Linear Proximity Gap for Linear Codes within the 1.5 Johnson Bound

Yiwen Gao*, Haibin Kan†and Yuan Li‡

November 28, 2024

Abstract

We establish a linear proximity gap for linear codes within the one-and-a-half Johnson bound. Specifically, we investigate the *proximity gap* for linear codes, revealing that any affine subspace is either entirely δ-close to a linear code or nearly all its members are δ-far from it. When δ is within the one-and-a-half Johnson bound, we prove an upper bound on the number of members (in the affine subspace) that are δ-close to the linear code for the latter case. Our bound is linear in the length of codewords. In comparison, Ben-Sasson, Carmon, Ishai, Kopparty and Saraf [FOCS 2020] work on Reed-Solomon codes and prove a linear bound when δ is within the unique decoding bound and a quadratic bound when δ is within the Johnson bound. Note that when the minimum relative distance of the linear code is bigger than 0.77, the one-and-a-half Johnson bound is better than the unique decoding bound.

Proximity gaps for linear codes have implications in various code-based protocols. In many cases, a stronger property than individual distance—known as *correlated agreement*—is required, i.e., functions in the affine subspace are not only δ-close to a linear code but also agree on the same evaluation domain. Our results support this stronger property. Furthermore, *mutual correlated agreement*, the further strengthening property, is also supported.

1 Introduction

Linear codes are a class of error-correcting codes. They are fundamental objects of study in algebraic coding theory and theoretical computer science. A linear code is an error-correcting code for which any linear combination of codewords is also a codeword. Linear codes have a wide range of applications. For example, many protocols in areas such as blockchain, distributed storage, and cryptography utilize different linear codes such as Reed-Solomon codes[RS60] and Reed-Muller codes[Mul54] as essential building blocks. In some protocols, the soundness relies on the existence of a series of vectors that are close to the linear code (in relative Hamming distance). Consequently, it is critical to efficiently identify vectors that are far from the linear code.

The *Proximity Testing* problem of linear codes involves a verifier determining whether a given codeword $\pi \in \mathbb{F}_q^n$ is a member of a given linear code C or is far from all the members of C. The verifier has limited query access to π, and an untrusted prover may assist the verifier. We consider this problem under the interactive oracle proofs of proximity (IOPP) model [BCS16] (also called

*School of Computer Science, Fudan University, Shanghai 200433, China. Email: ywgao21@m.fudan.edu.cn
†School of Computer Science, Fudan University, Shanghai 200433, China. Email: hbkan@fudan.edu.cn
‡School of Computer Science, Fudan University, Shanghai 200433, China. Email: yuan_li@fudan.edu.cn

probabilistically checkable interactive proofs of proximity in [RRR16]). This model combines aspects of probabilistically checkable proof (PCPs) and interactive proofs (IPs). The prover provides the verifier with auxiliary proofs, and the verifier has oracle access to the messages from the prover.

For a batch of codewords $\mathbf{u} = \{u_0, \ldots, u_l\} \subset \mathbb{F}_q^n$, one can implement a protocol for the proximity testing problem on each codeword to ensure that they are all close to C. However, this approach is inefficient. [RVW13] provides an approach: randomly choose a vector u' in the span of $\mathbf{u}$ (denoted by $\mathrm{span}(\mathbf{u})$) and check if u' is close to C. The soundness proof of this method raises an important question: *If $\exists u_i \in \mathbf{u}$ that is far from all the members of C, can we prove u' is far from C with high probability?* Many previous works have explored this question and provided positive answers. This property is referred to as the *proximity gap* for linear codes, as formally defined by Ben-Sasson, Carmon, Ishai, Kopparty, and Saraf in [Ben+20b].

1.1 Our results

We present proximity gaps for linear codes within the one-and-a-half Johnson bound. Our result is linear in the length of the codeword. We begin by considering a simplified case where $\mathbf{u} = \{u_0, u_1\}$. Here, $u' = u_0 + \alpha u_1, \alpha \in \mathbb{F}_q$ is over a line. We have the following result.

Theorem 1 (Informal). *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1 - \Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1 - \Delta_C)^{-\frac{1}{3}}$. Let $u_0, u_1 \in \mathbb{F}_q^n$ be two codewords. If there exists $i \in \{0, 1\}$ such that $\delta(\pi_i, C) > \delta$, then*

$$\mathbb{P}_{\alpha \in \mathbb{F}_q}(\delta(u_0 + \alpha u_1, C) \leq \delta) < \frac{2\Delta_C n}{9(1 - \Delta_C)\eta^3 |\mathbb{F}_q|}. \tag{1}$$

We prove that if either u_0 or u_1 is δ-far from the linear code, then in the probability that is linear in the code length, u' is δ-far from the linear code. The formal statement of this theorem is presented in Theorem 4, which describes the result in its contrapositive form. Furthermore, the formal theorem is stronger. We utilize the concept of *correlated agreement*, as defined in [Ben+20b]. A series of functions $u_0, \ldots, u_l$ have δ-correlated agreement with C if there exists a sufficiently large subdomain $D \subseteq \{1, \ldots, n\}$ and $v_0, \ldots, v_l \in C$ such that

$$|D| \geq (1 - \delta)n \text{ and } u_i[D] = c_i[D], 1 \leq i \leq l,$$

where $u_i[D](c_i[D])$ denotes the sub-codeword of $u_i(c_i)$ on D. The definition of correlated agreement is relevant in the context of real-world protocol applications. Notice that even if all of u_i are δ-close to C, they may not have δ-correlated agreement. Our formal theorem supports this stronger notion of agreement; specifically, if u_0 and u_1 do not have δ-correlated agreement, (1) holds. Furthermore, we prove the result in an improved version of the correlated agreement, called mutual correlated agreement in [Arn+24b]. This result can be found in Theorem 5.

We prove our result under the generalized case that $\mathbf{u} = \{u_0, \cdots, u_l\}$.

Theorem 2 (Informal). *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1 - \Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1 - \Delta_C)^{-\frac{1}{3}}$. Let $u_0, \cdots, u_l \in \mathbb{F}_q^n$ be a batch of codewords. If $\exists u_i$ such that $\delta(u_i, C) > \delta$, then*

$$\mathbb{P}_{\langle \alpha_1, \ldots, \alpha_l \rangle \in \mathbb{F}_q^l}(\delta(u_0 + \alpha_1 u_1 + \cdots + \alpha_l u_l, C) \leq \delta) < \left(\frac{2\Delta_C n}{9(1 - \Delta_C)\eta^3 |\mathbb{F}_q|}\right) \cdot l.$$

This result also supports the (mutual) correlated agreement version. The formal theorem can be found in Theorem 6.

1.2 Related work

Let $C \subseteq \mathbb{F}_q^n$ be any linear code and $\Delta_C \in [0,1]$ be its minimal relative distance. Suppose $u_i \in \mathbf{u}$ is δ-far from C in relative Hamming distance, denoted by $\delta(u_i, C) > \delta$. In [Ame+17], Ames, Hazay, Ishai, and Venkitasubramaniam proved that when $\delta < \Delta_C/4$, with high probability $u' \in \mathrm{span}(\mathbf{u})$ is δ-far from C. When $u' \in \mathrm{span}(\mathbf{u})$ is on a line, i.e., $\mathbf{u} = \{u_0, u_1\}$, Ben-Sasson, Kopparty, and Saraf [BKS18] demonstrated that when $\delta < 1 - \sqrt[4]{1-\Delta_C}$ (the double Johnson bound), with high probability (related to a small constant ϵ) $u' \in \mathrm{span}(\mathbf{u})$ is $(\delta - \epsilon)$-far from C. Later, Ben-Sasson, Goldberg, Kopparty, and Saraf improved the bound to $1 - \sqrt[3]{1-\Delta_C}$ (the 1.5 Johnson bound) in [Ben+20a]. Furthermore, they showed that their result is tight for certain Reed-Solomon (RS) codes. Especially, when C is an RS code, [Ben+20b] bounded the probability that u' is δ-close to C for the unique decoding bound $\Delta_C/2$ and the Johnson bound $\sqrt{1-\Delta_C}$ respectively.

	δ bound	u' distance	Probability	Code
[Ame+17]	$\Delta_C/4$	δ	$(\delta+1)/\lvert\mathbb{F}_q\rvert$	Linear code
[BKS18]	$J_\epsilon(J_\epsilon(\Delta_C))$	$\delta-\epsilon$	$2/(\epsilon^3\lvert\mathbb{F}_q\rvert)$	Linear code
[Ben+20a]	$1-\sqrt[3]{1-\Delta_C+\epsilon}$	$\delta-\epsilon$	$2/(\epsilon^2\lvert\mathbb{F}_q\rvert)$	Linear code
[Ben+20b]	$1-\sqrt{1-\Delta_C}-\epsilon$	δ	$((1-\Delta_C)^2n^2)/((2\epsilon)^7\lvert\mathbb{F}_q\rvert)$	RS
[Ben+20b]	$\Delta_C/2$	δ	$n/\lvert\mathbb{F}_q\rvert$	RS
[Zei24]	$J_\epsilon(J_\epsilon(\Delta_C))$	δ	$(n+2\sqrt{n})/(4\lvert\mathbb{F}_q\rvert)$	Linear code
This work	$1-\sqrt[3]{1-\Delta_C}-\epsilon$	δ	$2\Delta_C n/(9(1-\Delta_C)\epsilon^3\lvert\mathbb{F}_q\rvert)$	Linear code

Table 1: When $\mathbf{u} = \{u_0, u_1\}$ and $u_0(u_1)$ is δ-far from the code C, the provable probability that randomly chosen $u' \in \mathrm{span}(\mathbf{u})$ is δ (or $\delta - \epsilon$) close to C. δ bound is the upper bound. Δ_C is the minimal relative distance of C. $J_\epsilon(\Delta_C) = 1 - \sqrt{1-\Delta_C(1-\epsilon)}$ is the Johnson bound.

Table 1 compares our result with previous results. [Ben+20b] provides the linear proximity gap for RS codes under the unique decoding bound $\Delta_C/2$ and the quadratic proximity gap under the Johnson bound $1 - \sqrt{1-\Delta_C}$. When the one-and-a-half Johnson bound $1 - \sqrt[3]{1-\Delta_C}$ is better than the unique decoding bound (related to Δ_C), we improve the provable proximity gap to linear. Additionally, [Ben+20b] conjectures that we can prove the proximity gap when $\delta < \Delta_C$. We will briefly introduce this conjecture in Section 5.3. Figure 1 compares various bounds. When $\Delta_C \geq 0.77$, the one-and-a-half Johnson bound is better than the unique decoding bound. And we make improvements in this case.

This is the third version of our work. Our first version provided a framework for this work [GKL24]. However, there was a mistake in the proof and we withdrew it. A month later, we published the second version of this work, fixing the mistake. In the second version, we focused on RS codes and proved the linear proximity gap for RS codes when δ is under the 1.5 Johnson bound. [Zei24] was published a week after the publication of our second version. [Zei24] also uses similar combinatorial methods. Zeilberger's work proves the linear proximity gap for linear codes under the double Johnson bound and results in this work support the mutual correlated agreement (strong correlated agreement in this work) as well. [Zei24] pointed out that our previous work can be extended to general linear codes. Inspired by Zeilberger's observation, we extend our result from RS codes to any linear code using the same technique and have this latest version.

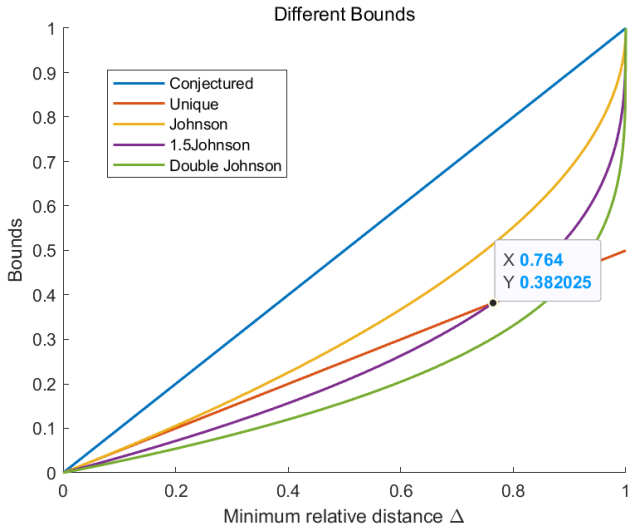

Figure 1: Bounds for linear codes

1.3 Applications

Reed-Solomon (RS) codes [RS60] are a class of linear error-correcting codes. Proximity gaps for RS codes provide provable soundness for a variety of protocols. For example, the Fast RS IOPP (known as FRI) [Ben+18] is a widely used IOPP for RS codes due to its high efficiency. FRI can be used to verify whether a given function belongs to an RS code or is far from it. It is implemented as a subprotocol in many (zk)SNARKs and real-world systems [Ben+19][KPV22][Sta23][Pol][Zha+20][Xie+22]. Our results fit the correlated agreement condition. The correlated agreement of FRI is used to prove the (knowledge) soundness of protocols that use FRI as a sub-protocol [Sta23], as well as to prove the round-by-round soundness of FRI [Sta23][Blo+23]. Furthermore, our results can also be applied to generalizations of FRI[ZCF23][Arn+24b][Arn+24a] that use foldable codes defined in [ZCF23]. The previous best provable mutual correlated agreement for general foldable codes is under the unique decoding bound[Arn+24b].

We introduce the application of our results in FRI, which improves the soundness analysis. Previously, Ben-Sasson, Carmon, Ishai, Kopparty, and Saraf established the best provable soundness of FRI in [Ben+20b], utilizing elegant mathematical techniques. They proved the soundness error bound of FRI is

$$\epsilon_{\text{FRI}} \leq \max\left\{O\left(\frac{(1-\Delta_C)^2}{\eta^7}\cdot\frac{n^2}{|\mathbb{F}_q|}\right),(1-\delta)^t\right\} \tag{2}$$

when $\delta \leq 1-\sqrt{1-\Delta_c}-\eta$. Let t represent the iteration time during the QUERY phase in FRI. It is important to note that the first term is a constant dependent on the parameters. For small values of t, the second term dominates the inequality. Furthermore, this term decreases as t increases. Consequently, when t becomes sufficiently large, the first term establishes a provable upper bound on the soundness error of FRI. We will introduce the protocol in detail in Section 6. We provide an alternative soundness error bound of FRI:

$$\epsilon_{\text{FRI}} \leq \max\left\{O\left(\frac{1}{(1-\Delta_C)\eta^3}\cdot\frac{n}{|\mathbb{F}_q|}\right),(1-\delta)^t\right\} \tag{3}$$

when $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$. When t is large, the first term dominates the soundness error bound. In practical applications, n is large and significantly influences the soundness error bound. Consequently, our bound indicates that FRI can provide enhanced security. However, when t is small, the previous bound is more advantageous. In practical applications, we can select the minimum of these two bounds.

2 Technical overview

In this section, we outline the overall idea behind our proof of the main result, Theorem 3. We take a new approach to address this problem through combinatorial methods. This section introduces the tools we employ and explains how this can be done.

2.1 Bad combining points

Let $\mathbb{F}_q$ be a finite field and $C \subseteq \mathbb{F}_q^n$ be a linear code with generator matrix $\mathbf{G} \in \mathbb{F}_q^k \times \mathbb{F}_q^n$, i.e. C is a linear subspace and for every message $\mathbf{m} \in \mathbb{F}_q^k$, there is a member $c \in C$ such that $c = \mathbf{m} \cdot \mathbf{G}$. Denote by Δ_C the minimum relative distance of C. For any codeword $\pi \in \mathbb{F}_q^n$, let $\delta(\pi, C)$ denote the relative Hamming distance between π and C.

Let π_1, π_2 be two codewords and $0 < \delta \leq 1$ such that at least one of the codewords is δ-far from C in relative Hamming distance, i.e., $\delta(\pi_1, C) > \delta$ or $\delta(\pi_2, C) > \delta$. (Actually, we only need π_1, π_2 do not have δ correlated agreement. See Section 3.2 for details.) Let $\alpha \in \mathbb{F}_q$ be a randomly chosen combining point and $\pi_1 + \alpha\pi_2$ be the combining result. There are cases that $\delta(\pi_1 + \alpha\pi_2, C) \leq \delta$. For example, let $C = \{\langle x, x, x, x\rangle \mid x \in \mathbb{F}_{17}\} \subseteq \mathbb{F}_{17}^4$ be a linear code containing codeword with the same value on all the locations. Let $\pi_1 = \langle 1, 1, 1, 2\rangle, \pi_2 = \langle 3, 3, 3, 2\rangle$. Then we have $\delta(\pi_1, C) = \delta(\pi_2, C) = \frac{1}{4}$. However, when $\alpha = 1$, we have $\pi_1 + \alpha\pi_2 = \langle 4, 4, 4, 4\rangle \in C$. Then we say $\alpha = 1$ is a *bad combining point*. More precisely, define the bad combining set to be

$$\mathrm{Bad}(\pi_1, \pi_2) \triangleq \{\alpha \in \mathbb{F}_q : \delta(\pi_1 + \alpha\pi_2, C) \leq \delta)\}.$$

Our goal is to prove that the number of bad combining points is linear in the codeword length, i.e., we prove that

$$|\mathrm{Bad}(\pi_1, \pi_2)| = O_{\Delta_C, \eta(n)}$$

when δ is under the one-and-a-half Johnson bound, i.e., $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$.

2.2 Partition the bad combining points into blocks

Many previous studies focus on combining results and use the list-decoding skill to restrict the number of bad combining points. We provide a new approach to deal with this problem. Instead of the combining result $\pi_1 + \alpha\pi_2$, we focus on the origin codewords π_1, π_2. We pay attention to some sub-codewords of π_1, π_2 and transform the problem into a combinatorics problem.

More precisely, let $\alpha \in \mathrm{Bad}(\pi_1, \pi_2)$ be a bad folding point, and $c_\alpha \in C$ is the closest codeword of $\pi_1 + \alpha\pi_2$. The following sub-domain is related to α :

$$B_\alpha^* = \{j \in \{1, \ldots, n\} \mid (\pi_1 + \alpha\pi_2)[j] = c_\alpha[j]\}.$$

For two distinct bad combining points $\alpha, \beta \in \mathrm{Bad}(\pi_1, \pi_2)$, we prove the following lemma:

Lemma 1. *[Informal] For any distinct $\alpha, \beta \in \mathrm{Bad}(\pi_1, \pi_2)$, there exists a pair of cordwords $c_1, c_2 \in C$ such that*

$$\pi_1[B_\alpha^* \cap B_\beta^*] = c_1[B_\alpha^* \cap B_\beta^*] \text{ and } \pi_2[B_\alpha^* \cap B_\beta^*] = c_2[B_\alpha^* \cap B_\beta^*].$$

This means the restrictions of π_1, π_2 on the intersection agree with restrictions of some elements in C respectively. Use the example $\pi_1 = \langle 1,1,1,2\rangle, \pi_2 = \langle 3,3,3,2\rangle$ again. When $\alpha = 1$, we have $\pi_1 + \alpha\pi_2 = \langle 4,4,4,4\rangle \in C$. This implies $B_\alpha^* = \{1,2,3,4\}$. Let $\beta = -1$, we have $\pi_1 + \beta\pi_2 = \langle -2,-2,-2,0\rangle$. The closest codeword of $\pi_1 + \beta\pi_2$ in C is $\langle -2,-2,-2,-2\rangle$ and $B_\beta^* = \{1,2,3\}$. Then, $B_\alpha^* \cap B_\beta^* = \{1,2,3\}$ and we have $c_1 = \langle 1,1,1,1\rangle, c_2 = \langle 3,3,3,3\rangle$.

Because of the use of Corrádi's lemma, which will be introduced later, we extract a series of subsections of $B_\alpha^*, \alpha \in \mathrm{Bad}(\pi_1, \pi_2)$ such that these subsections have the same size. For a given $0 < \delta < 1$, we define notice that $|B_\alpha^*| \geq (1-\delta)n$ for $\alpha \in \mathrm{Bad}(\pi_1, \pi_2)$. We define

$$B_\alpha : \text{The set of first } (1-\delta)n \text{ elements of } B_\alpha^*, \text{ i.e., } |B_\alpha| = (1-\delta)n.$$

Notice that $B_\alpha \cap B_\beta \subseteq B_\alpha^* \cap B_\beta^*$, Lemma 1 still holds on $B_\alpha \cap B_\beta$. Using these definitions, we can partition $\mathrm{Bad}(\pi_1, \pi_2)$ into blocks.

Our partition is based on some *long* codewords in C. More precisely, let $0 < \xi \leq 1$ and $D \subseteq \{1, \ldots, n\}$ satisfying $|D| \geq \xi n$. If $\pi_1[D], \pi_2[D]$ agrees with some pair of codewords $c_1, c_2 \in C$, i.e., $\pi_1[D] = c_1[D], \pi_2[D] = c_2[D]$, then we say c_1, c_2 is a pair of long codewords contained in π_1, π_2. There may be many such long pair of codewords $(c_1^{(1)}, c_2^{(1)}), \ldots, (c_1^{(s)}, c_2^{(s)})$, denote by $D_i, 1 \leq i \leq s$ the maximal agree domains of $(c_1^{(i)}, c_2^{(i)}), 1 \leq i \leq s$ and π_1, π_2, i.e., $\pi_1[D_i] = c_1^{(i)}[D_i], \pi_2[D_i] = c_2^{(i)}[D_i]$. We partition the set of bad combining points $\mathrm{Bad}(\pi_1, \pi_2)$ based on Algorithm 1.

Algorithm 1 Partition Bad Folding Points(Informal)

```
input: Bad(π1, π2)
initialization: r = 0
set X* = Bad(π1, π2)
while X* ≠ ∅ do
  r = r + 1
  pick an arbitrary x ∈ X* and let α_i = x
  let A_i = {β ∈ Bad(π1, π2) : |B_β ∩ B_{α_i}| ≥ ξn}
  X* = X* − A_i
end while
return A_1, ..., A_r and α_1, ..., α_r
```

Let $\{A_1, \ldots, A_r\}$ be the output blocks of the algorithm and $\{\alpha_1, \ldots, \alpha_r\}$ be the corresponding represent elements. For a block $A \in \{A_1, \ldots, A_r\}$ and its represent element α, we have $|B_\beta \cap B_\alpha| \geq \xi n$. Lemma 1 implies a pair of long codewords $c_1, c_2 \in C$ are contained in π_1, π_2 respectively on both B_α and B_β. On the other hand, for distinct represent elements $\alpha_i, \alpha_j \in \{\alpha_1, \ldots, \alpha_r\}$, we have $|B_{\alpha_i} \cap B_{\alpha_j}| < \xi n$.

The number of blocks is limited. Let $\delta \leq 1 - \sqrt[3]{\rho} - \eta$ and $\xi = (1-\delta)^2 - \eta'$, where $\eta' = \frac{3\sqrt[3]{\rho}\eta^2}{2}$, we prove $r \leq \frac{1-\rho}{\eta'}$. Focusing on the represent elements $\{\alpha_1, \ldots, \alpha_r\}$, we have

- $|B_{\alpha_i}| = (1-\delta)n, 1 \leq i \leq r$;

- $|B_{\alpha_i} \cap B_{\alpha_j}| < \xi n = \left((1-\delta)^2 - \eta'\right) n, 1 \leq i < j \leq r.$

Notice that $\bigcup_{i=1}^{r} B_{\alpha_i} \subseteq \{1, \ldots, n\}$. Using the following lemma by Corrádi, we can restrict the number of blocks.

Lemma 2 (Corrádi 1969 [Juk11]). *Let $B_1, \ldots, B_r$ be s-element sets. If $|B_i \cap B_j| \leq k$ for any distinct $i, j \in \{1, 2, \ldots, r\}$, then*

$$\left| \bigcup_{i=1}^{r} B_i \right| \geq \frac{s^2 r}{s + (r-1)k}.$$

Calculating directly, we have

$$r < \frac{1-\rho}{\eta'}.$$

The upper bound of the number of blocks is a constant and is independent of $|L|$, the length of the code word.

2.3 Partition the blocks into equivalence classes

We further restrict the number of elements in each block. Let $A \in \{A_1, \ldots, A_r\}$ be a block and α be the corresponding represent element, i.e., for all $\beta \in A$, we have $|B_\alpha \cap B_\beta| \geq \xi |L|$. Lemma 1 implies a pair of long codewords $c_1, c_2 \in C$ contained in π_1, π_2 on both B_α and B_β. Denote by $(c_1^{(1)}, c_2^{(1)}), \ldots, (c_1^{(s)}, c_2^{(s)})$ all the pairs of long codewords on B_α, i.e., $\exists D_i \subseteq B_\alpha$ such that $|D_i| \geq \xi |L|$ and $\pi_1[D_i] = c_1^{(i)}[D_i], \pi_2[D_i] = c_2^{(i)}[D_i]$. Furthermore, for all $\beta \in A \setminus \{\alpha\}$, we can find one and only one codeword pair $(c_1, c_2) \in \{(c_1^{(1)}, c_2^{(1)}), \ldots, (c_1^{(s)}, c_2^{(s)})\}$ such that $\pi_1[B_\alpha \cap B_\beta] = c_1[B_\alpha \cap B_\beta], \pi_2[B_\alpha \cap B_\beta] = c_2[B_\alpha \cap B_\beta]$. We define an equivalence relation $\mathcal{R}$ on $A \setminus \{\alpha\}$:

$(\beta_1, \beta_2) \in \mathcal{R} \iff$ The codewords pairings in C decided by $B_\alpha \cap B_{\beta_1}$ and $B_\alpha \cap B_{\beta_2}$ are the same.

The number of equivalence classes is limited. We restrict the number of long low-degree polynomials $p_1, \ldots, p_s$ contained in P_α. We have

- $|D_i| \geq \xi |L|, 1 \leq i \leq s$ according to the definition of p_i.
- $|D_i \cap D_j| < (1 - \Delta_C) n, 1 \leq i < j \leq s$ because c_i, c_j are distinct codewords in C and Δ_C is the minimum relative distance of C.
- $|D_i| \leq (1-\delta)|L|$ since f is δ-far from $\mathrm{RS}[\mathbb{F}_q, L, \rho]$.

The first two conditions are similar to the condition of Corrádi's lemma(Lemma 2). But we have $|D_i| \geq \xi |L|$ instead of $|D_i| = \xi |L|$ in this case. As a result, we use the third condition and follow the proof of Corrádi's lemma to prove

$$s < \frac{1}{3\eta \rho^{\frac{2}{3}}}$$

when $\delta \leq 1 - \sqrt[3]{\rho} - \eta, \xi = (1-\delta)^2 - \eta'$ and $\eta' = \frac{3\sqrt[3]{\rho}\eta^2}{2}$. The upper bound of the number of equivalence classes is a constant and is independent of n, the length of the codeword.

The size of each equivalence class is bounded. For any equivalence class $\{\beta_1, \ldots, \beta_t\}$. Suppose c_1, c_2 are the codewords related to the equivalence class and D is the maximal agree

domain of π_1, π_2 and c_1, c_2, i.e., $\pi_1[D] = c_1[D], \pi_2[D] = c_2[D]$. Since we have $\delta(\pi_1, C) > \delta$ or $\delta(\pi_2, C) > \delta$ (or π_1, π_2 do not have δ correlated agreement), $|D| < (1-\delta)n$. On the other hand, for any $\beta_i, 1 \leq i \leq t$ in the equivalence class, we have $|B_{\beta_i}| = (1-\delta)|L|$ according to the definition of P_{β_i}. As a result, $|B_{\beta_i} \setminus D| \geq 1$. We prove that

$$(B_{\beta_i} \setminus D) \cap \left(B_{\beta_j} \setminus D\right) = \emptyset, i \neq j.$$

For any $j \in B_{\beta_i} \setminus D$, we have $\pi_1[j] \neq c_1[j]$ or $\pi_2[j] \neq c_2[j]$. According to the definition of B_{β_i}, we have

$$(\pi_1 + \beta_i \pi_2)[j] = (c_1 + \beta_i c_2)[j].$$

The above equation implies that $\pi_2[j] \neq c_2[j]$, otherwise, we have $\pi_1[j] = c_1[j]$, a contradiction. Thus, we can transform the equation into

$$\beta_i = \frac{\pi_1[j] - c_1[j]}{c_2[j] - \pi_2[j]}.$$

Notice that once j is fixed, the right side of the above equation is fixed. Since β_j, β_i are distinct, $x \notin B_{\beta_j} \setminus D$.

Thus,

$$|\{1, \ldots, n\} \setminus D| \geq \left|\bigcup_{i=1}^{t} (B_{\beta_i} \setminus D)\right| = \sum_{i=1}^{t} |B_{\beta_i} \setminus D| \geq t.$$

The size of each equivalence class is bounded by $|[B] \setminus D|$.

Combining all these above, we finish our proof of the main theorem, i.e., the number of bad combining points is linear in the length of the code word.

3 Preliminaries

Let $\mathbb{F}_q$ denote the finite field with q elements. Denote by $\mathbb{F}_q^\times = \mathbb{F}_q \setminus \{0\}$ the cyclic multiplicative group. Let $n \in \mathbb{N}^+$ and $\pi \in \mathbb{F}_q^n$ be a codeword of length n. For $j \in \{1, \ldots, n\}$, denote by $\pi[j]$ the element in π on the j^{th} location. Let $D \subseteq \{1, \ldots, n\}$ be a subdomain, denote by $\pi[D]$ the sub-vector of π containing elements on these locations. When $C \subseteq \mathbb{F}_q^n$, we let $C[D] \triangleq \{c[D] : c \in C\}$.

3.1 Linear codes

Definition 1 (Linear codes)**.** *A linear code with message length k and codeword length n if it is a transformation from $\mathbb{F}_q^k$ to a linear subspace $C \subseteq \mathbb{F}_q^n$. The generator matrix $\mathbf{G} \in \mathbb{F}_q^k \times \mathbb{F}_q^n$ of the linear code is a matrix whose rows form a basis for the code, i.e., for any codeword $c \in C$, there exists a message $\mathbf{m} \in \mathbb{F}_q^k$ such that $c = \mathbf{m} \cdot \mathbf{G}$.*

Definition 2 (Relative distance)**.** *Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords, where $n \in \mathbb{N}^+$. Define the relative Hamming distance between π_1 and π_2 to be*

$$\delta(\pi_1, \pi_2) \triangleq \frac{|\{j \in \{1, \ldots, n\} : \pi_1[j] \neq \pi_2[j]\}|}{n}.$$

Let $C \subseteq \mathbb{F}_q^n$ be a set of codewords with length n and $\pi \in \mathbb{F}_q^n$ be a codeword. Define the relative Hamming distance between π and C to be

$$\delta(\pi, C) \triangleq \min_{c \in C} \delta(\pi, c).$$

Definition 3 (Minimum relative distance)**.** *Let $C \subseteq \mathbb{F}_q^n$ be a set of codewords with length n, where $n \in \mathbb{N}^+$. The minimum relative distance of C is defined as*

$$\Delta_C \triangleq \min_{c_1, c_2 \in C} \delta(c_1, c_2).$$

For $c_1, c_2 \in C$, if $\delta(c_1, c_2) < \Delta_C$, then we have $c_1 = c_2$ according to the definition of minimum relative distance. In other words, for two codewords $c_1, c_2 \in C$, if we can find a set $D \subseteq \{1, \ldots, n\}$ such that (1) $|D| > (1 - \Delta_C)n$ and (2) $c_1[D] = c_2[D]$, then $c_1 = c_2$.

3.2 Correlated agreement

Let $C \subseteq \mathbb{F}_q^n$ be a linear code. Let $\pi_0, \ldots, \pi_t \in \mathbb{F}_q^n$ be a series of codewords. If $\pi_0, \ldots, \pi_t$ are not only close to C individually but also share a common large agreement domain, then we say $\pi_0, \ldots, \pi_t$ have a *correlated agreement*.

Definition 4 (Correlated agreement [Ben+20b])**.** *Let $\pi_0, \ldots, \pi_t \in \mathbb{F}_q^n$ 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 $D \subseteq \{1, \ldots, n\}$ and $c_0, \ldots, c_t \in C$ satisfying*

- ***Density:*** *$|D| \geq (1 - \delta)n$, and*
- ***Agreement:*** *for all $i \in \{0, \ldots, t\}$, the codewords π_i and c_i agree on D.*

Then we say $\pi_0, \ldots, \pi_t$ have correlated agreement *with C of density $\geq 1 - \delta$.*

Definition 5 (Correlated distance)**.** *Let $\pi_0, \ldots, \pi_t \in \mathbb{F}_q^n$ be a sequence of codewords. Let $C \subseteq \mathbb{F}_q^n$ be a set of codewords. Define the* correlated distance *of $\pi_0, \ldots, \pi_t$ with respect to C to be*

$$\delta_{\text{corr}}(\{\pi_0, \ldots, \pi_t\}, C) \triangleq 1 - \max\left\{ \frac{|D|}{n} : D \subseteq \{1, \ldots, n\}, \pi_0[D], \ldots, \pi_t[D] \in C[D] \right\}.$$

4 Main proof

Let $C \subseteq \mathbb{F}_q^n$ be a linear code with minimum relative distance Δ_C. Let δ be a parameter under the one-and-a-half Johnson bound $1 - \sqrt[3]{1 - \Delta_C}$ and let $\eta > 0$ be the gap between δ and the 1.5 Johnson bound, i.e., $\delta \leq 1 - \sqrt[3]{1 - \Delta_C} - \eta$. Suppose $\pi_1, \pi_2 \in \mathbb{F}_q^n$ are codewords do not have correlated agreement with C of density $\geq 1 - \delta$, i.e., for every $D \subseteq \{1, \ldots, n\}$ such that $|D| \geq (1 - \delta)n$, we have $\pi_1[D] \notin C[D]$ or $\pi_2[D] \notin C[D]$. Our goal is to prove that with high probability, the combining codeword $\pi_1 + \alpha\pi_2$ is still δ-far from C.

Definition 6 (Bad combining points)**.** *Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$, where $n > k \geq 1$. Denote by Δ_C the minimum relative distance of C. Let $\delta \in (0, 1 - \Delta_C)$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be codewords such that $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, C) > \delta$. Define the* bad combining points *of (π_1, π_2) (with respect to the linear code C) to be*

$$\text{Bad}(\pi_1, \pi_2) \triangleq \{\alpha \in \mathbb{F}_q : \delta(\pi_1 + \alpha\pi_2, C) \leq \delta)\}.$$

Our main theorem limits the number of bad combining points when δ is within the 1.5 Johnson bound, i.e., $\delta < \sqrt[3]{1 - \Delta_C}$.

Theorem 3. *Let $\delta, \eta > 0$ satisfy $\delta \le 1 - \sqrt[3]{1-\Delta_C} - \eta$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be codewords such that $\delta_{\mathrm{corr}}(\{\pi_1, \pi_2\}, C) > \delta$. Then,*

$$|\mathrm{Bad}(\pi_1, \pi_2)| < \frac{2(1-\sqrt[3]{1-\Delta_C})\Delta_C}{9(1-\Delta_C)\eta^3} n + \frac{2\Delta_C}{3(1-\Delta_C)^{\frac{1}{3}}\eta^2} = O_{\Delta_C,\eta}(n).$$

Corollary 1. *Let $\delta, \eta > 0$ satisfy $\delta \le 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta \le \frac{1}{3}(1-\Delta_C)^{-\frac{1}{3}}$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be codewords such that $\delta_{\mathrm{corr}}(\{\pi_1, \pi_2\}, C) > \delta$. Then,*

$$|\mathrm{Bad}(\pi_1, \pi_2)| < \frac{2\Delta_C n}{9(1-\Delta_C)\eta^3}.$$

Proof. Since $n \ge 1$, we have $2n \ge 2 \ge 6(1-\Delta_C)^{\frac{1}{3}}\eta$. As a result, we have

$$\begin{aligned}
|\mathrm{Bad}(\pi_1, \pi_2)| &< \frac{2(1-\sqrt[3]{1-\Delta_C})\Delta_C}{9(1-\Delta_C)\eta^3} n + \frac{2\Delta_C}{3(1-\Delta_C)^{\frac{1}{3}}\eta^2} && \text{by Theorem 3} \\
&= \frac{\Delta_C\left(2n - (1-\Delta_C)^{\frac{1}{3}}(2n - 6(1-\Delta_C)^{\frac{1}{3}}\eta)\right)}{9(1-\Delta_C)\eta^3} \\
&\le \frac{2\Delta_C n}{9(1-\Delta_C)\eta^3}.
\end{aligned}$$

□

To prove the main theorem, we first introduce a few definitions.

Definition 7. *Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with message length k, where $n > k \ge 1$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords. Let $\alpha \in \mathbb{F}_q$.*

- *For a codeword $\pi \in \mathbb{F}_q^n$, let $\mathrm{Closest}(\pi, C) \in C$ denote the closest codeword to π (in terms of Hamming distance). If there are more than one codewords with the same minimal distance, choose the codeword with the smallest lexicographical order.*
- *$B_\alpha^* \triangleq \{j \in \{1, \dots, n\} : (\pi_1 + \alpha\pi_2)[j] = \mathrm{Closest}(\pi_1 + \alpha\pi_2, C)[j]\}$. That is, B_α^* is the set of locations where $\pi_1 + \alpha\pi_2$ agrees with its closest codeword.*
- *If $\alpha \in \mathrm{Bad}(\pi_1, \pi_2)$, define B_α to be the set of first $(1-\delta)n$ elements of B_α^*, i.e., $|B_\alpha| = (1-\delta)n$.*

Poof overview: The overall strategy is to *partition* the bad points $\mathrm{Bad}(\pi_1, \pi_2) \subseteq \mathbb{F}_q$ into r blocks, denoted by $A_1, \dots, A_r$, with representatives $\alpha_1 \in A_1, \dots, \alpha_r \in A_r$.

First, by using a combinatorial lemma (i.e., Corrádi's lemma), we prove $r = O_{\Delta_d,\eta}(1)$; then, we prove the size of each set A_i is bounded by $O_{\Delta_d,\eta}(n)$. Together, we can conclude the number of bad points is at most $O_{\Delta_d,\eta}(n)$.

Within each block A_i with representative α_i, by defining an equivalence relation, we further partition $A_i \setminus \{\alpha_i\}$ into s_i classes, such that

1) the number of equivalence classes is $O_{\Delta_C,\eta}(1)$, and

2) each equivalence class is of size $O(n)$.

1) and 2) would imply that each block A_i is of size $O_{\Delta_C,\eta}(n)$.

We use a simple greedy algorithm to find the partition of $\mathrm{Bad}(\pi_1, \pi_2)$ and the representatives.

Algorithm 2 Partition bad combining points

input: $\mathrm{Bad}(\pi_1, \pi_2)$
initialization: $r = 0$
set $X^* = \mathrm{Bad}(\pi_1, \pi_2)$
while $X^* \neq \emptyset$ **do**
  $r = r + 1$
  pick an arbitrary $x \in X^*$ and let $\alpha_i = x$
  let $A_i = \{\beta \in \mathrm{Bad}(\pi_1, \pi_2) : |B_\beta \cap B_{\alpha_i}| \geq \left((1-\delta)^2 - \eta'\right) n\}$
  $X^* = X^* - A_i$
end while
return $A_1, \ldots, A_r$ and $\alpha_1, \ldots, \alpha_r$

4.1 Upper bound on the number of blocks

In this subsection, we upper bound the number of blocks. When $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$, we prove $r \leq \frac{\Delta_C}{\eta'}$, where $\eta' = \frac{3}{2} \cdot \sqrt[3]{1-\Delta_C}\eta^2$.

Without loss of generality, we can set $\eta < 1$ since $0 < \delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$. Then we have

$$\begin{aligned}(1-\delta)^2 - \eta' &\geq ((1-\Delta_C)^{\frac{1}{3}} + \eta)^2 - \frac{3(1-\Delta_C)^{\frac{1}{3}}\eta^2}{2} \\ &= (1-\Delta_C)^{\frac{2}{3}} + (2(1-\Delta_C)^{\frac{1}{3}}\eta - \frac{3(1-\Delta_C)^{\frac{1}{3}}\eta^2}{2}) + \eta^2 \\ &> (1-\Delta_C)^{\frac{2}{3}} \geq 1 - \Delta_C. \end{aligned} \tag{4}$$

Using Corrádi's lemma (Lemma 2), we can restrict the number of blocks.

Lemma 3. *Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be any two codewords. Let $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta' = \frac{3}{2} \cdot \sqrt[3]{1-\Delta_C}\eta^2$. Let r denote the number of blocks given by Algorithm 2. We have*

$$r < \frac{\Delta_C}{\eta'}.$$

Proof. We associate a set B_{α_i} for each representative α_i. From Algorithm 2, we know

- $|B_{\alpha_i} \cap B_{\alpha_j}| < \left((1-\delta)^2 - \eta'\right) n$ for any distinct i, j.
- $|B_{\alpha_i}| = (1-\delta)n$ for any i.

By Lemma 2, we have

$$\left|\bigcup_{i=1}^{r} B_{\alpha_i}\right| \geq \frac{r(1-\delta)^2 n^2}{(1-\delta)n + (r-1)\left((1-\delta)^2 - \eta'\right) n}.$$

On the other hand, $\bigcup_{i=1}^{r} B_{\alpha_i}$ is a subset of $\{1, \ldots, n\}$. Thus,

$$n \geq \frac{r(1-\delta)^2 n}{(1-\delta) + (r-1)\left((1-\delta)^2 - \eta'\right)},$$

that is

$$(1-\delta)+(r-1)\left((1-\delta)^2-\eta'\right)\geq r(1-\delta)^2,$$

which implies

$$\begin{aligned} r &\leq \frac{1-\delta-\left((1-\delta)^2-\eta'\right)}{(1-\delta)^2-\left((1-\delta)^2-\eta'\right)} \\ &< \frac{1-\delta-(1-\Delta_C)}{\eta'} && \text{by (4)} \\ &< \frac{\Delta_C}{\eta'}. \end{aligned}$$

□

4.2 Upper bound on the size of each block

In this subsection, we bound the number of elements in each block A_i. Combining with Lemma 3, we can bound the size of $\mathrm{Bad}(\pi_1,\pi_2)$ and finish the proof of our main theorem.

Lemma 4. *For any distinct $\alpha_1,\alpha_2\in\mathbb{F}_q$, there exist codewords $c_1,c_2\in C$ such that*

$$\pi_1[B^*_{\alpha_1}\cap B^*_{\alpha_2}]=c_1[B^*_{\alpha_1}\cap B^*_{\alpha_2}] \text{ and } \pi_2[B^*_{\alpha_1}\cap B^*_{\alpha_2}]=c_2[B^*_{\alpha_1}\cap B^*_{\alpha_2}]. \tag{5}$$

Moreover, if $|B^*_{\alpha_1}\cap B^*_{\alpha_2}|>(1-\Delta_C)n$, codewords c_1 and c_2 are uniquely determined.

Proof. By Definition 7, there exist messages $\mathbf{m}_1,\mathbf{m}_2\in\mathbb{F}_q^k$ such that

$$(\pi_1+\alpha_1\pi_2)[B^*_{\alpha_1}]=(\mathbf{m}_1\cdot\mathbf{G})[B^*_{\alpha_1}] \text{ and } (\pi_1+\alpha_2\pi_2)[B^*_{\alpha_2}]=(\mathbf{m}_2\cdot\mathbf{G})[B^*_{\alpha_2}].$$

For any $j\in B^*_{\alpha_1}\cap B^*_{\alpha_2}$, solving two linear equations in $\pi_1[j]$ and $\pi_2[j]$, we have

$$\begin{cases} \pi_1[j]=\left(\dfrac{\alpha_1\mathbf{m}_2-\alpha_2\mathbf{m}_1}{\alpha_1-\alpha_2}\cdot\mathbf{G}\right)[j] \\ \pi_2[j]=\left(\dfrac{\mathbf{m}_1-\mathbf{m}_2}{\alpha_1-\alpha_2}\cdot\mathbf{G}\right)[j] \end{cases}.$$

When $|B^*_{\alpha_1}\cap B^*_{\alpha_2}|>(1-\Delta_C)n$, codewords $c_1=\frac{\alpha_1\mathbf{m}_2-\alpha_2\mathbf{m}_1}{\alpha_1-\alpha_2}\cdot\mathbf{G}$ and $c_2=\frac{\mathbf{m}_1-\mathbf{m}_2}{\alpha_1-\alpha_2}\cdot\mathbf{G}$ are uniquely determined, since C has minimum distance Δ_C. □

Proposition 1. *Let $\alpha,\beta\in\mathbb{F}_q$ be different points such that $|B^*_\alpha\cap B^*_\beta|>(1-\Delta_C)n$. Let $c_1,c_2\in C$ be the codewords uniquely determined such that $c_1[B^*_\alpha\cap B^*_\beta]=\pi_1[B^*_\alpha\cap B^*_\beta]$ and $c_2[B^*_\alpha\cap B^*_\beta]=\pi_2[B^*_\alpha\cap B^*_\beta]$ (by Lemma 4). Then*

$$\mathrm{Closest}(\pi_1+\beta\pi_2,C)=c_1+\beta c_2. \tag{6}$$

Proof. Since $c_1,c_2\in C$, we have $c_1+\beta c_2\in C$, because C is a linear code. Since both sides of (6) are codewords in C, whose minimum distance is Δ_C, it suffices to demonstrate $(1-\Delta_C)n+1$ points on which both sides of (6) are equal.

Let $j \in B^*_\alpha \cap B^*_\beta$. We have

$$\begin{aligned}(c_1 + \beta c_2)[j] &= c_1[j] + \beta c_2[j] \\ &= \pi_1[j] + \beta \pi_2[j] && \text{by our assumption} \\ &= (\pi_1 + \beta\pi_2)[j] \\ &= \text{Closest}(\pi_1 + \beta\pi_2, C)[j] && \text{by the definition of } B^*_\beta.\end{aligned}$$

Note that $|B^*_\alpha \cap B^*_\beta| > (1 - \Delta_C)n$. Thus, we have exhibited at least $(1 - \Delta_C)n + 1$ points where the evaluations of both sides of (6) are equal. □

Corollary 2. *Let $\alpha, \beta \in \mathbb{F}_q$ be distinct and $|B^*_\alpha \cap B^*_\beta| > (1 - \Delta_C)n$. Let $c_1, c_2 \in C$ be the uniquely determined codewords such that $\pi_1[B^*_\alpha \cap B^*_\beta] = c_1[B^*_\alpha \cap B^*_\beta]$ and $\pi_2[B^*_\alpha \cap B^*_\beta] = c_2[B^*_\alpha \cap B^*_\beta]$ (by Lemma 4). For any $j \in \{1, \ldots, n\}$, we have $j \in B^*_\alpha \cap B^*_\beta$ if and only if $\pi_1[j] = c_1[j]$ and $\pi_2[j] = c_2[j]$.*

Proof. The "only if" direction is obvious, which follows from the definitions of c_1 and c_2.

For the "if" direction, assuming $\pi_1[j] = c_1[j]$ and $\pi_2[j] = c_2[j]$, our goal is to prove $j \in B^*_\alpha \cap B^*_\beta$. We prove $j \in B^*_\alpha$, for $j \in B^*_\beta$ is similar to prove.

By Proposition 1, we have

$$c_1 + \alpha c_2 = \text{Closest}(\pi_1 + \alpha\pi_2, C). \tag{7}$$

Since $\pi_1[j] = c_1[j]$ and $\pi_2[j] = c_2[j]$, we have

$$(\pi_1 + \alpha\pi_2)[j] = (c_1 + \alpha c_2)[j]. \tag{8}$$

Combining (7) with (8), we have $\text{Closest}(\pi_1 + \alpha\pi_2, C)[j] = (\pi_1 + \alpha\pi_2)[j]$, which implies $j \in B^*_\alpha$. □

The following lemma proves a crucial combinatorial property for $B^*_\alpha, B^*_{\beta_1}, B^*_{\beta_2}$, where β_1, β_2 belongs to the same partition with the representative α. Whenever $B^*_\alpha \cap B^*_{\beta_1} \neq B^*_\alpha \cap B^*_{\beta_2}$, the size of the intersection of $B^*_\alpha, B^*_{\beta_1}, B^*_{\beta_2}$ is "small".

Lemma 5. *Let $\alpha, \beta_1, \beta_2 \in \mathbb{F}_q$ be different such that $|B^*_\alpha \cap B^*_{\beta_1}| > (1 - \Delta_C)n$ and $|B^*_\alpha \cap B^*_{\beta_2}| > (1 - \Delta_C)n$. Denote by $c_1, c_2 \in C$ the codewords uniquely determined by $\pi_1[B^*_\alpha \cap B^*_{\beta_1}], \pi_2[B^*_\alpha \cap B^*_{\beta_1}]$; denote by $c'_1, c'_2 \in C$ the codewords uniquely determined by $\pi_1[B^*_\alpha \cap B^*_{\beta_2}], \pi_2[B^*_\alpha \cap B^*_{\beta_2}]$ (by Lemma 4). Then exactly one of the followings holds:*

- *$c_1 = c'_1, c_2 = c'_2$ and $B^*_\alpha \cap B^*_{\beta_1} = B^*_\alpha \cap B^*_{\beta_2}$.*
- *There exists $i \in \{1, 2\}$ such that $c_i \neq c'_i$, and $|B^*_\alpha \cap B^*_{\beta_1} \cap B^*_{\beta_2}| \leq (1 - \Delta_C)n$.*

Proof. **Case 1**: $c_1 = c'_1$ and $c_2 = c'_2$. We want to prove $B^*_\alpha \cap B^*_{\beta_1} = B^*_\alpha \cap B^*_{\beta_2}$. Let us prove $B^*_\alpha \cap B^*_{\beta_1} \subseteq B^*_\alpha \cap B^*_{\beta_2}$ first. Let $j \in B^*_\alpha \cap B^*_{\beta_1}$. By Corollary 2, we have $\pi_1[j] = c_1[j]$ and $\pi_2[j] = c_2[j]$. Since $c_1 = c'_1, c_2 = c'_2$, we have $\pi_1[j] = c'_1[j]$ and $\pi_2[j] = c'_2[j]$. By Corollary 2, $j \in B^*_\alpha \cap B^*_{\beta_2}$. Thus, we have shown that $B^*_\alpha \cap B^*_{\beta_1} \subseteq B^*_\alpha \cap B^*_{\beta_2}$. The other direction, that is, $B^*_\alpha \cap B^*_{\beta_2} \subseteq B^*_\alpha \cap B^*_{\beta_1}$, is similar to prove.

Case 2: $c_1 \neq c'_1$ or $c_2 \neq c'_2$. Without loss of generality, suppose $c_1 \neq c'_1$, we want to prove $|B^*_\alpha \cap B^*_{\beta_1} \cap B^*_{\beta_2}| \leq (1 - \Delta_C)n$. Note that, for any $j \in B^*_\alpha \cap B^*_{\beta_1} \cap B^*_{\beta_2}$, by Corollary 2, $\pi_1[j] = c_1[j]$ and $\pi_1[j] = c'_1[j]$, which implies $c_1[j] = c'_1[j]$. Since $c_1, c'_1 \in C$ are different codewords and the minimum distance of C is Δ_C, we have $|B^*_\alpha \cap B^*_{\beta_1} \cap B^*_{\beta_2}| \leq (1 - \Delta_C)n$. □

To bound the size of each block, we further partition each block into equivalences classes by defining the following equivalence relation.

Definition 8 (Equivalence relation within a block)**.** *Let $A \subseteq \mathrm{Bad}(\pi_1, \pi_2)$ be a partition with representative α produced by Algorithm 2. Let $\beta_1, \beta_2 \in A \setminus \{\alpha\}$. We define an equivalence relation $\mathcal{R}_{A,\alpha}$ on $A \setminus \{\alpha\}$ as follows:*

$$(\beta_1, \beta_2) \in \mathcal{R}_{A,\alpha} \iff c_1 = c_1' \text{ and } c_2 = c_2',$$

where $c_1, c_2 \in C$ are the codewords uniquely determined by $\pi_1[B^*_\alpha \cap B^*_{\beta_1}], \pi_2[B^*_\alpha \cap B^*_{\beta_1}]$ and and $c_1', c_2' \in C$ are the codewords determined by $\pi_1[B^*_\alpha \cap B^*_{\beta_2}], \pi_2[B^*_\alpha \cap B^*_{\beta_2}]$. (Notice that for any $\beta \in A$, we have $|B^*_\beta \cap B^*_\alpha| \geq |B_\beta \cap B_\alpha| \geq \left((1-\delta)^2 - \eta'\right) n > 1 - \Delta_C$, where the last step is by (4).)

Let s be the number of equivalence classes. For convenience, let

$$[\beta_i] = \{\beta_{i,1}, \ldots, \beta_{i,t_i}\},$$

denote the equivalence classes, where $i \in [s]$. We have already shown

- (Lemma 5) Different equivalence classes correspond to different codeword pairs, denoted by $(c_1^{(1)}, c_2^{(1)}), \ldots, (c_1^{(s)}, c_2^{(s)})$.
- (Proposition 1) For any $i \in \{1, \ldots, s\}$, and for any $j \in \{1, \ldots, t_i\}$,

$$\mathrm{Closest}(\pi_1 + \beta_{i,j}\pi_2, C) = c_1^{(i)} + \beta_{i,j} c_2^{(i)}.$$

On the one hand, we can bound the number of equivalence classes by using the following generalization of Corrádi's lemma, whose proof is almost the same as Corrádi's.

Corollary 3 (A variant of the Corrádi lemma)**.** *Let $P_1, \ldots, P_r$ be r sets satisfying*

$$s_1 \leq |P_i| < s_2, 1 \leq i \leq r.$$

If $|P_i \cap P_j| \leq k$ for any distinct $i, j \in \{1, 2, \ldots, r\}$, then

$$\left|\bigcup_{i=1}^{r} P_i\right| > \frac{s_1^2 r}{s_2 + (r-1)k}.$$

Proof. We closely follow the proof of Corrádi's Lemma. For any $x \in \bigcup_{i=1}^r P_i$, denote by $d(x)$ the count of x, i.e., the number of P_i containing x, we have

$$\begin{aligned}\sum_{x \in P_i} d(x) &= \sum_{j=1}^{r} |P_i \cap P_j| = |P_i| + \sum_{j \neq i} |P_i \cap P_j| \\ &< s_2 + (r-1)k. \end{aligned} \tag{9}$$

Summing over all sets P_i, we have

$$\begin{aligned}\sum_{i=1}^{r}\sum_{x\in P_i} d(x) &= \sum_{x\in\bigcup_{i=1}^{r} P_i} d(x)^2 \\ &\geq \frac{1}{|\bigcup_{i=1}^{r} P_i|}\left(\sum_{x\in\bigcup_{i=1}^{r} P_i} d(x)\right)^2 && \text{by Cauchy–Schwarz inequality} \\ &= \frac{1}{|\bigcup_{i=1}^{r} P_i|}\left(\sum_{i=1}^{r}|P_i|\right)^2 \geq \frac{s_1^2 r^2}{|\bigcup_{i=1}^{r} P_i|}. && (10)\end{aligned}$$

Combining (9) and (10), we have

$$r\left(s_2+(r-1)k\right) > \frac{s_1^2 r^2}{|\bigcup_{i=1}^{r} P_i|} \Rightarrow \left|\bigcup_{i=1}^{r} P_i\right| > \frac{s_1^2 r}{s_2+(r-1)k}.$$

□

Corollary 4. *Let $\delta \leq 1-\sqrt[3]{1-\Delta_C}-\eta$ and $\eta' = \frac{3\sqrt[3]{1-\Delta_C}\eta^2}{2}$. Let $A \in \{A_1,\ldots,A_r\}$ be a set of combining points defined in Algorithm 2 and $\alpha \in A$ be the corresponding representative. Let s be the number of equivalence classes in $\mathcal{R}_{A,\alpha}$ (as defined in Definition 8). We have*

$$s < \frac{1}{3\eta(1-\Delta_C)^{\frac{2}{3}}}.$$

Proof. Let

$$[\beta_i] = \{\beta_{i,1},\ldots,\beta_{i,t_i}\},$$

be the equivalence classes, where $i \in [s]$. Notice that

- For any $i \in [s]$, $|B_\alpha \cap B_{\beta_{i,1}}| < (1-\delta)n$. By Lemma 2, π_1, π_2 agree with a pair of codewords simultaneously on $B^*_\alpha \cap B^*_{\beta_{i,1}}$. By our condition in Theorem 3, $\delta_{\text{corr}}(\{\pi_1,\pi_2\}, C) > \delta n$. Thus, $|B^*_\alpha \cap B^*_{\beta_{i,1}}| < (1-\delta)n$. Also notice that $B_\alpha \cap B_{\beta_{i,1}} \subseteq B^*_\alpha \cap B^*_{\beta_{i,1}}$.
- For any $i \in [s]$, $|B_\alpha \cap B_{\beta_{i,1}}| \geq \left((1-\delta)^2-\eta'\right)n$ according to Algorithm 2.
- For any distinct $i,j \in [s]$, $|B_\alpha \cap B_{\beta_{i,1}} \cap B_{\beta_{j,1}}| \leq |B^*_\alpha \cap B^*_{\beta_{i,1}} \cap B^*_{\beta_{j,1}}| \leq (1-\Delta_C)n$ by Lemma 5.
- $|\bigcup_{i=1}^{s}(B_\alpha \cap B_{\beta_i,1})| \leq |B_\alpha| = (1-\delta)n$.

Applying a variant of the Corrádi lemma, i.e., Corollary 3, we have

$$(1-\delta)n > \frac{\left(\left((1-\delta)^2-\eta'\right)n\right)^2 s}{(1-\delta)n+(s-1)(1-\Delta_C)n}.$$

Thus,

$$\begin{aligned} s &< \frac{(1-\delta-(1-\Delta_C))(1-\delta)}{((1-\delta)^2-\eta')^2-(1-\Delta_C)(1-\delta)} \\ &= \frac{(\Delta_C-\delta)(1-\delta)}{((1-\delta)^3-(1-\Delta_C))(1-\delta)-2\eta'(1-\delta)^2+\eta'^2} \\ &\leq \frac{(\Delta_C-\delta)(1-\delta)}{\left(3\eta^2(1-\Delta_C)^{\frac{1}{3}}+3\eta(1-\Delta_C)^{\frac{2}{3}}+\eta^3\right)(1-\delta)-2\eta'(1-\delta)^2+\eta'^2}. \end{aligned}$$

Notice that $\eta' = \frac{3(1-\Delta_C)^{\frac{1}{3}}\eta^2}{2}$, then

$$\begin{aligned} &3\eta^2(1-\Delta_C)^{\frac{1}{3}}(1-\delta)-2\eta'(1-\delta)^2 \\ =&3\eta^2(1-\Delta_C)^{\frac{1}{3}}(1-\delta)-3(1-\Delta_C)^{\frac{1}{3}}\eta^2(1-\delta)^2 \geq 0. \end{aligned}$$

Thus, we have

$$s < \frac{(\Delta_C-\delta)(1-\delta)}{3\eta(1-\Delta_C)^{\frac{2}{3}}(1-\delta)} = \frac{\Delta_C-\delta}{3\eta(1-\Delta_C)^{\frac{2}{3}}} \leq \frac{1}{3\eta(1-\Delta_C)^{\frac{2}{3}}}.$$

□

On the other hand, we can bound the number of elements in each equivalence class based on the following lemma.

Lemma 6. *Let $\alpha, \beta_1, \ldots, \beta_t \in \mathrm{Bad}(\pi_1, \pi_2)$ be distinct such that*

- $B^*_\alpha \cap B^*_{\beta_1} = B^*_\alpha \cap B^*_{\beta_2} = \ldots = B^*_\alpha \cap B^*_{\beta_t}$, *and*
- $|B^*_\alpha \cap B^*_{\beta_1}| > (1-\Delta_C)n$.

Then $B^*_{\beta_1}, \ldots, B^*_{\beta_t}$ form a sunflower with core $B^*_\alpha \cap B^*_{\beta_1}$. That is, for any distinct $i, j \in [t]$, we have

$$B^*_{\beta_i} \cap B^*_{\beta_j} = B^*_\alpha \cap B^*_{\beta_1}.$$

Proof. Let $i, j \in [t]$ be any two distinct number. Let $c_1, c_2 \in C$ denote the unique pair of codewords determined by $\pi_1[B^*_\alpha \cap B^*_{\beta_i}], \pi_2[B^*_\alpha \cap B^*_{\beta_i}]$ (by Lemma 4).

Observe that $B^*_\alpha \cap B^*_{\beta_i} = B^*_\alpha \cap B^*_{\beta_i} \cap B^*_{\beta_j} \subseteq B^*_{\beta_i} \cap B^*_{\beta_j}$. By Lemma 4, there are a pair of codewords $c'_1, c'_2 \in C$ uniquely determined by $\pi_1[B^*_{\beta_i} \cap B^*_{\beta_j}], \pi_2[B^*_{\beta_i} \cap B^*_{\beta_j}]$. Since $\pi_1[B^*_\alpha \cap B^*_{\beta_i}] = c_1[B^*_\alpha \cap B^*_{\beta_i}]$ and $\pi_1[B^*_{\beta_i} \cap B^*_{\beta_j}] = c'_1[B^*_{\beta_i} \cap B^*_{\beta_j}]$, where $|B^*_\alpha \cap B^*_{\beta_i}| > (1-\Delta_C)n$ and $|B^*_{\beta_i} \cap B^*_{\beta_j}| > (1-\Delta_C)n$, we have $c_1 = c'_1$. Similarly, we can show $c_2 = c'_2$.

For any $k \in B^*_{\beta_i} \cap B^*_{\beta_j}$, we have $\pi_1[k] = c'_1[k]$ and $\pi_2[k] = c'_2[k]$ by Corollary 2. Since $\pi_1[k] = c_1[k]$ and $\pi_2[k] = c_2[k]$, using Corollary 2 again, we have $k \in B^*_\alpha \cap B^*_{\beta_i} = B^*_\alpha \cap B^*_{\beta_i} \cap B^*_{\beta_j}$. Thus,

$$B^*_{\beta_i} \cap B^*_{\beta_j} = B^*_\alpha \cap B^*_{\beta_i} \cap B^*_{\beta_j} = B^*_\alpha \cap B^*_{\beta_1}.$$

Since i, j are arbitrary, we have completed the proof. □

Corollary 5. *The number of elements in each equivalence class in $\mathcal{R}_{A,\alpha}$ (as defined in Definition 8) is at most δn.*

Proof. Let

$$[\beta_i] = \{\beta_{i,1}, \ldots, \beta_{i,t_i}\}$$

be an equivalence class. By our definition of the equivalence relation $\mathcal{R}_{A,\alpha}$, we have $B^*_\alpha \cap B^*_{\beta_{i,1}} = \ldots = B^*_\alpha \cap B^*_{\beta_{i,t_i}}$. According to Algorithm 2, we have $|B^*_\alpha \cap B^*_{\beta_{i,1}}| \geq |B_\alpha \cap B_{\beta_{i,1}}| \geq \left((1-\delta)^2 - \eta'\right) n > (1-\Delta_C)n$. Lemma 6 tells us

$$B^*_{\beta_{i,j}} \cap B^*_{\beta_{i,k}} = B^*_\alpha \cap B^*_{\beta_{i,1}}, \quad 1 \leq j < k \leq t_i,$$

which implies

$$\left(B^*_{\beta_{i,j}} \setminus B^*_\alpha\right) \cap \left(B^*_{\beta_{i,k}} \setminus B^*_\alpha\right) = \emptyset, \quad 1 \leq j < k \leq t_i. \tag{11}$$

Thus,

$$\delta n \geq |\{1, \ldots, n\} \setminus B^*_\alpha| \geq \left|\bigcup_{j=1}^{t_i} \left(B^*_{\beta_{i,j}} \setminus B^*_\alpha\right)\right| = \sum_{j=1}^{t_i} \left|B^*_{\beta_{i,j}} \setminus B^*_\alpha\right| \tag{12}$$

For each $\beta_{i,j} \in [\beta_i]$, we have $|B^*_{\beta_{i,j}}| \geq (1-\delta)n$ according to the definition of bad combining points (Definition 6). On the other hand, Lemma 4 tells us there exist $c_1, c_2 \in C$ such that $\pi_1[B^*_\alpha \cap B^*_{\beta_{i,j}}] = c_1[B^*_\alpha \cap B^*_{\beta_{i,j}}]$ and $\pi_2[B^*_\alpha \cap B^*_{\beta_{i,j}}] = c_2[B^*_\alpha \cap B^*_{\beta_{i,j}}]$. Since $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, C) > \delta$, we must have $|B^*_\alpha \cap B^*_{\beta_{i,j}}| < (1-\delta)n$. As a result, we have

$$\left|B^*_{\beta_{i,j}} \setminus B^*_\alpha\right| \geq 1. \tag{13}$$

Plugging (13) into (12), we have $t_i \leq \delta n$. □

4.3 Proof of Theorem 3

Algorithm 2 gives a partition of $\text{Bad}(\pi_1, \pi_2)$, denoted by $A_1, \ldots, A_r \subseteq \text{Bad}(\pi_1, \pi_2)$. By Lemma 3, the number of blocks is at most $r < \frac{\Delta_C}{\eta'}$. The size of each block A_i is less than $st + 1$, which is bounded by Corollary 4 and Corollary 5. So

$$\begin{aligned}
|\text{Bad}(\pi_1, \pi_2)| &= \sum_{i=1}^{r} |A_i| \leq r \cdot (st+1) \\
&< \frac{\Delta_C}{\eta'} \cdot \left(\frac{1}{3\eta(1-\Delta_C)^{\frac{2}{3}}} \cdot \delta n + 1\right) && \text{Corollary 4 and Corollary 5} \\
&\leq \frac{\Delta_C}{\eta'} \cdot \left(\frac{1}{3\eta(1-\Delta_C)^{\frac{2}{3}}} \cdot \left(1 - (1-\Delta_C)^{\frac{1}{3}} - \eta\right) n + 1\right) \\
&\leq \frac{\Delta_C \left(1 - (1-\Delta_C)^{\frac{1}{3}}\right)}{\frac{3}{2}(1-\Delta_C)^{\frac{1}{3}}\eta^2 \cdot 3\eta(1-\Delta_C)^{\frac{2}{3}}} n + \frac{\Delta_C}{\frac{3}{2}(1-\Delta_C)^{\frac{1}{3}}\eta^2} \\
&= \frac{2\Delta_C \left(1 - (1-\Delta_C)^{\frac{1}{3}}\right)}{9(1-\Delta_C)\eta^3} n + \frac{2\Delta_C}{3(1-\Delta_C)^{\frac{1}{3}}\eta^2}
\end{aligned}$$

The following theorem is the contraposition of Theorem 3.

Theorem 4 (Linear proximity gap for linear codes over lines)**.** *Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with minimum relative distance $\Delta_C \in (0,1)$. Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1-\Delta_C)^{-\frac{1}{3}}$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords such that*

$$\mathbb{P}_{\alpha \in \mathbb{F}_q}(\Delta(\pi_1 + \alpha\pi_2, C) \leq \delta) \geq \frac{2\Delta_C n}{9(1-\Delta_C)\eta^3|\mathbb{F}_q|}.$$

Then π_1, π_2 are simultaneously δ-close to C, i.e., $\exists c_1, c_2 \in C$ such that

$$|\{j \in \{1,\ldots,n\} : \pi_1[j] = c_1[j] \text{ and } \pi_2[j] = c_2[j]\}| \geq (1-\delta)n.$$

5 Mutual correlated agreement for linear codes

This section provides stronger versions of Theorem 3. We first consider the list decoding of a given codeword and bound the number of combining points that *enlarge* the agree domains. This is called mutual correlated agreement in [Arn+24b] and strong correlated agreement in [Zei24], which is stronger than correlated agreement. After that, we further generalized the results to a batch of codewords, which can be used in many real-world applications. This batched version also supports the mutual correlated agreement.

5.1 Mutual correlated agreement for linear codes over lines

Instead of focusing on combining points that disclose the distance between the given codewords π_1, π_2 and a linear code C, we consider the combining points that *enlarge* the agree domains. More precisely, we consider the following domains:

Definition 9 (maximal δ-correlated-agree domains)**.** *Let $0 \leq \delta \leq 1$. Let $C \subseteq \mathbb{F}_q^n$ be a linear code. Let $\pi_1, \ldots, \pi_l \in \mathbb{F}_q^n$ be a series of codewords with $l \geq 1$. Let $D \subseteq \{1,\ldots,n\}$ be a domain satisfying:*

- ***Density:*** *$|D| \geq (1-\delta)n$;*
- ***Correlated agreement:*** *For $i \in \{1,\ldots,l\}$, $\exists c_i \in C$, such that $\pi_i[D] = c_i[D]$;*
- ***Maximal:*** *If $D \subsetneq D'$, then $\exists i \in \{1,\ldots,l\}$, for all $c \in C$, $\pi_i[D'] \neq c[D']$.*

We define such domain as a maximal δ-correlated-agree domain between $\{\pi_1,\ldots,\pi_l\}$ and C. Denote the set of all of the maximal δ-correlated-agree domains between $\{\pi_1,\ldots,\pi_l\}$ and C as

$$\mathcal{A}_{\delta,\{\pi_1,\ldots,\pi_l\},C} \triangleq \{D_1,\ldots,D_m\}.$$

Notice that $\mathcal{A}_{\delta,\{\pi_1,\ldots,\pi_l\},C}$ is unique and when $\delta_{\text{corr}}(\{\pi_1,\ldots,\pi_l\},C) > \delta$, $\mathcal{A}_{\delta,\{\pi_1,\ldots,\pi_l\},C}$ is empty.

Definition 10. *Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords. Function* Agree(π_1,π_2) *returns the locations where π_1 agrees with π_2, i.e.,*

$$\text{Agree}(\pi_1,\pi_2) = \{j \in [1,n] \mid \pi_1[j] = \pi_2[j]\} \subseteq [1,n].$$

Based on the above definitions, we can define the set of bad combining points that *enlarge* the list-decoding agree domains.

Definition 11. *Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$, where $n > k > 0$. Denote by Δ_C the minimum relative distance of C. Let $\delta \in (0, 1-\Delta_C)$. Let $\pi_1, \ldots, \pi_l \in \mathbb{F}_q^n$ be codewords with maximal δ-correlated-agree domains $\mathcal{A}_{\delta,\{\pi_1,\ldots,\pi_l\},C} = \{D_1, \ldots, D_m\}$, where $l \geq 2$. Define the* list-bad combining points *of $(\pi_1, \ldots, \pi_l)$ (with respect to the linear code C) to be*

$$\mathrm{Bad_L}(\pi_1, \ldots, \pi_l) \triangleq \{(\alpha_2, \ldots, \alpha_l) \in \mathbb{F}_q^{l-1} \mid \exists c \in C, \text{ such that } \delta(\pi_1 + \alpha_2\pi_2 + \ldots + \alpha_l\pi_l, c) \leq \delta \text{ and } \mathrm{Agree}(\pi_1 + \alpha_2\pi_2 + \ldots + \alpha_l\pi_l, c) \neq D_i, i = 1, \ldots, m\}.$$

Theorem 5. *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1-\Delta_C)^{-\frac{1}{3}}$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords with maximal δ-correlated-agree domains $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C} = \{D_1, \ldots, D_m\}$. Then we have*

$$|\mathrm{Bad_L}(\pi_1, \pi_2)| < \frac{2\Delta_C n}{9\rho\eta^3}.$$

Proof. We first modify the definition of some sets corresponding to Definition 7. For any $\alpha \in \mathrm{Bad_L}(\pi_1, \pi_2)$, $\exists c \in C$ such that $\delta(\pi_1 + \alpha\pi_2, c) \leq \delta$ and $\mathrm{Agree}(\pi_1 + \alpha\pi_2, c) \neq D_i, i = 1, \ldots, m$. If more than one codewords satisfy these conditions, choose the one with the smallest lexicographical order. Define

$$B^*_{L,\alpha} \triangleq \mathrm{Agree}(\pi_1 + \alpha\pi_2, c) \subseteq [1, n], \tag{14}$$

and $B_{L,\alpha} \subseteq B^*_{L,\alpha}$ is the of first $(1-\delta)n$ elements of $B^*_{L,\alpha}$. For each $B^*_{L,\alpha}$, we prove that there are two possible cases:

1. $|B^*_{L,\alpha} \cap D_i| \leq (1-\Delta_C)n, i \in \{1, \ldots, m\}$, or
2. $\exists D_i \in \mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}$, such that $B^*_{L,\alpha} \supsetneq D_i$.

This is because if $\exists D_i, |B^*_{L,\alpha} \cap D_i| > (1-\Delta_C)n$, denote by $c_1, c_2 \in C$ the codewords satisfying $c_1[D_i] = \pi_1[D_i], c_2[D_i] = \pi_2[D_i]$ according to the definition of $D_i \in \mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}$. Thus,

$$(c_1 + \alpha c_2)[D_i] = (\pi_1 + \alpha\pi_2)[D_i].$$

Notice that

$$(c_1 + \alpha c_2)[B^*_{L,\alpha} \cap D_i] = c[B^*_{L,\alpha} \cap D_i]$$

because of (14). Since $|B^*_{L,\alpha} \cap D_i| > (1-\Delta_C)n$, we have $c_1 + \alpha c_2 = c$. Then

$$(\pi_1 + \alpha\pi_2)[D_i] = (c_1 + \alpha c_2)[D_i] = c[D_i].$$

As a result, we have $B^*_{L,\alpha} = \mathrm{Agree}(\pi_1 + \alpha\pi_2, c) \supseteq D_i$. Definition 11 tells us $\mathrm{Agree}(\pi_1 + \alpha\pi_2, c) \neq D_i$, thus,

$$B^*_{L,\alpha} \supsetneq D_i.$$

Run Algorithm 2 on $\mathrm{Bad_L}(\pi_1, \pi_2)$ and the corresponding agree sets $B_{L,\alpha}, \alpha \in \mathrm{Bad_L}(\pi_1, \pi_2)$ to give a partition on $\mathrm{Bad_L}(\pi_1, \pi_2)$. Denote the output as $A'_1, \ldots, A'_r$ and $\alpha_1, \ldots, \alpha_r$. Lemma 3 still holds because our partition strategy is unchanged and we have $|B_{L,\alpha}| = (1-\delta)n$ for any $\alpha \in \mathrm{Bad_L}$. Thus, we have $r < \frac{\Delta_C}{\eta'}$. If we can bound the size of each block A_i, we finish our proof of the theorem.

For any $\alpha \in \{\alpha_1, \ldots, \alpha_r\}$, denote by A the block α is in.

For the first case that $|B^*_{L,\alpha} \cap D_i| \leq (1-\Delta_C)n, i = 1, \ldots, m$, we claim that $A = \{\alpha\}$. Otherwise, if $\exists \beta \in A$ and $\beta \neq \alpha$, we have π_1, π_2 agree with a pair of codewords $c_1, c_2 \in C$ on $B^*_{L,\alpha} \cap B^*_{L,\beta}$ by Lemma 4. Since $|B^*_{L,\alpha} \cap B^*_{L,\beta}| \geq |B_{L,\alpha} \cap B_{L,\beta}| > (1-\Delta_C)n$ according to the partition in Algorithm 2 and $|B^*_{L,\alpha} \cap D_i| \leq (1-\Delta_C)n, i = 1, \ldots, m$ according to our assumption, $B^*_{L,\alpha} \cap B^*_{L,\beta} \nsubseteq D_i, i = 1, \ldots, m$. A contradiction to Definition 9. So $|A| = 1$.

For the second case, if $|A| = 1$, we finish our proof of the theorem. Otherwise, we prove $|A| < \frac{\delta n}{3\eta(1-\Delta_C)^{\frac{2}{3}}} + 1$ in this case. Since Lemma 5 holds in this case, we still have the equivalent relation on $A \setminus \{\alpha\}$. Notice that the conditions of Corollary 4 hold in this case, we can bound the number of equivalence classes, i.e., $s < \frac{1}{3\eta(1-\Delta_C)^{\frac{2}{3}}}$. We slightly modified the proof of Corollary 5 to limit the number of elements in each equivalence class. Let

$$[\beta] = \{\beta_1, \ldots, \beta_t\}$$

be an equivalence class. Based on Lemma 6, we have

$$\delta n \geq |\{1, \ldots, n\} \setminus B^*_{L,\alpha}| \geq \left| \bigcup_{i=1}^{t} \left(B^*_{L,\beta_i} \setminus B^*_{L,\alpha} \right) \right| = \sum_{i=1}^{t} \left| B^*_{L,\beta_i} \setminus B^*_{L,\alpha} \right|. \tag{15}$$

According to Lemma 4, π_1, π_2 agree with a pair of codewords $c_1, c_2 \in C$ on $B^*_{L,\alpha} \cap B^*_{L,\beta_i}$. $\exists D \in \mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}$, such that $B^*_{L,\alpha} \cap B^*_{L,\beta_i} \subseteq D$. According to the definitions of $\mathrm{Bad_L}(\pi_1, \pi_2)$ and B^*_{L,β_i}, we have $B^*_{L,\beta_i} \nsubseteq D$. As a result,

$$|B^*_{L,\beta_i} \setminus B^*_{L,\alpha}| = |B^*_{L,\beta_i} \setminus (B^*_{L,\alpha} \cap B^*_{L,\beta_i})| \geq 1. \tag{16}$$

Plugging (16) into (15), we have

$$\delta n \geq t.$$

So we have $|A| \leq st + 1 < \frac{\delta n}{3\eta(1-\Delta_C)^{\frac{2}{3}}} + 1$ in this case. Combining with Corollary 1, we have the simplified expression in the theorem. □

5.2 Mutual correlated agreement over affine spaces

When there are a batch of codewords $\pi_0, \pi_1, \ldots, \pi_l$ with $l \geq 1$. We prove our results still hold.

Theorem 6. *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1-\Delta_C)^{-\frac{1}{3}}$. Let $\pi_0, \pi_1, \ldots, \pi_l \in \mathbb{F}_q^n$ be $l+1$ codewords with maximal δ-correlated-agree domains $\mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C} = \{D_1, \ldots, D_m\}$. Then we have*

$$|\mathrm{Bad_L}(\pi_0, \ldots, \pi_l)| < \frac{2\Delta_C n}{9\rho\eta^3} \cdot l.$$

Proof. We use induction to prove the result. When $l = 1$, the problem is reduced to Theorem 5. Suppose the result holds for $l-1$ codewords. For convenience, denote $\mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_i\},C}$ as $\mathcal{A}_i$ and denote $\langle \alpha_1, \ldots, \alpha_i \rangle \in \mathbb{F}_q^i$ as $\mathbf{z}_i$, $i = 1, \ldots, l$. For $1 \leq i \leq l$, define

$$S_i \triangleq \{\langle \alpha_1, \ldots, \alpha_i \rangle \in \mathbb{F}_q^i \mid \exists c \in C, \text{ such that } \delta(\pi_0 + \cdots + \alpha_i \pi_i, c) \leq \delta \\ \text{and } \mathrm{Agree}(\pi_0 + \cdots + \alpha_i \pi_i, c) \notin \mathcal{A}_i\}.$$

Then we have

$$\begin{aligned}
&\mathbb{P}_{\mathbf{z}_l\in\mathbb{F}_q^l}(\mathbf{z}_l \in S_l)\\
=&\mathbb{P}_{\mathbf{z}_l\in\mathbb{F}_q^l}(\mathbf{z}_l \in S_l \mid \mathbf{z}_{l-1} \in S_{l-1}) \cdot \mathbb{P}_{\mathbf{z}_{l-1}\in\mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in S_{l-1}) +\\
&\mathbb{P}_{\mathbf{z}_l\in\mathbb{F}_q^l}(\mathbf{z}_l \in S_l \mid \mathbf{z}_{l-1} \notin S_{l-1}) \cdot \mathbb{P}_{\mathbf{z}_{l-1}\in\mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \notin S_{l-1})\\
\leq&\mathbb{P}_{\mathbf{z}_{l-1}\in\mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in S_{l-1}) + \mathbb{P}_{\mathbf{z}_l\in\mathbb{F}_q^l}(\mathbf{z}_l \in S_l \mid \mathbf{z}_{l-1} \notin S_{l-1}). \qquad (17)
\end{aligned}$$

According to our assumption, the first item satisfies

$$\mathbb{P}_{\mathbf{z}_{l-1}\in\mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in S_{l-1}) < \frac{2\Delta_C n}{9\rho\eta^3|\mathbb{F}_q|} \cdot (l-1). \qquad (18)$$

For the second item, for any fixed $\mathbf{z}_{l-1} \notin S_{l-1}$, define the set

$$S_{\mathbf{z}_{l-1}} \triangleq \{\alpha_l \in \mathbb{F}_q : \langle \mathbf{z}_{l-1}, \alpha_l\rangle \in S_l\}.$$

For convenience, use π' to denote $\pi_0 + \cdots + \alpha_{l-1}\pi_{l-1}$. Consider the set

$$\begin{aligned}
S_{\delta,\{\pi',\pi_l\},C} = \{\alpha_l \in \mathbb{F}_q \mid &\exists c \in C, \text{ such that } \delta(\pi' + \alpha_l\pi_l, v) \leq \delta\\
&\text{and Agree}(\pi' + \alpha_l\pi_l, v) \notin \mathcal{A}_{\delta,\{\pi',\pi_l\},C}\}.
\end{aligned}$$

We want to prove

$$S_{\mathbf{z}_{l-1}} \subseteq S_{\delta,\{\pi',\pi_l\},C}. \qquad (19)$$

Notice that $|S_{\delta,\{\pi',\pi_l\},C}| < \frac{2\Delta_C n}{9\rho\eta^3}$ by Theorem 5. So if (19) holds, we have

$$|S_{\mathbf{z}_{l-1}}| < \frac{2\Delta_C n}{9\rho\eta^3}. \qquad (20)$$

Suppose $\mathcal{A}_{\delta,\{\pi',\pi_l\},C} = \{D'_1, \ldots, D'_{m'_1}\}$. Since $\mathbf{z}_{l-1} \notin S_{l-1}$, $\{D'_1, \ldots, D'_{m'_1}\}$ are δ-correlated-agree domains of $\pi_0, \cdots, \pi_{l-1}$ (may be not maximal). According to the definition of $\mathcal{A}_{\delta,\{\pi',\pi_l\},C}$, $\{D'_1, \ldots, D'_{m'_1}\}$ are δ-correlated-agree domains of $\pi_0, \cdots, \pi_l$. For any $D'_i \in \mathcal{A}_{\delta,\{\pi',\pi_l\},C}$, $\exists D \in \mathcal{A}_l$, such that $D'_i \subseteq D$. Since D is a correlated-agreement between π', π_l and C, we have $D'_i = D$. Thus,

$$\mathcal{A}_{\delta,\{\pi',\pi_l\},C} \subseteq \mathcal{A}_l. \qquad (21)$$

Fix an $\alpha \in S_{\mathbf{z}_{l-1}}$, $\langle \mathbf{z}_{l-1}, \alpha\rangle \in S_l$. According to the definition of S_l, $\exists c \in C$ such that

$$\text{Agree}(\pi' + \alpha\pi_l, c) \notin \mathcal{A}_l \text{ and } |\text{Agree}(\pi' + \alpha\pi_l, c)| \geq (1-\delta)n.$$

(21) tells us $\text{Agree}(\pi' + \alpha\pi_l, c) \notin \mathcal{A}_{\delta,\{\pi',\pi_l\},C}$. Thus,

$$\alpha \in S_{\delta,\{\pi',\pi_l\},C}$$

and we have proved (19).

Plugging (18) and (20) into (17), we have

$$\mathbb{P}_{\mathbf{z}_l\in\mathbb{F}_q^l}(\mathbf{z}_l \in S_l) < \frac{2\Delta_C n}{9\rho\eta^3|\mathbb{F}_q|} \cdot l.$$

□

Corollary 6 (Linear proximity gap for linear codes over affine spaces)**.** *Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$, where $n > k > 0$. Denote by Δ_C the minimum relative distance of C. Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C} - \eta$ and $\eta \leq \frac{1}{3}(1-\Delta_C)^{-\frac{1}{3}}$. Let $\pi_0, \ldots, \pi_l \in \mathbb{F}_q^n$ be $l+1$ codewords, where $l \geq 1$. Suppose*

$$\mathbb{P}_{\langle \alpha_1, \ldots, \alpha_l \rangle \in \mathbb{F}_q}(\Delta(\pi_0 + \ldots + \alpha_l \pi_l, C) \leq \delta) \geq \frac{2\Delta_C n}{9(1-\Delta_C)\eta^3 |\mathbb{F}_q|} \cdot l.$$

Then $\pi_0, \ldots, \pi_l$ are simultaneously δ-close to C, i.e., $\exists c_0, \ldots, c_l \in C$ such that

$$|\{j \in [1, n] : \pi_0[j] = c_0[j], \ldots, \pi_l[j] = c_l[j]\}| \geq (1-\delta)n.$$

5.3 Conjectured correlated agreement

Besides the provable correlated agreement for linear codes, conjectures are used in practice. [Ben+20b] provided a conjecture on the correlated agreement of Reed-Solomon codes. Later, [Arn+24b] gave a stronger version of the conjecture on mutual correlated agreement.

Conjecture 1 (Conjecture 8.4 in [Ben+20b])**.** *There exist universal constants $c_1, c_2 > 0$ such that the following holds. Let $u_0, u_1 : L \to \mathbb{F}_q$. Let $\delta, \eta > 0$ and $\delta \leq 1 - \rho - \eta$, and suppose*

$$\mathbb{P}_{z \in \mathbb{F}_q}(\Delta(u_0 + z u_1, \mathrm{RS}[\mathbb{F}_q, L, \rho]) \leq \delta) > \frac{1}{(\eta\rho)^{c_1}} \cdot \frac{|L|^{c_2}}{|\mathbb{F}_q|}.$$

Then u_0, u_1 are simultaneously δ-close to $\mathrm{RS}[\mathbb{F}_q, L, \rho]$, i.e. $\exists v_0, v_1 \in \mathrm{RS}[\mathbb{F}_q, L, \rho]$ such that

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

[Ben+20b] proved the conjecture when $c_2 = 2$ and δ is under the Johnson bound, i.e., $\delta \leq 1 - \sqrt{\rho} - \eta$. They state that "To the best of our knowledge, nothing contradicts setting $c_1 = c_2 = 2$" and "When limiting the scope to fields of characteristic greater than k (degree of the RS code), we are not aware of anything contradicting $c_1 = c_2 = 1$".

Theorem 6 provides proof for a part of the conjecture when setting $c_2 = 1$. Furthermore, our results hold for all linear codes and support the mutual correlated agreement, which is stronger than the standard correlated agreement. The parameter δ needs to be under the 1.5 Johnson bound in our setting, i.e., $\delta \leq 1 - \sqrt[3]{\rho} - \eta$.

The proof of the remaining part of the conjecture is still open.

6 Soundness of batched FRI

The proximity gap for linear codes is widely used in many code-based protocols. In this section, we applied our results to a widely used low-degree test protocol, called FRI[Ben+18]. FRI is used to verify whether a given codeword is close to an RS code or is far from it. We improve the soundness analysis of FRI. Our result can also be directly used to prove the round-by-round soundness of FRI.

Furthermore, our results can be applied in a series of generalizations of FRI. In particular, we provide the corresponding result in the version of foldable codes, which is formally defined in BaseFold[ZCF23]. BaseFold is field-agnostic while FRI only works for "smooth" domains.

6.1 Bad folding points

Folding is a crucial operation in FRI. The protocol applies this operation repeatedly to finish the verification. We first analyze the performance of one folding. We introduce the generalization definition of folding provided in [ZCF23], which can be operated on the *foldable* codes. *Foldable* codes are a subset of linear codes and RS codes are special cases.

Suppose $\pi \in \mathbb{F}_q^{2n}$ is a codeword with length $2n, n \in \mathbb{N}^+$ and $C \subseteq \mathbb{F}_q^{2n}$ is a given linear code. The folding result of π at α, denoted as $\mathrm{Fold}_\alpha(\pi)$, is a codeword with length n.

Definition 12 (Message folding [ZCF23])**.** *Let $c, k_0 \in \mathbb{N}, d \in \mathbb{N}^+$ and let $\mathbb{F}_q$ be a finite field. Let $\mathbf{G}_{d-1} \in \mathbb{F}_q^{k_0 2^{d-1} \times ck_0 2^{d-1}}$ be an arbitrary matrix and $C_{d-1} : \mathbb{F}_q^{k_0 \cdot 2^{d-1}} \to \mathbb{F}_q^{ck_0 \cdot 2^{d-1}}$ be the corresponding code. Let $C_d : \mathbb{F}_q^{k_0 \cdot 2^d} \to \mathbb{F}_d^{ck_0 \cdot 2^d}$ be a linear code with generator matrix $\mathbf{G}_d \in \mathbb{F}_q^{k_0 2^d \times ck_0 2^d}$ with structure*

$$\mathbf{G_d} = \begin{pmatrix} \mathbf{G}_{d-1} & \mathbf{G}_{d-1} \\ \mathbf{G}_{d-1} \cdot T_{d-1} & \mathbf{G}_{d-1} \cdot T'_{d-1} \end{pmatrix},$$

where $T_{d-1}, T'_{d-1} \in \mathbb{F}_q^{ck_0 \cdot 2^{d-1} \times ck_0 \cdot 2^{d-1}}$ are arbitrary diagonal matrices such that $T_{d-1}[j] \neq T'_{d-1}[j]$ for $j \in [1, ck_0 \cdot 2^d]$.

For all $\mathbf{m}_d \in \mathbb{F}_q^{k_0 \cdot 2^d}$, denote by $\mathrm{Enc}_d(\mathbf{m}_d) \in \mathbf{F}_q^{ck_0 \cdot 2^d}$ the encoding of $\mathbf{m}_d$, i.e., $\mathrm{Enc}_d(\mathbf{m}_d) = \mathbf{m}_d \cdot \mathbf{G}_d$. $\alpha \in \mathbb{F}_q$ is a folding point, then the folding result of $\mathrm{Enc}_d(\mathbf{m}_d)$ at point α is defined as:

$$\mathrm{CodeFold}_\alpha(\mathrm{Enc}_d(\mathbf{m}_d)) \triangleq (\mathbf{m_{d,l}} + \alpha \mathbf{m_{d,r}}) \cdot \mathbf{G}_{d-1}.$$

By definition, the folding result of an element in C_d at any point is an element in C_{d-1}.

Definition 13 (Codeword folding [ZCF23])**.** *Let $c, k_0 \in \mathbb{N}, d \in \mathbb{N}^+$ and let $\mathbb{F}_q$ be a finite field. Let $\mathbf{G}_{d-1} \in \mathbb{F}_q^{k_0 2^{d-1} \times ck_0 2^{d-1}}, T_{d-1}, T'_{d-1} \in \mathbb{F}_q^{ck_0 \cdot 2^{d-1} \times ck_0 \cdot 2^{d-1}}$ and $\mathbf{G}_d \in \mathbb{F}_q^{k_0 2^d \times ck_0 2^d}$ be matrices defined as above. Let $\pi_d \in \mathbb{F}_q^{ck_0 2^d}$ be any vector (may not be an element of C_d). $\alpha \in \mathbb{F}_q$ is a folding point, then the folding result of π_d at point α, denoted by $\mathrm{Fold}_\alpha(\pi_d) \in \mathbb{F}_q^{ck_0 2^{d-1}}$ is defined as:*

$$\mathrm{Fold}_\alpha(\pi_d)[j] \triangleq \frac{\mathrm{diag}(T_{d-1})[j]\pi_d[j + ck_0 2^{d-1}] - \mathrm{diag}(T'_{d-1})[j]\pi_d[j]}{\mathrm{diag}(T_{d-1})[j] - \mathrm{diag}(T'_{d-1})[j]} + \alpha \cdot \frac{\pi_d[j] - \pi_d[j + ck_0 2^{d-1}]}{\mathrm{diag}(T_{d-1})[j] - \mathrm{diag}(T'_{d-1})[j]},$$

where $j \in [1, ck_0 2^{d-1}]$.

Proposition 2 ([ZCF23])**.** *When $\pi_d \in C_d$, for any $\alpha \in \mathbb{F}_q$, we have*

$$\mathrm{CodeFold}_\alpha(\pi_d) = \mathrm{Fold}_\alpha(\pi_d).$$

The proof can be found in Lemma 5 in [ZCF23].

The above proposition implies that when we do the codeword folding operation on a codeword in C_d, the folding result is an element in C_{d-1}. As a result, we can do folding operations repeatedly and check whether the final folding result is a member of C_0. Notice that the final folding result has a short code length, making verification easy.

However, there exist "unlucky" cases (over the choice of folding point) that fold a codeword $\pi \notin C_d$ into a member of C_{d-1}. The following result limits the possibility of such cases. More precisely, we limit the possibility over the choice of folding point that *discloses* the distance between π and C_d.

Definition 14 (bad folding points)**.** *Let $c, k_0 \in \mathbb{N}, d \in \mathbb{N}^+$ and let $\mathbb{F}_q$ be a finite field. Let $2n = ck_0 2^d$ and $\rho = \frac{1}{c}$ be the rate. Let $C_{d-1} \subseteq \mathbb{F}_q^n$ be a linear code with generator matrix $\mathbf{G}_{d-1} \in \mathbb{F}_q^{\rho n \times n}$. $T_{d-1}, T'_{d-1} \in \mathbb{F}_q^{n \times n}$ are two arbitrary diagonal matrix such that $\mathrm{diag}(T_{d-1})[j] \neq \mathrm{diag}(T'_{d-1})[j]$ for every $j \in [1, n]$. Define*

$$\mathbf{G}_d = \begin{pmatrix} \mathbf{G}_{d-1} & \mathbf{G}_{d-1} \\ \mathbf{G}_{d-1} \cdot T_{d-1} & \mathbf{G}_{d-1} \cdot T'_{d-1} \end{pmatrix}$$

and $C_d : \mathbb{F}_q^{\rho(2n)} \to \mathbb{F}_q^{2n}$ is the corresponding linear code.

Let $0 < \delta < 1$. Let $\pi \in \mathbb{F}_q^{2n}$ be a codeword such that $\delta(\pi, C_d) > \delta(2n)$. Define the bad folding points *to be*

$$\mathrm{Bad}(\pi) = \{\alpha \in \mathbb{F}_q : \delta\left(\mathrm{Fold}_\alpha(\pi), C_{d-1}\right) \leq \delta)\}.$$

Denote by Δ_{d-1} the minimum relative distance of C_{d-1}. Using Corollary 1, we can limit the number of bad folding points when δ is within the 1.5 Johnson bound.

Theorem 7. *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1 - \Delta_{d-1}} - \eta$ and $\eta \leq \frac{1}{3}(1 - \Delta_{d-1})^{-\frac{1}{3}}$. Let $\pi \in \mathbb{F}_q^{2n}$ be a codeword such that $\delta(\pi, C_d) > \delta$. Then,*

$$|\mathrm{Bad}(\pi)| < \frac{\Delta_{d-1}(2n)}{9(1 - \Delta_{d-1})\eta^3}.$$

Proof. We use π to construct the following codewords with length n:

$$\begin{cases} \pi_1[j] = \dfrac{\mathrm{diag}(T_{d-1})[j]\pi[j+n] - \mathrm{diag}(T'_{d-1})[j]\pi[j]}{\mathrm{diag}(T_{d-1})[j] - \mathrm{diag}(T'_{d-1})[j]} \\ \pi_2[j] = \dfrac{\pi[j] - \pi[j+n]}{\mathrm{diag}(T_{d-1})[j] - \mathrm{diag}(T'_{d-1})[j]} \end{cases}, \tag{22}$$

where $j \in \{1, \ldots, n\}$.

We claim that π_1 and π_2 do not have δ-correlated agreement. Otherwise, suppose there exist a subdomain $D \subseteq \{1, \ldots, n\}$ and two codewords $c_1, c_2 \in C_{d-1}$ such that $|D| \geq (1 - \delta)n$ and $\pi_1[D] = c_1[D], \pi_2[D] = c_2[D]$. Let $\mathbf{m}_1, \mathbf{m}_2 \in \mathbb{F}_q^{\rho n}$ be the corresponding messages of c_1, c_2, i.e., $c_1 = \mathbf{m}_1 \cdot \mathbf{G}$ and $c_2 = \mathbf{m}_2 \cdot \mathbf{G}$. Combining with (22), we have

$$\begin{cases} \pi[j] = \pi_1[j] + \mathrm{diag}(T_{d-1})[j]\pi_2[j] = (\mathbf{m}_1 \cdot \mathbf{G}_{d-1})[j] + \mathrm{diag}(T_{d-1})[j](\mathbf{m}_2 \cdot \mathbf{G}_{d-1})[j] \\ \pi[j+n] = \pi_1[j] + \mathrm{diag}(T'_{d-1})[j]\pi_2[j] = (\mathbf{m}_1 \cdot \mathbf{G}_{d-1})[j] + \mathrm{diag}(T'_{d-1})[j](\mathbf{m}_2 \cdot \mathbf{G}_{d-1})[j] \end{cases},$$

for $j \in D$. Let $\mathbf{m} = \langle \mathbf{m}_1 \| \mathbf{m}_2 \rangle \in \mathbb{F}_q^{2n}$ be the concatenation of $\mathbf{m}_1$ and $\mathbf{m}_2$. Let $D' = \{j \in [2n] \mid j \in D \text{ or } j - n \in D\}$. Then we have

$$\pi[D'] = \mathbf{m} \cdot \mathbf{G}[D'].$$

This is a contradiction to $\delta(\pi, C_d) > \delta$. Thus, we have proved π_1 and π_2 do not have δ-correlated agreement.

Next, we prove that $\mathrm{Bad}(\pi) \subseteq \mathrm{Bad}(\pi_1, \pi_2)$. $\forall \alpha \in \mathrm{Bad}(\pi)$, notice that $\pi_1 + \alpha\pi_2 = \mathrm{Fold}_\alpha(\pi)$ according to Definition 13 and (22). Then we have

$$\delta(\pi_1 + \alpha\pi_2, C_{d-1}) = \delta(\mathrm{Fold}_\alpha(\pi), C_{d-1}) \leq \delta,$$

which implies $\mathrm{Bad}(\pi) \subseteq \mathrm{Bad}(\pi_1, \pi_2)$.

We have proved $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, C_d) > \delta$. By Theorem 3, $|\text{Bad}(\pi_1, \pi_2)| < \frac{\Delta_{d-1}(2n)}{9(1-\Delta_{d-1})\eta^3}$. As a result,

$$|\text{Bad}(\pi)| \leq |\text{Bad}(\pi_1, \pi_2)| < \frac{\Delta_{d-1}(2n)}{9(1-\Delta_{d-1})\eta^3}.$$

□

Remark 1. *[ZCF23] provides an analysis of the minimum relative distance of foldable codes. When applied to RS codes, the minimum relative distance is fixed, i.e., we have $\Delta_{d-1} = \Delta_d$.*

6.2 The batched FRI protocol

FRI[Ben+18] is an IOPP for testing proximity to the RS codes. It is used to help the verifier check whether a given function $f : L^{(0)} \to \mathbb{F}_q$ belongs to a given RS code or is far from it. In particular, FRI works for *smooth* evaluation sets. We say a set is smooth if it is a coset of a multiplicative group whose order is a power of 2. We apply our results to prove the soundness of FRI.

Let $L^{(0)}$ be a smooth domain, $0 < \rho < 1$. Let $\text{RS}[\mathbb{F}_q, L, \rho] : \mathbb{F}_q^{d+1} \to \mathbb{F}_q^{|L|}$ denote the Reed-Solomon code of degree strictly less than $\rho|L|$ evaluated on L, where $\text{RS}[\mathbb{F}_q, L, \rho]$ maps $(c_0, c_1, \ldots, c_d) \in \mathbb{F}_q^{d+1}$ to $\left(\sum_{i=0}^{k} c_i x^i\right)_{x \in L} \in \mathbb{F}_q^{|L|}$, and $d = \lceil \rho|L| \rceil - 1$. Notice that RS codes are linear codes. For a given function $f^{(0)} : L^{(0)} \to \mathbb{F}_q$, the verifier wants to know whether it is a member of $\text{RS}[\mathbb{F}_q, L^{(0)}, \rho]$. An untrusted prover may help the verifier while the verifier has query accesses to $f^{(0)}$. The verifier and the prover agree on a series of smooth evaluation domains $L^{(0)}, L^{(1)}, \cdots, L^{(n_r)}$, where n_r is the number of interactive rounds. For convenience, we will focus on a specific case of folding where $L^{(k+1)} = (L^{(k)})^2$. At this time, the matrices $\mathbf{G}_d, T_d, T'_d$ in Definition 12 are of a specific structure. And we fold a member of $\text{RS}[\mathbb{F}_q, L^{(d)}, \rho]$ into a member of $\text{RS}[\mathbb{F}_q, L^{(d-1)}, \rho]$. For general cases, the definition of folding can be found in [Ben+18], and we will not discuss those details here.

The FRI protocol has two phases, called COMMIT and QUERY.

In the COMMIT phase, the prover and the verifier work together round by round to *fold* the target function $f^{(0)}$ into a field element (or a short vector). Thus, the verifier can check the element easily. In the k^{th} round, the prover sends the oracle of a function $f^{(k)}$ to the verifier. The verifier randomly selects a folding parameter $\alpha^{(k)} \in \mathbb{F}_q$ and sends it to the prover. In this context, we assume that the folding parameter cannot be zero. Upon receiving $\alpha^{(k)}$, the prover folds $f^{(k)}$ using this parameter to obtain a new function $f^{(k+1)} : L^{(k+1)} \to \mathbb{F}_q$. If the prover is honest, the folding result is supposed to be

$$f^{(k+1)} = \text{Fold}_{\alpha^{(k)}}(f^{(k)}).$$

If $f^{(k)}$ is a member of $\text{RS}[\mathbb{F}_q, L^{(k)}, \rho]$, then the degree of $f^{(k+1)}$ is expected to be halved. Consequently, any member of $\text{RS}[\mathbb{F}_q, L^{(0)}, \rho]$ will be folded into a single element after $\log(\rho|L^{(0)}|)$ rounds.

In the QUERY phase, the verifier queries some random locations in $L^{(0)}$, and the prover responds with the queried elements as well as those involved in the folding path. The verifier then calculates the folding results to verify the correctness of the folding process.

Batching Batched FRI is a generalization of the FRI protocol. Instead of checking only one function $f^{(0)}$ is near $\text{RS}[\mathbb{F}_q, L^{(0)}, \rho]$, the prover is now required to prove a series of functions $f_0^{(0)}, \cdots, f_l^{(0)} : L^{(0)} \to \mathbb{F}_q$ are near $\text{RS}[\mathbb{F}_q, L^{(0)}, \rho]$. A trivial strategy is checking each function individually; however, this approach becomes inefficient when the number of functions is large. Batched FRI provides a way to do the verifications at one time.

Suppose the prover has a series of functions $f_0^{(0)}, \cdots, f_l^{(0)} \in \mathbb{F}_q^{L^{(0)}}$, and the verifier has oracle access to these functions. Before executing the FRI protocol, the verifier randomly selects $z_1, \cdots, z_l \in \mathbb{F}_q$ and sends them to the prover. The prover and the verifier then run the FRI protocol on the combined function defined as $f^{(0)} \triangleq f_0^{(0)} + z_1 f_1^{(0)} + \cdots + z_l f_l^{(0)}$. The COMMIT phase of the batched FRI protocol is the same as the basic FRI protocol, while the QUERY phase includes additional checks to verify that the combination is correct. More precisely, the batched FRI protocol works as follows:

BATCH **Phase:**

1. The verifier picks uniformly random $z_1, \ldots, z_l \in \mathbb{F}_q^{\times}$.
2. Set $f^{(0)} \triangleq f_0^{(0)} + z_1 f_1^{(0)} + \cdots + z_l f_l^{(0)}$.

COMMIT **Phase:**

1. For each $k \in [0, n_r - 1]$:
   (a) The verifier picks a uniformly random $\alpha^{(k)} \in \mathbb{F}_q^{\times}$.
   (b) The prover writes down a function
   $$f^{(k+1)} : L^{(k+1)} \to \mathbb{F}_q$$
   and sends the oracle of $f^{(k+1)}$ to the verifier. For an honest prover, we have $f^{(k+1)} = \mathrm{Fold}_{\alpha^{(k)}}(f^{(k)})$.
2. The prover writes down a value $C \in \mathbb{F}_q$.

QUERY **Phase:** Repeat t times:

1. The verifier picks a uniformly random $s^{(0)} \in L^{(0)}$.
2. If $f^{(0)}(s^{(0)}) \neq f_0^{(0)}(s^{(0)}) + z_1 f_1^{(0)}(s^{(0)}) + \cdots + z_l f_l^{(0)}(s^{(0)})$, REJECT.
3. For each $i \in [0, n_r - 1]$:
   (a) Define $s^{(k+1)} \in L^{(k+1)}$ by $s^{(k+1)} = (s^{(k)})^2$.
   (b) Compute $\mathrm{Fold}_{\alpha^{(k)}}(f^{(k)})(s^{(k+1)})$ by making queries to $f^{(k)}(s^{(k)})$ and $f^{(k)}(-s^{(k)})$.
   (c) If $\mathrm{Fold}_{\alpha^{(k)}}(f^{(k)})(s^{(k+1)}) \neq f^{(k+1)}(s^{(k+1)})$, REJECT.
4. If $f^{(n_r)}(s^{(n_r)}) \neq C$, REJECT.
5. ACCEPT.

6.3 Soundness of batched FRI

The soundness error of batched FRI consists of bad batching, bad folding, and prover's cheating.

The bad batching is restricted by Theorem 6 directly. We propose an analysis of the possibility of bad folding based on Theorem 7. Let $f^{(0)} : L^{(0)} \to \mathbb{F}_q$ be the initial function. The prover and the verifier have agreed on a series of "smooth" evaluation domains, $L^{(0)}, L^{(1)}, \ldots$. Suppose there are n_r rounds in the FRI protocol. Let

$$\delta^{(k)} \triangleq \delta(f^{(k)}, \mathrm{RS}[\mathbb{F}_q, L^{(k)}, \rho])$$

be the relative distance.

Let $B^{(k)} = \min\{\delta^{(k)} - \frac{1}{|L^{(k+1)}|}, 1 - \sqrt[3]{\rho} - \eta\}$. Let $\eta \leq \frac{1}{3}\rho^{-\frac{1}{3}}$. Define the k^{th} *Bad Event* $E^{(k)}, 0 \leq k \leq n_r - 1$ as the event:

$$E^{(k)} = \{\alpha^{(k)} \in \mathbb{F}_q : \delta(\mathrm{Fold}_{\alpha^{(k)}}(f^{(k)}), \mathrm{RS}[\mathbb{F}_q, L^{(k+1)}, \rho]) \leq B^{(k)}\}$$

where $\alpha^{(k)}$ is the random folding point chosen by the verifier in the k^{th} round. According to Corollary 1, we have

$$\mathbb{P}(E^{(k)}) < \frac{(1-\rho)|L^{(k)}|}{9\rho\eta^3|\mathbb{F}_q|}.$$

Notice that $|L^{(k+1)}| = \frac{|L^{(k)}|}{2}$. Then the possibility that in all of the n_r rounds, the bad events do not happen satisfies:

$$\begin{aligned}\mathbb{P}(\bigwedge_{k=0}^{n_r-1} \neg E^{(k)}) &\geq 1 - \sum_{k=0}^{n_r-1} \mathbb{P}(E^{(k)}) \\ &> 1 - \frac{2(1-\rho)|L^{(0)}|}{9\rho\eta^3|\mathbb{F}_q|}.\end{aligned}$$

Suppose bad batching and bad folding do not happen in all the n_r rounds. A dishonest prover may modify some locations of the codeword to pass the verification. However, modifications will be checked during the QUERY phase and can not increase the possibility of passing. As a result, we have the following soundness error bound of batched FRI.

Suppose that bad batching and bad folding do not occur in any of the n_r rounds. A dishonest prover may alter some positions of the codeword to pass verification; however, these modifications will be scrutinized during the QUERY phase and cannot increase the likelihood of passing. Consequently, we derive the following soundness error bound for batched FRI. Further details regarding the soundness of FRI can be found in [Ben+18], which offers a comprehensive soundness analysis.

Theorem 8 (Batched FRI soundness)**.** *Let $\mathbb{F}_q$ be a finite field. Let $L^{(0)} \subseteq \mathbb{F}_q$ be a smooth evaluation domain.*

Let $f_0^{(0)}, \cdots, f_l^{(0)} : L^{(0)} \to \mathbb{F}_q, 1 \leq l$ be a sequence of functions and let $V^{(0)} = \mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$ and ρ satisfies $\rho = 2^{-R}$ for a positive integer R. Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{\rho} - \eta$ and $\eta \leq \frac{1}{3}\rho^{-\frac{1}{3}}$. Furthermore, let t denote 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 a probability greater than

$$\epsilon_{\textit{Batched-FRI}} = \left(\frac{2(1-\rho)|L^{(0)}|}{9\rho\eta^3|\mathbb{F}_q|}\right) \cdot (l+1) + (1-\delta)^t \tag{23}$$

Then $f_0^{(0)}, \cdots, f_l^{(0)}$ *have correlated agreement with* $V^{(0)}$ *on a domain* $D \subseteq L^{(0)}$ *of density at least* $1 - \delta$.

Remark 2. *For general cases that* $q(X) = X^{2^k}, k \in \mathbb{N}^*$*, this error bound also holds. One folding in this case can be seen as* k *foldings of the special case with the same folding parameter.*

6.4 Numerical Example

We provide a numerical example to show the improvement in the provable soundness of FRI. Set $q = |\mathbb{F}_q| > 2^{183}$ (the extension field used in [Sta23]), $\rho = \frac{1}{8}$, $m = 3$, $\eta = 2^{-6}$, $|L^{(0)}| = 2^{24}$ and $l = 28$. $n_r = \log_2(|L^{(0)}|) = 24$ is the number of rounds. t is the number of QUERY times.

Let

$$\epsilon_c \triangleq \frac{(m + \frac{1}{2})^7 \cdot |L^{(0)}|^2}{2\rho^{3/2} q} + \frac{(2m+1) \cdot (|L^{(0)}| + 1)}{\sqrt{\rho}} \cdot \frac{\sum_{i=0}^{n_r - 1} l^{(i)}}{q},$$

where $l^{(i)} = \frac{|L^{(i)}|}{|L^{(i+1)}|} = 2$ in our example. And we have

$$2^{-122} < \epsilon_c < 2^{-121}.$$

The soundness error bound provided in [Ben+20b] is

$$\epsilon_{\text{Batched-FRI}} = \epsilon_c + \left(\sqrt{\rho}(1 + \frac{1}{2m})\right)^t. \tag{24}$$

This can reach 121 bits of security when $t \geq 97$, and can not reach 128 bits of security. For higher security levels, we can apply the FRI protocol in a bigger extension field. However, this will increase the cost of operations. Our soundness error bound is provided in (23). And we have

$$2^{-136} < \left(\frac{2(1-\rho)|L^{(0)}|}{9\rho\eta^3|\mathbb{F}_q|}\right) \cdot (l+1) < 2^{-135}.$$

We prove FRI can reach 128 bits of security in the current field when $t \geq 134$.

Acknowledgment

This is the third version of this work. We thank Swastik Kopparty for identifying a critical mistake in the first version and for his invaluable discussions. His insights have greatly assisted us in completing this work. We thank Giacomo Fenzi for suggesting that the second version could be extended to general linear codes.

References

[Ame+17] Scott Ames et al. "Ligero: Lightweight sublinear arguments without a trusted setup". In: *Proceedings of the 2017 acm sigsac conference on computer and communications security*. 2017, pp. 2087–2104.

[Arn+24a] Gal Arnon et al. "STIR: Reed–Solomon Proximity Testing with Fewer Queries". In: *Cryptology ePrint Archive* (2024).

[Arn+24b] Gal Arnon et al. "WHIR: Reed–Solomon Proximity Testing with Super-Fast Verification". In: *Cryptology ePrint Archive* (2024).

[Ben+20a] E Ben-Sasson et al. "DEEP-FRI: Sampling Outside the Box Improves Soundness". In: *11th Innovations in Theoretical Computer Science Conference, ITCS 2020*. LIPIcs Dagstuhl. 2020.

[BCS16] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. "Interactive oracle proofs". In: *Theory of Cryptography: 14th International Conference, TCC 2016-B, Beijing, China, October 31-November 3, 2016, Proceedings, Part II 14*. Springer. 2016, pp. 31–60.

[BKS18] Eli Ben-Sasson, Swastik Kopparty, and Shubhangi Saraf. "Worst-case to average case reductions for the distance to a code". In: *33rd Computational Complexity Conference (CCC 2018)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik. 2018.

[Ben+18] Eli Ben-Sasson et al. "Fast reed-solomon interactive oracle proofs of proximity". In: *45th international colloquium on automata, languages, and programming (icalp 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018.

[Ben+19] Eli Ben-Sasson et al. "Aurora: Transparent succinct arguments for R1CS". In: *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*. Springer. 2019, pp. 103–128.

[Ben+20b] Eli Ben-Sasson et al. "Proximity gaps for Reed–Solomon codes". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 900–909.

[Blo+23] Alexander R Block et al. "Fiat-Shamir security of FRI and related snarks". In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2023, pp. 3–40.

[GKL24] Yiwen Gao, Haibin Kan, and Yuan Li. "Improved Soundness Analysis of the FRI Protocol". In: *Cryptology ePrint Archive* (2024).

[Juk11] Stasys Jukna. *Extremal combinatorics: with applications in computer science*. Vol. 571. Springer, 2011.

[KPV22] Assimakis A Kattis, Konstantin Panarin, and Alexander Vlasov. "RedShift: transparent SNARKs from list polynomial commitments". In: *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 2022, pp. 1725–1737.

[Mul54] David E Muller. "Application of Boolean algebra to switching circuit design and to error detection". In: *Transactions of the IRE professional group on electronic computers* 3 (1954), pp. 6–12.

[Pol] Polygon. *Plonky2: Fast recursive arguments with plonk and fri*. https://github.com/mir-protocol/plonky2/tree/main/plonky2. URL: https://github.com/mir-protocol/plonky2/tree/main/plonky2.

[RS60] Irving S Reed and Gustave Solomon. "Polynomial codes over certain finite fields". In: *Journal of the society for industrial and applied mathematics* 8.2 (1960), pp. 300–304.

[RRR16] Omer Reingold, Guy N Rothblum, and Ron D Rothblum. "Constant-round interactive proofs for delegating computation". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016, pp. 49–62.

[RVW13] Guy N Rothblum, Salil Vadhan, and Avi Wigderson. "Interactive proofs of proximity: delegating computation in sublinear time". In: *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. 2013, pp. 793–802.

[Sta23] StarkWare. *ethSTARK Documentation v1.2*. Cryptology ePrint Archive, Paper 2021/582. https://eprint.iacr.org/2021/582. 2023. URL: https://eprint.iacr.org/2021/582.

[Xie+22] Tiancheng Xie et al. "zkbridge: Trustless cross-chain bridges made practical". In: *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 2022, pp. 3003–3017.

[Zei24] Hadas Zeilberger. "Khatam: Reducing the Communication Complexity of Code-Based SNARKs". In: *Cryptology ePrint Archive* (2024).

[ZCF23] Hadas Zeilberger, Binyi Chen, and Ben Fisch. "BaseFold: Efficient Field-Agnostic Polynomial Commitment Schemes from Foldable Codes". In: *Cryptology ePrint Archive* (2023).

[Zha+20] Jiaheng Zhang et al. "Transparent polynomial delegation and its applications to zero knowledge proof". In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2020, pp. 859–876.