Electronic Colloquium on Computational Complexity, Report No. 158 (2025)

Probabilistic Guarantees to Explicit Constructions: Local Properties of Linear Codes

Fernando Granha Jeronimo* Nikhil Shagrithaya†

October 7, 2025

Abstract

We present a general framework for derandomizing random linear codes with respect to a broad class of permutation-invariant properties, known as *local properties*, which encompass several standard notions such as distance, list-decoding, list-recovery, and perfect hashing. Our approach extends the classical Alon–Edmonds–Luby (AEL) construction through a modified formalism of local coordinate-wise linear (LCL) properties, introduced by Levi, Mosheiff, and Shagrithaya (2025). The main theorem demonstrates that if random linear codes satisfy the complement of an LCL property $\mathcal{P}$ with high probability, then one can construct explicit codes satisfying the complement of $\mathcal{P}$ as well, with an enlarged yet constant alphabet size. This gives the first explicit constructions for list recovery, as well as special cases (e.g., list recovery with erasures, zero-error list recovery, perfect hash matrices), with parameters matching those of random linear codes. More broadly, our constructions realize the full range of parameters associated with these properties at the same level of optimality as in the random setting, thereby offering a systematic pathway from probabilistic guarantees to explicit codes that attain them. Furthermore, our derandomization of random linear codes also admits efficient (list) decoding via recently developed expander-based decoders.

*University of Illinois, Urbana-Champaign. granha@illinois.edu.
†University of Michigan, Ann Arbor. nshagri@umich.edu.

ISSN 1433-8092

Contents

1 Introduction

Error-correcting codes play an important role in numerous areas [GRS23]. In addition to their more immediate applications to protect data against errors in transmission and storage, they have found various uses in diverse fields such as complexity theory [NW94], pseudorandomness [Vad12], and cryptography [GL89, YZ24]. The quest for codes approaching optimal parameter trade-offs for a given property has been a central theme of coding theory. This is typically a two-fold quest. First, one needs to understand what are the optimal parameter trade-offs. This step is often established via an existential proof, commonly using randomness and the probabilistic method. This can be far from trivial in some cases and is established via innovative randomized techniques. Secondly, one proceeds to search for an explicit construction approaching the ideal parameter trade-offs. There are many reasons why an explicit construction is desirable or needed in an application over, say, a randomized one. For instance, certifying the minimum distance and decoding can be computationally hard, making it unfit for use. One then faces the following natural question.

How to explicitly construct codes having proved an existential result?

A common challenge is that an existential proof may shed little to no light on how to explicitly construct such codes, and it may take decades, new ideas, and great ingenuity for the discovery of a corresponding explicit construction. The gulf separating existential and explicit code construction results has been prevalent throughout the history of coding theory. Shannon's seminal work [Sha48] established the existence of capacity-achieving codes through random constructions, but it was only many decades later that Arıkan introduced his breakthrough explicit construction of polar codes [Ari09]. In the case of list decoding, the capacity theorem of Zyablov and Pinsker [ZP82] was followed, decades later, by the explicit construction of Guruswami and Rudra [GR06a] (which was inspired by [PV05]). Recently, the seemingly stronger notion of list-decoding capacity was shown to indeed imply[1] Shannon's capacity on symmetric channels [PSW25]. Likewise, from the classical existential Gilbert–Varshamov bound [Gil52, Var57], it took many decades until Ta-Shma [TS17] obtained an explicit construction of binary (balanced) codes with near-optimal parameters. Despite the gulf between several existential and explicit results, we can try to remain hopeful and ask the ambitious question,

Is there a general procedure to convert existential code constructions into explicit ones?

Here, we show that this is indeed possible for a vast range of properties of random linear codes known as local properties, which include list decoding, list recovery, perfect hashing, and average pairwise distance, among many others. Random linear codes are widely used as a powerful yardstick for understanding parameter trade-offs of codes, often achieving the best possible trade-offs for many tasks. Our results provide a framework to convert existential guarantees on local properties into explicit ones, while preserving all the parameters attained by random linear codes, at the cost of increased alphabet size.

Local-to-Global Phenomena. A popular paradigm seen in several pseudorandom constructions is that of a local-to-global transfer of properties. Generally, this involves the coupling of two entities. The first is a constant-sized object possessing the property we desire, whose existence is guaranteed by probabilistic arguments and obtained via a brute-force search. The second is an infinite family of objects (typically expander graphs) whose construction is known through previous results. The novelty of such constructions often lies in identifying the appropriate manner of integrating the two objects so that the resulting construction inherits the desired property from the first object. One of the earliest constructions employing such techniques is the work of Tanner [Tan81], which details a construction of a family of error-correcting codes that involves integrating several copies of a single constant-sized code (having good rate-distance tradeoffs) with a bipartite graph having large girth. Sipser and Spielman [SS96] modified

[1] Under mild assumptions.

this construction by substituting the large girth graph for an expander graph, and using its expansion properties to prove a lower bound on the distance of the final code.

In recent years, there has been a resurgence of constructions utilizing the local-to-global phenomenon. Examples include quantum LDPC codes [PK22], locally testable codes [DEL$^+$22, PK22], unique neighbor expanders [AD24, Che25, HMMP24], lossless vertex expanders [Gol24, HLM$^+$25a, HLM$^+$25b], and erasure code ensembles [CCS25]. Our framework builds upon the Alon–Edmonds–Luby (AEL) construction, a classical instance of the local-to-global paradigm. We provide a few details about this construction.

Alon-Edmonds-Luby (AEL) Construction. The Alon-Edmonds-Luby (AEL) construction was first introduced by Alon, Edmonds, and Luby in [AEL95] to construct codes with constant alphabet size that approached the Singleton bound. The construction has three components: a constant-sized inner code having good minimum distance, found by a brute-force search, an explicit outer code having a sub-optimal rate-distance tradeoff, and a bipartite spectral expander graph. The construction can be described in two steps: the outer code is first concatenated with the inner one, upon which the symbols of codewords from the concatenated code are permuted, in a manner prescribed by the expander, to produce codewords in the final code. The expansion properties of the underlying graph are used to "lift" the minimum distance property of the inner code onto the final code.

Over the years, the AEL construction has been adapted and applied in numerous subsequent works. A recurring paradigm in these constructions is to employ a constant-sized inner code with strong parameters, combined with an outer code that may have weaker parameters, but is fully explicit. We use the term *AEL procedure* to refer to such constructions henceforth. In [GI02], Guruswami and Indyk gave explicit, linear time encodable and decodable codes for unique decoding that approached the Singleton bound, by utilizing the AEL procedure. In [KMRS17], Kopparty, Meir, Ron-Zewi, and Saraf, utilized it for constructions of locally testable codes and locally correctable codes. It was also leveraged in the work of Kopparty, Ron-Zewi, Saraf, and Wootters [KRSW23] to provide list-recoverable and list-decodable codes that matched the parameters achieved by Folded Reed-Solomon codes, while having constant alphabet size. Very recently, it was utilized by Jeronimo, Mittal, Srivastava, and Tulsiani in [JMST25] to give constructions of list-decodable codes that approached the Generalized Singleton bound over constant size alphabets.

1.1 Our Results

Local Properties. A local property, in the context of codes, is a property for which the existence of a constant number of codewords suffices as a "witness" to the code satisfying that property. For example, the complement of (ρ, L)-list-decodability is a local property, as a set of $L+1$ codewords within a Hamming ball of relative radius ρ serves as a witness for any code possessing the property. For a locality parameter L independent of the block length, a local property $\mathcal{P}$ can be informally defined by a collection of (pairwise distinct) vector sets of size L. A code is said to satisfy $\mathcal{P}$ if it contains all vectors in a vector set from the collection corresponding to $\mathcal{P}$. Typically, our objective is to understand codes that satisfy the complement of local properties, that is, codes that avoid containing any vector set from the collection defined by $\mathcal{P}$. For instance, a (ρ, L)-list-decodable code must avoid containing all pairwise distinct vector sets of size $L+1$ that lie entirely within a Hamming ball of relative radius ρ.

The concept of local properties for codes first originated in the work of Mosheiff, Resch, Ron-Zewi, Silas, and Wootters [MRR$^+$20], where they introduced the framework with the purpose of proving the existence of LDPC codes achieving list-decoding capacity. The existence follows from a more general result; the first step consists of proving a threshold result for local properties achieved by random linear codes, followed by the establishment of a transfer type result, which states that random LDPC codes achieve the same parameters for all local properties as random linear codes. The threshold result states that every local property has a threshold rate, above which random linear codes satisfy the property with exponentially high probability, and below which they do not. The framework was employed by Guruswami, Li, Mosheiff, Resch, Silas, and Wootters in [GLM$^+$22] to provide lower bounds for list sizes for list-decoding and list-recovery, and also by Guruswami and Mosheiff in [GM22] to prove that

punctured low-bias codes achieve the same parameters as random linear codes, with respect to local properties. Later, Guruswami, Mosheiff, Resch, Silas, and Wootters [GMR+22] gave a local properties framework for random codes as well.

While providing a powerful framework to investigate list-decoding and list-recovery in the low alphabet regime, the precise formulation of local properties in [MRR+20] did not allow one to study local properties in the large alphabet regime. There are two important random (linear) code families in this regime: random linear codes whose alphabet size is a large constant that is independent of the block length (but may depend on other parameters, such as gap to capacity), and random Reed-Solomon codes (whose alphabet size is at least the block length). As a consequence, these code families could not be analyzed in the context of local properties. Addressing this limitation required a new formulation tailored to the large alphabet regime, which was developed by Levi, Mosheiff, and Shagrithaya in [LMS25]. In this work, the authors establish a threshold result for local coordinate-wise linear (LCL) properties of large alphabet random linear codes, and leverage it to establish an equivalence between random Reed–Solomon codes and random linear codes, with respect to LCL properties.

As noted previously, we seek to understand codes that satisfy the complement of local properties. One approach to do so is through explicit constructions. Our main result demonstrates that it is possible to explicitly construct codes satisfying the complement of LCL properties, as long as the properties meet certain requirements.

Theorem 1.1 (Informal, see Corollary 6.1)**.** *For any reasonable LCL property $\mathcal{P}$, there exists a suitable (linear) inner code, a bipartite expander, and an outer code such that the AEL procedure, when instantiated with these components, yields an explicit linear code $\mathcal{C}_{AEL}$ that does not satisfy $\mathcal{P}$, and whose rate is arbitrarily close to the threshold rate.*

We observe that our explicit codes achieve parity with random linear codes in terms of all parameters associated with the LCL property.

Remark 1.2 (Tradeoffs)**.** *Two important tradeoffs arise in our construction: alphabet size and the underlying field of linearity. First, our construction incurs an exponential increase in alphabet size. This phenomenon is characteristic of all constructions based on the AEL procedure in the literature, and our setting is no exception. Second, while our codes are defined over a larger field, they remain linear only with respect to a subfield—namely, the field over which the inner code is defined. This limitation, however, does not pose difficulties for most applications.*

Remark 1.3 (Local Properties of Random Reed-Solomon Codes)**.** *[LMS25] also proved that random Reed-Solomon codes and random linear codes are equivalent with respect to local properties: that is, they have the same threshold rates for all reasonable local properties. This implies that our constructions also match the parameters attained by random Reed-Solomon codes for reasonable local properties, with the additional property of having constant alphabet size.*

List Decoding, List Recovery. We turn to discuss two important local properties studied in the literature: list-decoding and list-recovery. A code is (ρ, L)-list-decodable if for every vector y, the number of codewords that have (relative) Hamming distance less than ρ from y is at most L. A code $\mathcal{C} \subseteq \Sigma^n$ is (ρ, ℓ, L)-list-recoverable if for input lists $S_1, \ldots, S_n$ satisfying $|S_i| \leq \ell$ for all $i \in [n]$, we have that the output list size L is at most

$$|\{c \in \mathcal{C} \mid |\{i \in [n] \mid c[i] \in S_i\}| \geq (1-\rho)n\}| \leq L.$$

Clearly, $(\rho, 1, L)$-list recoverability is equivalent to (ρ, L)-list-decodability. One can think of these notions as generalizations of the notion of minimum distance, which requires every pair of distinct codewords to be far from one another.

List-decodable and list-recoverable codes have found uses in numerous areas of theoretical computer science, including pseudorandomness [Tre99, GUV09, LP20], compressed sensing [NPR12], and algorithms [LNNT19, DW22]. Their broad applicability has motivated the development of several explicit constructions spanning a wide range of parameter regimes. For example, list-recoverable codes have been used in constructions of list-decodable codes [GI03, GR06b, KRSW23], and locally decodable codes [HRW20].

On the other side of the coin, there has been a significant line of work investigating existential properties of linear codes through the probabilistic method. The linear structure of such codes means that the codewords of a random linear code are not mutually independent. This dependence introduced substantial obstacles in analyzing their list sizes for list decoding. For instance, the probabilistic argument of Zyablov and Pinsker [ZP82] provided list size upper bounds of $2^{O(1/\varepsilon)}$ and $O(1/\varepsilon)$ for random linear codes and random codes respectively, where ε is the gap to capacity. The exponential gap in list sizes was closed by Guruswami, Håstad, Sudan, and Zuckerman [GHSZ02], where they showed that random linear codes indeed achieve a list size of $O(1/\varepsilon)$, by means of a clever potential method argument. Their result only held in expectation, however, and not with very high probability; this was subsequently resolved by Guruswami, Håstad, and Kopparty [GHK11].

In the large alphabet regime, the works [Sud97, GS98] proved that full length Reed-Solomon codes are decodable upto the Johnson bound. However, it is known that the Johnson bound is not optimal, and an exciting line of work [ST20, GLS+24, BGM23, GZ23, AGL24] showed that randomly punctured Reed-Solomon codes approached the Generalized Singleton Bound [ST20], which is a tight bound on the radius ρ of list-decodable codes having rate R and list size L. The bound proves that

$$\rho \leq \frac{L}{L+1}(1-R).$$

In [AGL24], Alrabiah, Guruswami, and Li also proved that large alphabet random linear codes (with alphabet size $2^{O(1/\varepsilon^2)}$, where ε is the gap to capacity) approached the Generalized Singleton bound.

In the case of list-recovery, the list-recovery capacity theorem (see [Res20], Proposition 2.4.14 for a proof) states that there exist codes that are (ρ, ℓ, L)-list-recoverable, with

$$\rho \geq 1 - R - \varepsilon,$$

and $L \leq O(\ell/\varepsilon)$, as long as $|\Sigma| \geq \exp(\Omega(\log \ell/\varepsilon))$. However, this result is for random codes, and it was unclear whether random linear codes could achieve the same output list size. [LMS25] proved that this is not the case: the output list size is lower bounded by $L \geq \ell^{\Omega(R/\varepsilon)}$, and this bound was subsequently shown to hold for all linear codes by Li and Shagrithaya in [LS25].

A long line of works (e.g., [RW18], [LP20], [GLS+24], [LS25]) have studied list-recovery in the large alphabet regime. [GLS+24] showed that random Reed-Solomon codes are $(1 - R - \varepsilon, \ell, O(\ell/\varepsilon))$-list recoverable codes with rate $\Omega(\varepsilon/(\sqrt{\ell}\log(1/\varepsilon)))$. In [LS25], the authors showed for any rate R, random linear codes are $(1 - R - \varepsilon, \ell, L)$-list recoverable, where $L \leq \left(\frac{\ell}{\varepsilon}\right)^{O\left(\frac{\ell}{\varepsilon}\right)}$.

Explicit constructions of list-recoverable, list-decodable codes. We discuss results exhibiting explicit constructions of list-decoding and list-recoverable codes in the large alphabet regime. Parvaresh and Vardy [PV05] introduced the first family of error-correcting codes that was provably list-decodable beyond the Johnson bound. This was improved upon by Guruswami and Rudra in [GR06b], where they showed that Folded Reed-Solomon codes achieved list-decoding capacity, with polynomial list size. Further improvements in the analysis of the list size in a fruitful line of works [KRSW23, Tam24, Sri25, CZ25] proved that the list size matches the one implied by the Generalized Singleton bound. For list-recovery, the works of [KRSW23] and [Tam24] showed that Folded Reed-Solomon codes are $(1-R-\varepsilon, \ell, L)$-list-recoverable with output list sizes upper bounded by $(\ell/\varepsilon)^{O(\ell/\varepsilon)}$ and $(\ell/\varepsilon)^{(\log \ell/\varepsilon)}$, respectively.

In the constant-alphabet regime, existing works on capacity-achieving list-decodable and list-recoverable codes fall into two main categories. Both approaches employ the AEL procedure, where the inner code is a constant-alphabet list-decodable or list-recoverable code obtained via brute force. The distinction lies in the choice of the outer code. The first category, exemplified by [KRSW23], employs Folded Reed–Solomon codes, which are known to possess strong list-decoding and list-recovery guarantees. In contrast, the second category relies on outer codes with significantly weaker parameters, requiring only rate $1-\varepsilon$ and distance ε^3. As a result, the analysis in the latter case is more involved, but it yields improved bounds on the list sizes. Examples of works belonging to the second category are [JMST25] and [ST25].

For list-recoverability, [KRSW23] (Theorem 6.7), along with Tamo's analysis of the output list size of Folded Reed-Solomon codes [Tam24] (Theorem 4.5) gives explicit constructions of codes over alphabet Σ that are $(1-R-\varepsilon, \ell, L)$-list-recoverable, with $|\Sigma| \leq \ell^{O(1/\varepsilon^4)}$, and $L \leq (\ell/\varepsilon)^{\left(\frac{\ell}{\varepsilon}\right)^2 \cdot \log\left(\frac{\ell}{\varepsilon}\right)}$.[ST25] gives $(1-R-\varepsilon, \ell, L)$-list-recoverable codes, where $|\Sigma| = L \leq \exp\left((\ell/\varepsilon)^{O(\log(\ell/\varepsilon))}\right)$. Our result gives constructions with smaller output list sizes than either result, at the cost of increased alphabet size. All results mentioned below follow from Theorem 1.1 by specializing to the appropriate list recovery variant.

Theorem 1.4 (Informal, see Corollary 6.10, Corollary 6.9)**.** *There exist explicit constructions of linear codes of rate $R - 2\varepsilon$ that are $(1 - R - \varepsilon, \ell, L = L_{R,\varepsilon,\ell})$-list recoverable, where $L_{R,\varepsilon,\ell}$ denotes the smallest output list size attained by random linear codes of rate $R-\varepsilon$ that are list recoverable with radius $(1-R-\varepsilon)$ and input list size ℓ. The codes have an alphabet size that is at most $\exp((L/\varepsilon)^{O(L)})$.*

The nature of our result ensures that the output list sizes of our construction exactly match those attained by random linear codes. Consequently, any improvement establishing a tighter upper bound on the list sizes of random linear codes immediately carries over to our construction. In contrast to [KRSW23], which relies on an outer code with strong list-size guarantees—a potential bottleneck for future constructions which use this method—our approach requires no such assumption. Indeed, we only assume that the outer code has rate $1 - \varepsilon$ and distance ε^3.

Instantiating the codes with the best known upper bound on list sizes from [LS25], we get explicit $(1 - R - \varepsilon, \ell, L)$-list recoverable codes with rate $R - 2\varepsilon$, and $L \leq (\ell/\varepsilon)^{O(\ell/\varepsilon)}$, with alphabet size at most $\exp\left((\ell/\varepsilon)^{(\ell/\varepsilon)^{(\ell/\varepsilon)}}\right)$. The generality of our main result also yields explicit constructions for related notions such as zero-error list recovery and erasure list recovery. Zero-error list recovery is a special case of list recovery in which the decoding radius is zero. Such codes have found applications in the design of data structures for the heavy hitters problem [DW22]. In the case of erasure list recovery, some of the input lists may contain only the blank symbol, and the objective is to minimize the number of codewords that remain consistent with the non-blank input lists.

Theorem 1.5 (Informal, see Corollary 6.11)**.** *There exist explicit constructions of linear codes of rate $R - 2\varepsilon$ that are $(\ell, L = L_{R,\varepsilon,\ell})$-zero error list-recoverable, where $L_{R,\varepsilon,\ell}$ denotes the smallest output list size attained by random linear codes of rate $R - \varepsilon$ that are zero error list-recoverable with input list size ℓ.*

Theorem 1.6 (Informal, see Corollary 6.12)**.** *There exist explicit constructions of linear codes of rate $R - 2\varepsilon$ that are $(\sigma, \ell, L = L_{R,\sigma,\varepsilon,\ell})$-erasure list-recoverable, where $L_{R,\sigma,\varepsilon,\ell}$ denotes the smallest output list size attained by random linear codes of rate $R - \varepsilon$ that are erasure list-recoverable with erasure fraction σ, and input list size ℓ.*

Perfect Hash Matrices. An (n, m, t)-perfect hash matrix is defined as an $n \times m$ matrix with the following property: for every set of t columns, there exists at least one row in which the entries of those t columns are all distinct. Perfect hash matrices were first introduced by [Meh84] in the context of database management, and have since found applications in circuit complexity [NW95] and networking [LPB06]. Consequently, there has been significant work on the explicit construction of such matrices, including [FKS82, AN96, BW98, Bla00]. In particular, Blackburn and Wild [BW98] presented constructions of optimal linear perfect hash matrices, which are perfect hash matrices where the columns belong to a vector space. It is straightforward to observe that the set of codewords of a $(0, t-1, t-1)$-list-recoverable code of block length n and size m is equivalent to the set of columns of an (n, m, t)-perfect hash matrix. Thus, constructing linear perfect hash matrices is equivalent to constructing linear $(0, t-1, t-1)$-list-recoverable codes.

We recover the result of [BW98] by providing an alternate construction, which are close to optimal.

Theorem 1.7 (Informal, see Corollary 6.13)**.** *For an integer $t \geq 2$, there exist explicit constructions of linear $(n, Q^{\left(\frac{1}{t-1}-\varepsilon\right)n}, t)$-perfect hash matrices where the entries belong to $\mathbb{F}_Q$, where $Q = \exp(t^t)$.*

Lastly, we get a modest improvement in the alphabet size of explicit list-decodable codes approaching the Generalized Singleton Bound. Previously, [JMST25] constructed explicit codes with the same

parameters, except with alphabet size equal to $2^{\mathrm{poly}(L^L/\varepsilon)}$.

Theorem 1.8 (Informal, see Corollary 4.12)**.** *There exist explicit constructions of* $\left(\frac{L}{L+1}(1-R-\varepsilon), L\right)$*-list-decodable codes, having rate* $R-\varepsilon$ *and alphabet size at most* $2^{\mathrm{poly}(2^L/\varepsilon)}$*.*

The proof in [JMST25] is based on induction, and leverages the expander property of the graph used in the construction. In contrast, ours is a proof by contradiction, and relies on the sampling property of the underlying graph, thereby providing an alternative proof of essentially the same result. Although implied by Theorem 1.1, we include a proof of Theorem 1.8 as it more transparently illustrates the ideas behind the proof of the former theorem.

Efficient Decoding. Our derandomization of random linear codes from Theorem 1.1 has the added benefit of admitting efficient decoding algorithms. More precisely, the recently developed efficient (list) decoding algorithms for AEL using the Sum-of-Squares hierarchy [JMST25] and regularity lemmas [ST25, JS25] can also be used to decode our AEL based constructions. Those efficient decoding algorithms are possible thanks to the use of expander graphs in AEL rendering the decoding task tractable. In contrast, a considerable amount of evidence [DMS03, FM04, BLVW19] seems to point towards the problem of decoding for random linear codes being computationally inefficient.

1.2 Organization

In Section 2, we provide an overview of the proofs for the list-decoding case (Theorem 1.8) and the general result (Theorem 1.1). The detailed proofs appear in Section 4 and Section 5, respectively. Section 5 is independent of Section 4; therefore, readers interested solely in the general theorem's proof may proceed directly to it. Finally, the results pertaining to list-recovery variants and perfect hash matrices (Theorem 1.4, Theorem 1.5, Theorem 1.6, Theorem 1.7) are presented in Section 6.

2 Technical Overview

We begin with an overview of the proof of Theorem 1.8, which addresses the special case of list decoding. We instantiate the AEL procedure with the following three components: an inner code of constant block length, found through brute force, an explicit outer code having high rate and distance δ_{out}, where δ_{out} is a constant, and an explicit bipartite expander graph G. We note that G can equivalently be interpreted as a sampler, a viewpoint that has been leveraged in prior works (cf. [KMRS17, KRSW23]) to analyze AEL-based constructions. This sampling perspective naturally motivates a strategy for the explicit construction of codes with strong list-decoding parameters: perform a brute-force search to identify a constant block-length code with good list-decoding parameters, and then use this code as the inner code in the AEL procedure to obtain $\mathcal{C}_{\mathrm{AEL}}$. If $\mathcal{C}_{\mathrm{AEL}}$ has pairwise distinct codewords $c_1, \ldots, c_{L+1}$ close to some received word y, then the codewords agree with y at a large number of coordinates. The nature of these agreements can be encoded by an *agreement hypergraph* on $L+1$ vertices that contains n hyperedges, one for each coordinate. The hyperedge for a coordinate is simply the set of indices of codewords agreeing with y on that coordinate. It follows that if $c_1, \ldots, c_{L+1}$ agree with y on many coordinates, then the sum of the sizes of the hyperedges is large. Let these agreements be described by an agreement hypergraph denoted by $\mathcal{H}$. Upon invoking the sampling property of the graph G, it is observed that the precise pattern of these agreements is "ported over" to a significant number of vertices on the left. As a result, the projections of codewords $c_1, \ldots, c_{L+1}$ and y onto several left vertices have agreements that closely resemble the agreement pattern described by $\mathcal{H}$. Consequently, the codeword projections agree with the projection of y at a large number of coordinates within the inner code. But as the codeword projections are also codewords of the inner code, and because the inner code has good list-decoding parameters, its codewords do not have a lot of agreements with the local projection of y. Therefore, we have seemingly arrived at a contradiction.

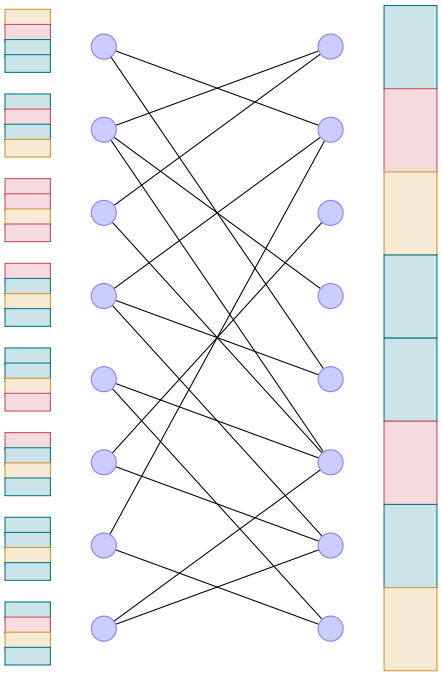

Figure 1: Several copies of the inner code witness vector sets satisfying the constraints in roughly the same proportion as the vector set on the right. Each type of constraint has a different color.

However, there is a flaw with this argument: the guarantee of the inner code applies only when the (inner) codewords are pairwise distinct. Since it cannot be guaranteed that the codeword projections are pairwise distinct, this argument fails. In order to overcome this obstacle, we use a concept known as a *weakly-partition-connected hypergraph*. This object was investigated in the context of list-decoding in [AGL24], where the authors established that (i) any agreement hypergraph with a sufficiently large hyperedge size sum contains a weakly-partition-connected hypergraph, and (ii) there exist linear codes that contain *no non-trivial* codeword sets satisfying weakly-partition-connected hypergraphs. By *non-trivial* codeword sets, we mean sets of codewords that are not all equal.

We now search for a linear code of constant block length satisfying the above property, and use it as an inner code. The analysis now proceeds in a manner analogous to the first approach. If $L+1$ codewords from $\mathcal{C}_{\text{AEL}}$ have a large number of agreements with some received word y, then the agreement hypergraph contains a weakly-partition-connected hypergraph $\mathcal{H}$. Consequently, a subset of those codewords satisfy the constraints as set forth by $\mathcal{H}$. We then use the sampling property of the graph G to port over information about the nature of the agreements to the left side, with the result that the inner codewords at several vertices on the left satisfy the constraints described by $\mathcal{H}$. From the parameters that we eventually set, it is seen that the number of such parts exceeds $n(1-\delta_{\text{out}})$ (where δ_{out} is the minimum distance of the outer code). Consequently, at least one set of inner codewords from such a part must be non-trivial, thereby contradicting our assumption on the inner code.

Before giving a proof sketch for the main result (Theorem 1.1), we present a high-level overview of LCL properties. The central idea is that the collection of vector sets corresponding to an LCL property $\mathcal{P}$ can be defined by specifying a family of linear constraint sets, known as local profiles. For a block length n and a locality parameter L, a local profile specifies a set of linear constraints, on vectors of length L, for each coordinate. The collection of vector sets associated with $\mathcal{P}$ is then defined as the column set of all matrices of dimension $n \times L$ whose rows satisfy the linear constraints described by at least one local profile associated with $\mathcal{P}$. A code is said to satisfy $\mathcal{P}$ if it contains a vector set from the collection associated with $\mathcal{P}$.

Every linear constraint set can be written down as a matrix of dimension $L \times L$. Two linear constraint

sets S and T are said to be of the same type if there exists a full rank linear transformation that maps the matrix associated with S to that associated with T. Even though the number of types can be exponential in L^2, it is independent of the block length of $\mathcal{C}_{\mathrm{AEL}}$. This fact ensures that if there are L pairwise distinct codewords in $\mathcal{C}_{\mathrm{AEL}}$ that satisfy the linear constraints described by some local profile M associated with $\mathcal{P}$, then upon arranging them in an $n \times L$ matrix, a significant fraction of the coordinates can be partitioned according to the constraint type satisfied by the row whose index is equal to the coordinate. Moreover, the fractional size of each of these sets will be a constant. Thus, the sampling property of the graph G can be utilized to port over the proportion of these types onto several projections on the left. The local projections, which are also codewords belonging to the inner code, consequently have codewords that satisfy (a close approximation of) the local profile M. If the AEL procedure were instantiated with an inner code that avoids vector sets satisfying close approximations of the local profiles associated with $\mathcal{P}$, one might hope to derive a contradiction. However, we encounter the same obstacle as before: the guarantee for the inner code applies only when the inner codewords are pairwise distinct, and we lack a mechanism to ensure that this condition is met.

We circumvent this obstacle by employing the concept of *implied types*, first introduced in [MRRZ+19] in the context of local properties. Implied types were used in their work for the purpose of pinning down the exact threshold rates of local properties of random linear codes in the low alphabet regime. They also appear implicitly in the work of [LMS25]. Informally speaking, implied vector sets can be thought of as "compressed" representations of the vector sets corresponding to a local property. For LCL properties, the corresponding notion is that of implied local profiles. Consider the implied local profile $I(M)$ corresponding to a local profile M. Denote by $V_{I(M)}, V_M$ the set of matrices associated with $I(M), M$ respectively. That is, $V_{I(M)}$ (respectively, V_M) is the set of matrices whose rows satisfy the constraints specified by $I(M)$ (respectively, M). Then, $V_{I(M)}$ can be obtained by applying an appropriate linear map on the rows of every matrix in V_M. Consequently, we see that if a linear code $\mathcal{C}$ contains a vector set that satisfies M, then by linearity, $\mathcal{C}$ also contains a vector set that satisfies $I(M)$.

Recall that we are interested in codes that satisfy the complement of an LCL property $\mathcal{P}$. We shall prove in Section 5 that if random linear codes of rate R satisfy the complement of $\mathcal{P}$, then random linear codes of rate R also satisfy the following property: with high probability, they do not contain any *non-zero matrices* satisfying any implied local profile $I(M)$, where M is any local profile associated with $\mathcal{P}$. Upon instantiating the AEL procedure with an inner code satisfying the aforementioned property, the analysis can be carried out in a manner analogous to that of the naive approach detailed above. If L pairwise distinct codewords in $\mathcal{C}_{\mathrm{AEL}}$ satisfy a local profile M associated with $\mathcal{P}$, then upon arranging these codewords in a $n \times L$ matrix and applying the appropriate linear map to each of its rows, the matrix obtained satisfies the constraints set forth by the implied local profile $I(M)$. Upon porting the constraint types to the left as before, we see that the projected matrices on to several left vertices are codewords belonging to the inner code, and moreover, they satisfy $I(M)$[2]. The parameters are instantiated in a way so as to ensure that this occurs on more than $(1 - \delta_{\mathrm{out}})n$ left vertices, and therefore there is one projected matrix that is non-zero. But because the columns of the projected matrices are codewords of the inner code, and because the inner code avoids containing all non-zero matrices satisfying $I(M)$ for any M associated with the LCL property $\mathcal{P}$, we successfully arrive at a contradiction.

3 Preliminaries

Let $[n]$ denote the set $\{1, \ldots, n\}$. For a vector $x \in (\Sigma^d)^n$ and $i \in [n]$, we denote $x(i) \in \Sigma^d$ as the ith entry of x. Furthermore, for $j \in [d]$, we denote $x(i)[j] \in \Sigma$ as the jth entry of $x(i)$. For two vectors $x, y \in \Sigma^n$, the *Hamming distance* is defined as $d(x, y) := |\{i \in [n] \colon x(i) \neq y(i)\}|$, that is, the number of indices at which x and y differ. For a set X, we denote 2^X to be the power set of X. For a matrix G, we use the notation $G[i][j]$ to index the entries of G, and use $G[i][]$ and $G[][j]$ to refer to the ith row and jth column of G, respectively. For a prime power q, let $\mathbb{F}_q$ be the finite field of order q. The notation $x \sim X$ means that the random variable x is being sampled uniformly from the set X.

[2]In reality, the projected matrices satisfy a close approximation of $I(M)$, instead of $I(M)$. Nevertheless, the inner codes we employ are chosen to be sufficiently robust to accommodate this technical subtlety.

For a vector space V, let $\mathcal{L}(V)$ denote the set of all subspaces of V. Furthermore, if V is an L-dimensional space, then we define $\mathcal{L}_{\mathsf{Dist}(V)}$ as

$$\mathcal{L}_{\mathsf{Dist}}(V) := \{U \in \mathcal{L}(V) \mid \forall 1 \leq i < j \leq L, \exists u \in U \text{ such that } u[i] \neq u[j]\}.$$

That is, $\mathcal{L}_{\mathsf{Dist}}(V)$ is the set of all subspaces that for each pair of distinct coordinates, contains at least one vector whose entries differ at those coordinates. We use $\ker \psi, \ker M$ to denote the kernel of a linear map ψ and the kernel of the linear map given by matrix M in the standard basis, respectively. The notation $\mathbf{0}$ is used to denote the all zeroes matrix.

3.1 Error-Correcting Codes

An error-correcting code $\mathcal{C}$ over alphabet Σ is a subset of Σ^n where every pair of distinct vectors in $\mathcal{C}$ has large Hamming distance. We denote by n the block length of the code. We say that a code has *(relative) distance* δ if:

$$\min_{\substack{x,y \in \mathcal{C} \\ x \neq y}} \frac{d(x,y)}{n} \geq \delta.$$

The *rate* of the code $\mathcal{C}$, usually denoted by R, is defined as:

$$R := \frac{\log_\Sigma |\mathcal{C}|}{n}.$$

In this paper, we concern ourselves with linear codes. For a finite field $\mathbb{F}_q$, a linear code $\mathcal{C}$ is a code that is a linear subspace of $\mathbb{F}_q^n$. For linear codes, the dimension of the corresponding subspace and the rate are related in the following manner:

$$R = \frac{\dim \mathcal{C}}{n}.$$

We say that a code $\mathcal{C}$ is a $[N, \delta, R]_\Sigma$ code if it has block length N, distance δ, rate R, and is over the alphabet Σ. For constants $\rho \in [0,1]$, $L \in \mathbb{N}$, a code $\mathcal{C} \subseteq \Sigma^n$ is (ρ, L)-list decodable if for every vector $y \in \Sigma^n$, we have

$$|\{c \in \mathcal{C} \mid d(c,y) \leq \rho n\}| \leq L.$$

A *random linear code* (RLC) $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate R is the kernel of a uniformly random matrix in $\mathbb{F}_q^{(1-R)n \times n}$.

Concatenated Codes. For integers N, d, $d < N$, let $\mathcal{C}_{\text{out}}$ be a $[N, R_{\text{out}}, \delta_{\text{out}}]_{\Sigma_{\text{out}}}$ code and let $\mathcal{C}_{\text{in}}$ be a $[d, R_{\text{in}}, \delta_{\text{in}}]_{\Sigma_{\text{in}}}$ code, satisfying

$$|\Sigma_{\text{out}}| = |\mathcal{C}_{\text{in}}| = |\Sigma_{\text{in}}|^{R_{\text{in}} \cdot d}. \tag{1}$$

Note that Eq. (1) allows us to construct an encoding function for $\mathcal{C}_{\text{in}}$ that is a bijection from Σ_{out} to $\mathcal{C}_{\text{in}}$, denoted by $\phi \colon \Sigma_{\text{out}} \to \mathcal{C}_{\text{in}}$. Then the concatenated code $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is a $[Nd, R_{\text{out}} \cdot R_{\text{in}}, \delta_{\text{out}} \cdot \delta_{\text{in}}]_{\Sigma_{\text{in}}}$ code defined as

$$\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}} = \left\{ v \in \Sigma_{\text{in}}^{([N] \times [d])} \mid \exists\, c \in \mathcal{C}_{\text{out}}, \left(\forall i \in [N], \phi(c[i]) \in \mathcal{C}_{\text{in}}\right) \wedge \left(\phi(c[i])\right)_{i \in [N]} = v \right\}.$$

That is, for a codeword $c \in \mathcal{C}_{\text{out}}$, we encode $c[i]$ for every $i \in [N]$ with the encoding map ϕ, thus producing $(\phi(c[i]))_{i \in [N]}$. This vector is the one obtained by concatenating the codewords of $\mathcal{C}_{\text{in}}$ corresponding to each entry of c. We perform this procedure for every codeword of $\mathcal{C}_{\text{out}}$, and collect them in the set $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$.

It is easy to see that the rate and distance of $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is equal to $R_{\text{out}} \cdot R_{\text{in}}$ and $\delta_{\text{out}} \cdot \delta_{\text{in}}$, respectively.

3.2 Alon-Edmonds-Luby (AEL) Construction

Let $G = (V_L \cup V_R, E)$ be a bipartite graph satisfying $|V_L| = |V_R| = N$, with both vertex sets having degree equal to d. For a vertex $v \in V_L \cup V_R$, denote $\Gamma(v)$ to be the neighborhood of v. For every vertex v, we will have an arbitrary, but fixed ordering on the edges incident on v. This allows us to define the ith neighbor of v: for every $i \in [d]$, $\Gamma_i(v) = w$ if $e = (v, w)$ is the ith edge of v. We will also fix an ordering, according to $[N]$, on V_L and V_R. With a slight abuse of notation, we will use ℓ (respectively, r) to refer to a vertex in V_L (respectively, V_R), and also an index in $[N]$.

Observe that the structure of G implies a bijection $\varphi_G \colon (V_L \times [d]) \to (V_R \times [d])$. Namely, $\varphi_G(\ell, i) = (r, j)$ for $i, j \in [d], \ell \in V_L$ and $r \in V_R$ if $\Gamma_i(\ell) = r$ and $\Gamma_j(r) = \ell$ both hold. For our applications, we will be dealing with matrices whose rows are indexed by (r, j) where $r \in [N]$ and $j \in [d]$. We shall now develop some notation that allows us to "project" the rows of these matrices from the right to the left.

Definition 3.1 (Projection Operation)**.** *For an integer L and a matrix $A \in \Sigma_{in}^{([N]\times[d])\times L}$, define the matrix $A^{\mathsf{proj}} \in \Sigma_{in}^{([N]\times[d])\times L}$ as follows: $\forall \ell \in [N], \forall i \in [d], \forall r \in [N], \forall j \in [d]$,*

$$A^{\mathsf{proj}}[(\ell, i)][] = A[(r, j)][] \iff \varphi_G(\ell, i) = (r, j).$$

We think of A as a matrix that creates a label (from its rows) for each outgoing edge from V_R, and A^{proj} as "collecting" the labels on the incoming edges into V_L. The graph G plays the role of "shuffling" these edge labelings. Thus, A^{proj} is created from A by permuting its rows, according to G.

Definition 3.2 (Projection on Vertices)**.** *For $\ell \in [N]$, define the matrix $A^{\mathsf{proj}}(\ell) \in \Sigma_{in}^{[d]\times L}$ to be the submatrix of A^{proj} satisfying*

$$\forall j \in [d], (A^{\mathsf{proj}}(\ell))[j][] = A^{\mathsf{proj}}[(\ell, j)][].$$

Definition 3.3 (Flattening Operation)**.** *For an integer n and a vector $h \in (\Sigma_{in}^d)^n$, we define $h_{\mathsf{fl}} \in \Sigma_{in}^{([n]\times[d])}$ to be the* flattened *vector corresponding to h: that is, $h_{\mathsf{fl}}[(i, j)] = h[i](j)$ for all $i \in [n], j \in [d]$.*

Note that h_{fl} can also be viewed as a matrix having just a single column, and therefore Definition 3.1 applies to h_{fl} as well. For codes $\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}$ where $\mathcal{C}_{\text{out}}$ is a $[N, R_{\text{out}}, \delta_{\text{out}}]_{\Sigma_{\text{out}}}$ code and $\mathcal{C}_{\text{in}}$ is a $[d, R_{\text{in}}, \delta_{\text{in}}]_{\Sigma_{\text{in}}}$ code, the code $\mathcal{C}_{\text{AEL}}(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}, G) \subseteq \left(\Sigma_{\text{in}}^d\right)^N$ is defined as follows

$$\mathcal{C}_{\text{AEL}}(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}, G) := \left\{ h \in \left(\Sigma_{\text{in}}^d\right)^N : h_{\mathsf{fl}}^{\mathsf{proj}} \in \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}} \right\}.$$

In all our constructions, the underlying graph G is obtained from the following result.

Claim 3.4 (Lemma 2.7, [KMRS17]. Also see Claim E.1, [KRSW23].)**.** *Let $\beta, \eta, \zeta \in [0, 1]$. For infinitely many integers N, there is a $d = O(1/\zeta\eta^2)$, so that the following holds. There exists a bipartite expander graph $G = (V_L, V_R, E)$ that can be constructed in time* poly(N)*, with N vertices on each side, degree d on both sides, and with the following property: for any set $Y \subseteq V_R$ of right-hand vertices with $|Y| = \beta N$, we have*

$$|\{v \in V_L : |\Gamma(v) \cap Y| < (\beta - \eta)d\}| \leq \zeta N.$$

The expander graph constructed in the preceding claim is a sampler in the sense that, for every sufficiently large subset of right vertices Y, a substantial fraction of left vertices have neighborhoods intersecting with Y in a proportion that is roughly equal to the density of Y.

Observation 3.5 (Explicitness of AEL Procedure)**.** *We observe that $\mathcal{C}_{AEL}(\mathcal{C}_{out}, \mathcal{C}_{in}, G) \subseteq \left(\Sigma_{in}^d\right)^N$ can be constructed in time* poly(N)*—that is, in time polynomial in the block length of the code—provided that each of the three components, namely, $\mathcal{C}_{in}$, $\mathcal{C}_{out}$, and G, can themselves be constructed in time* poly(N).

4 Warm Up: Construction of List-Decodable Codes

In this section, we give a proof for Theorem 1.8. We define the required concepts and definitions in Section 4.1, and state the proof in Section 4.2.

4.1 Preliminaries

A hypergraph $\mathcal{H} = (V, \mathcal{E})$ consists of a vertex set V, and a collection $\mathcal{E}$ of subsets of V. The subsets are known as hyperedges. For every hyperedge e, we define the weight of e as $\mathsf{wt}(e) := \max(|e| - 1, 0)$. Furthermore, the weight of the set of hyperedges is defined as the sum of hyperedge weights. That is, $\mathsf{wt}(\mathcal{E}) := \sum_{e \in \mathcal{E}} wt(e)$.

Definition 4.1 (Agreement Hypergraph)**.** *For an integer $n \in \mathbb{N}$ and a set of vectors $y, c_1, \ldots, c_t \in \Sigma^n$, we define an agreement hypergraph $\mathcal{H}(y, c_1, \ldots, c_t) = ([t], \mathcal{E})$ as follows: For each $i \in [n]$, construct hyperedge $\varepsilon_i \subseteq [t]$ by including the indexes of all vectors that agree with y at the ith coordinate. That is, $e_i := \{j \in [t] \colon c_j[i] = y[i]\}$.*

Definition 4.2 (Weak Partition Connectivity)**.** *For an integer $n \in \mathbb{N}$, and $R \in [0, 1]$, we say that a hypergraph $H = (V, \mathcal{E})$ is (R, n)-weakly-partition-connected if for every partition $\mathcal{P}$ of the vertex set V, the following holds:*

$$\sum_{e \in \mathcal{E}} \max\{|\mathcal{P}(e)| - 1, 0\} \geq Rn(|\mathcal{P}| - 1), \tag{2}$$

where $|\mathcal{P}|$ denotes the number of parts in $\mathcal{P}$ and $|\mathcal{P}(e)|$ denotes the number of parts that intersect non-trivially with e.

From Lemma 2.3 in [AGL24], we see that the agreement hypergraph corresponding to a bad list-decoding configuration contains a weakly-partition-connected sub-hypergraph.

Lemma 4.3 (Lemma 2.3, [AGL24])**.** *Suppose that for vectors $c_1, \ldots, c_{L+1} \in \Sigma^n$, the average Hamming distance of these vectors from a vector $y \in \Sigma^n$ is at most $\frac{L}{L+1}(1-R)n$. Then, for some subset $J \subseteq [L+1]$ where $|J| \geq 2$, the agreement hypergraph corresponding to vectors y and $\{c_j \colon j \in J\}$ is (R, n)-weakly-partition-connected.*

Proof. Let $\mathcal{H} = ([L+1], \mathcal{E})$ be the agreement hypergraph corresponding to vectors y and $c_1, \ldots, c_{L+1} \in \Sigma^n$. Since

$$\sum_{i \in [n]} \frac{(L+1) - |e_i|}{L+1} = \sum_{j \in [L+1]} \frac{\sum_{i \in [n]} \mathbb{1}[c_j(i) \neq y(i)]}{L+1} = \sum_{j \in [L+1]} \frac{d(y, c_i)}{L+1} \leq \frac{L}{L+1}(1-R)n, \tag{3}$$

we have

$$\mathsf{wt}(\mathcal{E}) = \sum_{i \in [n]} \mathsf{wt}(e_i) \geq -n + \sum_{i \in [n]} |e_i| \geq LRn.$$

The last inequality follows from Eq. (3).

Let $J \subseteq [L+1]$ be an inclusion minimal subset with $|J| \geq 2$ such that

$$\sum_{i \in [n]} \mathsf{wt}(e_i \cap J) \geq LRn. \tag{4}$$

The existence of such a J follows from the fact that $J = [L+1]$ satisfies Eq. (4). Let $\mathcal{H}' = (J, \mathcal{E}')$ be the hypergraph with vertex set J and edge set $\mathcal{E}' := \{J \cap e \mid e \in \mathcal{E}\}$.

We now prove that $\mathcal{H}'$ is (R, n)-weakly-partition-connected. Observe that Eq. (2) follows from Eq. (4) when $\mathcal{P}$ is the trivial partition with a single part. Now, consider a non-trivial partition $\mathcal{P} = P_1 \sqcup \ldots \sqcup P_p$.

We have

$$\begin{aligned}
\sum_{e\in\mathcal{E}'} \max\{|\mathcal{P}(e)|-1,0\} &= \sum_{\substack{e\in\mathcal{E}'\\ e\neq\emptyset}} \Big(-1+\sum_{b\in[p]} \mathbb{1}[|e\cap P_b|>0]\Big)\\
&= \sum_{\substack{e\in\mathcal{E}'\\ e\neq\emptyset}} \Big((|e|-1)-\sum_{b\in[p]} (|e\cap P_b|-\mathbb{1}[|e\cap P_b|>0])\Big)\\
&= \sum_{\substack{e\in\mathcal{E}'\\ e\neq\emptyset}} \Big(\max(|e|-1,0)-\sum_{b\in[p]} \max(|e\cap P_b|-1,0)\Big)\\
&= \sum_{e\in\mathcal{E}'} \mathsf{wt}(e)-\sum_{b\in[p]}\sum_{e\in\mathcal{E}'} \mathsf{wt}(e\cap P_b)\\
&\geq (|J|-1)Rn-\sum_{b\in[p]} (|P_b|-1)Rn\\
&= (p-1)Rn = (|\mathcal{P}|-1)Rn,
\end{aligned}$$

where the last inequality follows from the inclusion minimality of set J, and the fact that every P_b is a strict subset of J. □

Lemma 2.14 of [AGL24] proves the robustness of weakly-partition-connected hypergraphs to hyperedge deletions:

Lemma 4.4 (Robustness of weakly-partition-connected hypergraphs (Lemma 2.14, [AGL24]))**.** *Let $\mathcal{H}=([t],\mathcal{E})$ be a $(R+\varepsilon,n)$-weakly-partition-connected hypergraph. Then for all sets $\mathcal{E}'\subseteq\mathcal{E}$, $|\mathcal{E}'|\leq\varepsilon n$, the hypergraph $\mathcal{H}'=([t],\mathcal{E}\setminus\mathcal{E}')$ is (R,n)-weakly-partition-connected.*

Proof. Consider any partition $\mathcal{P}$ of $[t]$. We have

$$\begin{aligned}
\sum_{e\in\mathcal{E}\setminus\mathcal{E}'} \max(|\mathcal{P}(e)|-1,0) &= \sum_{e\in\mathcal{E}} (|\mathcal{P}(e)|-1,0)-\sum_{e\in\mathcal{E}'} (|\mathcal{P}(e)|-1,0)\\
&\geq (R+\varepsilon)(|\mathcal{P}|-1)-|\mathcal{E}'|\,(|\mathcal{P}|-1)\\
&= Rn(|\mathcal{P}|-1).
\end{aligned}$$

□

We now proceed to define another object used in the proofs of [AGL24]: a Reduced Intersection Matrix (RIM). Although several versions of the RIM appeared in other works such as [ST20, GLS+24, BGM23], this variant was first introduced in [GZ23].

In order to define the RIM, we need to set up some notation. For any integers k,m with $k\leq m$ and a finite field $\mathbb{F}_q$, define the symbolic matrix $\mathcal{G}\in\mathbb{F}_q(X_{1,1},\ldots,X_{k,n})^{k\times n}$ as

$$\mathcal{G} := \begin{bmatrix} X_{1,1} & \ldots & X_{1,n}\\ \vdots & \ddots & \vdots\\ X_{1,k} & \ldots & X_{k,n}\end{bmatrix}.$$

The ith column of $\mathcal{G}$ is denoted by $\mathcal{G}_i=[X_{1,i},\ldots,X_{k,i}]$.

Definition 4.5 (Reduced Intersection Matrix)**.** *The Reduced Intersection Matrix $\mathsf{RIM}_{\mathcal{H}}$ associated with a hypergraph $\mathcal{H}=([t],\mathcal{E}=(e_1,\ldots,e_n))$ is a $\mathsf{wt}(\mathcal{E})\times(t-1)k$ matrix with entries from $\mathbb{F}_q(X_{1,1},\ldots,X_{k,n})^{k\times n}$. We construct $\mathsf{RIM}_{\mathcal{H}}$ as follows: for every hyperedge $e_i\in\mathcal{E}$ containing vertices $j_1<j_2<\ldots<j_{|e_i|}$, add $\mathsf{wt}(e_i)=|e_i|-1$ rows for each $\ell=2,\ldots,|e_i|$. Each row has $(t-1)$ segments, each of length k, of the form $r_{i,\ell}=(r^{(1)},\ldots,r^{(t-1)})$. The segments are defined as follows:*

- *If $j = j_1$, then $r^{(j)} = \mathcal{G}_i^\top = [X_{1,i}, \ldots, X_{k,i}]$.*
- *If $j = j_\ell$, and $j_\ell \neq t$, then $r^{(j)} = \mathcal{G}_i^\top = -[X_{1,i}, \ldots, X_{k,i}]$.*
- *Otherwise, $r^{(j)} = 0^k$.*

Definition 4.6. *(Substituted RIM) For any matrix $G \in \mathbb{F}_q^{k \times n}$ and a $\mathsf{RIM}_{\mathcal{H}}$ associated with the hypergraph $\mathcal{H} = ([t], \mathcal{E})$, the* substituted Reduced Intersection Matrix*, denoted by $\mathsf{RIM}_{\mathcal{H}}(G)$ is defined to be the matrix in $\mathbb{F}_q^{\mathsf{wt}(\mathcal{E}) \times (t-1)k}$ obtained by substituting every indeterminate symbol in $\mathsf{RIM}_{\mathcal{H}}$ with the corresponding entry from G. That is, we obtain $\mathsf{RIM}_{\mathcal{H}}(G)$ from $\mathsf{RIM}_{\mathcal{H}}$ by replacing, for every $i \in [k], j \in [n]$, $X_{i,j}$ with $G[i][j]$.*

Lemma 4.7 (Column Rank of Reduced Intersection Matrices (Lemma 4.2, [AGL24]))**.** *Let $\mathcal{H}$ be the agreement hypergraph For a vector $y \in \mathbb{F}_q^n$, and a matrix $G \in \mathbb{F}_q^{k \times n}$, suppose that the corresponding agreement hypergraph for y and codewords $c_1, \ldots, c_t$ generated by G is equal to $\mathcal{H}$. Moreover, let $c_1, \ldots, c_t$ satisfy the property that they are not all equal. Then, the substituted reduced intersection matrix $\mathsf{RIM}_{\mathcal{H}}(G)$ does not have full column rank.*

Proof. Let $m_1, \ldots, m_t$ be the message vectors for codewords $c_1, \ldots, c_t$. That is, $c_i = m_i \cdot G$ for all $i \in [t]$. Then, we see that

$$\mathsf{RIM}_{\mathcal{H}}(G) \cdot \begin{bmatrix} m_1 - m_t \\ \vdots \\ m_{t-1} - m_t \end{bmatrix} = 0. \tag{5}$$

Because $c_1, \ldots, c_t$ were not all equal, the same applies for $m_1, \ldots, m_t$. Therefore the vector multiplied to $\mathsf{RIM}_{\mathcal{H}}(G)$ in Eq. (5) is non-zero. Because the agreement hypergraph corresponding to y and codewords $c_1, \ldots, c_t$ is the same as $\mathcal{H}$, every constraint in $\mathsf{RIM}_{\mathcal{H}}(G)$ is satisfied, and hence, $\mathsf{RIM}_{\mathcal{H}}(G)$ does not have full column rank. □

The following lemma, which is the crux of the main result of [AGL24], states that for a uniformly random matrix $G \in \mathbb{F}_q^{k \times n}$, the corresponding substituted reduced intersection matrices associated with all weakly partition-connected agreement hypergraphs have full column rank.

Lemma 4.8 (Existence of Linear Codes whose RIMs have Full Column Rank (Theorem 1.3, Lemma 4.6, [AGL24]))**.** *For integers n, L, where L is a constant independent of n, rate $R \in [0, 1]$ and a sufficiently small $\varepsilon > 0$, alphabet size $q \geq 2^{10L/\varepsilon}$, with probability at least $1 - 2^{-Ln}$, a uniformly random matrix $\mathbf{G} \in \mathbb{F}_q^{Rn \times n}$ has the following property: for every agreement hypergraph $\mathcal{H}$ on a vertex set of size $\leq L + 1$ that is $(R + \varepsilon/2, n)$-weakly-partition-connected, the substituted reduced intersection matrix $\mathsf{RIM}_{\mathcal{H}}(G)$ has full column rank.*

The proof of Lemma 4.8 follows from the proof of Theorem 1.3 in [AGL24]. We record an important corollary of the lemma.

Corollary 4.9. *For integers n, L such that $n > 1/L$, rate $R \in [0, 1]$ and a sufficiently small $\varepsilon > 0$, there exist $\mathbb{F}_q$-linear codes $\mathcal{C} \subseteq \mathbb{F}_q^n$, where $q = 2^{10L/\varepsilon}$, with the following property: for every agreement hypergraph $\mathcal{H}$ on a vertex set of size $\leq L + 1$ that is $(R + \varepsilon/2, n)$-weakly-partition-connected, the only set of codewords in $\mathcal{C}$ that satisfy every agreement in $\mathcal{H}$ is the trivial set of codewords that are all equal.*

Proof. By Lemma 4.8, we see that with a non-zero probability, a uniformly random matrix $\mathbf{G} \in \mathbb{F}_q^{Rn \times n}$ has the property of having full column rank on every $\mathsf{RIM}_{\mathcal{H}}(G)$, for all agreement hypergraphs $\mathcal{H}$ that are $(R + \varepsilon/2, n)$-weakly-partition-connected. Thus, there exists at least one matrix $G \in \mathbb{F}_q^{Rn \times n}$ satisfying the property. The contrapositive of Lemma 4.7 then implies that the only set of codewords in G that simultaneously satisfy every agreement in $\mathcal{H}$ is the set of codewords that are all equal. The corollary follows by considering the code $\mathcal{C}$ generated by the rows of G. □

4.2 Proof for List-Decoding

Recall that the bipartite graph G from Claim 3.4 has N vertices on each side. For the sake of simplifying the exposition, we omit floor and ceiling notation in the proof. In order to utilize the results listed in Section 4.1, we require the inner code to be defined over a finite field $\mathbb{F}_q$, and so we take $\Sigma_{\text{in}} = \mathbb{F}_q$. Consequently, we may interpret $\mathcal{C}_{\text{AEL}} \subseteq \left(\mathbb{F}_q^d\right)^N$ as a code over the extension field $\mathbb{F}_Q$, where $Q = q^d$.

Fix a list size parameter L, rate R, and slack $\varepsilon > 0$. Let $q = 2^{10L/\varepsilon}$. Recall that δ_{out} is defined to be the distance of the outer code $\mathcal{C}_{\text{out}}$, and take $\eta = \varepsilon/2^{(L+3)}$ and $\zeta < \delta_{\text{out}}/2^{(L+1)}$. Take $d = \max(O(1/\zeta\eta^2), 1/L)$. We take $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_q^d$ to be a $\mathbb{F}_q$-linear code having rate $R_{\text{in}} = R$. Let G be the bipartite graph from Claim 3.4 having N vertices on each side, and every vertex having degree $d \geq O(1/\zeta\eta^2)$. Let $\mathcal{C}_{\text{out}} \subseteq (\mathbb{F}_q^{R_{\text{in}}d})^N$ be the outer code. Note that by our definition, ϕ is now a map of the form $\phi\colon \mathbb{F}_q^{R_{\text{in}}d} \to \mathcal{C}_{\text{in}}$. We require $\mathcal{C}_{\text{out}}$ to be an $\mathbb{F}_q$-linear code and ϕ to be a $\mathbb{F}_q$-linear map.

We now prove the following result:

Theorem 4.10. *Let $\mathcal{C}_{in}, \mathcal{C}_{out}, G$ be as defined above. Furthermore, if $\mathcal{C}_{in}$ is a code satisfying the property in Corollary 4.9, then $\mathcal{C}_{AEL} \subseteq \left(\mathbb{F}_q^d\right)^N$ is an $\mathbb{F}_q$-linear code that is $\left(\frac{L}{L+1}(1-R-\varepsilon), L\right)$ average-radius list-decodable.*

Since both $\mathcal{C}_{\text{out}}$ and the map ϕ are $\mathbb{F}_q$-linear, it follows that $\mathcal{C}_{\text{AEL}}$ is itself an $\mathbb{F}_q$-linear code.

For the sake of contradiction, assume there is a vector $y \in (\mathbb{F}_q^d)^N$ and pairwise distinct codewords $c_1, \ldots, c_{L+1} \in \mathcal{C}_{\text{AEL}}$ such that their average Hamming distance from y is less than $\frac{L}{L+1}(1-R-\varepsilon)N$. That is

$$\sum_{i \in [L+1]} \frac{d(y, c_i)}{L+1} < \frac{L}{L+1}(1-R-\varepsilon)N.$$

Then, Lemma 4.3 implies that there is a subset $J \subseteq [L+1]$, $|J| \geq 2$ such that the agreement hypergraph corresponding to vectors y and $\{c_j : j \in J\}$ is $(R+\varepsilon, n)$-weakly-partition-connected. We state a lemma which proves that the local projections of the vectors y and $\{c_j : j \in J\}$ onto many left vertices also yields a weakly-partition-connected hypergraph. Let $y^{\mathsf{proj}} := y_{\mathsf{fl}}^{\mathsf{proj}}$ and $c_j^{\mathsf{proj}} := (c_j)_{\mathsf{fl}}^{\mathsf{proj}}$ for every $j \in J$ denote the flattened projections of y and the codewords $\{c_j : j \in J\}$, respectively. These vectors are obtained by performing a flattening operation (see Definition 3.3), followed by a projection operation (see Definition 3.1).

Lemma 4.11 (Local Projections are Weakly-Partition-Connected)**.** *If the agreement hypergraph corresponding to a vector $y \in (\mathbb{F}_q^d)^N$ and codewords $c_1, \ldots, c_t \in \mathcal{C}_{AEL}$ is $(R+\varepsilon, N)$-weakly-partition-connected, there exists a set $L^* \subseteq V_L$ satisfying $|L^*| > (1-\delta_{out})N$ such that for every $\ell \in L^*$, the agreement hypergraph corresponding to the local projections $y^{\mathsf{proj}}(\ell) \in \mathbb{F}_q^d$, $c_1^{\mathsf{proj}}(\ell), \ldots, c_t^{\mathsf{proj}}(\ell) \in \mathcal{C}_{in}$ is $(R + \frac{\varepsilon}{2}, d)$-weakly-partition-connected.*

We first prove Theorem 4.10 using this lemma.

Proof of Theorem 4.10 using Lemma 4.11. Fix a pair of codewords c_i, c_j, such that $i, j \in J$, and i, j are distinct (such a pair of indices exist since $|J| \geq 2$). Define

$$S := \left\{\ell \in V_L : c_i^{\mathsf{proj}}(\ell) \neq c_j^{\mathsf{proj}}(\ell)\right\}.$$

Because c_i and c_j are distinct codewords belonging to $\mathcal{C}_{\text{AEL}}$, we know that by construction, $|S| \geq \delta_{\text{out}}N$. This is because $c_i^{\mathsf{proj}}(\ell) \neq c_j^{\mathsf{proj}}(\ell)$ if and only if $\phi^{-1}(c_i^{\mathsf{proj}}(\ell)) \neq \phi^{-1}(c_j^{\mathsf{proj}}(\ell))$, and this holds for exactly those vertices on which the codewords in $\mathcal{C}_{\text{out}}$ corresponding to c_i, c_j differ. Upon applying Lemma 4.11 to the agreement hypergraph corresponding to vectors y and $\{c_j : j \in J\}$, there exists at least one left vertex $\ell \in S \cap L^*$. Because $\ell \in S$, $c_i^{\mathsf{proj}}(\ell) \neq c_j^{\mathsf{proj}}(\ell)$, and therefore, the local codewords $\left\{c_j^{\mathsf{proj}}(\ell) : j \in J\right\}$ are not all equal. Since $\ell \in L^*$, the agreement hypergraph corresponding to the local projections $y^{\mathsf{proj}}(\ell) \in \mathbb{F}_q^d$, $c_1^{\mathsf{proj}}(\ell), \ldots, c_T^{\mathsf{proj}}(\ell) \in \mathcal{C}_{\text{in}}$ is $(R + \frac{\varepsilon}{2}, d)$-weakly-partition-connected. This contradicts the property of $\mathcal{C}_{\text{in}}$ as described in Corollary 4.9. □

Proof of Lemma 4.11. We will associate a subset of $[t]$ with every $r \in V_R$, which we shall refer to as the *type* of r. We say that r is of *type* T for some subset $T \subseteq [t]$ if the set of indices of all codewords agreeing with y at r is equal to T. More formally, define $\mathsf{type}\colon V_R \to 2^{[t]}$, with

$$\mathsf{type}(r) := \{i \in [t] : c_i(r) = y(r)\}.$$

For $\beta \in [0,1]$, we say that a *type* $T \subseteq [t]$ is β-dense if $\mathsf{type}(r) = T$ for more than βN vertices $r \in V_R$. We shall only consider types that are β-dense for $\beta := \varepsilon/2^{(t+2)}$. Define $D_\beta \subseteq 2^{[t]}$ to be the set of all subsets of $[t]$ that are β-dense. Then the number of right vertices whose type is *not* β-dense is at most

$$|\{r \in V_R : \mathsf{type}(r) \notin D_\beta\}| \leq |2^{[t]} \setminus D_\beta| \cdot \beta N \leq \frac{\varepsilon N}{4}.$$

By the robustness of weakly-partitioned-hypergraphs (c.f. Lemma 4.4), we see that the agreement hypergraph created by deleting all hyperedges corresponding to types that are not β-dense is still $(R + \frac{3\varepsilon}{4}, N)$-weakly-partition-connected. Denote this agreement hypergraph by $\mathcal{H}'$. Combining the fact that (i) $\mathcal{H}'$ is $(R + 3\varepsilon/4, N)$-weakly-partition-connected and (ii) all hyperedges in $\mathcal{E}_{\mathcal{H}'}$ are associated with vertices whose type belongs to D_β, we see that for every partition $\mathcal{P}$ of $[t]$,

$$\sum_{T \in D_\beta} \sum_{\substack{r \in V_R \\ \mathsf{type}(r) = T}} \max\{|\mathcal{P}(T)| - 1, 0\} = \sum_{e \in \mathcal{E}_{\mathcal{H}'}} \max\{|\mathcal{P}(e)| - 1, 0\} \geq \left(R + \frac{3\varepsilon}{4}\right) N(|\mathcal{P}| - 1).$$

Denote the set of all vertices of type T by $S_T \subseteq V_R$, and denote its density by $\mu(S_T) := |S_T|/|V_R| = |S_T|/N \geq \beta$. Then,

$$\sum_{T \in D_\beta} \mu(S_T) \cdot \max\{|\mathcal{P}(T)| - 1, 0\} \geq \left(R + \frac{3\varepsilon}{4}\right)(|\mathcal{P}| - 1). \tag{6}$$

Fix some β-dense type $T \subseteq [t]$. Using the sampling property of the graph G, we now show that for a large number of left vertices $\ell \in V_L$, the fraction of edges entering ℓ that arise from right vertices of type T is roughly the same as the fraction of type T vertices on the right side. Quantitatively, by Claim 3.4, we see that

$$|\{\ell \in V_L : |\Gamma(\ell) \cap S_T|/d \leq \mu(S_T) - \eta\}| \leq \zeta N.$$

By applying a simple union bound argument over all β-dense types, the following holds

$$|\{\ell \in V_L : \exists T \in D_\beta : |\Gamma(\ell) \cap S_T|/d \leq \mu(S_T) - \eta\}| \leq |D_\beta|\,\zeta N.$$

Thus for at least $(1 - |D_\beta|\,\zeta)N > (1 - \delta_{\text{out}})N$ left vertices $\ell \in V_L$,

$$\forall T \in D_\beta : |\Gamma(\ell) \cap S_T| > (\mu(S_T) - \eta)d. \tag{7}$$

Denote this set by $L^* \subseteq V_L$. For a vertex $\ell \in L^*$ and a type $T \in D_\beta$, observe that for all indices $i \in [d]$ for which $\mathsf{type}(\Gamma_i(\ell))$ belongs to T, the local codewords $\left\{c_j^{\mathsf{proj}}(\ell) : j \in T\right\}$ agree with $y^{\mathsf{proj}}(\ell)$ at coordinate i. Therefore, we can speak of types for local coordinates as well, and by a slight abuse of notation, define $\mathsf{type}(i) := \mathsf{type}(\Gamma_i(\ell))$.

Additionally, the proportion of those coordinates is roughly equal to the proportion of right vertices on which the codewords $\{c_j : j \in T\}$ agree with y. Thus, we see that the agreements corresponding to all hyperedges that occur on more than β fraction of the right vertices are "ported over" to the vertices in L^*. Informally, this says that the agreement hypergraph corresponding to the local projections at every $\ell \in L^*$ is roughly equivalent to $\mathcal{H}'$. We will now prove this in a formal manner, in order to conclude that these local agreement hypergraphs are weakly-partition-connected.

Turning our attention over to a fixed $\ell \in L^*$ and denoting the set of hyperedges in the local agreement hypergraph of ℓ by $\mathcal{H}_\ell = ([t], \mathcal{E}_\ell)$, we see by Eq. (7) that for every partition $\mathcal{P}$ of $[t]$,

$$\begin{aligned}\sum_{e\in\mathcal{E}_\ell} \max\{|\mathcal{P}(e)|-1,0\} &\geq \sum_{T\in D_\beta} \sum_{\substack{i\in[d]\\ \mathsf{type}(i)=T}} \max\{|\mathcal{P}(T)|-1,0\} \\ &> \sum_{T\in D_\beta} (\mu(S_T)-\eta)d\cdot\max\{|\mathcal{P}(T)|-1,0\}.\end{aligned}$$

The last term can be expanded as

$$\sum_{T\in D_\beta} \mu(S_T)d\cdot\max\{|\mathcal{P}(T)|-1,0\} - \sum_{T\in D_\beta} \eta d\cdot\max\{|\mathcal{P}(T)|-1,0\}.$$

The first term is at least $(R+3\varepsilon/4)d(|\mathcal{P}|-1)$ by virtue of Eq. (6), and the second term is at most $\frac{\varepsilon}{4}d(|\mathcal{P}|-1)$, as $|D_\beta| \leq 2^{L+1}$ and $\eta = \varepsilon/2^{(L+3)}$. Putting everything together, we get

$$\sum_{e\in\mathcal{E}_\ell} \max\{|\mathcal{P}(e)|-1,0\} \geq (R+\varepsilon/2)d(|\mathcal{P}|-1).$$

Thus, $\mathcal{H}_\ell$ is $(R+\varepsilon/2, d)$-weakly-partition-connected. □

Corollary 4.12. *Let $\mathcal{C}_{in}, \mathcal{C}_{out}, G$ be as defined in Theorem 4.10. Furthermore, let $\mathcal{C}_{out}$ be a code with rate $R_{out} = 1-\varepsilon$ and distance $\delta_{out} \geq \varepsilon^3$. Denote $Q := q^d$. Then $\mathcal{C}_{AEL} \subseteq (\mathbb{F}_Q)^N$ is an $\mathbb{F}_q$-linear code that is $\left(\frac{L}{L+1}(1-R-\varepsilon), L\right)$ average-radius list-decodable, with rate $R_{AEL} \geq R-\varepsilon$, where d satisfies*

$$d \leq O(2^{3L}/\varepsilon^5).$$

Thus, $Q = \exp(2^{O(L)}/\varepsilon^5)$. Moreover, $\mathcal{C}_{AEL}$ is constructible in time $\mathrm{poly}(N)$.

Proof. The $\mathbb{F}_q$-linearity of $\mathcal{C}_{\mathrm{AEL}}$ its list-decodability parameters are proven in Theorem 4.10. The rate is given by

$$R_{\mathrm{AEL}} = \frac{\log_Q(|\mathcal{C}_{\mathrm{AEL}}|)}{N}.$$

Indeed, it is easy to see that $|\mathcal{C}_{\mathrm{AEL}}| = |\mathcal{C}_{\mathrm{out}}| = q^{R_{\mathrm{in}}R_{\mathrm{out}}dN}$. Because $Q = q^d$, a simple calculation gives $R_{\mathrm{AEL}} = R_{\mathrm{in}}\cdot R_{\mathrm{out}} = R\cdot(1-\varepsilon) > R-\varepsilon$. The value for d is obtained by recalling that $d = \max(O(1/\zeta\eta^2), 1/L)$, $\eta = \varepsilon/(2^{L+3})$, $\zeta = \delta_{\mathrm{out}}/2^{(L+1)}$, and plugging in the value for δ_{out} in ζ. The value for Q is obtained by recalling the fact that $q = 2^{10L/\varepsilon}$ from Corollary 4.9.

We note that explicit constructions of $\mathbb{F}_q$-linear codes having rate $1-\varepsilon$ and distance ε^3 that are constructible in time $\mathrm{poly}(N)$ can be obtained by using Tanner codes (see Corollary 11.4.8 in [GRS23]). Moreover, the graph G can be constructed in time $\mathrm{poly}(N)$, by Claim 3.4. By Corollary 4.9, $\mathcal{C}_{\mathrm{in}}$ exists and has a block length independent of N, therefore it can be found through brute force in constant time. Upon invoking Observation 3.5, we see that $\mathcal{C}_{\mathrm{AEL}}(\mathcal{C}_{\mathrm{in}}, \mathcal{C}_{\mathrm{out}}, G)$ can be constructed in time $\mathrm{poly}(N)$. □

5 Constructions for Local Properties

A *code property* $\mathcal{P}$ can informally be defined as a family of codes sharing some common characteristics. We focus on code properties that are

(i) local,

(ii) monotone-increasing, and

(iii) independent of coordinate permutations.

A local code property, informally speaking, is defined by the inclusion of bad sets of vectors. A *monotone-increasing* code property is one for which the following is true: if $\mathcal{C}$ is in $\mathcal{P}$, then every $\mathcal{C}'$ for which $\mathcal{C}' \supseteq \mathcal{C}$ holds, also lies in $\mathcal{P}$. We say that a property $\mathcal{P}$ is independent of coordinate permutations if for every code $\mathcal{C} \in \mathcal{P}$, $\mathcal{C}'$ also lies in $\mathcal{P}$, where $\mathcal{C}'$ is obtained by applying a permutation on the coordinates to every codeword in $\mathcal{C}$. An example of local, monotone-increasing code property that is independent of coordinate permutations is the complement of (ρ, L)-list-decodability.

Before we give a formal definition of local properties studied in this paper, we discuss aspects of similar definitions in previous works. Local properties were first defined in [MRR+20] in order to prove threshold type results for random linear codes, and to prove that LDPC codes achieve the same parameters as random linear codes. Their work focused on proving results for the small alphabet regime, and given their definition of local properties, it was not possible to extend the results to the large alphabet regime. This was accomplished in [LMS25], where the authors provided a new definition suitable for the large alphabet regime.

Since our work studies codes in the latter regime, we choose to adapt the definition in [LMS25], and give a brief overview of the same before discussing our modifications. For a locality parameter $L \in \mathbb{N}$ and block length n, a *L-local coordinate wise linear* (L-LCL) property $\mathcal{P}$ is defined as a collection of local profiles. A local profile is an ordered tuple of subspaces $\mathcal{V} = (\mathcal{V}_1, \ldots, \mathcal{V}_n)$, where $\mathcal{V}_i \in \mathcal{L}(\mathbb{F}_q^L)$ for each $i \in [n]$. A matrix $A \in \mathbb{F}_q^{n \times L}$ is said to be *contained in* $\mathcal{V}$ if the ith row of A belongs to $\mathcal{V}_i$, for all i. We say that a code satisfies property $\mathcal{P}$ if there exists a matrix $A \in \mathbb{F}_q^{n \times L}$ such that the columns of A are pairwise distinct codewords in $\mathcal{C}$, and A is contained in some local profile belonging to the collection of local profiles associated with $\mathcal{P}$.

We now state our modifications, and the justifications for introducing them. In a nutshell, our modifications are concerned with reconciling the (seemingly) different definitions of LCL properties for codes having differing field sizes and block lengths. The modifications are necessary in order to talk about the LCL properties being satisfied by the constant-sized inner code, while also being satisfied by the infinite code family produced by the AEL construction. Recall that in addition to having differing block lengths, these two codes also have different alphabet (field) sizes.

Our first modification is to define local profiles as tuples of matrices, instead of subspaces. We then require that in order for a matrix A to be contained in a local profile, each row of A should lie in the kernel of the matrix corresponding to the row index. This is necessary in order to address the problem of differing field sizes. Recall that the alphabet of $\mathcal{C}_{\mathrm{AEL}}$ is Σ_{in}^d, where Σ_{in} is the alphabet of the inner code. Thus, if our inner code is over a field $\mathbb{F}_q$, one can view the alphabet of $\mathcal{C}_{\mathrm{AEL}}$ as being equal to $\mathbb{F}_q^d$. Note that one can naturally view the vectors in $\mathbb{F}_q^d$ as elements in the extension field $\mathbb{F}_Q$, where $Q = q^d$. This fact ensures that the rows of matrices in $\mathbb{F}_q^{L \times L}$, when viewed as linear constraints, will be applicable to vectors in $\mathbb{F}_q^L$ and $\mathbb{F}_Q^L$ simultaneously, as $\mathbb{F}_Q$ is an extension field of $\mathbb{F}_q$.

Our second modification is to construct local profiles using a list of fractions, each corresponding to a matrix in $\mathbb{F}_q^{L \times L}$ and denoting the fraction of coordinates on which that matrix appears. This representation enables us to describe local properties in a manner that is independent of the block length of the codes. We remark that this modification is similar in spirit to the definition of local properties in [MRR+20], where the forbidden matrices were described by specifying the frequency of each vector from $\mathbb{F}_q^L$ in such matrices. In the sequel, we refer to this representation as a *local profile description*, and note that each such description defines a collection of local profiles rather than a single one.

5.1 Preliminaries

Fix a locality parameter $L \in \mathbb{N}$, and a finite field $\mathbb{F}_q$. Throughout this section, we restrict attention to $\mathbb{F}_q$-linear codes. Accordingly, the term "linear" will henceforth always refer to $\mathbb{F}_q$-linear.

Definition 5.1 (Local Profile Description)**.** *An* L-local profile description *is an unordered tuple of tuples of the form*

$$\mathcal{V} = \big((f_1, \mathbf{M}_1), \ldots, (f_T, \mathbf{M}_T)\big),$$

where for each $t \in [T]$, we have $f_t \in [0,1]$ and $\sum_{t\in[T]} f_t = 1$. The matrices $\mathbf{M}_t$ are not required to be pairwise distinct.

Definition 5.2 (Local Profile)**.** *Fix a block length $n \in \mathbb{N}$, and a L-local profile description $\mathcal{V}$ such that every f_t in $\mathcal{V}$ is a multiple of $1/n$. We define an L-*local profile *$M_n(\mathcal{V})$ created according to $\mathcal{V}$ as an ordered tuple of matrices*

$$M_n(\mathcal{V}) = (M_1, \ldots, M_n),$$

where $M_i \in \mathbb{F}_q^{L\times L}$ for each $i \in [n]$. Moreover, as prescribed by $\mathcal{V}$, for each $t \in [T]$, the matrix $\mathbf{M}_t$ appears in exactly $f_t \cdot n$ coordinates in $M_n(\mathcal{V})$.

We emphasize that $M_n(\mathcal{V})$ is not unique given a local profile description $\mathcal{V}$. In fact, it is easily seen that all permutations of the entries of $M_n(\mathcal{V})$ are valid local profiles that can be created using $\mathcal{V}$. We will denote the set of all local profiles that can be created from $\mathcal{V}$ (for a block length n) by $\mathcal{V}_n$.

For the rest of this subsection, fix a block length n, and an L-local profile description $\mathcal{V} = ((f_1, \mathbf{M}_1), \ldots, (f_t, \mathbf{M}_T))$, where every fraction f_t is a multiple of $1/n$.

Definition 5.3 (Satisfying Local Profile Descriptions)**.** *For a matrix $A \in \mathbb{F}_q^{n\times L}$, if there exists a local profile $M_n(\mathcal{V}) = (M_1, \ldots, M_n) \in \mathcal{V}_n$ such that $A[i][] \in \ker M_i$ for all i, then we say that A* satisfies *$\mathcal{V}$, and that $M_n(\mathcal{V})$ is a* witness *for A satisfying $\mathcal{V}$.*

Definition 5.4 (Containing Matrices)**.** *We say that a matrix A is* contained *in a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ if the columns of A are codewords of $\mathcal{C}$. Equivalently, we will use the shorthand $A \subseteq \mathcal{C}$.*

Definition 5.5 (Containing Local Profiles)**.** *A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is said to* contain *$\mathcal{V}$ if there exists a matrix $A \in \mathbb{F}_q^{n\times L}$ such that*

1. *$A \subseteq \mathcal{C}$,*
2. *A satisfies $\mathcal{V}$, and*
3. *A has pairwise distinct columns.*

Definition 5.6 (Local Coordinate wise Linear (LCL) Property)**.** *We define a L-*local coordinate wise linear *(L-LCL) property $\mathcal{P}$ to be a set of L-local profile descriptions $\mathcal{V}$.*

By abuse of notation, we will use the term $\mathcal{P}$ to refer to both the property itself, as well as the set of local profile descriptions that specify it.

Definition 5.7 (Satisfying LCL Properties)**.** *We say that a code $\mathcal{C} \subseteq \mathbb{F}_q^n$* satisfies *$\mathcal{P}$ if there is a $\mathcal{V} \in \mathcal{P}$ such that $\mathcal{C}$ contains $\mathcal{V}$.*

Definition 5.8 (Code contains $(\mathcal{V}, U)$)**.** *For a subspace $U \in \mathcal{L}(\mathbb{F}_q^L)$, we say that a code $\mathcal{C} \subseteq \mathbb{F}_q^n$* contains *$(\mathcal{V}, U)$ if there is a matrix $A \in \mathbb{F}_q^{n\times L}$ such that*

1. *$A \subseteq \mathcal{C}$,*
2. *A satisfies $\mathcal{V}$, and*
3. *the row span of A is equal to U.*

Furthermore, if $M_n(\mathcal{V}) \in \mathcal{V}_n$ is a witness for A satisfying $\mathcal{V}$, then we say that $M_n(\mathcal{V})$ is a witness for $\mathcal{C}$ containing $(\mathcal{V}, U)$.

Observation 5.9. *A code $\mathcal{C} \in \mathbb{F}_q^n$ contains $\mathcal{V}$ if and only if $\mathcal{C}$ contains $(\mathcal{V}, U)$ for some $U \in \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)$.*

We now state a simple fact regarding the inclusion of matrices in a random linear code.

Fact 5.10. *For a matrix $A \in \mathbb{F}_q^{n\times L}$, the probability that A is contained in rate R RLC $\mathcal{C} \subseteq \mathbb{F}_q^n$ is equal to*

$$\Pr_{\mathcal{C}}[A \subseteq \mathcal{C}] = q^{-(1-R)n \operatorname{rank} A}.$$

Proof. Since $\mathcal{C}$ by our definition is the kernel of a uniformly random matrix $K \in \mathbb{F}_q^{(1-R)n \times n}$, we can write

$$\Pr_{\mathcal{C}}[A \subseteq \mathcal{C}] = \Pr_{K \sim \mathbb{F}_q^{(1-R)n \times n}}[K \cdot A = \mathbf{0}] = \prod_{i \in [(1-R)n]} \Pr_{K \sim \mathbb{F}_q^{(1-R)n \times n}}[K[i][] \cdot A = \mathbf{0}] = q^{-(1-R)n \operatorname{rank} A}.$$

□

Definition 5.11 (Potential)**.** *For $R \in [0,1]$, and a subspace $U \in \mathcal{L}(\mathbb{F}_q^L)$, define the potential $\Phi(\mathcal{V}, U, R)$ as*

$$\Phi(\mathcal{V}, U, R) := \sum_{t \in [T]} f_t \cdot \dim(\ker(\mathbf{M}_t) \cap U) - (1-R) \dim U.$$

The following lemma provides some motivation as to why we require the definition.

Lemma 5.12. *For an L-local profile $M_n(\mathcal{V})$ created according to $\mathcal{V}$ and a subspace $U \in \mathbb{F}_q^L$, the probability that an RLC $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate R contains $(\mathcal{V}, U)$ with $M_n(\mathcal{V})$ as a witness is at most $q^{\Phi(\mathcal{V},U,R)n}$.*

Proof. The proof is given in section 4 of [LMS25], but we include it here for completeness. Define the set

$$\mathcal{M}_{(M_n(\mathcal{V}),U)} := \left\{ A \in \mathbb{F}_q^{n \times L} \mid \forall i \in [n], A[i][] \in \ker(M_i) \wedge \mathsf{row\text{-}span}(A) = U \right\}.$$

This is the set of all matrices that satisfy $\mathcal{V}$ by "complying" with the constraints specified by $M_n(\mathcal{V})$, while also having $\mathsf{row\text{-}span}(A) = U$. We now define a set similar to the one above, except we now require $\mathsf{row\text{-}span}(A) \subseteq U$.

$$\mathcal{M}^*_{(M_n(\mathcal{V}),U)} := \left\{ A \in \mathbb{F}_q^{n \times L} \mid \forall i \in [n], A[i][] \in \ker(M_i) \wedge \mathsf{row\text{-}span}(A) \subseteq U \right\}.$$

It is easy to see that $\mathcal{M}_{(M_n(\mathcal{V}),U)} \subseteq \mathcal{M}^*_{(M_n(\mathcal{V}),U)}$.

By a union bound, the probability that a rate R RLC $\mathcal{C} \subseteq \mathbb{F}_q^n$ contains $(\mathcal{V}, U)$ with $M_n(\mathcal{V})$ as a witness is at most:

$$\sum_{A \in \mathcal{M}_{(M_n(\mathcal{V}),U)}} q^{-(1-R)n \cdot \operatorname{rank}(A)} = \left|\mathcal{M}_{(M_n(\mathcal{V}),U)}\right| \cdot q^{-(1-R)n \cdot \operatorname{rank}(A)} \leq \left|\mathcal{M}^*_{(M_n(\mathcal{V}),U)}\right| \cdot q^{-(1-R)n \cdot \dim U}. \quad (8)$$

We now proceed to estimate $\left|\mathcal{M}^*_{(M_n(\mathcal{V}),U)}\right|$. Upon observing that this set is a linear subspace of $\mathbb{F}_q^{n \times L}$, we see that

$$\log_q\left(\left|\mathcal{M}^*_{(M_n(\mathcal{V}),U)}\right|\right) = \sum_{i \in [n]} \dim(\ker M_i \cap U) = \sum_{t \in [T]} f_t \cdot n \cdot \dim(\ker(\mathbf{M}_t) \cap U).$$

By plugging the value of $|\mathcal{M}^*_{(M_n(\mathcal{V}),U)}|$ in Eq. (8) we see that the probability of $\mathcal{C}$ containing $(\mathcal{V}, U)$ with $M_n(\mathcal{V})$ as a witness is at most $q^{\Phi(\mathcal{V},U,R)n}$. □

Definition 5.13. *We define $R_{\mathcal{V},U}$ to be*

$$R_{\mathcal{V},U} = \min\left\{R \in [0,1] \mid \Phi(\mathcal{V}, U, R) \geq \Phi(\mathcal{V}, W, R) \text{ for every linear subspace } W \subseteq U\right\}.$$

Note that the minimum exists, as $\Phi(\mathcal{V}, U, 0) \geq \Phi(\mathcal{V}, W, 0)$ for every subspace $W \subseteq U$, by the definition of the potential Φ.

Let us provide some motivation for the definition. We first state Proposition 4.3 from [LMS25].

Proposition 5.14 (RLC thresholds for local profiles (Proposition 4.3, [LMS25]))**.** *Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a RLC of rate $R \in [0,1]$. Fix some $M_n(\mathcal{V}) \in \mathcal{V}_n$, and a $U \in \mathcal{L}(\mathbb{F}_q^L) \setminus \{\{0\}\}$. Let*

$$\gamma := \min_{\substack{W \in \mathcal{L}(\mathbb{F}_q^L) \\ W \subsetneq U}} \left\{\Phi(\mathcal{V}, U, R) - \Phi(\mathcal{V}, W, R)\right\}.$$

The following then holds.

1. *If* $\gamma < 0$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *contains* $(\mathcal{V}, U)$ *with* $M_n(\mathcal{V})$ *as a witness*$] \leq q^{\gamma n}$.
2. *If* $\gamma > 0$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *contains* $(\mathcal{V}, U)$ *with* $M_n(\mathcal{V})$ *as a witness*$] \geq 1 - q^{-\gamma n + L^2}$.

Observe that the value of γ depends on the rate R, the local profile description $\mathcal{V}$, and subspace U, but is independent of the local profile $M_n(\mathcal{V})$. From Proposition 5.14, we see that upon union bounding over all local profiles $M_n(\mathcal{V}) \in \mathcal{V}_n$, we get:

Corollary 5.15. *Let* $\mathcal{C} \subseteq \mathbb{F}_q^n$ *be a RLC of rate* $R \in [0, 1]$. *For a* $U \in \mathcal{L}(\mathbb{F}_q^L) \setminus \{\{0\}\}$, *let* γ *be as defined in Proposition 5.14. Then the following holds.*

1. *If* $\gamma < 0$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *contains* $(\mathcal{V}, U)] \leq |\mathcal{V}_n| \cdot q^{\gamma n}$.
2. *If* $\gamma > 0$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *contains* $(\mathcal{V}, U)] \geq 1 - q^{-\gamma n + L^2}$.

Thus, Corollary 5.15 implies that for $R = R_{\mathcal{V},U} - \varepsilon$, where $0 < \varepsilon < R_{\mathcal{V},U}$ is a constant, a rate R RLC contains $(\mathcal{V}, U)$ with exponentially low probability (provided that $|\mathcal{V}_n|$ is sufficiently small), while for $R = R_{\mathcal{V},U} + \varepsilon$, a rate R RLC contains $(\mathcal{V}, U)$ with probability exponentially close to 1.

We now define a threshold rate with respect to an RLC containing $\mathcal{V}$. Invoking Observation 5.9, we define the threshold rate corresponding to $\mathcal{V}$ as

$$R_{\mathcal{V}} := \min_{U \in \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)} R_{\mathcal{V},U}.$$

For an L-LCL property $\mathcal{P}$, recall that a code $\mathcal{C}$ satisfies $\mathcal{P}$ if $\mathcal{C}$ contains some $\mathcal{V} \in \mathcal{P}$. We thus define:

Definition 5.16. *The threshold rate of an* L*-LCL property* $\mathcal{P}$ *is defined to be*

$$R_{\mathcal{P}} := \min_{\mathcal{V} \in \mathcal{P}} R_{\mathcal{V}} = \min_{\substack{\mathcal{V} \in \mathcal{P} \\ U \in \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)}} R_{\mathcal{V},U}. \tag{9}$$

Theorem 4.4 from [LMS25] proves that $R_{\mathcal{P}}$ as defined above is indeed a threshold rate.

Theorem 5.17 (Theorem 4.4, [LMS25])**.** *Let* $\mathcal{P}$ *be an* L*-LCL property. Let* $\mathcal{C} \subseteq \mathbb{F}_q^n$ *be an RLC of rate* R, *and let* $R_{\mathcal{P}}$ *be as defined in Definition 5.16. Let* $\varepsilon > 0$ *be a sufficiently small constant. The following now holds*

1. *If* $R \geq R_{\mathcal{P}} + \varepsilon$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *satisfies* $\mathcal{P}] \geq 1 - q^{-\varepsilon n + L^2}$.
2. *If* $R \leq R_{\mathcal{P}} - \varepsilon$, *then* $\Pr_{\mathcal{C}}[\mathcal{C}$ *satisfies* $\mathcal{P}] \leq \sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n| \cdot q^{-\varepsilon n + L^2}$.

At this point we provide a brief discussion about the LCL properties $\mathcal{P}$ considered henceforth. In general, we allow the fractions $f_1, \ldots, f_T$ to range over sub-intervals of $[0, 1]$, and regard all local profile descriptions with such fractions as belonging to $\mathcal{P}$. Consequently, $\mathcal{P}$ is uncountably infinite. This, however, does not pose a difficulty, because for a block length n, the set of local profile descriptions satisfying $|\mathcal{V}_n| > 0$ is itself finite, as by definition, each fraction in such local profile descriptions must be an exact multiple of $1/n$. Furthermore, for any such local profile description, the corresponding set of local profiles $\mathcal{V}_n$ is also finite. Specifically, $|\mathcal{V}_n|$ is equal to the number of ways we can arrange n objects in a row, where we have $f_t \cdot n$ objects of type t, for $t \in T$. It follows, therefore, that

$$\sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n|$$

is finite.

Definition 5.18. *We say an* L*-LCL property* $\mathcal{P}$ *is* reasonable *if for every block length* n, *the total number of associated local profiles is at most*

$$\sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n| \leq q^{\kappa_q(\mathcal{P}) \cdot n},$$

where the term $\kappa_q(\mathcal{P})$ *approaches* 0 *as* $q \to \infty$.

We note that $\kappa_q(\mathcal{P})$ is allowed to depend on L in an arbitrary fashion, although a worse dependence would result in q being large (with respect to L) as well. Most local properties considered in the literature, such as list-decoding, list-recovery and perfect hash matrices are *reasonable* local properties.

We require Lemma 4.5 from [LMS25]:

Lemma 5.19 (Lemma 4.5, part (3), [LMS25])**.** *For $R \in [0,1]$, denote*

$$\operatorname{argmax}\{\Phi(\mathcal{V},*,R)\} := \left\{ W \in \mathcal{L}(\mathbb{F}_q^L) \mid \Phi(\mathcal{V},U,R) = \max_{U \in \mathcal{L}(\mathbb{F}_q^L)} \Phi(\mathcal{V},U,R)) \right\}.$$

Then, $\operatorname{argmax}\Phi(\mathcal{V},*,R_{\mathcal{V}})$ *contains at least one element from* $\mathcal{L}(\mathbb{F}_q^L) \setminus \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)$*, and at least one element from* $\mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)$*.*

Claim 5.20. *There is a canonical $W_{\mathcal{V}} \in \mathcal{L}(\mathbb{F}_q^L) \setminus \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)$ such that for all $R \le R_{\mathcal{V}}$ and for every $U \in \mathcal{L}(\mathbb{F}_q^L)$ satisfying $\dim U \ge \dim W_{\mathcal{V}}$, the inequality*

$$\Phi(\mathcal{V},U,R) \le \Phi(\mathcal{V},W_{\mathcal{V}},R)$$

is true.

Proof. By Lemma 5.19, the set

$$\operatorname{argmax}\Phi(\mathcal{V},*,R_{\mathcal{V}}) \;\cap\; \left(\mathcal{L}(\mathbb{F}_q^L) \setminus \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)\right)$$

is non-empty. Fix an arbitrary total ordering on the subspaces of $\mathcal{L}(\mathbb{F}_q^L)$, and select the first subspace in this ordering from the above set. Denote this subspace by $W_{\mathcal{V}}$.

Fix a subspace $U \in \mathcal{L}(\mathbb{F}_q^L)$ satisfying $\dim U \ge \dim W_{\mathcal{V}}$, and some $R \le R_{\mathcal{V}}$. Then,

$$\begin{aligned}
\Phi(\mathcal{V},U,R) - \Phi(\mathcal{V},W_{\mathcal{V}},R) &= \sum_{t\in[T]} f_t \cdot (\dim(\ker(\mathbf{M}_t)\cap U) - \dim(\ker(\mathbf{M}_t)\cap W_{\mathcal{V}})) \\
&\quad - (1-R)(\dim U - \dim W_{\mathcal{V}}) \\
&\le \sum_{t\in[T]} f_t \cdot (\dim(\ker(\mathbf{M}_t)\cap U) - \dim(\ker(\mathbf{M}_t)\cap W_{\mathcal{V}})) \\
&\quad - (1-R_{\mathcal{V}})(\dim U - \dim W_{\mathcal{V}}) \\
&= \Phi(\mathcal{V},U,R_{\mathcal{V}}) - \Phi(\mathcal{V},W_{\mathcal{V}},R_{\mathcal{V}}) \\
&\le 0,
\end{aligned}$$

where the last inequality holds because $W_{\mathcal{V}} \in \operatorname{argmax}\Phi(\mathcal{V},*,R_{\mathcal{V}})$, and the first one holds because $R \le R_{\mathcal{V}}$ and $\dim U \ge \dim W_{\mathcal{V}}$. □

5.2 Implied Local Profile Descriptions

Fix an L-local profile description $\mathcal{V} = ((f_1,\mathbf{M}_1),\dots,(f_T,\mathbf{M}_T))$.

Definition 5.21 (Implied Local Profile Description)**.** *Let ψ be a linear map $\psi : \mathbb{F}_q^L \to \mathbb{F}_q^{L-K}$ for some integer $K \le L$. Then, the $(L-K)$-*implied local profile description *of $\mathcal{V}$ with respect to ψ, denoted by $\mathcal{V}^{\psi}$, is an $(L-K)$-local profile description defined as*

$$\mathcal{V}^{\psi} := ((f_1,\mathbf{M}_1^{\psi}),\dots,(f_T,\mathbf{M}_T^{\psi})).$$

Here, for each $t \in [T]$, $\mathbf{M}_t^{\psi}$ is a matrix satisfying $\ker \mathbf{M}_t^{\psi} = \psi(\ker \mathbf{M}_t)$.

We now give some intuition for Definition 5.21.

Observation 5.22. *For a linear map $\psi : \mathbb{F}_q^L \to \mathbb{F}_q^{L-K}$, let $\mathcal{V}^\psi$ be the corresponding $(L-K)$-*implied local profile description *of $\mathcal{V}$. If a linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ contains $\mathcal{V}$, then there is a matrix $B \in \mathbb{F}_q^{n \times (L-K)}$ (with possibly non-distinct columns) such that*

1. *$B \subseteq \mathcal{C}$, and*
2. *B satisfies $\mathcal{V}^\psi$.*

Proof. By Definition 5.5, if $\mathcal{C}$ contains $\mathcal{V}$, then there is a matrix $A \in \mathbb{F}_q^{n \times L}$ that (i) has pairwise distinct columns, (ii) is contained in $\mathcal{C}$, and (iii) satisfies $\mathcal{V}$. Now consider the matrix $B \in \mathbb{F}_q^{n \times (L-K)}$ constructed row by row, such that $B[i][] := \psi(A[i][])$. By the linearity of $\mathcal{C}$, $A \subseteq \mathcal{C}$ implies $B \subseteq \mathcal{C}$, and moreover, it is easy to verify that B satisfies $\mathcal{V}^\psi$. □

Let $W_\mathcal{V}$ be the canonical subspace guaranteed by Claim 5.20 for $\mathcal{V}$. Recall that for all $R \leq R_\mathcal{V}$ and for every $U \in \mathcal{L}(\mathbb{F}_q^L)$ satisfying $\dim U \geq \dim W_\mathcal{V}$, $W_\mathcal{V}$ satisfies

$$\Phi(\mathcal{V}, U, R) \leq \Phi(\mathcal{V}, W_\mathcal{V}, R).$$

Fix a full rank linear map $\psi_\mathcal{V} \colon \mathbb{F}_q^L \to \mathbb{F}_q^{L - \dim W_\mathcal{V}}$ such that $\ker \psi_\mathcal{V} = W_\mathcal{V}$.

Definition 5.23 (Canonical Implied Local Profile Description)**.** *Let $\psi_\mathcal{V}$ be as defined above. The $(L - \dim W_\mathcal{V})$-*canonical implied local profile description *of $\mathcal{V}$, denoted by $\mathcal{V}^{\mathsf{imp}}$, is an $(L - \dim W_\mathcal{V})$-local profile description defined as*

$$\mathcal{V}^{\mathsf{imp}} := ((f_1, \mathbf{M}_1^{\mathsf{imp}}), \ldots, (f_T, \mathbf{M}_T^{\mathsf{imp}})).$$

Here, $\mathbf{M}_t^{\mathsf{imp}}$ is a matrix satisfying $\ker \mathbf{M}_t^{\mathsf{imp}} = \psi_\mathcal{V}(\ker \mathbf{M}_t)$ for each $t \in [T]$.

Robust Local Profile Descriptions. We will now describe a general method to create robust counterparts to local profile descriptions.

Definition 5.24 (Robust Local Profile Description)**.** *Let $\Delta \in [0, 1]$ be a constant, and $\mathcal{V} = ((f_1, \mathbf{M}_1), \ldots, (f_T, \mathbf{M}_T))$ be an L-local profile description. Consider the set of all L-local profile descriptions of the form*

$$\Big((f_1 - \Delta_1,\ \mathbf{M}_1),\ \ldots,\ (f_T - \Delta_T,\ \mathbf{M}_T),\ \big(\textstyle\sum_{t \in [T]} \Delta_t,\ \mathbf{0}\big) \Big)$$

where for every $t \in [T]$, we have $0 \leq \Delta_t \leq f_t$ and $\sum_{t \in [T]} \Delta_t \leq \Delta$. Recall that $\mathbf{0}$ is the all zeroes matrix. We denote this set by $\mathrm{Rob}_\Delta(\mathcal{V})$, and an element from this set is known as a Δ-robust local profile description *of $\mathcal{V}$.*

Informally speaking, any code $\mathcal{C}$ that avoids containing matrices that satisfy local profile descriptions from $\mathrm{Rob}_\Delta(\mathcal{V})$ will consequently avoid containing matrices that only satisfy a $(1 - \Delta)$ fraction of the constraints set forth by $\mathcal{V}$.

Observation 5.25. *For every constant $\Delta \in [0, 1]$, a matrix satisfying $\mathcal{V}$ also satisfies every local profile description from $\mathrm{Rob}_\Delta(\mathcal{V})$.*

We now state and prove a lemma asserting that if the rate R is bounded away from the threshold rate $R_\mathcal{V}$ by ε, then the potential of the robust counterparts of $\mathcal{V}^{\mathsf{imp}}$ is bounded above by $-\varepsilon$, up to a small additive factor.

Lemma 5.26. *For $\varepsilon > 0$, let $R \leq R_\mathcal{V} - \varepsilon$. Let $\mathcal{V}_\Delta^{\mathsf{imp}} \in \mathrm{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$ be some Δ-robust local profile description of $\mathcal{V}^{\mathsf{imp}}$. Then for every $U' \in \mathcal{L}(\mathbb{F}_q^{(L - \dim W_\mathcal{V})}), U' \neq \{0\}$, we have*

$$\Phi(\mathcal{V}_\Delta^{\mathsf{imp}}, U', R) \leq -\varepsilon + \Delta \cdot L.$$

Proof. By the guarantee for $W_{\mathcal{V}}$ (see Claim 5.20), we have for all subspaces $U \in \mathcal{L}(\mathbb{F}_q^L)$ satisfying $\dim U \geq \dim W_{\mathcal{V}}$,

$$\begin{aligned}\Phi(\mathcal{V}, U, R) - \Phi(\mathcal{V}, W_{\mathcal{V}}, R) &= \Phi(\mathcal{V}, U, R_{\mathcal{V}}) - \Phi(\mathcal{V}, W_{\mathcal{V}}, R_{\mathcal{V}}) - \varepsilon(\dim U - \dim W_{\mathcal{V}}) \\ &\leq -\varepsilon(\dim U - \dim W_{\mathcal{V}}). \end{aligned} \tag{10}$$

Let $U' \in \mathcal{L}(\mathbb{F}_q^{(L-\dim W_{\mathcal{V}})})$ be such that $U' \neq \{0\}$, and let $U \in \mathcal{L}(\mathbb{F}_q^L)$ be such that $W_{\mathcal{V}} \subset U$ and $\psi_{\mathcal{V}}(U) = U'$, that is, $U = \psi_{\mathcal{V}}^{-1}(U') = \left\{u \in \mathbb{F}_q^L \mid \psi_{\mathcal{V}}(u) \in U'\right\}$. By the rank-nullity theorem, $\dim U' = \dim U - \dim W_{\mathcal{V}}$.

Consequently,

$$\begin{aligned}\Phi(\mathcal{V}_\Delta^{\mathsf{imp}}, U', R) &= \Big(\sum_{t\in[T]} (f_t - \Delta_t) \cdot \dim(\ker(\mathbf{M}_t^{\mathsf{imp}}) \cap U')\Big) \\ &\quad + \sum_{t\in[T]} \Delta_t \cdot \dim(\ker \mathbf{0}^{\mathsf{imp}}) - (1-R)\dim U' \\ &= \Big(\sum_{t\in[T]} (f_t - \Delta_t) \cdot \dim(\psi_{\mathcal{V}}(\ker(\mathbf{M}_t) \cap U))\Big) \qquad (11) \\ &\quad + \sum_{t\in[T]} \Delta_t \cdot \dim(\psi_{\mathcal{V}}(\ker \mathbf{0})) - (1-R)(\dim U - \dim W_{\mathcal{V}}) \\ &= \Big(\sum_{t\in[T]} (f_t - \Delta_t) \cdot (\dim(\ker(\mathbf{M}_t) \cap U) - \dim(\ker(\mathbf{M}_t) \cap W_{\mathcal{V}}))\Big) \qquad (12) \\ &\quad + \sum_{t\in[T]} \Delta_t \cdot \dim(\psi_{\mathcal{V}}(\ker \mathbf{0})) - (1-R)(\dim U - \dim W_{\mathcal{V}}) \\ &\leq \Phi(\mathcal{V}, U, R) - \Phi(\mathcal{V}, W_{\mathcal{V}}, R) + \Delta \cdot L \\ &\leq -\varepsilon + \Delta \cdot L, \end{aligned}$$

where Eq. (11) follows from Claim 5.27, stated and proved below. Moreover, Eq. (12) follows from an application of the rank-nullity theorem. The final inequality follows from Eq. (10), owing to the fact that $U \supset W_{\mathcal{V}}$, given that U' is a non-trivial subspace. □

Claim 5.27. *Let $W_{\mathcal{V}}, \psi_{\mathcal{V}}, U'$ and U be as defined in Lemma 5.26. Then, for every subspace K in $\mathbb{F}_q^L$, we have*

$$\psi_{\mathcal{V}}(K) \cap U' = \psi_{\mathcal{V}}(K) \cap \psi_{\mathcal{V}}(U) = \psi_{\mathcal{V}}(K \cap U).$$

Proof. First, $\psi_{\mathcal{V}}(K \cap U) \subseteq \psi_{\mathcal{V}}(K) \cap \psi_{\mathcal{V}}(U)$ is immediate: if $x \in \psi_{\mathcal{V}}(K \cap U)$, then there is a $x' \in K \cap U$ such that $\psi_{\mathcal{V}}(x') = x$. Then, $x' \in K$ implies $x = \psi_{\mathcal{V}}(x') \in \psi_{\mathcal{V}}(K)$, and $x' \in U$ implies $x = \psi_{\mathcal{V}}(x') \in \psi_{\mathcal{V}}(U)$.

For the other inclusion, take $y \in \psi_{\mathcal{V}}(K) \cap \psi_{\mathcal{V}}(U)$. Then, there exist $k \in K$ and $u \in U$ such that $\psi_{\mathcal{V}}(u) = y = \psi_{\mathcal{V}}(k)$. Hence

$$k - u \in \ker \psi_{\mathcal{V}} = W_{\mathcal{V}}.$$

But recall that $W_{\mathcal{V}} \subset U$, and thus every element of $\ker \psi_{\mathcal{V}}$ lies in U. Therefore, $k = u + (k - u) \in U$. Since $k \in K$ as well, $k \in K \cap U$, and so $y = \psi_{\mathcal{V}}(k) \in \psi_{\mathcal{V}}(K \cap U)$. Thus, $\psi_{\mathcal{V}}(K) \cap \psi_{\mathcal{V}}(U) \subseteq \psi_{\mathcal{V}}(K \cap U)$.

□

Lemma 5.28. *For $0 \leq \Delta \leq 1$, consider some $\mathcal{V}_\Delta^{\mathsf{imp}} \in \mathrm{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$. For a block length n, the number of local profiles associated with $\mathcal{V}_\Delta^{\mathsf{imp}}$ is at most $2^n\,|\mathcal{V}_n|$.*

Proof. Let $\mathcal{V}_\Delta^{\mathsf{imp}}$ be of the form

$$\mathcal{V}_\Delta^{\mathsf{imp}} = \Big((f_1 - \Delta_1,\ \mathbf{M}_1), \ldots, (f_T - \Delta_T,\ \mathbf{M}_T), \big(\textstyle\sum_{t\in[T]} \Delta_t,\ \mathbf{0}\big)\Big),$$

where for every $t \in [T]$, we have $0 \leq \Delta_t \leq f_t$ and $\sum_{t\in[T]} \Delta_t \leq \Delta$. Each local profile associated with $\mathcal{V}_\Delta^{\mathsf{imp}}$ can be created by following this recipe: fix some local profile $M_n(\mathcal{V})$ created according to $\mathcal{V}$. Then,

(i) For each coordinate, change the associated matrix from $\mathbf{M}_t$ to $\mathbf{M}_t^{\mathsf{imp}}$.

(ii) for each $t \in [T]$, select some $\Delta_t \cdot n$ coordinates whose associated matrix is $\mathbf{M}_t$ and change the associated matrix to $\mathbf{0}$.

The first step does not cause an increase in the number of local profiles. The number of ways to perform the second step is upper bounded by $\binom{n}{\leq \Delta \cdot n} \leq 2^n$. □

We now state and prove a lemma that establishes an upper bound on the probability with which an RLC of rate below the rate threshold $R_{\mathcal{P}}$ contains non-zero matrices satisfying a robust local profile description of $\mathcal{V}^{\mathsf{imp}}$.

Lemma 5.29. *Let $\mathcal{P}$ be an L-LCL property. Let $R = R_{\mathcal{P}} - \varepsilon$ for a constant $\varepsilon \in [0, R_{\mathcal{P}})$, and let Δ be a constant in $[0, 1]$. Then the probability that an RLC $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate R contains a non-zero matrix satisfying a local profile description from the set*

$$\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$$

is at most $\sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n| \cdot 2^n \cdot q^{L^2 - \varepsilon n + \Delta \cdot L n}$.

Proof. By Lemma 5.28, we see that for an L-LCL property $\mathcal{P}$, the total number of local profiles that can be created according to local profile descriptions from the set $\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$ is at most

$$\sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n| \cdot 2^n.$$

For some $\mathcal{V} \in \mathcal{P}$, consider a local profile $M_n(\mathcal{V}_{\Delta}^{\mathsf{imp}})$ created according to $\mathcal{V}_{\Delta}^{\mathsf{imp}} \in \mathrm{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$. Recall that $\mathcal{V}_{\Delta}^{\mathsf{imp}}$ is a $(L - \dim W_{\mathcal{V}})$-canonical implied local profile description. Fix a non-zero subspace $U \in \mathbb{F}_q^{(L - \dim W_{\mathcal{V}})}$. From Lemma 5.12, and the fact that

$$R \leq R_{\mathcal{P}} - \varepsilon \leq R_{\mathcal{V}} - \varepsilon \leq R_{\mathcal{V},U} - \varepsilon,$$

which follows from Eq. (9), we see that the probability that there is a non-zero matrix $A \in \mathbb{F}_q^{n \times (L - \dim W_{\mathcal{V}})}$ such that $A \subseteq \mathcal{C}$, and A satisfies $(\mathcal{V}_{\Delta}^{\mathsf{imp}}, U)$ with $M_n(\mathcal{V}_{\Delta}^{\mathsf{imp}})$ as a witness is at most

$$q^{\Phi(\mathcal{V}_{\Delta}^{\mathsf{imp}}, U, R) n} \leq q^{(-\varepsilon + \Delta \cdot L) n},$$

where the inequality follows from Lemma 5.26. The result follows by union bounding over all non-zero subspaces U, and all local profiles created according to local profile descriptions from the set $\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$. □

We now turn to analyze the upper bound on the probability in Lemma 5.29. The bound is non-trivial if and only if $\sum_{\mathcal{V} \in \mathcal{P}} |\mathcal{V}_n| \cdot 2^n \cdot q^{L^2 - \varepsilon n} \cdot q^{\Delta \cdot L n} < 1$. For *reasonable* L-LCL properties, this is achieved when

$$\kappa_q(\mathcal{P}) + \log_q 2 + \frac{L^2}{n} - \varepsilon + \Delta \cdot L < 0.$$

Upon restricting Δ to be at most $\varepsilon / 2L$ and n to be at least $4L^2/\varepsilon$, it suffices to have

$$\kappa_q(\mathcal{P}) + \log_q 2 - \frac{\varepsilon}{4} < 0. \tag{13}$$

Here, we are interested in the smallest prime power q that satisfies the above inequality. Such a q exists, since $\lim_{q \to \infty} (\kappa_q(\mathcal{P}) + \log_q 2) = 0$.

We summarize the above discussion as the following fact:

Fact 5.30. *For a* reasonable *L-LCL property $\mathcal{P}$ and $R \leq R_{\mathcal{P}} - \varepsilon$, there exists a minimum prime power q that is solely a function of ε, L, and $\kappa_q(\mathcal{P})$, such that the following holds: There exists a linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ for every block length $n \geq 4L^2/\varepsilon$ that contains no non-zero matrices satisfying local profile descriptions from the set $\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$.*

Lemma 5.31. *For a block length $n \geq 4L^2/\varepsilon$ and a reasonable L-LCL property $\mathcal{P}$, let $\mathcal{C}$ be a linear code $\mathcal{C}$ as described in Fact 5.30. Then,*

1. *the code $\mathcal{C}$ can be found in time $q^{\mathrm{poly}(n,L)}$, and*
2. *$\mathcal{C}$ does not satisfy $\mathcal{P}$.*

Proof. We first prove that $\mathcal{C}$ can be found in time $q^{\mathrm{poly}(n,L)}$. By our definition of random linear codes, a non-zero probability is placed on at most $q^{(1-R)n}$ linear subspaces of $\mathbb{F}_q^n$. For each such code, we perform a check of the following form: determine whether it contains a non-zero matrix satisfying a local profile description from the set $\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$. We return the first code $\mathcal{C}$ that does not contain any matrices of the described form. Such a code is guaranteed to exist by Fact 5.30. Let us analyze the runtime of this algorithm.

The number of non-zero matrices satisfying local profile descriptions from the set $\bigcup_{\mathcal{V} \in \mathcal{P}} \mathrm{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$ is at most the total number of matrices of dimensions $n \times L$, which is equal to q^{nL}. Given the parity check matrix of a linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$, we can check whether every column of a matrix $A \in \mathbb{F}_q^{n \times L}$ lies in $\mathcal{C}$ in time at most $n^2 L$. Therefore, for each linear code, we can perform the aforementioned check in time at most $n^2 L \cdot q^{nL}$. Consequently, the total runtime is at most

$$q^{(1-R)n} \cdot n^2 L \cdot q^{nL} \leq q^{\mathrm{poly}(n,L)}.$$

We now prove that $\mathcal{C}$ does not satisfy $\mathcal{P}$. Assume for contradiction that it does. Then, there is a $\mathcal{V} \in \mathcal{P}$ such that $\mathcal{C}$ contains $\mathcal{V}$. By Observation 5.22, $\mathcal{C}$ contains a matrix B that satisfies $\mathcal{V}^{\mathsf{imp}}$. The matrix B was obtained from some matrix $A \subseteq \mathcal{C}$ having pairwise distinct columns, by applying a map ψ to every row of A. By the definition of $\mathcal{V}^{\mathsf{imp}}$, there exist $i, j \in [L]$, where $i \neq j$, such that the kernel of ψ consists of a subspace whose vectors agree at coordinates i, j. This implies that B is non-zero. Furthermore, by Observation 5.25, B also satisfies every $\mathcal{V}_\Delta^{\mathsf{imp}} \in \mathrm{Rob}_\Delta(\mathcal{V})$. This contradicts the fact that $\mathcal{C}$ is of the form as specified in Fact 5.30. □

Recall that the code produced by the AEL construction, $\mathcal{C}_{\mathrm{AEL}}$, has alphabet Σ_{in}^d, where Σ_{in} is the alphabet of the inner code. We take the inner code to be over $\mathbb{F}_q$, and therefore interpret $\mathcal{C}_{\mathrm{AEL}}$ as a code over the extension field $\mathbb{F}_Q$, where $Q = q^d$. Furthermore, by inspection, the definitions of containment and satisfiability (Definition 5.3, Definition 5.4, Definition 5.5, Definition 5.7) also apply to matrices and codes over $\mathbb{F}_Q$. In particular, they apply to $\mathcal{C}_{\mathrm{AEL}} \subseteq \mathbb{F}_Q^N$.

5.3 Main Result

We now state and prove the central theorem of this section. Fix a *reasonable* L-LCL property $\mathcal{P}$, and $R_{\mathrm{in}} \leq R_{\mathcal{P}} - \varepsilon$ for some $\varepsilon > 0$. Let $\Delta = \varepsilon/2L$ be as defined above. Let $T_{\mathcal{P}}$ denote the maximum number of distinct matrices appearing in any local profile description $\mathcal{V} \in \mathcal{P}$. Let G be the bipartite graph obtained from Claim 3.4 by setting

$$\eta = \frac{\Delta}{4T_{\mathcal{P}}},$$

and

$$\zeta = \frac{\delta_{\mathrm{out}}}{2T_{\mathcal{P}}},$$

having degree

$$d = O\left(\frac{1}{\zeta \eta^2}\right) = O\left(\frac{T_{\mathcal{P}}^3}{\Delta^2 \cdot \delta_{\mathrm{out}}}\right).$$

For any integer N satisfying $N > T_{\mathcal{P}}/\delta_{\text{out}}$, take G to have N vertices on both sides. Let $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_q^d$ be an $\mathbb{F}_q$-linear code of rate R_{in}, where q is the minimum prime power guaranteed by Fact 5.30, and let $\mathcal{C}_{\text{out}} \subseteq (\mathbb{F}_q^{R_{\text{in}}d})^N$ be an $\mathbb{F}_q$-linear code having distance δ_{out}. By definition, ϕ is now an $\mathbb{F}_q$-linear map of the form $\phi\colon \mathbb{F}_q^{R_{\text{in}}d} \to \mathcal{C}_{\text{in}}$.

Theorem 5.32. *Let $\mathcal{C}_{in}, \mathcal{C}_{out}$, and G be as defined above. Furthermore, assume that $\mathcal{C}_{in}$ is a linear code whose existence is guaranteed by Fact 5.30. Then, $\mathcal{C}_{AEL}(\mathcal{C}_{in}, \mathcal{C}_{out}, G) \subseteq \mathbb{F}_Q^N$ is an $\mathbb{F}_q$-linear code that **does not** satisfy $\mathcal{P}$.*

Proof. Since both $\mathcal{C}_{\text{out}}$ and the map ϕ are $\mathbb{F}_q$-linear, it follows that $\mathcal{C}_{\text{AEL}}$ is itself an $\mathbb{F}_q$-linear code.

Assume for contradiction that $\mathcal{C}_{\text{AEL}}$ satisfies $\mathcal{P}$. Then, there is a $\mathcal{V} = (f_1, \mathbf{M}_1), \ldots, (f_T, \mathbf{M}_T)) \in \mathcal{P}$ such that $\mathcal{C}_{\text{AEL}}$ contains $\mathcal{V}$, which implies the existence of a matrix $A \in \mathbb{F}_Q^{N \times L}$, such that

1. $A \subseteq \mathcal{C}_{\text{AEL}}$,
2. A satisfies $\mathcal{V}$, and
3. A has pairwise distinct columns.

Let the columns of A be pairwise distinct codewords $c_1, \ldots, c_L \in \mathcal{C}_{\text{AEL}} \subseteq \mathbb{F}_Q^N$. Note that we can reinterpret these codewords as vectors in $(\mathbb{F}_q^d)^N$, and we choose to do so. We now perform three operations. Firstly, we create $A_{\mathsf{fl}} \in \mathbb{F}_q^{([N]\times[d])\times L}$ from A by *flattening* every column of A (see Definition 3.3) and then collecting them in a matrix A_{fl}. Therefore, for every $i \in [L]$, $A_{\mathsf{fl}}[][i] = (A[][i])_{\mathsf{fl}}$. Secondly, we create $A_{\mathsf{fl}}^{\mathsf{proj}} \in \mathbb{F}_q^{([N]\times[d])\times L}$ by performing the *projection* operation on A_{fl} (see Definition 3.1).

We now instantiate the objects required to perform the third operation. As $R \leq R_{\mathcal{P}} - \varepsilon \leq R_{\mathcal{V}} - \varepsilon$, by Claim 5.20 there exists a (canonical) subspace $W_{\mathcal{V}}$ for $\mathcal{V}$. In particular, because

$$W_{\mathcal{V}} \in (\mathcal{L}(\mathbb{F}_q^L) \setminus \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)),$$

$\dim W_{\mathcal{V}} \neq L$ holds, as the only subspace of dimension L, $\mathbb{F}_q^L$, lies in $\mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L)$. Define a linear map $\psi\colon \mathbb{F}_q^L \to \mathbb{F}_q^{(L - \dim W_{\mathcal{V}})}$ satisfying $\ker \psi = W_{\mathcal{V}}$. Note that because $\dim W_{\mathcal{V}} < L$, this map is well-defined. The third operation consists of applying ψ on each row of $A_{\mathsf{fl}}^{\mathsf{proj}}$. We denote the resultant matrix by $A_{\psi}^{\mathsf{proj}} \in \mathbb{F}_q^{([N]\times[d])\times(L - \dim W_{\mathcal{V}})}$. That is, A_{ψ}^{proj} satisfies

$$A_{\psi}^{\mathsf{proj}}[(\ell, i)][] = \psi(A_{\mathsf{fl}}^{\mathsf{proj}}[(\ell, i)][])$$

for all $\ell \in [N]$ and $i \in [d]$.

We recall Definition 3.2, and note that $A_{\psi}^{\mathsf{proj}}(\ell)$ denotes the submatrix of A_{ψ}^{proj} consisting of rows indexed by vertex-edge pairs corresponding to the vertex $\ell \in [N]$. This submatrix lives in $\mathbb{F}_q^{[d]\times(L - \dim W_{\mathcal{V}})}$.

We now prove a claim stating that the submatrices $A_{\psi}^{\mathsf{proj}}(\ell)$ corresponding to most left vertices in fact satisfy $\mathcal{V}_{\Delta}^{\mathsf{imp}}$, for some $\mathcal{V}_{\Delta}^{\mathsf{imp}} \in \text{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$.

Claim 5.33. *There exists a set of indices $L^* \subseteq [N]$ satisfying $|L^*| > (1 - \delta_{out})N$ such that for every $\ell \in L^*$, there exists a $\mathcal{V}_{\Delta}^{\mathsf{imp}} \in \text{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$ such that the matrix $A_{\psi}^{\mathsf{proj}}(\ell)$ satisfies $\mathcal{V}_{\Delta}^{\mathsf{imp}}$.*

We state the rest of the proof assuming Claim 5.33, whose proof is given below.

We now prove that because the columns of A are pairwise distinct, A_{ψ}^{proj} contains at least one non-zero row. This is seen as follows: $A \in (\mathbb{F}_q^d)^{N \times L}$ has pairwise distinct columns, which implies that A_{fl} has pairwise distinct columns as well. Because $A_{\mathsf{fl}}^{\mathsf{proj}}$ is obtained by applying a permutation on the rows of A_{fl}, the same applies for $A_{\mathsf{fl}}^{\mathsf{proj}}$. Therefore, for every $i_1, i_2 \in [L]$, there is at least one row in $A_{\mathsf{fl}}^{\mathsf{proj}}$ that has differing entries on i_1 and i_2. In particular, this is true for a pair of indices of $[L]$ on which every vector in $W_{\mathcal{V}}$ has identical entries, and such a pair of indices exists by virtue of $W_{\mathcal{V}} \in (\mathcal{L}(\mathbb{F}_q^L) \setminus \mathcal{L}_{\mathsf{Dist}}(\mathbb{F}_q^L))$.

This implies that there exists a $\ell \in [N]$ and a $j \in [d]$ such that $A_{\mathsf{fl}}^{\mathsf{proj}}[(\ell, j)][] \notin W_{\mathcal{V}}$. Thus, it is seen that $A_{\psi}^{\mathsf{proj}}[(\ell, j)][] \neq \mathbf{0}^{L - \dim W_{\mathcal{V}}}$.

This immediately implies that one of the columns of A_{ψ}^{proj} is non-zero. Denote the index of this column by $e \in [L - \dim W_{\mathcal{V}}]$. Observe that the columns of A_{ψ}^{proj} are codewords in $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$. Indeed, the columns of A_{fl} are (flattened) codewords of $\mathcal{C}_{\text{AEL}}$, and so the columns of $A_{\mathsf{fl}}^{\mathsf{proj}}$ are codewords in $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$. It follows directly from the $\mathbb{F}_q$-linearity of $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$ that $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is also $\mathbb{F}_q$-linear. Moreover, since A_{ψ}^{proj} is obtained by applying an $\mathbb{F}_q$-linear map to the rows of $A_{\mathsf{fl}}^{\mathsf{proj}}$, the columns of A_{ψ}^{proj} are $\mathbb{F}_q$-linear combinations of columns of $A_{\mathsf{fl}}^{\mathsf{proj}}$, and therefore are also codewords of $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$. In particular, this holds for the non-zero column $A_{\psi}^{\mathsf{proj}}[][e]$, and therefore is of the following form

$$(\phi(c[1]), \ldots, \phi(c[N]))^{\top}$$

where c is a codeword in $\mathcal{C}_{\text{out}}$. This in turn implies that at least $\delta_{\text{out}} N$ submatrices of the form $A_{\psi}^{\mathsf{proj}}(\ell)$ are non-zero, because the column indexed by e is non-zero for all such submatrices. Denote this set of indices by $S \subseteq [N]$.

By Claim 5.33, and the fact that $|S| \geq \delta_{\text{out}} N$, there is at least one index $\ell^* \in (S \cap L^*)$. Since the columns of A_{ψ}^{proj} are codewords of $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$, it follows that every column of the submatrix $A_{\psi}^{\mathsf{proj}}(\ell^*)$ is a codeword of $\mathcal{C}_{\text{in}}$. Therefore, $A_{\psi}^{\mathsf{proj}}(\ell^*)$ has the following properties:

1. $A_{\psi}^{\mathsf{proj}}(\ell^*) \subseteq \mathcal{C}_{\text{in}}$,
2. there exists a $\mathcal{V}_{\Delta}^{\mathsf{imp}} \in \text{Rob}_{\Delta}(\mathcal{V}^{\mathsf{imp}})$ such that $A_{\psi}^{\mathsf{proj}}(\ell^*)$ satisfies $\mathcal{V}_{\Delta}^{\mathsf{imp}}$ (by Claim 5.33 and the fact that $\ell^* \in L^*$), and
3. $A_{\psi}^{\mathsf{proj}}(\ell^*)$ is non-zero (as $\ell^* \in S$).

This contradicts the fact that $\mathcal{C}_{\text{in}}$ is a code satisfying the guarantee mentioned in Fact 5.30.

□

Proof of Claim 5.33. Since A satisfies $\mathcal{V} = (f_1, \mathbf{M}_1), \ldots, (f_T, \mathbf{M}_T))$, there is a *witness* $M_N(\mathcal{V}) \in \mathcal{V}_N$ such that

$$\forall r \in [N], A[r][] \in \ker M_r. \tag{14}$$

For every $t \in [T]$, define

$$S_t := \{r \in [N] \mid M_r = \mathbf{M}_t\}.$$

This is the set of right vertices whose corresponding matrix in $M_N(\mathcal{V})$ is equal to $\mathbf{M}_t$. By definition of $\mathcal{V}$, $|S_t| = f_t N$ for every $t \in [T]$. Denote

$$T_\beta := \{t \in [T] \mid f_t > \beta\}.$$

We now set $\beta := \Delta/(2T_{\mathcal{P}})$, and restrict our attention to elements in T_β. This is possible because the f_t corresponding to $t \notin T_\beta$ contribute only a small amount of mass. Precisely,

$$\sum_{t \in [T] \setminus T_\beta} f_t \leq |[T] \setminus T_\beta| \cdot \frac{\Delta}{2T_{\mathcal{P}}} \leq \frac{\Delta}{2}, \tag{15}$$

where the first inequality holds because $|[T] \setminus T_\beta| \leq T$, and $T \leq T_{\mathcal{P}}$ by definition.

Fix some $t \in T_\beta$. By the guarantee on graph G from Claim 3.4, we see that

$$|\{\ell \in V_L \mid |\Gamma(\ell) \cap S_t| < (f_t - \eta)d\}| \leq \zeta N = \frac{\delta_{\text{out}}}{2T_{\mathcal{P}}} \cdot N < \frac{\delta_{\text{out}}}{T} \cdot N.$$

By applying a union bound argument over all $t \in T_\beta$,

$$|\{\ell \in V_L \mid \forall t \in T_\beta, |\Gamma(\ell) \cap S_t| \geq (f_t - \eta)\, d\}| > (1 - \delta_{\text{out}})N. \tag{16}$$

Denote this set by $L^* \subseteq [N]$, and note that $|L^*| > (1 - \delta_{\text{out}})N$.

We now argue that for every $\ell \in L^*$, the matrix $A^{\mathsf{proj}}_\psi(\ell)$ satisfies some $\mathcal{V}^{\mathsf{imp}}_\Delta \in \text{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$. Informally, this can be seen as follows: fix some part $\ell \in L^*$. Then, the proportion of incoming edges that are adjacent to some right vertex set S_t is roughly equal to the fractional size of S_t, which is equal to f_t. This implies that the proportion of rows in $A^{\mathsf{proj}}(\ell)$ that lie in $\ker \mathbf{M}_t$ is also roughly f_t. Since $(A^{\mathsf{proj}}(\ell))[j][] \in \ker \mathbf{M}_t \Rightarrow (A^{\mathsf{proj}}_\psi(\ell))[j][] \in \ker \mathbf{M}^{\mathsf{imp}}_t$, we see that for all t, roughly f_t fraction of rows of $A^{\mathsf{proj}}_\psi(\ell)$ lie in $\ker \mathbf{M}^{\mathsf{imp}}_t$. Because we need to account for the fact that these fractions may not be exact, we see that the matrix satisfies a robust local profile description of $\mathcal{V}^{\mathsf{imp}}$.

We now formalize this argument. Firstly, note that Eq. (14) implies the following

$$\forall r \in [N], \forall j \in [d], A_{\mathsf{fl}}[(r, j)][] \in \ker M_r. \tag{17}$$

This can be seen by applying the following simple fact to each row of A.

Fact 5.34. *If for vectors $u \in \mathbb{F}_Q^L$ and $v \in \mathbb{F}_q^L$, $\langle u, v\rangle = 0$ holds, then $u_{\mathsf{fl}} \cdot v^\top = \mathbf{0}^d$ holds as well.*

Observe that Eq. (17), along with the definition of φ_G now implies

$$\forall \ell, r \in [N], \forall i, j \in [d], (\varphi_G(\ell, i) = (r, j) \wedge A_{\mathsf{fl}}[(r, j)][] \in \ker M_r) \iff A^{\mathsf{proj}}[(\ell, i)][] \in \ker M_r. \tag{18}$$

Fix an $\ell \in L^*$. From Eqs. (16) and (18), we see that

$$\forall t \in T_\beta, \left|\{i \in [d] \mid (A^{\mathsf{proj}}(\ell))[i][] \in \ker \mathbf{M}_t\}\right| \geq (f_t - \eta)\, d.$$

From the definition of A^{proj}_ψ,

$$\forall t \in T_\beta, \left|\{i \in [d] \mid (A^{\mathsf{proj}}_\psi(\ell))[i][] \in \ker \mathbf{M}^{\mathsf{imp}}_t\}\right| \geq (f_t - \eta)\, d. \tag{19}$$

Furthermore, we will denote the set

$$Z_t := \left\{i \in [d] \mid A^{\mathsf{proj}}_\psi(\ell)[i][] \in \ker \mathbf{M}^{\mathsf{imp}}_t\right\}.$$

Observe that $Z_t \subseteq [d]$, and moreover,

$$|Z_t| \geq \lfloor (f_t - \eta)\, d \rfloor. \tag{20}$$

Because $f_t > \Delta/(2T_{\mathcal{P}})$ and $d \geq 4T_{\mathcal{P}}/\Delta$, we see that Eq. (19) implies $|Z_t| \geq 1$ for all $t \in T_\beta$.

Now consider

$$\mathcal{V}^{\mathsf{imp}}_\Delta := \left((f_t - (\Delta/(4T_{\mathcal{P}})), \mathbf{M}^{\mathsf{imp}}_t)_{t \in T_\beta}, (0, \mathbf{M}^{\mathsf{imp}}_t)_{t \notin T_\beta}, (|T_\beta| \cdot \Delta)/4T_{\mathcal{P}} + \textstyle\sum_{t \notin T_\beta} f_t, \mathbf{0}\right).$$

As we have

$$\frac{|T_\beta| \cdot \Delta}{4T_{\mathcal{P}}} + \sum_{t \notin T_\beta} f_t \leq \frac{\Delta}{4} + \frac{\Delta}{2} \leq \Delta,$$

where the first inequality follows from Eq. (15), we see that $\mathcal{V}^{\mathsf{imp}}_\Delta$ belongs to $\text{Rob}_\Delta(\mathcal{V}^{\mathsf{imp}})$. We now prove that $A^{\mathsf{proj}}_\psi(\ell)$ satisfies $\mathcal{V}^{\mathsf{imp}}_\Delta$. We do so by creating a local profile $M_d(\mathcal{V}^{\mathsf{imp}}_\Delta) \in (\mathcal{V}^{\mathsf{imp}}_\Delta)_d$ and then showing that $A^{\mathsf{proj}}_\psi(\ell)$ satisfies $\mathcal{V}^{\mathsf{imp}}_\Delta$ with $M_d(\mathcal{V}^{\mathsf{imp}}_\Delta)$ as a witness. The local profile $M_d(\mathcal{V}^{\mathsf{imp}}_\Delta)$ is created as follows:

(i) For every $t \in T_\beta$, define $\mathbf{M}^{\mathsf{imp}}_t$ to be the corresponding matrix for some

$$\left\lfloor \left(f_t - \frac{\Delta}{4T_{\mathcal{P}}}\right) d \right\rfloor$$

rows indexed by $i \in Z_t$. These rows are guaranteed to exist by Eq. (20), and the fact that $\eta = \Delta/4T_{\mathcal{P}}$.

(ii) For the rest of the at most

$$\begin{aligned} d - \sum_{t \in T_\beta} \left\lfloor \left(f_t - \frac{\Delta}{4T_{\mathcal{P}}} \right) d \right\rfloor &\leq d - \sum_{t \in T_\beta} \left(\left(f_t - \frac{\Delta}{4T_{\mathcal{P}}} \right) d - 1 \right) \\ &= d - \sum_{t \in T_\beta} f_t \cdot d + \frac{|T_\beta| \cdot \Delta \cdot d}{4T_{\mathcal{P}}} + |T_\beta| \\ &\leq \left(1 - \sum_{t \in T_\beta} f_t \right) \cdot d + \frac{\Delta \cdot d}{4} + T_{\mathcal{P}} \\ &\leq \sum_{t \notin T_\beta} f_t \cdot d + \frac{\Delta \cdot d}{2} \\ &\leq \Delta \cdot d \end{aligned}$$

rows, let the corresponding matrix be $\mathbf{0}$. Note that the last inequality follows from Eq. (15).

It is now seen that $A_\psi^{\mathsf{proj}}(V_L^p)$ satisfies $\mathcal{V}_\Delta^{\mathsf{imp}}$, with $M(\mathcal{V}_\Delta^{\mathsf{imp}})_d$ as a witness. Indeed, this is seen to follow from Eq. (19) for the rows corresponding to item (i). For the other rows, note that every row in $A_\psi^{\mathsf{proj}}(\ell)$ belongs to $\psi(\mathbb{F}_q^L) \subseteq \mathbb{F}_q^{(L - \dim W_{\mathcal{V}})} = \ker \mathbf{0}^{(L - \dim W_{\mathcal{V}})}$.

Thus, we see that for each $\ell \in L^*$, $A_\psi^{\mathsf{proj}}(\ell)$ satisfies $\mathcal{V}_\Delta^{\mathsf{imp}}$. □

6 Consequences

We state an important corollary of Theorem 5.32.

Corollary 6.1. *For a reasonable L-LCL property $\mathcal{P}$, let $R_{in} = R_{\mathcal{P}} - \varepsilon$ for some $\varepsilon > 0$. Let $\mathcal{C}_{in}, \mathcal{C}_{out}$, and G be as defined in Theorem 5.32. Furthermore, let $\mathcal{C}_{out} \subseteq (\mathbb{F}_q^{R_{in} d})^N$ be an $\mathbb{F}_q$-linear code with rate $R_{out} = 1 - \varepsilon$ and distance $\delta_{out} \geq \varepsilon^3$. Then, $\mathcal{C}_{AEL}(\mathcal{C}_{in}, \mathcal{C}_{out}, G) \subseteq \mathbb{F}_Q^N$ is an $\mathbb{F}_q$-linear code that **does not** satisfy $\mathcal{P}$, has rate $R_{AEL} > R_{\mathcal{P}} - 2\varepsilon$, and*

$$d = O\left(\frac{L^2 \cdot T_{\mathcal{P}}^3}{\varepsilon^5} \right). \tag{21}$$

Additionally, $\mathcal{C}_{AEL}(\mathcal{C}_{in}, \mathcal{C}_{out}, G)$ can be constructed in time $\mathrm{poly}(N)$.

Proof. The $\mathbb{F}_q$-linearity of $\mathcal{C}_{\mathrm{AEL}}$, and the fact that it does not satisfy $\mathcal{P}$ are established in Theorem 5.32. The rate is given by

$$R_{\mathrm{AEL}} = \frac{\log_Q(|\mathcal{C}_{\mathrm{AEL}}|)}{N}.$$

Indeed, it is easy to see that $|\mathcal{C}_{\mathrm{AEL}}| = |\mathcal{C}_{\mathrm{out}}| = q^{R_{\mathrm{in}} R_{\mathrm{out}} N d}$. Recalling that $Q = q^d$, a simple calculation gives $R_{\mathrm{AEL}} = R_{\mathrm{in}} \cdot R_{\mathrm{out}} = (R_{\mathcal{P}} - \varepsilon)(1 - \varepsilon) > R_{\mathcal{P}} - 2\varepsilon$. The value for d is obtained by recalling that $d = O(1/\zeta\eta^2)$, $\eta = \Delta/(4T_{\mathcal{P}})$, $\zeta = \delta_{\mathrm{out}}/2T_{\mathcal{P}}, \Delta = \varepsilon/2L$, and plugging in the value for δ_{out} in ζ.

We note that explicit constructions of $\mathbb{F}_q$-linear codes having rate $1 - \varepsilon$ and distance ε^3 that are constructible in time $\mathrm{poly}(N)$ can be obtained by using Tanner codes (see Corollary 11.4.8 in [GRS23]). Moreover, the graph G can be constructed in time $\mathrm{poly}(N)$, by Claim 3.4. By Lemma 5.31, $\mathcal{C}_{\mathrm{in}}$ can be constructed in time $q^{\mathrm{poly}(d,L)} \leq \mathrm{poly}(N)$. Upon invoking Observation 3.5, we see that $\mathcal{C}_{\mathrm{AEL}}(\mathcal{C}_{\mathrm{in}}, \mathcal{C}_{\mathrm{out}}, G)$ can be constructed in time $\mathrm{poly}(N)$. □

We now present a general definition of list-recovery, from which the special cases of interest can be derived.

Definition 6.2 (List Recovery with Erasures)**.** *For integers ℓ, L, where $\ell \leq L$, and constants $0 \leq \sigma \leq \rho \leq 1$, we say that a code $\mathcal{C} \subseteq \Sigma^n$ is* (ρ, σ, ℓ, L)-average radius list-recoverable with erasures *if the following holds. For every collection of sets $S_1, \ldots, S_n \subseteq (\Sigma \cup \perp)$ satisfying*

1. *$\forall i \in [n], |S_i| \leq \ell$, and*
2. *$\exists J \subseteq [n]$ satisfying $|J| \leq \sigma n$, such that $S_j = \{\perp\}$ for all $j \in J$,*

we have for all pairwise distinct codewords $c_1, \ldots, c_{L+1} \in \mathcal{C}$:

$$\frac{\sum_{k \in [L+1]} \sum_{i \in [n] \setminus J} \mathbb{1}[c_k[i] \in S_i]}{L+1} \leq (1 - \rho - \sigma)n. \tag{22}$$

We say that $\mathcal{C}$ is (ρ, σ, ℓ, L)-list-recoverable with erasures *if*

$$\min_{k \in [L+1]} |\{i \in ([n] \setminus J) \mid c_k[i] \in S_i\}| \leq (1 - \rho - \sigma)n.$$

Clearly, a code that is (ρ, σ, ℓ, L)-average radius list-recoverable with erasures is also (ρ, σ, ℓ, L)-list-recoverable with erasures.

Definition 6.3 (List Recoverability)**.** *For integers ℓ, L, where $\ell \leq L$, and constant $0 \leq \rho \leq 1$, we say that $\mathcal{C}$ is* (ρ, ℓ, L)-average radius list-recoverable *if $\mathcal{C}$ is $(\rho, 0, \ell, L)$-average radius list-recoverable with erasures.*

Definition 6.4 (List Decodability)**.** *For an integer L and constants $0 \leq \sigma \leq \rho \leq 1$, we say that a code $\mathcal{C} \subseteq \Sigma^n$ is* (ρ, σ, L)-average radius erasure list-decodable *if $\mathcal{C}$ is $(\rho, \sigma, 1, L)$-average radius list-recoverable with erasures.*

We say that $\mathcal{C}$ is (ρ, L)-average radius list-decodable *if $\mathcal{C}$ is $(\rho, 1, L)$-average radius list-recoverable.*

Definition 6.5 (Zero Error List Recoverability)**.** *For integers ℓ, L, where $\ell \leq L$, we say that a code $\mathcal{C} \subseteq \Sigma^n$ is* (ℓ, L)-zero error list-recoverable *if $\mathcal{C}$ is $(0, \ell, L)$-average radius list-recoverable.*

Definition 6.6 (List Recovery from Erasures)**.** *For integers ℓ, L, where $\ell \leq L$, and a constant $0 \leq \sigma \leq 1$, we say that a code $\mathcal{C} \subseteq \Sigma^n$ is* (σ, ℓ, L)-erasure list-recoverable *if $\mathcal{C}$ is $(0, \sigma, \ell, L)$-list recoverable with erasures.*

Definition 6.7 (Perfect Hash Matrix)**.** *For integer $t \geq 2$, a code $\mathcal{C} \subseteq \Sigma^n$ of rate R is defined to be a* $(n, |\Sigma|^{Rn}, t)$-perfect hash matrix *if it is $(0, t-1, t-1)$-list-recoverable.*

For integers ℓ, L, where $\ell \leq L$, and constants $0 \leq \sigma \leq \rho \leq 1$, let $\mathcal{P}(\rho, \sigma, \ell, L)$ be the property of **not** being (ρ, σ, ℓ, L)-average radius list-recoverable with erasures.

Claim 6.8. *Property $\mathcal{P}(\rho, \sigma, \ell, L)$ is a reasonable $(L+1)$-LCL property. In particular, we have*

$$\kappa_q(\mathcal{P}(\rho, \sigma, \ell, L)) = \log_q(\ell + 1)^{(L+1)}.$$

Furthermore, $T_{\mathcal{P}(\rho,\sigma,\ell,L)} \leq (\ell+1)^{(L+1)}$.

Proof. For convenience, we use the shorthand $\mathcal{P}$ to denote $\mathcal{P}(\rho, \sigma, \ell, L)$ throughout this proof. In order to prove this claim, we need to construct a set $\mathcal{P}$ of $(L+1)$-local profile descriptions such that a code $\mathcal{C}$ is not (ρ, σ, ℓ, L)-average radius list-recoverable with erasures if and only if it contains some local profile description from $\mathcal{P}$. Note that the local profile descriptions we create will depend on the characteristic of the field over which the codes exist, and we will denote that characteristic by p.

For every subset $K \subseteq [L+1]$, denote the set of partitions of K consisting of at most ℓ parts by P_K. For every subset $K \in (2^{[L+1]} \setminus \emptyset)$ and partition $P \in P_K$, create the matrix $\mathbf{M}(K, P)$, whose rows belong to $\mathbb{F}_p^{L+1}$, and are linear constraints that, for any prime power q of p, are satisfied by exactly the vectors in the following set:

$$\left\{v \in \mathbb{F}_q^{L+1} \mid \forall i, j \in K, i, j \text{ belong to the same part of } P \Rightarrow v[i] = v[j]\right\}. \tag{23}$$

Next, create the matrix $\mathbf{0}$, whose kernel is all of $\mathbb{F}_q^{L+1}$, and associate with it a fraction $f_{\mathbf{0}}$.

We denote the fraction associated with the matrix $\mathbf{M}(K,P)$ by $f_{K,P}$. Create a local profile description for every set of associated fractions that satisfy

$$\forall K \in \left(2^{[L+1]} \setminus \emptyset\right), \forall P \in P_K, 0 \leq f_{K,P} \leq 1, \tag{24}$$

$$f_{\mathbf{0}} \geq \sigma, \tag{25}$$

$$\sum_{K \in (2^{[L+1]} \setminus \emptyset)} \sum_{P \in P_K} |K| \cdot f_{K,P} > (1-\rho-\sigma)(L+1), \tag{26}$$

and

$$f_{\mathbf{0}} + \sum_{K \in (2^{[L+1]} \setminus \emptyset)} \sum_{P \in P_K} f_{K,P} = 1. \tag{27}$$

Lastly, we collect these local profile descriptions in a set $\mathcal{P}$.

We first prove that if a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ contains some

$$\mathcal{V} = \left((f_{K,P_K}, \mathbf{M}(K,P_K))_{K \in (2^{[L+1]} \setminus \emptyset), P \in P_K}, (f_{\mathbf{0}}, \mathbf{0})\right) \in \mathcal{P},$$

then it is **not** (ρ, σ, ℓ, L)-average radius list-recoverable with erasures. The code $\mathcal{C}$ containing a $\mathcal{V} \in \mathcal{P}$ implies the existence of a matrix $A \in \mathbb{F}_q^{n \times (L+1)}$ such that

1. $A \subseteq \mathcal{C}$,
2. A satisfies $\mathcal{V}$, and
3. A has pairwise distinct columns.

By (1) and (3), the columns of A are pairwise distinct codewords in $\mathcal{C}$. By (2), we know that for every non-empty K and partition $P \in P_K$, there are a f_{K,P_K} fraction of rows in A whose entries agree according to the constraints set forth by $\mathbf{M}(K,P_K)$. The constraints say that the number of distinct elements appearing in the entries specified by K is at most ℓ. Thus for each coordinate $i \in [n]$ corresponding to the matrix $\mathbf{M}(K,P_K)$, we can create a subset $S_i \subseteq \mathbb{F}_q$ such that $|S_i| \leq \ell$, and for all $k \in K$, we have $A[i][k] \in S_i$. According to Eq. (27), such a subset S_i can be created for $(1-f_{\mathbf{0}})n$ coordinates. For the remaining $f_{\mathbf{0}}n$ coordinates, we create the set $\{\perp\}$. Denote this set of coordinates by J. According to Eq. (26),

$$\sum_{i \in [n] \setminus J} \sum_{k \in [L+1]} \mathbb{1}[A[i][k] \in S_i] \geq \sum_{K \in (2^{[L+1]} \setminus \emptyset)} \sum_{P \in P_K} |K| \cdot f_{K,P} \cdot n > (1-\rho-\sigma)(L+1)n.$$

Therefore, we see that for a set of $L+1$ pairwise distinct codewords in $\mathcal{C}$, Eq. (22) is not satisfied, hence $\mathcal{C}$ is **not** (ρ, σ, ℓ, L)-average radius list-recoverable with erasures.

We now prove the other direction. Suppose that the code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **not** (ρ, σ, ℓ, L)-average radius list-recoverable with erasures. Then, there exists a collection of sets $S_1, \ldots, S_n \subseteq (\mathbb{F}_q \cup \perp)$ satisfying

1. $\forall i \in [n], |S_i| \leq \ell$, and
2. $\exists J \subseteq [n]$, satisfying $|J| \leq \sigma n$, such that $S_j = \{\perp\}$ for all $j \in J$.

Furthermore, there exists a matrix $A \in \mathbb{F}_q^{n \times (L+1)}$ satisfying

1. $A \subseteq \mathcal{C}$,
2. A has pairwise distinct columns, and
3. $\sum_{i \in [n] \setminus J} \sum_{k \in [L+1]} \mathbb{1}[A[i][k] \in S_i] > (1-\rho-\sigma)(L+1)n$.

For every coordinate $i \in [n] \setminus J$, we can assign i a type (K, P), where K is a subset of $[L+1]$ and P is some partition in P_K. The type assigned to i will be $(K = \{k \in [L+1] \mid A[i][k] \in S_i\}, P)$, where P is a partition of K such that for every k_1, k_2 belonging to the same part, $A[i][k_1] = A[i][k_2]$ holds. We now create a matrix $\mathbf{M}(K, P)$ for each type (K, P), whose entries belong to $\mathbb{F}_p$ and whose rows consist of linear constraints that are satisfied by vectors in $\mathbb{F}_q^{L+1}$ that belong to the set described in Eq. (23). Clearly, row $A[i][]$ satisfies the constraints of $\mathbf{M}(K, P)$ if its type is (K, P). The fraction of rows having type (K, P) is denoted by $f_{K,P}$. Therefore, from (3), we see that

$$\sum_{K \in (2^{[L+1]} \setminus \emptyset)} \sum_{i \in [n] \setminus J} K \cdot \mathbb{1}[\{k \in [L+1] \mid A[i][k] \in S_i\} = K] > (1 - \rho - \sigma)(L+1)n. \tag{28}$$

We also see that if the equality $\{k \in [L+1] \mid A[i][k] \in S_i\} = K$ is satisfied for some coordinate i and K, then there is a matrix $\mathbf{M}(K, P)$ for some $P \in P_K$ such that row $A[i][]$ satisfies the constraints set by $\mathbf{M}(K, P)$. Thus, we see that Eq. (28), upon being divided by the block length n, is exactly the same as Eq. (26).

For those coordinates that are present in J, and also those coordinates for which $K = \emptyset$, we assign the matrix $\mathbf{0}$, and note that $f_{\mathbf{0}}$ is at least σ. We then see that Eq. (24), Eq. (25), and Eq. (27) are also satisfied, and thus A satisfies a local profile description that belongs to $\mathcal{P}$.

Lastly, we prove that $\mathcal{P}$ is a *reasonable* $(L+1)$-LCL property. This is seen by observing that the number of matrices of the form $\mathbf{M}(K, P)$, as constructed above, is at most $(\ell+1)^{(L+1)}$, which is an upper bound on the number of partitions consisting of at most ℓ parts, of every subset of $[L+1]$. Thus, $T_{\mathcal{P}} \leq (\ell+1)^{(L+1)}$, and the number of local profiles associated with this property is at most

$$(\ell+1)^{(L+1)n} = q^{\log_q(\ell+1)^{(L+1)} \cdot n},$$

and we see that $\kappa_q(\mathcal{P}) = \log_q(\ell+1)^{(L+1)}$ goes to zero as $q \to \infty$. □

We then obtain the following corollary.

Corollary 6.9. *For $\mathcal{P}(\rho, \sigma, \ell, L)$, let $\mathcal{C}_{in}, \mathcal{C}_{out}, G$ be as defined in Theorem 5.32. Furthermore, let $\mathcal{C}_{out} \subseteq (\mathbb{F}_q^{R_{in}d})^N$ be a code with rate $R_{out} = 1 - \varepsilon$ and distance $\delta_{out} \geq \varepsilon^3$. Then, $\mathcal{C}_{AEL}(\mathcal{C}_{in}, \mathcal{C}_{out}, G) \subseteq \mathbb{F}_Q^N$ is a $\mathbb{F}_q$-linear code that is (ρ, σ, ℓ, L)-average radius list-recoverable with erasures, has rate $R_{AEL} > R_{\mathcal{P}} - 2\varepsilon$, and can be constructed in time* $\mathrm{poly}(N)$. *Additionally, we have $q \leq (\ell+1)^{(L+1)\cdot\frac{8}{\varepsilon}}$, and $Q = q^d$, where*

$$d = O\left(\frac{L^2 \cdot (\ell+1)^{3(L+1)}}{\varepsilon^5}\right).$$

Proof. By Claim 6.8, $\mathcal{P}(\rho, \sigma, \ell, L)$ is a *reasonable* property. We now turn to set a value for q so that Eq. (13) is satisfied. For $q = (\ell+1)^{(L+1)\cdot\frac{8}{\varepsilon}}$, we see

$$\log_q(\ell+1)^{(L+1)} + \log_q 2 - \frac{\varepsilon}{4} = \frac{\varepsilon}{8} + \frac{1}{\log_2 q} - \frac{\varepsilon}{4} < 0$$

Therefore, Corollary 6.1 implies that $\mathcal{C}_{\mathrm{AEL}}$ does not satisfy $\mathcal{P}(\rho, \sigma, \ell, L)$, and thus, is (ρ, σ, ℓ, L)-average radius list-recoverable with erasures.

Finally, Eq. (21) from Corollary 6.1, along with the value of $T_{\mathcal{P}(\rho,\sigma,\ell,L)}$ from Claim 6.8 implies

$$d = O\left(\frac{L^2 \cdot (\ell+1)^{3(L+1)}}{\varepsilon^5}\right).$$

Plugging the value into $Q = q^d$ gives us the final alphabet size. □

We now record several results, corresponding to the special cases defined above. The proofs for them follow from Corollary 6.9. The first details parameters for list recovery at capacity.

Corollary 6.10 (Explicitly Achieving Capacity for List Recovery)**.** *For a fixed input list size ℓ, and a fixed radius $\rho > 0$, let $L_{\rho,\ell}$ be the smallest output list size such that $R_{\mathcal{P}(\rho,0,\ell,L_{\rho,\ell})} \geq 1-\rho$. For $\varepsilon > 0$, let $R := 1-\rho-\varepsilon \leq R_{\mathcal{P}(\rho,0,\ell,L_{\rho,\ell})} - \varepsilon$. The code $\mathcal{C}_{AEL}$ as in Corollary 6.9 is $(1-R-\varepsilon, \ell, L_{\rho,\ell})$-average radius list-recoverable, with rate $R_{AEL} > R_{\mathcal{P}(\rho,0,\ell,L_{\rho,\ell})} - 2\varepsilon$, and alphabet size at most $\exp\big((L_{\rho,\ell}/\varepsilon)^{O(L_{\rho,\ell})}\big)$.*

Corollary 6.11 (Explicit Zero Error List Recovery)**.** *For $\varepsilon > 0$ and a fixed input list size ℓ and rate R, let $L_{R,\ell}$ be the smallest output list size such that $R_{\mathcal{P}(0,0,\ell,L_{R,\ell})} - \varepsilon \geq R$. Then, the code $\mathcal{C}_{AEL}$ as in Corollary 6.9 is $(\ell, L_{R,\ell})$-zero error list recoverable, with rate $R_{AEL} > R_{\mathcal{P}(0,0,\ell,L_{R,\ell})} - 2\varepsilon$, and alphabet size at most $\exp\big((L_{R,\ell}/\varepsilon)^{O(L_{R,\ell})}\big)$.*

Corollary 6.12 (Explicit List Recovery from Erasures)**.** *For constants $\varepsilon > 0$, $0 \leq \sigma \leq 1$, a fixed input list size ℓ and rate R, let $L_{\sigma,R,\ell}$ be the smallest output list size such that $R_{\mathcal{P}(0,\sigma,\ell,L_{\sigma,R,\ell})} - \varepsilon \geq R$. Then, the code $\mathcal{C}_{AEL}$ as in Corollary 6.9 is $(\sigma, \ell, L_{\sigma,R,\ell})$-erasure list recoverable, with rate $R_{AEL} > R_{\mathcal{P}(0,\sigma,\ell,L_{\sigma,R,\ell})} - 2\varepsilon$, and alphabet size at most $\exp\big((L_{\sigma,R,\ell}/\varepsilon)^{O(L_{\sigma,R,\ell})}\big)$.*

Corollary 6.13 (Explicit Perfect Hash matrices)**.** *For an integer $t \geq 2$, let $\mathcal{C}_{AEL}$ be the code as in Corollary 6.9 for the property $\mathcal{P}(0, 0, t-1, t-1)$. Then, $\mathcal{C}_{AEL}$ is a code of rate $R = R_{\mathcal{P}(0,0,t-1,t-1)} - 2\varepsilon$ that is $(0, t-1, t-1)$-list recoverable. The alphabet size is at most $\exp((t/\varepsilon)^{O(t)})$ and moreover, the codewords of $\mathcal{C}_{AEL}$, when arranged as columns in a $N \times Q^{RN}$ matrix, form a perfect hash matrix.*

Proof. The list-recoverability of $\mathcal{C}_{\text{AEL}}$ follows from Corollary 6.9. We prove that the set of codewords of a $(0, t-1, t-1)$-list-recoverable code can be arranged as columns in a matrix to form a perfect hash matrix. Indeed, $(0, t-1, t-1)$-list recoverability implies that the number of pairwise distinct codewords having at most $t-1$ distinct entries in every row is at most $t-1$. This implies that for any set of t codewords, there exists at least one index on which the t codewords have pairwise distinct entries. □

7 Acknowledgements

We thank Jonathan Mosheiff for providing feedback and comments on a draft of the paper.

References

[AD24] Ron Asherov and Irit Dinur. Bipartite unique neighbour expanders via ramanujan graphs. *Entropy*, 26(4):348, 2024. 2

[AEL95] N. Alon, J. Edmonds, and M. Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519, 1995. 2

[AGL24] Omar Alrabiah, Venkatesan Guruswami, and Ray Li. Randomly punctured reed-solomon codes achieve list-decoding capacity over linear-sized fields. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1458–1469. ACM, 2024. 4, 7, 11, 12, 13

[AN96] Noga Alon and Moni Naor. Derandomization, witnesses for boolean matrix multiplication and construction of perfect hash functions. *Algorithmica*, 16(4/5):434–449, 1996. 5

[Ari09] E. Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009. 1

[BGM23] Joshua Brakensiek, Sivakanth Gopi, and Visu Makam. Generic reed-solomon codes achieve list-decoding capacity. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1488–1501, 2023. 4, 12

[Bla00] Simon R. Blackburn. Perfect hash families: Probabilistic methods and explicit constructions. *J. Comb. Theory A*, 92(1):54–60, 2000. 5

[BLVW19] Zvika Brakerski, Vadim Lyubashevsky, Vinod Vaikuntanathan, and Daniel Wichs. Worst-case hardness for lpn and cryptographic hashing via code smoothing. In *Advances in Cryptology – EUROCRYPT 2019*, 2019. 6

[BW98] Simon R. Blackburn and Peter R. Wild. Optimal linear perfect hash families. *J. Comb. Theory A*, 83(2):233–250, 1998. 5

[CCS25] Yeyuan Chen, Mahdi Cheraghchi, and Nikhil Shagrithaya. Optimal erasure codes and codes on graphs. *CoRR*, abs/2504.03090, 2025. 2

[Che25] Yeyuan Chen. Unique-neighbor expanders with better expansion for polynomial-sized sets. In Yossi Azar and Debmalya Panigrahi, editors, *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025*, pages 3335–3362. SIAM, 2025. 2

[CZ25] Yeyuan Chen and Zihan Zhang. Explicit Folded Reed-Solomon and Multiplicity Codes Achieve Relaxed Generalized Singleton Bounds. In *Proceedings of the 57th ACM Symposium on Theory of Computing*, 2025. 4

[DEL+22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022. 2

[DMS03] Ilya Dumer, Daniele Micciancio, , and Madhu Sudan. Hardness of approximating the minimum distance of a linear code. *IEEE Transactions on Information Theory*, 49(1):22–37, 2003. 6

[DW22] Dean Doron and Mary Wootters. High-probability list-recovery, and applications to heavy hitters. In Mikolaj Bojanczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022, July 4-8, 2022, Paris, France*, volume 229 of *LIPIcs*, pages 55:1–55:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. 3, 5

[FKS82] Michael L. Fredman, János Komlós, and Endre Szemerédi. Storing a sparse table with O(1) worst case access time. In *23rd Annual Symposium on Foundations of Computer Science, Chicago, Illinois, USA, 3-5 November 1982*, pages 165–169. IEEE Computer Society, 1982. 5

[FM04] Uriel Feige and Daniele Micciancio. The inapproximability of lattice and coding problems with preprocessing. *J. Comput. Syst. Sci.*, 69(1):45–67, 2004. 6

[GHK11] Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. *IEEE Trans. Inf. Theory*, 57(2):718–725, 2011. 4

[GHSZ02] Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Trans. Inf. Theory*, 48(5):1021–1034, 2002. 4

[GI02] Venkatesan Guruswami and Piotr Indyk. Near-optimal linear-time codes for unique decoding and new list-decodable codes over small alphabets. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 812–821, 2002. 2

[GI03] Venkatesan Guruswami and Piotr Indyk. Linear time encodable and list decodable codes. In *Proceedings of the 35th ACM Symposium on Theory of Computing*, 2003. 3

[Gil52] E.N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31:504–522, 1952. 1

[GL89] O. Goldreich and L. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the 21st ACM Symposium on Theory of Computing*, pages 25–32, 1989. 1

[GLM+22] 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 Trans. Inf. Theory*, 68(2):923–939, 2022. 2

[GLS+24] Zeyu Guo, Ray Li, Chong Shangguan, Itzhak Tamo, and Mary Wootters. Improved list-decodability and list-recoverability of reed–solomon codes via tree packings. *SIAM Journal on Computing*, 53(2):389–430, 2024. 4, 12

[GM22] Venkatesan Guruswami and Jonathan Mosheiff. Punctured low-bias codes behave like random linear codes. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 36–45. IEEE, 2022. 2

[GMR+22] Venkatesan Guruswami, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Threshold rates for properties of random codes. *IEEE Trans. Inf. Theory*, 68(2):905–922, 2022. 3

[Gol24] Louis Golowich. New explicit constant-degree lossless expanders. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 4963–4971. SIAM, 2024. 2

[GR06a] Venkatesan Guruswami and Atri Rudra. Explicit capacity-achieving list-decodable codes. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 1–10, 2006. 1

[GR06b] Venkatesan Guruswami and Atri Rudra. Explicit capacity-achieving list-decodable codes. In Jon M. Kleinberg, editor, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 1–10. ACM, 2006. 3, 4

[GRS23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. Book, 2023. 1, 16, 29

[GS98] Venkatesan Guruswami and Madhu Sudan. Improved decoding of reed-solomon and algebraic-geometric codes. In *39th Annual Symposium on Foundations of Computer Science, FOCS 1998, Palo Alto, California, USA, November 8-11, 1998*, pages 28–39. IEEE Computer Society, 1998. 4

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh-vardy codes. *J. ACM*, 56(4):20:1–20:34, 2009. 3

[GZ23] Zeyu Guo and Zihan Zhang. Randomly punctured reed-solomon codes achieve the list decoding capacity over polynomial-size alphabets. In *Proceedings of the 64rd IEEE Symposium on Foundations of Computer Science*, 2023. 4, 12

[HLM+25a] Jun-Ting Hsieh, Ting-Chun Lin, Sidhanth Mohanty, Ryan O'Donnell, and Rachel Yun Zhang. Explicit two-sided vertex expanders beyond the spectral barrier. In Michal Koucký and Nikhil Bansal, editors, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*, pages 833–842. ACM, 2025. 2

[HLM+25b] Jun-Ting Hsieh, Alexander Lubotzky, Sidhanth Mohanty, Assaf Reiner, and Rachel Yun Zhang. Explicit lossless vertex expanders. *CoRR*, abs/2504.15087, 2025. 2

[HMMP24] Jun-Ting Hsieh, Theo McKenzie, Sidhanth Mohanty, and Pedro Paredes. Explicit two-sided unique-neighbor expanders. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 788–799. ACM, 2024. 2

[HRW20] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes and applications. *SIAM J. Comput.*, 49(4), 2020. 3

[JMST25] Fernando Granha Jeronimo, Tushant Mittal, Shashank Srivastava, and Madhur Tulsiani. Explicit codes approaching generalized singleton bound using expanders. In *Proceedings of the 57th ACM Symposium on Theory of Computing*, 2025. 2, 4, 5, 6

[JS25] Fernando Granha Jeronimo and Aman Singh. List decoding expander-based codes via fast approximation of expanding csps: I, 2025. 6

[KMRS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):11:1–11:42, 2017. 2, 6, 10

[KRSW23] Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved list decoding of folded reed-solomon and multiplicity codes. *SIAM J. Comput.*, 2023. 2, 3, 4, 5, 6, 10

[LMS25] Matan Levi, Jonathan Mosheiff, and Nikhil Shagrithaya. Random reed-solomon codes and random linear codes are locally equivalent, 2025. 3, 4, 8, 17, 19, 20, 21

[LNNT19] Kasper Green Larsen, Jelani Nelson, Huy L. Nguyen, and Mikkel Thorup. Heavy hitters via cluster-preserving clustering. *Commun. ACM*, 62(8):95–100, 2019. 3

[LP20] Ben Lund and Aditya Potukuchi. On the list recoverability of randomly punctured codes. In Jaroslaw Byrka and Raghu Meka, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*, volume 176 of *LIPIcs*, pages 30:1–30:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. 3, 4

[LPB06] Yi Lu, Balaji Prabhakar, and Flavio Bonomi. Perfect hashing for network applications. In *Proceedings 2006 IEEE International Symposium on Information Theory, ISIT 2006, The Westin Seattle, Seattle, Washington, USA, July 9-14, 2006*, pages 2774–2778. IEEE, 2006. 5

[LS25] Ray Li and Nikhil Shagrithaya. Near-optimal list-recovery of linear code families. *CoRR*, abs/2502.13877, 2025. 4, 5

[Meh84] Kurt Mehlhorn. *Data Structures and Algorithms 1: Sorting and Searching*, volume 1 of *EATCS Monographs on Theoretical Computer Science*. Springer, 1984. 5

[MRR+20] Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. LDPC codes achieve list decoding capacity. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 458–469. IEEE, 2020. 2, 3, 17

[MRRZ+19] Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. LDPC codes achieve list decoding capacity, 2019. 8

[NPR12] Hung Q. Ngo, Ely Porat, and Atri Rudra. Efficiently decodable compressed sensing by list-recoverable codes and recursion. In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, volume 14 of *LIPIcs*, pages 230–241. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2012. 3

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994. Preliminary version in *Proc. of FOCS'88*. 1

[NW95] Ilan Newman and Avi Wigderson. Lower bounds on formula size of boolean functions using hypergraph entropy. *SIAM J. Discret. Math.*, 8(4):536–542, 1995. 5

[PK22] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022. 2

[PSW25] Francisco Pernice, Oscar Sprumont, and Mary Wootters. List-decoding capacity implies capacity on the q-ary symmetric channel. In *Proceedings of the 57th ACM Symposium on Theory of Computing*, 2025. 1

[PV05] Farzad Parvaresh and Alexander Vardy. Correcting errors beyond the Guruswami-Sudan radius in polynomial time. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, pages 285–294, 2005. 1, 4

[Res20] Nicholas Resch. List-decodable codes:(randomized) constructions and applications. *PhD thesis, Carnegie Mellon University, 2020*, 2020. 4

[RW18] Atri Rudra and Mary Wootters. Average-radius list-recoverability of random linear codes. In *Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms*, 2018. 4

[Sha48] Claude Shannon. A mathematical theory of communications. *Bell System Technical Journal*, 27:379–423, 623–656, 1948. 1

[Sri25] Shashank Srivastava. Improved list size for folded reed-solomon codes. In Yossi Azar and Debmalya Panigrahi, editors, *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2025, New Orleans, LA, USA, January 12-15, 2025*, pages 2040–2050. SIAM, 2025. 4

[SS96] M. Sipser and D. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996. Preliminary version in *Proc. of FOCS'94*. 1

[ST20] Chong Shangguan and Itzhak Tamo. Combinatorial list-decoding of reed-solomon codes beyond the johnson radius. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 538–551, 2020. 4, 12

[ST25] Shashank Srivastava and Madhur Tulsiani. List decoding expander-based codes up to capacity in near-linear time, 2025. 4, 5, 6

[Sud97] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *J. Complex.*, 13(1):180–193, 1997. 4

[Tam24] Itzhak Tamo. Tighter list-size bounds for list-decoding and recovery of folded reed-solomon and multiplicity codes. *IEEE Trans. Inf. Theory*, 2024. 4, 5

[Tan81] Robert Michael Tanner. A recursive approach to low complexity codes. *IEEE Trans. Inf. Theory*, 27(5):533–547, 1981. 1

[Tre99] Luca Trevisan. Construction of extractors using pseudo-random generators (extended abstract). In Jeffrey Scott Vitter, Lawrence L. Larmore, and Frank Thomson Leighton, editors, *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing, May 1-4, 1999, Atlanta, Georgia, USA*, pages 141–148. ACM, 1999. 3

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, STOC 2017, pages 238–251, New York, NY, USA, 2017. ACM. 1

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Now Publishers Inc., 2012. 1

[Var57] R.R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk SSSR*, 117:739–741, 1957. 1

[YZ24] Takashi Yamakawa and Mark Zhandry. Verifiable quantum advantage without structure. *J. ACM*, 2024. 1

[ZP82] Victor V. Zyablov and Mark S. Pinsker. List cascade decoding. *Problems of Information Transmission*, 1982. 1, 4

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il