the result above incurs a $2\times$ degradation in the proximity parameter δ: the worst-case assumption — that some u_i is δ-far from V — implies an average-case distance that is only $\delta/2$. Eliminating the proximity degradation is easy when the field size is exponential in the code length. More concretely, if $q \gg 2^{nH(\delta)}$, where H is the binary entropy function, then a union bound over agreement sets shows that for $\delta < \delta_V$, if u_i is δ-far from V then so are nearly all $u' \in \mathrm{span}(\mathbf{u})$. However, exponential field size is prohibitively large in the context of the motivating applications. Obtaining similar results over fields of sub-exponential size appears to be much more challenging.

A number of works looked at this question and were able to remove the degradation in δ with polynomial field size. Ames et al. [1] showed that for proximity parameters δ that are smaller than half of the unique-decoding radius of V (i.e., when $\delta < \delta_V/4$), nearly all $u' \in \mathrm{span}(\mathbf{u})$ are δ-far from V. The proximity bound was subsequently improved to $\delta < \delta_V/3$ by Roth and Zémor [34]. Ben-Sasson et al. [12] showed similar results for δ above the unique decoding radius, holding for any $\delta < 1 - \sqrt{1 - \delta_V}$, and the state of the art[1] was given in [11], holding for any $\delta < 1 - \sqrt[3]{1 - \delta_V}$. In fact, this latter result was shown to be tight for certain RS codes, in particular, of maximal blocklength $n = q$.

Ames et al., who were the first to show that in certain cases the average-case distance of $u' \in \mathrm{span}(\mathbf{u})$ from V is nearly-always equal to the worst-case distance of $u_i \in \mathbf{u}$ from V, also raised the following intriguing question, which is at the focus of our investigation here: For which codes and what range of δ does the following statement hold?

If some $u^* \in \mathrm{span}(\mathbf{u})$ is δ-far from V, then so are nearly all $u' \in \mathrm{span}(\mathbf{u})$.

One implication of our main result is that when V is an RS code over a sufficiently large field — polynomially large in the code's blocklength — and when δ is smaller than the Johnson/Guruswami–Sudan list decoding bound, the above phenomenon holds. We refer to it as a *proximity gap*, as explained next.

1.1 Gaps and proximity gaps

When a "gap" is mentioned in theoretical computer science, it usually refers to a situation where all objects under consideration must fall into one of two categories, and these categories display a large gap according to some metric. Striking examples are given by PCP reductions whose outputs are constraint satisfaction problems

[1]We note that these improvements give a roughly $2\times$ improvement to the protocol of [35] in which this question was originally studied, when that protocol is instantiated with codes of sufficiently large relative distance (see Theorem 3.4 there).

that lie in one of two categories: satisfiable instances in which some assignment satisfies all constraints, and unsatisfiable instances in which all assignments fail to satisfy more than an ϵ fraction of constraints. Another gap example underlies randomized algorithms. For instance, the Miller–Rabin primality test relies on a gap between primes and composites: in the latter case (composites), at least three-quarters of the integers serve as composite witnesses whereas for primes none do, leading to a "gap" of measure $3/4$.

Our main result can be phrased as a *proximity gap* according to the following definition.

Definition 1.1 (Proximity gap)**.** *Let* $\mathsf{P} \subset \Sigma^n$ *be a property and* $\mathsf{C} \subset 2^{\Sigma^n}$ *be a collection of sets. Let* Δ *be a distance measure on* Σ^n*. We say that* C *displays a* (δ, ϵ)*-proximity gap with respect to* P *under* Δ *if every* $\mathsf{S} \in \mathsf{C}$ *satisfies exactly one of the following:*

(1) $\Pr_{s \in \mathsf{S}}[\Delta(s, \mathsf{P}) \leq \delta] = 1$.
(2) $\Pr_{s \in \mathsf{S}}[\Delta(s, \mathsf{P}) \leq \delta] \leq \epsilon$.

We call δ *the* proximity parameter *and* ϵ *is the* error parameter*. By default,* Δ *denotes the relative Hamming distance measure.*

Using this definition, our main result can be informally stated as follows: if $V \subset \mathbb{F}^n$ is an RS code and $A \subset \mathbb{F}^n$ is an affine space, then either all elements of A are close to V, or otherwise, nearly all elements of A are far from V. In other words, there is no affine A in which roughly half of the elements are close to V while the other half are far from V.

Throughout this paper, $\mathbb{F}_q$ denotes the field of size q, and $\mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k]$ is the RS code of *dimension* $k+1$ and *blocklength* $n = |\mathcal{D}|$ containing as its codewords the polynomials of degree $\leq k$, evaluated on $\mathcal{D}$. Formally,

$$\mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k] = \{f : \mathcal{D} \to \mathbb{F}_q \mid \deg(f) \leq k\}, \tag{1.1}$$

where $\deg(f)$ denotes the degree of the *interpolant* of f – the unique polynomial of degree stricly less than $|\mathcal{D}|$ whose evaluation on $\mathcal{D}$ is f.

We use ρ to denote the *rate* $\rho = \frac{k+1}{n}$ of the code. The letter δ will denote distance in a metric space, often referring to the relative Hamming distance between a function and an ambient RS code, and ϵ will denote an error parameter, the probability that a "bad event" occurs (with varying definitions of the term "bad event").

The following result has two parts and each part has its own proof. The first part holds only below the unique decoding radius but has a smaller error parameter, denoted ϵ_{U}, which is in fact essentially tight; the second part holds for proximity parameters up to the Johnson/Guruswami–Sudan bound (which is greater than the unique decoding bound) but has a larger error bound ϵ_{J} (the proof of the second part is also significantly harder).

Theorem 1.2 (Proximity gap for RS codes). *For all prime powers* q *and* $\mathcal{D} \subseteq \mathbb{F}_q$*, the collection* $\mathsf{C}_{\mathsf{Affine}}$ *of affine spaces in* $\mathbb{F}_q^{\mathcal{D}}$ *displays a* (δ, ϵ)*-proximity gap with respect to the RS code* $V := \mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k]$ *of blocklength* n *and rate* $\rho = \frac{k+1}{n}$*, for any* $\delta \in (0, 1 - \sqrt{\rho})$*, and* $\epsilon = \epsilon(q, n, \rho, \delta)$ *defined as the following piecewise function:*

- **Unique decoding bound:** *For* $\delta \in \left(0, \frac{1-\rho}{2}\right]$*, the error parameter* ϵ *is*

$$\epsilon = \epsilon_{\mathsf{U}} = \epsilon_{\mathsf{U}}(q, n) := \frac{n}{q}. \tag{1.2}$$

- **Johnson bound:** *For* $\delta \in \left(\frac{1-\rho}{2}, 1 - \sqrt{\rho}\right)$*, setting* $\eta := 1 - \sqrt{\rho} - \delta$*, the error parameter* ϵ *is*

$$\epsilon = \epsilon_{\mathsf{J}} = \epsilon_{\mathsf{J}}(q, n, \rho, \delta) := \frac{(k+1)^2}{\left(2 \min\left(\eta, \frac{\sqrt{\rho}}{20}\right)\right)^7 q} = O\left(\frac{1}{(\eta\rho)^{O(1)}} \cdot \frac{n^2}{q}\right) \tag{1.3}$$

There are two striking aspects to this result. First, the proximity parameter δ can take any value smaller than the famous Johnson/Guruswami–Sudan bound, which is the largest distance for which we know of efficient (list) decoding algorithms. (Looking ahead, the Guruswami–Sudan algorithm will play a crucial, though non-algorithmic, role in our proofs.) Second, the size of the field needed to achieve this result is relatively small — linear in the blocklength when δ is below the unique decoding radius $\delta < (1-\rho)/2$ and, for fixed rate, quadratic in blocklength for larger δ up to the list decoding bound.

Remark 1.1 (On the tightness of our results). We make several comments on the margins for possible improvement.

- **Proximity parameter δ.** The maximal proximity parameter δ for which Theorem 1.2 applies happens to coincide with the Johnson/Guruswami–Sudan list-decoding bound $(1-\sqrt{\rho})$. This evidently follows from the techniques we use here, which rely on list-decoding algorithms that reach that bound. However, we conjecture that Theorem 1.2 holds even for larger proximity parameters, up to capacity $(1-\rho)$. See Conjecture 8.4 and the discussion there.
- **Field size q.** The bound in Eq. (1.3) which reaches the Johnson bound becomes nontrivial only for fields of size q that are at least quadratically larger than the blocklength n. In contrast, the bound for smaller proximity parameters, below the unique decoding radius, works for $q = O(n)$ (see Eq. (1.2)). We point out that for certain combinations of fields and rate parameters one cannot hope to reach the Johnson bound with linear size fields, as this would contradict prior results from [12].
- **Error parameter ϵ.** In the unique decoding regime, our bound on the error ϵ is sharp in the sense that affine spaces do not all display a proximity gap with $q \cdot \epsilon$ being sublinear in n, for fixed distance parameter δ. A simple example is of the affine line $\{u_0 + zu_1 : z \in \mathbb{F}_q\}$, where $u_0, u_1 : \mathcal{D} \to \mathbb{F}_q$ are such that on a set $\mathcal{D}' \subset \mathcal{D}$ of size $|\mathcal{D}'| = n(1-\delta) - 1$ we have $u_0|_{\mathcal{D}'} = u_1|_{\mathcal{D}'} = 0$, and on the complement we have that $u_1|_{\mathcal{D}\setminus\mathcal{D}'} = 1$, and u_0 takes $\delta n + 1$ distinct non-zero values. We then have $\Delta(u_0 + zu_1, V) \le \delta$ for each of the $\delta n + 1$ values of $z \in \mathbb{F}_q$ for which $-z$ is in the image of $u_0|_{\mathcal{D}\setminus\mathcal{D}'}$, but that $\Delta(u_0, V) = \delta + \frac{1}{n} > \delta$, thus this line does not display a $(\delta, \frac{\delta n}{q})$-proximity gap with respect to the code.

1.2 Concentration bounds

Theorem 1.2 implies the following concentration bound, saying that for any affine space in which the element farthest from the RS code is within the Johnson/Guruswami–Sudan radius, nearly all elements are at exactly the same distance from the code(!).

For two sets $U, V \subset \Sigma^n$ define the *divergence*[2] of U from V as $\mathrm{D}(U,V) := \max_{u\in U} \Delta(u,V)$.

Corollary 1.3 (Concentration bounds)**.** *Let $V, q, n, k, \mathcal{D}$ and ρ be as defined in Theorem 1.2. Let $U \subset \mathbb{F}_q^{\mathcal{D}}$ be an affine space over $\mathbb{F}_q$ and denote $\delta^* := \mathrm{D}(U,V)$. If δ^* is smaller than the Johnson/Guruswami–Sudan bound, then nearly all elements of U have distance* exactly *δ^* from the code. In other words, if $\delta^* \in (0, 1-\sqrt{\rho})$, then*

$$\Pr_{u\in U}\left[\Delta(u,V) \neq \delta^*\right] \le \epsilon,$$

where $\epsilon = \epsilon(q, n, \rho, \delta^*)$ is as defined in Theorem 1.2.

Proof. Define $\delta = \delta^* - \frac{1}{n} < \delta^*$. Note that because the values of Δ are integer multiples of $\frac{1}{n}$, so are δ^* and δ, and we have that for all $u \in U$, $\Delta(u,V) < \delta^* \iff \Delta(u,V) \le \delta$. On the other hand by the maximality of δ^*, we have $\Delta(u,V) \neq \delta^* \iff \Delta(u,V) < \delta^*$, i.e.

$$\Pr_{u\in U}\left[\Delta(u,V) \neq \delta^*\right] = \Pr_{u\in U}\left[\Delta(u,V) < \delta^*\right] = \Pr_{u\in U}\left[\Delta(u,V) \le \delta\right].$$

[2]Note that divergence is not symmetric as can be seen, e.g., when U is a strict subset of V.

This probability cannot equal 1, since *some* $u \in U$ exists with $\Delta(u, V) = \delta^*$, by definition. Thus the proximity gap from Theorem 1.2 gives

$$\Pr_{u \in U}[\Delta(u, V) \neq \delta^*] = \Pr_{u \in U}[\Delta(u, V) \leq \delta] \leq \epsilon(q, n, \rho, \delta) \leq \epsilon(q, n, \rho, \delta^*) = \epsilon,$$

where the last inequality is due to ϵ being monotone non-decreasing as a function of the δ parameter. □

When the divergence of U from the RS code V is greater than the Johnson/Guruswami–Sudan bound ($\delta^* > 1 - \sqrt{\rho}$) we may still use Theorem 1.2 to conclude that nearly all elements of U are $\approx (1 - \sqrt{\rho})$-far from V, but what remains an interesting open problem is whether nearly all members of U are maximally far (δ^*-far) from V. An example from [11] shows that this need not be the case for RS codes where $q = O(n)$.

1.3 Correlated agreement

Next, we state the main technical theorem proved in the paper. Consider two vectors $u_0, u_1 \in \mathbb{F}^{\mathcal{D}}$. The result says that if sufficiently many elements in the 1-dimensional affine space $A = \{u_0 + zu_1 : z \in \mathbb{F}\}$ are sufficiently close (δ-close) to the RS code V, then there must be a nontrivial subdomain $\mathcal{D}' \subset \mathcal{D}$ of density $1 - \delta$ in $\mathcal{D}$, such that restricting u_0, u_1 to $\mathcal{D}'$ gives a valid RS codeword (evaluated over $\mathcal{D}'$). We refer to the property that such a $\mathcal{D}'$ exists as *correlated agreement*, in the sense that u_0, u_1 and the elements of A do not only have large agreement with the RS code individually, but also share a common large agreement set. The result has two ranges of parameters, as in prior statements in this paper. For proximity parameters in the unique decoding regime this is proved in Theorem 4.1, and for proximity parameters in the list decoding regime this is proved in Theorem 5.1.

Theorem 1.4 (Main Theorem — Correlated agreement over lines). *Let $V, q, n, k, \mathcal{D}$ and ρ be as defined in Theorem 1.2. For $u_0, u_1 \in \mathbb{F}_q^{\mathcal{D}}$, if $\delta \in (0, 1 - \sqrt{\rho})$ and*

$$\Pr_{z \in \mathbb{F}_q}[\Delta(u_0 + z \cdot u_1, V) \leq \delta] > \epsilon,$$

where ϵ is as defined in Theorem 1.2, then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, v_1 \in V$ satisfying

- **Density:** $|\mathcal{D}'|/|\mathcal{D}| \geq 1 - \delta$, *and*
- **Agreement:** v_0 *agrees with* u_0 *and* v_1 *agrees with* u_1 *on all of* $\mathcal{D}'$.

Furthermore, in the unique decoding regime $\delta \in \left(0, \frac{1-\rho}{2}\right]$, there exists a unique maximal *$\mathcal{D}'$ satisfying the above, with unique v_i.*

Remark 1.2 (Sampling from extension fields). One may sample z from a finite extension field $\mathbb{F}_{q'}$ of $\mathbb{F}_q$. In this case, the statement above holds with ϵ_{U} and ϵ_{J} modified by replacing q with q' in the denominators of Eqs. (1.2) and (1.3), respectively. Note that even in this setting, the vectors v_0, v_1 deduced to exist in Theorem 1.4 belong to $\mathrm{RS}[\mathbb{F}_q, \mathcal{D}, k]$, not just in $\mathrm{RS}[\mathbb{F}_{q'}, \mathcal{D}, k]$, because v_0, v_1 have high agreement with $u_0, u_1 \in \mathbb{F}_q^{\mathcal{D}}$. The ability to sample from a larger field (and incur smaller error) applies to the other statements of this section but for simplicity we state all of them using a single field $\mathbb{F}_q$ to both define V and sample z from.

Motivated by applications (described later), we generalize the theorem above to two interesting cases: (i) low-degree parameterized curves, and (ii) higher-dimensional affine spaces; details follow.

Correlated agreement over parameterized curves. The first extension of Theorem 1.4 extends it from the case of a "line" passing through u_0 and u_1 (the line being $\{u_0 + zu_1 : z \in \mathbb{F}\}$) to a "low-degree curve" with coefficients $u_0, u_1, \ldots, u_l$, as described below. This result is of particular importance for two reasons. First, it leads to derandomized testing of verifiable secret sharing and other multi-party protocols (cf. Section 8.1). Second, it improves the soundness analysis of the Fast RS IOPP (FRI) protocol [5], which is used in concretely efficient and transparent (public coin) proof systems [6–9, 17]. We discuss this application in Sections 3.2 and 8.2.

Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$. The *parameterized curve* of degree l that is generated by $\mathbf{u}$ is the following collection of vectors in $\mathbb{F}_q^{\mathcal{D}}$:

$$\mathsf{curve}(\mathbf{u}) := \left\{ u_z := \sum_{i=0}^{l} z^i \cdot u_i \;\middle|\; z \in \mathbb{F}_q \right\}.$$

Theorem 1.5 (Correlated agreement for low-degree parameterized curves). *Let V, q, n, k, $\mathcal{D}$ and ρ be as defined in Theorem 1.2. Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$. If $\delta \in (0, 1 - \sqrt{\rho})$ and*

$$\Pr_{u \in \mathsf{curve}(\mathbf{u})} [\Delta(u, V) \leq \delta] > l \cdot \epsilon,$$

where ϵ is as defined in Theorem 1.2, then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, \ldots, v_l \in V$ satisfying

- **Density:** *$|\mathcal{D}'|/|\mathcal{D}| \geq 1 - \delta$, and*
- **Agreement:** *for all $i \in \{0, \ldots, l\}$, the functions u_i and v_i agree on all of $\mathcal{D}'$.*

Furthermore, in the unique decoding regime $\delta \in \left(0, \frac{1-\rho}{2}\right]$, there exists a unique maximal *$\mathcal{D}'$ satisfying the above, with unique v_i.*

Correlated agreement for affine spaces. The second generalization of our Main Theorem 1.4, extends it from the 1-dimensional case (affine line) to an affine space of arbitrary dimension. Theorem 1.2 follows directly from the following statement. Note that Main Theorem 1.4 is actually a case of the following result (for 1-dimensional spaces). However, we stated that special case separately because we prove it first, and from it deduce the more general case (see Section 6.3).

Theorem 1.6 (Correlated agreement over affine spaces). *Let $V, q, n, k, \mathcal{D}$ and ρ be as defined in Theorem 1.2. For $u_0, u_1, \ldots u_l \in \mathbb{F}_q^{\mathcal{D}}$ let $U = u_0 + \mathrm{span}\{u_1, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ be an affine subspace. If $\delta \in (0, 1 - \sqrt{\rho})$ and*

$$\Pr_{u \in U} [\Delta(u, V) \leq \delta] > \epsilon,$$

where ϵ is as defined in Theorem 1.2, then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, \ldots, v_l \in V$ satisfying

- **Density:** *$|\mathcal{D}'|/|\mathcal{D}| \geq 1 - \delta$, and*
- **Agreement:** *for all $i \in \{0, \ldots, l\}$, the functions u_i and v_i agree on all of $\mathcal{D}'$.*

Furthermore, in the unique decoding regime $\delta \in \left(0, \frac{1-\rho}{2}\right]$, there exists a unique maximal *$\mathcal{D}'$ satisfying the above, with unique v_i.*

Correlated agreement (Theorem 1.6) is a sufficient condition for proximity gaps with the same error and proximity parameters (Theorem 1.2). We leave as open problems (i) whether correlated agreement is also a necessary condition for a proximity gap. And, if the answer to this question is negative, an intriguing possibility arises: (ii) obtaining proximity gaps for $\delta > 1 - \sqrt{\rho}$ while bypassing the correlated agreement approach we took here.

Organization of the rest of the paper: We start with an overview of the proof of Main Theorem 1.4 in Section 2. In Section 3 we survey several applications of our results. Section 4 gives the (simpler) proof of the unique decoding radius part of Main Theorem 1.4. Section 5 gives the proof of the (harder) list decoding radius part of that theorem, by reducing it to a different, more parameterized format (Appendix A provides the preliminary algebraic setup for the proof). In Section 6 we prove the generalizations of Main Theorem 1.4 to curves (Theorem 1.5) and higher dimensional affine spaces (Theorem 1.6). In Section 7 we state and prove weighted versions of our theorems. Section 8 concludes with more details on selected applications — verifiable secret sharing (VSS) and Fast RS IOPs of Proximity (FRI).

Acknowledgments. We thank Venkatesan Guruswami and Amnon Ta-Shma for carefully auditing this paper and suggesting valuable improvements to it, and the Ethereum Foundation for funding their audit.

We thank Ronald Cramer and Jade Nardi for pointing out a flaw in the cited version of the Polishchuk–Spielman lemma and suggesting possible fixes.

2 PROOF OVERVIEW

In this section, we give an overview of our proof strategy of our main result, Theorem 1.4.

Recall the setup. $V = \mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k]$ is the Reed-Solomon code of of degree $\le k$ polynomials evaluated at the points of $\mathcal{D} \subseteq \mathbb{F}_q$, where $|\mathcal{D}| = n$. We have functions $u_0, u_1 : \mathcal{D} \to \mathbb{F}_q$ such that for many $z \in \mathbb{F}_q$, the function $u_0 + zu_1$ is δ-close to V. We want to deduce that u_0 and u_1 are themselves close to V.

The main conceptual idea of our analysis is to work with the function field $\mathbb{K} = \mathbb{F}_q(Z)$ with a formal variable Z, and to study the various received words $u_0 + zu_1$ for the code V simultaneously by considering the *formal received word* $w = u_0 + Zu_1 : \mathcal{D} \to \mathbb{K}$ for the (big field) Reed–Solomon code $\mathsf{RS}[\mathbb{K}, \mathcal{D}, k]$. It turns out that showing that w is close to a (well-structured) codeword of this Reed–Solomon code is sufficient to show that u_0 and u_1 are both close to the original Reed–Solomon code V. With this viewpoint, our proof strategy is to *run a decoding algorithm for Reed–Solomon codes* on this received word $w = u_0 + Zu_1$. Our goal is to analyze the execution of this algorithm to show that it succeeds in finding a nearby Reed–Solomon codeword. We do such an analysis by relating it to the execution of that decoding algorithm on the various received words $u_0 + zu_1$ for the Reed–Solomon code V over the small field $\mathbb{F}_q$.

This strategy is instantiated with two different decoding algorithms for Reed–Solomon codes: the Berlekamp–Welch unique decoding algorithm, and the Guruswami–Sudan list decoding algorithm [22]. Both instantiations give rise to intriguing algebraic questions about polynomials, which we resolve using nontrivial tools from algebraic geometry and the theory of algebraic function fields.

Instantiation with the Berlekamp–Welch Algorithm. Over a field $\mathbb{F}$ and an evaluation domain $\mathcal{D}$, given a received word $r : \mathcal{D} \to \mathbb{F}$, the Berlekamp–Welch decoding algorithm for finding the (unique) nearby polynomial $P(X) \in \mathbb{F}[X]$ close to r works as follows. First it searches for low-degree polynomials $A(X), B(X) \in \mathbb{F}[X]$ such that for each $x \in \mathcal{D}$:

$$A(x)r(x) = B(x).$$

Then the nearby polynomial $P(X)$ is recovered as $B(X)/A(X)$ (which a priori may be a rational function).

In our setting, we first run the Berlekamp–Welch algorithm with received word $w = u_0 + Zu_1 : \mathcal{D} \to \mathbb{K}$ over the big field $\mathbb{K} = \mathbb{F}_q(Z)$ (we will sometimes view this as a function $w(x, z)$ with $w : \mathcal{D} \times \mathbb{F}_q \to \mathbb{F}_q$). Our goal is to find a nearby Reed–Solomon codeword (low-degree polynomial) $P(X) \in \mathbb{K}[X]$ which has the special form $P_0(X) + ZP_1(X)$, where each $P_i(X) \in \mathbb{F}_q[X]$. The first step of the Berlekamp–Welch algorithm gives us $A(X), B(X) \in \mathbb{K}[X] = \mathbb{F}_q(Z)[X]$. Making the Z dependence explicit, we write these as $A(X, Z), B(X, Z)$. This gives us a candidate, namely $A(X, Z)/B(X, Z)$, for being a Reed–Solomon codeword close to w. We will show two things: that $A(X, Z)/B(X, Z)$ is a polynomial in $\mathbb{F}_q(Z)[X]$ (a priori it is only a rational function), and that it is close to w.

The crucial step is to substitute $Z = z$ into $A(X, Z)$ and $B(X, Z)$ for various values of $z \in \mathbb{F}_q$. Letting $w_z = u_0 + zu_1 : \mathcal{D} \to \mathbb{K}$ (the result of substituting $Z = z$ into w), it turns out that $A(X, z), B(X, z) \in \mathbb{F}_q[X]$ are what we would get if we run the Berlekamp–Welch algorithm (over the small field $\mathbb{F}_q$) on received word w_z. In particular, for many z we get that $B(X, z)$ is divisible by $A(X, z)$ in $\mathbb{F}_q[X]$, and $B(X, z)/A(X, z)$ equals the Reed–Solomon codeword close to w_z. This then allows us to use the Polishchuk–Spielman lemma (a strengthening of the classical Bezout theorem, which deduces divisibility of bivariate polynomials from divisibility of univariate restrictions) to conclude that $B(X, Z)/A(X, Z)$ is in fact a polynomial $P(X, Z)$ in $\mathbb{K}[X]$ of low degree in X. The final step is to show that $P(X, Z)$, when viewed as a function from $\mathcal{D}$ to $\mathbb{K}$, is close to w, and that $P(X, Z)$ is of the form

$P_0(X) + ZP_1(X)$. This is again achieved by considering Z substitutions. We know that for many z, $P(X, z)$ is the degree at most k polynomial $P_z(X)$ that is close to w_z. This means that the X-degree of $P(X, z)$ is at most k, and that for many $x \in \mathcal{D}$ and z there is agreement between $P(x, z)$ and $w_z(x) = w(x, z)$. On the other hand, for any $x \in \mathcal{D}$, $w(x, \cdot)$ is a linear function, and $P(x, \cdot)$ is a low degree rational function, and so they cannot agree on too many points unless the low degree rational function $P(x, \cdot)$ formally equals the linear function $w(x, \cdot)$. Therefore this formal equality must happen for many $x \in \mathcal{D}$, i.e., $P(\cdot, Z)$ is close to w. Finally, by simple linear algebra, if $P(x, Z)$ is linear in Z for many x, we conclude that $P(X, Z)$ is linear in Z. This gives us our desired conclusion.

Instantiation with the Guruswami–Sudan Algorithm. Over a field $\mathbb{F}$ and an evaluation domain $\mathcal{D}$, given a received word $r : \mathcal{D} \to \mathbb{F}$, the Sudan and Guruswami–Sudan decoding algorithms for finding all nearby polynomials $P(X) \in \mathbb{F}[X]$ close to r work as follows. First one searches for a low-degree polynomial $Q(X, Y) \in \mathbb{F}[X, Y]$ such that for each $x \in \mathcal{D}$,

$$Q(x, r(x)) = 0.$$

(This is the Sudan algorithm; for the Guruswami–Sudan algorithm we ask that Q vanishes at each $(x, r(x))$ with high multiplicity.) Then every nearby polynomial $P(X)$ turns out to have the property that $Y - P(X)$ divides $Q(X, Y)$ in the bivariate polynomial ring $\mathbb{F}_q[X, Y]$. This means that all such $P(X)$ can be found by factoring $Q(X, Y)$.

In our setting, we run the Guruswami–Sudan algorithm with received word $w = u_0 + Zu_1 : \mathcal{D} \to \mathbb{K}$ over the big field $\mathbb{K} = \mathbb{F}_q(Z)$. Our goal is to find a nearby low-degree polynomial $P(X) \in \mathbb{K}[X]$ which has the special form $P_0(X) + ZP_1(X)$, where each $P_i(X) \in \mathbb{F}_q[X]$. The first step of the Guruswami–Sudan algorithm gives us a bivariate polynomial $Q(X, Y) \in \mathbb{K}[X, Y]$ such that $Q(x, w(x)) = 0$ for each $x \in \mathcal{D}$. Again, we write $Q(X, Y)$ as $Q(X, Y, Z) \in \mathbb{F}_q(Z)[X, Y]$ to make the Z dependence explicit (and we can clear denominators in Z without affecting the vanishing property).

Substituting $Z = z$, we get that $Q(x, w_z(x), z) = 0$ for each $x \in \mathcal{D}$. This means that the polynomial $Q_z(X, Y) \in \mathbb{F}_q[X, Y]$ given by $Q_z(X, Y) = Q(X, Y, z) \in \mathbb{F}_q[X, Y]$ is the bivariate polynomial we would have found while running the Guruswami–Sudan algorithm with received word $w_z : \mathcal{D} \to \mathbb{F}_q$ over the small field $\mathbb{F}_q$. Since for many $z \in \mathbb{F}_q$ we have that w_z is close to some codeword $P_z(X) \in \mathbb{F}_q[X]$ of the Reed–Solomon code V, we get that $Y - P_z(X)$ divides $Q(X, Y, z)$ for many $z \in \mathbb{F}_q$. We would like to deduce from this that over the big field $\mathbb{K}$ there is a low-degree polynomial $P(X) \in \mathbb{K}[X]$ such that $Y - P(X)$ divides $Q(X, Y)$ in $\mathbb{K}[X, Y]$ (and furthermore, this $P(X)$ is close to w and has a simple Z dependence).

This is the most involved (and interesting) part of the analysis. We will factor $Q(X, Y, Z)$ completely into linear factors in Y.

$$Q(X, Y, Z) = C(X, Z)(Y - \gamma_1(X, Z))(Y - \gamma_2(X, Z)) \cdots (Y - \gamma_D(X, Z)). \tag{2.1}$$

This is natural to do, because we are searching for factors that are linear in Y. Then we substitute $Z = z$ into this, and we should see $P_z(X)$ as one of the factors.

However, getting such a factorization for $Q(X, Y, Z)$ may not be possible with polynomials $\gamma_i(X, Z)$, and we have to look (far) beyond. What kind of objects should we think of the γ_i as? After getting the $\gamma_i(X, Z)$, we would like to (a) argue about when $\gamma_i(X, Z)$ is a polynomial in X, and (b) substitute $Z = z$ into it and inspect the resulting object. To enable these, we will express $\gamma_i(X, Z)$ in the ring $R = \overline{\mathbb{K}}[[X]]$, the ring of power series in X, whose coefficients are in the algebraic closure of $\mathbb{K} = \mathbb{F}_q(Z)$. The power series in X representation allows us to see when γ_i is a polynomial in X, and the coefficients being simply algebraic functions in Z (such as $\sqrt{Z^3 + Z + 1}$) allows us to reason about substitutions $Z = z$. Having decided on R, it is a simple application of Hensel lifting (after possibly a random shift) to show that a factorization as in (2.1) is possible with the $\gamma_i \in R$.

Rather than describe what happens in full generality, we just sketch what would happen in a special case with most of the action. Suppose $\mathbb{F}_q$ is not of characteristic 2, and we have:

$$Q(X, Y, Z) = Y^2 - (Z^3 + Z + 1)(1 - ZX).$$

Going to the ring R, and letting $\alpha = \sqrt{Z^3 + Z + 1} \in \overline{\mathbb{K}}$, it turns out that $Q(X, Y, Z)$ factors as:

$$\begin{aligned} Q(X, Y, Z) &= \left(Y - \sqrt{Z^3 + Z + 1}\sqrt{1 - ZX}\right) \cdot \left(Y + \sqrt{Z^3 + Z + 1}\sqrt{1 - ZX}\right) \\ &= \left(Y - \left(\alpha - \frac{\alpha \cdot Z}{2}X - \frac{\alpha \cdot Z^2}{16}X^2 + \ldots\right)\right)\left(Y + \left(\alpha - \frac{\alpha \cdot Z}{2}X - \frac{\alpha \cdot Z^2}{16}X^2 + \ldots\right)\right) \end{aligned}$$

where we used the Taylor series expansion for $\sqrt{1 - ZX}$. Now substitute $Z = z$ for $z \in \mathbb{F}_q$. Substituting values into algebraic functions like α is a slightly delicate operation (which square root do you choose? how do you make these choices consistent for different algebraic functions?), but it can be done using basic concepts from the theory of algebraic function fields. Another tool that we need from the theory of algebraic function fields is an analogue of the degree of a polynomial, to measure complexity of algebraic functions and bound the number of their zeroes. In this sketch we avoid going into any such details.

Doing the substitution gives us:

$$\begin{aligned} Q_z(X, Y) &= Q(X, Y, z) \\ &= \left(Y - \left(\alpha(z) - \frac{\alpha(z) \cdot z}{2}X + \cdots + c_i\alpha(z)z^iX^i + \cdots\right)\right) \times \\ &\quad \left(Y + \left(\alpha(z) - \frac{\alpha(z) \cdot z}{2}X + \cdots + c_i\alpha(z)z^iX^i + \cdots\right)\right). \end{aligned}$$

By properties of the Guruswami–Sudan decoding algorithm, we know for all "good" $z \in \mathbb{F}_q$ where w_z is close to some low degree polynomial P_z, we must have that $Y - P_z(X)$ divides $Q_z(X, Y)$. Given the factorization above, one of the following must occur:

(1) $P_z(X) = \left(\alpha(z) - \frac{\alpha(z)\cdot z}{2}X - \frac{\alpha(z)\cdot z^2}{16}X^2 + \cdots + c_i\alpha(z)z^iX^i + \cdots\right)$,

(2) $P_z(X) = -\left(\alpha(z) - \frac{\alpha(z)\cdot z}{2}X - \frac{\alpha(z)\cdot z^2}{16}X^2 + \cdots + c_i\alpha(z)z^iX^i + \cdots\right)$.

Whichever power series ends up equaling $P_z(X)$, the coefficient of X^{k+1} in that power series must equal 0. In our particular example, we deduce that $c_{k+1}\alpha(z)z^{k+1} = 0$ for some constant c_{k+1}. Assuming c_{k+1} is nonzero in $\mathbb{F}_q$, we get that $\alpha(z)z^{k+1} = 0$ for every good z. Finally we use the fact that a *nonzero* algebraic functions of low "degree" like $\alpha(Z)Z^{k+1} = \sqrt{Z^3 + Z + 1} \cdot Z^{k+1}$ cannot vanish at too many points z. This means that there cannot be too many good z, contradicting our hypothesis. We conclude that $Q(X, Y, Z)$ cannot equal $Y^2 - (Z^3 + Z + 1)(1 - ZX)$!

A very similar argument derives a contradiction unless $Q(X, Y, Z)$ has a factor of the form $Y - P(X)$ for some $P(X) \in \overline{\mathbb{K}}[X]$ of degree at most k. The only twist is that we may have to focus on the coefficient of some different power X^{k+c} in the power series than the coefficient of X^{k+1} (in case the coefficient of X^{k+1} in the power series is identically 0). To make this argument work, we need to estimate the "degree" of the algebraic functions that appear as coefficients in these power series. This involves a careful study of the Hensel lifting process, especially its effect on the complexity of its coefficients.

The final part of the argument, showing that some $Y - P(X)$ factor of $Q(X, Y, Z)$ is such that $P(X)$ has high agreement with w and all the coefficients of $P(X)$ are linear polynomials in Z, is similar to what happened in the unique decoding case. Instead of using the fact that a low degree rational function and a linear function cannot have high agreement unless they are equal, we use the fact that a low degree algebraic function and a linear function cannot have high agreement unless they are equal. This completes our sketch of the proof.

Technical issues. When we actually implement the argument, there are some technical changes we make (both for simplicity and for optimizing parameters). First, we do not do the proof by contradiction, but instead show how to find the factor of the form $Y - P(X)$. Next, instead of directly doing Hensel lifting with Q, we factor Q into irreducible factors over $\mathbb{F}_q[X, Y, Z]$ and focus on a single irreducible factor that is "responsible" for many of the P_z. This helps in that we do not need to factor arbitrarily messy Q's completely into linear factors, but only those which have the property that $Q(X, Y, z)$ has a linear factor of the form $Y - P_z(X)$. Finally, instead of arguing over the algebraic closure $\overline{\mathbb{K}}$, we go to a small algebraic extension $\mathbb{L}$ of $\mathbb{K}$ which is rich enough to express all the coefficients of the relevant power series. These changes lead to some simplifications and quantitative improvements in our proofs.

Relationship with the Arora-Sudan low degree test [2]. A beautiful and fundamental paper of Arora and Sudan [2], analyzed the "line vs. line" low degree test for multivariate polynomials in the high error regime. The heart of their paper is a theorem that says that if a function $f : \mathbb{F}_q^2 \to \mathbb{F}_q$ is such that for most lines L given by $Y = aX + b$ in $\mathbb{F}_q^2$ the univariate function obtained from restricting f to L (denoted $f|_L$) is close to a low degree univariate polynomial, then f is itself close to a low degree bivariate polynomial. This is closely related to our theorem which deduces a similar conclusion about a received word $w : \mathcal{D} \times \mathbb{F}_q \to \mathbb{F}_q$, also based on restrictions to lines. Our proof is heavily influenced by the proof in [2] (which in turn builds on fundamental results on polynomial factorization and the Hilbert irreducibility theorem by Kaltofen [28, 29]). There is one crucial difference in our proof. Our approach is spearheaded by the idea of running all arguments over the big field $\mathbb{K} = \mathbb{F}_q(Z)$ (as opposed to treating Z as another variable over $\mathbb{F}_q$ just like X and Y, as is done in [2]). This difference affects our proofs in a tangible sense: our proofs are based on bivariate interpolation over the big field $\mathbb{K}$ rather than trivariate interpolation over the small field $\mathbb{F}_q$. Inside the analysis, our proofs use power series in one variable over function fields rather than power series in two variables over finite extensions of $\mathbb{F}_q$. This leads to more involved algebraic tools being needed for our proof (most seriously the use of algebraic function fields), but also yields three improvements. First, our result is about axis parallel restrictions $Z = z$ (for $z \in \mathbb{F}_q$) instead of more general linear restrictions $Z = aX + b$ (for $a, b \in \mathbb{F}_q$). This simpler form of restriction is important for our applications. Second, our result deduces structure all the way up to the Johnson radius, while the result in Arora-Sudan is to a smaller radius (polynomially worse in terms of agreement parameter). Third, our result works over fields that are quadratic in the degree of the polynomials involved whereas the Arora-Sudan result requires fields that are quartic (at least) in the degree.

3 APPLICATIONS

Our proximity gap results are motivated by the following general setting. There are several purported codewords $\mathbf{u} = \{u_1, \ldots, u_l\} \subset \mathbb{F}_q^n$ of an RS code V. A verifier would like to be assured that they are all close to V. This is done by taking a random linear combination of the u_i and checking its proximity to V. The analysis of this simple test, which is useful in a variety of application scenarios, turns out to be surprisingly challenging. Indeed, it is closely related to the proximity gap problem we study in this work.

This batch verification problem arises in two kinds of settings: a *distributed* setting, where entries of $\mathbf{u}$ are split between multiple parties ("servers") and may not be known to any single entity, and a *centralized* setting, where $\mathbf{u}$ is entirely known to a prover and can be queried by a verifier. We briefly explain the role of proximity gaps in these two types of applications.

In the distributed setting, the coefficients of the random linear combination is either generated by a single verifier or jointly via a distributed coin tossing protocol. Each server then responds with its own share of the output. Verification succeeds if the joint output is a codeword, or alternatively it is close to the code. Examples for applications in the distributed setting include verifiable secret sharing (see Section 8.1) and secure multiparty computation protocols, such as those from [19, 27]. These applications typically rely on unique decoding and can

thus benefit from our near-optimal analysis for this regime. In this type of applications, the main challenge is protecting against an *adaptive* adversary who may choose which servers to corrupt after seeing the coefficients of the random linear combination. To defeat such an adversary, we need to ensure that if at least one of the u_i is far from the code, then (with high probability) so is their random linear combination. If this were not the case, an adaptive adversary could eliminate all inconsistencies by corrupting a small set of servers whose identity is determined by the linear combination. Proximity gaps rule out this kind of attack.

In the centralized setting, $\mathbf{u}$ is known to a prover and can be queried by the verifier. A typical realization is using a tree-based succinct cryptographic commitment that binds the prover to a uniquely defined $\mathbf{u}$ and yet enables efficient local opening of symbols queried by the verifier. In this case, the verifier challenges the prover by choosing the coefficients r_i of the random linear combination. The prover, who claims that all u_i are codewords in V, must respond with a valid codeword $u \in V$. The verifier checks that u agrees with $u' = r_1u_1 + \ldots + r_lu_l$ by querying a random entry of u and the corresponding entries of $\mathbf{u}$ and checking their consistency. (To amplify soundness, the verifier can query several random entries of u.) Here too, proximity gaps guarantee that if one of the u_i is far from V, then (with high probability) so is u'. This ensures that the verifier detects an inconsistency with high probability. Examples for applications in the centralized setting include communication-efficient proof systems [1, 5, 35], homomorphic commitment schemes [14], and secure two-party computation protocols [23, 26]. See more in Section 3.2 below.

An appealing feature of the simple "random linear combination" test is that it can be implemented with low communication and computation costs. In particular, in the distributed setting it suffices for each server to send a single field element to the verifier. In both settings, communicating the l random coefficients r_i is typically not a bottleneck. This random challenge can be made shorter either by using a cryptographic pseudorandom generator or unconditionally by using simple derandomization techniques. In particular, one can generate all coefficients as distinct powers of a single random field elements and appeal to the parameterized curves variant of the proximity gap theorem (Theorem 1.5).

Our new proximity gaps imply a tighter analysis of applications that test proximity to RS codes. Generally speaking, in the distributed setting the improved proximity gap bounds imply a constant-factor improvement in the *resilience threshold*, namely the number of corrupted parties that can be tolerated. In the centralized setting, one typically gets constant-factor savings in the overall communication and computation costs. While often ignored in theory-oriented research, the latter kind of improvements can be very significant in the context of practical succinct proof systems.

Why RS codes? Reed–Solomon codes are commonly used in distributed storage, efficient proof systems, and cryptographic protocols. They are useful because of their MDS property, near-linear encoding, and efficient (list)-decoding algorithms. A more qualitative feature of RS codes, which is commonly used in proof systems and cryptography, is the following *multiplication-friendliness* property: when $n = |\mathcal{D}| > 2k$, the pointwise products of codewords in $V = \mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k]$ span a linear code that has nontrivial minimal distance, namely the code $\mathsf{RS}[\mathbb{F}_q, \mathcal{D}, 2k]$.

We now give more concrete examples of applying proximity gaps to analyze batch-verification tasks that arise in different application scenarios.

3.1 Distributed storage and cryptography

Distributed storage. Consider a scenario in which l users encode their inputs using a length-n RS code $V = \mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k]$, where server i stores the i-th symbol of each of the l codewords. Suppose that some of the nl symbols were corrupted, say by a transient malware that overwrites a subset of the symbols before being discovered and eliminated. A verifier would like to get a quick estimate of the amount of damage caused by the malware. A natural idea is to have the servers communicate a random linear combination u' of the potentially corrupted

codewords u_j. Using the basic proximity gap result (Theorem 1.2), if at least one of u_j is δ-far from the code (for $\delta \leq \frac{1-\rho}{2}$ or $\delta < 1 - \sqrt{\rho}$), then u' is δ-far from the code except with small failure probability (at most n/q for $\delta \leq \frac{1-\rho}{2}$). Thus, for sufficiently large $\mathbb{F}_q$, the distance of u' from V provides a reliable upper bound on the maximal relative distance of a vector u_i from V within the proximity bounds of Theorem 1.2. This estimate is not too pessimistic in the sense that if only a μ-fraction of the servers were affected, the upper bound obtained by the test is no bigger than μ.

Distributed proximity test for Interleaved RS codes. The above analysis leaves something to be desired: if u' is within (sufficiently small) distance δ from V, the verifier is only assured that each u_j is *individually* within distance δ from V. In some applications, we would like to get the stronger guarantee that in such an event there is a δ-fraction of the coordinates whose removal makes *all* u_j consistent with V. Moreover, we would like to identify this set of coordinates, which is uniquely defined in the unique decoding regime. This is useful even in the above distributed storage scenario, but will be even more useful for the applications we discuss next. The stronger feature can be conveniently captured using the notion of an *Interleaved Reed–Solomon* (IRS) code. In an $\mathrm{IRS}(V, l)$ code, the codewords are $l \times n$ matrices in which each row is a codeword in V. The symbols of such a codeword are the matrix columns. Namely, a codeword consists of n symbols in $\mathbb{F}_q^\ell$. The following theorem, which follows easily from Theorem 1.6, phrases the stronger guarantee provided by the refined analysis in terms of proximity testing for IRS codes. We state it for the unique decoding regime, which suffices (and is sometimes required) for the applications we discuss next. For u within the unique decoding radius of V, we denote by $\Gamma(u, V)$ the set of coordinates on which u disagrees with the closest codeword from V.

Theorem 3.1 (Distributed proximity test for Interleaved RS codes). *Let $V = \mathrm{RS}[\mathbb{F}_q, \mathcal{D}, k]$ for $|\mathcal{D}| = n$ and $\mathbf{V} = \mathrm{IRS}(V, l)$. We view codewords in V and $\mathbf{V}$ as vectors in $\mathbb{F}_q^n$ and matrices in $\mathbb{F}_q^{l\times n}$ respectively. Let $\rho = \frac{k+1}{n}$ and $\delta \leq \frac{1-\rho}{2}$. Let $\mathbf{u} \in \mathbb{F}_q^{l\times n}$ and let $u' = r^T\mathbf{u}$ where $r \in_R \mathbb{F}_q^l$.*

- *Completeness: If $\Delta(\mathbf{u}, \mathbf{V}) \leq \delta$ then $\Pr[\Delta(u', V) \leq \delta] = 1$ and moreover $\Pr[\Gamma(u', V) \neq \Gamma(\mathbf{u}, \mathbf{V})] \leq n/q$.*
- *Soundness: If $\Delta(\mathbf{u}, \mathbf{V}) > \delta$ then $\Pr[\Delta(u', V) \leq \delta] \leq n/q$.*

We refer to the above test as distributed because it can be implemented with low communication complexity in the distributed setting, where each server holds a different column of $\mathbf{u}$. One can similarly obtain an affine version with the same guarantee, where $\mathbf{u}$ has an additional row u_0 that is always added to u' (i.e., with coefficient $r_0 = 1$), and the code $\mathbf{V}$ is extended to $\mathrm{IRS}(V, l + 1)$. This affine version is useful for zero-knowledge variants of the test, where a single random $u_0 \in V$ is used for blinding $u_1, \ldots, u_l$. This is used in the cryptographic applications we discuss next.

General cryptographic protocols. Theorem 3.1 serves as a useful tool for analyzing cryptographic protocols in the presence of an *adaptive* adversary who can dynamically choose the set of corrupted parties. For instance, it shows that secure multiparty computation protocols from [19, 27] are adaptively secure when the adversary can corrupt roughly 1/3 of the parties. The best previous proximity gaps from [11, 12, 34] could only get up to 1/4 corruption threshold in the same setting. Adaptive security, in turn, is crucial for the general transformation from [25, 26] of these honest-majority protocols to two-party protocols and protocols for dishonest majority. Indeed, this is the context that gave rise to proximity gap in the analysis of the Ligero zero-knowledge proof system [1], which applies a variant of the transformation from [25] to a variant of the protocol from [19]. We give a detailed exposition of the application of proximity gaps to *verifiable secret sharing*, which serves as a basis for the above results on secure multiparty computation, in Section 8.

3.2 Soundness of the Fast RS IOPP (FRI) protocol

FRI is an Interactive Oracle Proof of Proximity (IOP of Proximity, or IOPP) as defined in [10, 33]. An IOP is an interactive protocol in which the verifier has oracle access to messages sent by the prover, so she need not read and store those messages but may query random entries of them. FRI is one of a family of protocols for testing proximity to the RS code (an "RS proximity testing" (RPT) protocol). Its purpose is to check whether a received word $f : \mathcal{D} \to \mathbb{F}_q$ belongs to a pre-specified RS code $V := \mathrm{RS}[\mathbb{F}_q, \mathcal{D}, k]$ and to reject words that are δ-far from the code with high probability and low query complexity. Due to its efficiency it is used as a building block in several recent succinct zero knowledge protocols including scalable and transparent (public coins) arguments of knowledge (STARKs) [6, 7], Aurora [9] and its succinct version [8], and Fractal [17], to name a few. These systems have been shown by Chiesa et al. to be sound in the quantum random oracle model (hence are "plausibly post-quantum secure") [16]. Therefore, understanding the concrete soundness error of FRI, denoted ϵ_{FRI}, is of significant practical value, in addition to being a theoretically interesting question.

Consider the case of f that is maximally far from V, i.e., $\Delta(f, V) \approx 1 - \rho$ (this holds, e.g., for random f, with high probability). Fix a target soundness error bound $2^{-\lambda}$ (in concrete settings, λ is the "security parameter", often fixed to $\lambda = 128$). The communication complexity of FRI is dominated by the number t of iterations of the QUERY phase, so the question at hand is:

How many iterations t of the QUERY phase are needed to obtain $\epsilon_{\mathsf{FRI}} \leq 2^{-\lambda}$?

The initial analysis of [5] required a number t that is quite large, and does not tend to 0 even for tiny rates ρ. This was improved by [12] to $t \approx 4\lambda/\log\frac{1}{\rho}$, and then by [11] to $t \approx 3\lambda/\log\frac{1}{\rho}$. Sadly, that paper also showed that this bound is tight, at least when the field size q equals the code's blocklength n. Our main result regarding FRI (Theorem 8.3) shows that for $q \gg n^2$ we can reduce the number t of iterations by 33% to $t \approx 2\lambda/\log\frac{1}{\rho}$, which leads to communication complexity that is at least 33% shorter, for provable soundness settings. The actual savings in the provable soundness case are likely larger, due to smaller field size and the ability of the improved analysis to operate with any sequence of oracle sizes in the FRI COMMIT phase (as discussed after the statement of Theorem 8.3).

4 CORRELATED AGREEMENT OVER LINES — UNIQUE DECODING RADIUS

In this section we prove the correlated agreement result for proximity parameters that are below the unique decoding radius, corresponding to the $\epsilon = \epsilon_{\mathsf{U}}$ part of Theorem 1.4. In this case, where $\delta \in \left(0, \frac{1-\rho}{2}\right]$, our result holds even with fields that are merely *linear* in the blocklength of the code. More importantly, the proof will present several ideas, in simplified form, that will appear again in the proof of harder, list decoding regime, result (Theorem 5.1).

As usual, let $\mathbb{F}_q$ be the finite field of size q, let $\mathcal{D} \subseteq \mathbb{F}_q$ be an evaluation domain of size $|\mathcal{D}| = n$, let $k \leq n$, and let $V = \mathrm{RS}[\mathbb{F}_q, \mathcal{D}, k]$ be the Reed–Solomon code of rate $\rho = \frac{k+1}{n}$.

Theorem 4.1. *Suppose $\delta \leq (1-\rho)/2$. Let $u_0, u_1 : \mathcal{D} \to \mathbb{F}_q$ be functions. Let*

$$S = \{z \in \mathbb{F}_q : \Delta(u_0 + zu_1, V) \leq \delta\}$$

and suppose $|S| > n$. Then $S = \mathbb{F}_q$. Furthermore there are $v_0, v_1 \in V$ such that for all $z \in \mathbb{F}_q$,

$$\Delta(u_0 + zu_1, v_0 + zv_1) \leq \delta$$

and in fact

$$|\{x \in \mathcal{D} : (u_0(x), u_1(x)) \neq (v_0(x), v_1(x))\}| \leq \delta|\mathcal{D}|.$$

Remark 4.1. Since $\delta \leq \frac{1-\rho}{2}$ is within the unique decoding regime, the above v_0, v_1, which are simultaneously and separately δ-close to u_0, u_1, are also unique.

4.1 The Berlekamp–Welch decoder

Our proof will be based on the Berlekamp–Welch decoding algorithm. Let $\mathbb{F}$ be a (general) field and $\mathcal{D} \subseteq \mathbb{F}$. For an integer k, consider the Reed–Solomon code $V = \mathrm{RS}[\mathbb{F}, \mathcal{D}, k]$. We will be instantiating the Berlekamp–Welch decoder for RS codes over two different fields: the "standard" field $\mathbb{F}_q$ and the field of rational functions $\mathbb{K} = \mathbb{F}_q(Z)$ in the formal variable Z. We now give a quick description of the Berlekamp–Welch decoding algorithm and some useful aspects of it.

Given a received word $w : \mathcal{D} \to \mathbb{F}$, where $\mathcal{D} \subseteq \mathbb{F}$, and an error parameter $e = \lfloor \delta n \rfloor \leq \frac{n-k-1}{2}$, the Berlekamp–Welch decoder finds the unique (if any) polynomial $P(X) \in \mathbb{F}[X]$ such that $\Delta(w, P) \leq e$.

The first step of the Berlekamp–Welch decoder is to set up a homogeneous system of linear equations to find polynomials $A(X), B(X) \in \mathbb{F}[X]$ with $\deg(A) \leq e$, $\deg(B) \leq k + e$ such that:

$$A(x)w(x) = B(x)$$

for all $x \in \mathcal{D}$.

Lemma 4.2. *The homogeneous system of linear equations above has the following properties:*

(1) *Suppose $\Delta(w, V) \leq \delta$. Then the system of equations has a nonzero solution.*
(2) *Suppose $\Delta(w, V) \leq \delta$. Then for any nonzero solution $A(X), B(X)$ to the system of equations, we have that $A(X)$ divides $B(X)$ (in $\mathbb{F}[X]$), and the polynomial $B(X)/A(X)$ belongs to V and is at distance at most δ from w.*
(3) *If $A(X), B(X)$ is a nonzero solution to the system of equations such that $A(X)$ divides $B(X)$ (in $\mathbb{F}[X]$), then the polynomial $P(X) = B(X)/A(X)$ has the property[3] that $\Delta(w, P) \leq \delta$.*

These properties are well known; see [21] for a proof.

Note that in our setting we have $k + 2e < n$. This may seem to be off if one is used to seeing the condition $k + 2e \leq n$. The difference is that we use k here to denote the degree of the polynomials rather than the dimension of the code, and they are indeed off by 1 from each other.

After writing down the system of linear equations and establishing it has a solution, we will make use of the following linear algebra lemma to obtain bounds on non-trivial solutions:

Lemma 4.3. *Consider a homogenous linear system of m equations in n variables, given by a matrix $A = (a_{ij})_{i \leq m, j \leq n} \in M_{m,n}(\mathbb{F})$, with rank $r < n$. Then there exists a non-trivial solution to the system $(x_1, \ldots, x_n) \in \mathbb{F}^n$, where for each $1 \leq j \leq n$, either $x_j = 0$ or $x_j = \pm \det A_{I,J}$, where $A_{I,J}$ is the I, J minor of A, $I \subseteq \{1, \ldots, m\}$, $J \subset \{1, \ldots, n\} \setminus \{j\}$ and $|I| = |J| = r$.*

Proof. Without loss of generality, we may assume that $m = r$: otherwise, identify r linearly independent rows of A and remove all the remaining rows. This does not affect the set of solutions.

Given the $m \times n$ matrix A with rank $r = m < n$, identify r linearly independent columns of A. Let B be the $r \times r$ submatrix of A corresponding to these r columns, and let C be any $r \times (r + 1)$ submatrix of A which contains B as a submatrix. Cramer's rule gives us a linear combination of the columns of C which equals 0, with coefficients being the determinants of the $r \times r$ minors of C. This yields the desired solution x. Note that x is indeed non-trivial, since one of its elements is the determinant of the minor B, which is non-zero by construction.

□

4.2 The Polishchuk–Spielman lemma

Another ingredient that will appear in our proof is a version of the Polishchuk–Spielman lemma [31]. The version we state below is a variation of [37, Lemma 4.2.18], with some differences in the precise conditions, and we include a derivation in Appendix D.

[3]Note that $P(X)$ may have degree larger than k. The best we can say about the degree of $P(X)$ is that it is at most $k + e$.

Lemma 4.4. *Let $\mathbb{F}$ be a field, and let $a_X, a_Z, b_X, b_Z, d_X, d_Z, n_X, n_Z$ be non-negative integers satisfying $d_X = b_X - a_X$, $d_Z = b_Z - a_Z$. Let $A(X,Z), B(X,Z) \in \mathbb{F}[X,Z]$ be polynomials with degrees bounded by*

$$\deg_X(A) \le a_X, \quad \deg_X(B) \le b_X, \quad \deg_Z(A) \le a_Z, \quad \deg_Z(B) \le b_Z.$$

Suppose that for at least n_X values of $x \in \mathbb{F}$, there exist polynomials $P_{Z,x}(Z) \in \mathbb{F}[Z]$ with $\deg_Z(P_{Z,x}) \le d_Z$ such that $B(x,Z) = P_{Z,x}(Z)A(x,Z)$, and similarly suppose that for at least n_Z values of $z \in \mathbb{F}$, there exist polynomials $P_{X,z}(X) \in \mathbb{F}[X]$ with $\deg_X(P_{X,z}) \le d_X$ such that $B(X,z) = P_{X,z}(X)A(X,z)$.

Finally, suppose

$$\frac{b_X}{n_X} + \frac{b_Z}{n_Z} < 1. \qquad (\star)$$

If all conditions hold, then $A(X,Z) \mid B(X,Z)$ as polynomials in $\mathbb{F}_q[X,Z]$. Furthermore, the quotient $P(X,Z) = \frac{B(X,Z)}{A(X,Z)}$ satisfies $\deg_X(P) \le d_X, \deg_Z(P) \le d_Z$, as well as $P(X,z) = P_{X,z}(X)$ for at least $n_Z - a_Z$ values of z, and $P(x,Z) = P_{Z,x}(Z)$ for at least $n_X - a_X$ values of x.

4.3 Proof of Theorem 4.1

By definition of S, for each $z \in S$, we have a polynomial $P_z(X) \in \mathbb{F}_q[X]$ with $\deg(P_z) \le k$ such that $\Delta(u_0 + zu_1, P_z) \le \delta$.

Our strategy is to run the Berlekamp–Welch decoder over the field $\mathbb{K} = \mathbb{F}_q(Z)$ of rational functions in the formal variable Z.

First define a received word

$$w : \mathcal{D} \to \mathbb{K}$$

given by:

$$w(x) = u_0(x) + Zu_1(x).$$

We sometimes also use the notation $w(x,Z)$ to denote $u_0(x) + Zu_1(x)$.

We will try to find a polynomial $P(X,Z) \in \mathbb{F}_q[X,Z]$ of the form $P(X,Z) = v_0(X) + Zv_1(X)$, where $\deg_X(P) \le k$, such that

$$P(x,Z) = w(x)$$

for at least $n - e$ choices of $x \in \mathcal{D}$.

4.3.1 **Step 1: Finding $A(X,Z), B(X,Z)$.** The first step of the Berlekamp–Welch algorithm is to find nonzero $A(X,Z), B(X,Z) \in \mathbb{K}[X]$ of degrees $\le e$ and $\le k + e$ (in the variable X) respectively such that

$$A(x,Z)w(x) = B(x,Z) \qquad (4.1)$$

for all $x \in \mathcal{D}$, where $e = \lfloor \delta n \rfloor$ as before. Setting this up as a homogeneous linear system over $\mathbb{K}$, we get an $n \times (k + 2e + 2)$ matrix $M(Z)$ with entries being polynomials in Z, of degree ≤ 1 for the $e + 1$ columns of the A-variables, and degree 0 for the $k+e+1$ columns of the B variables. Explicitly, writing $A(X,Z) = \sum_i A_i(Z)X^i$ and $B(X,Z) = \sum_i B_i(Z)X^i$, in the row corresponding to $x \in \mathcal{D}$, the entry of $M(Z)$ corresponding to the coefficient of $A_i(Z)$ is $u_0(x)x^i + u_1(x)x^iZ$, and the entry corresponding to the coefficient of $B_i(Z)$ is simply $-x^i$.

We now show that $M(Z)$ has rank $< k + 2e + 2$ over $\mathbb{K}$. Fix any $(k + 2e + 2) \times (k + 2e + 2)$ minor of $M(Z)$, and consider its determinant $R(Z) \in \mathbb{F}_q[Z]$. We will show that $R(Z) = 0$. This then implies that $M(Z)$ has rank $< k + 2e + 2$ over $\mathbb{K}$.

For any $z \in S$, consider $M(z)$. This is the homogeneous linear system that arises when we run the Berlekamp–Welch decoder with received word $u_0 + zu_1 \in \mathbb{F}_q^n$ over the field $\mathbb{F}_q$. By definition of S we know that $\Delta(u_0 + zu_1, V) \le \delta$, and so Item 1 of Lemma 4.2 tells us that this linear system has a nonzero solution. Therefore $M(z)$ has rank $< k + 2e + 2$. Thus for each $z \in S$, $R(z) = 0$. Now note that $\deg(R) \le e + 1$, since the $e + 1$ columns of $M(Z)$

corresponding to coefficients of A_i are linear in Z, and the rest of the columns are degree 0. On the other hand we have $|S| > n \geq e + 1$, hence we conclude that $R(Z) = 0$ formally, as desired.

We now know that the system of equations $M(Z)$ has a non-trivial solution, with $A_i(Z), B_i(Z) \in \mathbb{F}_q(Z)$. We wish to show that in particular there is a solution in which $A_i(Z), B_i(Z)$ are not just rational functions, but polynomials of bounded Z-degree. Indeed, this follows easily from Lemma 4.3, which exactly states that there is a non-trivial solution where each $A_i(Z)$ or $B_i(Z)$ is either 0 or some determinant of a submatrix of $M(Z)$. These determinants are clearly polynomials, as the entries of $M(Z)$ are polynomials. Moreover, the $e + 1$ columns of $M(Z)$ corresponding to A variables are linear in Z, and the columns corresponding to B variables are constant in Z. Thus each product in the determinants may use at most $e + 1$ degree 1 entries, and have degree at most $e + 1$; and for the determinants yielding the A variables, Lemma 4.3 also tells us that the column corresponding to the variable must be excluded, and so there will be at most e degree 1 terms in the product. Thus, our obtained solution will satisfy $\deg B_i(Z) \leq e + 1$, $\deg A_i(Z) \leq e$.

Writing now $A(X, Z) = \sum_{i=0}^{e} A_i(Z)X^i$, $B(X, Z) = \sum_{i=0}^{k+e} B_i(Z)X^i$, we find that $A(X, Z), B(X, Z) \in \mathbb{F}_q[X, Z]$ are polynomials with $\deg_Z(A) \leq e$, $\deg_Z(B) \leq e + 1$, $\deg_X(A) \leq e$, $\deg_X(B) \leq k + e$ and:

$$A(x, Z)w(x) = B(x, Z)$$

for all $x \in \mathcal{D}$. Using our alternate notation for w, we get:

$$A(x, Z)w(x, Z) = B(x, Z)$$

for all $x \in \mathcal{D}$.

4.3.2 **Step 2: Dividing $B(X, Z)$ by $A(X, Z)$ in $\mathbb{F}_q[X, Z]$.** Now if $z \in S$, we know that the function $w(\cdot, z) : \mathcal{D} \to \mathbb{F}_q$ has distance $\leq e$ from V. Consider the Berlekamp–Welch system of linear equations associated with this received word: namely, we consider the space of all pairs of polynomials $E(X), F(X)$ of degrees $\leq e, \leq k + e$ respectively, such that

$$E(x)w(x, z) = F(x)$$

for all $x \in \mathcal{D}$. We see that any solution $E(X), F(X)$ of this system falls into one of two cases:

- $E(X) = 0$, in which case $F(x) = 0$ for all $x \in \mathcal{D}$, and so $F(X) = 0$ too.
- $E(X)$ is nonzero, in which case Item 2 of Lemma 4.2 tells us that $E(X)$ divides $F(X)$, and $\frac{F(X)}{E(X)} = P_z(X)$.

In both these cases, $E(X)$ divides $F(X)$, and $F(X) = P_z(X)E(X)$. For any $z \in S$, $A(X, z), B(X, z)$ are polynomials that satisfy the properties of E, F above, thus $B(X, z) = P_z(X)A(X, z)$.

Let $a_X = a_Z = e$, $b_X = k + e$, $b_Z = e + 1$, $d_X = k$, $d_Z = 1$, $n_X = n$ and $n_Z = |S| > 1$. We have seen that $B(x, Z) = w(x, Z)A(x, Z)$ with $w(x, Z)$ of degree at most $d_Z = 1$ for $|\mathcal{D}| = n_X$ values of x, and $B(X, z) = P_z(X)A(X, z)$ with $P_z(X)$ of degree at most $d_X = k$ for $|S| = n_Z$ values of z. The X and Z degrees of A and B are bounded by a_X, a_Z, b_X, b_Z, and finally

$$\frac{b_X}{n_X} + \frac{b_Z}{n_Z} = \frac{k + e}{n} + \frac{e + 1}{|S|} < \frac{k + 2e + 1}{n} \leq 1.$$

We may therefore apply the Polishchuk–Spielman lemma (Lemma 4.4). We find that $A(X, Z)$ divides $B(X, Z)$ in $\mathbb{F}_q[X, Z]$, and the quotient $P(X, Z) = B(X, Z)/A(X, Z)$ is of degree at most $d_Z = 1$ in Z and $d_X = k$ in X, and so can be written as $v_0(X) + Zv_1(X)$, where v_0, v_1 are of degree at most k. Furthermore the lemma implies that

$$v_0(x) + Zv_1(x) = P(x, Z) = w(x, Z) = u_0(x) + Zu_1(x)$$

for at least $n_X - a_X = n - e = \lceil n(1 - \delta) \rceil$ values of $x \in \mathcal{D}$. Thus

$$\Delta(u_0 + Zu_1, v_0 + Zv_1) \leq \delta$$

as claimed. This completes the proof of Theorem 4.1. □

5 CORRELATED AGREEMENT OVER LINES — LIST DECODING RADIUS

In this section we prove the large distance part of the correlated agreement theorem (Theorem 1.4), corresponding to larger proximity parameters $\delta \in \left(\frac{1-\rho}{2}, 1-\sqrt{\rho}\right)$. First, we state the theorem in a slightly different form that will be easier to work with.

Theorem 5.1. *Let $u_0, u_1 : \mathcal{D} \to \mathbb{F}_q$, let $m \geq 3$, define*

$$\delta_0(\rho, m) := 1 - \sqrt{\rho} - \frac{\sqrt{\rho}}{2m}, \tag{5.1}$$

and let $\delta \leq \delta_0(\rho, m)$. Define

$$S = \{z \in \mathbb{F}_q : \Delta(u_0 + zu_1, V) \leq \delta\} \tag{5.2}$$

and suppose

$$|S| > \frac{(1+\frac{1}{2m})^7 m^7}{3\rho^{3/2}} n^2. \tag{5.3}$$

Then u_0, u_1 are simultaneously δ-close to V, i.e. $\exists v_0, v_1 \in V$ such that

$$|\{x \in \mathcal{D} : (u_0(x), u_1(x)) = (v_0(x), v_1(x))\}| \geq (1-\delta)|\mathcal{D}|.$$

The above version easily implies the large distance part of Theorem 1.4 with the coarser

$$\epsilon_J = O\left(\frac{1}{(\eta\rho)^{O(1)}} \cdot \frac{n^2}{q}\right)$$

bound by setting $m = O\left(\frac{\sqrt{\rho}}{\eta}\right)$. For the more precise bound on ϵ_J, we need to be a little careful, and we do this in the following theorem.

Theorem 5.2 (Correlated agreement over lines — alternative formulation). *Let $u_0, u_1 : \mathcal{D} \to \mathbb{F}_q$. Let $\delta, \eta > 0$ satisfy $\eta \leq \frac{\sqrt{\rho}}{20}$ and $\delta \leq \delta_0(\rho, \eta) := 1 - \sqrt{\rho} - \eta$, and suppose*

$$\mathbb{P}_{z \in \mathbb{F}_q}(\Delta(u_0 + zu_1, V) \leq \delta) > \frac{\rho^2 n^2}{(2\eta)^7 q} =: \epsilon_J. \tag{5.4}$$

Then u_0, u_1 are simultaneously δ-close to V, i.e. $\exists v_0, v_1 \in V$ such that

$$|\{x \in \mathcal{D} : (u_0(x), u_1(x)) = (v_0(x), v_1(x))\}| \geq (1-\delta)|\mathcal{D}|.$$

Proof of Theorem 5.2 from Theorem 5.1. Set $m = \left\lceil \frac{\sqrt{\rho}}{2\eta} \right\rceil \geq 10$, and note that $\delta \leq \delta_0(\rho, \eta) < \delta_0(\rho, m)$. Define S as in Theorem 5.1, and observe that (5.3) is satisfied:

$$|S| > \epsilon_J q = (2\eta)^{-7}\rho^2 n^2 > \left(\frac{m-1}{\sqrt{\rho}}\right)^7 \rho^2 n^2 = \left(1 - \tfrac{1}{m}\right)^7 \frac{m^7}{\rho^{3/2}} n^2 > \frac{(1+\frac{1}{2m})^7 m^7}{3\rho^{3/2}} n^2,$$

where in the last step we use $3\left(1-\frac{1}{m}\right)^7 > \left(1+\frac{1}{2m}\right)^7$, which holds for $m \geq 10$. Thus we may apply Theorem 5.1, and conclude (u_0, u_1) is δ-close to (v_0, v_1), as claimed. □

5.1 The Guruswami–Sudan decoder

Our proof will be based on the Guruswami–Sudan decoding algorithm. Let $\mathbb{F}$ be a (general) field and $\mathcal{D} \subseteq \mathbb{F}$. Let V be the Reed–Solomon code $\mathrm{RS}[\mathbb{F}, \mathcal{D}, k]$. Let $\rho = \frac{k+1}{n}$ denote its rate. We will be instantiating the Guruswami–Sudan decoder for RS codes over two different fields: the "standard" field $\mathbb{F}_q$ and the field of rational functions $\mathbb{K} = \mathbb{F}_q(Z)$ in the formal variable Z. We now give a quick description of the Guruswami–Sudan decoding algorithm and some useful aspects of it.

First, some definitions related to bivariate polynomials: The (a, b)-weighted degree of a monomial $X^i Y^j$ is $ai + bj$. The (a, b)-weighted degree of a polynomial $Q(X, Y) \in \mathbb{F}[X, Y]$ is the maximal (a, b)-weighted degree of all its non-zero monomials. The vanishing multiplicity of a polynomial $Q(X, Y) \in \mathbb{F}[X, Y]$ at a point $(x, y) \in \mathbb{F}^2$ is the smallest m such that the shifted polynomial $Q(x + X, y + Y)$, written as:

$$Q(x + X, y + Y) = \sum_{i,j} a_{ij} X^i Y^j$$

has $a_{ij} = 0$ for all (i, j) with $i + j < m$. We denote the vanishing multiplicity of Q at (x, y) by $\mathrm{mult}(Q, (x, y))$.

Given a received word $w : \mathcal{D} \to \mathbb{F}$ and a multiplicity parameter m, the Guruswami–Sudan decoder first solves a homogeneous system of linear equations to find a nonzero polynomial $Q(X, Y) \in \mathbb{F}[X, Y]$ with $(1, k)$-weighted degree less than $D_X(m)$ (for a certain function $D_X(m)$, specified later), such that:

$$\mathrm{mult}(Q, (x, w(x)) \geq m$$

for all $x \in \mathcal{D}$.

The key properties of this system of linear equations that enable decoding are given by the following lemma, proved by [22].

Lemma 5.3. *Let $\delta_0(\rho, m) = 1 - \sqrt{\rho} - \frac{\sqrt{\rho}}{2m}$. With $D_X(m) = (m + \frac{1}{2})\sqrt{\rho} n$, the system of linear equations set up above has the following properties:*

(1) *The system has a nonzero solution $Q(X, Y)$.*
(2) *For any nonzero solution $Q(X, Y)$ of the above system, and for any polynomial $P(X) \in V$ such that $\Delta(w, P) \leq \delta_0(\rho, m)$, we have that $Y - P(X)$ divides $Q(X, Y)$ in the polynomial ring $\mathbb{F}[X, Y]$.*

Note that these choices of δ_0 and D_X are not quite optimal. The optimal values are only slightly better, but their formulas are longer and messier, and we opt for simplicity in favor of optimization.

5.2 Proof of Theorem 5.1

By definition of S, for each $z \in S$, we have a polynomial $P_z(X) \in \mathbb{F}_q[X]$ with $\deg(P_z) \leq k$ such that $\Delta(u_0 + zu_1, P_z) \leq \delta$.

Our strategy is to run the Guruswami–Sudan decoder over the field $\mathbb{K} = \mathbb{F}_q(Z)$ of rational functions in the formal variable Z.

First define a received word

$$w : \mathcal{D} \to \mathbb{K}$$

given by:

$$w(x) = u_0(x) + Z u_1(x).$$

We sometimes also use the notation $w(x, Z)$ to denote $u_0(x) + Z u_1(x)$.

We will try to find a polynomial $P(X, Z) \in \mathbb{F}_q[X, Z]$ of the form $P(X, Z) = v_0(X) + Z v_1(X)$, where $\deg_X(P) \leq k$, such that

$$P(x, Z) = w(x)$$

for at least than $n - e$ choices of $x \in \mathcal{D}$, where $e = \lfloor n\delta \rfloor$ is the decoder's error parameter.

5.2.1 Step 1: Interpolating $Q(X, Y, Z)$. Let $D_X = D_X(m) = (m + \frac{1}{2})\sqrt{\rho}n$. The first step of the Guruswami–Sudan decoding algorithm is to find a nonzero polynomial $Q(X, Y) \in \mathbb{K}[X, Y]$:

$$Q(X, Y) = \sum_{i+k\cdot j<D_X} Q_{ji}(Z)X^iY^j,$$

where each $Q_{ji}(Z)$ lies in the big field $\mathbb{K}$, such that $Q(X, Y)$ has a zero of multiplicity m at $(x, w(x))$ for each $x \in \mathcal{D}$.

This is possible when the number of available monomials, which is at least $\frac{k}{2}\left(\left(\frac{D_X}{k} + \frac{1}{2}\right)^2 - \frac{1}{4}\right)$ (see Claim B.1 in Appendix B.2), exceeds the number of homogeneous linear equations $\binom{m+1}{2}n$. Indeed, this happens for our choice of D_X.

Solving this system of equations for a nonzero solution using Cramer's rule, and clearing Z denominators, we get such a $Q(X, Y) \in \mathbb{K}[X, Y]$ where each coefficient $Q_{ji}(Z)$ is in fact an element of $\mathbb{F}_q[Z]$ (i.e., a polynomial instead of just a rational function) with controlled degree.

Explicitly, we get:

Claim 5.4. *There is a nonzero polynomial $Q(X, Y) \in \mathbb{K}[X, Y]$ with $(1, k)$-weighted degree less than D_X such that for each $x \in \mathcal{D}$, we have:*

$$\mathsf{mult}(Q, (x, w(x))) \geq m,$$

and furthermore:

- $$\deg_X(Q) < D_X = (m + \tfrac{1}{2})\sqrt{\rho}n. \tag{5.5}$$
- *$D_Y \coloneqq \deg_Y(Q)$ satisfies:*
$$D_Y < \frac{D_X}{k} = \frac{m + \frac{1}{2}}{\sqrt{\rho}}. \tag{5.6}$$
- *Each coefficient $Q_{ji}(Z)$ of $Q(X, Y)$ is in $\mathbb{F}_q[Z]$.*
- *$D_{YZ} \coloneqq \deg_{Y,Z}(Q)$ (which is the total Y, Z degree of Q) satisfies:*
$$D_{YZ} \leq \frac{(m + \frac{1}{2})^3}{6\sqrt{\rho}}n. \tag{5.7}$$

Only the bound on D_{YZ} needs to be discussed, and it is explained and proven in detail in Appendix B.1.

Claim 5.4 allows us to express the lower bound on $|S|$ from (5.3) in terms of D_X, D_Y, D_{YZ}, as

$$|S| > \frac{(1 + \frac{1}{2m})^7 m^7}{3\rho^{3/2}}\, n^2 = 2\left(\frac{(m + \frac{1}{2})}{\sqrt{\rho}}\right)^3 \left((m + \tfrac{1}{2})\sqrt{\rho}\, n\right)\left(\frac{(m + \frac{1}{2})^3}{6\sqrt{\rho}}\, n\right) \geq 2D_Y^3 D_X D_{YZ}. \tag{5.8}$$

Over the next subsections, we will prove the following proposition, which is the core of our proof:

Proposition 5.5. *There exist a subset $S' \subset S$ and a polynomial $P(X, Z) \in \mathbb{F}_q[X, Z]$ with the following properties:*

$$|S'| > \frac{|S|}{2D_Y}, \tag{5.9}$$

$$\forall z \in S',\ P_z(X) = P(X, z), \tag{5.10}$$

$$\deg_X(P) \leq k,\ \deg_Z(P) \leq 1. \tag{5.11}$$

5.2.2 Step 2: $Q(X, Y, z)$ is divisible by $Y - P_z(X)$ for many z. Recall that for every $z \in S$ there exists a polynomial $P_z(X) \in \mathbb{F}_q[X]$ of degree at most k with distance at most δ from $u_0 + zu_1$, i.e. $(x, P_z(x))$ equals $(x, w_z(x))$ for at least $n(1-\delta)$ values of $x \in \mathcal{D}$. In each such point of agreement, the univariate polynomial $Q(X, P_z(X), z) \in \mathbb{F}_q[X]$ must then have a zero of order m, thus it has at least $mn(1-\delta)$ roots counted with multiplicity. On the other hand, since Q is chosen to have $(1, k)$-weighted degree less than D_X, and $\deg P_z(X) \le k$, we have $\deg_X(Q(X, P_z(X), z)) < D_X$. Thus the polynomial must be identically zero if $D_X \le mn(1-\delta)$. Indeed this holds, because

$$1 - \delta \ge 1 - \delta_0 = \left(1 + \frac{1}{2m}\right)\sqrt{\rho} = \frac{D_X}{mn}.$$

Thus $Q(X, Y, z)$ is divisible by $Y - P_z(X)$ for each $z \in S$.

5.2.3 Step 3: Finding a good x_0 to start Hensel lifting. We now begin the process of finding a power series solution $Y = \gamma(X) \in \overline{\mathbb{K}}[[X]]$ to $Q(X, Y, Z) = 0$ (thought of as a bivariate equation $Q(X, Y) = 0$ with coefficients in $\mathbb{K}$). To find the power series solution, we will start at a suitable solution (x_0, α_0) of $Q(X, Y) = 0$, and then use Hensel lifting. In this section, our goal is to find such a "suitable" (x_0, α_0).

Denote by $\mathrm{disc}_Y(\cdot)$ the Y-discriminant of a given polynomial (which may itself be a polynomial, if the original polynomial was multivariate). Similarly, denote by $\mathrm{res}_Y(\cdot, \cdot)$ the Y-resultant of two polynomials.

Considering $Q(X, Y, Z)$ as a polynomial in Y over $\mathbb{F}_q[X, Z]$, it can be uniquely factored as

$$Q(X, Y, Z) = C(X, Z) \prod_i R_i(X, Y^{p^{f_i}}, Z)^{e_i}, \tag{5.12}$$

where p is the characteristic of $\mathbb{F}_q$, $f_i \ge 0$, $e_i \ge 1$, and each $R_i(X, Y, Z)$ is irreducible and separable[4]. This factorization is a combination of two well-known facts: Unique factorization in the polynomial ring $\mathbb{F}_q[X, Z][Y]$, which gives a factorization of the form $C(X, Z) \prod_i \widetilde{R}_i(X, Y, Z)$, with $\widetilde{R}_i(X, Y, Z)$ irreducible in $\mathbb{F}_q[X, Z][Y]$ but not necessarily separable in Y; and the fact that every such irreducible $\widetilde{R}_i$ is of the form $R_i(X, Y^{p^{f_i}}, Z)$ for an appropriate f_i and separable irreducible R_i, which follows inductively from the fact that every Y-inseparable irreducible polynomial must in fact be in $\mathbb{F}_q[X, Z][Y^p]$, since its Y-derivative must vanish. For more details on the two facts, see [30, Chapter IV, Theorem 2.3] and [30, Chapter V, Proposition 6.1].

In this section we prove the following claim:

Claim 5.6. *There exists $x_0 \in \mathbb{F}_q$ such that for all i,*

$$\mathrm{disc}_Y(R_i(X, Y, Z))(x_0) \neq 0 \in \mathbb{F}_q[Z].$$

Before we prove the claim, let us explain its use and motivation. The Hensel lift (described in more detail in Appendix A.4) shows that any *simple* root[5] $Y = \alpha_0 \in \overline{\mathbb{K}}$ of $Q(x_0, Y, Z)$ can be uniquely lifted to a power series solution $\gamma(X) \in \overline{\mathbb{K}}[[X - x_0]]$ with free coefficient α_0, by iteratively finding solutions to $Q(X, Y, Z) \equiv 0 \pmod{(X - x_0)^s}$ with increasing s.

However, it may be that $Q(X, Y)$ has no simple roots. This could happen, for example, if the factors of $Q(X, Y)$ appear with multiplicity. To resolve this, we instead focus on an irreducible factor R_i of Q. Even after focusing on R_i, it might still be the case that the particular root α_0 of $R_i(x_0, Y, Z)$ which we wish to lift is non-simple. We avoid this issue by requiring that $R_i(x_0, Y, Z)$ is separable in Y (i.e., all of its roots in Y are simple). This happens if $\mathrm{disc}_Y R_i(x_0, Y, Z) \neq 0$. Claim 5.6 exactly guarantees the existence of an element x_0 such that this occurs for all possible R_i. For all future sections, we will fix any such x_0 arbitrarily.

[4] $R(X, Y, Z)$ being separable in Y means it does not have repeated roots in the variable Y, in any extension field. This is equivalent to $\mathrm{disc}_Y(R(X, Y, Z)) \neq 0$. For an irreducible polynomial in Y, it is also equivalent to the Y-derivative being not identically 0, or to the polynomial not being representable as a polynomial in Y^p.

[5] A root is simple when it has multiplicity 1.

Henceforth, we will assume for simplicity that Q does not have inseparable irreducible factors, i.e. that $f_i = 0$ for all i. Note that since any inseparable factor has Y-degree at least p, this is necessarily the case if the characteristic is larger than D_Y, for example in the case where $\mathbb{F}_q$ is a prime field with $p = q$: (5.8) together with the trivial bound $|S| \leq q$ immediately yield $D_Y < q^{1/3}$. The general case, including the possibility of inseparable factors, is very similar, but has more technicalities, which are discussed in detail in Appendix C.

Proof of Claim 5.6. Since all of the polynomials $R_i(X, Y, Z)$ are separable in Y, the discriminants $\mathrm{disc}_Y(R_i(X, Y, Z)) \in \mathbb{F}_q[X, Z]$ are non-zero polynomials. We need to find an $x_0 \in \mathbb{F}_q$ which makes all these discriminants evaluate to nonzero polynomials in $\mathbb{F}_q[Z]$. This will simply follow from a bound on the sum of degrees of all the R_i, which we would like to show is less than q. A crude bound on the sum of degrees (which would require a stronger bound on $|S|$) is easy to give. The rest of the proof is just a more careful bound.

Define

$$\mathrm{disc}^*_Y(Q) \coloneqq \prod_i \mathrm{disc}_Y(R_i(X, Y, Z)) \in \mathbb{F}_q[Z][X].$$

It suffices to show that $\deg_X \mathrm{disc}^*_Y(Q) < q$.

In order to bound the X-degree of $\mathrm{disc}^*_Y(Q)$, we will instead bound the X-degree of $\mathrm{disc}_Y(Q)$, and see that it serves as an upper bound. There is a subtle issue here, as $\mathrm{disc}_Y(Q)$ may very well be the zero polynomial, for example if any $e_i > 1$. However, note that by expressing Q as polynomial in Y, i.e. $Q = \sum_{j \leq D_Y} Q_j(X, Z)Y^j$ with $Q_j(X, Z) = \sum_{i < D_X - k \cdot j} Q_{ji}(Z)X^i$, there is a generic formula for the discriminant $\mathrm{disc}_Y(Q)$, which is a polynomial in the $Q_j(X, Z)$ coefficients, given (up to sign) by the determinant of the Sylvester matrix of Q and $\frac{\partial Q}{\partial Y}$, divided by the leading coefficient Q_{D_Y}. The generic polynomial is non-zero, and though it might vanish for the particular substitution of Q_j, we can still compute its formal X-degree as the maximal degree that would appear when expanding all algebraic expressions in the $Q_j(X, Z)$, before cancellations. Furthermore, this polynomial will be formally divisible by $\mathrm{disc}^*_Y(Q)$, in the sense that the quotient can be expressed as a generic polynomial in X, Z (which again might vanish for the particular substitution). This is due to the R_i dividing Q, so that we may write $Q(X, Y, Z) = U(X, Y, Z) \cdot \prod_i R_i(X, Y, Z)$ where U is the quotient. We then use the fact that the discriminant of a product of polynomials is given by a product of their discriminants and resultants, in this case yielding

$$\mathrm{disc}_Y(Q) = \mathrm{disc}_Y(U) \cdot \prod_i \mathrm{disc}_Y(R_i) \cdot \prod_i \mathrm{res}_Y(R_i, U)^2 \prod_{i<i'} \mathrm{res}_Y(R_i, R_{i'})^2$$

where the right hand side contains $\mathrm{disc}^*_Y(Q) = \prod_i \mathrm{disc}_Y R_i$ as well as other terms, and all discriminants and resultants can be expressed as polynomials in the Y-coefficients of Q, U and the R_i. Thus we deduce $\deg_X(\mathrm{disc}^*_Y(Q)) \leq \deg_X(\mathrm{disc}_Y(Q))$.

Finally, we wish to evaluate the formal X-degree of $\mathrm{disc}_Y(Q)$ from the Sylvester matrix and the bounds $\deg_X Q_j < D_X - k \cdot j$. The crux is to observe that every term in the determinant expansion of the Sylvester matrix has the same degree bound, and then to evaluate the degree bound for a convenient term (which is the principal diagonal in a natural ordering).

For each row in this matrix, the non-zero coefficients are simply $Q_{D_Y}, \ldots Q_0$ (or multiplied by constants, for rows of $\frac{\partial Q}{\partial Y}$). Thus we see that shifting a column to the right always increases the degree bound by the same constant k, so that in general the entry degree bound is given by the sum of a function of the row index and a linear function of the column index. Since the determinant always uses products with one item from each row and from each column, every non-zero product in the expression will have the exact same degree bound, and we may compute this bound "by example", considering any one such expansion, as claimed above. It is particularly pleasant to take the leading coefficient Q_{D_Y} from each of the D_Y copies of $\frac{\partial Q}{\partial Y}$, and the free coefficient of Q_0 from the $D_Y - 1$ copies of Q, and of course dividing once by Q_{D_Y}, as is part of the discriminant formula, giving the term $Q_{D_Y}^{D_Y - 1} Q_0^{D_Y - 1}$. We thus get

$$\begin{aligned}\deg_X(\mathrm{disc}_Y^*(Q)) &\le \deg_X(\mathrm{disc}_Y(Q)) = (D_Y-1)(\deg_X(Q_{D_Y})+\deg_X(Q_0))\\ &< (D_Y-1)(2D_X-kD_Y) < \frac{(kD_Y)(2D_X-kD_Y)}{k}\\ &< \frac{D_X^2}{k} = \frac{(m+\frac{1}{2})^2\rho\, n^2}{\rho\, n} = (m+\tfrac{1}{2})^2 n.\end{aligned}$$

Note that from (5.3), we in particular have $q \ge |S| > \frac{\left(1+\frac{1}{2m}\right)^7 m^7}{3\rho^{3/2}} n^2 > (m+\frac{1}{2})^2 n$, and we have our desired inequality. □

5.2.4 Step 4: Focusing on a useful factor $R_i(X,Y,Z)$. Our main goal in the following sections is to show that at least one of the factors $R_i(X,Y,Z)$ is of the form $Y - P(X,Z)$, with $P \in \mathbb{F}_q[X,Z]$ a polynomial of X-degree at most k and Z-degree at most 1. Note that for such a factor, we will have in particular that for any $z \in \mathbb{F}_q$, $Y - P(X,z)$ is a factor of $R_i(X,Y,z)$, and $P(x_0,z)$ is a rational root of $R_i(x_0,Y,z)$. We will see that a converse is also true: if such factors and roots of R_i exist for sufficiently many z's, then R_i must be divisible by (and thus of the form) $Y - P(X,Z)$. In this section we will make use of this fact in order to focus on a useful R_i. More precisely, we prove the following claim:

Claim 5.7. *There exists a factor $R = R_i$ of Q, and an irreducible factor $H(Y,Z)$ of $R(x_0,Y,Z)$, such that the set $S_{x_0,R,H}$ of z values for which both R and H vanish at P_z, i.e.*

$$S_{x_0,R,H} = \{z \in S : R(x_0, P_z(x_0), z) \equiv 0 \text{ and } H(P_z(x_0), z) = 0\},$$

is sufficiently large; more precisely, such that

$$|S_{x_0,R,H}| \ge \frac{|S|}{D_Y} > 2D_Y^2 D_X D_{YZ}. \tag{5.13}$$

Proof. After substituting $X \mapsto x_0$, each of the irreducible $R_i(X,Y,Z)$ can be factored as

$$R_i(x_0,Y,Z) = C_i(Z)\prod_j H_{ij}(Y,Z),$$

where $H_{ij} \in \mathbb{F}_q[Z][Y]$ are irreducible, separable in Y, and with positive Y-degree. For Q, this yields the factorization

$$Q(x_0,Y,Z) = \left(C(x_0,Z)\prod_i C_i(Z)\right)\prod_{i,j} H_{ij}(Y,Z).$$

In particular, the number of H_{ij} is at most D_Y.

For any $z \in S$, the polynomial $P_z(X)$ satisfies $Q(X,P_z(X),z) = 0$, i.e. $Y - P_z(X) \mid Q(X,Y,z)$, thus there is some i such that $Y - P_z(X) \mid R_i(X,Y,z)$, or equivalently $R_i(X,P_z(X),z) = 0$. Substituting $X \mapsto x_0$ yields also $Y - P_z(x_0) \mid R_i(x_0,Y,z)$, thus there is some j such that $Y - P_z(x_0) \mid H_{ij}(Y,z)$ and equivalently $H_{ij}(P_z(x_0),z) = 0$. Therefore we have $z \in S_{x_0,R_i,H_{ij}}$, by definition. Let (i,j) be the most common pair appearing in this process and set $R = R_i, H = H_{ij}$. Since the total number of pairs (i,j) is at most D_Y, and using (5.8), by the pigeonhole principle we find that

$$|S_{x_0,R,H}| \ge \frac{|S|}{D_Y} > 2D_Y^2 D_X D_{YZ},$$

as claimed. □

As mentioned above, our proof will *eventually* show that both R and H must in fact be linear in Y, with $R(X, Y, Z) = Y - P(X, Z)$, where P is linear in Z and of degree k in X and $H = R(x_0, Y, Z) = Y - P(x_0, Z)$. We will also see that $P(x_0, z) = P_z(x_0)$ for almost all $z \in S_{x_0,R,H}$. We will reach this point only later — for now we assume R, H have Y-degrees $d, d_H \leq D_Y$ correspondingly, that their total Y, Z degrees at most $D = D_{YZ}$, and are not necessarily monic.

5.2.5 Step 5: Interlude — the algebraic function field $\mathbb{L}$ and the power series $\gamma(X)$. Our next step is to find a root of H (if needed, by artificially adding it to the field $\mathbb{F}_q(Z)$), and then to lift it to a power series solution $Y = \gamma(X)$ to $R(X, Y, Z) = 0$. This process is carried out in Appendix A, which also provides the required setup and definitions from the theory of algebraic extensions of function fields.

We strongly encourage reading Appendix A at this point, as the analysis of the aforementioned γ in the next sections will make use of many objects introduced and discussed there in length. We list here only the notations and basic properties of each object:

- $W(Z)$ — the leading coefficient of $H(Y, Z)$ as a polynomial in Y.
- $\widetilde{H}(T, Z) = W(Z)^{d_H - 1} H(T/W(Z), Z)$ — a monicized version of H (monic in T).
- $\mathbb{L} = \mathbb{F}_q(Z)[T]/(\widetilde{H}(T, Z))$ — an algebraic extension of $\mathbb{F}_q(Z)$, containing a root $\alpha_0 = \frac{T}{W(Z)}$ of $H(Y, Z)$.
- $\mathcal{O} = \mathbb{F}_q[Z][T]/(\widetilde{H}(T, Z))$ — the ring of "regular" elements in $\mathbb{L}$, which is a subring of its ring of integers.
- $\Lambda(\cdot)$ — a weight function from $\mathcal{O}$ to non-negative integers, which satisfies $\Lambda(Z) = 1$, $\Lambda(T) = D - d_H + 1$, and $\Lambda(\alpha\beta) \leq \Lambda(\alpha) + \Lambda(\beta)$ for any $\alpha, \beta \in \mathcal{O}$.
- $\zeta = \frac{\partial R}{\partial Y}(x_0, \alpha_0, Z) \in \mathbb{L}$ — the Y derivative of R_{x_0} at the root α_0, which is non-zero by the separability of R_{x_0}.
- $\xi = W(Z)^{d_H - 2}\zeta \in \mathcal{O}$ — the "regular part" (or numerator) of ζ.
- $\gamma = \sum_{t=0}^{\infty} \alpha_t (X - x_0)^t \in \mathbb{L}[[X - x_0]]$ — a power series which is the Hensel lift of α_0, satisfying $R(X, \gamma, Z) = 0$. The coefficients can be written as $\alpha_t = \frac{\beta_t}{W^{t+1}\xi^{e_t}}$, with $e_t = \max(0, 2t - 1)$ and $\beta_t \in \mathcal{O}$ (see Claim A.2).
- $\pi_z : \mathcal{O} \to \mathbb{F}_q^2$ — substitution maps, which are ring homomorphisms mapping the generators T, Z to $t_z, z \in \mathbb{F}_q$ satisfying $\widetilde{H}(t_z, z) = 0$.
- We will also make use of Lemma A.1, which states that if $\beta \in \mathcal{O}$ satisfies $\pi_z(\beta) = 0$ for more than $d_H \Lambda(\beta)$ different substitutions π_z, then $\beta = 0$.

At the end of the day, the power series $\gamma(X)$ will be shown to be of the form $P(X, Z) \in \mathbb{F}_q[X, Z]$ with X- and Z-degrees at most k and 1, respectively. However, to reach that point, we will analyze $\gamma(X)$ as having coefficients in the algebraic extension $\mathbb{L}$ of $\mathbb{F}_q(Z)$, and unbounded X-degree.

5.2.6 Step 6: Bounding the X-degree of γ. In this section we show that the power series solution $\gamma = \sum_{t=0}^{\infty} \alpha_t (X - x_0)^t$ to $R(X, Y, Z) = 0$ is in fact a finite polynomial in X of degree k. In other words, we prove:

Claim 5.8. *For all $t > k$, $\alpha_t = 0$. Equivalently,*

$$\gamma = \gamma_k = \sum_{t=0}^{k} \alpha_t (X - x_0)^t.$$

The claim is proved in two steps: first for all $k < t < D_X$, by showing that $\pi_z(\alpha_t)$ is well-defined and vanishes for sufficiently many substitutions; then for all $t \geq D_X$ as well, by observing that γ_k is already a root of $R(X, Y, Z)$.

In the course of the proof we will also define the set S' appearing in Proposition 5.5, and show that it satisfies (5.9).

Proof. For each $z \in S_{x_0,R,H}$, from $H(P_z(x_0), z) = 0$ we also get $\widetilde{H}(W(z)P_z(x_0), z) = 0$, and thus we have a substitution map $\pi_z : \mathcal{O} \to \mathbb{F}_q$ with $t_z = W(z)P_z(x_0)$. The denominators appearing in the α_t are all powers of W

and ξ, so we can evaluate α_t at any z which is not a root of W or ξ, i.e. in the set

$$S' = S_{x_0,R,H} \setminus \{z : W(z) = 0 \text{ or } \pi_z(\xi) = 0\}.$$

Using Lemma A.1 and the bounds on $\deg W, \Lambda(\xi)$ from Claim A.2, as well as (5.13), we find

$$\begin{aligned}
|S'| &\geq |S_{x_0,R,H}| - (\deg W + d_H \Lambda(\xi)) \\
&\geq |S_{x_0,R,H}| - ((d-1)d_H + 1)(D - d_H + 1) + 1 \\
&> |S_{x_0,R,H}| - d_H dD \geq \frac{|S|}{D_Y} - D_Y^2 D_{YZ} \\
&> 2D_Y^2 D_X D_{YZ} - D_Y^2 D_{YZ} = D_Y^2 D_{YZ}(2D_X - 1). \qquad (5.14)
\end{aligned}$$

Note that to apply the lemma we also used $\xi \neq 0$ in $\mathbb{L}$. Since furthermore $D_Y^2 D_{YZ} < \frac{|S|}{2D_Y}$, we also get the bound

$$|S'| > \frac{|S|}{D_Y} - D_Y^2 D_{YZ} > \frac{|S|}{2D_Y}$$

as we claimed in (5.9).

Now for each $z \in S'$, we may apply π_z to γ, since we removed all poles of γ from S'. We wish to show that $\pi_z(\gamma) = P_z(X)$. Note that $\pi_z(\gamma) \in \mathbb{F}_q[[X - x_0]]$, and the polynomial $P_z(X)$ can also be considered as a (finite) power series in $\mathbb{F}_q[[X - x_0]]$. We have that $Y = P_z(X), \pi_z(\gamma)$ both are roots of $R(X, Y, z) = 0$: For $P_z(X)$ this follows by definition from $z \in S_{x_0,R,H}$, and for $\pi_z(X)$ we have $R(X, \pi_z(\gamma), z) = \pi_z(R(X, \gamma, Z)) = \pi_z(0) = 0$. Additionally, modulo $X - x_0$ we have $\pi_z(\alpha_0) = \pi_z(T/W) = t_z/W(z) = P_z(x_0)$. Furthermore, $\pi_z(\zeta)$ is well-defined and non-zero (since $\pi_z(\xi) \neq 0$), and equal to

$$\pi_z(\zeta) = \pi_z\left(\frac{\partial R}{\partial Y}\left(x_0, \frac{T}{W}, Z\right)\right) = \frac{\partial R}{\partial Y}\left(x_0, \frac{t_z}{W(z)}, z\right) = \frac{\partial R}{\partial Y}(x_0, P_z(x_0), z).$$

In particular, $\frac{\partial R}{\partial Y}(x_0, P_z(x_0), z), \neq 0$, i.e. $P_z(x_0) = \pi_z(\alpha_0)$ is a simple root of $R(x_0, Y, z)$.

Putting all of the above together, we see that both $\pi_z(\gamma), P_z(X) \in \mathbb{F}_q[[X - x_0]]$ are power series solutions to $R(X, Y, z) = 0$, with the same free coefficient $\pi_z(\alpha_0) = P_z(x_0)$ modulo $X - x_0$, which is a simple root of $R(x_0, Y, z)$. Thus, they must be identically equal, by the uniqueness of the Hensel lift with a given starting simple root. In other words, we have

$$\mathbb{F}_q[[X - x_0]] \ni \sum_{t=0}^{\infty} \pi_z(\alpha_t)(X - x_0)^t = \pi_z(\gamma) = P_z(X) \in \mathbb{F}_q[X - x_0]$$

and in particular, $\pi_z(\alpha_t) = 0$ for all $z \in S'$ and all $t > k$, since $\deg(P_z) \leq k$. Thus we also find $\pi_z(\beta_t) = 0$ for all $t > k$. Restricting to $k < t < D_X$, we additionally have by Claim A.2

$$\Lambda(\beta_t) < (2t + 1)dD \leq dD(2D_X - 1),$$

and from (5.14) it follows that

$$|S'| > D_Y^2 D_{YZ}(2D_X - 1) \geq d_H dD(2D_X - 1) > d_H \Lambda(\beta_t).$$

We can therefore apply Lemma A.1 to find that indeed $\beta_t = 0$ and $\alpha_t = 0$ in $\mathbb{L}$.

We thus have that the degree k polynomial

$$\gamma_k = \sum_{t=0}^{k} \alpha_t (X - x_0)^t = \sum_{t=0}^{D_X - 1} \alpha_t (X - x_0)^t = \gamma_{D_X - 1} \in \mathbb{L}[X]$$

satisfies $\gamma \equiv \gamma_k \pmod{(X - x_0)^{D_X}}$, and therefore

$$R(X, \gamma_k, Z) \equiv 0 \pmod{(X - x_0)^{D_X}};$$

but, $R(X, \gamma_k, Z) \in \mathbb{L}[X]$ is a polynomial of degree $< D_X$, since by construction Q has $(1, k)$-weighted degree less than D_X and so do its factors, and therefore $R(X, \gamma_k, Z) = 0$ identically. By the uniqueness of the lifting, we thus find $\gamma = \gamma_k \in \mathbb{L}[X]$, as claimed. □

5.2.7 Step 7: Bounding the Z-degree of γ. In the previous section we have seen that $\gamma = \gamma_k \in \mathbb{L}[X]$ is a polynomial of degree at most k in X, whose coefficients lie in $\mathbb{L}$, an extension field of $\mathbb{F}_q(Z)$. We have also seen that $\pi_z(\gamma) = P_z(X)$ for all $z \in S'$. In this section we will show that the coefficients of γ are all in fact simply linear polynomials in Z, and thus obtain that $\gamma = P(X, Z) \in \mathbb{F}_q[X, Z]$ with X-degree at most k and Z-degree at most 1:

Claim 5.9. *The exists degree $\le k$ polynomials $v_0, v_1 \in \mathbb{F}_q[X]$, such that*

$$\gamma = v_0(X) + Z \cdot v_1(X) =: P(X, Z).$$

Proving Claim 5.9 will also complete the proof of Proposition 5.5, as $\gamma = P$ satisfies (5.11), the set S' satisfies (5.9), and together they satisfy (5.10), since for each $z \in S'$,

$$P(X, z) = \pi_z(P(X, Z)) = \pi_z(\gamma) = P_z(X).$$

We prove Claim 5.9 by showing that $\gamma(x)$ agrees with the linear function $w(x, Z)$ on at least $k + 1$ values of x, using sufficiently many Z-substitutions at each x, and then use the fact γ can be interpolated from such values of $\gamma(x)$, and this interpolation is also linear in Z. Details follow.

We consider good pairs of $x \in \mathcal{D}, z \in S'$ satisfying $w(x, z) = P_z(x)$. We define the sets of x's which are good for each $z \in S'$ and vice versa, that is,

$$\mathcal{D}_z = \{x \in \mathcal{D} : w(x, z) = P_z(x)\},$$

$$S'_x = \{z \in S' : w(x, z) = P_z(x)\} = \{z \in S' : x \in \mathcal{D}_z\}.$$

By the definitions of S and P_z, we have $|\mathcal{D}_z| \ge n - e$ for each $z \in S'$, where $e = \lfloor \delta n \rfloor$. We make the following claims regarding the sizes of the sets S'_x:

Claim 5.10. *Suppose $|S'_x| > (2k + 1)d_H dD$. Then $\gamma(x) = w(x, Z)$, and in particular, $\gamma(x)$ is a linear polynomial in $\mathbb{F}_q[Z]$.*

Claim 5.11. *There exists a set $\mathcal{D}_{\text{top}} = \{x_1, \ldots, x_{k+1}\} \subset \mathcal{D}$ of $k + 1$ points of $\mathcal{D}$, satisfying $|S'_{x_j}| > (2k + 1)d_H dD$ for all $1 \le j \le k + 1$.*

Before proving the two claims, let us first deduce Claim 5.9 from them:

Proof Claim 5.9. Observe that from Claim 5.10 and Claim 5.11 it follows that $\gamma(x) = w(x, Z) = u_0(x) + Z \cdot u_1(Z)$ is linear in Z for every $x = x_j \in \mathcal{D}_{\text{top}}$. But, since $\gamma(X)$ is a polynomial of degree at most k, it can be interpolated from its values in any $k + 1$ points, and this interpolation only involves operations over $\mathbb{F}_q$. Thus the interpolated polynomial will also have coefficients which are linear in Z.

More concretely, let $v_0(X), v_1(X) \in \mathbb{F}_q[X]$ be the unique polynomials of degree at most k interpolating $u_0(x), u_1(x)$ at the points of $\mathcal{D}_{\text{top}}$. Then $\gamma(X)$ and $v_0(X) + Z \cdot v_1(X)$ are two polynomials in $\mathbb{L}[X]$ of degree at most k which agree on at least $k + 1$ evaluations, since

$$\gamma(x_j) = w(x_j, Z) = u_0(x_j) + Z \cdot u_1(x_j) = (v_0(X) + Z \cdot v_1(X))(x_j)$$

for each $x_j \in \mathcal{D}_{\text{top}}$. It follows that γ and $v_0 + Z \cdot v_1$ are identically equal as polynomials in $\mathbb{L}[X]$, as claimed. □

We now proceed to prove the claims:

Proof of Claim 5.10. Since $\pi_z(\gamma) = P_z(X)$ for each $z \in S'$, by definition of S'_x we have

$$\pi_z(\gamma(x)) = P_z(x) = w(x,z) = u_0(x) + z \cdot u_1(x)$$

for each $z \in S'_x$, or equivalently

$$\pi_z\big(\gamma(x) - (u_0(x) + Z \cdot u_1(x))\big) = 0. \tag{5.15}$$

On the other hand, we can write

$$\begin{aligned}\gamma(x) - (u_0(x) + Z \cdot u_1(x)) &= \left(\frac{1}{W^{k+1}\xi^{e_k}} \sum_{t=0}^{k} \beta_t (x - x_0)^t W^{k-t} \xi^{e_k - e_t}\right) - (u_0(x) + Z \cdot u_1(x)) \\ &= \frac{1}{W^{k+1}\xi^{e_k}} \Big(\beta(x) - (u_0(x) + u_1(x) \cdot Z)\ W^{k+1}\xi^{e_k}\Big),\end{aligned} \tag{5.16}$$

where $\beta(x) \coloneqq \sum_{t=0}^{k} \beta_t (x - x_0)^t W^{k-t} \xi^{e_k - e_t} \in \mathcal{O}$, which by Claim A.2 has weight

$$\begin{aligned}\Lambda(\beta(x)) &\le \max_{t=0,\ldots,k} (\Lambda(\beta_t) + (k-t)\Lambda(W) + (e_k - e_t)\Lambda(\xi)) \\ &\le \max_{t=0,\ldots,k} ((1 + (t+1)\Lambda(W) + e_t\Lambda(\xi)) + (k-t)\Lambda(W) + (e_k - e_t)\Lambda(\xi)) \\ &= 1 + (k+1)\Lambda(W) + e_k\Lambda(\xi) \le (2k+1)dD,\end{aligned}$$

and so does $\widetilde{\beta}(x) \coloneqq \beta(x) - (u_0(x) + u_1(x) \cdot Z)\ W^{k+1}\xi^{e_k}$. From (5.15) and (5.16) we have $\pi_z(\widetilde{\beta}(x)) = 0$ for all $z \in S'_x$, with $|S'_x| > (2k+1)d_H dD \ge d_H \Lambda(\widetilde{\beta}(x))$, by assumption. By Lemma A.1 it follows that $\widetilde{\beta}(x) = 0$, and thus $\gamma(x) = u_0(x) + Z \cdot u_1(x) = w(x,Z)$ identically in $\mathbb{L}$, as claimed. □

Proof of Claim 5.11. Let $\mathcal{D}_{\text{top}} = \{x_1, \ldots, x_{k+1}\} \subset \mathcal{D}$ be the set of the $x_j \in \mathcal{D}$ with the $k+1$ largest sizes of $|S'_{x_j}|$, breaking ties arbitrarily. We first claim that for each $1 \le j \le k+1$,

$$|S'_{x_j}| \ge \frac{n-k-1-e}{n-k-1}|S'| \ge \frac{1-\rho-\delta}{1-\rho}|S'|.$$

This follows by way of contradiction, which we get by double counting the set of bad pairs $(x,z) \in \mathcal{D} \times S'$ with $w(x,z) \ne P_z(x)$. The contrary assumption implies that $|S'_x| < \frac{n-k-1-e}{n-k-1}|S'|$ for all $x \notin \mathcal{D}_{\text{top}}$, giving many bad z-s for each such x, but on the other hand each z is only paired with a few bad x-s. More precisely, we get the following contradiction:

$$\begin{aligned}e|S'| \ge \sum_{z \in S'} |\mathcal{D} \setminus \mathcal{D}_z| = \sum_{x \in \mathcal{D}} |S' \setminus S'_x| &\ge \sum_{x \in \mathcal{D} \setminus \mathcal{D}_{\text{top}}} (|S'| - |S'_x|) \\ &> (n-k-1)\left(1 - \frac{n-k-1-e}{n-k-1}\right)|S'| = e|S'|.\end{aligned} \tag{5.17}$$

From (5.14) we have

$$|S'_{x_j}| \ge \frac{1-\rho-\delta}{1-\rho}|S'| > \frac{1-\rho-\delta}{1-\rho} D_Y^2 D_{YZ}(2D_X - 1) \ge \frac{1-\rho-\delta}{1-\rho}(2D_X - 1)d_H dD,$$

so to conclude $|S'_{x_j}| > (2k+1)d_H dD$ it suffices to show that $2D_X - 1 > \frac{1-\rho}{1-\rho-\delta}(2k+1)$. And indeed for $m \geq 3$, we have

$$\begin{aligned} 2D_X - 1 = (2m+1)\sqrt{\rho}\, n - 1 > 2 \cdot 3\sqrt{\rho}\, n - 1 > \frac{1+\sqrt{\rho}}{\sqrt{\rho}}(3\rho\, n) - 1 \\ > \frac{1-\rho}{\sqrt{\rho}(1-\sqrt{\rho})}(2k+1) > \frac{1-\rho}{\sqrt{\rho}+\frac{\sqrt{\rho}}{2m}-\rho}(2k+1) \\ = \frac{1-\rho}{1-\delta_0(\rho,m)-\rho}(2k+1) \geq \frac{1-\rho}{1-\rho-\delta}(2k+1) \end{aligned}$$

as needed. □

5.2.8 Step 8: Proving the correlated agreement between u_i and v_i. We have found a polynomial $\gamma = v_0(X) + Z \cdot v_1(X)$ of the required degrees satisfying $Q(X, \gamma, Z) = 0$. To finish the proof of Theorem 5.1, it now remains only to be seen that $\gamma(x, Z)$ and $w(x, Z)$ agree identically on all but $\delta|\mathcal{D}|$ values of $x \in \mathcal{D}$.

Define $\mathcal{D}' = \{x \in \mathcal{D} : |S'_x| \geq 2\}$. Note that for each $x \in \mathcal{D}'$, we must have $v_0(x) = u_0(x)$, $v_1(x) = u_1(x)$ (and thus $\gamma(x, Z) = w(x, Z)$), since

$$\begin{aligned} v_0(x) + z \cdot v_1(x) = \pi_z(\gamma(x)) = \pi_z(\gamma)(x) = P_z(x) = w(x, z) \\ = u_0(x) + z \cdot u_1(x) \end{aligned}$$

for at least 2 different values of $z \in S'_x$. Since $|S'_x| \leq 1$ for every $x \in \mathcal{D} \setminus \mathcal{D}'$, double counting the number of bad pairs $(x, z) \in \mathcal{D} \times S'$ with $w(x, z) \neq P_z(x)$, as in (5.17), gives

$$e|S'| \geq \sum_{z \in S'} |\mathcal{D} \setminus \mathcal{D}_z| = \sum_{x \in \mathcal{D}} |S' \setminus S'_x| \geq \sum_{x \in \mathcal{D} \setminus \mathcal{D}'} (|S'| - 1) = (|S'| - 1)|\mathcal{D} \setminus \mathcal{D}'|$$

and therefore $|\mathcal{D} \setminus \mathcal{D}'| \leq \frac{e|S'|}{|S'|-1}$. On the other hand, from (5.3), (5.6), (5.9), as well as $m \geq 2$ and $\rho < 1$ we have

$$|S'| > \frac{|S|}{2D_Y} > \frac{(m+\frac{1}{2})^6}{6\rho} n^2 > 2n^2, \tag{5.18}$$

thus

$$|\mathcal{D} \setminus \mathcal{D}'| \leq \frac{e|S'|}{|S'|-1} < \left(1 + \frac{2}{|S'|}\right)e < \left(1 + \frac{1}{n^2}\right)e < e + \frac{1}{n} < e + 1,$$

i.e. $|\mathcal{D} \setminus \mathcal{D}'| \leq e \leq \delta n$, which we finally rewrite as

$$|\{x \in \mathcal{D} : (u_0(x), u_1(x)) = (v_0(x), v_1(x))\}| \geq |\mathcal{D}'| \geq (1-\delta)|\mathcal{D}|$$

as we wanted to show. □

Remark 5.1. Note that from $\gamma = v_0(X) + Z \cdot v_1(X) = P(X, Z)$ solving $R(X, \gamma, Z) = 0$ we find that $Y - P(X, Z) \mid R(X, Y, Z)$. But since R is irreducible, it indeed follows that $R = Y - P(X, Z)$ is monic and linear in Y, and so is H, as was mentioned earlier; and, as mentioned, we only reach this conclusion near the end of the proof. It is an interesting open problem whether this conclusion can be reached without passing through the various extension fields as our proof required.

6 CORRELATED AGREEMENT IN GENERALIZED SETTINGS

Both Theorems 4.1 and 5.1 considered two functions $u_0, u_1 : \mathcal{D} \to \mathbb{F}$ and the corresponding affine line $\{u_0 + zu_1 : z \in \mathbb{F}_q\}$ inside the linear plane $\text{span}\{u_0, u_1\}$. A generalization of these theorems that is particularly important to the soundness analysis of the FRI protocol in Theorem 8.3 is obtained by considering $l+1$ functions $u_0, \ldots, u_l : \mathcal{D} \to \mathbb{F}$, and the 1-dimensional, degree l parameterized curve

$$\{u_0 + zu_1 + z^2u_2 + \cdots + z^l u_l : z \in \mathbb{F}_q\}$$

inside the linear space $\text{span}\{u_0, \ldots, u_l\}$. This curve can also be viewed as a function $w(\cdot, Z) : \mathcal{D} \to \mathbb{K}$ given by

$$w(x, Z) = u_0(x) + u_1(x)Z + u_2(x)Z^2 + \cdots + u_l(x)Z^l.$$

The two theorems can then be generalized as follows:

Theorem 6.1. *Suppose $\delta \le (1-\rho)/2$. Let $u_0, u_1, \ldots, u_l : \mathcal{D} \to \mathbb{F}_q$ be functions. Let*

$$S = \{z \in \mathbb{F}_q : \Delta(u_0 + zu_1 + \cdots + z^l u_l, V) \le \delta\}$$

and suppose $|S| > l \cdot n$. Then for all $z \in \mathbb{F}_q$ we have

$$\Delta(u_0 + zu_1 + \cdots + z^l u_l, V) \le \delta,$$

and furthermore there are $v_0, \ldots, v_l \in V$ such that for all $z \in \mathbb{F}_q$,

$$\Delta(u_0 + zu_1 + \cdots + z^l u_l, v_0 + zv_1 + \cdots + z^l v_l) \le \delta$$

and in fact

$$|\{x \in \mathcal{D} : (u_0(x), \ldots, u_l(x)) \ne (v_0(x), \ldots, v_l(x))\}| \le \delta|\mathcal{D}|.$$

Theorem 6.2. *Let $u_0, u_1, \ldots, u_l : \mathcal{D} \to \mathbb{F}_q$, let $m \ge 3$, define $\delta_0(\rho, m) := 1 - \sqrt{\rho} - \frac{\sqrt{\rho}}{2m}$, and let $\delta \le \delta_0(\rho, m)$. Define*

$$S = \{z \in \mathbb{F}_q : \Delta(u_0 + zu_1 + \cdots + z^l u_l, V) \le \delta\}$$

and suppose

$$|S| > \frac{(1 + \frac{1}{2m})^7 m^7}{3\rho^{3/2}} n^2 l. \tag{6.1}$$

Then $u_0, \ldots, u_l$ are simultaneously δ-close to V, i.e. $\exists v_0, \ldots, v_l \in V$ such that

$$|\{x \in \mathcal{D} : \forall 0 \le i \le l, u_i(x) = v_i(x)\}| \ge (1-\delta)|\mathcal{D}|.$$

These generalizations do not greatly affect the proofs, which were presented in the previous sections in the special case $l = 1$ only for the purposes of simplicity. We will thus not repeat the arguments in full, but only detail the required changes in subsections 6.1 and 6.2.

Another generalization of interest is to correlated agreement in the entire affine space, stated as Theorem 1.6, restated here for completeness:

Theorem 1.6 (Correlated agreement over affine spaces). *Let $V, q, n, k, \mathcal{D}$ and ρ be as defined in Theorem 1.2. For $u_0, u_1, \ldots u_l \in \mathbb{F}_q^{\mathcal{D}}$ let $U = u_0 + \text{span}\{u_1, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ be an affine subspace. If $\delta \in (0, 1 - \sqrt{\rho})$ and*

$$\Pr_{u \in U}[\Delta(u, V) \le \delta] > \epsilon,$$

where ϵ is as defined in Theorem 1.2, then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, \ldots, v_l \in V$ satisfying

- **Density:** $|\mathcal{D}'|/|\mathcal{D}| \ge 1 - \delta$, *and*
- **Agreement:** *for all $i \in \{0, \ldots, l\}$, the functions u_i and v_i agree on all of $\mathcal{D}'$.*

Furthermore, in the unique decoding regime $\delta \in \left(0, \frac{1-\rho}{2}\right]$, *there exists a unique* maximal $\mathcal{D}'$ *satisfying the above, with unique* v_i.

To prove this theorem we will make use of the following lemma:

Lemma 6.3. *Let* V, q, n, k *and* ρ *be as defined in Theorem 1.2. For* $u_0, u_1, \ldots u_l \in \mathbb{F}_q^{\mathcal{D}}$ *let* $U = u_0 + \mathrm{span}\{u_1, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ *be an affine subspace, and let* $U' = \mathrm{span}\{u_1, \ldots, u_l\}$ *be the corresponding linear subspace (so that* $U = u_0 + U'$ *and* $U' = U - U$*). If* $\delta \in (0, 1 - \sqrt{\rho})$ *and*

$$\Pr_{u \in U}\left[\Delta(u, V) \leq \delta\right] > \epsilon,$$

where ϵ *is as defined in Theorem 1.2, then we have* $\Delta(u', V) \leq \delta$ *for every* $u' \in U'$.

Note that Lemma 6.3 is very similar to Theorem 1.2, the difference being only that the consequent $\Delta(u', V) \leq \delta$ is stated for elements of the linear space U' instead of elements of the affine space U. In particular, when U itself is linear and $U' = U$, they are equivalent. Moreover, Theorem 1.2 can be proven from Lemma 6.3, and in fact we will pass through such a proof in the course of proving Theorem 1.6. Both Lemma 6.3 and Theorem 1.6 are proven in Section 6.3.

6.1 Proof of Theorem 6.1

Following the proof in Section 4, the first change is in the analysis of the matrix M, where the entries in the columns corresponding to the coefficients of A are of degree $\leq l$ in Z instead of 1 (and the columns corresponding to the coefficients of B remain of degree 0). The determinants of minors $R(Z)$ are then of degree at most $l(e+1)$, which is less than $|S|$, so are still identically 0, and the matrix has rank $< k + 2e + 2$.

Similarly, the solution A, B of the equation system will have Z-degrees $\deg_Z(A) \leq l \cdot e$, $\deg_Z(B) \leq l(e+1)$, so a_Z, b_Z grow by a factor of l, as does $d_Z = l$ (which remains the degree of the $w(x, Z)$), and similarly $n_Z = |S| > l \cdot n$. We again have

$$\frac{b_X}{n_X} + \frac{b_Z}{n_Z} < \frac{k+e}{n} + \frac{l(e+1)}{l \cdot n} < \frac{k + 2e + 1}{n} \leq 1$$

and the Polishchuk–Spielman lemma remains applicable, and the rest of the conclusions are derived as before with no further changes, save that $P(X, Z)$ is now of degree l in Z, and thus written as $v_0(X) + Zv_1(X) + \cdots Z^l v_l$.

6.2 Proof of Theorem 6.2

As in the $l = 1$ case, we apply the Guruswami–Sudan decoder to the $\mathbb{K}^{\mathcal{D}}$ word $w(x, Z) = \sum_{i=0}^{l} u_i(x) Z^i$. We use exactly the same parameters, and the degrees D_X, D_Y are unaffected. In the entries of the equation system, every appearance of (a power of) the previously linear $w(x, Z)$ is replaced by (a power of) a polynomial of degree l, thus the bounds on D_C, D_R, D, D_{YZ} are all multiplied by l, and specifically $D_{YZ} < \frac{(m+1/2)^3}{6\sqrt{\rho}} n\, l$.

Also note that instead of the Y, Z degrees being graded in Q such that the coefficient $Q_{ji}(Z)$ of $X^i Y^j$ is of degree at most $D - j$ in Z, it is of degree at most $D - l \cdot j$, i.e. when Y and Z are assigned weights l and 1 (instead of 1 and 1), the total weight of Q, and thus also of R and H, is at most D. Thus all bounds on Z-degrees and weights are henceforth simply multiplied by l. The leading coefficient W of H has degree at most $D - l \cdot d$, and the variable $T = W(Z)Y$ is assigned the l times larger weight $D - l(d-1)$, which makes the weight of the monic polynomial $\widetilde{H}$ dominated by its leading monomial T^d, thus the weight never increases on reductions modulo $\widetilde{H}$. The bounds on weights of ξ, β_t are multiplied by l. Since for $\mathbb{F}_q[Z]$ polynomials the weight corresponds to degree, Lemma A.1 remains valid; in all its applications, the upper bounds on weights of the regular elements are multiplied by l; and so are the lower bounds on the sizes of the sets of vanishing substitutions, i.e. the sets S, $S_{x_0,R,H}$, S', S'_{x_j}. Thus we can still deduce that γ is of degree k, and then by interpolating $u_0, \ldots, u_l$ by polynomials $v_0, \ldots, v_l$ at $k+1$ appropriately chosen points, to deduce that $\gamma = \sum_{i=0}^{l} v_i(X) Z^i$ identically.

For the final argument we again note that $\sum_{i=0}^{l} u_i(x)z^i = \sum_{i=0}^{l} v_i(x)z^i$ for all $x \in \mathcal{D}$ and $z \in S'_x$. Since for a fixed x both sides of the equation are degree l polynomials in z, if $|S'_x| > l$ then we must have $u_i(x) = v_i(x)$ for all $0 \le i \le l$. We have $|S'| > 2l \cdot n^2$, or equivalently $\frac{l}{|S'|} < \frac{2}{n^2}$, and thus the number of $x \in \mathcal{D}$ for which $|S'_x| \le l$ is at most $\frac{e|S'|}{|S'|-l} < e+1$, and again we find that u_i and v_i all agree on the set $\mathcal{D}' = \{x \in \mathcal{D} : |S'_x| > l\}$ which is of size at least $(1-\delta)n$, as claimed. □

6.3 Proofs of Lemma 6.3 and Theorem 1.6

Proof of Lemma 6.3. Let $u' \in U'$ be an arbitrary element. If $u' = 0$, then clearly $\Delta(u', V) = 0 \le \delta$. Otherwise, consider the partition of U into affine lines parallel to u'; formally, write $U = \widetilde{U} \oplus \mathrm{span}\{u'\}$, where $\widetilde{U} \subset U$ is some direct complement of $\mathrm{span}\{u'\}$ in U' shifted by u_0. Then

$$\mathbb{E}_{\tilde{u} \in \widetilde{U}} \Pr_{z \in \mathbb{F}_q} [\Delta(\tilde{u} + z \cdot u', V) \le \delta] = \Pr_{u \in U} [\Delta(u, V) \le \delta] > \epsilon,$$

and in particular there exists some $\tilde{u} \in \widetilde{U}$ for which $\Pr_{z \in \mathbb{F}_q} [\Delta(\tilde{u} + z \cdot u', V) \le \delta] > \epsilon$. We can thus apply Theorem 1.4 to the line $\{\tilde{u} + z \cdot u' : z \in \mathbb{F}_q\}$, which in particular implies $\Delta(u', V) \le \delta$, as claimed. □

Proof of Theorem 1.6. Let $\delta^* = \mathrm{D}(U, V)$, where D is the divergence as defined in subsection 1.2, and let $u^* \in U$ be an element with $\Delta(u^*, V) = \delta^*$.

We first show that $\delta^* \le \delta$. Let $\overline{U} = \mathrm{span}(U)$ be the linear space spanned by U, which either equals U when $0 \in U$, or otherwise is the disjoint union of $U' = U - U$ and $\{z \cdot U : z \in \mathbb{F}_q \setminus \{0\}\}$. By Lemma 6.3 we have $\Delta(u', V) \le \delta$ for all $u' \in U'$, and we also have $\Delta(z \cdot u, V) = \Delta(u, V)$ for all $z \ne 0$ and $u \in U$. Thus in both cases, whether $\overline{U} = U$ or $\overline{U} = U' \cup \bigcup_{z \ne 0}(z \cdot U)$, we have

$$\Pr_{\bar{u} \in \overline{U}} [\Delta(\bar{u}, V) \le \delta] \ge \Pr_{u \in U} [\Delta(u, V) \le \delta] > \epsilon.$$

We can therefore apply Lemma 6.3 to $\overline{U}$, and in particular for $u^* \in U \subset \overline{U}$ we get $\delta^* = \Delta(u^*, V) \le \delta$, as claimed. As previously noted, this is in fact the content of Theorem 1.2, which is now proved.

Let $\{v_1^*, \ldots, v_L^*\} \subset V$ be all possible codewords at distance at most δ^* from u^*. Note that they are all in fact at distance *exactly* δ^*, not less, since otherwise we would have $\Delta(u^*, V) < \delta^*$. For each $1 \le i \le L$ define the agreement set $\mathcal{D}'_i = \{x \in \mathcal{D} : u^*(x) = v_i^*(x)\}$ which has size exactly $(1-\delta^*)|\mathcal{D}|$. Note that in the unique decoding regime $\delta^* \le \frac{1-\rho}{2}$ we must have $L = 1$, and in the general Johnson/Guruswami–Sudan regime $\delta^* < 1 - \sqrt{\rho} - \eta$ we have $L < q$: indeed, from $\epsilon = \epsilon_J < 1$ and $\eta < \frac{\sqrt{\rho}}{16}$ we must have $q > (2\eta)^{-7} > \frac{2^{17}}{\eta}, \frac{2^{21}}{\sqrt{\rho}}$, and from the analysis of the Guruswami–Sudan algorithm, using (5.6) and $m = \left\lceil \frac{\sqrt{\rho}}{2\eta} \right\rceil < \frac{\sqrt{\rho}}{2\eta} + 1$ we find

$$L \le D_Y < \frac{2m+1}{2\sqrt{\rho}} < \frac{1}{2\eta} + \frac{3}{2\sqrt{\rho}} < 2^{-17} q < q.$$

For each $i \le L$, let $U_i = \{u \in U : u|_{\mathcal{D}'_i} \in V|_{\mathcal{D}'_i}\}$ be the set of all functions in U which agree with some codeword in V on all of $\mathcal{D}'_i$. Note that this condition is linear, therefore each U_i is an affine subspace of U. We claim that $U \subset \bigcup_{i=0}^{L} U_i$, i.e. every element of U belongs to at least one U_i. This is obvious for u^*, which belongs to all U_i. Consider any $u \in U \setminus \{u^*\}$, and the affine line $\{u^* + z \cdot (u - u^*) : z \in \mathbb{F}_q\} \subset U$ containing u^* and u. By definition of δ^*, we have

$$\Pr_{z \in \mathbb{F}_q} [\Delta(u^* + z \cdot (u - u^*), V) \le \delta^*] = 1 > \epsilon.$$

Since $\delta^* \le \delta$, by Theorem 1.4 we have *correlated* agreement in the line, i.e. there exists $\mathcal{D}' \subset \mathcal{D}$ with $|\mathcal{D}'| \ge (1-\delta^*)|\mathcal{D}|$, and codewords $v^*, v \in V$ which respectively agree with u^*, u on $\mathcal{D}'$. In particular $\Delta(u^*, v^*) \le \delta^*$,

thus v^* must be one of the L decodings $\{v_1^*, \ldots, v_L^*\}$, as the list was exhaustive, and $\mathcal{D}' = \mathcal{D}_i'$. This exactly implies $u \in U_i \subset \bigcup_{i=0}^{L} U_i$, as claimed.

Comparing sizes, we find that the largest U_i must then satisfy $|U_i| \geq \frac{|U|}{L} > \frac{|U|}{q}$. But U_i is a subspace of U, and it thus follows that $U_i = U$. Now setting $\mathcal{D}' = \mathcal{D}_i'$, we have in particular that the restrictions of $u_0, u_1, \ldots, u_l$ to $\mathcal{D}'$ are codewords, since they are elements or differences of elements in U_i. Then setting $v_0, v_1, \ldots, v_l \in V$ to be the unique extensions of these codewords from $\mathcal{D}'$ to $\mathcal{D}$, we find that $\mathcal{D}', v_0, \ldots, v_l$ satisfy both conditions. Again note that in the unique decoding regime, $L = 1$ and this $\mathcal{D}'$ is therefore uniquely defined. □

7 CORRELATED WEIGHTED AGREEMENT

For certain applications, like analyzing the soundness of the FRI protocol (Section 8.2), a *weighted* version of Theorem 1.5 is necessary, and we provide it in this section.

For a given weight vector $\mu : \mathcal{D} \to [0, 1]$, the (relative) *$\mu$-agreement* between words u, v is defined as

$$\operatorname{agree}_\mu(u, v) \coloneqq \frac{1}{|\mathcal{D}|} \sum_{x : u(x) = v(x)} \mu(x).$$

Note that for $\mu \equiv 1$ the notion of μ-agreement is equivalent to the standard notion of relative agreement, which is defined as $\operatorname{agree}(u, v) = 1 - \Delta(u, v)$. The agreement between a word u and a linear code V is the maximal agreement between u and a codeword of V,

$$\operatorname{agree}_\mu(u, V) \coloneqq \max_{v \in V} \operatorname{agree}_\mu(u, v).$$

We also define the weighted size of a subdomain $\mathcal{D}' \subset \mathcal{D}$ as

$$\mu(\mathcal{D}') \coloneqq \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}'} \mu(x).$$

Thus the agreement satisfies $\operatorname{agree}_\mu(u, v) = \mu(\{x \in \mathcal{D} : u(x) = v(x)\})$. Finally, for $\mathbf{u} = \{u_0, \ldots, u_l\}$, $u_i \in \mathbb{F}^{\mathcal{D}}$ a set of words, the *μ-weighted correlated agreement* of $\mathbf{u}$ and V is the maximal μ-weighted size of a subdomain $\mathcal{D}' \subset \mathcal{D}$ such that the restriction of $\mathbf{u}$ to $\mathcal{D}'$ belongs to $V|_{\mathcal{D}'}$, i.e., for each $i = 0, \ldots, l$ there exist $v_i \in V$ such that $u_i|_{\mathcal{D}'} = v_i|_{\mathcal{D}'}$. When μ is unspecified, it is set to the constant weight function 1, which recovers the notion of correlated agreement measure discussed in Section 1.3.

In what follows we shall assume the weight function μ has some structure, specifically, all weights $\mu(x)$ are the form $\mu(x) = \frac{a_x}{M}$ for varying integers a_x and common denominator M. This assumption indeed holds for the special case of FRI soundness (where M equals the blocklength of the RS code to which the FRI protocol is applied). The following is the weighted generalization of Theorem 1.5.

Theorem 7.1 (Weighted correlated agreement over curves – Version I). *Let V, q, n, k and ρ be as defined in Theorem 1.2. Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$. Let $\alpha \in (\sqrt{\rho}, 1)$ and let $\mu : \mathcal{D} \to [0, 1]$ be a vector of weights, whose values all have denominator M. Suppose*

$$\Pr_{u \in \operatorname{curve}(\mathbf{u})} \left[\operatorname{agree}_\mu(u, V) \geq \alpha \right] > l \cdot \epsilon,$$

where ϵ is as defined in Theorem 1.2 (with $\eta = \min(\alpha - \sqrt{\rho}, \frac{\sqrt{\rho}}{20})$), and additionally suppose

$$\Pr_{u \in \operatorname{curve}(\mathbf{u})} \left[\operatorname{agree}_\mu(u, V) \geq \alpha \right] \geq \frac{l(M|\mathcal{D}| + 1)}{q} \left(\frac{1}{\eta} + \frac{3}{\sqrt{\rho}} \right).$$

Then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, \ldots, v_l \in V$ satisfying

- *μ-**Density:** $\mu(\mathcal{D}') \geq \alpha$, and*

- **Agreement:** *for all $i \in \{0, \ldots, l\}$, the functions u_i and v_i agree on all of $\mathcal{D}'$.*

A slightly more precise form, only for the Johnson bound regime, is the following:

Theorem 7.2 (Weighted correlated agreement over curves – Version II). *Let V, q, n, k and ρ be as defined in Theorem 1.2. Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$. Let $\mu : \mathcal{D} \to [0,1]$ be a vector of weights, whose values all have denominator M. Let $m \geq 3$ and let*

$$\alpha \geq \alpha_0(\rho, m) \coloneqq \sqrt{\rho} + \frac{\sqrt{\rho}}{2m}.$$

Let

$$S = \{z \in \mathbb{F}_q : \mathrm{agree}_\mu(u_0 + zu_1 + \cdots + z^l u_l, V) \geq \alpha\}$$

and suppose

$$|S| > \max\left(\frac{(1+\frac{1}{2m})^7 m^7}{3\rho^{3/2}}\, n^2 l,\ \frac{2m+1}{\sqrt{\rho}}(M \cdot n + 1)l\right). \tag{7.1}$$

Then $u_0, \ldots, u_l$ have at least α correlated μ-agreement with V, i.e. $\exists v_0, \ldots, v_l \in V$ such that

$$\mu(\{x \in \mathcal{D} : \forall 0 \leq i \leq l, u_i(x) = v_i(x)\}) \geq \alpha.$$

Similarly, we can also prove a weighted version of the theorem for affine spaces. As was the case in the unweighted version, here the lower bounds on the probability or size of S are the same as they are for affine lines, which can be considered as curves with degree $l = 1$. Again we give two versions, with η and with m:

Theorem 7.3 (Weighted correlated agreement over affine spaces). *Let V, q, n, k and ρ be as defined in Theorem 1.2. Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ and let $U = u_0 + \mathrm{span}\{u_1, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ be an affine subspace. Let $\alpha \in (\sqrt{\rho}, 1)$ and let $\mu : \mathcal{D} \to [0,1]$ be a vector of weights, whose values all have denominator M. Suppose*

$$\Pr_{u \in U}\left[\mathrm{agree}_\mu(u, V) \geq \alpha\right] > \epsilon,$$

where ϵ is as defined in Theorem 1.2 (with $\eta = \min(\alpha - \sqrt{\rho}, \frac{\sqrt{\rho}}{20})$), and additionally suppose

$$\Pr_{u \in U}\left[\mathrm{agree}_\mu(u, V) \geq \alpha\right] \geq \frac{M|\mathcal{D}| + 1}{q}\left(\frac{1}{\eta} + \frac{3}{\sqrt{\rho}}\right).$$

Then there exist $\mathcal{D}' \subset \mathcal{D}$ and $v_0, \ldots, v_l \in V$ satisfying

- *μ-**Density:** $\mu(\mathcal{D}') \geq \alpha$, and*
- *__Agreement:__ for all $i \in \{0, \ldots, l\}$, the functions u_i and v_i agree on all of $\mathcal{D}'$.*

Theorem 7.4 (Weighted correlated agreement over affine spaces – Version II). *Let V, q, n, k and ρ be as defined in Theorem 1.2. Let $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ and let $U = u_0 + \mathrm{span}\{u_1, \ldots, u_l\} \subset \mathbb{F}_q^{\mathcal{D}}$ be an affine subspace. Let $\mu : \mathcal{D} \to [0,1]$ be a vector of weights, whose values all have denominator M. Let $m \geq 3$ and let*

$$\alpha \geq \alpha_0(\rho, m) \coloneqq \sqrt{\rho} + \frac{\sqrt{\rho}}{2m}.$$

Suppose

$$\Pr_{u \in U}\left[\mathrm{agree}_\mu(u, V) \geq \alpha\right] > \max\left(\frac{(1+\frac{1}{2m})^7 m^7}{3\rho^{3/2}} \cdot \frac{n^2}{q},\ \frac{2m+1}{\sqrt{\rho}} \cdot \frac{M \cdot n + 1}{q}\right). \tag{7.2}$$

Then $u_0, \ldots, u_l$ have at least α correlated μ-agreement with V, i.e. $\exists v_0, \ldots, v_l \in V$ such that

$$\mu(\{x \in \mathcal{D} : \forall 0 \leq i \leq l, u_i(x) = v_i(x)\}) \geq \alpha.$$

7.1 Proof of Theorems 7.1 and 7.2

The proofs of the theorems rely on the proofs of Theorems 6.1 and 6.2, which in turn were generalizations of the proofs of Theorems 4.1 and 5.1. Let S be as in Theorem 7.2, so that for each $z \in S$ there is $P_z \in V$ with $\text{agree}_\mu(\sum_{j=0}^{l} z^j u_j, P_z) \geq \alpha$. By definition, the unweighted agreement is at least the weighted agreement, thus $\Delta(\sum_{j=0}^{l} z^j u_j, P_z) \leq 1-\alpha$. It follows that we can immediately apply Theorems 6.1 and 6.2 to the set S, as all other assumptions hold, and deduce that $u_0, \ldots, u_l$ have at least α correlated agreement with codewords $v_0, \ldots, v_l$. This is not sufficient, however, as our goal is to show μ-agreement, which is stronger.

To continue, we observe that in the course of our proofs, the codewords $v_0, \ldots, v_l$ that we found had the property that the identity $v_0 + zv_1 + \cdots + z^l v_l = P_z$ was satisfied for every $z \in S'$, where $S' \subset S$ was a fairly large set. In the unique decoding regime this is even true for $S' = S$, because P_z and $\sum_{j=0}^{l} z^j v_j$ are both decodings of $\sum_{j=0}^{l} z^j u_j$, and the decoding is unique. In the Johnson bound regime, this is exactly the content of Proposition 5.5, with $P(X, Z)$ (of degree l in Z) written explicitly as $P(X, Z) = \sum_{j=0}^{l} Z^j v_j$, and a set S' of size greater than $\frac{|S|}{2D_Y}$ by (5.9). By (5.6) we have $2D_Y \leq \frac{2m+1}{\sqrt{\rho}}$, and in the setting of Theorem 7.1 we use the definition $m = \left\lceil \frac{\sqrt{\rho}}{2\eta} \right\rceil$ to get

$$2D_Y < \frac{2m+1}{\sqrt{\rho}} < \frac{\frac{\sqrt{\rho}}{\eta} + 3}{\sqrt{\rho}} = \frac{1}{\eta} + \frac{3}{\sqrt{\rho}}.$$

We thus find that in all cases the additional assumptions give us $|S'| \geq (M \cdot n + 1)l$. The crucial aspect here is that these P_z which for $z \in S'$ are linear combinations of the v_j, are the same P_z-s with which we started—which in this setting were assumed to also have high μ-agreement with linear combinations of the words u_j, and not just regular agreement. In other words, we know that $\sum_{j=0}^{l} z^j u_j(x)$ and $\sum_{j=0}^{l} z^j v_j(x)$ have high μ-agreement for every $z \in S'$. We can now use this fact to deduce that the v_j also must have correlated μ-agreement with u_j. The necessary argument is very similar to that of Section 5.2.8, but in higher generality. We apply it through the following two lemmas, first getting the correlated agreement with a small loss, then showing this loss must be 0, if $|S'|$ is large enough in comparison to the denominator M of the weights – as we had assumed.

Lemma 7.5. *Let V, q, n, k and ρ be as defined in Theorem 1.2. Let $u_0, \ldots, u_l \in \mathbb{F}_q^{\mathcal{D}}$, $v_0, \ldots, v_l \in V$, let μ be a weight vector, and let $\alpha \geq 0$. Denote*

$$w(x, z) = \sum_{j=0}^{l} z^j u_j(x), \quad \widetilde{w}(x, z) = \sum_{j=0}^{l} z^j v_j(x)$$

for all $x \in \mathcal{D}$, $z \in \mathbb{F}_q$, and suppose that there exists a set $S' \subset \mathbb{F}_q$ with $|S'| > l$ and such that

$$\forall z \in S', \ \text{agree}_\mu(w(\cdot, z), \widetilde{w}(\cdot, z)) \geq \alpha. \tag{7.3}$$

Then the correlated agreement domain of the (u_j) and (v_j)

$$\mathcal{D}' = \{x \in \mathcal{D} : (u_0(x), \ldots, u_l(x)) = (v_0(x), \ldots, v_l(x))\}$$

satisfies

$$\mu(\mathcal{D}') > \alpha - \frac{l}{|S'| - l}.$$

Lemma 7.6. *Let $V, q, n, k, \rho, \alpha, \mu, u_j, v_j, S', \mathcal{D}'$ be as in Lemma 7.5 and with the same assumptions. Assume additionally that μ only takes rational values with denominator (dividing) M, and that $|S'| \geq M|\mathcal{D}|l + l$. Then*

$$\mu(\mathcal{D}') \geq \alpha.$$

As we have seen that S' satisfies the assumptions of Lemma 7.5 and $|S'| \geq (M \cdot n + 1)l = M|\mathcal{D}|l + l$ holds, our claimed correlated μ-agreement follows immediately from Lemma 7.6. It remains to prove the lemmas, which we do in the next subsection.

7.2 Proofs of Lemmas 7.5 and 7.6

Proof of Lemma 7.5. Denote

$$S'_x = \{z \in S' : w(x,z) = \widetilde{w}(x,z)\}, \quad \forall x \in \mathcal{D},$$
$$\mathcal{D}_z = \{x \in \mathcal{D} : w(x,z) = \widetilde{w}(x,z)\}, \quad \forall z \in S'.$$

Note that for any given $x \in \mathcal{D}$, $w(x,z), \widetilde{w}(x,z)$ are both polynomials in z of degree at most l. It follows that if they agree on more than l values of z, then they must be identical as polynomials and agree for *every* z. Thus, either $|S'_x| \leq l$ or $|S'_x| = |S'|$. Moreover, since $|S'| > l$, the set $\mathcal{D}'$ is precisely the set of $x \in \mathcal{D}$ for which $|S'_x| = |S'|$, and $\mathcal{D} \setminus \mathcal{D}'$ is the set on which $|S'_x| \leq l$. On the other hand, for each $z \in S'$, $\mathcal{D}_z$ is the agreement set of $w(\cdot,z), \widetilde{w}(\cdot,z)$, so by (7.3), it has μ-weighted size at least α. Thus by double counting we get

$$\begin{aligned}\alpha|S'| &\leq \sum_{z \in S'} \mathrm{agree}_\mu(w(\cdot,z), \widetilde{w}(\cdot,z)) = \sum_{z \in S'} \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}_z} \mu(x) \\ &= \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} \mu(x)|S'_x| \leq \mu(\mathcal{D}')|S'| + \mu(\mathcal{D} \setminus \mathcal{D}') \cdot l = \mu(\mathcal{D}')(|S'| - l) + \mu(\mathcal{D}) \cdot l.\end{aligned}$$

Here, the first inequality is from the lower bound on the μ-agreement of $w(\cdot,z), \widetilde{w}(\cdot,z)$, the next equality is by definition of μ-agreement, the next equality is double counting (changing the order of summation), and the final inequality is from bounding the size of $|S'_x|$ inside and outside $\mathcal{D}'$. Rearranging, we get

$$\mu(\mathcal{D}') \geq \frac{\alpha|S'| - \mu(\mathcal{D}) \cdot l}{|S'| - l} = \alpha - \frac{(\mu(\mathcal{D}) - \alpha) \cdot l}{|S'| - l} > \alpha - \frac{l}{|S'| - l}$$

as claimed. Here the final step uses the simple bound $\mu(\mathcal{D}) - \alpha < 1$. □

Proof of Lemma 7.6. By the assumption on the values of μ, it follows that agree_μ only takes rational values with denominator $M|\mathcal{D}|$, and similarly so does $\mu(\mathcal{D}')$. We may therefore round α up to the nearest multiple of $\frac{1}{M|\mathcal{D}|}$ without affecting the validity of the assumption (7.3), nor of the wanted consequent $\mu(D') \geq \alpha$. The assumption $|S'| \geq M|\mathcal{D}|l + l$ is equivalent to $\frac{l}{|S'|-l} \leq \frac{1}{M|\mathcal{D}|}$, and thus from Lemma 7.5 we get

$$\mu(\mathcal{D}') > \alpha - \frac{l}{|S'| - l} \geq \alpha - \frac{1}{M|\mathcal{D}|}.$$

But, since both α and $\mu(\mathcal{D}')$ are integer multiples of $\frac{1}{M|\mathcal{D}|}$, it follows that $\mu(\mathcal{D}') \geq \alpha$, as claimed. □

7.3 Proof of Theorems 7.3 and 7.4

As was the case in the unweighted version, these theorems are obtained by reduction to the respective theorems on affine lines (which are special cases of Theorems 7.1 and 7.2 with $l = 1$). The reduction is exactly as done in Section 6.3, with the distance Δ replaced by the weighted agreement agree_μ; the parameters δ, δ^* replaced by α, α^*; all inequalities on agreement being in the opposite direction to those on distances, thus $\alpha^* \geq \alpha$ is the *smallest* μ-agreement with the code of any $u \in U$; and the sizes of the agreement sets $\mathcal{D}_i, \mathcal{D}'$ are replaced by their μ-sizes, which are now equal to α^* instead of $(1 - \delta^*)|\mathcal{D}|$.

For this reduction, there are no new nuances introduced by the move to μ-weights, and the proof can be reapplied verbatim, with the above mentioned replacements. We thus omit duplicating it here.

8 APPLICATIONS TO VERIFIABLE SECRET SHARING AND FRI SOUNDNESS

In this section we give some more details on applications that were briefly described in Section 3.

8.1 Verifiable Secret Sharing

We start with an application of distributed proximity testing of Interleaved Reed–Solomon Codes (Theorem 3.1) to *verifiable secret sharing* (VSS) [18]. VSS serves as a building block both for useful special tasks, such as simultaneous broadcast and fair coin-flipping, and for general protocols for secure multiparty computation with an honest majority [4, 15, 32].

A VSS protocol with security threshold t is a two-phase protocol involving a dealer, n servers, and an honest output client. The *sharing phase* starts with the dealer distributing the secret s among the servers by sending a share $s^{\langle i \rangle}$ to server i and is followed by a verification protocol. In the *reconstruction* phase, each server sends its share $s^{\langle i \rangle}$ to the output client, who reconstructs s from the shares. Both phases of the protocol are attacked by a malicious adversary who may corrupt at most t servers and possibly also the dealer. Communication proceeds in synchronous rounds and may use secure point-to-point channels. We also assume the availability of a common broadcast medium and a source of unpredictable public coins. These assumptions, which can be eliminated at a small amortized cost, make the protocol simpler. Finally, while the communication is synchronous, the adversary has a *rushing* capability, in the sense that it can wait to receive messages from uncorrupted parties before sending its own messages.

Here we consider a simplified variant of VSS that allows a denial-of-service attack only in the sharing phase, but not in the reconstruction phase. The latter "guaranteed output delivery" requirement makes the protocol suitable for applications that rely on independence, such as simultaneous broadcast and fair coin-tossing, as they prevent the adversary from making the protocol selectively fail based on information obtained in the final round via rushing.

A VSS protocol as above should satisfy the following properties:

- Completeness: if the dealer is honest and the sharing phase succeeds, the output client outputs s. Moreover, the sharing phase succeeds if the adversary does not attack it.
- Secrecy: if the dealer is honest, the adversary learns nothing about s.
- Unique reconstruction: even if the dealer is dishonest, if the sharing phase succeeds then the messages sent in the sharing phase define a unique s^* such that (except with small failure probability) the output client will output s^* in the end of the reconstruction phase.

Implementing VSS efficiently is a challenging task. Here we consider the following simple approach for simultaneously sharing l secrets $s_1, \ldots, s_l$. This approach underlies the efficient MPC protocols from [19, 24, 27]. A centralized variant of this protocol (in a relaxed setting that allows reconstruction to fail) is a building block in efficient two-party protocols for zero knowledge [1, 25] and secure computation [23, 26].

We will assume here that each secret s_j is a single field element and all secrets originate from the same dealer. However, the protocol is even more attractive when the secrets originate from different dealers and when each secret s_j is a vector of $\ell < n$ field elements that are simultaneously shared via so-called "packed secret sharing" [20]. Moreover, while the communication complexity of the protocol beats all competing approaches we are aware of in the *amortized* case when l is large, it is potentially useful (and nontrivial to analyze) even with $l = 1$. In fact, if we count public coins and broadcast as normal communication, the communication complexity of the protocol is nontrivial to match even in this case.

The two phases of the protocol proceed as follows.

Sharing. In the sharing phase, the dealer uses Shamir's secret sharing scheme [36] for sharing each secret, with secrecy threshold t. Viewed abstractly, each s_j is randomly mapped to a codeword u_j of an $[n, t+1, d]$ RS code

V over $\mathbb{F}_q$, with $d = n - t$. The dealer distributes u_j between the servers, together with an additional random codeword $u_0 \in_R V$ that is used for blinding. Following a public random challenge $r \in \mathbb{F}_q^l$, each server i broadcasts its view of $u' = r^T \mathbf{u}$, namely $u'(i) = u_0(i) + \sum_{j=1}^{l} r(i)u_j(i)$. (As discussed in Section 3, this challenge can be compressed using either cryptography or simple derandomization techniques.) The sharing phase succeeds if $u' \in V$.

Reconstruction. In the reconstruction phase, each server sends its shares to an output client, who recovers the secrets s_j by error-correcting the (potentially) corrupted codewords u_j^*.

We assume in the following that $n > 3t$, implying that the minimal distance of the underlying RS code satisfies $d = n - t > 2t$. We start by considering the case of a *static* adversary, who decides in advance which t servers to corrupt. Completeness follows from the fact that $t < d/2$ and each u_j^* is t-close to V. Secrecy follows from the secrecy property of Shamir's scheme and from the fact that u_0 blinds the information exchanged during the verification process. For the unique reconstruction property, consider the shares of the $n - t > 2t$ uncorrupted servers. If they are not consistent with V, then (by a simple analysis) the sharing phase will fail except with $1/q$ probability. If they are consistent with V, then (as before) u_j^* is within the unique decoding radius and the outputs will be correct.

What goes wrong when the adversary can be adaptive? In this case, the dealer could potentially distribute badly formed vectors $\mathbf{u}$ that have the following devious property: there is u_j which is very far from the code, and yet u' is with high probability (say, 1/2) t-close to the code. Now, the adaptive adversary can corrupt only those servers in $T = \Gamma(u', V)$ and send on their behalf field elements that make u' consistent with V. This in turn makes the sharing phase succeed. But since u_j is not within the unique decoding distance from V, we lose the unique decoding guarantee. Theorem 1.2 rules out the existence of such a devious $\mathbf{u}$. But this is not enough. We actually need to ensure that $\mathbf{u}$ is consistent with V when restricted to the $n - t$ servers that are not corrupted during reconstruction phase.

This is ensured by the stronger guarantee of Theorem 3.1. The analysis proceeds as follows. If $\mathbf{u}$ is t-far from the interleaved code $\mathbf{V}$, then u' will be t-far from V except with $\leq n/q$ probability. In this event, even an adaptive corruption cannot make u' look consistent with V, and the sharing phase fails. If $\mathbf{u}$ is within distance $t' \leq t$ from $\mathbf{V}$, then except with $\leq n/q$ probability we have that $\Gamma(u', V) = \Gamma(\mathbf{u}, \mathbf{V})$. Denoting this set by T', the adversary must corrupt the entire set T' for the sharing phase to succeed, and may additional corrupt $t - t'$ more servers. In this case, $\mathbf{u}$ restricted to the $n - t$ uncorrupted servers is fully consistent with V, guaranteeing unique reconstruction.

We finally note that security against an adaptive adversary can be reduced to security against a static adversary via a generic union bound argument, taking the union over all $\binom{n}{t}$ sets of servers that can be eventually be corrupted [13]. (Alternatively, this follows from the simple derivation of proximity gaps over exponentially large fields discussed in the Introduction.) However, this would require the field size q to satisfy $q = 2^{\Omega(n)}$, which would make communication grow quadratically (rather than linearly) with the number of servers n. In this parameter regime the proximity gap approach does not seem useful, as there are VSS protocols that meet the quadratic communication bound with *perfect* (rather than statistical) security against an adaptive adversary [4].

8.2 Soundness of the Fast RS IOP of Proximity (FRI) protocol for batched instances

In this section we use Theorems 7.2 and 7.4, the weighted and sharper versions of Theorems 1.5 and 1.6, to improve the soundness of the FRI protocol [5], when applied to a batch of functions. We start by briefly recalling the essential components of the protocol needed to state our theorem, assuming familiarity with the protocol (see [5] for a more thorough explanation of it). We also explain below the meaning of "batching" of FRI instances.

The FRI protocol. As explained in Section 3.2, the purpose of FRI is to verify, in the IOP model, the proximity of a received word $f^{(0)} : \mathcal{D}^{(0)} \to \mathbb{F}$ to an RS code $V^{(0)} := \mathsf{RS}[\mathbb{F}, \mathcal{D}^{(0)}, k^{(0)}]$. FRI works for any evaluation domain

$\mathcal{D}^{(0)}$ that is a coset of 2-*smooth* group, i.e., for any $\mathcal{D}^{(0)}$ that is a coset of a group (additive or multiplicative) of size 2^s, for integer s. Henceforth we assume the group $\mathcal{D}^{(0)}$ is multiplicative[6]. The FRI protocol has two phases, called COMMIT and QUERY. During the COMMIT phase, a sequence of functions $f^{(1)} : \mathcal{D}^{(1)} \to \mathbb{F}, f^{(2)} : \mathcal{D}^{(2)} \to \mathbb{F}, \ldots, f^{(\mathsf{r})} : \mathcal{D}^{(\mathsf{r})} \to \mathbb{F}$ is generated over a finite number r of interactive rounds. With each iteration the domain size $|\mathcal{D}^{(i)}|$ shrinks. Assuming an honest prover and $f^{(0)}$ being low-degree, the low-degreeness property is maintained for each $f^{(i)}$ (see Claim 8.1). At the beginning of the i-th round, the prover message $f^{(i)} : \mathcal{D}^{(i)} \to \mathbb{F}$ has already been created (and the verifier has oracle access to it). The verifier now sends a uniformly random $z^{(i)} \in \mathbb{F}$ and the prover replies with a new function $f^{(i+1)} : \mathcal{D}^{(i+1)} \to \mathbb{F}$ where $\mathcal{D}^{(i+1)}$ is a (2-smooth) strict subgroup of $\mathcal{D}^{(i)}$.

$\mathcal{D}^{(i+1)}$ partitions $\mathcal{D}^{(i)}$ into cosets of size $l^{(i)} := |\mathcal{D}^{(i)}|/|\mathcal{D}^{(i+1)}|$. Let $C_g^{(i)}$ denote the coset corresponding to $g \in D^{(i+1)}$, namely

$$C_g^{(i)} := \{g' \in \mathcal{D}^{(i)} \mid (g')^{l^{(i)}} = g\}. \tag{8.1}$$

For each coset $C_g^{(i)}$, the *interpolation map* $M_g^{(i)}$ is the invertible linear map $M_g^{(i)} : \mathbb{F}^{C_g^{(i)}} \to \mathbb{F}^{l^{(i)}}$ over $\mathbb{F}$ that maps $f^{(i)}\big|_{C_g^{(i)}} : C_g^{(i)} \to \mathbb{F}$ — the restriction of $f^{(i)}$ to domain $C_g^{(i)} \subset \mathcal{D}^{(i)}$ — to the vector $\mathbf{u}^{(i)}(g) = \left(u_0^{(i)}(g), \ldots, u_{l^{(i)}-1}^{(i)}(g)\right)$ of coefficients of the polynomial $P_{\mathbf{u}^{(i)}(g)}(Z) = \sum_{j<l^{(i)}} Z^j \cdot \left(u_j^{(i)}(g)\right)$ that interpolates $f^{(i)}\big|_{C_g^{(i)}}$. In other words, $M_g^{(i)}$ is the inverse of the Vandermonde matrix generated by $C_g^{(i)}$, which implies that $\left(M_g^{(i)}\right)^{-1} \cdot (u_0, \ldots, u_{l^{(i)}-1})$ is the evaluation of the polynomial $P_{\mathbf{u}}(X) = \sum_{i<^{(i)}} u_i X^i$ on the coset $C_g^{(i)}$.

The following claim is a restatement of [5, Section 4.1], using our notation (and working over a multiplicative rather than additive group). For the sake of completeness a proof appears in Appendix E.

Claim 8.1. *Suppose that $f^{(i)} \in \mathsf{RS}[\mathbb{F}, \mathcal{D}^{(i)}, k^{(i)}]$ where $k^{(i)}+1$ is an integral power of 2. Then, for any $z^{(i)} \in \mathbb{F}$, letting $\mathbf{z}^{(i)} = \left(\left(z^{(i)}\right)^0, \left(z^{(i)}\right)^1, \ldots, \left(z^{(i)}\right)^{l^{(i)}-1}\right)$, the function $f^{(i+1)}_{f^{(i)},z^{(i)}} : \mathcal{D}^{(i+1)} \to \mathbb{F}$ defined on $g \in \mathcal{D}^{(i+1)}$ by*

$$f^{(i+1)}_{f^{(i)},z^{(i)}}(g) := \left(\mathbf{z}^{(i)}\right)^\top \cdot \mathbf{u}^{(i)}(g) = \left(\mathbf{z}^{(i)}\right)^\top \cdot M_g^{(i)} \cdot f^{(i)}\Big|_{C_g^{(i)}} \tag{8.2}$$

is a valid codeword of $V^{(i+1)} := \mathsf{RS}[\mathbb{F}, \mathcal{D}^{(i+1)}, k^{(i+1)}]$ where $k^{(i+1)} := \frac{k^{(i)}+1}{l^{(i)}} - 1$.

Batching. In certain cases the first prover oracle $f^{(0)}$ is sampled from an affine space $F \subset \mathbb{F}^{\mathcal{D}^{(0)}}$ of functions, which serves as our input,

$$F = \left\{ f_0^{(0)} + \sum_{i=1}^{t} x_i \cdot f_i^{(0)} | x_i \in \mathbb{F}, f_i^{(0)} : \mathcal{D}^{(0)} \to \mathbb{F} \right\}. \tag{8.3}$$

This is the case when the FRI protocol is used to "batch" several different instances of the low degree testing problem, combining them all using a random linear combination. In this batching setting we assume the prover has committed to $f_1^{(0)}, \ldots, f_t^{(0)}$ (notice we set $f_0^{(0)} = 0$ in this case), and the batched FRI verifier samples uniformly random $x_1, \ldots, x_t \in \mathbb{F}$, the prover replies with $f^{(0)}$ which supposedly equals $f_0^{(0)} + \sum_i x_i \cdot f_i^{(0)}$, and the FRI protocol is now applied to $f^{(0)}$. Accordingly, the batched FRI QUERY phase is extended so that each time a query to $f^{(0)}(g)$ is requested, the verifier also queries $f_0^{(0)}(g), \ldots, f_t^{(0)}(g)$ and verifies that indeed $f^{(0)}(g) = f_0^{(0)}(g) + \sum_i x_i \cdot f_i^{(0)}(g)$.

[6]The FRI protocol in [5] is stated for cosets of additive 2-smooth groups; Remark 3.1 in the online version of [5] translates the results to multiplicative groups (cf. Remark 1.4 and Section 2.1 there). Generalizing the results here to (i) additive groups, (ii) t-smooth groups for larger constant t and (iii) cosets of groups, is straightforward, using [5]; we omit these generalizations to simplify the exposition.

The (batched) FRI QUERY phase. Claim 8.1 implies that an honest prover may construct from a codeword $f^{(i)} \in V^{(i)}$ a new codeword $f^{(i+1)} \in V^{(i+1)}$ (for any value $z^{(i)}$ picked by the verifier), doing so by computing Eq. (8.2) for each $g \in \mathcal{D}^{(i+1)}$. Henceforth we shall always assume $f^{(\mathrm{r})} \in V^{(\mathrm{r})}$, say, by assuming the verifier always queries the first $k^{(\mathrm{r})}$ elements of $f^{(\mathrm{r})}$ (according to some canonical order) and identifies $f^{(\mathrm{r})}$ with the interpolating polynomial of this function.

Claim 8.1 suggests a natural test for consistency between $f^{(i)}$ and $f^{(i+1)}$, and the QUERY phase of FRI follows this natural test by applying it iteratively from "top" ($f^{(\mathrm{r})}$) to "bottom" ($f^{(0)}$), according to the following process

A single invocation of the batched QUERY phase:

(1) Pick uniformly random $g^{(\mathrm{r})} \in \mathcal{D}^{(\mathrm{r})}$. For $i = \mathrm{r}, \ldots, 1$, sample $g^{(i-1)}$ uniformly at random from the coset $C^{(i-1)}_{g^{(i)}}$.
(2) If $f^{(0)}(g^{(0)}) \neq f_0^{(0)}(g^{(0)}) + \sum_{i=1}^{t} x_i \cdot f_i^{(0)}(g^{(0)})$, then reject.
(3) If, for any $i \in \{0, \ldots, \mathrm{r}-1\}$, we have $f^{(i+1)}(g^{(i+1)}) \neq \left(\mathbf{z}^{(i)}\right)^{\top} \cdot M^{(i)}_{g^{(i+1)}} \cdot f^{(i)}\big|_{C^{(i)}_{g^{(i+1)}}}$, then reject.
(4) Otherwise — when equality holds in all cases mentioned in the bullets above, then accept.

Summary of the batched FRI protocol: Let us summarize the essential properties recounted thus far, as they will be used in our soundness analysis below.

(1) At the end of the protocol's COMMIT phase the verifier has oracle access to a sequence of functions $f^{(0)} : \mathcal{D}^{(0)} \to \mathbb{F}, \ldots, f^{(\mathrm{r})} : \mathcal{D}^{(\mathrm{r})} \to \mathbb{F}$ where $\mathcal{D}^{(0)} \supsetneq \ldots, \supsetneq \mathcal{D}^{(\mathrm{r})}$ is a sequence of 2-smooth groups and $f^{(i)}$ depends arbitrarily on $z^{(0)}, \ldots, z^{(i)}$ (and $f^{(0)}, \ldots, f^{(i-1)}$). We assume that $f^{(\mathrm{r})} \in V^{(\mathrm{r})}$.
(2) There exists a set of $l^{(i)} \times l^{(i)}$ invertible matrices $\{M^{(i)}_{g^{(i+1)}} : g^{(i+1)} \in \mathcal{D}^{(i+1)}\}$, so that applying $M^{(i)}_{g^{(i+1)}}$ to $f^{(i)}\big|_{C^{(i)}_{g^{(i+1)}}}$ maps $f^{(i)}$ to a sequence of vectors $\mathbf{u} = \mathbf{u}^{(i)} = \{u_0^{(i)}, \ldots, u_{l^{(i)}}^{(i)}\} \subset \mathbb{F}^{\mathcal{D}^{(i+1)}}$, where
$$\mathbf{u}^{(i)}\left(g^{(i+1)}\right) = \left(u_0^{(i)}\left(g^{(i+1)}\right), \ldots, u_{l^{(i)}-1}^{(i)}\left(g^{(i+1)}\right)\right) = M^{(i)}_{g^{(i+1)}} \cdot f^{(i)}\Big|_{C^{(i)}_{g^{(i+1)}}}. \tag{8.4}$$
Furthermore, if $f^{(i)}$ is a valid RS codeword over $\mathcal{D}^{(i)}$ of rate ρ, then each vector on the parameterized curve passing through $\mathbf{u}^{(i)}$ is a valid RS codeword over $\mathcal{D}^{(i+1)}$ of the same rate ρ.
(3) Each iteration of the QUERY phase checks whether $f^{(i+1)}$ was constructed from $f^{(i)}$ via Eq. (8.2) and (in the batched case) whether $f^{(0)}$ was computed correctly via Eq. (8.3).

Soundness. We now bound the soundness error of the batched FRI protocol, using Theorems 7.1 and 7.2. Recall the notion of correlated agreement from Section 1.3 and its generalization to μ-weighted correlated agreement defined at the beginning of Section 7.

Lemma 8.2 (batched FRI error bound)**.** *Let $V^{(0)} = \mathrm{RS}[\mathbb{F}, \mathcal{D}^{(0)}, k^{(0)}]$ where $\mathcal{D}^{(0)}$ is a coset of a 2-smooth multiplicative group, and $k^{(0)} + 1$ is a power of 2; set $\rho = (k^{(0)} + 1)/|\mathcal{D}^{(0)}|$.*

Let $F \subseteq \mathbb{F}^{\mathcal{D}^{(0)}}$ be a space of functions as defined in Eq. (8.3) *whose correlated agreement density with $V^{(0)}$ is at most α. For integer $m \geq 3$, let*
$$\alpha^{(0)}(\rho, m) = \max\{\alpha, \sqrt{\rho}(1 + 1/2m)\}.$$
Assume the FRI protocol is used with r *rounds, and let $l^{(i)} = |\mathcal{D}^{(i)}|/|\mathcal{D}^{(i+1)}|$ denote the ratio between prover messages (oracles) i and $i+1$. Let ϵ_{Q} denote the probability that the verifier accepts a single FRI QUERY invocation. Then,*
$$\Pr_{x_1,\ldots,x_t,z^{(0)},\ldots,z^{(\mathrm{r}-1)}}\left[\epsilon_{\mathrm{Q}} > \alpha^{(0)}(\rho, m)\right] \leq \epsilon_{\mathrm{C}}, \tag{8.5}$$

where

$$\epsilon_{\mathsf{C}} = \frac{\left(m+\frac{1}{2}\right)^7 \cdot |\mathcal{D}^{(0)}|^2}{2\rho^{3/2}|\mathbb{F}|} + \frac{(2m+1)\cdot(|\mathcal{D}^{(0)}|+1)}{\sqrt{\rho}} \cdot \frac{\sum_{i=0}^{\mathsf{r}-1} l^{(i)}}{|\mathbb{F}|}.$$

In words: For any interactive FRI prover P^*, the probability that the oracles $f^{(0)}, \ldots, f^{(\mathsf{r})}$ sent by P^* will pass a single invocation of the batched FRI QUERY test with probability greater than $\alpha^{(0)}(\rho, m)$, is smaller than ϵ_{C}. The probability is over the random variables $x_1, \ldots, x_t$ used to sample $f^{(0)}$ from F and over the random messages $z^{(0)}, \ldots, z^{(\mathsf{r}-1)}$ sent by the verifier during the COMMIT phase.

This lemma is proved in Section 8.2.1. At the high level there are two ingredients: one is the reduction of a batched FRI claim (about a tuple of functions) to a single FRI claim (about a single function), and the other is a better analysis of FRI for a single function. The first ingredient is a direct application of the proximity gaps phenomenon. The second ingredient, in principle, comes from the previously observed [11, 12] connection that proximity gaps give FRI soundness. However, there is an error in the proof of this connection in [12], and the proof of this connection in [11] is in a slightly more complicated setting. Finally, the proof from [11] of the connection becomes significantly cleaner and quantitatively stronger because of the *exact* distance preservation (not approximate distance preservation) that our proximity gaps have, and which was not present in previous works. Thus we take the opportunity to present a complete analysis of the FRI protocol (following and simplifying [11]) in the quantitatively strongest form that we know, which involves our weighted proximity gaps theorems.

Lemma 8.2 gives us the following result.

Theorem 8.3 (Batched FRI Soundness). *Let $f_0^{(0)}, \ldots, f_t^{(0)} : \mathcal{D}^{(0)} \to \mathbb{F}$ be a sequence of functions and let $V^{(0)} = \mathsf{RS}[\mathbb{F}, \mathcal{D}^{(0)}, k^{(0)}]$ where $\mathcal{D}^{(0)}$ is a coset of a 2-smooth group of size $n^{(0)} = |\mathcal{D}^{(0)}|$, and $\rho = \frac{k^{(0)}+1}{n^{(0)}}$ satisfies $\rho = 2^{-\mathsf{R}}$ for positive integer R. Let $\alpha = \sqrt{\rho}(1+1/2m)$ for integer $m \geq 3$ and ϵ_{C} be as defined in Lemma 8.2.*

Assume the FRI protocol is used with r rounds. Let $l^{(i)} = |\mathcal{D}^{(i)}|/|\mathcal{D}^{(i+1)}|$ denote the ratio between prover messages (oracles) i and $i+1$. Assume furthermore that s is the number of invocations of the FRI QUERY step.

Suppose there exists a batched FRI prover P^* that interacts with the batched FRI verifier and causes it to output "accept" with probability greater than

$$\epsilon_{\mathsf{FRI}} := \epsilon_{\mathsf{C}} + \alpha^s = \frac{\left(m+\frac{1}{2}\right)^7 \cdot |\mathcal{D}^{(0)}|^2}{2\rho^{3/2}|\mathbb{F}|} + \frac{(2m+1)\cdot(|\mathcal{D}^{(0)}|+1)}{\sqrt{\rho}} \cdot \frac{\sum_{i=0}^{\mathsf{r}-1} l^{(i)}}{|\mathbb{F}|} + \left(\sqrt{\rho}\cdot\left(1+\frac{1}{2m}\right)\right)^s.$$

Then $f_0^{(0)}, \ldots, f_t^{(0)}$ have correlated agreement with $V^{(0)}$ on a domain $\mathcal{D}' \subset \mathcal{D}^{(0)}$ of density at least α.

Numerical Example: Suppose $q \geq 2^{256}$, $n = |\mathcal{D}^{(0)}| = 2^{20}$, $\rho = 2^{-4}$, so $k+1 = 2^{16}$, an entirely reasonable choice for practical applications (In STARKs, k corresponds to the length of a computation for which a STARK proof is being generated). Set $m = 2^{11} - 1$ and notice $\sum(l^{(i)} - 1) \leq n$, so

$$\epsilon_{\mathsf{C}} < \frac{(2^{11})^7 \cdot 2^{40}}{2 \cdot 2^{-6} \cdot 2^{256}} + \frac{2^{11} \cdot 2^{21}}{2^{-2}} \cdot \frac{2^{20}}{2^{256}} < 2^{-133}.$$

Assume the maximal correlated agreement density of $f_0^{(0)}, \ldots, f_t^{(0)}$ with $\mathsf{RS}[\mathbb{F}, \mathcal{D}^{(0)}, k+1]$ is at most

$$\alpha = \sqrt{\rho}(1+1/2m) \approx 0.25006\ldots\ldots \approx \frac{1}{4}.$$

Invoking the QUERY protocol for $s = 65$ invocations gives

$$\alpha^s \approx 0.25006^{65} < 2^{-129.97}$$

So the total FRI error is bounded by

$$\epsilon_{\mathsf{FRI}} \leq \epsilon_{\mathsf{C}} + \alpha^s < 2^{-128}.$$

In words, if the FRI protocol accepts with probability greater than 2^{-128} then $f_0^{(0)}, \ldots, f_t^{(0)}$ have correlated agreement with $V^{(0)}$ of density greater than α. Thus we get 128-bits of provable soundness for the FRI protocol for this setting of parameters and 65 invocations of the QUERY phase.

Proof of Theorem 8.3. The proof follows directly from Lemma 8.2 by way of contradiction. Suppose the maximal correlated agreement of $f_0^{(0)}, \ldots, f_t^{(0)}$ with $V^{(0)}$ is less than $\alpha^{(0)}(\rho, m) = \sqrt{\rho}(1+1/2m)$, yet the probability of acceptance is greater than $\epsilon_{\mathrm{C}} + (\alpha^{(0)}(\rho, m))^s$.

Let E be the event that each FRI QUERY accepts with probability greater than $\alpha^{(0)}(\rho, m)$. This event depends on $x_1, \ldots, x_t, f^{(0)}, z^{(0)}, f^{(1)}, \ldots, z^{(\mathrm{r}-1)}, f^{(\mathrm{r})}$, where each $f^{(i)}$ is generated by P^* in response to prior verifier messages. By Lemma 8.2, for any prover P^*, the probability of the event E is bounded by ϵ_{C}. When E does not hold, then the probability of the event that s independent invocations of the FRI QUERY all return "accept", is bounded by $(\alpha_0(\rho, m))^s$.

We conclude the probability of acceptance of the FRI verifier is bounded by $\epsilon_{\mathrm{C}} + (\alpha^{(0)}(\rho, m))^s$, contradicting our assumption. □

Discussion and Two Open Questions. Theorem 8.3 improves on the previous state of the art, due to [11, 12], in several ways. First, as explained in Section 3.2, the prior state of the art required a proximity parameter that is smaller than the "one-and-a-half Johnson bound": $\delta < 1 - \sqrt[3]{\rho}$; the current result pushes the proximity parameter (for large fields) up to the Johnson/Guruswami–Sudan bound. Second, the error parameter during the COMMIT phase was worse, and the analysis incurred an additional loss in the proximity parameter during the QUERY phase, which led to worse concrete soundness bounds across all proximity parameter settings. Last, but not least, the prior bounds on δ (above the unique decoding radius) were only valid for the case where the "folding parameters" $l^{(0)}, \ldots, l^{(\mathrm{r}-1)}$ were all equal to the fixed value $l^\star = 2$, and deteriorated swiftly for larger $l^{(i)}$ (they only work up to radius $1 - \sqrt[(l^{(i)}+1)]{\rho}$). The current bounds deteriorate much more slowly with $l^{(i)}$, and this is important because large values of $l^{(i)}$ are often preferable in practice.

Ben-Sasson et al. showed in [5] that for the FRI protocol to achieve security parameter λ (i.e., $\epsilon_{\mathsf{FRI}} \leq 2^{-\lambda}$), we need to use at least $s \geq \lambda / \log \frac{1}{\rho}$ invocations of the QUERY phase, and conjectured that this lower bound on s is also sufficient (for sufficiently large fields). As noted earlier in Section 3.2 our results imply that taking $s \approx 2\lambda / \log \frac{1}{\rho}$ suffices to get security parameter λ for quadratically large fields. Closing the gap between the provable upper and lower bounds on s is left as an interesting open problem. Concretely, the following conjectured improvement to our main correlated agreement theorem (Theorem 1.4) will imply the conjecture of [5]. To the best of our knowledge, nothing contradicts setting $c_1 = c_2 = 2$ in the conjecture below. When limiting the scope to fields of characteristic greater than k (degree of the RS code), we are not aware of anything contradicting $c_1 = c_2 = 1$; note that [11, Appendix B] shows these smaller exponents cannot hold for fields of characteristic two.

Conjecture 8.4. *There exist constants c_1, c_2 such that the following statements hold for all $\eta > 0$.*

- *Theorems 1.2, 1.4 and 1.6 hold for proximity parameter $\delta \leq 1 - \rho - \eta$ with error*
$$\epsilon \leq \frac{1}{(\eta\rho)^{c_1}} \cdot \frac{n^{c_2}}{q}.$$
- *Theorem 1.5 holds for proximity parameter $\delta \leq 1 - \rho - \eta$ and parameterized curves of degree l with error*
$$\epsilon \leq \frac{1}{(\eta\rho)^{c_1}} \cdot \frac{(l \cdot n)^{c_2}}{q}.$$

DEEP FRI is another Reed-Solomon Proximity Testing (RPT) protocol that is closely related to FRI (as its name suggests). Introduced in [11], it is slightly less efficient (in terms of prover and verifier complexity) than FRI because it requires more from the prover, making it harder for a malicious prover to cheat. Prior to this work, the

extra complexity of DEEP FRI led to improved soundness, which reaches the Johnson/Guruswami–Sudan bound. But Theorem 8.3 shows that FRI has soundness which also reaches the same bound. Moreover, when approaching the Johnson bound, DEEP FRI requires cubic size fields for the arguments to work, whereas FRI is shown here to require only quadratic size fields. Thus, according to our new understanding, in terms of both complexity and field size, FRI dominates DEEP FRI, even though DEEP FRI demands strictly more from the prover side.[7] This raises an interesting second question: understanding how the techniques developed here may be combined with the techniques of the DEEP FRI protocol to derive better soundness bounds for DEEP FRI and new, improved RPT protocols.

8.2.1 Proof of Lemma 8.2. Recall that the prover sends a function $f^{(i+1)}$ in response to the random choice of $z^{(i)}$. In the FRI QUERY phase, the function $f^{(i+1)}$ will be checked for consistency with $f^{(i)}$. We now introduce a way to keep track of the consistencies and inconsistencies.

Define a sequence of weight functions $\mu^{(i)} : \mathcal{D}^{(i)} \to [0,1]$ and $\nu^{(i)} : \mathcal{D}^{(i)} \to [0,1]$ inductively for $i = 0, \ldots, \mathsf{r}$. For $i = 0$, we assign $\{0,1\}$ weights indicating whether $f^{(0)}(g)$ is computed correctly:

$$\mu^{(0)}(g) = \begin{cases} 1 & \text{if } f^{(0)}(g) = f_0^{(0)}(g) + \sum_{i=1}^{t} x_i f_i^{(0)}(g), \\ 0 & \text{otherwise.} \end{cases}$$

Now, we inductively define an auxiliary weight $\nu^{(i+1)} : \mathcal{D}^{(i+1)} \to [0,1]$. Recall the coset $C_g^{(i)} \subset \mathcal{D}^{(i)}$ from Eq. (8.1). Then

$$\nu^{(i+1)}(g) = \mathbb{E}_{g' \in C_g^{(i)}}[\mu^{(i)}(g')]. \tag{8.6}$$

In words, $\nu^{(i+1)}(g)$ is the expected $\mu^{(i)}$ weight of a member of the coset $C_g^{(i)}$. Finally, we define the weight function $\mu^{(i+1)}$ thus for each $g \in \mathcal{D}^{(i+1)}$:

$$\mu^{(i+1)}(g) = \begin{cases} \nu^{(i+1)}(g) & \text{if } f^{(i+1)}(g) = f^{(i+1)}_{f^{(i)},z^{(i)}}(g), \\ 0 & \text{otherwise.} \end{cases}$$

The key property of the above definition of the $\mu^{(i)}$ is that $\mu^{(i)}(g)$ is a measure of the success probability of the FRI QUERY phase conditioned on g being queried from $f^{(i)}$. This is the underlying reason behind the following claim.

Claim 8.5. *The probability ϵ_{Q} that a single invocation of the batched FRI QUERY accepts $f^{(0)}, \ldots, f^{(\mathsf{r})}$, where $f^{(\mathsf{r})} \in \mathsf{RS}[\mathbb{F}, \mathcal{D}^{(\mathsf{r})}, k^{(\mathsf{r})}]$, satisfies*

$$\epsilon_{\mathsf{Q}} = \mathbb{E}_{g^{(\mathsf{r})} \in \mathcal{D}^{(\mathsf{r})}} \left[\mu^{(\mathsf{r})} \left(g^{(\mathsf{r})} \right) \right].$$

Proof. Recall that the FRI QUERY invocation picks a random sequence $g^{(\mathsf{r})}, \ldots, g^{(0)}$ as described above, where $g^{(i-1)}$ is sampled uniformly at random from $C^{(i-1)}_{g^{(i)}}$. We prove by induction on $i = 0, \ldots, \mathsf{r}$ that

$$\mathbb{E}_{g^{(i)} \in \mathcal{D}^{(i)}} \left[\mu^{(i)} \left(g^{(i)} \right) \right]$$

equals the probability that, when sampling a uniformly random $g^{(i)}$ and generating from it the random sequence $g^{(i-1)} \in C^{(i-1)}_{g^{(i)}}, \ldots, g^{(0)} \in C^{(0)}_{g^{(1)}}$, all tests associated with $g^{(i)}$ and its induced sequence accept.

The base case follows from the definition of $\mu^{(0)}$. For the inductive case, notice $\mu^{(i)}(g^{(i)})$ equals 0 when $f^{(i)}(g^{(i)})$ is not computed correctly as per Eq. (8.2), and otherwise it is the average of the values of $\mu^{(i-1)}$ on

[7] The one remaining virtue of DEEP FRI is that its soundness is closely connected to *a classically studied problem* (the list-decodability of Reed–Solomon codes), and under a simple, plausible conjecture about that problem, it achieves the optimal bound on the required repetition parameter $t = (1 + o(1))\lambda \log \frac{1}{\rho}$.

the coset $C^{(i-1)}_{g^{(i)}} \subseteq \mathcal{D}^{(i-1)}$, which, by the inductive assumption, is the expectation that the tests associated with $g^{(i-1)}, \ldots, g^{(0)}$ are all accepted. □

Proof of Lemma 8.2. In light of Claim 8.5 it suffices to prove that with probability $1 - \epsilon_C$ over the random choices of the verifier,

$$\mathbb{E}_{g \in \mathcal{D}^{(r)}} \left[\mu^{(r)}(g) \right] \leq \alpha^{(0)}(\rho, m). \tag{8.7}$$

We define a sequence of bad events $E^{(0)}, \ldots, E^{(r)}$ and bound the sum of their probabilities by ϵ_C. Assuming none of the bad events occurred, we shall show that Eq. (8.7) holds.

Let $E^{(0)}$ be the event

$$\text{agree}_{\mu^{(0)}} \left(f^{(0)}, V^{(0)} \right) > \alpha^{(0)}(\rho, m).$$

The definition of $\mu^{(0)}$ implies that the event $E^{(0)}$ is

$$\text{agree} \left(f_0^{(0)} + \sum_{i=1}^{t} x_i f_i^{(0)}, V^{(0)} \right) > \alpha^{(0)}(\rho, m) = \max(\alpha, \sqrt{\rho}(1 + 1/2m)).$$

This event depends on $x_1, \ldots, x_t$. By assumption the maximal correlated agreement density of $(f_0^{(0)}, \ldots f_t^{(0)})$ with $V^{(0)}$ is at most α. So Theorem 7.4 (with $\alpha = \alpha^{(0)}(\rho, m)$ and $\mu \equiv 1$, $M = 1$) implies:

$$\Pr_{x_1, \ldots, x_t} \left[E^{(0)} \right] \leq \epsilon, \text{ where } \epsilon = \frac{\left(m + \frac{1}{2}\right)^7}{3\rho^{3/2}} \cdot \frac{n^2}{q}. \tag{8.8}$$

Now fix $i \in \{0, \ldots, r-1\}$. We define $E^{(i+1)}$ to be the event that:

$$\text{agree}_{\nu^{(i+1)}} \left(f^{(i+1)}_{f^{(i)}, z^{(i)}}, V^{(i+1)} \right) > \max \left(\text{agree}_{\mu^{(i)}} \left(f^{(i)}, V^{(i)} \right), \sqrt{\rho}(1 + 1/2m) \right). \tag{8.9}$$

Having fixed $f^{(i)}$ and $\mu^{(i)}$, the event $E^{(i+1)}$ depends on $z^{(i)}$. By definition, we have

$$\text{agree}_{\mu^{(i+1)}} \left(f^{(i+1)}, V^{(i+1)} \right) \leq \text{agree}_{\nu^{(i+1)}} \left(f^{(i+1)}_{f^{(i)}, z^{(i)}}, V^{(i+1)} \right)$$

so when $E^{(i+1)}$ does not hold we conclude from Eq. (8.9) that

$$\text{agree}_{\mu^{(i+1)}} \left(f^{(i+1)}, V^{(i+1)} \right) \leq \max \left(\text{agree}_{\mu^{(i)}} \left(f^{(i)}, V^{(i)} \right), \sqrt{\rho}(1 + 1/2m) \right). \tag{8.10}$$

Let $\alpha = \max \left(\text{agree}_{\mu^{(i)}} \left(f^{(i)}, V^{(i)} \right), \sqrt{\rho}(1 + 1/2m) \right)$. Opening up the definition of $f^{(i+1)}_{f^{(i)}, z^{(i)}}$, we get that $E^{(i+1)}$ is the event that:

$$\text{agree}_{\nu^{(i+1)}} \left(u_0 + z^{(i)} u_1 + \ldots + (z^{(i)})^{l^{(i)}-1} u_{l^{(i)}-1}, V^{(i+1)} \right) > \alpha,$$

where $u_0, \ldots, u_{l^{(i)}-1} : D^{(i+1)} \to \mathbb{F}$ are the functions obtained from $f^{(i)}$ in the definition of the FRI protocol (cf. Claim 8.1). This is precisely the situation addressed by Theorem 7.2. Moreover, $\nu^{(i+1)}$ has a common denominator $M = |\mathcal{D}^{(0)}|/|\mathcal{D}^{(i+1)}|$, so, using the notation there $M \cdot n = |\mathcal{D}^{(0)}|$. So Theorem 7.2 tells us that *if*

$$\Pr_{z^{(i)}} \left[E^{(i+1)} \right] \geq (l^{(i)} - 1) \cdot \left(\epsilon^{(i)} + \frac{2m+1}{\sqrt{\rho}} \cdot \frac{|\mathcal{D}^{(0)}| + 1}{|\mathbb{F}|} \right)$$

then there is an $S \subseteq \mathcal{D}^{(i+1)}$ and codewords $v_0, \ldots, v_{l^{(i)}-1} \in V^{(i+1)}$ such that the u_i and v_i agree on S, and $\nu^{(i+1)}(S) > \alpha$. Here $\epsilon^{(i)} = \frac{|\mathcal{D}^{(i+1)}|^2}{|\mathcal{D}^{(0)}|^2} \epsilon = \frac{\epsilon}{(l^{(0)} \cdots l^{(i)})^2}$ where ϵ is given in Eq. (8.8), since the required probability is quadratic in the size of the target domain $\mathcal{D}^{(i+1)}$. Recall from Eq. (8.4) that the mapping of $f^{(i)}\big|_{C^{(i)}_{g^{(i+1)}}}$ to $\mathbf{u}^{(i)}(g^{(i+1)})$

is the invertible interpolation map. Apply the inverse map, i.e., the evaluation map, to $v_0(g^{(i+1)}), \ldots, v_{l^{(i)}-1}(g^{(i+1)})$ for each $g^{(i+1)} \in \mathcal{D}^{(i+1)}$, to get a function $h^{(i)} : \mathcal{D}^{(i)} \to \mathbb{F}$ that, on $g^{(i)} \in C_{g^{(i+1)}}$, satisfies:

$$h^{(i)}(g^{(i)}) = \sum_{j=0}^{l^{(i)}-1} \left(g^{(i)}\right)^j \cdot v_j\left(g^{(i+1)}\right) = \sum_{j=0}^{l^{(i)}-1} \left(g^{(i)}\right)^j \cdot v_j\left(\left(g^{(i)}\right)^{l^{(i)}}\right).$$

Therefore, since $v_j \in V^{(i+1)}$, we conclude that $h^{(i)} \in V^{(i)}$. Moreover, by definition we have

$$\operatorname{agree}_{\mu^{(i)}}\left(f^{(i)}, V^{(i)}\right) \geq \operatorname{agree}_{\mu^{(i)}}\left(f^{(i)}, h^{(i)}\right) = \nu^{(i+1)}(S) > \alpha,$$

contradicting our definition of α. The equality above arises from the definition of $\mu^{(i)}$, $\nu^{(i+1)}$ and $h^{(i)}$, noticing $\nu^{(i+1)}(g^{(i+1)}) \neq 0$ implies that $h^{(i)}\big|_{C^{(i)}_{g^{(i+1)}}} = f^{(i)}\big|_{C^{(i)}_{g^{(i+1)}}}$.

Thus, if none of the events $E^{(i+1)}$ happen, we deduce via Eq. (8.10) that for all $i \in 0, 1, \ldots, \mathrm{r}-1$:

$$\operatorname{agree}_{\mu^{(i+1)}}(f^{(i+1)}, V^{(i+1)}) \leq \max\left(\operatorname{agree}_{\mu^{(i)}}\left(f^{(i)}, V^{(i)}\right), \sqrt{\rho}(1 + 1/2m)\right).$$

Recalling the definition of ϵ from Eq. (8.8), the probability that $E^{(0)}$ or some $E^{(i+1)}$ happens is bounded by

$$\begin{aligned}
\Pr_{x_1,\ldots,x_t}\left[E^{(0)}\right] + \sum_{i=0}^{\mathrm{r}-1} \Pr_{z^{(i)}}\left[E^{(i+1)}\right] &\leq \epsilon + \sum_{i=0}^{\mathrm{r}-1}(l^{(i)} - 1)\left(\epsilon^{(i)} + \frac{2m+1}{\sqrt{\rho}} \cdot \frac{|\mathcal{D}^{(0)}| + 1}{|\mathbb{F}|}\right) \\
&= \left(1 + \sum_{i=0}^{\mathrm{r}-1} \frac{l^{(i)} - 1}{(l^{(0)} \cdots l^{(i)})^2}\right)\epsilon + \frac{2m+1}{\sqrt{\rho}} \cdot \frac{|\mathcal{D}^{(0)}| + 1}{|\mathbb{F}|} \cdot \sum_{i=0}^{\mathrm{r}-1}(l^{(i)} - 1) \\
&< \frac{3\epsilon}{2} + \frac{2m+1}{\sqrt{\rho}} \cdot \frac{|\mathcal{D}^{(0)}| + 1}{|\mathbb{F}|} \cdot \sum_{i=0}^{\mathrm{r}-1} l^{(i)}.
\end{aligned}$$

Here we bounded

$$\sum_{i=0}^{\mathrm{r}-1} \frac{l^{(i)} - 1}{(l^{(0)} \cdots l^{(i)})^2} = \sum_{i=0}^{\mathrm{r}-1}\left(\frac{1}{(l^{(0)} \cdots l^{(i-1)})^2 l^{(i)}} - \frac{1}{(l^{(0)} \cdots l^{(i)})^2}\right) < \frac{1}{2},$$

which is immediate from $l^{(i)} \geq 2$.

Putting everything together, we get that except on a set with probability strictly less than

$$\frac{3\epsilon}{2} + \frac{2m+1}{\sqrt{\rho}} \cdot \frac{|\mathcal{D}^{(0)}| + 1}{|\mathbb{F}|} \cdot \sum_{i=0}^{\mathrm{r}-1} l^{(i)} = \epsilon_{\mathrm{C}},$$

it holds that

$$\operatorname{agree}_{\mu^{(\mathrm{r})}}(f^{(\mathrm{r})}, V^{(\mathrm{r})}) = \mathbb{E}_{g^{(\mathrm{r})} \in \mathcal{D}^{(\mathrm{r})}}\left[\mu^{(\mathrm{r})}(g^{(\mathrm{r})})\right] \leq \max\left(\alpha, \sqrt{\rho}(1 + 1/2m)\right) = \alpha^{(0)}(\rho, m).$$

The first equality above holds by the assumption that $f^{(\mathrm{r})} \in V^{(\mathrm{r})}$. This completes our proof. □

REFERENCES

[1] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. 2017. Ligero: Lightweight Sublinear Arguments Without a Trusted Setup. In *Proceedings of the 24th ACM Conference on Computer and Communications Security*.

[2] Sanjeev Arora and Madhu Sudan. 2003. Improved Low-Degree Testing and its Applications. *Combinatorica* 23, 3 (2003), 365–426. Preliminary version appeared in STOC '97..

[3] Olivier Bégassat. 2019. *Bivariate polynomial divisibility test of Spielman.* https://mathoverflow.net/questions/348657/bivariate-polynomial-divisibility-test-of-spielman

[4] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. 1988. Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation (Extended Abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC '88)*. 1–10.

[5] Eli Ben-Sasson, Iddo Bentov, Ynon Horesh, and Michael Riabzev. 2018. Fast Reed-Solomon Interactive Oracle Proofs of Proximity. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP)*. https://eccc.weizmann.ac.il/report/2017/134 Available online as Report 134-17 on Electronic Colloquium on Computational Complexity.

[6] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. 2018. Scalable, transparent, and post-quantum secure computational integrity. *IACR Cryptology ePrint Archive* 2018 (2018), 46. http://eprint.iacr.org/2018/046

[7] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. 2019. Scalable Zero Knowledge with No Trusted Setup. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 11694)*, Alexandra Boldyreva and Daniele Micciancio (Eds.). Springer, 701–732. https://doi.org/10.1007/978-3-030-26954-8_23

[8] Eli Ben-Sasson, Alessandro Chiesa, Lior Goldberg, Tom Gur, Michael Riabzev, and Nicholas Spooner. 2019. Linear-Size Constant-Query IOPs for Delegating Computation. In *Theory of Cryptography - 17th International Conference, TCC 2019, Nuremberg, Germany, December 1-5, 2019, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 11892)*, Dennis Hofheinz and Alon Rosen (Eds.). Springer, 494–521. https://doi.org/10.1007/978-3-030-36033-7_19

[9] Eli Ben-Sasson, Alessandro Chiesa, Michael Riabzev, Nicholas Spooner, Madars Virza, and Nicholas P. Ward. 2018. Aurora: Transparent Succinct Arguments for R1CS. *IACR Cryptology ePrint Archive* 2018 (2018), 828. https://eprint.iacr.org/2018/828

[10] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. 2016. Interactive Oracle Proofs. In *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part II*. 31–60. https://doi.org/10.1007/978-3-662-53644-5_2

[11] Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. 2020. DEEP-FRI: Sampling Outside the Box Improves Soundness. In *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA (LIPIcs, Vol. 151)*, Thomas Vidick (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 5:1–5:32. https://doi.org/10.4230/LIPIcs.ITCS.2020.5

[12] Eli Ben-Sasson, Swastik Kopparty, and Shubhangi Saraf. 2018. Worst-Case to Average Case Reductions for the Distance to a Code. In *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*. 24:1–24:23. https://doi.org/10.4230/LIPIcs.CCC.2018.24

[13] Ran Canetti, Ivan Damgård, Stefan Dziembowski, Yuval Ishai, and Tal Malkin. 2004. Adaptive versus Non-Adaptive Security of Multi-Party Protocols. *J. Cryptology* 17, 3 (2004), 153–207. https://doi.org/10.1007/s00145-004-0135-x

[14] Ignacio Cascudo, Ivan Damgård, Bernardo David, Nico Döttling, and Jesper Buus Nielsen. 2016. Rate-1, Linear Time and Additively Homomorphic UC Commitments. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*. 179–207. https://doi.org/10.1007/978-3-662-53015-3_7

[15] David Chaum, Claude Crépeau, and Ivan Damgård. 1988. Multiparty Unconditionally Secure Protocols (Extended Abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*. 11–19. https://doi.org/10.1145/62212.62214

[16] Alessandro Chiesa, Peter Manohar, and Nicholas Spooner. 2019. Succinct Arguments in the Quantum Random Oracle Model. In *Theory of Cryptography*, Dennis Hofheinz and Alon Rosen (Eds.). Springer International Publishing, Cham, 1–29.

[17] Alessandro Chiesa, Dev Ojha, and Nicholas Spooner. 2019. Fractal: Post-Quantum and Transparent Recursive Proofs from Holography. *IACR Cryptology ePrint Archive* 2019 (2019), 1076. https://eprint.iacr.org/2019/1076 To appear in Eurocrypt 2020.

[18] Benny Chor, Shafi Goldwasser, Silvio Micali, and Baruch Awerbuch. 1985. Verifiable Secret Sharing and Achieving Simultaneity in the Presence of Faults (Extended Abstract). In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*. 383–395. https://doi.org/10.1109/SFCS.1985.64

[19] Ivan Damgård and Yuval Ishai. 2006. Scalable Secure Multiparty Computation. In *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings (Lecture Notes in Computer Science, Vol. 4117)*, Cynthia Dwork (Ed.). Springer, 501–520. https://doi.org/10.1007/11818175_30

[20] Matthew K. Franklin and Moti Yung. 1992. Communication Complexity of Secure Computation (Extended Abstract). In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis (Eds.). ACM, 699–710. https://doi.org/10.1145/129712.129780

[21] Peter Gemmell and Madhu Sudan. 1992. Highly Resilient Correctors for Polynomials. *Inform. Process. Lett.* 43 (1992), 169–174.
[22] Venkatesan Guruswami and Madhu Sudan. 1999. Improved decoding of Reed-Solomon and algebraic-geometry codes. *IEEE Trans. Information Theory* 45, 6 (1999), 1757–1767. https://doi.org/10.1109/18.782097
[23] Carmit Hazay, Yuval Ishai, Antonio Marcedone, and Muthuramakrishnan Venkitasubramaniam. 2019. LevioSA: Lightweight Secure Arithmetic Computation. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, Lorenzo Cavallaro, Johannes Kinder, XiaoFeng Wang, and Jonathan Katz (Eds.). ACM, 327–344. https://doi.org/10.1145/3319535.3354258
[24] Martin Hirt and Ueli M. Maurer. 2001. Robustness for Free in Unconditional Multi-party Computation. In *CRYPTO 2001*, Joe Kilian (Ed.). 101–118.
[25] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. 2009. Zero-Knowledge Proofs from Secure Multiparty Computation. *SIAM J. Comput.* 39, 3 (2009), 1121–1152. https://doi.org/10.1137/080725398
[26] Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. 2008. Founding Cryptography on Oblivious Transfer - Efficiently. In *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings (Lecture Notes in Computer Science, Vol. 5157)*, David A. Wagner (Ed.). Springer, 572–591. https://doi.org/10.1007/978-3-540-85174-5_32
[27] Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. 2009. Secure Arithmetic Computation with No Honest Majority. In *Theory of Cryptography, 6th Theory of Cryptography Conference, TCC 2009, San Francisco, CA, USA, March 15-17, 2009. Proceedings (Lecture Notes in Computer Science, Vol. 5444)*, Omer Reingold (Ed.). Springer, 294–314. https://doi.org/10.1007/978-3-642-00457-5_18
[28] Erich Kaltofen. 1985. Polynomial-Time Reductions from Multivariate to Bi- and Univariate Integral Polynomial Factorization. *SIAM J. Comput.* 14, 2 (May 1985), 469–489.
[29] Erich Kaltofen. 1995. Effective Noether Irreducibility Forms and Applications. *J. Comput. Syst. Sci* 50, 2 (1995), 274–295.
[30] Serge Lang. 2002. *Algebra* (3 ed.). Springer.
[31] Alexander Polishchuk and Daniel A. Spielman. 1994. Nearly-linear size holographic proofs. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing (STOC '94)*. 194–203.
[32] Tal Rabin and Michael Ben-Or. 1989. Verifiable Secret Sharing and Multiparty Protocols with Honest Majority (Extended Abstract). In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA*. 73–85. https://doi.org/10.1145/73007.73014
[33] Omer Reingold, Guy N. Rothblum, and Ron D. Rothblum. 2016. Constant-round interactive proofs for delegating computation. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*. 49–62. https://doi.org/10.1145/2897518.2897652
[34] Ron M. Roth and Gilles Zémor. 2018. Personal communication.
[35] Guy N. Rothblum, Salil Vadhan, and Avi Wigderson. 2013. Interactive proofs of proximity: delegating computation in sublinear time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 793–802.
[36] Adi Shamir. 1979. How to Share a Secret. *Commun. ACM* 22, 11 (1979), 612–613.
[37] Daniel A. Spielman. 1995. *Computationally Efficient Error-Correcting Codes and Holographic Proofs.* Ph. D. Dissertation. MIT.

A ALGEBRAIC EXTENSIONS OF $\mathbb{F}_q(Z)$

In this section we develop preliminaries that will be necessary for the proof of Theorem 5.1. In the proof, we will have a trivariate polynomial $Q(X, Y, Z)$, with an irreducible factor $R(X, Y, Z)$. When evaluated at a certain $X = x_0 \in \mathbb{F}_q$, $R(x_0, Y, Z)$ will have an irreducible factor $H(Y, Z)$. For both R and H, irreducible means irreducible in the respective ring of polynomials over $\mathbb{F}_q$. Our goal will be to show that Q has a factor of the form $Y - P(X, Z)$, where P has low X and Z degree, and in fact R is this factor. Considering Q and R as polynomials in Y over $\mathbb{F}_q[X, Z]$, this is equivalent to finding a rational root $P(X, Z)$ of Q, which is also a root of R, and $P(x_0, Z)$ is thus a root of H. To do so, we will instead start by understanding roots of H, not necessarily of the required form or even lying in $\mathbb{F}_q(Z)$, lifting them to roots of R (and Q), and then investigating these lifts to prove that they are indeed of the required form $P(X, Z)$.

We will therefore need to understand the roots of $H(Y, Z)$. Such roots can be realized in $\mathbb{F}_q(Z)[T]/(H(T, Z))$, which is a finite algebraic extension of $\mathbb{F}_q(Z)$. The field will perhaps contain only one root of H, which is sufficient, since H's irreducibility implies all of its roots are equivalent. We will construct this field, in a slightly different way, and also introduce its ring of *regular* elements, in Appendix A.1. In this field, we will need a concept of *weight*, which is an extension of the concept of "degree" from $\mathbb{F}_q(Z)$, and is introduced in Appendix A.2. A major tool that we will use in several ways is of substitution maps from this field to $\mathbb{F}_q$, extending the concept of substituting $Z = z$ in a rational function in $\mathbb{F}_q(Z)$. These substitutions and a useful lemma regarding them are introduced in Appendix A.3. The final tool we need is the lifting of roots of $H(Y, Z)$ to roots of $R(X, Y, Z)$, which is known as Hensel lifting. We describe this process and state an important lemma on the weights of the field elements appearing in this lift in Appendix A.4.

A.1 The algebraic extension and its regular elements

Let $d \geq 1$ be an integer, and let

$$H(Y, Z) = h_0(Z)Y^d + h_1(Z)Y^{d-1} + \cdots + h_d(Z) \in \mathbb{F}_q[Y, Z]$$

be an irreducible bivariate polynomial, expressed as a polynomial in Y over $\mathbb{F}_q[Z]$. The leading coefficient of H will feature prominently throughout the section, and so we introduce the additional notation $W = h_0$, to emphasize its role, and slightly simplify notations.

We wish to understand the field $\mathbb{F}_q(Z)[T]/(H(T, Z))$, of polynomials in T over $\mathbb{F}_q(Z)$ modulo H. The presence of the leading coefficient W means that T is not an integer in this field, and makes the arithmetic modulo H unpleasant to keep track of, with possible emergence of high powers of W in denominators whenever a reduction modulo H is performed. To avoid this, we first define a "monicized" version of H, denoted $\widetilde{H}$, which is a monic, irreducible polynomial in $\mathbb{F}_q[Z][T]$ generating the same field as H:

$$\begin{aligned}\widetilde{H}(T, Z) &= W(Z)^{d-1}H(T/W(Z), Z) \\ &= T^d + h_1(Z)T^{d-1} + h_2(Z)W(Z)T^{d-2} + \cdots + h_d(Z)W(Z)^{d-1}.\end{aligned}$$

We now denote $\mathbb{L} = \mathbb{F}_q(Z)[T]/(\widetilde{H}(T, Z))$, and observe that $Y = \frac{T}{W(Z)}$ is a root of $H(Y, Z)$ in $\mathbb{L}$; this also establishes that $T \mapsto W(Z) \cdot Y$ and $Y \mapsto \frac{T}{W(Z)}$ are isomorphisms between $\mathbb{L}$ and $\mathbb{F}_q(Z)[Y]/(H(Y, Z))$. We say that an element of $\mathbb{L}$ is *regular* if it can be expressed as a polynomial in T with coefficients only in the ring $\mathbb{F}_q[Z]$ instead of the field $\mathbb{F}_q(Z)$; equivalently, if this is true for its canonical form as a polynomial in T of degree less than d. We denote the set of regular elements by $\mathcal{O} = \mathbb{F}_q[Z][T]/(\widetilde{H}(T, Z))$. The regular elements are in fact a ring, and a subring of the ring of integers of $\mathbb{L}$ (we will not be interested in non-regular integers, for our purposes).

A.2 Algebraic weights

Let $D \geq d$ be an upper bound on the total Y, Z degree of H, so that $\deg_Z h_k \leq D + k - d$ for all k. We define a *weight* function Λ on $\mathbb{F}_q[T, Z]$ by assigning $\Lambda(Z) = 1$ and $\Lambda(T) = D + 1 - d$, extended additively to monomials, i.e. $\Lambda(T^a Z^b) = a\Lambda(T) + b\Lambda(Z)$, and the weight of a polynomial is the maximal weight of all monomials appearing in it with non-zero coefficients (with the weight of the 0 polynomial being $-\infty$). Note that Λ is fully additive on $\mathbb{F}_q[T, Z]$, i.e. for any $A, B \in \mathbb{F}_q[T, Z]$, $\Lambda(AB) = \Lambda(A) + \Lambda(B)$. Also note that when restricted to $\mathbb{F}_q[Z]$, $\Lambda = \deg_Z$.

Observe that $\widetilde{H}(T, Z)$ has weight $\Lambda(\widetilde{H}) = d(D + 1 - d) = dD - d(d - 1)$, with the leading monomial being of this exact weight and every other monomial bounded by it. It follows that any simple modulo $\Lambda(\widetilde{H})$ operation of the form

$$T^{d+k} \to -\sum_{i=1}^{d} h_i(Z) W(Z)^{i-1} T^{d+k-i}$$

never increases the weight Λ, and so does complete reduction modulo $\widetilde{H}$.

We now define the weight $\Lambda(\alpha)$ of a regular element $\alpha \in \mathcal{O}$ as the weight of the canonical representative of α with degree less than d, which by the above is also the minimal value of Λ over all representatives of α. It also follows that for any $\alpha, \beta \in \mathcal{O}$, if $A(T, Z), B(T, Z)$ are their canonical representatives, and $C(T, Z) = A(T, Z)B(T, Z) \bmod \widetilde{H}(T, Z)$ is the canonical representative of $\gamma = \alpha\beta$, then

$$\Lambda(\gamma) = \Lambda(C) \leq \Lambda(AB) = \Lambda(A) + \Lambda(B) = \Lambda(\alpha) + \Lambda(\beta).$$

In other words, Λ is sub-additive in $\mathcal{O}$.

A.3 Rational substitutions

Let $z \in \mathbb{F}_q$ be such that $\widetilde{H}(T, z)$ has a rational root $T = t_z$. In other words, $(t_z, z) \in \mathbb{F}_q^2$ is a root of $\widetilde{H}$, with t_z considered as depending on z (in our applications, it will in fact be given as a function of z). For any such root-pair, we define the *substitution* π_z, which is the homomorphism $\pi_z : \mathcal{O} \to \mathbb{F}_q$ given by $\pi_z(Z) = z, \pi_z(T) = t_z$. The homomorphism is well-defined since $\mathcal{O} = \mathbb{F}_q[T, Z]/(\widetilde{H}(T, Z))$ and $\pi_z(\widetilde{H}(T, Z)) = \widetilde{H}(t_z, z) = 0$. The substitution π_z can be extended naturally to any element of $\mathbb{L}$ for which z is not a pole, i.e. elements of the form $\frac{\beta}{C(Z)}$ with $\beta \in \mathcal{O}$ and z not a root of C, by $\pi_z\left(\frac{\beta}{C(Z)}\right) = \frac{\pi_z(\beta)}{C(z)}$.

The following lemma gives an upper bound on the number of substitutions in which an element $\beta \in \mathcal{O}$ can vanish in terms of its weight. It is analogous to the statement that a polynomial of degree at most d which vanishes for more than d evaluations must be the 0 polynomial.

Lemma A.1. *Let $\beta \in \mathcal{O}$ be regular with weight $\Lambda(\beta)$. Let*

$$S_\beta = \{z \in \mathbb{F}_q : \exists t_z \in \mathbb{F}_q, \widetilde{H}(t_z, z) = 0 \text{ and } \pi_z(\beta) = 0\}$$

and suppose $|S_\beta| > d \cdot \Lambda(\beta)$. Then $\beta = 0 \in \mathbb{L}$.

Proof. Let $\beta = \sum_{i=0}^{d-1} \beta_i(Z) T^i \in \mathbb{F}_q[Z][T]$ be the canonical representative, with

$$\deg_Z(\beta_i) = \Lambda_Z(\beta_i) \leq \Lambda(\beta) - i\Lambda(T) = \Lambda(\beta) - i(D + 1 - d).$$

Consider the resultant $R(Z) = \mathrm{res}_T(\beta, \widetilde{H})$. From the Sylvester matrix, or considering weights of roots, we find $\deg_Z R \leq d \cdot \Lambda(\beta) < |S_\beta|$. On the other hand, every $z \in S_\beta$ is a root of R, since $\widetilde{H}(T, z)$ and $\beta(T, z)$ have a common root t_z. It follows that R is identically 0, i.e. β and $\widetilde{H}$ are not coprime, but since $\widetilde{H}$ is irreducible and $\deg_T(\beta) < \deg_T(\widetilde{H})$, we have $\beta = 0$, as claimed. □

A.4 Hensel lifts

Suppose $H(Y,Z)$ as above is a factor of $R_{x_0}(Y,Z) = R(x_0,Y,Z)$, where $x_0 \in \mathbb{F}_q$, and $R(X,Y,Z)$ is irreducible in $\mathbb{F}_q[X,Y,Z]$. Additionally assume that R and R_{x_0} are both separable in Y, which means that they do not have double roots when considered as polynomials in Y, or equivalently, that they are coprime to their Y-derivatives. Note that this applies not only to roots in the fields over which they are defined, but over any extension field as well. Denoting $\alpha_0 = \frac{T}{W} \in \mathbb{L}$, we find $R(x_0,\alpha_0,Z) = 0 \in \mathbb{L}$, or equivalently in $\mathbb{L}[X]$,

$$R(X,\alpha_0,Z) \equiv 0 \pmod{X - x_0}.$$

Since α_0 is a root of the separable polynomial R_{x_0}, it must be a simple root, or equivalently must satisfy $\zeta = \frac{\partial R}{\partial Y}(x_0,\alpha_0,Z) \neq 0$. In other words $\zeta \in \mathbb{L}$ is invertible. This fact allows us to iteratively lift the root α_0 of $R(X,Y,Z) \pmod{X - x_0}$ to a root

$$\alpha_0 + \alpha_1(X - x_0) \quad \text{of} \quad R(X,Y,Z) \pmod{(X - x_0)^2}$$

by solving the equation $R(X,\alpha_0 + \alpha_1(X - x_0),Z) \equiv 0 \pmod{(X - x_0)^2}$, in which α_1 appears with coefficient ζ, after expansion. We then lift again to a root

$$\alpha_0 + \alpha_1(X - x_0) + \alpha_2(X - x_0)^2 \quad \text{of} \quad R(X,Y,Z) \pmod{(X - x_0)^3},$$

and so on. At each step the lifting is unique, and determined by an algebraic equation in which the new α_t appears linearly with the same coefficient ζ.

We obtain an infinite sequence $(\alpha_t)_{t=0}^{\infty}$, such that at each step s the truncated series $\gamma_s = \sum_{t=0}^{s} \alpha_t (X - x_0)^t \in \mathbb{L}[X]$ satisfies

$$R(X,\gamma_s,Z) \equiv 0 \pmod{(X - x_0)^{s+1}}.$$

Equivalently, for the infinite formal power series $\gamma = \sum_{t=0}^{\infty} \alpha_t (X - x_0)^t \in \mathbb{L}[[X - x_0]]$, we have $R(X,\gamma,Z) = 0 \in \mathbb{L}[[X - x_0]]$. Here $\mathbb{L}[[X - x_0]]$ is the ring of formal power series in $X - x_0$ with coefficients in $\mathbb{L}$. This power series γ is the Hensel lift of α_0, and the process by which it is computed is the Hensel lifting. Its existence and uniqueness both follow solely from α_0 being a simple root modulo $X - x_0$.

We make a slight change to the notation, and henceforth d will be the Y-degree of R, and $d_H \leq d$ will denote the Y-degree of H (previously denoted by d). We will also assume that D is an upper bound not only on the total degree of H but also of R. Note that W, the leading coefficient of H, divides the leading coefficient of R_{x_0}, and has weight $\Lambda(W) \leq D - d_H$.

The following claim describes the coefficients α_t appearing in the Hensel lift, and bounds their denominators and the weights of their numerators:

Claim A.2. *α_t is of the form $\frac{\beta_t}{W^{t+1}\xi^{e_t}}$, where*

- *$\xi = W(Z)^{d-2}\zeta \in O$ with $\zeta = \frac{\partial R}{\partial Y}(x_0, \frac{T}{W}, Z) \in \mathbb{L}$, and*
$$\Lambda(\xi) \leq (D-1) + (d-2)\Lambda(W) \leq (d-1)(D - d_H + 1),$$
- *$e_t = \max(0, 2t-1)$, i.e. $e_0 = 0$ and $e_t = 2t-1$ for $t \geq 1$,*
- *$\beta_t \in O$ with*

$$\begin{aligned}\Lambda(\beta_t) &\leq 1 + (t+1)\Lambda(W) + e_t\Lambda(\xi)\\ &\leq 1 + (t+1)(D - d_H) + e_t(d-1)(D - d_H + 1)\\ &= ((d-1)\cdot e_t + t + 1)(D - d_H + 1) - t\\ &\leq ((2t(d-1) + t + 1)(D - d_H + 1) - t < (2t+1)dD.\end{aligned}$$

The claim is proven by straight-forwardly expanding $R(X, \gamma, Z)$ as a series in $X - x_0$, comparing each coefficient to 0, and using induction on t. The existence and uniqueness of the Hensel lift γ, and how each new coefficient α_t is computed from the previous ones, will also follow from the proof.

Proof. We prove the statement by induction on t. For $t = 0$, we have simply $\alpha_0 = \frac{T}{W} \bmod \widetilde{H}$, i.e. $\beta_0 = T \bmod \widetilde{H}$ and indeed $\Lambda(T) = \Lambda(W) + 1$.

Let $R_{ji}(Z)$ be the coefficient of $X^i Y^j$ in $R(X, Y, Z)$, i.e. $R(X, Y, Z) = \sum_{i,j} R_{ji}(Z) X^i Y^j$. We wish to write $R(X, \gamma, Z)$ as a power series in $X - x_0$. A *partition* of t is a sequence of non-negative integers $\lambda = (\lambda_l)_{l \geq 1}$ with $\sum_l l \cdot \lambda_l = t$. Such sequences are non-zero only finitely many times, and we trim any trailing 0s in writing, e.g. $(1) = (1, 0, 0, 0, \dots)$. We also denote $|\lambda| = \sum_{l \geq 1} \lambda_l$. Let $\mathcal{P}(t)$ be the set of partitions of t. For any $t, i_1 \leq t$, and $\lambda \in \mathcal{P}(t - i_1)$, denote

$$\begin{aligned}
A_{i_1,\lambda} &= \sum_{\substack{j_0 \\ j = j_0 + |\lambda|}} \binom{j}{j_0, \lambda_1, \dots, \lambda_l, \dots} \alpha_0^{j_0} \sum_{\substack{i_0 \\ i = i_0 + i_1}} \binom{i}{i_0, i_1} R_{ji}(Z) x_0^{i_0} \\
&= \sum_{\substack{i_0, j_0 \\ i = i_0 + i_1 \\ j = j_0 + |\lambda|}} \binom{i}{i_0, i_1} \binom{j}{j_0, \lambda_1, \dots, \lambda_l, \dots} R_{ji}(Z) x_0^{i_0} \alpha_0^{j_0} \\
&= \binom{|\lambda|}{\lambda_1, \dots, \lambda_l, \dots} \sum_{i,j} \binom{i}{i_1} \binom{j}{|\lambda|} R_{ji}(Z) x_0^{i - i_1} \alpha_0^{j - |\lambda|} \\
&= \binom{|\lambda|}{\lambda_1, \dots, \lambda_l, \dots} \left(\Delta_X^{i_1} \Delta_Y^{|\lambda|} R(X, Y, Z) \right) \Big|_{(x_0, \alpha_0, Z)}
\end{aligned}$$

where the sums are taken only over non-zero terms, i.e. with $i + kj < D_X$, and Δ_V^k is the k-th Hasse derivative with respect to the variable V. In particular, whenever λ is such that $|\lambda| = 1$, we have

$$A_{0,\lambda} = (\Delta_Y R)(x_0, \alpha_0, Z) = \frac{\partial R}{\partial Y}(x_0, \alpha_0, Z) = \zeta.$$

Since the maximal degree of $\alpha_0 = \frac{T}{W}$ appearing in $A_{i_1,\lambda}$ is $d - |\lambda|$, we can generally write $A_{i_1,\lambda} = \frac{B_{i_1,\lambda}}{W^{d-|\lambda|}}$, where $B_{i_1,\lambda} \in \mathcal{O}$ has weight $\Lambda(B_{i_1,\lambda}) = (D - |\lambda|) + (d - |\lambda|)\Lambda(W)$. In the special case $i_1 = 0$, the coefficient of $\alpha_0^{d-|\lambda|}$ is a multiple of $\sum_i R_{di}(Z) x_0^i$, which is the leading coefficient of R_{x_0}, hence divisible by W, and thus we can save a little and write $A_{0,\lambda} = \frac{B_{0,\lambda}}{W^{d-1-|\lambda|}}$, where $B_{0,\lambda} \in \mathcal{O}$ has weight $(D - |\lambda|) + (d - 1 - |\lambda|)\Lambda(W)$. When $|\lambda| = 1$ we then get $\zeta = A_{0,\lambda} = \frac{B_{0,\lambda}}{W^{d-2}} = \frac{\xi}{W^{d-2}}$ where $\xi \in \mathcal{O}$ has weight $(D - 1) + (d - 2)\Lambda(W)$ as stated. To generalize the two cases, we may write $A_{i_1,\lambda} = \frac{B_{i_1,\lambda}}{W^{d - \delta_{i_1,0} - |\lambda|}}$, where $\delta_{i_1,0} = 1$ if $i_1 = 0$ and $\delta_{i_1,0} = 0$ otherwise.

Now, expanding $R(X, \gamma, Z)$, we get

$$\begin{aligned}
0 = R(X, \gamma, Z) &= R\Big(x_0 + (X - x_0), \sum_l \alpha_l (X - x_0)^l, Z\Big) \\
&= \sum_{ij} R_{ji}(Z)(x_0 + (X - x_0))^i \Big(\sum_l \alpha_l (X - x_0)^l \Big)^j \\
&= \sum_{ij} R_{ji}(Z) \Bigg(\sum_{i_0 + i_1 = i} \binom{i}{i_0, i_1} x_0^{i_0} (X - x_0)^{i_1} \Bigg)
\end{aligned}$$

$$\times \sum_{j_0+|\lambda|=j} \binom{j}{j_0,\lambda} \left(\alpha_0^{j_0} \prod_l \alpha_l^{\lambda_l}\right) (X-x_0)^{\sum_l l\cdot\lambda_l}\Bigg)$$
$$= \sum_{t=0}^{\infty} (X-x_0)^t \sum_{\substack{i_1 \\ \lambda \in \mathcal{P}(t-i_1)}} A_{i_1,\lambda} \prod_{l\geq 1} \alpha_l^{\lambda_l}$$

Note that α_t appears for the first time in the term corresponding to $(X-x_0)^t$, and only with $i_1 = 0$ and $\lambda = \lambda^{(t)}$ defined by $\lambda_t^{(t)} = 1$, $\lambda_l^{(t)} = 0$ for $l \neq t$, with the coefficient $A_{0,\lambda^{(t)}} = \zeta$. All other summands in the coefficient of $(X-x_0)^t$ involve only α_l with $l < t$, so we may apply the induction. Comparing the coefficient to 0, we get

$$\begin{aligned}
\alpha_t &= -\frac{1}{\zeta} \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} A_{i_1,\lambda} \prod_{l\geq 1} \alpha_l^{\lambda_l} \\
&= -\frac{W^{d-2}}{\xi} \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} \frac{B_{i_1,\lambda}}{W^{d-\delta_{i_1,0}-|\lambda|}} \prod_{l\geq 1} \left(\frac{\beta_l}{W^{l+1}\xi^{e_l}}\right)^{\lambda_l} \\
&= \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} \frac{B_{i_1,\lambda} \prod_l \beta_l^{\lambda_l}}{W^{2-\delta_{i_1,0}-|\lambda|+\sum_l (l+1)\lambda_l} \xi^{1+\sum_l (2l-1)\lambda_l}} \\
&= \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} \frac{B_{i_1,\lambda} \prod_l \beta_l^{\lambda_l}}{W^{t-i_1-\delta_{i_1,0}+2} \xi^{2t-2i_1-|\lambda|+1}}, \\
&= \frac{1}{W^{t+1}\xi^{2t-1}} \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} W^{i_1+\delta_{i_1,0}-1} \xi^{2i_1+|\lambda|-2} B_{i_1,\lambda} \prod_l \beta_l^{\lambda_l},
\end{aligned}$$

thus we have

$$\beta_t = \sum_{\substack{i_1;\lambda \in \mathcal{P}(t-i_1) \\ \lambda \neq \lambda^{(t)}}} W^{i_1+\delta_{i_1,0}-1} \xi^{2i_1+|\lambda|-2} B_{i_1,\lambda} \prod_l \beta_l^{\lambda_l} \tag{A.1}$$

which is indeed regular, as the W, ξ, the B's and the β's are all regular and the exponents are non-negative: for the exponent of W, it is always the case that $i_1 + \delta_{i_1,0} \geq 1$. For the exponent of ξ, for $i_1 = 0$, every $\lambda \in \mathcal{P}(t)$ with $\lambda \neq \lambda^{(t)}$ indeed has $|\lambda| \geq 2$, and for $i_1 \geq 1$ we even have $2i_1 + |\lambda| - 2 \geq 1$, since $|\lambda| \geq 1$.

The upper bound on the weight of β can be shown by induction using the recursion (A.1), but an easier way to understand it is by considering the weight of α_t: Since $\gamma = \sum_{t=0}^{\infty} \alpha_t (X-x_0)^t$ is a solution to $R(X,Y,Z) = 0$, γ has the same weight as Y; since X, x_0 have weight 0, each α_t also has weight $\Lambda(\alpha_t) = \Lambda(Y) = 1$. Thus

$$\Lambda(\beta_t) = \Lambda(\alpha_t W^{t+1} \xi^{e_t}) \leq 1 + (t+1)\Lambda(W) + e_t \Lambda(\xi)$$

as claimed.

□

B MISCELLANEOUS COMPUTATIONS

B.1 The Y, Z-degree of Q

In this section we prove inequality (5.7) of Claim 5.4, which claims that the total Y, Z-degree of $Q(X, Y, Z)$ is bounded by $\frac{(m+1/2)^3}{6\sqrt{\rho}}\, n$.

The bound on $D_{YZ} = \deg_{Y,Z}(Q)$ comes from minors of the matrix M representing the system of equations defining Q. These equations are all of the form "the (m_X, m_Y)-th derivative of Q vanishes at $(x, w(x))$", for $x \in \mathcal{D}$ and non-negative integers m_X, m_Y with $m_X + m_Y < m$. Computing the derivative[8] and substituting $(x, w(x))$, we see that the (x, m_X, m_Y)-th equation is

$$\Delta_X^{m_X} \Delta_Y^{m_Y} Q(x, w(x, Z), Z) = \sum_{i+k\cdot j<D_X} Q_{ji}(Z) \binom{i}{m_X} \binom{j}{m_Y} x^{i-m_X} (u_0(x) + Z u_1(x))^{j-m_Y} = 0.$$

The coefficient of $Q_{ji}(Z)$ in this equation is therefore

$$\binom{i}{m_X} \binom{j}{m_Y} x^{i-m_X} (u_0(x) + Z u_1(x))^{j-m_Y},$$

and this coefficient appears in the matrix M at row (x, m_X, m_Y) and column ji. Note that as a polynomial in Z, it has degree at most $j - m_Y$, which is the sum of j, which is determined by the column, and $-m_Y$, determined by the row. We thus call j and $-m_Y$ the contributions of the column and the row, respectively, to the Z-degree of the matrix's entry.

Let r be the rank of M, which is bounded from above by the number of rows $\binom{m+1}{2} n$. As in Section 4.3.1, we apply Lemma 4.3 to obtain a non-trivial solution to the system, which assigns to each Q_{ji} either 0 or a determinant of some $r \times r$ submatrix, obtained from fixed $r \times (r+1)$ submatrix by removing the ji-th column. Such a determinant is a sum over products corresponding to permutations, each containing a single entry from each row and each column of the $r \times r$ submatrix. The degree of each such product is thus bounded by the sum of the degree contributions from all columns and rows of the $r \times r$ submatrix, regardless of the permutation, or of all columns and rows of the $r \times (r+1)$ matrix, minus that of the ji-th column. In other words, it is at most $D - j$, where $D = D_C - D_R$ is the sum of the contributions of all columns and (negative) contributions of all rows from the $r \times (r+1)$ matrix. Thus the total Y, Z degree of the monomial $Q_{ji} X^i Y^j$ is at most $(D - j) + j \le D$, and hence the total Y, Z degree of Q is at most D, i.e. $D_{YZ} \le D$.

It remains to bound $D = D_C - D_R$ from above. We do this by simply using $D_R \ge 0$ and bounding D_C by the sum of column contributions of the entire matrix. [9]

Write $s = \left\lfloor \frac{D_X}{k} \right\rfloor$, $t = \left\{ \frac{D_X}{k} \right\} \in [0, 1)$ with $D_X = k(s + t)$. Then

$$\begin{aligned} D_{YZ} = \deg_{Y,Z}(Q) \le D = D_C - D_R \le D_C &\le \sum_{i+kj<D_X} j \\ &= \sum_{j<\frac{D_X}{k}} j(D_X - j\cdot k) = D_X \sum_{j=0}^{s} j - k \sum_{j=0}^{s} j^2 = k(s+t)\frac{s(s+1)}{2} - k\frac{s(s+1)(2s+1)}{6} \end{aligned}$$

[8] We use the Hasse derivatives $\Delta_X^m(X^i) := \binom{i}{m} X^{i-m}$ instead of the regular derivatives to avoid complications due to the characteristic of the field.

[9] It is possible to improve this bound by finding the worst r for which the sum of the largest $r+1$ column contributions minus the sum of the smallest r row contributions is maximal, and computing these sums. We opt for the simpler bound, since the optimal bound requires a more technical and involved computation, and only ends up improving on the simple bound by a small constant factor, with an unpleasant dependence on ρ.

$$= \frac{ks(s+1)(s+3t-1)}{6} = \frac{k}{6}(s^3+3ts^2+(3t-1)s) = \frac{k}{6}((s+t)^3-(1-3t+3t^2)s-t^3)$$
$$= \frac{k}{6}((s+t)^3-(1-t)^3 s-(s+1)t^3) < \frac{k}{6}(s+t)^3 = \frac{D_X^3}{6k^2} \le \frac{(m+\frac{1}{2})^3}{6\sqrt{\rho}}\, n,$$

as claimed in (5.7). □

B.2 The number of variables in the Guruswami–Sudan decoder

Recall that the Guruswami–Sudan decoder searched for a polynomial $Q(X,Y) = \sum_{i,j} Q_{ji} X^i Y^j$, with variable coefficients Q_{ji} coming from the set $\{(i,j) : i,j \ge 0, i+k\cdot j < D_X\}$.

Claim B.1. *The number of variables in the Guruswami–Sudan decoder above is at least*

$$\frac{k}{2}\left(\left(\frac{D_X}{k}+\frac{1}{2}\right)^2 - \frac{1}{4}\right) = \frac{D_X(D_X+k)}{2k}.$$

Proof. For this computation, we may assume D_X is an integer: if is not, then replacing it by $\lceil D_X \rceil$ does not change the definition of the set of indices (since the inequality is strict), and only increases the lower bound we wish to prove. Dividing D_X by k with remainder, we write $D_X = k\cdot a + r$, with $a = \left\lfloor \frac{D_X}{k} \right\rfloor$, $0 \le r < k$. The size of the index set is

$$\sum_{i+k\cdot j<D_X} 1 = \sum_{j=0}^{a} |\{0 \le i < D_X - kj\}| = \sum_{j=0}^{a} (D_X - kj) = (a+1)D_X - k\frac{a(a+1)}{2}$$
$$= (a+1)\left(ka + r - \frac{ka}{2}\right) = \frac{(ka+k)(ka+2r)}{2k} \ge \frac{(ka+r)(ka+r+k)}{2k}$$
$$= \frac{D_X(D_X+k)}{2k},$$

where the inequality $(ka+k)(ka+2r) > (ka+r)(ka+r+k)$, is equivalent after expansion to $2rk > r(r+k) \ge r^2+rk$, or simply $r(k-r) \ge 0$, which follows from $0 \le r < k$.

□

C THE INSEPARABLE FACTOR CASE IN THE LIST DECODING REGIME

Recall that in Section 5.2.3, we had assumed that in the decomposition

$$Q(X,Y,Z) = C(X,Z)\prod_i R_i(X, Y^{p^{f_i}}, Z)^{e_i},$$

the factors $R_i(X, Y^{p^{f_i}}, Z)$ were all separable, i.e. $f_i = 0$. This assumption was in fact necessary only for the factor R on which we focused in Section 5.2.4. We now consider the case where $f = f_i > 0$, and our relevant factor is of the form $R(X, Y^{\mathfrak{p}}, Z)$, where $\mathfrak{p} = p^f$ and $R(X, \widetilde{Y}, Z)$ is separable and irreducible in $\widetilde{Y}$. Note that we still have that $R(x_0, \widetilde{Y}, Z)$ is separable. The elements of $S_{x_0,R,H}$ now satisfy $Y - P_z(X) \mid R(X, Y^{\mathfrak{p}}, Z)$, and equivalently $\widetilde{Y} = P_z(X)^{\mathfrak{p}}$ is a root of $R(X, \widetilde{Y}, Z)$ and $Y^{\mathfrak{p}} - P_z(X)^{\mathfrak{p}} \mid R(X, Y^{\mathfrak{p}}, Z)$. Similarly, $\widetilde{Y} = P_z(x_0)^{\mathfrak{p}}$ is a root of the irreducible factor $H(\widetilde{Y}, Z)$ of $R(x_0, \widetilde{Y}, Z)$ — and a simple root of both, since they are separable.

Note that $\mathfrak{p} \le \deg_Y(R(X, Y^{\mathfrak{p}}, Z)) \le \deg_Y(Q(X,Y,Z)) = D_Y$ and therefore

$$\deg_X(P_z(X)^{\mathfrak{p}}) \le p^f k \le kD_Y < D_X.$$

We construct the field $\mathbb{L}$ for the polynomial $H(\widetilde{Y}, Z)$ exactly as before, but noting that $\widetilde{Y} = Y^{\mathfrak{p}}$ will have weight $\mathfrak{p}$, which correspondingly affects the weight of $T = W(Z)\widetilde{Y} = W(Z)Y^{\mathfrak{p}}$, now defined as $D - (d_H - 1)\mathfrak{p}$. Note that the upper bound on d_H, d (the $\widetilde{Y}$ degrees of R, H) is also changed, and is now $\frac{D_Y}{\mathfrak{p}}$ instead of simply D_Y. With these adaptations in mind, we perform the Hensel lift, lifting the root $\alpha_0 = \frac{T}{W} \in \mathbb{L}$ of $R(x_0, \widetilde{Y}, Z)$ to the power series root $\gamma \in \mathbb{L}[[X - x_0]]$ of $R(X, \widetilde{Y}, Z)$. Claim A.2 still holds and gives us

$$\begin{aligned}\Lambda(\xi) &\leq (D - \mathfrak{p}) + (d-2)\Lambda(W) \leq (d-1)(D - (d_H - 1)\mathfrak{p}),\\ \Lambda(\beta_t) &\leq \mathfrak{p} + (t+1)\Lambda(W) + e_t\Lambda(\xi)\\ &\leq ((d-1)\cdot e_t + t + 1)(D - (d_H - 1)\mathfrak{p}) - \mathfrak{p}t < (2t+1)dD.\end{aligned}$$

As in Section 5.2.6, the substitution $\pi_z(\gamma)$ for $z \in S'$ is a root of $R(X, \widetilde{Y}, z)$ which is the lift of the simple root $\widetilde{Y} = P_z(x_0)^{\mathfrak{p}}$ of $R(x_0, \widetilde{Y}, Z)$. Since $P_z(X)^{\mathfrak{p}}$ is also of this form, by the uniqueness of the lifting we get $\pi_z(\gamma) = P_z(X)^{\mathfrak{p}}$, and in particular $\pi_z(\alpha_t) = 0$ and $\pi_z(\beta_t) = 0$ for all t except for $0 \leq t \leq \mathfrak{p}k < D_X$ which are divisible by $\mathfrak{p}$. As before, we have $|S'| \geq d_H\Lambda(\beta_t)$ for all $t < D_X$: the right hand side is bounded from above by $\frac{D_Y^2 D_{YZ}(2D_X - 1)}{\mathfrak{p}^2}$, which only decreases as $\mathfrak{p}$ increases. It thus follows that $\beta_t = 0$ and $\alpha_t = 0$ for all $t < D_X$ except for those which are at most $\mathfrak{p}k$ and divisible by $\mathfrak{p}$. In other words we have

$$\gamma_{D_X - 1} = \sum_{t=0}^{k} \alpha_{\mathfrak{p}t}(X - x_0)^{\mathfrak{p}t}.$$

Out next goal is to show that γ_{D_X-1} is a $\mathfrak{p}$-th power of a polynomial of degree k (and later that $\gamma = \gamma_{D_X-1}$, and that this polynomial is in fact in $\mathbb{F}_q[Z][X - x_0]$, with coefficients linear in Z). This part did not appear in Section 5, as it is trivial for $\mathfrak{p} = 1$. This polynomial should naturally be the $\mathfrak{p}$-th root of γ_{D_X-1} — but in order to construct such roots, we will need some more preliminaries about the field in which they live.

Let $\sigma : \mathbb{F}_q \to \mathbb{F}_q$ be the automorphism mapping each element a to its unique $\mathfrak{p}$-th root $\sigma(a) = a^{1/\mathfrak{p}}$. Let $\widehat{\mathbb{L}}$ be the inseparable algebraic field extension of $\mathbb{L}$ with elements $\widehat{T}, \widehat{Z}$ satisfying $\widehat{T}^{\mathfrak{p}} - T = \widehat{Z}^{\mathfrak{p}} - Z = 0$; equivalently, $\widehat{T} = T^{1/\mathfrak{p}}$ and $\widehat{Z} = Z^{1/\mathfrak{p}}$. Note that $\widehat{\mathbb{L}}$ can also be defined directly as the field $\mathbb{F}_q(\widehat{Z})[\widehat{T}]/(\widehat{H}(\widehat{T}, \widehat{Z}))$, where $\widehat{H} = \sigma(\widetilde{H})$ is the (irreducible) polynomial obtained by applying σ to the coefficients of $\widetilde{H}$, which satisfies

$$\widehat{H}(\widehat{T}, \widehat{Z})^{\mathfrak{p}} = \widetilde{H}\left(\widehat{T}^{\mathfrak{p}}, \widehat{Z}^{\mathfrak{p}}\right) = \widetilde{H}(T, Z).$$

Since the monomials of $\widehat{H}$ have the same $\widehat{T}$- and $\widehat{Z}$-degrees as the T- and Z-degrees of $\widetilde{H}$, a weight $\widehat{\Lambda}$ can be defined for regular elements in $\widehat{\mathbb{L}}$ in exactly the same way as in $\mathbb{L}$. Additionally, σ can be extended to a map $\widehat{\sigma} : \mathbb{L} \to \widehat{\mathbb{L}}$ satisfying $\widehat{\sigma}(\alpha)^{\mathfrak{p}} = \alpha$ for all $\alpha \in \mathbb{L}$ by defining $\widehat{\sigma}(T) = \widehat{T}$ and $\widehat{\sigma}(Z) = \widehat{Z}$. Note that $\Lambda(\beta) = \widehat{\Lambda}(\widehat{\sigma}(\beta))$ for any $\beta \in \mathcal{O}$, since $\widehat{\sigma}$ preserves degrees. The substitution maps π_z for $z \in S'$ can also be extended to $\widehat{\mathbb{L}}$ by setting $\pi_z(\widehat{Z}) = \sigma(z) = z^{1/\mathfrak{p}}$ and $\pi_z(\widehat{T}) = \sigma(t_z) = t_z^{1/\mathfrak{p}} = \widehat{W}(z^{1/\mathfrak{p}})P_z(x_0)$, where $\widehat{W} = \sigma(W)$.

Define

$$\widehat{\gamma} = \sum_{t=0}^{k} \widehat{\sigma}(\alpha_{\mathfrak{p}t})(X - x_0)^t \in \widehat{\mathbb{L}}[X - x_0]$$

which indeed satisfies $\gamma_{D_X-1} = \widehat{\gamma}^{\mathfrak{p}}$, and therefore $R(X, \widehat{\gamma}^{\mathfrak{p}}, Z) \equiv 0 \pmod{(X - x_0)^{D_X}}$. Since $R(X, Y^{\mathfrak{p}}, Z)$ is a divisor of $Q(X, Y, Z)$, it follows as before that $\deg_X R(X, \widehat{\gamma}^{\mathfrak{p}}, Z) < D_X$, and therefore $R(X, \widehat{\gamma}^{\mathfrak{p}}, Z) = 0$ identically, and $\gamma = \gamma_{D_X-1} = \widehat{\gamma}^{\mathfrak{p}}$ by the uniqueness of the lifting. Furthermore, for every $z \in S'$, since $\pi_z(\widehat{\gamma})^{\mathfrak{p}} = \pi_z(\widehat{\gamma}^{\mathfrak{p}}) = \pi_z(\gamma) = P_z(X)^{\mathfrak{p}}$, we also have $\pi_z(\widehat{\gamma}) = P_z(X)$.

As in Section 5.2.7, our next goal is now to show that the coefficients $\widehat{\alpha}_t = \widehat{\sigma}(\alpha_{\mathfrak{p}t})$ of $\widehat{\gamma}$ are linear polynomials in $\mathbb{F}_q[Z]$, rather than general elements of $\widehat{\mathbb{L}}$. This is done in exactly the same way, by comparing the values of

$\pi_z(\widehat{\gamma}(x))$ and $w(x,z)$ at every $x = x_j \in \mathcal{D}_{\text{top}}$ and $z \in S'_x$, deducing that $\widehat{\gamma}(x)$ and $w(x,Z)$ must be equal in $\widehat{\mathbb{L}}$ since

$$d_H\widehat{\Lambda}(\widehat{\beta}_k) = d_H\widehat{\Lambda}(\widehat{\sigma}(\beta_{\mathfrak{p}k})) = d_H\Lambda(\beta_{\mathfrak{p}k}) \le d_H(2\mathfrak{p}k+1)dD < \frac{(2k+1)D_Y^2 D_{YZ}}{\mathfrak{p}}$$
$$< (2k+1)D_Y^2 D_{YZ} < |S'_x|,$$

Finally, having shown that $\widehat{\gamma} \in \mathbb{F}_q[X,Z]$, the arguments of Section 5.2.8 can be applied without any further changes, only with $\widehat{\gamma}$ in the role γ, concluding the proof.

D PROOF OF LEMMA 4.4

It has been recently discovered that there is a subtle flaw in the proof of [37, Lemma 4.2.18], and indeed, there are counter examples to the Lemma as originally stated (cf. [3]). We thank Ronald Cramer and Jade Nardi for pointing out the precise flaw and suggesting possible fixes; the fix chosen and described in this paper was independently discovered by Ronald Cramer.

The main difference between the version of the lemma used in this paper and the original version is the condition that the quotients $P_{Z,x}$ and $P_{X,z}$ have bounded degrees, with the bound matching the difference in the degrees of A and B. This extra assumption suffices to make the original proof valid, and is fortunately satisfied in our application.

For completeness, we provide here a proof of the lemma, essentially following [37, Lemma 4.2.18]. During the proof, we will make several reductions to smaller and simpler cases by changing the a, b, n variables and the polynomials A, B, but note that none of the reductions will change d_X, d_Z or $\frac{B(X,Z)}{A(X,Z)}$.

Clearly, we may assume B is not the zero polynomial (otherwise, the results are trivial). It then follows that A is also not the zero polynomial (otherwise, $B(x,Z) = 0$ for $n_X > b_X \ge \deg_X(B)$ values of x, so B is identically 0). We may also assume that either $\deg_X(A) = a_X$ or $\deg_X(B) = b_X$. If not, then we may decrease a_X and b_X by 1 without changing $d_X = b_X - a_X$ or any other assumption (in particular, the LHS of ($\star$) strictly decreases), and repeat until at least one equality holds.

In fact, once we are done subtracting, we will necessarily have $\deg_X(A) = a_X$. This fact follows from the bound on $\deg_X(P_{X,z})$ as well as the inequality $b_Z < n_Z$ (which follows from ($\star$)). We omit the full details, as this fact will not be used in the proof.

Similarly, by reducing a_Z and b_Z if necessary, we may assume $\deg_Z(A) = a_Z$ (or $\deg_Z(B) = b_Z$).

Next, let $G(X,Z) \in \mathbb{F}[X,Z]$ be the greatest common divisor of A and B; we aim to prove that $G = A$. Let $g_X = \deg_X(G)$, $g_Z = \deg_Z(G)$, and replace both A and B with the quotients A/G and B/G. The degrees $\deg_X(A)$, $\deg_X(B)$ and their bounds a_X, b_X decrease by exactly g_X, and $\deg_Z(A)$, $\deg_Z(B)$, a_Z, b_Z decrease by g_Z; it follows that d_X, d_Z are unchanged, and neither is the quotient $\frac{B}{A} = \frac{B/G}{A/G}$. The number of good substitutions n_X also decreases by (at most) g_X, and similarly n_Z decreases by at most g_Z: indeed, at most g_X of the $x \in \mathbb{F}$ can satisfy $G(x,Z) = 0$, and in any of at least $n_X - g_X$ good substitution with $G(x,Z) \ne 0$, we can divide both sides of $B(x,Z) = P_{Z,x}(Z)A(x,Z)$ by $G(x,Z)$ to obtain $\frac{B}{G}(x,Z) = P_{Z,x}(Z)\frac{A}{G}(x,Z)$, and the assumption $\deg_Z(P_{Z,x}) \le d_Z$ isn't affected. Since $\frac{b_X}{n_X} < 1$ and $g_X \ge 0$, we will have $\frac{b_X - g_X}{n_X - g_X} \le \frac{b_X}{n_X}$ (with equality only if $g_X = 0$), and thus

$$\frac{b_X - g_X}{n_X - g_X} + \frac{b_Z - g_Z}{n_Z - g_Z} \le \frac{b_X}{n_X} + \frac{b_Z}{n_Z} < 1,$$

and it follows that assumption ($\star$) still holds.

We have seen that after dividing by the GCD, the polynomials still satisfy all of the Lemma's assumptions. Additionally, they are now coprime. We will prove that in this case we must have $a_X = a_Z = 0$, i.e. A is a constant function.

Write $A(X,Z) = A_0(Z) + A_1(Z)X + \cdots A_{a_X}(Z)X^{a_X}$, and similarly $B(X,Z) = B_0(Z) + B_1(Z)X + \cdots B_{b_X}(Z)X^{b_X}$. Note that at least one of A_{a_X}, B_{b_X} is non-zero by the early assumption on equality of degrees. Let

$$M_X(Z) = \mathrm{Syl}_X(A,B) = \begin{pmatrix} B_{b_X}(Z) & \cdots & \cdots & B_1(Z) & B_0(Z) & & \\ & \ddots & \cdots & \cdots & \ddots & \ddots & \\ & & B_{b_X}(Z) & \cdots & \cdots & B_1(Z) & B_0(Z) \\ A_{a_X}(Z) & \cdots & A_1(Z) & A_0(Z) & & & \\ & A_{a_X}(Z) & \cdots & A_1(Z) & A_0(Z) & & \\ & & \ddots & \cdots & \ddots & \ddots & \\ & & & A_{a_X}(Z) & \cdots & A_1(Z) & A_0(Z) \end{pmatrix}$$

be the Sylvester matrix of A and B, considered as polynomials of degrees a_X and b_X in X over $\mathbb{F}[Z]$. It is a square matrix of order $a_X + b_X$, with the first a_X rows being shifts of the coefficient vector of B, and the last b_X rows being shifts of the coefficients of A.

Let $R_X(Z) = \det M_X(Z)$ be the determinant of the Sylvester matrix, which is not identically 0, since the polynomials are coprime[10]. By the bounds on the Z-degrees of A_i, B_j, it is immediate to see that $\deg_Z(R_X) \le a_X b_Z + a_Z b_X$. We will denote $D \coloneqq a_X b_Z + a_Z b_X$, and note that the expression is symmetric in X and Z.

For any of the n_Z good values of $z \in \mathbb{F}$, the fact that $B(X,z) = P_{X,z}(X)A(X,z)$ with $\deg_X(P) \le d_X = b_X - a_X$ implies that each of the first a_X rows of $M_X(z)$, corresponding to the coefficients of $B(X,z)$, is a linear combination of (at most) $d_X + 1$ of the b_X rows corresponding to coefficients of $A(X,z)$. It follows that the last b_X rows span the whole row space, i.e. $\mathrm{rk}\, M_X(z) \le b_X$, from which it follows that $R_X, R'_X, \ldots, R_X^{(a_X-1)}$ all vanish at $Z = z$, i.e. R_X has a zero of multiplicity at least a_X at every such z (see also [37, Proposition 4.2.17]). Note that the assumption $\deg_X(P_{X,z}) \le d_X$ is critical: if the degree were larger, than the first $\deg_X(P_{X,z}) - d_X$ of the B rows would not have been spanned by the A rows.

We have obtained that R_X is a non-vanishing polynomial of degree at most D, with at least n_Z roots of multiplicity a_X each, thus $a_X n_Z \le D$. Repeating the argument with the roles of X and Z switched, we find $a_Z n_X \le D$ as well. Thus

$$D = a_X b_Z + a_Z b_X = \frac{b_Z}{n_Z} a_X n_Z + \frac{b_X}{n_X} a_Z n_X \le \frac{b_X}{n_X} D + \frac{b_Z}{n_Z} D = \left(\frac{b_X}{n_X} + \frac{b_Z}{n_Z}\right) D.$$

However, ($\star$) yields $\frac{b_X}{n_X} + \frac{b_Z}{n_Z} < 1$, and thus the obtained inequality is only possible for $D = 0$. Since $n_X, n_Z > 0$, and $a_X n_Z, a_Z n_X \le D = 0$, we must have $a_X = a_Z = 0$, as claimed.

From here the rest of the claim is immediate: $a_X = a_Z = 0$ imply that A is a constant polynomial, so clearly $A \mid B$, and furthermore the quotient $P = B/A$ is proportional to B, thus $\deg_X(P) = \deg_X(B) \le b_X = b_X - a_X = d_X$, and similarly $\deg_Z(P) \le d_Z$. Since d_X, d_Z and $P = B/A$ did not change during our reduction steps, the same results hold for the original polynomials and quotient.

For all n_Z substitutions $z \in \mathbb{F}$, we have $P(X,z)A(X,z) = B(X,z) = P_{X,z}(X)A(X,z)$, and if $A(X,z) \neq 0$ then we can cancel it to obtain $P(X,z) = P_{X,z}(X)$. At most $\deg_Z(A) \le a_Z$ values of z satisfy $A(X,z) = 0$, hence for at least $n_Z - a_Z$ values we have $P(X,z) = P_{X,z}(X)$, and similarly for values of x, as claimed.

□

[10]This polynomial is almost, but not quite, the resultant of A and B, which is non-zero iff they are coprime; the difference comes from the possibility that A, B are not exactly of X-degrees a_X, b_X, and in fact $R_X(Z)$ is equal to the resultant times $A_{a_X}(Z)^{b_X - \deg_X(B)} B_{b_X}(Z)^{a_X - \deg_X(A)}$. The fact that at least one of $\deg_X(A) = a_X$ or $\deg_X(B) = b_X$ holds is equivalent to this extra factor being non-vanishing. In case one of the inequalities is not exact, it is important for the proof *not* to modify the degrees and to keep the extra factor, to get the right order of vanishing.

E PROOF OF CLAIM 8.1

Proof of Claim 8.1. By assumption $f^{(i)}$ is the evaluation of a polynomial $P(X)$ of degree strictly less than $k^{(i)}$ and $k^{(i)}$ is an integral power of 2. Recall $k^{(i+1)} = \frac{k^{(i)}+1}{2} - 1$ noticing $k^{(i+1)} + 1$ is an integral power of 2. Let

$$Q(X, Y) = P(X) \bmod Y - X^{l^{(i)}}.$$

By definition $\deg_X(Q) < l^{(i)}$ and $\deg_Y(Q) < k^{(i+1)}$. We claim $f^{(i+1)}_{f^{(i)},z^{(i)}}$ is the evaluation of the polynomial $Q(z^{(i)}, Y) \in \mathbb{F}[Y]$ on $\mathcal{D}^{(i+1)}$. To see this recall that $M^{(i)}_g$ is the interpolation map over $f^{(i)}\big|_{C^{(i)}_g}$ and so

$$M^{(i)}_g \cdot f^{(i)}\Big|_{C^{(i)}_g} = Q(X, g).$$

Hence, using Eq. (8.2), we have

$$f^{(i+1)}_{f^{(i)},z^{(i)}}(g) = \left(\mathbf{z}^{(i)}\right)^{\top} \cdot M^{(i)}_g \cdot f^{(i)}\Big|_{C^{(i)}_g} = Q(z^{(i)}, g).$$

So $f^{(i+1)}_{f^{(i)},z^{(i)}}$ is the evaluation of the polynomial $Q(z^{(i)}, Y)$, which has degree less than $k^{(i+1)}$, on the domain $\mathcal{D}^{(i+1)}$. This completes the proof. □