Locality of Curve-Decoding and Improved Proximity Gaps

Rohan Goyal
Massachusetts Institute of Technology, Cambridge, MA, USA

Venkatesan Guruswami
University of California, Berkeley, CA, USA

Yihang Sun
Stanford University, CA, USA

Mary Wootters
Stanford University, CA, USA

Abstract

Proximity gaps are a property of error correcting codes that arise in the study of Interactive Oracle Proofs (IOPs) and Succinct Non-interactive Arguments of Zero Knowledge (SNARKs). Informally, we say that a code $C \subset \Sigma^n$ exhibits a *proximity gap* (with respect to degree-ℓ curves) if for any degree-ℓ curve $u(x) \in \Sigma^n$, either *every* point on $u(x)$ is close to C, or else *most* of them are far from C.

Recent work [16] has established near-optimal proximity gaps for many families of codes, including subspace design codes, as well as random ensembles like random linear codes, Reed-Solomon codes with random evaluation points, and Gallager's ensemble of LDPC codes. However, the parameters for these latter randomized ensembles are worse than the parameters for subspace design codes, and degrade as the degree ℓ increases.

In this work, we obtain improved proximity gaps for random ensembles of codes, including random linear codes, Reed-Solomon codes with random evaluation points, and Gallager's ensemble. Quantitatively, our results for these random ensembles match the results that [16] attained for subspace design codes. In fact, our techniques are a black-box transference from subspace design codes: Any progress on subspace design codes will automatically lead to analogous progress for these random ensembles.

To obtain our results, we extend the Local Coordinate-wise Linear (LCL) property framework developed in [9, 27] to a *row-span constrained* version. This allows us to cast *curve-decodability* – a property that implies proximity gaps – directly as an (row-span constrained) LCL property, and make use of that machinery. In contrast, because curve-decodability is not obviously a (vanilla) LCL property, prior work had worked with a proxy property instead, leading to the aforementioned parameter losses. In addition, we extend the framework to also show an equivalence theorem for Gallager's ensemble of random LDPC codes and random linear codes for our row-span constrained LCL properties.

2012 ACM Subject Classification Theory of computation → Error-correcting codes; Theory of computation → Interactive proof systems

Keywords and phrases Proximity gaps, random codes, curve decoding, local properties

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.61

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2607.08516` [18]

Funding *Rohan Goyal*: R.G. is supported by (Yael Tauman Kalai's) grant from Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-25-C-0300. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA).

Venkatesan Guruswami: V.G. is supported by a Simons Investigator award, NSF grant CCF-2211972, and ONR grant N00014-24-1-2491.
Yihang Sun: Y.S. is funded by the NSF Graduate Research Fellowship and the Stanford Graduate Fellowship.
Mary Wootters: M.W. is partially funded by NSF grants CCF-2231157 and CNS-2321489.

Acknowledgements The authors would like to thank Dan Boneh for helpful conversations.

1 Introduction

Proximity gaps are a property of error correcting codes that arise in the study of Interactive Oracle Proofs (IOPs) and Succinct Non-interactive Arguments of Zero Knowledge (SNARKs), both of which are relevant for blockchains and related technologies. Let $\mathbb{F}$ be a finite field, and let Σ be a vector space over $\mathbb{F}$. An $\mathbb{F}$-*additive code* C (with alphabet Σ and block length n) is an $\mathbb{F}$-linear subspace $C \subset \Sigma^n$. Informally, we say that C exhibits a *proximity gap* (for degree-ℓ curves, which is the focus of this paper) if the following holds. For any degree-ℓ curve $u(x) = \sum_{i=0}^{\ell} u_i x^i$ in Σ^n given by $u_0, \ldots, u_\ell \in \Sigma^n$ and $x \in \mathbb{F}$, either *every* point $u(\alpha)$ on the curve is close to C, or else a $1-\varepsilon$ fraction of the points on the curve are far from C. Formally, we have the following definition. Below, $\Delta(x,y) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}[x_i \neq y_i]$ is the (relative) Hamming distance between $x, y \in \Sigma^n$, and $\Delta(x, C) := \min_{y \in C} \Delta(x,y)$ for code C.

▶ **Definition 1** (Proximity Gaps (PG)). *An $\mathbb{F}$-additive code $C \subset \Sigma^n$ is said to have $(\ell, \delta, \varepsilon, \gamma)$ proximity gap if for all $u_0, u_1, \cdots, u_\ell \in \Sigma^n$, and for all $\delta' \leq \delta$,*

$$\mathbb{P}_{\alpha \in \mathbb{F}}\left(\Delta\left(\sum_{j=0}^{\ell} u_j \alpha^j, C\right) \leq \delta'\right) > \varepsilon \implies \forall \alpha \in \mathbb{F},\ \Delta\left(\sum_{j=0}^{\ell} u_j \alpha^j, C\right) \leq \frac{\delta'}{1-\gamma}\ .$$

When $\gamma = 0$, we say that C has an $(\ell, \delta, \varepsilon)$ proximity gap.

Proximity gaps, especially for *Reed-Solomon (RS) Codes*,[1] have been a hot topic lately, with the Ethereum Foundation offering $1M in prizes for resolving (or disproving) conjectures related to proximity gaps for RS codes [14]. See [2] for a recent survey on proximity gaps and problems relevant to the proximity prize.

Until recently, the best proximity gap guarantees required $\delta < 1 - \sqrt{R}$, where $R = \frac{\log_{|\Sigma|} |C|}{n}$ is the *rate* of the code $C \subset \Sigma^n$. In recent work [16], Guruswami and Goyal established near-optimal proximity gaps – with δ approaching $1 - R$ – for several families of codes. For *Folded Reed-Solomon Codes, Multiplicity Codes*,[2] and more generally any *Subspace Design Code* (Definition 15), they established $(\ell, \delta, \varepsilon)$ proximity gaps with $\delta = 1 - R - 2\eta$ for any $\eta > 0$, provided that

$$\varepsilon q \geq \frac{n\ell}{\eta} + O\left(\frac{\ell^2}{\eta^3}\right), \tag{1}$$

[1] Reed-Solomon codes are a classical family of codes based on low-degree polynomials. Given fixed distinct evaluation points $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q =: \Sigma$, the corresponding Reed-Solomon code $C \subset \Sigma^n$ of length n and dimension $k \leq n$ is given by $C = \{(f(\alpha_1), \ldots, f(\alpha_n)) : f \in \mathbb{F}_q[x], \deg(f) < k\}$.

[2] Informally, a *folded Reed-Solomon code* is obtained by "folding" (aka, bundling symbols together) an RS code $C \in \mathbb{F}_q^n$ to obtain a code $C' \in (\mathbb{F}_q^m)^{n/m}$. A (univariate) *multiplicity code* is similar to an RS code, except each symbol contains not just $f(\alpha_i)$, but also $f^{(j)}(\alpha_i)$ for $j = 1, 2, \ldots, m-1$, where $f^{(j)}$ denotes the j'th (Hasse) derivative. In either case, the alphabet size increases from q to q^m. By a result of [20], both are special cases of Subspace Design Codes, defined in Definition 15.

where q is the size of the base field. Thus, the left-hand side εq is a bound on the number of points on the curve $u(x)$ that can still be close to C, in the case that not all q points are. We'd like for this bound to be as small as possible.

Via a framework from [9, 27] that establishes a connection between random linear codes and subspace design codes, the work [16] extended their results for subspace design codes to several random ensembles of codes, including random linear codes, Reed-Solomon codes with random evaluation points, and Gallager's ensemble of random LDPC codes. However, some parameters degraded in this extension: The final result that [16] established for these random ensembles was $(\ell, \delta, \varepsilon)$ proximity gaps with $\delta = 1 - R - 2\eta$, provided that

$$\varepsilon q \geq \frac{n\ell(1-R)}{\eta} + \left(\frac{\ell}{\eta^2}\right)^{O(\ell)}. \tag{2}$$

When ℓ is large, on the order of $\Omega(\log n)$, the error term $(\ell/\eta^2)^{O(\ell)}$ in (2) swamps the main term, and the requirement on εq quickly becomes large. This leaves us with the question of whether a bound like (1) is achievable for these random ensembles of codes. This question has practical as well as theoretical motivations: Larger values of ℓ can occur in practice, for example in WHIR [3]; random ensembles of codes are also practically interesting, for example in Blaze [10] or Bolt [19];[3] and understanding RS codes with random evaluation points may deepen our understanding of explicit RS codes, relevant to the proximity prize [14].

1.1 Our Contributions

Our main contribution is to improve the bound (2) for random ensembles of codes to match the bound (1) that was already established for subspace design codes. We establish these improved bounds for random linear codes, RS codes with random evaluation points, and Gallager's ensemble of LDPC codes. See Definitions 5–7 for formal definitions of these codes.

▶ **Theorem 2** (Informal). *Fix any $\ell \in \mathbb{N}$ and $\eta > 0$. Let $C \in \mathbb{F}_q^n$ be a random linear code, random Reed-Solomon code, or random LDPC code*[4] *Then, with probability at least $2/3$ over the choice of C, C has $(\ell, 1 - R - 2\eta, \varepsilon)$ proximity gaps (with respect to degree-ℓ curves), provided that*

$$\varepsilon q \geq \frac{n\ell(1-R)}{\eta} + O\left(\frac{\ell^2}{\eta^3}\right). \tag{3}$$

In fact, our results establish not just proximity gaps, but also stronger notions known as *Correlated Agreement* (CA, Definition 8) and *Mutual Correlated Agreement* (MCA, Definition 9); see the full version for the full statements [18].

Our framework, discussed more below in Section 1.2, is a black-box transference method from subspace design codes. Thus, any future improvements for subspace design codes will immediately lead to an analogous improvement in Theorem 2.

In order to establish our results, we generalize the *Local Coordinate-wise Linear* (LCL) framework of [27] to a new "row-constrained" version of the framework, described below. We hope that this generalization may find other applications going forward.

[3] In more detail, Blaze considers a random ensemble of *RAA* codes, which use random permutations to decide the parity checks. Bolt uses *Gallager's ensemble* of LDPC codes, which we also study in this paper.

[4] Technically, the result for random LDPC codes requires the leading term $n\ell(1-R)/\eta$ to be slightly larger; see full version for details.

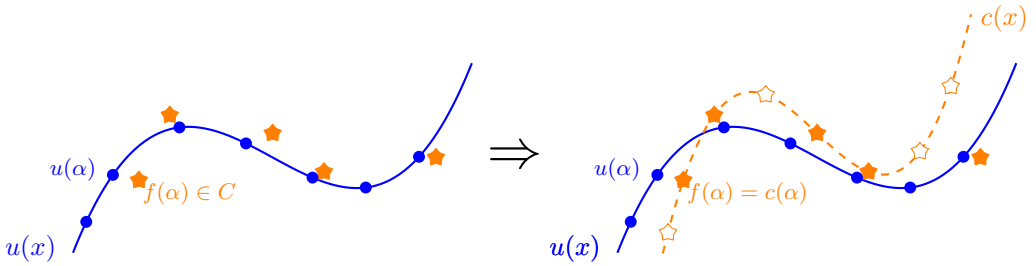

Figure 1 Depiction of curve decodability (Definition 10). The orange stars represent $f(\alpha) \in C$ for $\alpha \in A$, and the blue dots represent $u(\alpha) \in \Sigma^n$ for $\alpha \in \mathbb{F}_q$, where $u(x)$ is a degree-ℓ curve. Informally, a code $C \subseteq \Sigma^n$ is (ℓ, δ, a, b)-curve-decodable if the following holds: If there are at least a orange stars δ-close to distinct blue dots, then at least b of those orange stars lie on a degree-ℓ curve $c(x)$. In the picture, $a = 5$, and $b = 4$.

▶ Remark 3. [27] showed an equivalence for LCL properties between random linear and random Reed-Solomon codes; we state its row-span constrained extension here. The analogous statement for Gallager's ensemble, based on [28], is recorded in Section 3.3.3; the proofs and LDPC-specific details appear in the full version [18].

1.2 Techniques: Curve-Decoding as a Row-Span Constrained LCL Property

Following [16], we study *curve-decoding*. For the rest of this discussion, assume for simplicity that the alphabet Σ is equal to $\mathbb{F}_q$, so C is a linear subspace of $\mathbb{F}_q^n$.

Informally, such a code C is (ℓ, δ, a, b) *curve-decodable* if, for any degree-ℓ curve $u(x) = \sum_{i=0}^{\ell} u_i x^i$ with $u_i \in \Sigma^n$, the following holds. Suppose that there is a set $A \subset \mathbb{F}_q$ of size a so that $u(\alpha)$ is δ-close to some codeword $f(\alpha) \in C$ for all $\alpha \in A$. Then, there should be a "codeword curve" $c(x) = \sum_{i=0}^{\ell} c_i x^i$, where $c_i \in C \subset \Sigma^n$, so that $f(\alpha) = c(\alpha)$ for at least b values of $\alpha \in A$. See Definition 10 for a formal definition; the property is illustrated in Figure 1.

The work [16] showed that curve-decodability is a sufficient condition to establish proximity gaps (as well as stronger notions like correlated agreement and mutual correlated agreement); see Theorem 12 and Theorem 13. They then established that subspace design codes have curve-decodability (Theorem 18), which led to their results about subspace design codes.

Our work comes in the next step. To obtain their results for random ensembles of codes, [16] applied a framework of [9, 27]. This framework roughly shows that any *Local Coordinate-wise Linear (LCL)* property that is satisfied by subspace design codes is also satisfied by these random ensembles. We will explain what LCL properties are in a moment, but the main challenge with this approach is that "being curve-decodable" is in fact *not* (or, not obviously) an LCL property. To get around this, [16] defined another property, which they called $\bigvee$*-decoding*, which *is* captured by the LCL framework. Then existing results imply that the random ensembles of codes are just as $\bigvee$-decodable as subspace design codes, and translating back to curve-decoding gives the final result. Unfortunately, going through $\bigvee$-decoding results in a loss in parameters, degrading (1) to (2). We avoid this loss by *extending* the LCL framework to apply directly to curve-decoding.

The LCL Framework. Before explaining our extension, we first explain the LCL framework of [27]. Intuitively, an r-local LCL property is one defined by the exclusion of r-sized bad sets of codewords. For example, the property of "having good distance" might be defined by the

exclusion of all pairs ($r = 2$) of codewords that agree in too many places. The LCL framework allows for any definition of "bad" that can be captured by a family $\mathcal{F}$ of coordinate-wise linear relationships.

In more detail, let $C \subset \mathbb{F}_q^n$ be a linear code, and consider a set of r distinct "bad" codewords, stacked together to form the columns of a matrix $M \in \mathbb{F}_q^{n \times r}$. Let $\mathcal{F}$ be a family of *profiles* $\mathbf{V} = (V_1, \ldots, V_n)$, where each $V_i \in \mathcal{L}(\mathbb{F}_q^r)$. (Here, $\mathcal{L}(W)$ denotes the collection of subspaces of W.) In the LCL framework, the "badness" of a matrix M is captured by the following criterion: There is some $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{F}$ so that for all $i \in [n]$, the i'th row of M is contained in V_i.

Thus, an LCL property defined by $\mathcal{F}$ is the property of *avoiding* all "bad" matrices M that satisfy the constraints given by some $\mathbf{V} \in \mathcal{F}$ (we say that C *avoids* M if the column span of M is not contained in C). See Definition 20 for a formal definition.

Capturing Curve-Decoding and the Approach of [16]. The LCL framework captures many familiar properties, for example distance, list-decoding, and list-recovery. Unfortunately, it does not seem to capture curve-decoding. To see why, let us consider two approaches.

For the first – and probably most natural – approach, suppose that we have a bad example for (ℓ, δ, a, b) curve-decoding. That is, suppose we have a set $A \subset \mathbb{F}_q$ of size a, a degree-ℓ curve $u(x) \in \mathbb{F}_q^n$, and a function $f : \mathbb{F}_q \to C$ so that $\Delta(u(\alpha), f(\alpha)) \leq \delta$ for all $\alpha \in A$; but suppose that there is *no* degree-ℓ "codeword curve" $c(x) = \sum_{i=0}^{\ell} c_i x^i$ for $c_i \in C$ that passes through at least b of the $f(\alpha)$.

We will attempt to construct a "witness" matrix $M \in \mathbb{F}_q^{n \times A}$ whose columns are $f(\alpha)$ for all $\alpha \in A$. In order to certify this as a bad example, we should be able to come up with some LCL profile $\mathbf{V} = (V_1, \ldots, V_n)$ that "explains" the badness of the example. There are two properties that make a "bad example" bad:

(A) Each of the $f(\alpha)$ (the columns of M) agree in many coordinates with some ambient curve $u(x)$ of degree ℓ.

(B) No b columns of M themselves lie on the same degree-ℓ curve.

Item (A) can be captured with an a-local profile $\mathbf{V} = (V_1, \ldots, V_n)$. Indeed, let $A_i = \{\alpha \in A : u(\alpha)_i = f(\alpha)_i\}$, and then define V_i to be the subspace of all $y \in \mathbb{F}_q^A$ so that y (when viewed as a function $y : A \to \mathbb{F}_q$) agrees with some degree-ℓ polynomial on A_i. Then consider the class $\mathcal{F}$ of all profiles $\mathbf{V}$ generated this way (over all appropriate choices of A_i). Item (A) holds if and only if C contains some $\mathbf{V} \in \mathcal{F}$.

However, it is not clear how to capture item (B) with the LCL framework, as (B) is a more global property of the matrix, not easily captured with coordinate-wise constraints. Thus, another approach seems needed.

As a second possible approach, we describe what [16] did, and briefly explain why it incurs an exponential loss in the parameter ℓ. The idea is to consider first the special case that $b = \ell + 2$. Given a bad example for curve-decoding, consider a much larger "witness" matrix $M \in \mathbb{F}_q^{n \times r}$ for $r = \binom{a}{\ell+1}$, as follows. The columns of M now correspond to subsets $S \in \binom{A}{\ell+1}$, and the columns correspond to polynomials $g_S(x)$ defined by interpolating through the points $(\alpha, f(\alpha))$ for all $\alpha \in S$. It turns out that if $b = \ell + 2$, then all of these polynomials must be distinct: Otherwise, if $g_S = g_{S'}$ for some $S \neq S'$, then this gives a degree-ℓ curve passing through at least $|S \cup S'| \geq \ell + 2 = b$ of the codewords $f(\alpha)$, contradicting the badness of this example. Then, if $\mathbb{F}_q$ is large enough, there is some β so that $\left\{g_S(\beta) \in C : S \in \binom{A}{\ell+1}\right\}$ are also all distinct, and these can be the columns of M. However, this requires us to take $\ell = b + 2$; otherwise, we can't be guaranteed that the columns are distinct, which is required by the LCL framework. The insight of [16] was to essentially transform the general-ℓ case

to the $\ell = b + 2$ case via a new property called $\bigvee$-decoding; but this came at a cost of an exponential dependence on ℓ. (We emphasize that the exponential loss comes from the transformation to the $b = \ell + 2$ case, not the larger size of the witness matrix, although we will be able to overcome both issues in this work.)

Our Approach: Row-span Constrained LCL Properties. Instead of trying to manipulate curve-decodability to fit into an LCL property, we extend the definition of an LCL property. That is, we return to the first, more straightforward, attempt above. Recall that the first property (A) – that each column of M is close to some degree-ℓ curve $u(x)$ –is easily captured by an a-local LCL profile. The harder part is dealing with item (B), that no b columns of M should themselves lie on a degree-ℓ curve. This is a global constraint on the matrix M, and thus is not immediately amenable to the coordinate-wise framework of [27].

In order to capture item (B), we introduce a *row-span* constraint. One example of such a constraint was already present in prior works [9, 27]. Indeed, that framework requires that the columns of the witness matrix M be distinct, which turns out to be equivalent to the property that the row-span U of M does not have any identically equal coordinates (that is, there should not be any $i \neq j$ so that $u_j = u_i$ for all $u \in U$). Thus, [9, 27] restricted the row-span of their witness matrices M to the set $\mathcal{L}_{\text{dist}}(\mathbb{F}_q^r)$ of *distinct* subspaces U, those without any identically equal coordinates.

Our observation is that item (B) can also be captured in this way. In Lemma 50 and Corollary 51, we show that item (B) is equivalent to the following: For any $B \in \binom{A}{b}$, there is some u in the row-span U of M so that $u|_B$ does *not* agree with any degree-ℓ curve $p : B \to \mathbb{F}_q$. We define a subspace U with this property to be (ℓ, A, b) *curve-free* (Definition 49). Thus, in addition to an a-local LCL profile $\mathbf{V}$ to enforce the condition (A), we also have a condition on the row-span – namely that the row-span U of M should be (ℓ, A, b) curve free – to enforce condition (B).

Formally, we implement this by defining the set $\mathcal{F}$ to be a set of *pairs* $(\mathbf{V}, U)$, where $\mathbf{V}$ is an r-local profile, and $U \subset \mathbb{F}_q^r$ is a subspace. Then a witness matrix M with columns in C is "bad" if there is some $(\mathbf{V}, U) \in \mathcal{F}$ so that the i-th row of M is contained in V_i for all i; and additionally the row-span of M is contained in U. We say that C *avoids* $\mathcal{F}$ if there is no such witness matrix.

Putting it Together. In order to implement the ideas above, we first re-derive the LCL framework and the connections in [9, 27] in the row-span constrained setting. We do this in Section 3. To do so, in Section 3, we introduces row-span constrained LCL profiles and states the corresponding threshold and transfer theorems for random linear, random Reed–Solomon, random LDPC, and subspace-design codes. In particular, Theorems 34 and 45 and Corollaries 37 and 48 identify the relevant threshold rate and its transfer to subspace-design codes.

Next, in Section 4, we characterize *not* being (ℓ, δ, a, b) curve-decoding as a row-span constrained a-local LCL property, as described above. The local profile records the agreement of the nearby codewords with an ambient degree-ℓ curve (property (A)), while the row-span constraint records the absence of a degree-ℓ codeword curve through any b nearby codewords (property (B)). This is stated formally in Theorem 52 and Corollary 53.

Finally, we apply our row-span constrained LCL theory to curve-decoding and to three random ensembles of codes – random linear codes, RS codes with random evaluation points, and Gallager's ensemble of LDPC codes – to prove Theorem 2. To recap, we know from [16] that subspace design codes satisfy curve-decodability; we have shown in Section 3 that

random ensembles of codes behave similarly to subspace design codes for any *row-span constrained* LCL property; and we have shown in Section 4 that curve-decodability is such a property. Therefore random ensembles of codes exhibit good curve-decodability as well. Section 5 states the resulting proximity-gap, CA, and MCA guarantees for the three random ensembles.

Due to space constraints, we omit some proofs and calculations; the reader can find these in the full version of this paper [18].

1.3 Related Works

Proximity Gaps. Proximity gaps (and related notions, including correlated agreement and mutual correlated agreement) arise naturally in the study of interactive protocols. The goal in these settings is to test whether or not *every* vector in a set – say, a line or a curve – is close to a code C. If that code exhibits a proximity gap, then it suffices to choose a random vector in that set and use its distance from the code as a proxy for the minimum distance to C over the whole set. This connection was observed in [29], and since then a long line of work has been devoted to improving the parameters [1, 3, 6–8, 16] and establishing limitations [5, 12, 13, 26]. See the recent survey [2] for an overview of the work on proximity gaps and related notions. Prior to the work of [16], the best known proximity gap results required $\delta < 1 - \sqrt{R}$, corresponding to the *Johnson radius* in list-decoding. As discussed above, the work [16] improved this to allow δ to approach $1 - R$, the best trade-off possible. However, as noted earlier, the results can still be improved, notably in how small the parameter ε can be. The current best results for explicit codes – including folded RS codes, multiplicity codes, and all subspace design codes – follow from [16] and work under the condition given by (1). In this work we improve their results for random ensembles of codes from (2) to (1).

Local properties and subspace design codes. A recent line of work [21, 22, 27, 28] has shown that a wide variety of random ensembles of codes – including random linear codes, random LDPC codes, and Reed-Solomon codes with random evaluation points – all behave similarly with respect to "local" properties: properties that are characterized by the exclusion of small bad sets. The paper [27] introduced the LCL framework discussed above, and [9] connected this to another line of work on *subspace design codes*. The study of subspace design codes implicitly began with the introduction of Folded Reed-Solomon codes [23]. Over nearly two decades, it was observed that a key property of these codes, as well as of univariate multiplicity codes, is that they satisfy a *subspace design property* [20], which has been leveraged and generalized in subsequent works [4, 9, 11, 17, 25, 30, 31]. Notably, [11] showed that *any* such subspace design codes enjoy optimal list-decodability. The work [9] then connected this line of work to the LCL framework discussed above, showing that in fact the same thresholds are achieved by subspace design codes. The work [16] leveraged this connection to study curve-decoding, and we do the same in this work.

1.4 Discussion and Open Problems

We conclude this section with a few open questions.

- Our approach can automatically turn any improvements for subspace design codes into analogous improvements for random ensembles of codes. Thus, the natural open question is to further improve the trade-offs between ε, q, ℓ and η for subspace design codes.

- Our results imply that Reed-Solomon codes with random evaluation points attain near-optimal proximity gaps, even for large ℓ. While it is known that there exist Reed-Solomon codes that do *not* achieve optimal proximity gaps [5, 12, 13, 26], it is not known which choices of evaluation points are "good." Can we identify explicit Reed-Solomon codes with near-optimal proximity gaps? This question is related to the open problem of identifying explicit Reed-Solomon codes that achieve near-optimal list-decoding.
- We hope that our techniques will be useful beyond the present application. Are there other natural properties that can be captured with the row-span constrained LCL framework, but not with the original LCL framework?

2 Preliminaries

We begin by introducing the basic coding-theoretic definitions we will be using. For any two vectors x, y in Σ^n where Σ is some alphabet, we define the fractional Hamming distance $\Delta(x,y) = \frac{1}{n}|\{i \in [n] : x_i \neq y_i\}|$ to be the fraction of coordinates where they differ. For a set $S \subset \Sigma^n$, we define $\Delta(x,S) = \min_{y \in S} \Delta(x,y)$ to be the fractional distance of x to its closest vector in S. Throughout this paper, all distances are taken to be fractional unless stated otherwise.

The two fundamental quantities associated with a code are its rate and distance. For a code $C \subset \Sigma^n$, we define its (relative) distance as $\delta(C) = \min_{x,y \in C, x \neq y} \Delta(x,y)$. Moreover, its rate $R(C)$ is defined as $R(C) = \frac{1}{n}\log_{|\Sigma|}|C|$.

In this paper, we will focus on additive codes over a finite field, defined as follows:

▶ **Definition 4** (Additive codes). *Let $\mathbb{F}$ be a finite field and let $\Sigma = \mathbb{F}^s$ for some positive integer s. A code $C \subset \Sigma^n$ is said to be $\mathbb{F}$-additive (or just additive when the field $\mathbb{F}$ is clear from context) if C is an $\mathbb{F}$-linear subspace of Σ^n.*

We focus on three families of random codes; the first is a random linear code.

▶ **Definition 5** (Random Linear Code (RLC) and Random $\mathbb{F}_q$-Additive Code). *A random linear code $C \subset \mathbb{F}_q^n$ of rate R is given by taking $C = \{(\mathsf{Enc}_1(x), \cdots, \mathsf{Enc}_n(x)) \mid x \in \mathbb{F}_q^{Rn}\}$ where $\mathsf{Enc}_i : \mathbb{F}_q^{Rn} \to \mathbb{F}_q$ are independent uniformly random linear maps.* [5] *More generally, a random $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ of rate R is given by taking $C = \{(\mathsf{Enc}_1(x), \ldots, \mathsf{Enc}_n(x)) \mid x \in \mathbb{F}_q^{Rsn}\}$ where $\mathsf{Enc}_i : \mathbb{F}_q^{Rsn} \to \mathbb{F}_q^s$ are independent uniformly random linear maps.*

The next family is a random Reed-Solomon code, which is just a Reed-Solomon code where the evaluation points are chosen uniformly at random.

▶ **Definition 6** (Random Reed-Solomon Code (RRS)). *A Reed-Solomon code over a field $\mathbb{F}_q$ with evaluation points $\alpha_1, \cdots, \alpha_n$ is defined as follows*

$$\mathsf{RS}_{\mathbb{F}_q}(\alpha_1, \ldots, \alpha_n, k) = \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f \in \mathbb{F}_q[X], \deg f < k\}$$

A random Reed-Solomon code of rate R is one in which $k = Rn$ and α_i are independent uniformly random elements of $\mathbb{F}_q$.

[5] Note that, defined this way, there is some small chance that the rate $\log_q |C|/n$ of a random linear code may not actually be R, if there are too many linear dependencies between the maps Enc_i. However, it is more convenient to work with i.i.d. encoding maps, and the probability of the rate not being R is negligible, so anything that holds with high probability for this model also holds with high probability for other models of random linear codes, for example the model where one chooses C to be a uniformly random subspace of dimension Rn, or the model where C is the kernel of a random matrix in $\mathbb{F}_q^{(1-R)n \times n}$. See, e.g., [27, Appendix A].

Finally, we introduce our last family of codes, random Low-Density Parity-Check (LDPC) codes. In this paper, by "random LDPC codes" we mean *Gallager's ensemble* of random LDPC codes [15]. In general, a rate R, s-sparse LDPC code can be described as follows: Pick a bipartite graph $G = (L \sqcup R, E)$ with $|L| = n$, $|R| = (1-R)n$ and right degree s. Each edge $e \in E$ has a weight $w_e \in \mathbb{F}_q$, the code C is given as follows:

$$C = \Big\{ c \in \mathbb{F}_q^L : \sum_{u \in N(v)} w_{(u,v)} c_u = 0 \text{ for all } v \in R \Big\}.$$

Above, $N(v)$ denotes the neighborhood of v in G.

To define a random LDPC code, we thus just need to describe the underlying bipartite graph and the edge weights w_e. Gallager's ensemble [15] defines an ensemble of random graphs (and hence random LDPC codes) as follows.[6]

▶ **Definition 7** (Random Low Density Parity Check Code (RLDPC)). *Let $t = (1-R)s$, and let $G_i = (V, W_i, E_i)$ for $i = 1, \cdots, t$ be uniformly random $(1, s)$ regular bipartite graphs with a shared left vertex set and disjoint right vertex sets W_i of size n/s. The graph G is then defined as $\bigsqcup_i G_i$; that is, $G = (V, W, E)$, where $W = \bigsqcup_i W_i$ and $E = \bigsqcup_i E_i$. The edge weights w_e are sampled as independent uniformly random elements of $\mathbb{F}_q^*$. A* Random LDPC Code *C of rate R is then the LDPC code derived from G and the weights $\{w_e\}$, as described above.*

2.1 Proximity Gaps and (Mutual) Correlated Agreement

Proximity Gaps are related to two stronger notions, *Correlated Agreement* (CA) and *Mutual Correlated Agreement* (MCA). Our results actually establish both CA and MCA as well as proximity gaps.

Informally, we say that a code C has $(\ell, \delta, \varepsilon, \gamma)$-correlated agreement if, for all degree-ℓ curves $u(x) = \sum_{i=0}^{\ell} u_i x^i$ with $u_i \in \Sigma^n$, the following holds: Either at most an ε fraction of the points on $u(x)$ are δ-close to C; or there is some "codeword curve" $c(x) = \sum_{i=0}^{\ell} c_i x^i$ with $c_i \in C$ so that $u(x)$ agrees identically with $c(x)$ on a set $S \subset [n]$ of at least $\delta n/(1-\gamma)$ coordinates. Formally, we have the following definition.

▶ **Definition 8** (Correlated Agreement (CA)). *An $\mathbb{F}$-additive code $C \subset \Sigma^n$ is said to have $(\ell, \delta, \varepsilon, \gamma)$ correlated agreement if for all $u_0, u_1, \cdots, u_\ell \in \Sigma^n$ and $\delta' \le \delta$*

$$\mathbb{P}_{\alpha \in \mathbb{F}}\Big[\Delta\big(\sum_{j=0}^{\ell} u_j \alpha^j, C\big) \le \delta'\Big] > \varepsilon \implies \exists c_0, c_1, \cdots c_\ell \in C : \Big|\Big\{ i : \bigvee_{j=0}^{\ell} c_{j,i} \neq u_{j,i} \Big\}\Big| \le \frac{\delta' n}{1-\gamma}.$$

In particular, CA implies that for all $\alpha \in \mathbb{F}$, we have that $\Delta\Big(\sum_{j=0}^{\ell} u_j \alpha^j, \sum_{j=0}^{\ell} c_j \alpha^j\Big) \le \delta'/(1-\gamma)$. Thus, $(\ell, \delta, \varepsilon, \gamma)$ correlated agreement implies an $(\ell, \delta, \varepsilon, \gamma)$proximity gap.

Next, we define Mutual Correlated Agreement (MCA). Informally, we say that an $\mathbb{F}$-additive code C has $(\ell, \delta, \varepsilon)$ MCA if for all degree-ℓ curves $u(x)$ in Σ^n, the following holds. For all sets $S \subset [n]$ of size at least $(1-\delta)n$, and for at least an $1-\varepsilon$ fraction of $\alpha \in \mathbb{F}$, if there is some $c \in C$ that agrees with $u(\alpha)$ on S, then there is some "codeword curve" $c(x) = \sum_{i=0}^{\ell} c_i x^i$ for $c_i \in C$ so that $c(x)$ and $u(x)$ agree identically on S. Formally, we have the following definition, which captures the contrapositive of the intuition above.

[6] Technically, this way of generalizing to larger alphabets is slightly different than [15]. Our definition coincides with [15] for $q = 2$; for larger q, [15] chooses weights w_e to all be 1, while we choose them uniformly in $\mathbb{F}_q^*$; this matches the ensemble studied in [28].

▶ **Definition 9** (Mutual Correlated Agreement (MCA)). *An $\mathbb{F}$-additive code $C \subset \Sigma^n$ is said to have $(\ell, \delta, \varepsilon)$ mutual correlated agreement if for all $u_0, u_1, \cdots, u_\ell \in \Sigma^n$ and $\delta' \leq \delta$*

$$\mathbb{P}_{\alpha \in \mathbb{F}}\left(\exists S \subset [n], c \in C, j \in \{0,1,\ldots,\ell\} : |S| \geq (1-\delta')n, c_{|S} = \sum_{j=0}^{\ell} \alpha^j u_{j|S}, c'_{|S} \neq u_{j|S} \forall c' \in C \right) \leq \varepsilon.$$

We note that $(\ell, \delta, \varepsilon)$ MCA implies $(\ell, \delta, \varepsilon, \gamma = 0)$ CA and hence $(\ell, \delta, \varepsilon, \gamma = 0)$ PG. Indeed, suppose that C satisfies MCA, and suppose there is some codeword $c \in C$ that is δ'-close to $u(x)$ on at least εq points, call them $A \subset \mathbb{F}$. MCA implies that there is a set $B \subset \mathbb{F}$ of size at least $(1-\varepsilon)q$ so that the event inside the $\mathbb{P}_\alpha[\cdot]$ statement holds. Then there must be at least one point $\alpha \in A \cap B$. Letting $S \subset [n]$ be the set of coordinates (of size at least $(1-\delta')n$) on which $u(\alpha)$ agrees with c, we conclude from the definition of MCA that there is some "codeword curve" $c(x)$ so that $c(x)$ and $u(x)$ identically agree on S, establishing CA.

Finally, we formally define curve-decodability, which we informally defined in Section 1 (see Figure 1). The following definition was given in [16], and the case of $\ell = 1$ was independently defined in [24] as *collinearity of proximates*. We adopt the convention from [16].

▶ **Definition 10** (Curve-Decodability). *An additive code $C \subset \Sigma^n$ is (ℓ, δ, a, b) curve-decodable if for every $u_0, u_1, \cdots, u_\ell \in \Sigma^n$, all functions $f : \mathbb{F}_q \to C$, whenever the set $A = \left\{\alpha \in \mathbb{F} : \Delta\left(\sum_{j=0}^{\ell} u_j \alpha^j, f(\alpha)\right) \leq \delta\right\}$ has at least a elements, there exist $c_0, c_1, \cdots, c_\ell \in C$ such that $\left|\left\{\alpha \in A : f(\alpha) = \sum_{j=0}^{\ell} c_j \alpha^j\right\}\right| \geq b$.*

The work [16] showed a transformation to boost results about the curve-decodability of a code assuming that the underlying code is also list-decodable.

▶ **Theorem 11** ([16, Theorem 3.6]). *Suppose an $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ is $\left(\delta\left(1+\frac{\ell}{b-\ell}\right), L\right)$-list-decodable and (ℓ, δ, a, b)-curve-decodable. Then C is also $(\ell, \delta, (T-1)L + a, T)$ curve-decodable for any T.*

The key motivation to define curve-decodability is due to the following implication for correlated and mutual correlated agreement.

▶ **Theorem 12** (Correlated agreement from curve-decodability, [16, Theorem 3.4]). *Let $C \subset \Sigma^n$ be an $\mathbb{F}_q$-additive code. Let $\ell, t > \ell, a \in \mathbb{N}$ and $\delta \in (0,1)$. Suppose that $\mathcal{C}$ is (ℓ, δ, a, t) curve-decodable. Then C has $(\ell, \delta, a/q, \ell/t)$ correlated agreement.*

▶ **Theorem 13** (Mutual correlated agreement from curve-decodability, [16, Theorem 3.5]). *Let $C \subset \Sigma^n$ be an $\mathbb{F}_q$-additive code. Let $\ell, a \in \mathbb{N}$ and $\delta \in (0,1)$. Suppose that $\mathcal{C}$ is $(\ell, \delta, a, \ell n + 1)$ curve-decodable. Then C has $(\ell, \delta, a/q)$ mutual correlated agreement.*

▶ **Corollary 14.** *If an $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ is $\left(\delta\left(1+\frac{\ell}{b-\ell}\right), L\right)$-list-decodable and (ℓ, δ, a, b)-curve-decodable, then C has:*

1. *for any $t \geq 1$, $(\ell, \delta, (t-1)L + a, t)$ curve-decodability,*
2. *for any $m > 1$, $\left(\ell, \delta, \frac{\ell m L + a}{q}, \frac{1}{m}\right)$ correlated agreement and proximity gap,*
3. *$(\ell, \delta, \frac{\ell n L + a}{q})$ mutual correlated agreement.*

Enabled by the above reductions, to prove an additive code C has correlated agreement, it suffices to prove list-decodability and curve-decodability of the code. In this work, and related works, we always work in the regimes where the codes in consideration are already known to be list-decodable with good parameters. Thus, the key innovation lies in establishing curve-decodability.

2.2 Subspace-Design Codes

We introduce notation and results from [16] that we will use.

▶ **Definition 15** (Subspace-Design Code). *For any function $\tau : \mathbb{N} \to \mathbb{R}_{\leq 1}$, an $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ is said to be a τ-subspace design code if for every $r \in \mathbb{N}$, and every $\mathbb{F}_q$-linear subspace A of C of dimension at most r, $\frac{1}{n}\sum_{i=1}^{n} \dim A_i \leq \dim(A) \cdot \tau(r)$, where $A_i = \{a \in A \mid a_i = 0\}$.*

▶ **Definition 16.** *For an additive code $C \subset \left(\mathbb{F}_q^s\right)^n$ and dimension r, let*

$$\sigma(r) := \max\left\{\mathbb{E}_{i\in[n]}\left[\frac{\dim V_i}{\dim V}\right] : V \subset C, \dim V = r\right\} \tag{4}$$

where $V_i = \{x \in V : x_i = 0\}$. In the message space with encoding maps $\mathrm{Enc}_i : \mathbb{F}_q^k \to \mathbb{F}_q^s$, we have

$$\sigma(r) := \max\left\{\mathbb{E}_{i\in[n]}\left[\frac{\dim(A \cap \ker(\mathrm{Enc}_i))}{\dim A}\right] : A \subset \mathbb{F}_q^k, \dim A = r\right\} \tag{5}$$

Following the notation of [16], we can define the monotone envelope $\tau(r) := \max\{\sigma(1), \ldots, \sigma(r)\}$. Observe that with this choice, C is by definition a τ-subspace design code.

▶ Remark 17 (Comparison to [9]). The paper [9] uses slightly different language; they say that C is *r-subspace designable if*, for every $d \in [r]$, $\sigma(d) \leq R + \frac{1}{nd}$. We recall that $\sigma(r) \geq R - r/sn$ always. Thus, having the subspace design property with "good" parameters means that we should have $\tau(d) \leq R + \eta$ for some "small" $\eta > 0$. [9] proved that random $\mathbb{F}_q$-additive codes $C \subset (\mathbb{F}_q^s)^n$ are r-subspace designable whenever $s = \Omega(rn)$.

Finally, we state a theorem from [16] which shows that any subspace-design code is curve-decodable.

▶ **Theorem 18** ([16, Theorem 4.7]). *Fix any integers a, ℓ and $\eta > 0$, and let $d = \lceil(\ell + 1)/\eta\rceil$. Then for any τ-subspace design code $C \subset (\mathbb{F}_q^s)^n$, C is $\left(\ell, 1 - \tau(d) - \eta, a, \frac{\eta a}{d+\eta}\right)$-curve-decodable.*

3 Row-span Constrained Local Properties

3.1 Generalizing Local Profiles

We begin by introducing some notation from [9, 27], and explaining how we generalize it to our row-span constrained versions. As described in Section 1.2, a *local profile* gives a set of coordinate-wise linear constraints that a "bad" witness matrix M might satisfy. This section states the threshold and transfer results needed later; their proofs appear in the full version [18].

▶ **Definition 19** (r-Local Profile, [27]). *Let $\mathbb{F}_q$ be a finite field of q elements and let n and r be positive integers. A tuple $\mathbf{V} = (V_1, \cdots, V_n) \in \mathcal{L}(\mathbb{F}_q^r)^n$ is called a r-local profile.*

Next, we define what it means for a code to *contain* a local profile (with a particular row-span).

▶ **Definition 20** (Containing a local profile, [27]). *A code $C \subset \mathbb{F}_q^n$ is said to* contain *a local profile $\mathbf{V}$ with span $U \in \mathcal{L}(\mathbb{F}_q^r)$ (often written "C* contains *$(\mathbf{V}, U)$") if there exists an $M \in \mathbb{F}_q^{n\times r}$ such that:*

- *each column of M is contained in C;*
- *the i-th row of M is contained in V_i; and*
- *the row-span of M is U.*

We call such an M a witness matrix*, and columns of M* witness columns*.*

In [27], an LCL property is given by a collection $\mathcal{F}$ of r-local profiles $\mathbf{V}$. Our key generalization is to also record the corresponding row-spans in the family. That is, $\mathcal{F}$ is no longer a collection of profiles $\mathbf{V}$; now it is a collection of *pairs* $(\mathbf{V}, U)$, where U constrains the row-span of a witness matrix.

▶ **Definition 21** (Row-span Constrained LCL Property)**.** *A* row-span constrained r-local LCL family $\mathcal{F}$ *is a collection of* r-local pairs $(\mathbf{V}, U)$ *where* $\mathbf{V} = (V_1, \ldots, V_n) \in \mathcal{L}(\mathbb{F}_q^r)^n$ *is an r-local profile and* $U \in \mathcal{L}(\mathbb{F}_q^r)$ *is the row-span constraint of the witness matrix.*

A code C contains *$\mathcal{F}$ if it contains some pair $(\mathbf{V}, U) \in \mathcal{F}$. It* avoids *$\mathcal{F}$ otherwise. We say that the* row-span constrained r-local property *given by $\mathcal{F}$ is the property of* avoiding *$\mathcal{F}$.*[7]

▶ Remark 22. Definition 21 generalizes the definition of an LCL property in [27], in the sense that they restrict the columns of their witness matrices M to be distinct; as they observe, this is the same as restricting the row-span of the witness matrices to lie in $\mathcal{L}_{\text{dist}}(\mathbb{F}_q^r)$, where $\mathcal{L}_{\text{dist}}(\mathbb{F}_q^r)$ denotes all subspaces U of $\mathbb{F}_q^r$ so that no two distinct coordinates $i, j \in [r]$ are identically equal on U.

Next, we define what it means for an $\mathbb{F}_q$-additive code to contain $(\mathbf{V}, U)$.

▶ **Definition 23** (Folded containment)**.** *Let $C \subset (\mathbb{F}_q^s)^n$ be an $\mathbb{F}_q$-additive code. For an r-local profile $\mathbf{V} = (V_1, \ldots, V_n) \in (\mathcal{L}(\mathbb{F}_q))^n$, define the* duplicated profile $\mathbf{V}^{\otimes s} \in (\mathcal{L}(\mathbb{F}_q))^{sn}$ *by*

$$V^{\otimes s}_{(i,j)} = V_i \tag{6}$$

for any $i \in [n]$ and $j \in [s]$. We say that C contains $(\mathbf{V}, U)$ if the "unfolded" version $C \subset \mathbb{F}_q^{sn}$ contains $(\mathbf{V}^{\otimes s}, U)$. Here, by the "unfolded code," we mean to embed $(\mathbb{F}_q^s)^n$ in $\mathbb{F}_q^{sn}$ in the natural way.

For a family $\mathcal{F}$ of pairs, define $\mathcal{F}^{\otimes s} = \{(\mathbf{V}^{\otimes s}, U) : (\mathbf{V}, U) \in \mathcal{F}\}$. Thus the $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ contains $\mathcal{F}$ if and only if the "unfolded" version $C \subset \mathbb{F}_q^{sn}$ contains $\mathcal{F}^{\otimes s}$.

▶ Remark 24. Note that $\mathcal{F}^{\otimes s}$ does not encode the same property as $\mathcal{F}$ applied to unfolded codes. For example, proximity is not preserved: being far as an $\mathbb{F}_q$-additive code over $\Sigma = \mathbb{F}_q^s$ does not mean it is far when we unfold, since we can change one coordinate per block of s. The onus is on the property proving step, to show for any additive code, the desired property profiles take the form (6) for every s.

Now, we define the potential function and the threshold of row-span constrained r-local LCL properties identically as [27], up to slightly modified notational conventions and normalization by n. We emphasize that these definitions are code-independent.

▶ **Definition 25** (Potential and Threshold, [27])**.** *For an r-local pair $(\mathbf{V}, U)$, and a real number R, define the* potential function

$$\phi_{\mathbf{V}}(U, R) \coloneqq (R-1)\dim U + \mathbb{E}_{i \in [n]} \dim(U \cap V_i). \tag{7}$$

[7] We note that other sources, like [27], define an r-local property to be the property of *containing* $\mathcal{F}$. Since we are concerned with the property of avoiding a family $\mathcal{F}$, we define it this way.

For $W \subsetneq U$, define

$$R_{\mathbf{V}}(U, W) \coloneqq 1 - \mathbb{E}_{i\in[n]}\left[\frac{\dim(U \cap V_i) - \dim(W \cap V_i)}{\dim U - \dim W}\right]. \tag{8}$$

Define threshold *of the r-local pair $(\mathbf{V}, U)$ as*

$$R_{\mathbf{V}}(U) \coloneqq \max_{W\subsetneq U} R_{\mathbf{V}}(U, W). \tag{9}$$

For a row-span constrained r-local LCL property, define

$$R_{\mathcal{F}} \coloneqq \min_{(\mathbf{V},U)\in\mathcal{F}} R_{\mathbf{V}}(U). \tag{10}$$

The only difference between the definition above and that in [27] is the fact that the minimum in (10) is over profile-subspace pairs $(\mathbf{V}, U) \in \mathcal{F}$ rather than over profiles $\mathbf{V} \in \mathcal{F}$.

Next, we unpack the definitions to interpret the threshold as a root of the potential function.

▶ **Lemma 26** (Threshold interpretation). *For every r-local profile $\mathbf{V}$, $W \subsetneq U \in \mathcal{L}(\mathbb{F}_q^r)$, and $R \in \mathbb{R}$,*

$$\phi_{\mathbf{V}}(U, R) - \phi_{\mathbf{V}}(W, R) = (\dim U - \dim W)(R - R_{\mathbf{V}}(U, W)). \tag{11}$$

Let $\mathcal{F}$ be a row-span constrained r-local LCL family and let $\eta \geq 0$. We observe that

1. *If $R \leq R_{\mathcal{F}} - \eta$, then for every $(\mathbf{V}, U) \in \mathcal{F}$ there exists $W \subsetneq U$ such that $\phi_{\mathbf{V}}(U, R) - \phi_{\mathbf{V}}(W, R) \leq -\eta$.*
2. *If $R \geq R_{\mathcal{F}} + \eta$, then there exists $(\mathbf{V}, U) \in \mathcal{F}$ such that for every $W \subsetneq U$, $\phi_{\mathbf{V}}(U, R) - \phi_{\mathbf{V}}(W, R) \geq \eta$.*

We also see that by our definition of profile containment for additive (rather than linear) codes, the threshold rate R_V is preserved for profiles taking tensor form in (6) as s changes. Intuitively, this is because $(U, \mathbf{V}, W)$ are all independent of s, so potential function $\phi_{\mathbf{V}^{\otimes s}}(U, R) = \phi_{\mathbf{V}}(U, R)$. Then, $R_{\mathbf{V}^{\otimes s}}(U, W)$ is simply its root. We verify this directly.

▶ **Proposition 27.** *Let $(\mathbf{V}, U)$ be any r-local pair on plain codes in $\mathbb{F}_q^n$ and let $W \subsetneq U$. Then, for every $s \in \mathbb{N}$, $R_{V^{\otimes s}}(U, W) = R_V(U, W)$. Consequently, for any row-span constrained local LCL family, $R_{\mathcal{F}} = R_{\mathcal{F}^{\otimes s}}$ is independent of s.*

3.2 Behavior Under Quotients

We next record how local profile containment behaves under quotient maps. Again, omitted proofs can be found in [18].

▶ Remark 28 (Comparison to [17]). In [17], the spaces V_i, U can lie in an arbitrary vector space over $\mathbb{F}_q$. In contrast, our definition of r-local pairs require $V_i, U \subset \mathbb{F}_q^r$ and is thus more concrete. However, this is at the expense of a less elegant quotient behavior: since our profiles are coordinate-dependent, a quotient by $W \subsetneq U$ is implemented by choosing a surjective linear map $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ where $\ker \pi = W$. This is the same as quotienting by W and choosing coordinates on the quotient.

▶ **Definition 29** (Quotient profile). *Let $\mathbf{V} = (V_1, \ldots, V_n)$ be an r-local pair and let $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ be a surjective linear map, so $r' \leq r$. Define the r'-local profile $\pi\mathbf{V} \coloneqq (\pi V_1, \ldots, \pi V_n)$. For an r-local pair $(\mathbf{V}, U)$ and surjective linear map π, if $W = \ker \pi \subsetneq U \subset \mathbb{F}_q^r$, then we say $(\pi\mathbf{V}, \pi U)$ is a quotient pair of $(\mathbf{V}, U)$. Then, the condition $W \subsetneq U$ is equivalent to $\pi U \neq 0$.*

While the locality is changed in a quotient pair which in particular means it is no longer in $\mathcal{F}$, we will see next that the potential, threshold, and code-containment are well-behaved.

▶ **Lemma 30** (Quotient Containment). *Suppose code C contains r-local pair $(\mathbf{V}, U)$ and $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ is a surjective linear map such that $\pi U \neq 0$. Then, C contains $(\pi\mathbf{V}, \pi U)$.*

We remark that this definition of code containment works for all folding s.

▶ **Lemma 31** (Quotient Potential). *Let $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ be a surjective linear map with kernel $W \subsetneq U \in \mathcal{L}(\mathbb{F}_q^r)$. Then, for any $R \in \mathbb{R}$, $\phi_{\pi\mathbf{V}}(\pi U, R) = \phi_{\mathbf{V}}(U, R) - \phi_{\mathbf{V}}(W, R)$.*

Finally, we present a useful corollary of the low-rate code profile containment via the quotient characterization. We emphasize that this is code-independent, and reduces the low-rate directions of the threshold proofs of various codes into simple code-specific counting.

▶ **Lemma 32** (Maximal Quotient). *If C contains an r-local pair $(\mathbf{V}, U)$ and $R \leq R_{\mathbf{V}}(U) - \eta$ for some $R, \eta \geq 0$, then there exists a surjective linear map $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ such that*

- *$W \coloneqq \ker \pi \subsetneq U$,*
- *C contains r'-local pair $(\pi\mathbf{V}, \pi U)$,*
- *$\phi_{\pi\mathbf{V}}(\pi U, R) \leq -\eta \dim \pi U$.*

Moreover, we can choose π such that $\phi_{\pi\mathbf{V}}(U', R) \leq -\eta \dim U'$ for all nonzero subspaces $U' \subset \pi U$.

3.3 Thresholds for Random Ensembles of Codes

In this subsection, we state fixed-pair and family-level threshold results for the random ensembles considered here; omitted proofs can be found in [18]. The quantity $R_{\mathbf{V}}(U)$ is the fixed-pair threshold, and a union bound over pairs gives the corresponding statement for a row-span constrained LCL family $\mathcal{F}$.

3.3.1 Random Linear Codes

▶ **Lemma 33.** *Let $(\mathbf{V}, U)$ be an r-local pair, let $\mathcal{M}_{\mathbf{V}}(U)$ be the set of matrices $M \in \mathbb{F}_q^{n \times r}$ with row-span U that satisfy $\mathbf{V}$, and let $\mathcal{M}^{\star}_{\mathbf{V}}(U)$ be those with row-span contained in U. For any $R \in \mathbb{R}$, $|\mathcal{M}_{\mathbf{V}}(U)| \leq |\mathcal{M}^{\star}_{\mathbf{V}}(U)| = q^{n((1-R)\cdot \dim U + \phi_{\mathbf{V}}(U,R))}$.*

▶ **Theorem 34** (Fixed-Pair Random Linear Code Threshold). *Let $(\mathbf{V}, U)$ be an r-local pair and let $C \subset \mathbb{F}_q^n$ be a random linear code of rate R.*

1. *If $R \leq R_{\mathbf{V}}(U) - \eta$ for some $\eta > 0$, then $\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) \leq q^{-\eta n}$.*
2. *If $R \geq R_{\mathbf{V}}(U) + \eta$ for some $\eta > 0$, then $\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) \geq 1 - q^{r^2 - \eta n}$.*

The low-rate direction (1) will essentially follow from Lemma 32 and a few simple lemmas that we state first.

▶ **Lemma 35** ([27, Lemma 4.1]). *If $C \subset \mathbb{F}_q^n$ is an RLC of rate R, then for any $M \in \mathbb{F}_q^{n \times r}$*

$$\mathbb{P}(M_{\star j} \subset C \,\forall\, j \in [r]) = q^{(R-1)n \cdot \mathrm{rank}(M)} \tag{12}$$

The idea behind Lemma 35 is to view C as the kernel of $(1-R)n$ many linear constraints in $\mathbb{F}_q^n$ which column-span of M needs to satisfy. Each of $\mathrm{rank}(M)$ basis vectors satisfies them with probability $q^{-(1-R)n}$. We note that this is the only RLC-specific ingredient we use.

Combining with Lemma 33, we obtain the following corollary, which intuitively explains why $\phi_{\mathbf{V}}(U, R)$ characterizes the containment of $(\mathbf{V}, U)$ in a random linear code of rate C.

▶ **Corollary 36.** *The expected number of witnesses matrices M for the containment of an r-local pair $(\mathbf{V}, U)$ in a random linear code C with rate $R \in [0,1]$ is at most $q^{n\phi_{\mathbf{V}}(U,R)}$.*

With these lemmas, we deduce Theorem 34.

Proof of Theorem 34. For the low-rate case (1), if C contains $(\mathbf{V}, U)$, then by Lemma 32, there is a surjective linear map $\pi = \pi_{(\mathbf{V},U)} : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ such that C contains $(\pi\mathbf{V}, \pi U)$ and $\phi_{\pi\mathbf{V}}(\pi U, R) \le -\eta$. By a union bound and Corollary 36,

$$\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) \le \mathbb{P}(C \text{ contains } (\pi\mathbf{V}, \pi U)) \le q^{n\phi_{\pi\mathbf{V}}(\pi U, R)} \le q^{-\eta n}. \tag{13}$$

For the high-rate case (2), we know that $R \ge R_{\mathbf{V}}(U, W) + \eta$ for every $W \subsetneq U$. Then, $\phi_{\mathbf{V}}(U, R) \ge \phi_{\mathbf{V}}(W, R) + \eta$ for all $W \subsetneq U$. Let random variable X_W be the number of witnesses matrices M for the containment by C of an r-local pair $(\mathbf{V}, W)$. By Corollary 36, we know that

$$\mathbb{E}[X_W] \le q^{n\phi_{\mathbf{V}}(W,R)} \tag{14}$$

Now, let us fix any code C of rate $R \in [0,1]$ and consider matrices $\mathcal{M}^{\star}_{\mathbf{V}}(U)$ that satisfies $\mathbf{V}$ and has row-span contained in U. By Lemma 33, we know that

$$|\mathcal{M}^{\star}_{\mathbf{V}}(U)| = q^{n((1-R)\cdot \dim U + \phi_{\mathbf{V}}(U,R))}. \tag{15}$$

Now, for any realization of C, requiring each column of M to lie in C imposes at most $(1-R)n \dim U$ independent constraints, since C has codimension $(1-R)n$ and the row-span is contained in U. Therefore, the number of $M \in \mathcal{M}^{\star}_{\mathbf{V}}(U)$ contained in C is

$$\sum_{W \in \mathcal{L}(U)} X_W \ge |\mathcal{M}^{\star}_{\mathbf{V}}(U)| q^{(R-1)n \dim U} = q^{n\phi_{\mathbf{V}}(U,R)} \tag{16}$$

deterministically for every realization C. Now, C contains $(\mathbf{V}, U)$ if and only if $X_U > 0$, so

$$\begin{aligned}
\mathbb{P}(C \text{ does not contain } (\mathbf{V}, U)) &= \mathbb{P}(X_U = 0) \\
&\le \mathbb{P}\left(\sum_{W \subsetneq U} X_W \ge q^{n\phi_{\mathbf{V}}(U,R)} \right) \\
&\le \frac{\sum_{W \subsetneq U} \mathbb{E}[X_W]}{q^{n\phi_{\mathbf{V}}(U,R)}} \\
&\le \sum_{W \subsetneq U} q^{n\phi_{\mathbf{V}}(W,R) - n\phi_{\mathbf{V}}(U,R)} \\
&\le |\mathcal{L}(U)| q^{-\eta n}
\end{aligned} \tag{17}$$

where we apply Markov's inequality and (14). Now, the conclusion follows $|\mathcal{L}(U)| \le q^{r^2}$. ◀

We now lift Theorem 34 to the whole row-span constrained LCL family $\mathcal{F}$.

▶ **Corollary 37** (Random Linear Code Threshold)**.** *Let $\mathcal{F}$ be a row-span constrained r-local property and let $C \subset \mathbb{F}_q^n$ be a random linear code of rate R.*

1. *If $R \le R_{\mathcal{F}} - \eta$ for some $\eta > 0$, then $\mathbb{P}(C \text{ contains } \mathcal{F}) \le |\mathcal{F}| q^{-\eta n}$.*
2. *If $R \ge R_{\mathcal{F}} + \eta$ for some $\eta > 0$, then $\mathbb{P}(C \text{ contains } \mathcal{F}) \ge 1 - q^{r^2 - \eta n}$.*

Therefore, when $|\mathcal{F}| = q^{o(n)}$[8], we know that $R_{\mathcal{F}}$ is the threshold for RLCs for large n and q: RLCs of rate less than $R_{\mathcal{F}}$ avoid $\mathcal{F}$ with high probability; RLCs of rate greater than $R_{\mathcal{F}}$ contain $\mathcal{F}$ with high probability.

[8] Some previous literature, e.g. [27], call such $\mathcal{F}$ *reasonable*

3.3.2 Random Reed-Solomon Codes

We next record the analogue of the low-rate local threshold Corollary 37(1) for random Reed-Solomon code: if the rate is below the threshold of a row-span constrained LCL family, the random Reed-Solomon code is unlikely to contain the family. We use the following result from [27] as a black box.

▶ **Lemma 38** ([27, Proposition 6.1]). *Let $n \le q$ with q a prime power, and let $r \in \mathbb{N}$. Let $\mathbf{V} = (V_1, \cdots, V_n)$ be an r-local profile, $U \in \mathcal{L}(\mathbb{F}_q^r)$, and let $0 \le R \le 1$ such that for all $U' \in \mathcal{L}(U)$, $\phi_{\mathbf{V}}(U', R) \le -\varepsilon \cdot \dim U'$. Then, with probability taken over uniformly random $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$, we have*

$$\mathbb{P}\big(C = \mathsf{RS}_{\mathbb{F}_q}(\alpha_1, \ldots, \alpha_n, k) \text{ contains } V \text{ with span } U \neq \{0\}\big) \le (2^r - 1)\left(\frac{(4r)^{4r} k}{\varepsilon q}\right)^{\varepsilon n / 2r}.$$

▶ **Corollary 39** (Random Reed-Solomon Code Threshold: Low Rate Direction). *Let $\mathcal{F}$ be a row-span constrained r-local property and let $C \subset \mathbb{F}_q^n$ be a random Reed-Solomon code of rate R. If $R \le R_{\mathcal{F}} - \eta$ for some $\eta > 0$, then*

$$\mathbb{P}(C \text{ contains } \mathcal{F}) \le |\mathcal{F}|(2^r - 1)\left(\frac{(4r)^{4r} R n}{\eta q}\right)^{\eta n / 2r} \tag{18}$$

3.3.3 Random LDPCs (Gallager's Ensemble)

We next record the analogue of the low-rate local threshold Corollary 37(1) and Corollary 39 for random LDPC codes drawn from Gallager's Ensemble (Definition 7).

There is one important difference from random linear codes. For RLCs, the probability that a fixed matrix M is contained in the code is exactly $q^{(R-1)n \operatorname{rank}(M)}$ by Lemma 35. For LDPC codes C, this exact identity is false. Instead, [28] proves that, after conditioning on the usual good-distance event defined for $C = \mathsf{RLDPC}(n, q, s, R)$ as

$$\mathsf{Good} := \left\{\Delta(C) \ge \frac{1}{2} h_q^{-1}(1 - R)\right\}, \tag{19}$$

then C contains any fixed matrix with essentially the same probability as a random linear code, up to a factor $q^{\varepsilon n}$. This is the only LDPC-specific input we need.

We first recall the lemma from [28] saying that Good happens with high probability.

▶ **Lemma 40** ([28, Theorem 2.14]). *For any finite field $\mathbb{F}_q$ and $R \in [0, 1 - 1/q)$, there exists $s_0 = s_0(R, q)$ such that for every $s > s_0$, with probability over $C = \mathsf{RLDPC}(n, q, s, R)$, we have $\mathbb{P}(\mathsf{Good}) := \mathbb{P}\big(\Delta(C) \ge \frac{1}{2} h_q^{-1}(1 - R)\big) \ge 1 - o_{n\to\infty}(1)$.*

The next lemma is the fixed-pair containment estimate for row-span constrained local profiles. This lemma follows [28]; we restate it in our framework and refer to the full version [18] for the proof.

▶ **Lemma 41** (RLDPCs containing local profiles). *For any $\varepsilon, R \in (0, 1)$, positive integer r, finite field $\mathbb{F}_q$, there exists an $s_0 = s_0(\varepsilon, r, R, q)$ such that the following holds for all odd $s > s_0$ and sufficiently large n. Let $\mathbf{V} = (V_1, \ldots, V_n)$ be an r-local profile and let $U \in \mathcal{L}(\mathbb{F}_q^r)$. For $C = \mathsf{RLDPC}(n, q, s, R)$, $\mathbb{P}(C \text{ contains } (\mathbf{V}, U) | \mathsf{Good}) \le q^{n\phi_{\mathbf{V}}(U,R) + \varepsilon n}$, where $\mathsf{Good} := \big\{\Delta(C) \ge \frac{1}{2} h_q^{-1}(1 - R)\big\}$.*

The LDPC threshold statement differs from its RLC analogue only by conditioning on the good-distance event Good and by the additional $q^{\eta n/2}$ slack from Lemma 41. The complete derivation appears in the full version [18].

▶ **Theorem 42** (Fixed-Pair Random LDPC Threshold). *Let $(\mathbf{V}, U)$ be an r-local pair and let $C = \mathsf{RLDPC}(n, q, s, R)$ be a random LDPC code from Gallager's ensemble. Define the event* $\mathsf{Good} := \left\{\Delta(C) \geq \frac{1}{2} h_q^{-1}(1-R)\right\}$.

For every $R, \eta \in (0,1)$, positive integer r, and finite field $\mathbb{F}_q$, there exists $s_0 = s_0(\eta, r, R, q)$ such that for every odd $s > s_0$, the following holds for all sufficiently large n:

1. *If $R \leq R_{\mathbf{V}}(U) - \eta$, then $\mathbb{P}(C \text{ contains } (\mathbf{V}, U) | \mathsf{Good}) \leq q^{-\eta n/2}$, and so*

$$\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) \leq q^{-\eta n/2} + o_{n\to\infty}(1). \tag{20}$$

2. *If $R \geq R_{\mathbf{V}}(U) + \eta$, then $\mathbb{P}(C \text{ contains } (\mathbf{V}, U) | \mathsf{Good}) \geq 1 - q^{r^2 - \eta n/2}$, and so*

$$\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) \geq 1 - q^{r^2 - \eta n/2} - o_{n\to\infty}(1). \tag{21}$$

▶ Remark 43. Only part (1) of Theorem 42 is needed later; part (2) is recorded for completeness, and the proofs of both parts appear in the full version [18].

This should not be confused with a converse to the local-property transfer theorem for RLDPC codes ([28, Theorem 1.9]), which is false at that level of generality, as observed in [28]; our argument uses the additional linear coordinate-wise profile structure of row-span constrained LCL properties.

As before, the corresponding family-level statement is the following.

▶ **Corollary 44** (Random LDPC Code Threshold). *Let $\mathcal{F}$ be a row-span constrained r-local property and let $C = \mathsf{RLDPC}(n, q, s, R)$ be a random LDPC code from Gallager's ensemble (Definition 7). For every $R, \eta \in (0,1)$, positive integer r, and finite field $\mathbb{F}_q$, there exists $s_0 = s_0(\eta, r, R, q)$ such that for every odd $s > s_0$, the following holds for all sufficiently large n:*

1. *If $R \leq R_{\mathcal{F}} - \eta$, then $\mathbb{P}(C \text{ contains } \mathcal{F}) \leq |\mathcal{F}| q^{-\eta n/2} + o_{n\to\infty}(1)$*
2. *If $R \geq R_{\mathcal{F}} + \eta$, then $\mathbb{P}(C \text{ contains } \mathcal{F}) \geq 1 - q^{r^2 - \eta n/2} - o_{n\to\infty}(1)$*

3.4 Threshold for Subspace Design Codes

Informally, for ordinary r-local properties where every row-span $U \in \mathcal{L}_{\text{dist}}(\mathbb{F}_q^r)$ is permitted for each profile $\mathbf{V}$, [9, Theorems 3.1, 4.2] shows that the threshold of containing $\mathbf{V}$ is dictated by $\tau(r)$ versus $R_{\mathbf{V}}$. For good subspace design codes, like folded Reed-Solomon codes, we have $\tau(r) \approx R$, which means that good subspace design codes and random linear codes share the same threshold.

We generalize these results to row-span constrained local properties. We emphasize that the transfer works at the level of any pair $(\mathbf{V}, U)$ via the following theorem.

▶ **Theorem 45** (Fixed-Pair Subspace Design Code Threshold). *Let $(\mathbf{V}, U)$ be an r-local pair.*

1. *Let $C \subset (\mathbb{F}_q^s)^n$ be an additive τ-subspace design code. If $\tau(r) < R_{\mathbf{V}}(U)$, then C does not contain $(\mathbf{V}, U)$.*
2. *For any $\eta > 0$, there exists a τ-subspace design additive code $C \subset (\mathbb{F}_q^s)^n$ with $\tau(r) \leq R_{\mathbf{V}}(U) + \eta$ that contains $(\mathbf{V}, U)$.*

The statement (1) in Theorem 45 essentially follows from Lemma 32 and the subspace design condition.

Proof of Theorem 45(1). Suppose $\tau(r) \leq R_{\mathbf{V}}(U) - \eta$, then by Lemma 32 there exists surjective linear map $\pi : \mathbb{F}_q^r \to \mathbb{F}_q^{r'}$ such that $\phi_{\pi\mathbf{V}}(\pi U, \tau(r)) < 0$ and C contains $(\pi\mathbf{V}, \pi U)$. The latter implies the existence of a witness matrix M. Let the columns of the witness matrix M for $(\pi\mathbf{V}, \pi U)$ be $c_j = \text{Enc}(a_j)$ and define the message subspace $A := \text{span}(a_j)$. Since the encoder is injective and the witness matrix has row-span πU, $d := \dim A = \dim(\pi U)$.

For each coordinate $i \in [n]$, the image $\mathrm{Enc}_i(A) \subset \mathbb{F}_q^s$ is the column span of the i-th "block" of the witness matrix. The rows of that block lie in $\pi U \cap \pi V_i$, hence

$$\dim(\pi U \cap \pi V_i) \geq \dim \mathrm{Enc}_i(A) = \dim A - \dim(A \cap \ker(\mathrm{Enc}_i)). \tag{22}$$

where the equality follows from rank-nullity. Averaging over i and applying the subspace design condition Definition 15 with $d = \dim A \leq r$, we obtain that

$$d - \mathbb{E}_{i\in[n]} \dim(\pi U \cap \pi V_i) \leq \mathbb{E}_{i\in[n]} \dim(A \cap \ker(\mathrm{Enc}_i)) \leq \tau(r)d. \tag{23}$$

Rearranging gives $\phi_{\pi\mathbf{V}}(\pi U, \tau(r)) \geq 0$, a contradiction. Thus, C does not contain $(\mathbf{V}, U)$. ◀

For the statement (2) in Theorem 45, we use the following theorem from [9] that shows RLCs are good subspace design codes when viewed as an $\mathbb{F}_q$-additive code with alphabet $\mathbb{F}_q^s$.

▶ **Theorem 46** ([9, Theorem 4.1]). *Fix any $\eta > 0$. A random $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ with rate $R = k/sn$ satisfies*

$$\mathbb{P}(\tau(r) \leq R + \eta) \geq 1 - rq^{-\eta srn + 3(r^2+1)n}. \tag{24}$$

Proof of Theorem 45(2). Fix $\eta > 0$ and let $R = R_{\mathbf{V}}(U) + \eta/2$. For sufficiently large s, we know by Theorem 46 that, for a random $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$ of rate R,

$$\mathbb{P}\Big(\tau(r) \leq R_{\mathbf{V}}(U) + \eta = R + \frac{\eta}{2}\Big) \geq 1 - o_{s\to\infty}(1), \tag{25}$$

On the other hand, $R = R_{\mathbf{V}}(U) + \eta/2 = R_{\mathbf{V}^{\otimes s}}(U) + \eta/2$ by Proposition 27. By Theorem 34,

$$\mathbb{P}(C \text{ contains } (\mathbf{V}, U)) = \mathbb{P}(\text{unfolded } C \text{ contains } (\mathbf{V}^{\otimes s}, U)) \geq 1 - q^{r^2 - \eta sn/2} = 1 - o_{s\to\infty}(1). \tag{26}$$

Therefore, we can pick sufficiently large s such that with probability $2/3$, C contains $(\mathbf{V}, U)$ and is a τ-subspace design code with $\tau(r) \leq R_{\mathbf{V}}(U) + \eta$. This proves (2). ◀

▶ Remark 47. Compared to [9, Theorem 4.2], we take the subspace design parameter d in $\tau(d)$ to be $d = r$ for simplicity.

We can obtain analogous threshold statements for $R_{\mathcal{F}}$ by taking the minimum over $\mathcal{F}$.

▶ **Corollary 48** (Subspace Design Code Threshold). *Let $\mathcal{F}$ be any row-span constrained r-local LCL family. Then*

1. *Let $C \subset (\mathbb{F}_q^s)^n$ be an additive τ-subspace design code. If $\tau(r) < R_{\mathcal{F}}$, then C avoids $\mathcal{F}$.*
2. *For any $\eta > 0$, there exists a τ-subspace design additive code $C \subset (\mathbb{F}_q^s)^n$ with $\tau(r) \leq R_{\mathcal{F}} + \eta$ that contains $\mathcal{F}$.*

4 Curve-Decoding as a Row-Span Constrained LCL Property

Recall the definition of curve-decoding:

▶ **Definition 10** (Curve-Decodability). *An additive code $C \subset \Sigma^n$ is (ℓ, δ, a, b) curve-decodable if for every $u_0, u_1, \cdots, u_\ell \in \Sigma^n$, all functions $f : \mathbb{F}_q \to C$, whenever the set $A = \Big\{\alpha \in \mathbb{F} : \Delta\big(\sum_{j=0}^{\ell} u_j \alpha^j, f(\alpha)\big) \leq \delta\Big\}$ has at least a elements, there exist $c_0, c_1, \cdots, c_\ell \in C$ such that $\Big|\Big\{\alpha \in A : f(\alpha) = \sum_{j=0}^{\ell} c_j \alpha^j\Big\}\Big| \geq b$.*

As observed in [16] and discussed in Section 1.2, curve-decoding does not seem to fit immediately into the LCL framework. This section states its exact casting as a *row-span constrained* LCL property and refers to the full version [18] for the proof. Before we describe the correct row-span constraint, we define the degree-ℓ Reed-Solomon evaluation subspace on B by

$$\mathrm{RS}_\ell(B) \coloneqq \{(p(\alpha))_{\alpha\in B} : p \in \mathbb{F}_q[x],\ \deg p \le \ell\} \subset \mathbb{F}_q^B.$$

Thus $\mathrm{RS}_\ell(B)$ is the space of scalar degree-ℓ evaluation patterns on the parameter set B.

▶ **Definition 49** (Curve-free Subspaces). *Let $A \subset \mathbb{F}_q$. A subspace $U \in \mathcal{L}(\mathbb{F}_q^A)$ is (ℓ, A, b)* curve-free *if $\pi_B(U) \not\subset \mathrm{RS}_\ell(B)$ for every $B \in \binom{A}{b}$, where $\pi_B : \mathbb{F}_q^A \to \mathbb{F}_q^B$ denote coordinate restriction to B.*

For this definition to be non-trivial, we assume $\ell + 1 \le b \le |A|$. Now, we observe the following lemma, which translates (ℓ, A, b) curve-freeness of the row-span of a matrix to the condition that no b columns lie on a degree-ℓ curve; see [18] for the proof.

▶ **Lemma 50.** *Let $B \subset A \subset \mathbb{F}_q$ and let N be a positive integer.*[9] *Fix a matrix $M \in \mathbb{F}_q^{N\times A}$ with row-span $U \in \mathcal{L}(\mathbb{F}_q^A)$. Then, $\pi_B(U) \subset \mathrm{RS}_\ell(B)$ if and only if there exists a degree-ℓ curve $g : \mathbb{F}_q \to \mathbb{F}_q^N$ such that $g(\alpha) = M_{\star\alpha}$ for every $\alpha \in B$.*

Observe that latter condition is exactly the kind of codeword condition we wish to impose on columns of witness matrices for the failure of (ℓ, δ, a, b) curve-decoding. Lemma 50 gives an equivalent row-span constraint. We make this more formal below.

▶ **Corollary 51.** *Let $A \subset \mathbb{F}_q$ and let $M \in \mathbb{F}_q^{N\times A}$ be a matrix with row-span U whose columns lie in a linear code $C \subset \mathbb{F}_q^N$. If $\ell < b \le |A|$, then U is (ℓ, A, b) curve-free if and only if no b columns of M lie on a degree-ℓ "codeword curve." (That is, if and only if there is no curve $c(x) = \sum_{i=0}^{\ell} c_i x^i$ for $c_i \in C$ and no set $B \in \binom{A}{b}$ so that $c(\alpha) = M_{\star\alpha}$ for all $\alpha \in B$.)*

We now cast curve-decoding as a row-span constrained LCL property, thereby bypassing the notion of $\bigvee$-decodability from [16], and the parameter losses that arose from it.

▶ **Theorem 52** (Locality of curve-decoding). *Let $\ell, a, b \in \mathbb{N}$ with $\ell + 1 \le b \le a \le q$ and let $\delta \in (0, 1)$. Then there exists a row-span constrained a-LCL family $\mathcal{F}$ with $|\mathcal{F}| \le q^{a+a^2} 2^{an}$ such that every $\mathbb{F}_q$-additive code C is not (ℓ, δ, a, b) curve-decodable if and only if C contains $\mathcal{F}$.*

Proof. Fix indexing sets $A \in \binom{\mathbb{F}_q}{a}$ and $I_\alpha \subset [n]$ with $|I_\alpha| \ge (1-\delta)n$ for each $\alpha \in A$. For each choice of $A, \{I_\alpha\}_{\alpha\in A}$, we will define an r-local profile $\mathbf{V} = \mathbf{V}(A, \{I_\alpha\}_{\alpha\in A}) = (V_1, \ldots, V_n)$ as follows. For each $i \in [n]$, let $A_i \coloneqq \{\alpha \in A : i \in I_\alpha\}$. Define

$$V_i \coloneqq \left\{y \in \mathbb{F}_q^A : (y_\alpha)_{\alpha\in A_i} \in \mathrm{RS}_\ell(A_i)\right\}. \tag{27}$$

Equivalently, $y \in V_i$ if there exists a polynomial $p \in \mathbb{F}_q[x]$ of degree at most ℓ such that $p(\alpha) = y_\alpha$ for every $\alpha \in A_i$. This is clearly a linear subspace of $\mathbb{F}_q^A$. Now define

$$\mathcal{F} \coloneqq \left\{(\mathbf{V}(A, \{I_\alpha\}_{\alpha\in A}), U) : A \in \binom{\mathbb{F}_q}{a},\ I_\alpha \in \binom{[n]}{\ge (1-\delta)n},\ U \in \mathcal{L}(\mathbb{F}_q^A) \text{ is } (\ell, A, b) \text{ curve-free}\right\}. \tag{28}$$

We show the bound on the size of $\mathcal{F}$: There are $\binom{q}{a} \le q^a$ choices for A, at most 2^n choices for each $I_\alpha \subset [n]$, and the number of choices of U is at most $|\mathcal{L}(\mathbb{F}_q^a)| \le q^{a^2}$.

Next, we prove that not being curve-decodable is equivalent to containing $\mathcal{F}$.

[9] Think of N as either the block length $N = n$ of a linear code $C \subset \mathbb{F}_q^n$; or the block length $N = sn$ of an "unfolded" $\mathbb{F}_q$-additive code $C \subset (\mathbb{F}_q^s)^n$, thought of as $C \subset \mathbb{F}_q^{sn}$.

Not curve-decodable ⇒ Contains $\mathcal{F}$. If $C \subset (\mathbb{F}_q^s)^n$ is not (ℓ, δ, a, b) curve-decodable, then there exists $A \in \binom{\mathbb{F}_q}{a}$, a degree-$\ell$ curve $u : \mathbb{F}_q \to (\mathbb{F}_q^s)^n$, and codewords $f(\alpha) \in C$ such that $\Delta(u(\alpha), f(\alpha)) \leq \delta$ for every $\alpha \in A$, but no degree-ℓ codeword curve passes through $f(\alpha)$ for b-many of α. For $\alpha \in A$, define the agreement set $I_\alpha \coloneqq \{i \in [n] : u(\alpha)_i = f(\alpha)_i\}$, so $|I_\alpha| \geq (1-\delta)n$. Form the matrix $M \in \mathbb{F}_q^{sn \times A}$ with columns $M_{\star\alpha} = f(\alpha) \in C$. Index the rows by $[n] \times [s]$ and let U be the row-span of M.

We check the local profile constraints. For any $(i, j) \in [n] \times [s]$, the corresponding row $M_{ij,\star} = (f(\alpha)_{ij})_{\alpha \in A}$. Since $f(\alpha)_{ij} = u(\alpha)_{ij}$ whenever $i \in I_\alpha$, we observe that $M_{ij,\star}$ agrees with a polynomial $x \mapsto u(x)_{ij}$ of degree at most ℓ on A_i, so $M_{ij,\star} \in V_i$ by the definition of V_i in (27).

By assumption, no b of the columns $f(\alpha)$ lie on a degree-ℓ codeword curve. By Corollary 51, the row-span U is (A, ℓ, b) curve-free. Hence M witnesses that C contains $(\mathbf{V}(A, \{I_\alpha\}_{\alpha \in A}), U) \in \mathcal{F}$.

Contains $\mathcal{F}$ ⇒ Not curve-decodable. Suppose C contains the a-local pair $(\mathbf{V}(A, \{I_\alpha\}_{\alpha \in A}), U)$ via witness matrix $M \in \mathbb{F}_q^{sn \times A}$. Let the columns of M be $f(\alpha) \in C$ for $\alpha \in A$.

We construct a degree-ℓ curve close to these codewords. For $(i, j) \in [n] \times [s]$ the row $M_{ij,\star}$ lies in V_i, so by definition (27) of V_i, there exists a polynomial u_{ij} of degree at most ℓ such that $u_{ij}(\alpha) = f(\alpha)_{ij}$ for every $\alpha \in A_i \coloneqq \{\alpha \in A : i \in I_\alpha\}$. Collecting these polynomials over all (i, j) defines a degree-ℓ curve $u : \mathbb{F}_q \to (\mathbb{F}_q^s)^n$. For each $\alpha \in A$, if $i \in I_\alpha$ so $\alpha \in A_i$, then $u_{ij}(\alpha) = f(\alpha)_{ij}$ for every $j \in [s]$. Hence, $u(\alpha)_i = f(\alpha)_i$ for all $i \in I_\alpha$. For every $\alpha \in A$, $|I_\alpha| \geq (1-\delta)n$, so $\Delta(u(\alpha), f(\alpha)) \leq \delta$.

Finally, since U is (A, ℓ, b) curve-free, Corollary 51 implies that no b of the columns $f(\alpha)$ lie on a degree-ℓ codeword curve. Together, u and $f(\alpha)$ witness that C is not (ℓ, δ, a, b) curve-decodable. ◀

We can plug in specific choices of parameters from Theorem 18 for which subspace design codes are curve-decodable, thereby obtaining a lower bound on the threshold rate via Corollary 48.

▶ **Corollary 53.** *Let $\delta, \eta \in (0, 1)$ and integers $\ell, a, b \in \mathbb{N}$ such that*

$$d \coloneqq \left\lceil \frac{\ell + 1}{\eta} \right\rceil \leq a \quad \text{and} \quad b \leq \frac{\eta a}{d + \eta}. \tag{29}$$

Let $\mathcal{F}$ be the row-span constrained a-local LCL family from Theorem 52, so avoiding $\mathcal{F}$ is equivalent to being (ℓ, δ, a, b) curve-decodable. Then, $R_{\mathcal{F}} \geq 1 - \delta - \eta$. Additionally, $|\mathcal{F}| \leq q^{a+a^2} 2^{an}$.

We conclude with a specialization for the common choice of curve-decoding parameters.

▶ **Corollary 54.** *Let $\ell \in \mathbb{N}$ and $\delta, \eta \in (0, 1)$. Define $a, b, d \in \mathbb{N}$ by*

$$d \coloneqq \left\lceil \frac{\ell + 1}{\eta} \right\rceil, \qquad b \coloneqq \left\lceil \frac{4\ell}{\eta} + \ell \right\rceil, \qquad a \coloneqq \left\lceil \frac{100\ell^2}{\eta^3} \right\rceil. \tag{30}$$

Let $\mathcal{F}$ be the row-span constrained a-local LCL family from Theorem 52, so avoiding $\mathcal{F}$ is equivalent to being (ℓ, δ, a, b) curve-decodable. Then $R_{\mathcal{F}} \geq 1 - \delta - \eta$ and $|\mathcal{F}| \leq q^{a+a^2} 2^{an}$.

5 Improved Proximity Gaps

Finally, we establish proximity gaps (as well as CA and MCA) for three ensembles of codes. For each ensemble, we use Corollary 54 to prove curve-decodability, then we combine this with list-decodability via Corollary 14. We record the final statements below, and refer the reader to [18] for proofs. We begin with a statement for random linear codes.

▶ **Theorem 55** (PG, CA, and MCA for Random Linear Codes). *Let $\ell \in \mathbb{N}$ and $\eta, \delta \in (0,1)$. Let $C \subset \mathbb{F}_q^n$ be a random linear code of rate $R \le 1 - \delta - 2\eta$. If $q \ge \exp(\Omega(\ell^2/\eta^4))$ and $n \ge \Omega(\ell^4/\eta^7)$, then with high probability:*

1. *For every $T \ge 1$, C is curve-decodable with parameters $\left(\ell, \delta, (T-1)\left\lceil \frac{1-R}{\eta} \right\rceil + O\left(\frac{\ell^2}{\eta^3}\right), T\right)$.*
2. *For every $m > 1$, C has correlated agreement, and hence proximity gap, with parameters $\left(\ell, \delta, \frac{\ell m \lceil (1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}, \frac{1}{m}\right)$.*
3. *C has mutual correlated agreement with parameters $\left(\ell, \delta, \frac{\ell n \lceil (1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}\right)$.*

Next, we state our main result for Random Reed-Solomon Codes.

▶ **Theorem 56** (PG, CA, and MCA for Random Reed-Solomon Codes). *Let $\ell \in \mathbb{N}$ and $\eta, \delta \in (0,1)$. Let $C \subset \mathbb{F}_q^n$ be a random Reed-Solomon code of rate $R \le 1 - \delta - 2\eta$. If $q \ge n \cdot \exp(\Omega(\ell^4/\eta^7))$ and $n \ge \Omega(\ell^4/\eta^7)$, then with high probability:*

1. *For every $T \ge 1$, C is curve-decodable with parameters $\left(\ell, \delta, (T-1)\left\lceil \frac{1-R}{\eta} \right\rceil + O\left(\frac{\ell^2}{\eta^3}\right), T\right)$.*
2. *For every $m > 1$, C has correlated agreement, and hence proximity gap, with parameters $\left(\ell, \delta, \frac{\ell m \lceil (1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}, \frac{1}{m}\right)$.*
3. *C has mutual correlated agreement with parameters $\left(\ell, \delta, \frac{\ell n \lceil (1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}\right)$.*

Finally, we state our result for random LDPC codes.

▶ **Theorem 57** (PG, CA, and MCA for Random LDPC Codes). *Let $\ell \in \mathbb{N}$ and $\eta, \delta \in (0,1)$. There exists $s_0 = s_0(\ell, \eta, \delta, q)$ such that the following holds for every odd $s > s_0$. Let $C = \mathsf{RLDPC}(n, q, s, R)$ be a random LDPC code from Gallager's ensemble with rate $R \le 1 - \delta - 2\eta$. If $q \ge \exp(\Omega(\ell^2/\eta^4))$ and n is sufficiently large, then with high probability:*

1. *For every $T \ge 1$, C is curve-decodable with parameters $\left(\ell, \delta, (T-1)\left\lceil \frac{2(1-R)}{\eta} \right\rceil + O\left(\frac{\ell^2}{\eta^3}\right), T\right)$.*
2. *For every $m > 1$, C has correlated agreement, and hence proximity gap, with parameters $\left(\ell, \delta, \frac{\ell m \lceil 2(1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}, \frac{1}{m}\right)$.*
3. *C has mutual correlated agreement with parameters $\left(\ell, \delta, \frac{\ell n \lceil 2(1-R)/\eta \rceil + O(\ell^2/\eta^3)}{q}\right)$.*

References

1 Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. Ligero: Lightweight sublinear arguments without a trusted setup. In *Proceedings of the 2017 acm sigsac conference on computer and communications security*, pages 2087–2104, 2017. doi:10.1145/3133956.3134104.

2 Gal Arnon, Dan Boneh, and Giacomo Fenzi. Open problems in list decoding and correlated agreement. Cryptology ePrint Archive, Paper 2026/680, 2026. URL: https://eprint.iacr.org/2026/680.

3 Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. WHIR: Reed-Solomon proximity testing with super-fast verification. In *Advances in Cryptology–EUROCRYPT 2025-44th Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2025, Proceedings*, pages 214–243. Springer Science and Business Media Deutschland GmbH, 2025. doi:10.1007/978-3-031-91134-7_8.

4 Vikrant Ashvinkumar, Mursalin Habib, and Shashank Srivastava. Algorithmic improvements to list decoding of folded reed-solomon codes. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 880–898. SIAM, 2026. doi:10.1137/1.9781611978971.35.

5 Eli Ben-Sasson, Dan Carmon, Ulrich Haböck, Swastik Kopparty, and Shubhangi Saraf. On proximity gaps for reed–solomon codes. Cryptology ePrint Archive, Paper 2025/2055, 2025. URL: https://eprint.iacr.org/2025/2055.

6 Eli Ben-Sasson, Dan Carmon, Yuval Ishai, Swastik Kopparty, and Shubhangi Saraf. Proximity gaps for reed–solomon codes. *Journal of the ACM*, 70(5):1–57, 2023. doi:10.1145/3614423.

7 Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. DEEP-FRI: Sampling Outside the Box Improves Soundness. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:32. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ITCS.2020.5.

8 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)*, volume 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.CCC.2018.24.

9 Joshua Brakensiek, Yeyuan Chen, Manik Dhar, and Zihan Zhang. From random to explicit via subspace designs with applications to local properties and matroids. *arXiv preprint arXiv:2510.13777*, 2025. doi:10.48550/arXiv.2510.13777.

10 Martijn Brehm, Binyi Chen, Ben Fisch, Nicolas Resch, Ron D Rothblum, and Hadas Zeilberger. Blaze: fast snarks from interleaved raa codes. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 123–152. Springer, 2025. doi:10.1007/978-3-031-91134-7_5.

11 Yeyuan Chen and Zihan Zhang. Explicit folded reed-solomon and multiplicity codes achieve relaxed generalized singleton bounds. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1–12, 2025. doi:10.1145/3717823.3718114.

12 Elizabeth Crites and Alistair Stewart. On Reed–Solomon proximity gaps conjectures. *Cryptology ePrint Archive*, 2025.

13 Benjamin E Diamond and Angus Gruen. On the distribution of the distances of random words. *Cryptology ePrint Archive*, 2025.

14 Ethereum Foundation. The proximity prize: Reed–solomon challenge, 2026. Accessed: 2026-04-29. URL: https://proximityprize.org/.

15 Robert G Gallager. Low-density parity-check codes. *Low Density Parity Check Codes*, 1963.

16 Rohan Goyal and Venkatesan Guruswami. Optimal proximity gaps for subspace-design codes and (random) reed-solomon codes. *Cryptology ePrint Archive*, 2025.

17 Rohan Goyal, Venkatesan Guruswami, and Jun-Ting Hsieh. Explicit constant-alphabet subspace design codes. *arXiv preprint arXiv:2604.15218*, 2026. doi:10.48550/arXiv.2604.15218.

18 Rohan Goyal, Venkatesan Guruswami, Yihang Sun, and Mary Wootters. Locality of curve-decoding and improved proximity gaps, 2026. arXiv:2607.08516.

19 Kobi Gurkan, Andrija Novakovic, and Ron D Rothblum. Bolt: Faster snarks from sketched codes. *Cryptology ePrint Archive*, 2026.

20 Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016. doi:10.1007/s00493-014-3169-1.

21 Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear codes. *IEEE Transactions on Information Theory*, 68(2):923–939, 2021. doi:10.1109/TIT.2021.3127126.

22 Venkatesan Guruswami, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Threshold rates for properties of random codes. *IEEE Transactions on Information Theory*, 68(2):905–922, 2021. doi:10.1109/TIT.2021.3123497.

23 Venkatesan Guruswami and Atri Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Transactions on information theory*, 54(1):135–150, 2008. doi:10.1109/TIT.2007.911222.

24 Ulrich Haböck. Basefold in the list decoding regime. Cryptology ePrint Archive, Paper 2024/1571, 2024. URL: https://eprint.iacr.org/2024/1571.

25 Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved decoding of folded reed-solomon and multiplicity codes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 212–223. IEEE, 2018. doi:10.1109/FOCS.2018.00029.

26 Dmitry Krachun, Stepan Kazanin, and Ulrich Haböck. Failure of proximity gaps close to capacity. Cryptology ePrint Archive, Paper 2026/782, 2026. URL: https://eprint.iacr.org/2026/782.

27 Matan Levi, Jonathan Mosheiff, and Nikhil Shagrithaya. Random reed-solomon codes and random linear codes are locally equivalent. In *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2097–2131. IEEE, 2025. doi:10.1109/FOCS63196.2025.00112.

28 Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. Low-density parity-check codes achieve list-decoding capacity. *SIAM Journal on Computing*, 53(6):FOCS20–38–FOCS20–73, November 2021. doi:10.1137/20m1365934.

29 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*, pages 793–802, 2013. doi:10.1145/2488608.2488709.

30 Shashank Srivastava. Improved list size for folded reed-solomon codes. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2040–2050. SIAM, 2025. doi:10.1137/1.9781611978322.64.

31 Itzhak Tamo. Tighter list-size bounds for list-decoding and recovery of folded reed-solomon and multiplicity codes. *IEEE Transactions on Information Theory*, 70(12):8659–8668, 2024. doi:10.1109/TIT.2024.3402171.