Linear Proximity Gaps for Linear Codes within the 1.5 Johnson Bound

Yiwen Gao*, Haibin Kan†and Yuan Li‡

March 30, 2025

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 provide 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.

Keywords— Proximity gap, Linear codes.

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

*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

this problem under the interactive oracle proofs of proximity (IOPP) model [BCS16] (also called 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 $\pi = \{\pi_0, \ldots, \pi_l\} \subseteq \mathbb{F}_q^n$, 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 π' in the span of π (denoted by $\mathrm{span}(\pi)$) and check if π' is close to C. The soundness proof of this method raises an important question: *If $\exists \pi_i \in \pi$ that is far from all the members of C, can we prove π' 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 $\pi = \{\pi_0, \pi_1\}$. Here, $\pi' = \pi_0 + \alpha\pi_1, \alpha \in \mathbb{F}_q$ is over a line. For a given linear code C with minimal relative distance Δ_C, we have the following result.

Theorem 1 (Informal)**.** *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1 - \Delta_C + \eta}$. Let $\pi_0, \pi_1 \in \mathbb{F}_q^n$ be two codewords. If there exists $i \in \{0, 1\}$ such that $\delta(\pi_i, C) > \delta$, then*

$$\Pr_{\alpha \in \mathbb{F}_q}(\delta(\pi_0 + \alpha\pi_1, C) \leq \delta) < \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1 - \Delta_C + \eta} - \sqrt{1 - \Delta_C + \eta})}\right) \frac{1}{|\mathbb{F}_q|} = O(\frac{n}{\eta|\mathbb{F}_q|}). \quad (1)$$

We prove that if either π_0 or π_1 is δ-far from the linear code, then with the probability that is linear in the code length, π' is δ-far from the linear code. The formal statement of this theorem is presented in Theorem 3, which is stronger in the following two aspects.

First, we utilize the concept of *correlated agreement*, as defined in [Ben+20b]. A series of functions $\pi_0, \ldots, \pi_l$ have δ-correlated agreement with C if there exists a sufficiently large subdomain $D \subseteq [1, n]$ and $c_0, \ldots, c_l \in C$ such that

$$|D| \geq (1 - \delta)n \text{ and } \pi_i[D] = c_i[D], 0 \leq i \leq l,$$

where $\pi_i[D](c_i[D])$ denotes the sub-codeword of $\pi_i(c_i)$ on D. The definition of correlated agreement is relevant in the context of real-world protocol applications. Notice that even if all of π_i are δ-close to C, they may not have δ-correlated agreement. Our formal theorem supports this stronger notion of agreement; specifically, if π_0 and π_1 do not have δ-correlated agreement, (1) holds.

Second, our result not only contains the combining points α that explicitly disclose the distance, but also accounts for those that have *potential* to disclose the distance. More precisely, for a given $\delta \leq 1 - \sqrt[3]{1 - \Delta_C + \eta}$, π_1, π_2 may have correlated agreement on $[1, n]$. Let $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C} = \{D_1, D_2, \ldots, D_m\}$ denote all the correlated-agree-domains satisfying $|D_i| \geq (1 - \delta)n$. We demonstrate with high probability ((1)) that the δ-agree-domains preserved in the combining result $\pi_1 + \alpha\pi_2$, i.e., coincide with $\{D_1, D_2, \ldots, D_m\}$.

There are some unlucky choices of $\alpha \in \mathbb{F}_q$ that enlarge the agree-domains (notice that $\{D_1, D_2, \ldots, D_m\}$ are δ-agree-domains of $\pi_1 + \alpha\pi_2$ for all $\alpha \in \mathbb{F}_q$). Figure 1 provides an example. We say these α

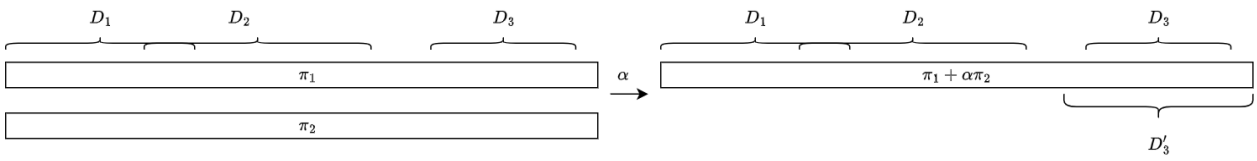


Figure 1: π_1, π_2 are two given codeword with δ-correlated-agree-domains D_1, D_2, D_3. α is a bad combining points such that the δ-agree-domains of $\pi_1 + \alpha\pi_2$ is D_1, D_2, D'_3, where $D_3 \subsetneq D'_3$.

have the potential to disclose the distance. This is because, intuitively, for a pair of codewords π_1, π_2, let $D \subseteq [1, n]$ be the biggest correlated-agree-domain of π_1, π_2 and $|D| < (1 - \delta)n$. Then the bigger D is, the more likely that there exists a combining point α that $\pi_1 + \alpha\pi_2$ is δ-close to C. In the general case of combining a series of codewords $\pi_0, \ldots, \pi_l$, let $\{D_1, D_2, \ldots, D_m\}$ be all the δ-correlated-agree-domains. Suppose $\pi_0, \ldots, \pi_l$ do not have $\delta_1 < \delta$ correlated agreement on $[1, n]$, we want to ask, with what probability (over the choice of combining points $\alpha_1, \ldots, \alpha_l$), the combining result $\pi_0 + \alpha_1\pi_1 + \ldots + \alpha_l\pi_l$ is δ_1-close to C? Notably, certain combining point α_1 may fail to induce δ_1-correlated agreement among $\pi_0 + \alpha_1\pi_1, \pi_2, \ldots, \pi_l$ on $[1, n]$. However, their δ-agree-domains might experience volumetric expansion, thereby enhancing the likelihood of subsequent combinations achieving δ_1-closeness. Theorem 3 bounds the number of such points with bad case potential. This is also defined as *mutual correlated agreement* in [Arn+24a] and *strong correlated agreement* in [Zei24].

With this stronger property, we generalize our result to a batch of codewords that $\pi = \{\pi_0, \cdots, \pi_l\}$.

Theorem 2 (Informal)**.** *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1 - \Delta_C + \eta}$. Let $\pi_0, \cdots, \pi_l \in \mathbb{F}_q^n$ be a batch of codewords. If $\exists \pi_i$ such that $\delta(\pi_i, C) > \delta$, then*

$$\Pr_{\langle \alpha_1, \ldots, \alpha_l \rangle \in \mathbb{F}_q^l}(\delta(\pi_0 + \alpha_1\pi_1 + \cdots + \alpha_l\pi_l, C) \leq \delta) < \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1 - \Delta_C + \eta} - \sqrt{1 - \Delta_C + \eta})} \right) \frac{l}{|\mathbb{F}_q|}.$$

This result also supports the stronger version, and the formal theorem can be found in Theorem 4.

Technique overview. Let C be a linear code and $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two given codewords. We partition the problem into two parts: the low-error case and the high-error case.

In the low-error case, π_1, π_2 have good correlated agreement, and combining $\pi_1 + \alpha\pi_2$ only makes little progress. We select the parameter $\delta_0 = \sqrt{1 - \Delta_C + \eta}$ (the Johnson bound) to define the low-error case, i.e., π_1, π_2 have a correlated-agree-domain larger than $(1 - \delta_0)n$ and the agree-domain of $\pi_1 + \alpha\pi_2$ is enlarged from this. Using a similar skill as the list-decoding, we bound the number of correlated-agree-domains larger than $(1 - \delta_0)n$. Furthermore, for each large correlated-agree-domain, we prove that the number of combining points that can enlarge it is limited. Thus, we bound the probability in this case.

In the high-error case, we follow the approach of [Ben+20a] to bound the probability. The basic idea is to use a combinatorial skill. Let $\alpha \in \mathbb{F}_q$ be a combining point. Let $c \in C$ be the closest codeword of $\pi_1 + \alpha\pi_2$, define the agree-domain to be

$$B_\alpha = \{i \in [1, n] \mid (\pi_1 + \alpha\pi_2)[i] = c[i]\}.$$

They observe that for two distinct $\alpha, \beta \in \mathbb{F}_q$, π_1 and π_2 have correlated agreement on $B_\alpha \cap B_\beta$. Since this is the high-error case, we have $|B_\alpha \cap B_\beta| < (1 - \delta_0)n$. On the other hand, $|B_\alpha| \geq (1 - \delta)n$

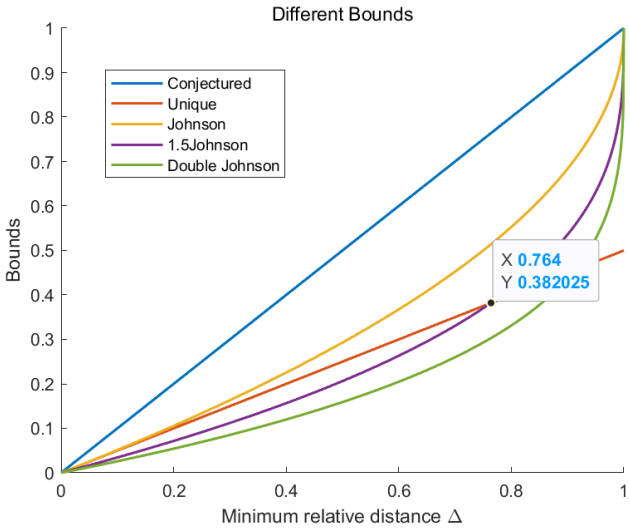


Figure 2: Bounds for linear codes

according to the definition of bad combining points. Thus, the probability can be bounded in this case.

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 $\pi_i \in \pi$ is δ-far from C in relative Hamming distance, denoted by $\delta(\pi_i, C) > \delta$. In [Ame+17], Ames, Hazay, Ishai, and Venkitasubramaniam proved that when $\delta < \Delta_C/4$, with high probability $\pi' \in \text{span}(\pi)$ is δ-far from C. When $\pi' \in \text{span}(\pi)$ is on a line, i.e., $\pi = \{\pi_0, \pi_1\}$, 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) $\pi' \in \text{span}(\pi)$ 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. Setting $\epsilon = \frac{1}{n}$, their results implies with high probability, $\pi' \in \text{span}(\pi)$ is δ-far from C and the probability is linear in n. Our work shows that when $\pi = \{\pi_0, \ldots, \pi_l\}$, the probability is also linear in n under the 1.5 Johnson bound. Especially, when C is an RS code and $\pi = \{\pi_0, \ldots, \pi_l\}$, [Ben+20b] bounded the probability that π' is δ-close to C for the unique decoding bound $\Delta_C/2$ and the Johnson bound $\sqrt{1-\Delta_C}$ respectively.

The generalized definition of bad combining introduced in the last subsection (formal defined in Definition 8), is also called mutual correlated agreement in WHIR[Arn+24a] and can be used to prove the soundness of it. [Arn+24a] provides the mutual correlated agreement for linear codes under the unique decoding bound. Figure 2 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.

[Zei24] provides a definition called *strong correlated agreement*. In their latest version, they provide a strong correlated agreement for linear codes under the one-and-a-half Johnson bound. Setting one $\epsilon = \frac{1}{n}$ in Theorem 2 in their work, they bound the probability to $\frac{2n}{\eta|\mathbb{F}_q|}$. Our methodology achieves a superior constant factor while generalizing the scope to multiple codeword combinations:

Their analysis is confined to a pairwise codewords $\{\pi_1, \pi_2\}$ while our result works for arbitrary linear combinations of codewords $\{\pi_0, \ldots, \pi_l\}$.

A previous version of this work focused on Reed-Solomon codes. And the first version of [Zei24], published a week later, works for linear codes while it only meets the double Johnson bound. They pointed out that our work can be extended to linear codes. Inspired by this, we published another version that works for linear codes and meets the one-and-a-half Johnson bound. This latest version has a better probability than our previous versions.

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]. In particular, the soundness of [Arn+24a] is based on the looser definition of bad combining provided in this work.

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 Section 6. We provide an alternative soundness error bound of FRI:

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

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 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 jth 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 $C[D] \triangleq \{c[D] : c \in C\}$.

2.1 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*

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

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 $c_1 = c_2$.

2.2 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*

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

Definition 6 (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.

When $l = 1$, i.e., there is only one given codeword, the above definition degenerates to the common definition of maximal δ-agree-domains between π_1 and C. For consistency, we use $\mathcal{A}_{\delta,\{\pi_1\},C}$ to denote the set of agree-domains. When neglecting δ, maximal correlated-agree-domains refer to the domains with no density limitation.

Definition 7. *Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords. Function* $\text{Agree}(\pi_1, \pi_2)$ *returns the locations where π_1 agrees with π_2, i.e.,*

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

3 The properties of agree-domains

Let $\mathbb{F}_q$ be a finite field. Let $C \subseteq \mathbb{F}_q^n$ be a linear code with message length k, where $n > k \geq 1$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two codewords with $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, C) \geq 0$. Our goal is to prove that with high probability over the choice of combining points $\alpha \in \mathbb{F}_q$, the combining result $\pi_1 + \alpha\pi_2$ is $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, C)$-far from C. (We consider a generalized case, which will be formally discussed in Section 4 and Section 5.)

This section focuses on the properties of maximal agree-domains between the combining results and C, which will be used to bound the probability. More precisely, let $\alpha \in \mathbb{F}_q$ be any combining point, we define $B_\alpha \subseteq [1, n]$ to be a maximal agree-domain between $\pi_1 + \alpha\pi_2$ and C if there is a codeword $c \in C$ such that $B_\alpha = \text{Agree}(\pi_1 + \alpha\pi_2, c)$. Notice that when $|B_\alpha| \leq (1 - \Delta_C)n$, it is always an agree-domain between $\pi_1 + \alpha\pi_2$ and C but may not be maximal.

Lemma 1. *For any distinct $\alpha_1, \alpha_2 \in \mathbb{F}_q$, let $B_{\alpha_1} \subseteq [1, n]$ be a maximal agree-domain between $\pi_1 + \alpha_1\pi_2$ and C and $B_{\alpha_2} \subseteq [1, n]$ be a maximal agree-domain between $\pi_1 + \alpha_2\pi_2$ and C. Then $B_{\alpha_1} \cap B_{\alpha_2}$ is a correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C, i.e., there exist codewords $c_1, c_2 \in C$ such that*

$$\pi_1[B_{\alpha_1} \cap B_{\alpha_2}] = c_1[B_{\alpha_1} \cap B_{\alpha_2}] \text{ and } \pi_2[B_{\alpha_1} \cap B_{\alpha_2}] = c_2[B_{\alpha_1} \cap B_{\alpha_2}]. \tag{4}$$

Moreover, if $|B_{\alpha_1} \cap B_{\alpha_2}| > (1 - \Delta_C)n$, codewords c_1 and c_2 are uniquely determined and we have

$$B_{\alpha_1} = \text{Agree}(\pi_1 + \alpha_1\pi_2, c_1 + \alpha_1 c_2) \text{ and } B_{\alpha_2} = \text{Agree}(\pi_1 + \alpha_2\pi_2, c_1 + \alpha_2 c_2).$$

Proof. Denote by $\mathbf{G} \in \mathbb{F}_q^k \times \mathbb{F}_q^n$ the generator matrix of C. Let $d_1 = \mathbf{m}_1 \cdot G \in C$ be the codeword such that $B_{\alpha_1} = \text{Agree}(\pi_1 + \alpha_1 \pi_2, d_1)$. Similarly, we have $d_2 = \mathbf{m}_2 \cdot G \in C$ and $B_{\alpha_2} = \text{Agree}(\pi_1 + \alpha_2 \pi_2, d_2)$. For any $j \in B_{\alpha_1} \cap B_{\alpha_2}$, we have

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

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. Notice that $d_1 = c_1 + \alpha_1 c_2, d_2 = c_1 + \alpha_2 c_2$ are uniquely determined, and we have finished our proof. □

Corollary 1. *Let $\alpha, \beta \in \mathbb{F}_q$ be distinct combining points. Let $B_\alpha, B_\beta \subseteq [1, n]$ be maximal agree-domains such that $|B_\alpha \cap B_\beta| > (1 - \Delta_C)n$. Then $B_\alpha \cap B_\beta$ is a maximal correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C.*

Proof. By Lemma 1, $B_\alpha \cap B_\beta$ is a correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C. And let $c_1, c_2 \in C$ be uniquely determined codewords such that

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

Furthermore, $B_\alpha = \text{Agree}(\pi_1 + \alpha \pi_2, \pi_1 + \alpha \pi_2)$ and $B_\beta = \text{Agree}(\pi_1 + \beta \pi_2, c_1 + \beta c_2)$. It remains to prove $B_\alpha \cap B_\beta$ is "maximal".

Let $D \subseteq [1, n]$ be a maximal agree-domain between $\{\pi_1, \pi_2\}$ and C such that $B_\alpha \cap B_\beta \subseteq D$. $|B_\alpha \cap B_\beta| > (1 - \Delta_C)n$ implies

$$\pi_1[D] = c_1[D] \text{ and } \pi_2[D] = c_2[D].$$

For any $j \in D$, we have

$$(\pi_1 + \alpha \pi_2)[j] = (c_1 + \alpha c_2)[j] \text{ and } (\pi_1 + \beta \pi_2)[j] = (c_1 + \beta c_2)[j].$$

Thus, $j \in \text{Agree}(\pi_1 + \alpha \pi_2, \pi_1 + \alpha \pi_2) = B_\alpha$ and $j \in \text{Agree}(\pi_1 + \beta \pi_2, \pi_1 + \beta \pi_2) = B_\beta$. As a result, we have $D \subseteq B_\alpha \cap B_\beta$. □

The following lemma proves a crucial combinatorial property for $B_{\alpha_1}, B_{\alpha_2}, B_{\beta_1}, B_{\beta_2}$, where $\alpha_1 \neq \beta_1, \alpha_2 \neq \beta_2$ and $|B_{\alpha_1} \cap B_{\beta_1}| > (1 - \Delta_C)n, |B_{\alpha_2} \cap B_{\beta_2}| > (1 - \Delta_C)n$. We prove that whenever $B_{\alpha_1} \cap B_{\beta_1} \neq B_{\alpha_2} \cap B_{\beta_2}$, the size of the intersection of $B_{\alpha_1}, B_{\alpha_2}, B_{\beta_1}, B_{\beta_2}$ is "small".

Lemma 2. *Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{F}_q, \alpha_1 \neq \beta_1, \alpha_2 \neq \beta_2$ be combining points. $B_{\alpha_1}, B_{\alpha_2}, B_{\beta_1}, B_{\beta_2}$ are maximal agree-domains such that $|B_{\alpha_1} \cap B_{\beta_1}| > (1 - \Delta_C)n$ and $|B_{\alpha_2} \cap B_{\beta_2}| > (1 - \Delta_C)n$. Denote by $c_1, c_2 \in C$ the codewords uniquely determined by $\pi_1[B_{\alpha_1} \cap B_{\beta_1}], \pi_2[B_{\alpha_2} \cap B_{\beta_1}]$; denote by $c_1', c_2' \in C$ the codewords uniquely determined by $\pi_1[B_{\alpha_1} \cap B_{\beta_2}], \pi_2[B_{\alpha_2} \cap B_{\beta_2}]$ (by Lemma 1). Then exactly one of the followings holds:*

- $c_1 = c_1', c_2 = c_2'$ *and* $B_{\alpha_1} \cap B_{\beta_1} = B_{\alpha_2} \cap B_{\beta_2}$.
- *There exists* $i \in \{1, 2\}$ *such that* $c_i \neq c_i'$, *and* $|B_{\alpha_1} \cap B_{\alpha_2} \cap B_{\beta_1} \cap B_{\beta_2}| \leq (1 - \Delta_C)n$.

Proof. **Case 1**: $c_1 = c_1'$ and $c_2 = c_2'$. $B_{\alpha_1} \cap B_{\beta_1}$ and $B_{\alpha_2} \cap B_{\beta_2}$ are maximal correlated-agree-domains by Corollary 1. Thus, when $c_1, c_2 (c_1', c_2')$ are determined, $B_{\alpha_1} \cap B_{\beta_1} (B_{\alpha_2} \cap B_{\beta_2})$ is determined.

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 $|B_{\alpha_1} \cap B_{\alpha_2} \cap B_{\beta_1} \cap B_{\beta_2}| \leq (1 - \Delta_C)n$. Note that, for any $j \in B_{\alpha_1} \cap B_{\alpha_2} \cap B_{\beta_1} \cap B_{\beta_2}$, by Lemma 1, $\pi_1[j] = c_1[j]$ 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 $|B_{\alpha_1} \cap B_{\alpha_2} \cap B_{\beta_1} \cap B_{\beta_2}| \leq (1 - \Delta_C)n$. □

For any $\alpha \in \mathbb{F}_q$, let $B_\alpha \subseteq [1, n]$ be a maximal agree-domain between $\pi_1 + \alpha\pi_2$ and c for some $c \in C$. Notice that π_1, π_2 may not have correlated agreement with C on B_α. In this case, let $D \subsetneq B_\alpha$ be a non-trivial maximal correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C, i.e., $|D| \geq (1 - \Delta_C)n$ and $\exists c_1, c_2 \in C$ such that $\pi_1[D] = c_1[D], \pi_2[D] = c_2[D]$. We say B_α *expands* D. For a fixed maximal correlated-agree-domain D, the following results bound the number of B_α that expand D.

Lemma 3. *Let* $|D| > (1 - \Delta_C)n$ *be a maximal correlated-agree-domain between* $\{\pi_1, \pi_2\}$ *and* C *and* $c_1, c_2 \in C$ *be the codewords that* $\pi_1[D] = c_1[D]$ *and* $\pi_2[D] = c_2[D]$. *Let* $\beta_1, \ldots, \beta_t \in \mathrm{Bad}_\delta(\pi_1, \pi_2)$ *be distinct and* $B_{\beta_1}, \ldots, B_{\beta_t}$ *are maximal agree-domains such that* $D \subseteq B_{\beta_i}$, *for all* $1 \leq i \leq t$. *Then* $B_{\beta_1}, \ldots, B_{\beta_t}$ *form a sunflower with core* D. *That is, for any distinct* $i, j \in [t]$, *we have*

$$B_{\beta_i} \cap B_{\beta_j} = D.$$

Proof. Let $1 \leq i, j \leq t$ be distinct. Since $D \subseteq B_{\beta_i} \cap B_{\beta_j}$ and $|D| > (1 - \Delta_C)n$, we have $|B_{\beta_i} \cap B_{\beta_j}| > (1 - \Delta_C)n$. By Lemma 1, $B_{\beta_i} \cap B_{\beta_j}$ is a correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C. Since D is maximal, $B_{\beta_i} \cap B_{\beta_j} = D$.

□

Lemma 4. *Let* $|D| > (1 - \Delta_C)n$ *be a maximal correlated-agree-domain between* $\{\pi_1, \pi_2\}$ *and* C *and* $c_1, c_2 \in C$ *be the codewords that* $\pi_1[D] = c_1[D]$ *and* $\pi_2[D] = c_2[D]$. *Let* $\beta_1, \ldots, \beta_t \in \mathrm{Bad}_\delta(\pi_1, \pi_2)$ *be distinct and* $B_{\beta_1}, \ldots, B_{\beta_t}$ *are maximal agree-domains such that* $D \subsetneq B_{\beta_i}$, *for all* $1 \leq i \leq t$. *We have*

$$t \leq n - |D|.$$

Proof. Lemma 3 implies

$$(B_{\beta_i} \setminus D) \cap (B_{\beta_j} \setminus D) = \emptyset \text{ for all distinct } i, j. \tag{5}$$

We have

$$\begin{aligned} n - |D| = |[1, n] \setminus D| &\geq \left| \bigcup_{i=1}^{t} B_{\beta_i} \setminus D \right| \\ &= \sum_{i=1}^{t} |B_{\beta_i} \setminus D| && \text{by (5)} \\ &\geq \sum_{i=1}^{t} 1 = t && \text{by } D \subsetneq B_{\beta_i} \text{ for all } i. \end{aligned} \tag{6}$$

□

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 whose maximal δ-correlated-agree-domain with C is $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C} = \{D_1, \ldots, D_m\}$. Our goal is to prove that with high probability, the δ-agree-domains between the combining codeword $\pi_1 + \alpha\pi_2, \alpha \in \mathbb{F}_q$ is unchanged. Formally, we define the bad combining points as follows:

Definition 8 (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 $n > k > 0$. 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 with maximal δ-correlated-agree-domains $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C} = \{D_1, \ldots, D_m\}$. Define the* bad combining points *of (π_1, π_2) (with respect to the linear code C) to be*

$$\begin{aligned}\mathrm{Bad}_\delta(\pi_1,\pi_2) \triangleq \{\alpha \in \mathbb{F}_q \mid &\exists c \in C, \text{ such that } \delta(\pi_1+\alpha\pi_2, c) \leq \delta \\ &\text{and } \mathrm{Agree}(\pi_1+\alpha\pi_2, c) \notin \mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}\}.\end{aligned}$$

Notice that for any $D \in \mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}, \alpha \in \mathbb{F}_q$, there always exists a $c \in C$ that $(\pi_1 + \alpha\pi_2)[D] = c[D]$ according to the definition of correlated agreement and linear codes. Thus the bad combining points either *construct* a new δ-agree-domain or *enlarge* some domain in $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C}$.

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

Theorem 3. *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C+\eta}$. Let $\pi_1, \pi_2 \in \mathbb{F}_q^n$ be two given codewords, we have*

$$|\mathrm{Bad}_\delta(\pi_1,\pi_2)| < \frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}.$$

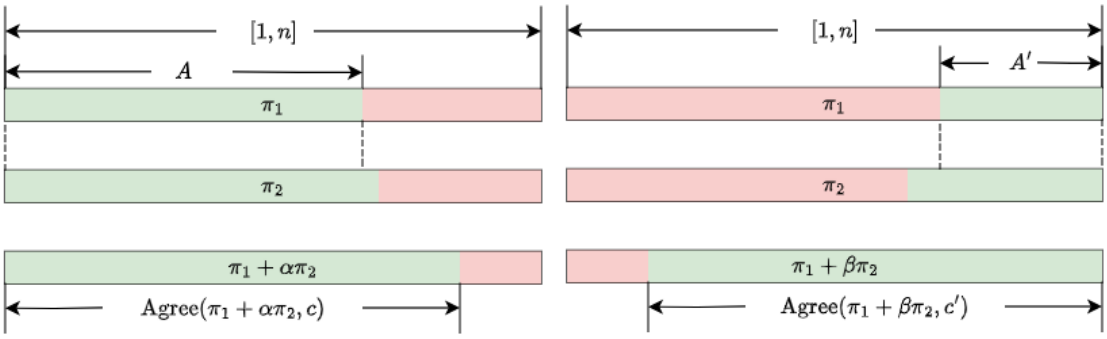
To count the number of bad combining points, we partition $\mathrm{Bad}_\delta(\pi_1, \pi_2)$ into two parts. The first part contains the bad combining points whose combining result is combined from some low-error codeword; the second part considers the high-error case. We use the Johnson bound to define the "low-error" and "high-error" cases. More precisely, let $\delta_0 = 1 - \sqrt{1-\Delta_C+\eta}$[1] and $\mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}$ be all the δ_0 correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C, notice that $\mathcal{A}_{\delta,\{\pi_1,\pi_2\},C} \subseteq \mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}$. Figure 3 provides an example to explain the partition.

Definition 9.

$$\begin{aligned}&\textit{(Low-error case)}\ \mathrm{Bad}^1_{\delta,\delta_0}(\pi_1,\pi_2) \triangleq \{\alpha \in \mathrm{Bad}_\delta(\pi_1,\pi_2) \mid \exists c \in C, A \in \mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C} \text{ such that} \\ &\qquad\qquad \delta(\pi_1+\alpha\pi_2) \leq \delta \text{ and } A \subsetneq \mathrm{Agree}(\pi_1+\alpha\pi_2, c)\}; \\ &\textit{(High-error case)}\ \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2) \triangleq \mathrm{Bad}_\delta(\pi_1,\pi_2) \setminus \mathrm{Bad}^1_{\delta,\delta_0}(\pi_1,\pi_2).\end{aligned}$$

We put upper bounds on these two sets respectively.

[1] We point out that the choice of δ_0 is not optimal.



(a) Low-error case. (b) High-error case.

Figure 3: Let $C \subseteq \mathbb{F}_q^n$ be a linear code with minimal relative distance Δ_C. $\pi_1, \pi_2 \subseteq \mathbb{F}_q^n$ are two codewords. The green part stands for an agree-domain, while the red is the disagree-domain. (a): For an $\alpha \in \mathrm{Bad}^1_{\delta,\delta_0}(\pi_1, \pi_2)$, there exists a code word $c \in C$ such that $|\mathrm{Agree}(\pi_1 + \alpha\pi_2, c)| \geq (1-\delta)n$. Furthermore, there is a "big" maximal correlated-agree-domain A contained in $\mathrm{Agree}(\pi_1 + \alpha\pi_2, c)$, i.e., $(1-\delta_0)n \leq |A| < |\mathrm{Agree}(\pi_1 + \alpha\pi_2, c)|$. (b): For an $\beta \in \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1, \pi_2)$, there also exists a code word $c' \in C$ such that $|\mathrm{Agree}(\pi_1 + \beta\pi_2, c')| \geq (1-\delta)n$. However, even the "biggest" correlated-agree-domain A' contained in $\mathrm{Agree}(\pi_1 + \beta\pi_2, c')$ is "small", i.e., $|A'| < (1-\delta_0)n$.

4.1 Bound the low-error case

This section bounds the number of bad combining points in $\mathrm{Bad}^1_{\delta,\delta_0}(\pi_1, \pi_2)$. We have:

Lemma 5.

$$|\mathrm{Bad}^1_{\delta,\delta_0}(\pi_1, \pi_2)| \leq \frac{n}{\eta}.$$

To prove the above lemma, we further partition $\mathrm{Bad}^1_{\delta,\delta_0}(\pi_1, \pi_2)$ into groups related to $A \in \mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}$. More precisely, denote by

$$\mathrm{Bad}_A \triangleq \{\alpha \in \mathrm{Bad}^1_{\delta,\delta_0}(\pi_1, \pi_2) \mid \exists c \in C \text{ such that } \delta(\pi_1 + \alpha\pi_2, c) \leq \delta \text{ and } A \subsetneq \mathrm{Agree}(\pi_1 + \alpha\pi_2, c)\}$$

the set of bad combining points corresponding to A. The number of $\mathrm{Bad}_A, A \in \mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}$ is bounded by $|\mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}|$. We prove:

Lemma 6.

$$|\mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C}| \leq \frac{1}{\eta}.$$

Proof. Let $\mathcal{A}_{\delta_0,\{\pi_1,\pi_2\},C} = \{A_1, \ldots, A_N\}$. We have

1. $|A_i| \geq (1-\delta_0)n = \sqrt{1-\Delta_C+\eta} \cdot n$ for all $1 \leq i \leq N$;
2. $|A_i \cap A_j| \leq (1-\Delta_C)n$ for $1 \leq i < j \leq N$.

The first item is directly from Definition 6. We prove the second item. For distinct $1 \leq i, j \leq N$, let $c_i^0, c_i^1 \in C$ be the codewords such that

$$c_i^0[A_i] = \pi_0[A_i], c_i^1[A_i] = \pi_1[A_i], c_i^0[A_j] = \pi_1[A_j], \text{ and } c_j^1[A_j] = \pi_1[A_j].$$

Thus,

$$c_i^0[A_i \cap A_j] = \pi_0[A_i \cap A_j] = c_j^0[A_i \cap A_j] \text{ and } c_i^1[A_i \cap A_j] = \pi_1[A_i \cap A_j] = c_j^1[A_i \cap A_j].$$

Suppose $|A_i \cap A_j| > (1 - \Delta_C)n$, then we have $c_i^k = c_j^k, k = 0, 1$ according to the definition of minimum relative distance. This implies $A_i = A_j$, a contradiction.

We follow the proof of Corrádi's lemma ([Juk11]), counting the elements in A_i to bound N. For any $x \in \bigcup_{i=1}^N A_i$, denote by $d(x)$ the *count* of x, i.e., the number of A_i containing x. We have

$$\sum_{x \in A_i} d(x) = |A_i| + \sum_{j \neq i} |A_i \cap A_j| \leq |A_i| + (N-1)(1 - \Delta_C)n. \tag{7}$$

Summing over all sets B_{α_i}, we get

$$\begin{aligned}
\sum_{i=1}^N \sum_{x \in A_i} d(x) &= \sum_{x \in \bigcup_{i=1}^N A_i} d(x)^2 \\
&\geq \frac{1}{\left|\bigcup_{i=1}^N A_i\right|} \left(\sum_{x \in \bigcup_{i=1}^N A_i} d(x) \right)^2 \qquad \text{by Cauchy–Schwarz inequality} \\
&= \frac{1}{\left|\bigcup_{i=1}^N A_i\right|} \left(\sum_{i=1}^N |A_i| \right)^2 . \qquad (8)
\end{aligned}$$

Combining (7) and (8), we have (8), we have

$$\frac{1}{\left|\bigcup_{i=1}^N A_i\right|} \left(\sum_{i=1}^N |A_i| \right)^2 \leq \sum_{i=1}^N |A_i| + N(N-1)(1 - \Delta_C)n.$$

Thus,

$$\begin{aligned}
&\frac{1}{\left|\bigcup_{i=1}^N A_i\right|} \left(\sqrt{1 - \Delta_C + \eta} Nn \right)^2 \leq Nn + N^2(1 - \Delta_C)n \\
\Rightarrow N \leq &\frac{1}{\frac{n}{|\bigcup_{i=1}^N A_i|}(1 - \Delta_C + \eta) - (1 - \Delta_C)} \leq \frac{1}{\eta}.
\end{aligned}$$

□

For any $A \in \mathcal{A}_{\delta_0, \{\pi_1, \pi_2\}, C}$, we bound the number of elements in Bad_A.

Corollary 2. *For any $A \in \mathcal{A}_{\delta_0, \{\pi_1, \pi_2\}, C}$, we have $|\mathrm{Bad}_A| \leq n$.*

Proof. Notice that A is a maximal correlated-agree-domain between $\{\pi_1, \pi_2\}$ and C with $|A| \geq (1 - \delta_0)n > (1 - \Delta_C)n$. Lemma 4 leads to the result. □

Combining Lemma 6 and Corollary 2, we have proved Lemma 5.

4.2 Bound the high-error case

This subsection bounds the number of combining points that combine the codewords with high error into a codeword with low error, i.e., we bound $|\mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)|$.

[Ben+20a] provides an analysis on the general distance preservation case:

Lemma 7 (Lemma 3.1 in [Ben+20a])**.** *Let $\eta < 1/3$ and $\delta < 1 - \sqrt[3]{1-\Delta_C+\eta}$. Suppose $\pi_0 \in \mathbb{F}_q^n$ is such that $\delta(\pi_0, C) > \delta + \eta$. Then for all $\pi_1 \in \mathbb{F}_q^n$, there are at most $2/\eta^2$ values of $\alpha \in \mathbb{F}_q$ such that $\delta(\pi_0 + \alpha\pi_1) < \delta$.*

We follow the approach of [Ben+20a] to bound $|\mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)|$.

Lemma 8.

$$|\mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)| \le \max\left\{\frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta}-\sqrt{1-\Delta_C+\eta})}, \frac{6}{\eta}\right\}.$$

Proof. We prove by contradiction. Suppose we have

$$|\mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)| > \max\left\{\frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta}-\sqrt{1-\Delta_C+\eta})}, \frac{6}{\eta}\right\}.$$

For any $\alpha \in \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)$, let B_α be $\mathrm{Agree}(\pi_1+\alpha\pi_2, c)$ for some $c \in C$ such that $|B_\alpha| \ge (1-\delta)n$. For α,β,γ picked uniformly from $\mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)$ and x picked uniformly from $[1,n]$, we have:

$$\begin{aligned}
\mathrm{E}_{\alpha,\beta,\gamma}[|B_\alpha \cap B_\beta \cap B_\gamma|/n] &= \mathrm{E}_x, \alpha, \beta, \gamma[1_{x\in B_\alpha \cap B_\beta \cap B_\gamma}] \\
&= \mathrm{E}_x[\mathrm{E}_\alpha[1_{x\in B_\alpha}]^3] \\
&\ge \mathrm{E}_{x,\alpha}[1_{x\in B_\alpha}]^3 \qquad \text{by Jensen's inequality} \\
&\ge (1-\delta)^3 \ge 1-\Delta_C+\eta.
\end{aligned}$$

This implies $\Pr_{\alpha,\beta,\gamma}(|B_\alpha \cap B_\beta \cap B_\gamma| > (1-\Delta_C)n) \ge \eta$. Since

$$\Pr_{\alpha,\beta,\gamma}(\alpha,\beta,\gamma \text{ are not all distinct}) < 3/|\mathrm{Bad}_\delta(\pi_1,\pi_2)| < 2/\eta,$$

we have

$$\Pr_{\alpha,\beta,\gamma}(|B_\alpha \cap B_\beta \cap B_\gamma| > (1-\Delta_C)n \text{ and } \alpha,\beta,\gamma \text{ are all distinct}) \ge \frac{2}{\eta}.$$

Then there exist distinct $\alpha_0, \beta_0 \in \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)$ such that

$$\Pr_\gamma(|B_{\alpha_0} \cap B_{\beta_0} \cap B_\gamma| > (1-\Delta_C)n) \ge \frac{2}{\eta}. \tag{9}$$

By Corollary 1, $B_{\alpha_0} \cap B_{\beta_0}$ is a maximal correlated-agree-domain between π_1, π_2 and C. According to Lemma 2, we have $B_{\alpha_0} \cap B_{\beta_0} \subseteq B_\gamma$ for all the γ that $|B_{\alpha_0} \cap B_{\beta_0} \cap B_\gamma| > (1-\Delta_C)n$. (9) provides a lower bound on the number of B_γ containing $B_{\alpha_0} \cap B_{\beta_0}$, i.e.,

$$|\{\gamma : B_{\alpha_0} \cap B_{\beta_0} \subseteq B_\gamma\}| \ge \frac{1}{\sqrt[3]{1-\Delta_C+\eta}-\sqrt{1-\Delta_C+\eta}}.$$

Since $\alpha_0, \beta_0 \in \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2)$, we have $|B_{\alpha_0} \cap B_{\beta_0}| < (1-\delta_0)n = \sqrt{1-\Delta_C+\eta}\cdot n$. On the other hand, $\gamma \in \mathrm{Bad}^2_{\delta,\delta_0}(\pi_1,\pi_2) \subseteq \mathrm{Bad}_\delta(\pi_1,\pi_2)$ implies $|B_\gamma| \ge (1-\delta)n \ge \sqrt[3]{1-\Delta_C+\eta}\cdot n$. Modifying (6) in the proof of Lemma 4, we have

$$|\{\gamma : B_{\alpha_0} \cap B_{\beta_0} \subseteq B_\gamma\}| \le \frac{n - |B_{\alpha_0} \cap B_{\beta_0}|}{(\sqrt[3]{1-\Delta_C+\eta}-\sqrt{1-\Delta_C+\eta})n} < \frac{1}{\sqrt[3]{1-\Delta_C+\eta}-\sqrt{1-\Delta_C+\eta}},$$

a contradiction. □

5 Bad combining over affine spaces

In real-world applications, we always need to combine more than two codewords and want to preserve the distance. This section generalizes our main result, Theorem 3, to a batch of codewords $\pi_0, \pi_1, \ldots, \pi_l$ with $l \geq 1$.

We define bad combining points for a batch of codewords:

Definition 10 (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 $n > k > 0$. Denote by Δ_C the minimum relative distance of C. Let $\delta \in (0, 1-\Delta_C)$. Let $\pi_0, \ldots, \pi_l \in \mathbb{F}_q^n$ be codewords with maximal δ-correlated-agree-domains $\mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C} = \{D_1, \ldots, D_m\}$. Define the* bad combining points *of $(\pi_0, \ldots, \pi_l)$ (with respect to the linear code C) to be*

$$\mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l) \triangleq \{\langle \alpha_1, \ldots, \alpha_l\rangle \in \mathbb{F}_q^l \mid \exists c \in C, \text{ such that } \delta(\pi_0 + \cdots + \alpha_l\pi_l, c) \leq \delta \text{ and } \mathrm{Agree}(\pi_0 + \cdots + \alpha_l\pi_l, c) \notin \mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C}\}.$$

Theorem 4. *Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{1-\Delta_C+\eta}$. 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*

$$\Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)) < \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}\right) \cdot \frac{l}{q}.$$

Proof. We use induction to prove the result. When $l = 1$, the problem is reduced to Theorem 3. Suppose the result holds for $l-1$ codewords. For convenience, denote $\langle \alpha_1, \ldots, \alpha_i\rangle \in \mathbb{F}_q^i$ as $\mathbf{z}_i$, $i = 1, \ldots, l$. For $1 \leq i \leq l$, we have

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

Then we have

$$\begin{aligned}
&\Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)) \\
=&\Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l) \mid \mathbf{z}_{l-1} \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) \cdot \Pr_{\mathbf{z}_{l-1} \in \mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) \\
+&\Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l) \mid \mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) \cdot \Pr_{\mathbf{z}_{l-1} \in \mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) \\
\leq&\Pr_{\mathbf{z}_{l-1} \in \mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) + \Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l) \mid \mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})).
\end{aligned} \tag{10}$$

According to our assumption, the first item satisfies

$$\Pr_{\mathbf{z}_{l-1} \in \mathbb{F}_q^{l-1}}(\mathbf{z}_{l-1} \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})) < \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}\right) \cdot \frac{l-1}{q}. \tag{11}$$

For the second item, for any fixed $\mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})$, define the set

$$\mathrm{Bad}_{\mathbf{z}_{l-1}} \triangleq \{\alpha_l \in \mathbb{F}_q : \langle \mathbf{z}_{l-1}, \alpha_l\rangle \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)\}.$$

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

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

We want to prove

$$\mathrm{Bad}_{\mathbf{z}_{l-1}} \subseteq \mathrm{Bad}_\delta(\pi', \pi_l). \tag{12}$$

Notice that $|\mathrm{Bad}_\delta(\pi', \pi_l)| < \frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}$ by Theorem 3. So if (12) holds, we have

$$|\mathrm{Bad}_{\mathbf{z}_{l-1}}| < \frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}. \tag{13}$$

Suppose $\mathcal{A}_{\delta,\{\pi',\pi_l\},C} = \{D'_1, \ldots, D'_{m'}\}$. Since $\mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})$, $\{D'_1, \ldots, D'_{m'}\}$ are δ-correlated-agree-domains of $\pi_0, \cdots, \pi_{l-1}$ (may be not maximal). Otherwise, suppose, without loss of generality, D'_1 is not a δ-correlated-agree-domain between $\{\pi_0, \ldots, \pi_{l-1}\}$ and C, then $\mathbf{z}_{l-1}$ expands the correlated-agree-domain. This is a contradiction to $\mathbf{z}_{l-1} \notin \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_{l-1})$.

Thus, $\{D'_1, \ldots, D'_{m'}\}$ are δ-correlated-agree-domains of $\pi_0, \cdots, \pi_l$ according to the definition of $\mathcal{A}_{\delta,\{\pi',\pi_l\},C}$. As a result, for any $D' \in \mathcal{A}_{\delta,\{\pi',\pi_l\},C}$, $\exists D \in \mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C}$, such that $D' \subseteq D$. On the other hand, D' is maximal when considering π' and π_l according to the definition of $\mathcal{A}_{\delta,\{\pi',\pi_l\},C}$. Then we have $D' = D$ since D is a correlated-agreement between π', π_l and C. Thus,

$$\mathcal{A}_{\delta,\{\pi',\pi_l\},C} \subseteq \mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C}. \tag{14}$$

Fix an $\alpha \in \mathrm{Bad}_{\mathbf{z}_{l-1}}$, we have $\langle \mathbf{z}_{l-1}, \alpha\rangle \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)$. According to the definition of $\mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)$, $\exists c \in C$ such that

$$\mathrm{Agree}(\pi' + \alpha\pi_l, c) \notin \mathcal{A}_{\delta,\{\pi_0,\ldots,\pi_l\},C} \text{ and } |\mathrm{Agree}(\pi' + \alpha\pi_l, c)| \ge (1-\delta)n.$$

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

$$\alpha \in \mathrm{Bad}_\delta(\pi', \pi_l)$$

and we have proved (12).

Plugging (11) and (13) into (10), we have

$$\Pr_{\mathbf{z}_l \in \mathbb{F}_q^l}(\mathbf{z}_l \in \mathrm{Bad}_\delta(\pi_0, \ldots, \pi_l)) < \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}\right) \cdot \frac{l}{q}.$$

□

Corollary 3 (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 $\delta \le 1 - \sqrt[3]{1-\Delta_C+\eta}$. Let $\pi_0, \ldots, \pi_l \in \mathbb{F}_q^n$ be $l+1$ codewords, where $l \ge 1$. Suppose*

$$\Pr_{\langle\alpha_1,\ldots,\alpha_l\rangle \in \mathbb{F}_q^l}(\delta(\pi_0 + \ldots + \alpha_l\pi_l, C) \le \delta) \ge \left(\frac{n+6}{\eta} + \frac{2}{\eta(\sqrt[3]{1-\Delta_C+\eta} - \sqrt{1-\Delta_C+\eta})}\right) \cdot \frac{l}{q}.$$

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

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

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 and generalizations of FRI such as [ZCF23][Arn+24b][Arn+24a].

6.1 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 with length n. $0 < \rho < 1$ is the rate and $d = \lceil \rho n \rceil - 1$. Let $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho] : \mathbb{F}_q^{d+1} \to \mathbb{F}_q^n$ denote the Reed-Solomon code that is consisted of evaluations of polynomials with degree d on the evaluation domain $L^{(0)}$, i.e., $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$ maps coefficients $(c_0, c_1, \ldots, c_d) \in \mathbb{F}_q^{d+1}$ to $\left(\sum_{i=0}^{k} c_i x^i\right)_{x \in L^{(0)}} \in \mathbb{F}_q^n$.

For a given function $f^{(0)} : L^{(0)} \to \mathbb{F}_q$, the verifier wants to know whether it is a member of $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$. An untrusted prover may help the verifier while the verifier has query accesses to $f^{(0)}$. The crucial operation in FRI is called *folding*. For a randomly chosen folding parameter $\alpha \in \mathbb{F}_q$, the folding result of $f^{(0)}$ is a function $f^{(1)} : {L^{(0)}}^2 \to \mathbb{F}_q$, where ${L^{(0)}}^2 = \{x^2 \mid x \in L^{(0)}\}$. Denote by $\mathrm{Fold}_\alpha(f^{(0)})$ the folding result of $f^{(0)}$ at point α. When $f^{(0)}$ is a member of $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$, $\mathrm{Fold}_\alpha(f^{(0)})$ is a member of $\mathrm{RS}[\mathbb{F}_q, L^{(1)}, \rho]$ for any folding point $\alpha \in \mathbb{F}_q$.

More precisely, for any $x^2 \in {L^{(0)}}^2$, define

$$\mathrm{Fold}_\alpha(f^{(0)})(x^2) = \frac{f^{(0)}(x) + f^{(0)}(-x)}{2} + \alpha \cdot \frac{f^{(0)}(x) - f^{(0)}(-x)}{2x}.$$

When $f^{(0)} \in \mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$ and let

$$p^{(0)}(X) = c_0 + c_1 X + \ldots + c_d X^d = \sum_{\text{even } i} c_i X^i + X \sum_{\text{odd } i} c_i X^{i-1}$$

be the corresponding low-degree polynomial. Then $f^{(1)} = \mathrm{Fold}_\alpha(f^{(0)}) : {L^{(0)}}^2 \to \mathbb{F}_q$ is the evaluation of polynomial

$$p^{(1)}(X^2) = \sum_{\text{even } i} c_i X^i + \alpha \sum_{\text{odd } i} c_i X^{i-1}. \tag{15}$$

Notice that the degree of $p^{(1)}(X)$ is a half of $p^{(0)}(X)$. Denote by $f^{(i+1)} = \mathrm{Fold}_{\alpha_i}(f^{(i)})$ the folding result of $f^{(i)}$ at folding point α_i, $i \geq 0$. Applying the folding operation $\log d$ times, the folding result of the function (if it is a member of the RS code) is supposed to be a constant on the evaluation domain and can be checked easily.

The time complexity of folding a function of length n is $O(n)$. In the FRI protocol, the verifier will not do the folding operation. Instead, the prover calculates the folding result and sends an oracle to the verifier. The verifier randomly chooses some locations and checks whether the folding

is accurately calculated. Notice that there are $O(\log d) = O(\log \rho n) = O(\log n)$ rounds., The total running time of the verifier is reduced to $O(\log n)$.

Batching. In real-world applications, we want to verify whether a series of functions $f_0^{(0)}, \ldots, f_l^{(0)}$ are all members of $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$. 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 $f^{(0)} \triangleq f_0^{(0)} + z_1 f_1^{(0)} + \cdots + z_l f_l^{(0)}$. Besides checking the accuracy of folding, the verifier also needs to check that the combining is done correctly at selected locations.

6.2 Soundness of batched FRI

The soundness error of batched FRI consists of:

1. Bad combining: the choice of combining parameters is "bad" that the combining result is closer to $\mathrm{RS}[\mathbb{F}_q, L^{(0)}, \rho]$.
2. Bad folding (in the ith round): the choice of folding parameter α_i is "bad", i.e., the folding result $f^{(i+1)} = \mathrm{Fold}_{\alpha_i}(f^{(i)})$ is closer to (with respect to relative distance) $\mathrm{RS}[\mathbb{F}_q, L^{(i+1)}, \rho]$.
3. The prover's cheating: the prover can do the folding operation dishonestly or modify some location of the function directly.

In each round, the verifier will randomly check some locations of the function to prevent the prover from cheating (called the QUERY phase in FRI). The probability of bad combining is restricted by Theorem 4 directly. We propose an analysis of the possibility of bad folding based on Theorem 3.

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, $L^{(0)}, L^{(1)}, \ldots$ that $L^{(i+1)} = {L^{(i)}}^2$. Suppose there are n_r rounds in the FRI protocol. Let

$$\delta^{(k)} \triangleq \delta(f^{(k)}, \mathrm{RS}[\mathbb{F}_q, L^{(k)}, \rho]), 0 \le k \le n_r$$

be the relative distance.

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

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

where α_k is the random folding point chosen by the verifier in the kth round. By (15), the folding operation can be seen as the combination of two codewords. Thus, we have:

Lemma 9 (Bad folding)**.**

$$\Pr(E^{(k)}) < \left(\frac{|L^{(k+1)}| + 6}{\eta} + \frac{2}{\eta(\sqrt[3]{\rho + \eta} - \sqrt{\rho + \eta})} \right) \cdot \frac{1}{q}.$$

Proof. Construct two functions on $L^{(k+1)}$, for every $x^2 \in L^{(k+1)}$,

$$\pi_1(x^2) = \frac{f^{(k)}(x) + f^{(k)}(-x)}{2} \text{ and } \pi_2(x^2) = \frac{f^{(k)}(x) - f^{(k)}(-x)}{2x}.$$

Then we have $f^{(k)}(x) = \pi_1(x^2) + x\pi_2(x^2)$ for all $x \in L^{(k)}$.

We claim that $\delta_{\text{corr}}(\{\pi_1, \pi_2\}, \text{RS}[\mathbb{F}_q, L^{(k+1)}, \rho]) \geq \delta^{(k)}$. Otherwise, suppose we have $p_1(X), p_2(X) \in \mathbb{F}_q^{<\rho|L^{(k+1)}|}[X]$ and $D \subseteq L^{(k+1)}$ such that $|D| > (1-\delta^{(k)})|L^{(k+1}|$ and

$$p_1[D] = \pi_1[D] \text{ and } p_2[D] = \pi_2[D].$$

Let $\sqrt{D} \subseteq L^{(k)}$ be the set of square roots of elements in D, i.e., $\sqrt{D} = \{x \mid x^2 \in L^{(k+1)}\}$. Then we have $|\sqrt{D}| = 2|D|$. Define $p(X) = p_1(X^2) + Xp_2(X^2)$ to be a polynomial of degree $\leq 2\rho(|L^{(k+1)}| - 1) + 1 = \rho|L^{(k)}| - 1$. For any $x \in \sqrt{D}$, we have

$$f^{(k)}(x) = \pi_1(x^2) + x\pi_2(x^2) = p_1(x^2) + xp_2(x^2) = p(x).$$

Thus, $\delta(f^{(k)}, \text{RS}[\mathbb{F}_q, L^{(k)}, \rho]) \leq \delta(f^{(k)}, p[L^{(k)}]) < \delta^{(k)}$. A contradiction to the definition of $\delta^{(k)}$.

We have

$$\Pr(E^{(k)}) = \Pr_{\alpha_k \in \mathbb{F}_q}(\delta(\pi_1 + \alpha\pi_2, \text{RS}[\mathbb{F}_q, L^{(k+1)}, \rho]) < \delta^{(k)}) < \left(\frac{|L^{(k+1)}| + 6}{\eta} + \frac{2}{\eta(\sqrt[3]{\rho+\eta} - \sqrt{\rho+\eta})}\right) \cdot \frac{1}{q}.$$

The inequality is by Theorem 3, where the set $\mathcal{A}_{\delta^{(k)} - \frac{1}{|L^{(k+1)}|}, \{\pi_1, \pi_2\}, \text{RS}[\mathbb{F}_q, L^{(k+1)}, \rho]}$ is empty. □

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:

$$\begin{aligned}\Pr(\bigwedge_{k=0}^{n_r-1} \neg E^{(k)}) &\geq 1 - \sum_{k=0}^{n_r-1} \Pr(E^{(k)}) \\ &> 1 - \left(\frac{|L^{(0)}| + 6n_r}{\eta} + \frac{2n_r}{\eta(\sqrt[3]{\rho+\eta} - \sqrt{\rho+\eta})}\right) \cdot \frac{1}{q}.\end{aligned}$$

In application, we have $n_r \leq \log n$. Consequently, we derive the following soundness error bound for batched FRI. Further details regarding the soundness of FRI can be found in [Ben+18][Ben+20b], which offer a comprehensive soundness analysis.

Theorem 5 (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 $V^{(0)} = \text{RS}[\mathbb{F}_q, L^{(0)}, \rho]$ and ρ satisfies $\rho = 2^{-R}$ for a positive integer R. Let $\delta, \eta > 0$ satisfy $\delta \leq 1 - \sqrt[3]{\rho+\eta}$. 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

$$\epsilon_{\textit{Batched-FRI}} = \left(\frac{|L^{(0)}|(l+1) + 6(l + \log n)}{\eta} + \frac{2(l + \log n)}{\eta(\sqrt[3]{\rho+\eta} - \sqrt{\rho+\eta})}\right) \cdot \frac{1}{q} + (1-\delta)^t \tag{16}$$

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 $1 - \delta$.

6.3 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

$$\epsilon_{\text{Batched-FRI}} = \epsilon_c + \left(\sqrt{\rho}(1+\frac{1}{2m})\right)^t. \tag{17}$$

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 (16). And we have

$$2^{-149} < \left(\frac{|L^{(0)}|(l+1)+6(l+\log n)}{\eta} + \frac{2(l+\log n)}{\eta(\sqrt[3]{\rho+\eta}-\sqrt{\rho+\eta})}\right)\cdot\frac{1}{q} < 2^{-148}.$$

We prove FRI can reach 128 bits of security in the current field when $t \geq 136$ and can reach 148 bits of security when $t \geq 157$.

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.

[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.