On Reed–Solomon Proximity Gaps Conjectures

Elizabeth Crites and Alistair Stewart

Web3 Foundation
firstname@web3.foundation

December 19, 2025

Abstract. We disprove a range of conjectures for Reed-Solomon codes underpinning the security and efficiency of many modern proof systems, including SNARKs based on FRI (Ben-Sasson-Bentov-Horesh-Riabzev, ICALP'18), DEEP-FRI (Ben-Sasson-Goldberg-Kopparty-Saraf, ITCS'20), STIR (Arnon-Chiesa-Fenzi-Yogev, CRYPTO'24), and WHIR (Arnon-Chiesa-Fenzi-Yogev, preprint). Concretely, we prove that the following conjectures are false:

1. The correlated agreement up-to-capacity conjecture of Ben-Sasson-Carmon-Ishai-Kopparty-Saraf (J. ACM'23),
2. The mutual correlated agreement up-to-capacity conjecture of WHIR,
3. The list-decodability up-to-capacity conjecture of DEEP-FRI, which follows from existing results in the literature.

We then propose minimal modifications to these conjectures up to the list-decoding capacity bound.

Our second main contribution is a proof that correlated agreement with small enough error probability implies list decoding of Reed-Solomon codes. Thus, any future positive results on our correlated agreement conjectures with small enough error probability would imply similar results in classical list decoding. A reduction from proximity gaps to list-decodability was heretofore a natural open problem.

1 Introduction

A proximity gap is a property of a linear error-correcting code whereby each affine subspace of words is almost entirely "close" to the code (by Hamming distance) or almost entirely "far" from it, with no in-between.

Proximity gaps are central to the security and efficiency of hash-based succinct non-interactive arguments of knowledge (SNARKs), such those based on FRI [BBHR18], DEEP-FRI [BGKS20], STIR [ACFY24a], and WHIR [ACFY24b]. A gap ensures that if a prover claims to be using a valid Reed-Solomon code, they cannot cheat, as a verifier can detect if a claimed codeword is not from the code. Moreover, a gap guarantees that if a *linear combination* is close to the code, then the original words are also very likely to be close to the code. This property enables exceptionally efficient batch verification of multiple statements simultaneously, reducing the computational overhead of these proof systems dramatically. However, this comes with an important caveat: the soundness of this approach

hinges critically on proximity gaps whose existence for the relevant regimes has only been conjectured. These conjectures are the focus of the prize, in particular those on correlated agreement, the stronger mutual correlated agreement, and list-decodability. Almost all of the nearly two dozen zero-knowledge virtual machines (zkVMs) tracked by EthProofs assume these conjectures. The million-dollar question is whether these hyper-optimized batch verification techniques can be pushed right up to the information-theoretic capacity bound, without sacrificing soundness.

We resolve up-to-capacity proximity gaps conjectures for Reed–Solomon codes in the negative. Concretely, we prove that the correlated agreement, mutual correlated agreement, and list-decodability conjectures up to capacity are all false.

Correlated Agreement. Ben-Sasson, Carmon, Ishai, Kopparty, and Saraf [BCI+23] introduce the notion of correlated agreement for Reed–Solomon codes. A set of codewords $u_0, \ldots, u_\ell$ have δ-correlated agreement with a code C if there exists a sufficiently large subdomain D' of the code's domain D, $|D| = n$, and codewords $v_0, \ldots, v_\ell$ such that $|D'|/n \geq 1 - \delta$ and $u_i = v_i$ on all of D' (Definition 2). (We provide background on codes in Section 2.) Correlated agreement requires that $u_0, \ldots, u_\ell$ (usually within an affine subspace) all agree with their respective codewords $v_0, \ldots, v_\ell$ on the same large subdomain. The existence of a proximity gap means that an affine subspace either entirely consists of words that are close to the code (with correlated agreement), or almost no words are close. We show that Conjecture 8.4 from [BCI+23] (Section 3.2) on correlated agreement of Reed-Solomon codes does not hold up to capacity ($\delta < 1 - \rho$ for rate ρ); rather, it fails beyond the list-decoding capacity bound ($\delta < 1 - H_q(\rho)$, where H_q is the q-ary entropy function; see Section 3.1). The soundness of FRI is analyzed under the assumption of correlated agreement.

Mutual Correlated Agreement. WHIR contains Conjecture 8.4 from [BCI+23] adapted to the stronger mutual correlated agreement setting (Definition 7, Conjecture 4.12, Section 3.2). Garreta, Mohnblatt, and Wagner [GMW25] present a simplified proof of soundness for FRI based on mutual correlated agreement.

List-Decodability. Instead of correlated agreement assumptions, DEEP-FRI introduces a conjecture on the classical list-decodability of Reed–Solomon codes (Conjecture 2.3, Section 3.1). List-decodability asks to find all codewords that are δ-close to a codeword (Definition 1). Conjecture 5.6 from STIR combines the conjectures from [BCI+23] and DEEP-FRI.

We observe that the failure of the list-decodability up-to-capacity conjecture (Conjecture 2.3) follows from the classical result of Elias [Eli57].

This leads to our second main contribution: a proof that correlated agreement with small enough error probability implies list decoding of Reed-Solomon codes. Thus, any future positive results on our correlated agreement conjectures with small enough error probability would imply similar results in classical list decoding. A reduction from proximity gaps to list-decodability was heretofore a natural open problem.

We then propose minimally modified conjectures for list-decodability, correlated agreement, and mutual correlated agreement up to the list-decoding capacity bound. Practically speaking, our conjectures represent the best standing conjectures to date, and deployed proof systems adjusting parameters to fall within the list-decoding capacity regime would incur a loss of $1/\log_2 q$ in the error rate, which is only $1/31$ even for the smallest field sizes used in practice.

Related work. Our work was inspired by the attack of Crites and Stewart [CS25] for threshold Schnorr signatures with a Shamir secret-shared key. The connection between Shamir secret sharing [Sha79] and Reed-Solomon codes was first observed by McEliece and Sarwate [MS81]. The Schnorr signature setting assumes much larger fields than those used in FRI-based proof systems.

We refer to [BCI+23] for a comprehensive treatment of proximity gaps for Reed-Solomon codes. Our work builds upon the fundamental research of countless coding theorists, mathematicians, cryptographers, and complexity theorists.

Proximity gaps are false up to capacity: a high-level overview. To begin with, observe that there are q^k polynomials of degree at most $k-1$, so there are q^k codewords in the Reed-Solomon code of rate $\rho = k/n$. Each of these is at distance exactly f from $\binom{n}{f}$ codewords. For a random codeword v, the expected number of sets A of size $n-f$ such that v agrees with a polynomial of degree at most $k-1$ on A is exactly $\binom{n}{f}/q^k$. Let the random variable X be the number of such sets.

We want to show that the probability p that v is distance of at most f from the code is relatively large (e.g., non-negligible or close to 1); that is, that there exists such an agreement set A, i.e., that the number of such A, X, is at least 1. The expectation of X being large does not imply this, of course; there might be some v that work with many A, e.g., when the distance is much smaller than f so perhaps the distribution of X has a heavy tail and not a high probability of being over its expectation. To rule this out, we compute the variance of X and show it is not too large and lower bounded by p via Cantelli's inequality.

Now, to attain a proximity gap failure, consider a line between a word u that is farther than f from the code, e.g., a deep hole of the code, and a uniformly random v. For any non-zero λ and a uniformly random v, $(1-\lambda)u + \lambda v$ is a uniformly random codeword, so it has probability at least p of being at most distance f to the code. By linearity of expectation, the expected number of words of distance at most f from a random point on the line is at least $(q-1)p$. Some concrete v gives at least this expectation, and for high enough p, the proximity gap fails. We capture this formally in Theorem 1.

For an adversary to prove a false statement in a SNARK, they would make one of their supposedly degree at most $k-1$ polynomial commitments a random string in order to hide another one which commits to a higher-degree polynomial than claimed. Then, for a random linear combination , there probably exists a degree at most $k-1$ polynomial that is close. However, the adversary would need to find such a polynomial to mount an attack. They might not be able to implement list decoding because, beyond the list-decoding capacity bound, the list of close codewords might be exponentially large, ruling out the existence of

efficient decoding algorithms. The easier problem of returning *one* out of the possibly many codewords within distance f, if one exists, is the bounded distance decoding problem [RS60]. Bounded distance decoding, indeed list decoding, is known to be feasible up the Johnson bound ($\delta < 1 - \sqrt{\rho}$), but it is an open question whether it is feasible beyond this bound. (Open Question 12.2.1 of [GRS14] is this for list decoding.) The bounded distance decoding problem is shown to be NP-hard by Guruswami and Vardy [GV05] for $n - k - f = 1$ which was extended to slightly lower f by Gandikota, Ghazi, and Grigorescu [GGG18]. It has also been shown to be hard for discrete log in an extension field for the region of our attack $\binom{n}{f} \geq q^{n-k-f}$ [CW04]. However, none of these results apply to the fields typically used in SNARKs, i.e., cryptographically small fields. The reduction in [GV05] applies to binary fields. [GGG18] used fields which are exponentially large in n. The bounded distance decoding result in [CW04] only applies to the domain being the entire field, which requires a smaller field. Moreoever, it is not clear to what extent quantum attacks on bounded distance decoding have been considered. This remains an important open question in practice, as hash-based SNARKs are being proposed as post-quantum solutions on Ethereum and beyond.

Correlated agreement implies list-decodability: a high-level overview. We show that if the Reed-Solomon code of rate k/n satisfies correlated agreement with error probability $\epsilon < 1/k$, then the Reed-Solomon code of rate $(k+1)/n$ is list decodable. Given a codeword u of a Reed-Solomon code of rate $(k+1)/n$ and elements of the code $v^{(1)}, \ldots, v^{(L)}$ within Hamming distance f, we want to define a line between codewords $u^{(0)}$ and $u^{(1)}$ with a lot of points close to the Reed-Solomon code of rate k without correlated agreement. For some random $a \in \mathbb{F}_q$, we define $u^{(1)}$ to be the evaluation of $1/(x-a)$, which is far from any Reed-Solomon code (indeed, it is a deep hole, i.e., as far as possible), and $u^{(0)}$ to be the result of pointwise scaling u by the evaluation of $1/(x-a)$. Now, consider some $v^{(\ell)}$ which is the evaluation of a degree at most k polynomial $p^{(\ell)}$. By the polynomial remainder theorem, $p^{(\ell)}(x) = q^{(\ell)}(x)(x-a) + p(a)$ for some degree at most $k-1$ polynomial q. Now, if we multiply the codeword $v^{(\ell)}$ pointwise by the evaluation of $1/(x-a)$, we get a codeword close to $u^{(0)}$ which is the evaluation of $p^{(\ell)}(x)/(x-a) = q^{(\ell)}(x) + p^{(\ell)}(a)/(x-a)$. It follows that $u^{(0)} - p^{(\ell)}(a)u^{(1)}$ is close to the evaluation of $q^{(\ell)}(x)$, a Reed-Solomon codeword. Thus, we just need to show that $p^{(\ell)}(a)$ takes many different values for different $1 \leq \ell \leq L$ to show that many points on the line are close to the code. Distinct polynomials of degree k only agree on k evaluation points, so for a random $a \in \mathbb{F}_q$, each pair of elements of the list only agree with probability k/q. If $L \ll q/k$, then there are few collisions and the error probability for correlated agreement ϵ is about L/q. If $L \gg q/k$ and there are ϵq values of $p^{(\ell)}(a)$ in expectation, then the chance of a specific two of them being identical would be at least $1/(\epsilon q)$, and that needed to be at most k/q, so $\epsilon \geq 1/k$. So, if in fact there is a proximity gap for correlated agreement with $\epsilon \ll 1/k$, then we must be in the first case with $L \lessapprox \epsilon q$. We capture this formally in Theorem 2.

Directions for future research. Given the results of this work, we suggest the following research directions for the community:

1. Pursuing (quantum) attacks on bounded distance decoding.
2. Proving or disproving our conjectures.
3. Pursuing large error probabilities for (mutual) correlated agreement.

Regarding 2, a positive result here would be challenging. Our second main result (Theorem 2) proves that any such result would imply a similar result in list decoding, where decades of research have been conducted.

Thus, our second main result indicates that the community should be looking for large error probabilities for (mutual) correlated agreement, to avoid implying list-decodability.

2 Preliminaries

General Notation. We use $[n]$ to represent the set $\{1, \ldots, n\}$ and $[0..n]$ to represent the set $\{0, \ldots, n\}$. For a non-empty set S, let $x \leftarrow\!\!\$\ S$ denote sampling an element of S uniformly at random and assigning it to x.

Reed-Solomon Codes [RS60]. The Reed-Solomon code $RS(\mathbb{F}, D, k)$ is the set of evaluations on some domain D of polynomials in $\mathbb{F}[x]$ of degree at most $k-1$. The code has rate $\rho = k/|D|$. A codeword is a function $u : D \to \mathbb{F}$, however we will normally enumerate D as $x_1, \ldots, x_n$ with $n = |D|$ so we can consider a codeword as a vector $u \in \mathbb{F}^n$. $RS(\mathbb{F}, D, k)$ is then a k-dimensional subspace. The Hamming distance $\Delta(u, v)$ between codewords u, v is $|\{i \in [|D|] : u_i \neq v_i\}|$ and the distance $\Delta(u, V)$ of a codeword from a set V is $\min_{v \in V} \Delta(u, v)$.

Definition 1 (List Decoding [Eli57]). *Let $C \subseteq \mathbb{F}_q^n$ be a code. We say C is (δ, L)-list decodable if for every codeword $u \in \mathbb{F}_q^n$, we have:*

$$|\{v \in C \mid \Delta(u, v) \leq \delta \cdot n\}| \leq L.$$

Definition 2 (Correlated Agreement [BCI+23]). *Let $u_0, \ldots, u_\ell \in \mathbb{F}_q^n$ be a sequence of codewords. Let $C \subseteq \mathbb{F}_q^n$ be a set of codewords. Let $0 < \delta \leq 1$. If there exists a subdomain $D' \subseteq D$ and $v_0, \ldots, v_\ell \in C$ satisfying:*

- *Density: $|D'|/|D| \geq 1 - \delta$, and*
- *Agreement: for all $i \in [0..\ell]$, the codewords u_i and v_i agree on D',*

then we say $u_0, \ldots, u_\ell$ have correlated agreement with C of density $\geq 1 - \delta$.

Polynomial Interpolation. A polynomial $p(x) = m_0 + m_1 x + m_2 x^2 + \ldots + m_{k-1} x^{k-1}$ of degree $k-1$ over a field $\mathbb{F}$ can be interpolated by k points. Let $S \subseteq [n]$ be the list of k distinct indices corresponding to the x-coordinates $x_i \in \mathbb{F}, i \in S$, of these points. Then the Lagrange polynomial $L_i(x)$ has the form $L_i(x) = \prod_{j \in S; j \neq i} \frac{x - x_j}{x_i - x_j}$. Given a set of k points $(x_i, p(x_i))_{i \in S}$, any point $p(x_\ell)$ on the polynomial p can be determined by Lagrange interpolation as $p(x_\ell) = \sum_{k \in S} p(x_k) \cdot L_k(x_\ell)$.

Definition 3 (Polynomial Remainder (Little Bézout's) Theorem). *Let $p \in R[x]$ be a non-zero polynomial of degree $d \geq 0$ over an integral domain R. For all $r \in R$, we have:*

$$p(x) = (x - r)p^*(x) + p(r),$$

where $p^*(x)$ is a polynomial of degree $d - 1$.

Definition 4 (Schwartz–Zippel Lemma [DL78,Zip79,Sch80][1]). *Let $p \in R[x_1, \ldots, x_n]$ be a non-zero polynomial of total degree $d \geq 0$ over an integral domain R. Let S be a finite subset of R and let $r_1, \ldots, r_n \leftarrow\!\!\$\, S$. Then:*

$$\Pr[p(r_1, \ldots, r_n) = 0] \leq \frac{d}{|S|}.$$

Definition 5 (Cauchy–Bunyakovsky-Schwarz Inequality). *For any vectors $x, y \in \mathbb{R}^n$, we have:*

$$|\langle x, y \rangle| \leq \|x\|_2 \cdot \|y\|_2.$$

Definition 6 (Chebyshev-Cantelli Inequality). *Let X be a real-valued random variable. Let $a > 0$ and let $Var(X)$ be the variance of X. Then:*

$$\Pr[X - E(X) \geq a] \leq \frac{Var[X]}{Var[X] + a^2},$$

where $E(X)$ is the expected value of X. Applying the above to $-X$ gives:

$$\Pr[X - E(X) \leq -a] \leq \frac{Var[X]}{Var[X] + a^2}.$$

3 Proofs of Conjecture Failures

We prove that list-decodability and proximity gaps fail beyond the list-decoding capacity bound. We reproduce the relevant conjectures verbatim, show that they are false, and suggest minimally modified conjectures. The idea is to replace capacity in the conjecture statements with the list-decoding capacity. Thus, before looking at the consequences of our results for correlated agreement, we first consider list decoding.

[1] Versions of this lemma were proven in [Ore22,Mul54].

3.1 List-Decodability

To begin with, we observe that the conjecture from [BGKS20] on the list-decodability of Reed-Solomon codes up to capacity does not hold, due to the following theorem of Elias [Eli57].

Theorem 7.4.1 (List-Decoding Capacity). [GRS14,Eli57] *Let $q \geq 2, 0 \leq \delta < 1 - \frac{1}{q}$, and $\eta > 0$ be a small enough real. Then the following holds for codes of large enough block length n:*

(i) *If $\rho \leq 1 - H_q(\delta) - \eta$, then there exists a $(\delta, O(\frac{1}{\eta}))$-list decodable code.*
(ii) *If $\rho \geq 1 - H_q(\delta) + \eta$, every (δ, L)-list decodable code has $L \geq q^{\Omega(\eta n)}$.*

Thus, the list-decoding capacity is $1 - H_q(\delta)$, where δ is the fraction of errors and $H_q(x)$ is the q-ary entropy function: $H_q(x) = x \log_q(q-1) - x \log_q(x) - (1-x) \log_q(1-x)$.

We have the following claim.

Claim 1 *When $\delta \leq 1 - 1/(q-1)$, we have:*

$$0 \leq \frac{H_2(\delta)}{\log_2 q} - \frac{\delta}{(\ln 2)(q-1)(\log_2 q)} \leq H_q(\delta) - \delta \leq \frac{H_2(\delta)}{\log_2 q} \leq \frac{1}{\log_2 q}.$$

Moreover, $H_2(\delta) \geq 4\delta(1-\delta)$.

Proof. Since the derivative of $\ln x$ is $1/x$, by the mean value inequality, $\ln q - \ln(q-1) \leq 1/(q-1)$. So, we can bound the first term of $H_q(x)$ by:

$$\delta - \frac{\delta}{(q-1)\ln q} \leq \delta \log_q(q-1) \leq \delta.$$

Note that for $q = 2$, $\log_q(q-1) = \log_2 1 = 0$, so $H_2(\delta) = -\delta \log_2 \delta + -(1-\delta) \log_2(1-\delta)$. For the other two terms of $H_q(\delta)$, changing the basis of logarithm gives:

$$-x \log_q x - (1-x) \log_q(1-x) = \frac{x \log_2 x - (1-x) \log_2(1-x)}{\log_2 q} = \frac{H_2(\delta)}{\log_2 \delta}.$$

By the convexity of $-\log_2 x$, $H_2(\delta) = -\delta \log_2 \delta + -(1-\delta) \log_2(1-\delta) \leq -\log_2(1/2) = 1$. Putting these together gives:

$$\frac{H_2(\delta)}{\log_2 q} - \frac{\delta}{(\ln 2)(q-1)(\log_2 q)}) \leq H_q(\delta) - \delta \leq \frac{H_2(\delta)}{\log_2 q} \leq \frac{1}{\log_2 q}.$$

Next, we need to show that $H_2(\delta) \geq \delta/((\ln 2)(q-1))$. The term $-\delta \log_2 \delta$ is smaller than $\delta/((\ln 2)(q-1))$ when $\delta \leq \exp(-1/(q-1)) \leq 1 - 1/(q-1)$.

It remains to show that $H_2(\delta) \geq 4\delta(1-\delta)$. For this, we consider $H_2(\delta)/(\delta(1-\delta)) = -\log_2 \delta/(1-\delta) - \log_2(1-\delta)/\delta$. This has value 4 at $\delta = 1/2$. Its derivative is $\log_2 \delta/(1-\delta)^2 - \ln 2/(\delta(1-\delta)) - \log_2(1-\delta)/\delta^2 + \ln 2/(\delta(1-\delta)) = \log_2 \delta/(1-\delta)^2 - \log_2(1-\delta)/\delta^2$. This derivative is 0 at $\delta = 1/2$. Since the terms $\log_2 \delta, 1/(1-\delta)^2, -\log_2(1-\delta)$, and $-1/\delta^2$ are monotonically increasing, so is the derivative. Therefore, $H_2(\delta)/(\delta(1-\delta))$ achieves its minimum value of 4 at $\delta = 1/2$, so we have $H_2(\delta) \geq 4\delta(1-\delta)$. □

Theorem 7.4.1 directly contradicts the following conjecture, since $H_q(\delta) > \delta$ by Claim 1.

Conjecture 2.3 (List-Decodability of Reed-Solomon Codes up to Capacity). [BGKS20] *For every $\rho > 0$, there is a constant C_ρ such that every Reed-Solomon code of length n and rate ρ is list-decodable from $\delta \leq 1 - \rho - \eta$ fraction errors with list size $(\frac{n}{\eta})^{C_\rho}$. That is:*

$$\mathcal{L}(\mathbb{F}_q, D, d = \rho|D|, 1 - \rho - \eta) \leq \frac{|D|}{\eta}^{C_\rho}.$$

We propose the following minimally modified conjecture for prime fields.

Our Conjecture 1. (List-Decodability of Reed-Solomon Codes up to List-Decoding Capacity for Prime Fields). *Conjecture 2.3 from [BGKS20] holds with $\delta \leq 1 - \rho - \eta$ replaced by $H_q(\delta) \leq 1 - \rho - \eta$ when q is prime.*

Claim 1 implies that the reduction in distance δ required to get the same L is at most only $1/\log_2 q$. Fields used in SNARKs have $\log_2 q \gtrapprox 31$, so in practice this is a small difference. However, SNARKs would also usually use $n \gg \log_2 q$, so the list-decoding capacity bound should rule out the unmodified conjecture for some integral values of k, f.

3.2 Proximity Gaps

For our proximity gaps results, we prove the following.

Theorem 1. *Suppose $f < n-k$ and $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) > f$. Then there exists a $u^{(0)}$ such that there are at least $\frac{\binom{n}{f}}{\binom{n}{f}+q^{n-f-k}g(f(n-f)/q)}$ values of $\lambda \in \mathbb{F}_q$ such that $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f$, where $g(x) = \begin{cases} \exp(x) \text{ when } x \leq 3/2 \\ \frac{\exp(2\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}} \text{ when } x > 3/2. \end{cases}$*

We begin by proving that a random codeword is close to the code beyond the list-decoding capacity bound.

Lemma 1. *For a uniformly random $u \in \mathbb{F}_q^n$, it holds that:*

$$\frac{\binom{n}{f}}{\binom{n}{f} + q^{n-f-k}g(f(n-f)/q)} \leq \Pr(\Delta(u, RS(\mathbb{Z}_q, D, k)) \leq f) \leq q^{H_q(f/n)n-n+k}.$$

Proof. Given a random $u \in \mathbb{F}_q^n$, what is the probability that $\Delta(u, RS(\mathbb{F}_q, D, k)) \leq f$? If it is, then there is some *disagreement set* $F \subset n$ with $|F| = f$ such that there exists a polynomial p with $\deg p \leq k-1$ and $p(x_i) = u_i$ for $i \in [n] \setminus F$. Let X_F be the indicator variable that F is such a set, i.e., $X_F = 1$ if and only if there exists a polynomial p with $\deg p \leq k-1$ and $p(x_i) = u_i$ for $i \in [n] \setminus F$. Let $X = \sum_F X_F$

with the sum taken over all $\binom{n}{f}$ subsets of $[n]$ of size f. $\Delta(u, RS(\mathbb{F}_q, D, k)) \leq f$ if and only if $X > 0$, so we are interested in $\Pr[X > 0]$.

To compute the expectation and variance of X, we need to have expressions for $\Pr[X_F = 1]$ and $\Pr[X_F = 1 \wedge X'_F] = 1$ for each pair of sets $F, F' \subset n$, $|F|, |F'| = f$, given by the following lemma.

Lemma 2. *Given two subsets $F, F' \subset n$ of size f and u sampled uniformly from $\mathbb{F}_q^n$, let p_F, p'_F be the polynomials of degree at most $n - f$ obtained by Lagrange interpolating $p_F(x_i) = u_i$ for $i \in [n]/F$ and $p_{F'}(x_i) = u_i$ for $i \in [n]/F'$, respectively. X_F and $X_{F'}$ are indicator variables for $p_F(x)$ or $p_{F'}(x)$ respectively having degree $k - 1$ or lower. Then, we have:*

(i) $\Pr[X_F = 1] = q^{-(n-f-k)}$.
(ii) $\Pr[X_F = 1 \wedge X_{F'} = 1 \wedge p_F(x) \equiv p_{F'}(x)] = q^{-(n-k-|F\cap F'|)}$.
(iii) If $|F \cup F'| \leq n - k$, $X_F = 1$, *and* $X_{F'} = 1$, *then* $p_F(x) \equiv p_{F'}(x)$.
(iv) When $|F \cup F'| \leq n - k$, $\Pr[X_F = 1 \wedge X_{F'} = 1] = q^{-(n-k-|F\cap F'|)}$.
(v) When $|F \cup F'| \geq n - k$, $p_F(x) = p_{F\cup F'}(x) + z_{F\cup F'}(x)a_{F/F'}(x)$ *and* $p_F(x) = p_{F\cup F'}(x)+z_{F\cup F'}(x)a_{F'\setminus F}(x)$, *where* $z_{F\cup F'}(x) = \prod_{i\in F\cup F'}(x-x_i)$ *and* $p_{F\cup F'}(x)$, $a_{F\setminus F'}(x), a_{F'\setminus F}(x)$ *are distributed indepenently and uniformly at random from* $\mathbb{F}_q[x]^{\leq n-|F\cup F'|-1}$, $\mathbb{F}_q[x]^{\leq |F\setminus F'|-1}$, $\mathbb{F}_q[x]^{\leq |F'\setminus F|-1}$ *respectively, where* $\mathbb{F}_q[x]^{\leq \ell}$ *is the set of polynomials with coefficients in $\mathbb{F}_q$ and degree up to ℓ.*
(vi) When $|F \cup F'| \geq n - k$, $\Pr[X_F = 1 \wedge X_{F'} = 1] = q^{-2(n-f-k)}$, *i.e., the events $X_F = 1$ and $X_{F'} = 1$ are independent.*

Proof. For (i), note that p_F is distributed uniformly at random from $\mathbb{F}_q[x]^{\leq n-f-1}$. The $n - f - k$ coefficients of $x^k, x^{k+1}, \ldots, x^{n-f-1}$ are independent and uniformly random, and the probability that they are all zero is $q^{-(n-f-k)}$. This is the probability that $X_F = 1$.

For (ii), note that if $X_F = 1, X_{F'} = 1$, and $p_F(x) \equiv p_{F'}(x)$, then there is a polynomial of degree at most $k - 1$ that agrees with u on $[n] \setminus (F \cap F')$. The same reasoning as (i) gives that this happens with probability $q^{-(n-k-|F\cap F'|)}$.

For (iii), if $v_F, v_{F'}$ are the evaluations of $p_F(x), p_{F'}(x)$ and $X_F = 1, X_{F'} = 1$, then $v_F, v_{F'}$ are codewords of $RS(\mathbb{F}_q, D, k)$, which has minimum distance $n-k+1$. Thus, either $\Delta(v_F, v_{F'}) \geq n - k + 1$ or $\Delta(v_F, v_{F'}) = 0$. Both $p_F(x)$ and $p_{F'}(x)$ agree with u outside of $F \cup F'$, so $\Delta(v_F, v_{F'}) \leq |F \cap F'|$. Since $|F \cap F'| \leq n - k$, $v_F = v_{F'}$ and $p_F(x) \equiv p_{F'}(x)$.

(iv) follows from combining (ii) and (iii).

For (v), we define $p_{F\cup F'}(x)$ as the polynomial obtained by Lagrange interpolating $p_{F\cup F'}(x_i) = u_i$ on $i \in [n] \setminus (F \cup F)$. Since this subvector of u is uniformly distributed and Lagrange interpolation is a bijective linear transformation, $p_{F\cup F'}(x)$ is uniformly distributed on $\mathbb{F}_q[x]^{\leq n-|F\cup F'|-1}$.

We define $a_{F\setminus F'}(x)$ to be the polynomial obtained by Lagrange interpolating $a_{F\setminus F'}(x_i) = \frac{u_i - p_{F\cup F'}(x_i)}{z_{F\cup F'}(x_i)}$ for $i \in F \setminus F'$, noting that the denominator is never 0. Conditioning on the values of u_i on $[n] \setminus (F \cup F')$, since the values of u_i for $i \in F \setminus F'$ are independent and uniformly distributed, so are the values of $\frac{u_i - p_{F\cup F'}(x_i)}{z_{F\cup F'}(x_i)}$ for $i \in F \setminus F'$. Thus, using bijectivity of Lagrange interpolation

again, $a_{F\setminus F'}(x)$ is uniformly distributed on $\mathbb{F}_q[x]^{\leq |F\setminus F'|-1}$ conditioned on the values of u_i on $i \in [n] \setminus (F \cup F')$. Since the distribution of $a_{F\setminus F'}(x)$ is the same for any values of u_i on $i \in [n] \setminus (F \cup F')$, it is in fact independent of these values of u_i and so also of $p_{F\cup F'}(x)$, which is determined by them.

We define $a_{F'\setminus F}(x)$ to be the polynomial obtained by Lagrange interpolating $a_{F'\setminus F}(x_i) = \frac{u_i - p_{F\cup F'}(x_i)}{z_{F\cup F'}(x_i)}$ for $i \in F' \setminus F$, noting that the denominator is never 0. Now, we can use a similar argument as for $a_{F\setminus F'}(x)$, but this time conditioned on the values of u_i on $[n] \setminus F'$, to show that $a_{F\setminus F'}(x)$ is uniformly distributed on $\mathbb{F}_q[x]^{\leq |F'\setminus F|-1}$ and that it is independent of the values of u_i on $[n] \setminus F'$. Since the values of u_i on $[n] \setminus F'$ determine $p_{F\cup F'}(x)$ and $a_{F\setminus F'}(x)$, it holds that $a_{F'\setminus F}(x)$ is independent of them.

The polynomials $p_F(x)$ and $p_{F\cup F'}(x) + z_{F\cup F'}(x)a_{F\setminus F'}(x)$ agree on $n-f$ points x_i for $i \in [n] \setminus F$ and have degree at most $n-f$, so they are equal. Similarly, we obtain $p_{F'}(x) = p_{F\cup F'}(x) + z_{F\cup F'}(x)a_{F'\setminus F}(x)$. This completes the proof of (v).

For (vi), $X_F = 1$ occurs when $p_F(x)$ has degree at most $k-1$. Using the formula $p_F(x) = p_{F\cup F'}(x) + z_{F\cup F'}(x)a_{F\setminus F'}(x)$, since $p_{F\cup F'}(x)$ has degree at most $n - |F \cup F'| \leq k-1$, $X_F = 1$ occurs exactly when $\deg a_{F\setminus F'}(x) \leq k - |F\cup F'| - 1$. Similarly, $X_{F'}$ occurs exactly when $\deg a_{F'\setminus F}(x) \leq k - |F \cup F'| - 1$. Since these polynomials $a_{F\setminus F'}(x)$ and $a_{F'\setminus F}(x)$ are independent, so are the events $X_F = 1$ and $X_{F'} = 1$. By (i), $\Pr[X_F = 1] = \Pr[X_{F'} = 1] = q^{-(n-f-k)}$, so since these are independent, $\Pr[X_F = 1 \wedge X_{F'} = 1] = q^{-2(n-f-k)}$ as required. □

We will use Cantelli's inequality (Definition 6) to bound $\Pr[X > 0]$, for which we will need the variance of X. By Lemma 2 (vi), when $|F \cap F'| \leq 2f - (n-k)$, X_F,$X_{F'}$ are independent and $Cov[X_F, X_{F'}] = 0$. Using (i) and (iv), if $|F \cap F'| > 2f - (n-k)$, then:

$$Cov[X_F, X'_F] = E[X_F X'_F] - E[X_F]E[X_{F'}] = q^{-(n-k-|F\cap F'|)} - q^{-2(n-f-k)}.$$

For any F, the number of F' with $|F \cap F'| = f - \ell$ is $\binom{f}{\ell}\binom{n-f}{\ell}$ since there are $\binom{f}{\ell}$ choices of ℓ points to remove from F to get $F \cap F'$ and then $\binom{n-f}{\ell}$ choices of ℓ points outside of F to add to get F'. We have $Cov[X_F, X_{F'}] = 0$ when $|F \cap F'| \leq 2f - (n-k)$ and so $\ell \geq n - k - f$. Thus, we have that:

$$\begin{aligned}
Var[X] &= \sum_F Var[X_F] + \sum_{F\neq F'} Cov[X_F, X_{F'}] \\
&= \binom{n}{f} q^{-(n-f-k)}(1 - q^{-(n-f-k)}) + \sum_{\ell=1}^{n-f-k-1} \binom{n}{f}\binom{f}{\ell}\binom{n-f}{\ell}(q^{-(n-k-f+\ell)} - q^{-2(n-f-k)}) \\
&\leq \binom{n}{f} q^{-(n-f-k)} + \sum_{\ell=1}^{n-k-f-1} \binom{n}{f}\binom{f}{\ell}\binom{n-f}{\ell} q^{-(n-k-f+\ell)} \\
&= \binom{n}{f} q^{-(n-f-k)} \sum_{\ell=0}^{n-k-f-1} \binom{f}{\ell}\binom{n-f}{\ell} q^{-\ell}
\end{aligned}$$

$$\leq \binom{n}{f} q^{-(n-f-k)} \sum_{\ell=0}^{\infty} f^{\ell}(n-f)^{\ell} q^{-\ell}/\ell!^2$$
$$= E[X] \sum_{\ell=0}^{\infty} f^{\ell}(n-f)^{\ell} q^{-\ell}/\ell!^2,$$

where we used that $E[X] = \sum_F E[X_F] = \binom{n}{f} q^{-(n-f-k)}$. We need a bound on this series.

Claim 2 $\sum_{\ell=0}^{\infty} x^{\ell}/\ell!^2 \leq g(x) = \begin{cases} \exp(x) \text{ when } x \leq 3/2 \\ \frac{\exp(2\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}} \text{ when } x > 3/2. \end{cases}$

Proof. For the $x \leq 3/2$ expression, we have $\sum_{\ell=0}^{\infty} x^{\ell}/\ell!^2 \leq \sum_{\ell=0}^{\infty} x^{\ell}/\ell! = \exp(x)$. Now we assume $x > 3/2$ and consider the other expression. We have $\sum_{\ell=0}^{\infty} x^{\ell}/\ell!^2 \leq \left(\sum_{\ell=0}^{\infty} x^{\ell/2}/\ell!\right)\left(\max_{\ell=0}^{\infty} x^{\ell/2}/\ell!\right)$. Since $\sum_{\ell=0}^{\infty} x^{\ell/2}/\ell! = \exp(\sqrt{x})$, it remains to bound the maximum.

The ratio of the ℓ term to the $\ell-1$ term is $(x^{\ell/2}/\ell!)/(x^{(\ell-1)/2}/(\ell-1)!) = \sqrt{x}/\ell$. It follows that the maximum is achieved at $\ell = \lfloor\sqrt{x}\rfloor$, with value $x^{\lfloor\sqrt{x}\rfloor}/\lfloor\sqrt{x}\rfloor!$. Using a standard lower bound for Stirling's approximation, $i! \geq \sqrt{2\pi i}(i/e)^i$ and $x^{\lfloor\sqrt{x}\rfloor/2}/\lfloor\sqrt{x}\rfloor! \leq (\sqrt{x}e/\lfloor\sqrt{x}\rfloor)^{\lfloor\sqrt{x}\rfloor}/\sqrt{2\pi\lfloor\sqrt{x}\rfloor}$. Now, $(ae/x)^x$ has derivative of its logarithm $\ln\ln(ae/x) - 1 = \ln(a/x)$, which is positive for $x < a$ and negative for $x > a$, so $(ae/x)^x$ achieves its maximum value at $x = a$, when it is e^x. Thus, we have $(\sqrt{x}e/\lfloor\sqrt{x}\rfloor)^{\lfloor\sqrt{x}\rfloor} \leq \exp(\sqrt{x})$. Putting these together, we have $\max_{\ell=0}^{\infty} x^{\ell/2}/\ell! \leq \frac{\exp(\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}}$.

Thus, we have:

$$\sum_{\ell=0}^{\infty} x^{\ell}/\ell!^2 \leq \left(\sum_{\ell=0}^{\infty} x^{\ell/2}/\ell!\right)\left(\max_{\ell=0}^{\infty} x^{\ell/2}/\ell!\right) \leq \exp(\sqrt{x}) \cdot \frac{\exp(\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}} = \frac{\exp(2\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}}.$$

□

Now, we have $Var[X] \leq E[X]g(f(n-f)/q)$. Applying Cantelli's inequality, we obtain:

$$\begin{aligned} \Pr[X > 0] &= 1 - \Pr[X - E[X] \leq -E[X]] \\ &\geq 1 - \frac{Var[X]}{Var[X] + E[X]^2} \\ &= \frac{E[X]^2}{Var[X] + E[X]^2} \\ &\geq \frac{E[X]}{g(f(n-f)/q) + E[X]} \\ &= \frac{\binom{n}{f} q^{-(n-f-k)}}{g(f(n-f)/q) + \binom{n}{f} q^{-(n-f-k)}} \end{aligned}$$

$$= \frac{\binom{n}{f}}{\binom{n}{f} + q^{n-f-k} g(f(n-f)/q)}.$$

This proves the lower bound for $\Pr(\Delta(u, C) \leq f)$, where $C = RS(\mathbb{F}_q, D, k)$. We need to show the upper bound. For this, we can follow the standard proof of list-decodability, such as that in [GRS14]. We define the Hamming ball $B(x, f) = \{y \in \mathbb{F}_q^n : \Delta(x, y) \leq f\}$. We will cite Proposition 3.3.3 of [GRS14] for the classical bound that for any codeword x and $f \leq (1 - 1/q)n$, it holds that:

$$|B(x, f)| = \sum_{i=0}^{f} \binom{n}{i} (q-1)^i \leq q^{H_q(f/n)n}.$$

We note that for Reed-Solomon codes, $n \leq q$, so any $f < n$ has $(1 - 1/q)n \geq n - 1 \geq f$.

We note that $y \in B(x, f)$ implies that $y - x \in B(0, f)$ and that, for any fixed v, $u + v$ is uniformly distributed, so $\Pr[u + v] \in C = \Pr[u \in C] = q^{k-n}$. Then, by a union bound:

$$\begin{aligned}
\Pr(\Delta(u, RS(\mathbb{F}_q, D, k)) \leq f) &\leq \sum_{v \in B(0,f)} \Pr[u + v \in C] \\
&= |B(0, f)| q^{k-n} \\
&\leq q^{H_q(f/n)n} q^{k-n} = q^{H_q(f/n)n - n + k}
\end{aligned}$$

as required. □

Proof (of Theorem 1). What does the above lemma imply for proximity gaps? The idea is to look at a line through a random codeword and a codeword that is farther from the code than a random codeword. The application to threshold Schnorr signatures [CS25] used a deep hole that is given by $u_i^{(1)} = 1/x_i$ with $0 \notin D$. (A deep hole is a codeword that is the largest distance possible to the code. For $RS(\mathbb{F}_q, D, k)$ with $|D| = n$, this is distance $n - k$.) So, let us just assume that $0 \notin D$ and use that. This is without loss of generality, as we could shift the domain. We then let $u^{(0)}$ be a random codeword. Now, for any $\lambda \in \mathbb{F}_q$, $u^{(0)} + \lambda u^{(1)}$ conditioned on λ but not $u^{(1)}$ is also a uniformly random codeword, so we have:

$$\Pr[\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f \mid \lambda \neq 0] \geq \frac{\binom{n}{f}}{\binom{n}{f} + q^{n-f-k} g(f(n-f)/q)}.$$

It follows that the expected number of λ such that $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f$ is at least $\frac{\binom{n}{f}}{\binom{n}{f} + q^{n-f-k} g(f(n-f)/q)}$. Some $u^{(1)}$ has at least this expectation. This completes the proof of Theorem 1. □

How far beyond the list-decoding capacity does correlated agreement fail? Because one quantity in the expression is exponential in k with base q, quite close to the list-decoding capacity, all points on the affine line are within distance f of the code.

Corollary 1. *Given n, q with $q \geq 10, q \geq n$, if we have:*

$$n(1 - H_q(f/n)) + 2 + \sqrt{nH_q(f/n) - f} \leq k \leq n - f - 2,$$

then there exists $u^{(0)}, u^{(1)} \in \mathbb{F}_q^n$ such that $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) > f$, and for every $\lambda \in \mathbb{F}_q$, we have $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f$. Indeed, there is a $u^{(0)}$ for which this holds for any $u^{(1)} \in \mathbb{F}_q^n$ with $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) > f$.

Note that $n(1 - H_q(f/n)) + nH_q(f/n) - f = n - f$, so when the list-decoding capacity $n(1-H_q(f/n))$ is significantly below the capacity $n-f$, i.e., $nH_q(f/n)-f \gg 1$, then $nH_q(f/n) - f \gg \sqrt{nH_q(f/n) - f}$, so the minimum k to which this bound applies is close to $n(1 - H_q(f/n))$.

In appplications, it is normally k that is fixed and f that can vary. Since the attack applies when $n - f - k \leq n/\log_2 q$, when $k/n \gg \log_2 q$, a similar result holds with $q(1 - H_q(k/n)) \lessapprox f \leq n - k - 1$.

Proof. We need to show that Theorem 1 applies. We can explicitly take $u^{(1)}$ to be the evaluation of x^k, which is not in $RS(\mathbb{F}_q, D, k)$ but is in $RS(\mathbb{F}_q, D, k+1)$. $RS(\mathbb{F}_q, D, k+1)$ has minimum separation $n - k - 2 > f$, and every codeword of $RS(\mathbb{F}_q, D, k)$ is in $RS(\mathbb{F}_q, D, k+1)$ and not equal to $u^{(1)}$, so $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) \geq f$.

Now, it remains to show that these parameters imply that $\frac{\binom{n}{f}}{\binom{n}{f}+q^{n-f-k}g(f(n-f)/q)} > 1 - 1/q$. Then, we could apply Theorem 1 to obtain the corollary.

This condition is implied by showing $\binom{n}{f} > q^{n-f-k+1}g(x)$. We will use the bound:

$$\binom{n}{f} \geq \sqrt{\frac{n}{8f(n-f)}} 2^{nH_2(f/n)}.$$

We need to show that $2^{nH_2(f/n)} > q^{n-f-k+1}g(f(n-f)/q)\sqrt{8f(n-f)/n)}$ or, taking logarithms, that $nH_2(f/n) > (n - f - k + 1)\log_2 q + \log_2 g(f(n-f)/q) + \log_2(8f(n-f)/n)/2$. Rearranging to get the k term on one side, we need $(n - f - k + 1)\log_2 q < nH_2(f/n) - \log_2 g(f(n-f)/q) - \log_2(8f(n-f)/n)/2$.

Splitting $g(f(n-f)/q)$ into its cases, when $f(n-f)/q \leq 3/2$, $\log_2 g(f(n-f)/q) = \frac{f(n-f)}{(\ln 2)q} < (2/\ln 2)\sqrt{f(n-f)/q}$, since for $y < 4$, $y < 2\sqrt{y}$. When $f(n-f)/q > 3/2$, $\log_2 g(f(n-f)/q) < (2/\ln 2)\sqrt{f(n-f)/q}$.

We have:

$$\begin{aligned}
&nH_2(f/n) - \log_2 g(f(n-f)/q) - \log_2(8f(n-f)/n)/2 \\
&> nH_2(f/n) - (2/\ln 2)\sqrt{f(n-f)/q} - \log_2(8q)/2 \\
&\geq nH_2(f/n) - (2/\ln 2)\sqrt{nH_2(f/n)/4} - 3/2 - \log_2 q/2 \\
&= nH_2(f/n) - \sqrt{nH_2(f/n)}/\ln 2 - 3/2 - \log_2 q/2 \\
&\geq n(H_q(f/n) - f/n)\log_2 q - \sqrt{nH_2(f/n)}/\ln 2 - 3/2 - \log_2 q/2 \\
&\geq (nH_q(f/n) - f)\log_2 q - \sqrt{(nH_q(f/n) - f)\log_2 q}/\ln 2 - 3/2 - \log_2 q/2
\end{aligned}$$

$$\geq (nH_q(f/n) - f)\log_2 q - \sqrt{nH_q(f/n) - f}n\log_2 q - \log_2 q.$$

This is strictly smaller than $(n - f - k + 1)\log_2 q$ when:

$$(k - n(1 - H_q(f/n)))\log_2 q \geq \log_2 q + \sqrt{nH_q(f/n) - f}\log_2 q + \log_2 q,$$

so when

$$k \geq n(1 - H_q(f/n)) + 2 + \sqrt{nH_q(f/n) - f}$$

as required. □

This contradicts the following conjectures, reproduced verbatim.

Conjecture 8.4. [BCI+23] *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*
$$\epsilon \leq \frac{1}{(\eta\rho)^{c_1}} \cdot \frac{(\ell \cdot n)^{c_2}}{q}.$$

For this conjecture, we are only interested in its application to the following.

Theorem 1.2. (Proximity gap for RS codes). [BCI+23] *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 $V := \mathsf{RS}[\mathbb{F}_q, \mathcal{D}, k+1]$ 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 $\delta \in (0, \frac{1-\rho}{2}]$, the error parameter ϵ is*
$$\epsilon = \epsilon_{\mathsf{U}} = \epsilon_{\mathsf{U}}(q, n) := \frac{n}{q}.$$
- *Johnson bound: For $\delta \in (\frac{1-\rho}{2}, 1 - \sqrt{\rho})$, setting $\eta := 1 - \sqrt{\rho} - \delta$, the error parameter ϵ is*
$$\epsilon = \epsilon_{\mathsf{J}} = \epsilon_{\mathsf{J}}(q, n, \rho, \delta) := \frac{(k+1)^2}{\left(2\min\left(\eta, \frac{\sqrt{\rho}}{20}\right)\right)^7 q} = O\left(\frac{1}{(\eta\rho)^{O(1)}} \cdot \frac{n^2}{q}\right).$$

Theorem 1.4. (Correlated agreement over lines). [BCI+23] *Let V, q, n, k 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 defined as 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* $\mathcal{D}'$.

Theorems 1.5 and 1.6 are generalizations of correlated agreement over lines (Theorem 1.4) to low-degree parameterized curves and affine spaces, respectively. Theorem 1.2 follows from Theorem 1.4, so we only need to show that Conjecture 8.4 applied to Theorem 1.4 is incorrect.

In the regime where Corollary 1 applies, $n(1-H_q(f/n))+2+\sqrt{nH_q(f/n)-f} \leq k \leq n-f-2$, the proximity gap can not have $\epsilon < 1$. The conjecture gives $\epsilon \leq \frac{1}{(\eta\rho)^{c_1}} \cdot \frac{(2n)^{c_2}}{q} \leq \frac{(2n)^{c_1+c_2}}{q}$. It is then enough to additionally take $q > (2n)^{c_1+c_2}$ for the conjecture to fail.

We propose the following minimally modified conjecture.

Our Conjecture 2. (Proximity gap and correlated agreement for RS codes for prime fields). *Conjecture 8.4 from [BCI+23] holds with* $\delta \leq 1-\rho-\eta$ *replaced by* $\delta \leq 1-H_q(\delta)-1/n-\eta$ *when* q *is prime.*

The construction in Theorem 1 gave the error probability for correlated agreement as the probability that a random codeword is within Hamming distance f of the code, which Lemma 1 bounds by $q^{H_q(\delta)n-n+k} = q^{(H_q(\delta)-1+\rho)n}$. Under the assumptions of the modified conjecture, $\rho \leq 1-H_q(\delta)-1/n$, and this probability is bounded by $q^{(H_q(\delta)-1+\rho)n} \leq q^{(1/n)n} = 1/q$. This only guarantees one point on the line is close to the code, which is fewer than the bound in the conjecture. Similarly to the list-decoding case, by Claim 1, the distance δ required to get the same parameters is only lowered by $1/\log_2 q + 1/n$.

Mutual correlated agreement implies correlated agreement, so the equivalent conjecture below fails.

Definition 7. (Mutual correlated agreement). [ACFY24b] *Let* $\mathcal{C} \subseteq \mathbb{F}^n$ *be a linear code. We say that* Gen *is a proximity generator for* $\mathcal{C}$ *with mutual correlated agreement with proximity bound* B^* *and error* err^* *if for every* $f_1, \ldots, f_\ell : [n] \to \mathbb{F}$ *and* $\delta \in (0, 1-\mathsf{B}(\mathcal{C},\ell))$ *the following holds:*

$$\Pr_{(r_1,\ldots,r_\ell)\leftarrow\mathsf{Gen}(\ell)}[\exists S \subseteq [n] \text{ s.t. } |S| \geq (1-\delta)\cdot n \wedge \exists u \in \mathcal{C}, u(S) = \sum_{j\in[\ell]} r_j \cdot f_j(S)$$
$$\wedge\, \exists i \in [\ell], \forall u' \in \mathcal{C}, u'(S) \neq f_i(S)] \leq \mathsf{err}^*(\mathcal{C},\ell,\delta).$$

Conjecture 4.12. (Mutual correlated agreement). [ACFY24b] *The function* $\mathsf{Gen}(\ell;\alpha) := (1,\alpha,\ldots,\alpha^{\ell-1}))$ *is a proximity generator with mutual correlated agreement for every smooth Reed-Solomon code* $\mathcal{C} := \mathsf{RS}[\mathbb{F},\mathcal{L},m]$ *(with rate* $\rho := 2^m/|\mathcal{L}|$*). We give two conjectures, for the parameters of the proximity bound* B^* *and the error* err^**:*

1. *Up to the Johnson bound:* $\mathsf{B}^*(\mathcal{C},\ell) := \sqrt{\rho}$, *and*

$$\mathsf{err}(\mathcal{C},\ell,\delta) := \frac{(\ell-1)\cdot 2^{2m}}{|\mathbb{F}|\cdot(2\cdot\min\{1-\sqrt{\rho}-\delta, \frac{\sqrt{\rho}}{20}\})^7}.$$

2. *Up to capacity:* $\mathsf{B}^*(\mathcal{C}, \ell) := \rho$, *and there exist constants* $c_1, c_2, c_3 \in \mathbb{N}$ *such that for every* $\eta > 0$ *and* $0 < \delta < 1 - \rho - \eta$*:*

$$\mathsf{err}^*(\mathcal{C}, \ell, \delta) := \frac{(\ell - 1)^{c_2} \cdot d^{c_2}}{\eta^{c_1} \cdot \rho^{c_1 + c_2} \cdot |\mathbb{F}|}.$$

We propose the following modified conjecture.

Our Conjecture 3. (Mutual correlated agreement for prime fields). *Conjecture 4.12 from [ACFY24b] holds with* $0 < \delta < 1 - \rho - \eta$ *replaced by* $0 < H_q(\delta) < 1 - 1/n - \rho - \eta$ *when* $\mathbb{F} = \mathbb{F}_q$ *for a prime* q*.*

4 Correlated Agreement of Reed-Solomon Codes Implies List Decoding

We now present our second main result.

Theorem 2. *If* $RS(\mathbb{F}_q, D, k)$ *satisfies correlated agreement over lines with* $f < n - k - 1$ *errors with error parameter* $\epsilon < (q - n)/kq$*, then* $RS(\mathbb{F}_q, D, k + 1)$ *is* $(f/n, L)$*-list decodable, where:*

$$L = \lceil \frac{\epsilon q(q - n)}{q - n - k\epsilon q} \rceil.$$

Note that for $\epsilon < (q - n)/2kq$, this gives the simple bound $L \leq 2\epsilon q$.

Correlated agreement results with $\epsilon > 1/k$ are useful; indeed, the only known results for small fields and large k, e.g., $q \approx 2^{32}$, $k \approx 2^{20}$, that are used in SNARKs in practice fall into this regime. For these parameters, Theorem 2 does not demonstrate anything.

On the other hand, conjectures about correlated agreement with $\epsilon = F(n, k, f)/q$ for some function F will imply similar conjectures about list decoding for fields larger than $2kF(n, k, f)$.

Proof. We begin by showing that the following implies Theorem 2.

Claim 3 *For any* L*, if there is a codeword* $u \in \mathbb{F}_q^n$ *such that there are* L *distinct elements* $v^{(1)}, \ldots, v^{(L)} \in RS(\mathbb{F}_q, D, k + 1)$ *with* $\Delta(u, v^{(i)}) \leq f$*, then there exist codewords* $u^{(0)}, u^{(1)} \in \mathbb{F}_q^n$ *such that there are* E *values of* λ *such that* $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f$*, where:*

$$E \geq \frac{(L - 1)(q - n)}{q - n + (L - 1)k}.$$

Suppose for a contradiction that $RS(\mathbb{F}_q, D, k + 1)$ is not $(f/q, L)$-list decodable for

$$L = 1 + \lceil \frac{k\epsilon q(q - n)}{q - n - k\epsilon q} \rceil.$$

Then there exists a $u \in \mathbb{F}_q^n$ such that there are L distinct elements $v^{(1)}, \ldots, v^{(L)} \in RS(\mathbb{F}_q, D, k+1)$ with $\Delta(u, v^{(i)}) \leq f$. By Claim 3, there exist codewords $u^{(0)}, u^{(1)} \in \mathbb{F}_q^n$ such that there are E values of λ such that $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_q, D, k)) \leq f$, where $E \geq \frac{(L-1)(q-n)}{q-n+(L-1)k}$. Rearranging the expression for E, we have:

$$1 - kE/(q-n) \leq \frac{q-n}{q-n+(L-1)k}.$$

Therefore:

$$q - n - (L-1)k \leq \frac{(q-n)^2}{q-n-kE}$$

$$L - 1 \geq \frac{E(q-n)}{q-n-kE}.$$

Substituting $E = \epsilon q$ gives the contradiction of the proximity gap. □

Proof (of Claim 3). The idea is to let $u^{(0)}$ be the codeword with coordinates $u_i/(x_i - a)$ for some $a \in \mathbb{F}_q$. This is within Hamming distance f of the codeword with coordinates $v_i^{(j)}/(x_i - a)$, which we will call $v'^{(j)}$. Now since $v^{(j)} \in RS(\mathbb{F}_q, D, k+1)$, there is a polynomial $p^{(j)}(x)$ of degree at most k such that $v_i^{(j)} = p^{(j)}(x_i)$. Now, by the polynomial remainder theorem (Definition 3), we have:

$$p^{(j)}(x)/(x-a) = \frac{p^{(j)}(x) - p^{(j)}(a)}{x-a} + p^{(j)}(a)/(x-a)$$

where $\frac{p^{(j)}(x)-p^{(j)}(a)}{x-a}$ is a degree-k polynomial.

Let $u^{(1)}$ be the codeword with $u_i^{(1)} = -1/(x_i - a)$. Now, $v'^{(j)} + p^{(j)}(a)u^{(1)}$ has coordinates $\frac{p^{(j)}(x)-p^{(j)}(a)}{x-a}$ evaluated at x_i, so $v'^{(j)} + p^{(j)}(a)u^{(1)} \in RS(\mathbb{F}_q, D, k)$. Thus,

$$\begin{aligned}
\Delta(u^{(0)} + p^{(j)}(a)u^{(1)}, RS(\mathbb{F}_q, D, k)) &\leq \Delta(u^{(0)} + p^{(j)}(a)u^{(1)}, v'^{(j)} + p^{(j)}(a)u^{(1)}) \\
&= \Delta(u^{(0)}, v'^{(j)}) \\
&= \Delta(u, v^{(j)}) \\
&\leq f.
\end{aligned}$$

By the Schwartz–Zippel lemma (Definition 4), for a random $a \in \mathbb{F}_q$, the probability for distinct j, j' that $p^{(j)}(a) = p^{(j')}(a)$ is at most k/q since these are distinct polynomials. So, the number of distinct values $p^{(j)}(a)$ is probably not much smaller than L.

We need to show that $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) > f$. For this, we multiply pointwise with $x_i - a$. Let $v \in RS(\mathbb{F}_q, D, k)$ be a codeword with $\Delta(u^{(1)}, v) = \Delta(u^{(1)}, RS(\mathbb{F}_q, D, k))$. The codeword v' with coordinates $v'_i = v_i(x_i - a)$ is in $RS(\mathbb{F}_q, D, k+1)$. The all-1's codeword $\mathbf{1}$ with coordinates $1 = u^{(1)}(x_i - a)$ is also in $RS(\mathbb{F}_q, D, k+1)$ and v' is the evaluation of a non-constant polynomial,

so $\Delta(\mathbf{1}, v')$ is at least the minimum distance of $RS(\mathbb{F}_q, D, k+1)$, which is $n-k$. So, we have $\Delta(u^{(1)}, RS(\mathbb{F}_q, D, k)) = \Delta(u^{(1)}, v) = \Delta(\mathbf{1}v') \geq n-k$.

To prove Claim 3 and therefore Theorem 2, it remains to show the following.

Claim 4 *There exists an $a \in \mathbb{F}_q \setminus D$ such that:*

$$|\{p^{(j)}(a) : 1 \leq j \leq L\}| \geq \frac{(L-1)(q-n)}{q-n+(L-1)k}.$$

Note that we need to restrict a to $\mathbb{F}_q \setminus D$ to avoid dividing by 0.

Proof (of Claim 4). Consider $\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L], a \leftarrow \mathbb{F}_q \setminus D]$. With probability $1/L$, $j = j'$, so $p^{(j)}(a) = p^{(j')}(a)$. Otherwise, since $p^{(j)}(x) - p^{(j')}(x)$ is a non-zero polynomial of degree at most k, $p^{(j)}(a) = p^{(j')}(a)$ occurs with probability at most $k/|\mathbb{F}_q \setminus D| = k/(q-n)$. Thus, we have:

$$\begin{aligned}\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L], a \leftarrow \mathbb{F}_q \setminus D] &\leq 1/L + (1-1/L)(k/(q-n)) \\ &= \frac{q-n+(L-1)k}{L(q-n)}.\end{aligned}$$

If we sample $a \leftarrow\!\!\$\, \mathbb{F}_q \setminus D$, then the above is the expectation of $\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L]]$ over a. There exists an a where this probability is under its expectaion, i.e., we have:

$$\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L]] \leq \frac{q-n+(L-1)k}{L(q-n)}.$$

Consider the distribution of $Y = p^{(j)}(a)$ for $j \leftarrow\!\!\$\, [L]$. It takes values in the set $S = \{p^{(j)}(a) : 1 \leq j \leq L\}$. $\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L]]$ is the probability that two samples from Y are equal. So, we have:

$$\Pr[p^{(j)}(a) = p^{(j')}(a) : j, j' \leftarrow\!\!\$\, [L]] = \sum_{y \in S} \Pr[Y = y]^2.$$

By the Cauchy–Schwarz inequality (Definition 5):

$$\begin{aligned}1 &= \left(\sum_{y \in S} \Pr[Y = y]\right)^2 \\ &\leq \left(\sum_{y \in S} \Pr[Y = y]^2\right)\left(\sum_{y \in S} 1\right) \\ &\leq \left(\frac{q-n+(L-1)k}{L(q-n)}\right)(|S|).\end{aligned}$$

Thus, we have:

$$|S| \geq \frac{(L-1)(q-n)}{q-n+(L-1)k}$$

as required. □

5 Extension Fields and When the Domain is in a Subfield

In this section, we consider two fields $\mathbb{F}_{q_{sub}} \subset \mathbb{F}_{q_{ext}}$ where the field $\mathbb{F}_{q_{ext}}$ is an extension of $\mathbb{F}_{q_{sub}}$ and so $q_{ext} = q_{sub}^d$ for some $d \geq 2$. We will be considering codes of the form $RS(\mathbb{F}_{q_{ext}}, D, k)$ when $D \subseteq \mathbb{F}_{q_{sub}}$.

Codewords in $RS(\mathbb{F}_{q_{sub}}, D, k)$ are also in $RS(\mathbb{F}_{q_{ext}}, D, k)$. The following lemma immediately follows.

Lemma 3. *If $RS(\mathbb{F}_{q_{sub}}, D, k)$ is not (δ, L)-list decodable for some δ, L, then nor is $RS(\mathbb{F}_{q_{ext}}, D, k)$.*

Combining this with the list decoding capacity bound gives the following corollary.

Corollary 2. *If $k/n \geq 1 - H_{q_{sub}}(\delta) + \eta$, then $RS(\mathbb{F}_{q_{ext}}, D, k)$ is not (δ, L)-list decodable unless $L \geq q^{\Omega(\eta n)}$.*

Fenzi and Sanso [FS25] show that this has consequences for the security of SNARKs that use small fields, e.g., $q_{sub} \approx 2^{32}$. It is clear that our Conjecture 1 does not apply with $H_q(\delta) \leq 1 - \rho - \eta$ for a prime power q. We conjecture the following.

Our Conjecture 4. (List-Decodability of Reed-Solomon Codes up to List-Decoding Capacity for Extension Fields). *Conjecture 2.3 from [BGKS20] holds with $\delta \leq 1 - \rho - \eta$ replaced by $H_p(\delta) \leq 1 - \rho - \eta$ when $q = p^d$ for some prime p and $d \geq 1$.*

This conjecture is plausible; we do not know of a proof that this does not hold if all domains are allowed. It is not clear that it is maximal for the case when the characteristic p is small, e.g., $p = 2$. In that regime, there could be a conjecture with a dependence on the domain size; however, we cannot confidently give one.

What about proximity gaps for correlated agreement? Applying Theorem 1 and Corollary 1 to $RS(\mathbb{F}_{q_{sub}}, D, k)$ gives that there are $u^{(0)}, u^{(1)} \in \mathbb{F}_{q_{sub}}^n$ such that there are at least $q_{sub}/2$ values of $\lambda \in \mathbb{F}_{q_{sub}}$ with $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_{q_{sub}}, D, k)) \leq f$ for some choice of parameters. Embedding $\mathbb{F}_{q_{sub}}$ into the extension field $\mathbb{F}_{q_{ext}}$, this gives $q_{sub}/2$ values of λ out of q_{ext}, which is a small fraction $\epsilon \approx 1/2q_{sub}^{d-1}$. It would be better to combine Corollary 2 with Theorem 2, which would give $\epsilon \approx 1/k$. We can do better still by combining the techniques of Theorem 1 and Theorem 2.

Theorem 3. *Suppose $f < n - k$. Then there exists a $u^{(0)}, u^{(1)}$ such that $\Delta(u^{(1)}, RS(\mathbb{F}_{q_{ext}}, D, k)) > f$ and there are at least*

$$q_{ext} - \frac{q_{ext}^2 q_{sub}^{n-f-k} g(x)}{\binom{n}{f}}$$

values of $\lambda \in \mathbb{F}_{q_{ext}}$ *such that* $\Delta(u^{(0)} + \lambda u^{(1)}, RS(\mathbb{F}_{q_{ext}}, D, k)) \leq f$*, where*

$$g(x) = \begin{cases} \exp(x) \text{ when } x \leq 3/2 \\ \frac{\exp(2\sqrt{x})}{\sqrt{2\pi\lfloor\sqrt{x}\rfloor}} \text{ when } x > 3/2. \end{cases}$$

Proof. We will take u to be a random word from the small field, i.e., uniformly at random from $\mathbb{F}_{q_{sub}}^n$ as in the proofs of Lemma 1 and Theorem 1, but conditioned on $\Delta(u, RS(\mathbb{F}_{q_{sub}}, D, k+1)) \leq f$. (Lemma 1 proves that this happens with significant probability in the region we care about.) Then, we define $u^{(0)}$ to be the codeword with $u_i^{(1)} = u_i/(x_i - a)$ and $u^{(1)}$ be the codeword with $u_i^{(1)} = -1/(x_i - a)$ for some a, similar to the proof of Theorem 2. Now, we will take a to be an arbitrary element of $\mathbb{F}_{q_{ext}}$ that is not in any subfield.

If $u^{(0)}$ is within Hamming distance f of the evaluation of $p(x)/(x-a)$ for some polynomial p of degree at most $k-1$, then $u^{(0)} - p(a)u^{(1)}$ is within Hamming distance f of the evaluation of $(p(x) - p(a))/(x-a)$, which is a polynomial of degree at most $k-1$, so $\Delta(u^{(0)} - p(a)u^{(1)}, RS(\mathbb{F}_{q_{ext}}, D, k)) \leq f$. We need to show that there are many such polynomials with distinct evaluations $p(a)$. The condition $\Delta(u, RS(\mathbb{F}_{q_{sub}}, D, k+1)) \leq f$ ensures that there is at least one such polynomial.

We are interested in the list $v'^{(1)}, \ldots, v'^{(L)}$ of codewords $RS(\mathbb{F}_q, D, k+1)$ with $\Delta(u, v^{(i)}) \leq f$ for which the condition $\Delta(u, RS(\mathbb{F}_{q_{sub}}, D, k+1)) \leq f$ ensures $L \geq 1$. Rather than sample uniformly from this list, we will instead sample a disagreement set $F \subset [n]$ uniformly from S_f ,where S_f is the set of $F \subset [n]$ with $|F| = f$, conditioned on there being a codeword $v_F \in RS(\mathbb{F}_q, D, k+1)$ that agrees with u outside of F. In the terminology of the proof of Lemma 1, this is sampling an F uniformly from S_f conditioned on $X_F = 1$. v_F is the evaluation of a polynomial $p_F(x)$ of degree at most k when $X_F = 1$.

Next, we consider sampling F, F' independently and uniformly from S_f and conditioning on $X_F = 1$ and $X'_F = 1$ and then what is the probability that $p_F(a) = p_{F'}(a)$. Considering sampling u, F, F' independently and uniformly from $\mathbb{F}_{q_{sub}}^n, S_f, S_f$, we can write this as $\Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1]$. We can separate the case when the polynomials are identical as:

$$\begin{aligned} &\Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1] \\ &= \Pr[p_F(x) \equiv p_{F'}(x) | X_F = 1, X_{F'} = 1] + \Pr[p_F(a) = p_{F'}(a) \wedge p_F(x) \not\equiv p_{F'}(x) | X_F = 1, X_{F'} = 1] \,. \end{aligned}$$

Consider this second term $\Pr[p_F(a) = p_{F'}(a) \wedge p_F(x) \not\equiv p_{F'}(x) | X_F = 1, X_{F'} = 1]$ for fixed F, F'. Using Lemma 2 (iii), in the case $|F \cup F'| \leq n - k$, if $X_F = 1$ and $X_{F'} = 1$, then $p_F(x) \equiv p_{F'}(x)$ and so $\Pr[p_F(a) = p_{F'}(a) \wedge p_F(x) \not\equiv p_{F'}(x) | X_F = 1, X_{F'} = 1] = 0$. So, we assume that we are in the case $|F \cup F'| > n - k$. Then Lemma 2 (v) applies, so: $p_F(x) = p_{F \cup F'}(x) + z_{F \cup F'}(x) a_{F/F'}(x)$ and $p_F(x) = p_{F \cup F'}(x) + z_{F \cup F'}(x) a_{F' \setminus F}(x)$, where $z_{F \cup F'}(x) = \prod_{i \in F \cup F'}(x - x_i)$ and $p_{F \cup F'}(x), a_{F \setminus F'}(x), a_{F' \setminus F}(x)$ are distributed independently and uniformly at random from $\mathbb{F}_{q_{sub}}[x]^{\leq n - |F \cup F'| - 1}$, $\mathbb{F}_{q_{sub}}[x]^{\leq |F \setminus F'| - 1}$, $\mathbb{F}_q[x]^{\leq |F' \setminus F| - 1}$ respectively. Conditioning on $X_F = 1$ and $X_{F'} = 1$ is equivalent to conditioning on $p_F(X)$

and $p_{F'}(x)$ having degree at most $k-1$. Since $p_{F\cup F'}(x)$ has degree at most $n-|F\cup F'|-1 \leq k-1$ and $z_{F\cup F'}(x)$ has degree $|F\cup F'|$, conditioning on $X_F = 1 \wedge X_{F'} = 1$ is equivalent to conditioning on $\deg a_{F\setminus F'}(x), \deg a_{F'\setminus F}(x) \leq k-|F\cup F'|-1$. So, conditioning on $X_F = 1 \wedge X_{F'} = 1$, we have that $a_{F\setminus F'}(x)$ and $a_{F'\setminus F}(x)$ are independent and uniformly distributed on $\mathbb{F}_{q_{sub}}[x]^{\leq k-|F\cup F'|-1}$. We can write $p_F(x) - p_{F'}(x) = z_{F\cup F'}(x)b(x)$, where $b(x) = a_{F\setminus F'}(x) - a_{F'\setminus F}(x)$ is uniformly distributed on $\mathbb{F}_{q_{sub}}[x]^{\leq k-|F\cup F'|-1}$. The polynomial $z_{F\cup F'}(x) = \prod_{i\in|F\cup F'}(x-x_i)$ is not identically zero and is non-zero at a since $a \notin \mathbb{F}_{q_{sub}}$, but each $x_i \in D \subset \mathbb{F}_{q_{sub}}$. Thus, $p_F(x) - p_{F'}(x)$ is not identically 0 or has a root at a if and only if $b(x)$ is not identically 0 or has a root at a, respectively. If $b(a) = 0$, then the minimal polynomial $m_a(x)$ of a over $\mathbb{F}_{q_{sub}}$ divides $b(x)$. Since a is not in any subfield, $m_a(x)$ has degree exactly d, recalling that $q_{ext} = q_{sub}^d$. If $k-|F\cup F'|-1 < d$, then $b(a) = 0$ only when $b(x) \equiv 0$ and $\Pr[p_F(a) = p_{F'}(a) \wedge p_F(x) \not\equiv p_{F'}(x)|X_F = 1, X_{F'} = 1] = 0$. Otherwise, the elements of $F_{q_{sub}}[x]^{\leq k-|F\cup F'|-1}$ that divide $m_a(x)$ are precisely those of the form $m_a(x)c(x)$ for $c(x) \in F_{q_{sub}}[x]^{\leq k-|F\cup F'|-1-d}$. There are $q_{sub}^{k-|F\cup F'|-d}$ such elements out of $q_{sub}^{k-|F\cup F'|}$ elements of $F_{q_{sub}}[x]^{\leq k-|F\cup F'|-1}$, and since $b(x)$ is distributed uniformly, the probability that it is one of them is exactly $q_{sub}^{-d} = 1/q_{ext}$. Thus, we have $\Pr[p_F(a) = p_{F'}(a) \wedge p_F(x) \not\equiv p_{F'}(x)|X_F = 1, X_{F'} = 1] \leq 1/q_{ext}$ in all cases.

Substituting, we now have:

$$\Pr[p_F(a) = p_{F'}(a)|X_F = 1, X_{F'} = 1] \leq \Pr[p_F(x) \equiv p_{F'}(x)|X_F = 1, X_{F'} = 1] + 1/q_{ext}\ .$$

We now turn our attention to the first term. Bayes' rule gives:

$$\Pr[p_F(x) \equiv p_{F'}(x) \wedge X_F = 1 \wedge X_{F'} = 1] = \frac{\Pr[p_F(x) \equiv p_{F'}(x)|X_F = 1, X_{F'} = 1]}{\Pr[X_F = 1, X_{F'} = 1]},$$

and we must deal with the numerator and denominator separately. For fixed F, F', Lemma 2 (ii),(iv), and (vi) bound the equivalent probabilities as $\Pr[X_F = 1 \wedge X_{F'} = 1 \wedge p_F(x) \equiv p_{F'}(x)] = q^{-(n-k-|F\cap F'|)}$ and $\Pr[X_F = 1 \wedge X_{F'} = 1] = q_{sub}^{-\min\{n-k-|F\cap F'|, 2(n-k-f)\}}$. For the denominator, we can simply argue that for fixed F, F', $\Pr[X_F = 1 \wedge X_{F'} = 1] \geq q_{sub}^{-2(n-k-f)}$, so taking F, F' as random variables, we also have $\Pr[X_F = 1 \wedge X_{F'} = 1] \geq q_{sub}^{-2(n-k-f)}$. For the numerator, we need to sum over F, F', giving a series very similar to that in the proof of Lemma 1:

$$\begin{aligned}
\binom{n}{f}^2 \Pr[X_F = 1 \wedge X_{F'} = 1 \wedge p_F(x) \equiv p_{F'}(x)] &= \sum_{F,F'} q_{sub}^{-(n-k-|F\cap F'|)} \\
&= \sum_{\ell=0}^{\min\{f,n-f\}} \binom{n}{f}\binom{f}{\ell}\binom{n-f}{\ell} q_{sub}^{-(n-k-f+\ell)} \\
&= \binom{n}{f} q_{sub}^{-(n-f-k)} \sum_{\ell=0}^{\min\{f,n-f\}} \binom{f}{\ell}\binom{n-f}{\ell} q_{sub}^{-\ell}
\end{aligned}$$

$$\leq \binom{n}{f} q_{sub}^{-(n-f-k)} \sum_{\ell=0}^{\infty} f^{\ell}(n-f)^{\ell} q^{-\ell}/\ell!^2$$
$$\leq \binom{n}{f} q_{sub}^{-(n-f-k)} g(f(n-f)/q)),$$

where the final inequality makes use of Claim 2. Putting these together, we obtain:

$$\Pr[p_F(x) \equiv p_{F'}(x) | X_F = 1, X_{F'} = 1] \leq \frac{q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}} .$$

Substituting this gives:

$$\Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1] \leq \frac{q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}} + 1/q_{ext} .$$

The probability $\Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1]$ is an expectation over values of u with $\Delta(u, RS(\mathbb{F}_{q_{sub}}, D, k+1)) \leq f$. There must be some concrete u with $\Delta(u, RS(\mathbb{F}_{q_{sub}}, D, k+1)) \leq f$ for which this probability is at most this expectation. For this fixed u we still have:

$$\Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1] \leq \frac{q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}} + 1/q_{ext} .$$

Let S be the set of values $-p_F(a)$ for some $F \in S_F$. By the Cauchy–Schwarz inequality (Definition 5):

$$\begin{aligned}
1 &= \left(\sum_{y \in S} \Pr[Y = y] \right)^2 \\
&\leq \left(\sum_{y \in S} \Pr[Y = y]^2 \right) \left(\sum_{y \in S} 1 \right) \\
&= \left(\sum_{y \in S} \Pr[p_F(a) = p_{F'}(a) = -y | X_F = 1, X_{F'} = 1] \right) \left(\sum_{y \in S} 1 \right) \\
&= \Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1] |S|,
\end{aligned}$$

so we have:

$$\begin{aligned}
|S| &\geq \Pr[p_F(a) = p_{F'}(a) | X_F = 1, X_{F'} = 1] \\
&\geq 1 / \left(\frac{q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}} + 1/q_{ext} \right) \\
&= q_{ext} / \left(\frac{q_{ext} q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}} + 1 \right)
\end{aligned}$$

$$
\begin{aligned}
&= q_{ext} \frac{\binom{n}{f}}{q_{ext} q_{sub}^{n-f-k} g(f(n-f)/q_{sub}) + \binom{n}{f}} \\
&= q_{ext} - \frac{q_{ext}^2 q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{q_{ext} q_{sub}^{n-f-k} g(f(n-f)/q_{sub}) + \binom{n}{f}} \\
&\leq q_{ext} - \frac{q_{ext}^2 q_{sub}^{n-f-k} g(f(n-f)/q_{sub})}{\binom{n}{f}}
\end{aligned}
$$

□

In light of Theorem 3, we again suggest replacing δ with $H_p(\delta)$ for characteristic p. Again, there may be a more maximal but still plausible conjecture for $p \ll n$; however, in light of Theorem 1.6 of Ben-Sasson, Carmon, Haböck, Kopparty and Saraf [BSCH+25], we cannot confidently give one.

Our Conjecture 5. (Proximity gap and correlated agreement for RS codes for extension fields)). *Conjecture 8.4 from [BCI+23] holds with $\delta \leq 1 - \rho - \eta$ replaced by $\delta \leq 1 - H_p(\delta) - 1/n - \eta$ when $q = p^d$ for some prime p and $d \geq 1$.*

Our Conjecture 6. (Mutual correlated agreement for extension fields). *Conjecture 4.12 from [ACFY24b] holds with $0 < \delta < 1 - \rho - \eta$ replaced by $0 < H_p(\delta) < 1 - 1/n - \rho - \eta$ when $q = p^d$ for some prime p and $d \geq 1$.*

Acknowledgements. We would like to thank Felipe Voloch, Gal Arnon, Mahdi Sedaghat, and Angus Gruen for their careful proofreading and Yuval Ishai for an insightful discussion and helping us understand the relation of our work to his.

References

ACFY24a. Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. STIR: Reed-solomon proximity testing with fewer queries. In Leonid Reyzin and Douglas Stebila, editors, *CRYPTO 2024, Part X*, volume 14929 of *LNCS*, pages 380–413. Springer, Cham, August 2024.

ACFY24b. Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. STIR: Reed–solomon proximity testing with fewer queries. Cryptology ePrint Archive, Report 2024/390, 2024.

BBHR18. Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast reed-solomon interactive oracle proofs of proximity. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *ICALP 2018*, volume 107 of *LIPIcs*, pages 14:1–14:17. Schloss Dagstuhl, July 2018.

BCI+23. Eli Ben-Sasson, Dan Carmon, Yuval Ishai, Swastik Kopparty, and Shubhangi Saraf. Proximity gaps for reed-solomon codes. *J. ACM*, 70(5):31:1–31:57, 2023.

BGKS20. Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. DEEP-FRI: Sampling outside the box improves soundness. In Thomas Vidick, editor, *ITCS 2020*, volume 151, pages 5:1–5:32. LIPIcs, January 2020.

BSCH+25. Eli Ben-Sasson, Dan Carmon, Ulrich Haböck, Swastik Kopparty, and Shubhangi Saraf. On proximity gaps for reed–solomon codes. Cryptology ePrint Archive, Paper 2025/2055, 2025.

CS25. Elizabeth C. Crites and Alistair Stewart. A plausible attack on the adaptive security of threshold schnorr signatures. In Yael Tauman Kalai and Seny F. Kamara, editors, *CRYPTO 2025, Part VI*, volume 16005 of *LNCS*, pages 457–479. Springer, Cham, August 2025.

CW04. Qi Cheng and Daqing Wan. On the list and bounded distance decodibility of the reed-solomon codes (extended abstract). In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 335–341. IEEE Computer Society, 2004.

DL78. Richard A. DeMillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Inf. Process. Lett.*, 7(4):193–195, 1978.

Eli57. Peter Elias. List decoding for noisy channels. *Research Laboratory of Electronics, Massachusetts Institute of Technology*, 1957.

FS25. Giacomo Fenzi and Antonio Sanso. Small-field hash-based SNARGs are less sound than conjectured. Cryptology ePrint Archive, Paper 2025/2197, 2025.

GGG18. Venkata Gandikota, Badih Ghazi, and Elena Grigorescu. Np-hardness of reed-solomon decoding, and the prouhet-tarry-escott problem. *SIAM J. Comput.*, 47(4):1547–1584, 2018.

GMW25. Albert Garreta, Nicolas Mohnblatt, and Benedikt Wagner. A simplified round-by-round soundness proof of FRI. *IACR Cryptol. ePrint Arch.*, page 1993, 2025.

GRS14. Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*. Cambridge University Press, 2014.

GV05. Venkatesan Guruswami and Alexander Vardy. Maximum-likelihood decoding of reed-solomon codes is NP-hard. In *16th SODA*, pages 470–478. ACM-SIAM, January 2005.

MS81. Robert J. McEliece and Dilip V. Sarwate. On sharing secrets and reed-solomon codes. *Commun. ACM*, 24(9):583–584, 1981.

Mul54. David E. Muller. Application of boolean algebra to switching circuit design and to error detection. *Trans. I R E Prof. Group Electron. Comput.*, 3(3):6–12, 1954.

Ore22. Øystein Ore. On higher congruences. *Norsk Mat. Forenings Skrifter Ser. I*, 1922.

RS60. I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, 1960.

Sch80. Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

Sha79. Adi Shamir. How to share a secret. *Communications of the Association for Computing Machinery*, 22(11):612–613, November 1979.

Zip79. Richard Zippel. Probabilistic algorithms for sparse polynomials. In Edward W. Ng, editor, *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, Marseille, France, June 1979, Proceedings*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer, 1979.