

List-Recovery of Random Linear Codes over Small Fields

Dean Doron   

Ben-Gurion University of the Negev, Beer-Sheva, Israel

Jonathan Mosheiff   

Ben-Gurion University of the Negev, Beer-Sheva, Israel

Nicolas Resch   

Informatics Institute, University of Amsterdam, The Netherlands

João Ribeiro   

Instituto de Telecomunicações and Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract

We study list-recoverability of random linear codes over small fields, both from errors and from erasures. We consider codes of rate ε -close to capacity, and aim to bound the dependence of the output list size L on ε , the input list size ℓ , and the alphabet size q . Prior to our work, the best upper bound was $L = q^{O(\ell/\varepsilon)}$ (Zyablov and Pinsker, Prob. Per. Inf. 1981).

Previous work has identified cases in which *linear* codes provably perform worse than non-linear codes with respect to list-recovery. While there exist non-linear codes that achieve $L = O(\ell/\varepsilon)$, we know that $L \geq \ell^{\Omega(1/\varepsilon)}$ is necessary for list recovery from erasures over fields of small characteristic, and for list recovery from errors over large alphabets.

We show that in other relevant regimes there is no significant price to pay for linearity, in the sense that we get the correct dependence on the gap-to-capacity ε and go beyond the Zyablov–Pinsker bound for the first time. Specifically, when q is constant and ε approaches zero,

- For list-recovery from erasures over *prime fields*, we show that $L \leq C_1/\varepsilon$. By prior work, such a result cannot be obtained for low-characteristic fields.
- For list-recovery from errors over *arbitrary fields*, we prove that $L \leq C_2/\varepsilon$.

Above, C_1 and C_2 depend on the decoding radius, input list size, and field size. We provide concrete bounds on the constants above, and the upper bounds on L improve upon the Zyablov–Pinsker bound whenever $q \leq 2^{(1/\varepsilon)^c}$ for some small universal constant $c > 0$.

2012 ACM Subject Classification Theory of computation → Error-correcting codes

Keywords and phrases List recovery, random linear codes

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2025.57

Category RANDOM

Related Version *Full Version:* <https://www.arxiv.org/abs/2505.05935> [3]

Funding This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing, supported by DOE grant #DE-SC0024124.

Dean Doron: Supported in part by NSF-BSF grant #2022644.

Jonathan Mosheiff: Supported by Israel Science Foundation grant 3450/24 and an Alon Fellowship.

Nicolas Resch: Supported by an NWO (Dutch Research Council) grant with number C.2324.0590.

João Ribeiro: Supported by FCT/MECI through national funds and when applicable co-funded EU funds under UID/50008: Instituto de Telecomunicações.



© Dean Doron, Jonathan Mosheiff, Nicolas Resch, and João Ribeiro;
licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025).

Editors: Alina Ene and Eshan Chattopadhyay; Article No. 57; pp. 57:1–57:18



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Error-correcting codes enable reliable communication over noisy channels by encoding messages $m \in \Sigma^k$ as codewords $c \in \Sigma^n$. A code $\mathcal{C} \subseteq \Sigma^n$ of rate $R = k/n$ and minimum (relative) Hamming distance δ allows for reliable communication over an adversarial noisy channel that corrupts up to a $\delta/2$ -fraction of codeword symbols. To tolerate more corruptions, one can relax unique decoding to *list decoding*, where the decoder outputs all codewords within a given Hamming radius.

This notion is further generalized by *list recovery (from errors)*, which models scenarios where the receiver gets a small list of possible values for each symbol. Formally, a code \mathcal{C} is said to be (ρ, ℓ, L) -*list-recoverable* if for every sequence of sets $T_1, \dots, T_n \subseteq \Sigma$ with $|T_i| \leq \ell$, we have

$$|\mathcal{C} \cap B_\rho(T_1 \times \dots \times T_n)| \leq L,$$

where $B_\rho(T_1 \times \dots \times T_n)$ denotes the Hamming ball consisting of all words in Σ^n that agree with T_1, \dots, T_n in at least $(1 - \rho)n$ coordinates. When $\ell = 1$, list-recovery from errors reduces to standard list-decoding (from errors).

A related variant is *list recovery from erasures*, where some coordinates are entirely unknown, modeled by setting $T_i = \Sigma$. A code is said to be (α, ℓ, L) -*list-recoverable from erasures* if

$$|\mathcal{C} \cap (T_1 \times \dots \times T_n)| \leq L$$

whenever $|T_i| \leq \ell$ for at least $(1 - \alpha)n$ positions. Here too, the case $\ell = 1$ corresponds to list-decoding from erasures.

List recoverable codes are used as a building block for list-decodable and uniquely decodable codes [13, 14, 15, 16, 26, 9, 21]. They have also gained a significant independent interest, in part due to their applications in pseudorandomness [38, 19, 4, 24], algorithms (in particular, for heavy hitters, compressed sensing, and combinatorial group testing [23, 34, 28, 7, 5]), and cryptography [20, 22].

For both list-recovery from errors and from erasures, there exists a well-defined *capacity threshold* that characterizes the maximal achievable rate for which bounded list-size decoding is possible. Specifically, given parameters ρ , ℓ , and alphabet size q , there is a critical rate $R^* = R^*(\rho, \ell, q)$ such that:

- For every $\varepsilon > 0$ and any large enough block length, there exist codes of rate $R^* - \varepsilon$ that are (ρ, ℓ, L) -list-recoverable (from errors or erasures) with

$$L = O_{\ell, \varepsilon}(1). \tag{1}$$

- For every $\varepsilon > 0$ and any large enough block length n , no code of rate $R^* + \varepsilon$ is (ρ, ℓ, L) -list-recoverable for $L = q^{o(n)}$.

The exact threshold depends on the recovery model:

- For (ρ, ℓ, L) -list-recoverability from *errors*, the threshold rate is

$$R_{\text{errors}}^* = 1 - h_{q, \ell}(\rho),$$

where

$$h_{q, \ell}(\rho) = \rho \log_q \left(\frac{q - \ell}{\rho} \right) + (1 - \rho) \log_q \left(\frac{\ell}{1 - \rho} \right),$$

valid for $0 \leq \rho \leq 1 - \frac{\ell}{q}$ [35, Theorem 2.4.12].

- For (α, ℓ, L) -list-recoverability from *erasures*, the corresponding threshold is

$$R_{\text{erasures}}^* = (1 - \alpha) \cdot (1 - \log_q \ell).$$

The dependence of the list size L on the parameters ℓ and ε (see Equation (1)) is often critical, and has been the focus of extensive research (e.g., [36, 37, 32, 10, 17, 8, 39, 2, 31]).

Using the probabilistic method, it is easy to show that plain random codes achieve $L = O(\ell/\varepsilon)$, and this dependence is often viewed as the optimal benchmark. Codes achieving this tradeoff are said to match the *Elias bound for list-recovery*, in reference to the analogous threshold in list-decoding [6] (see also [33]).

To set expectations for list-recovery of *linear* codes, we recall the classic argument of Zyablov and Pinsker [40], adapted to the setting of list-recovery by Guruswami [11]. Since any set of $L + 1$ vectors has a linearly independent subset of size $\log_q(L + 1)$, and since the events that linearly independent vectors lie in a linear code are stochastically independent, the argument for plain random codes gives $L = q^{O(\ell/\varepsilon)}$. This naturally raises the question: what is the actual *price of linearity* in list-recovery? While various forms of degradation are possible, we seek to understand how the requirement of linearity affects the list-size achievable near capacity.

Prior work

Most previous results concerning the output list size have focused on the *large alphabet regime*, where q is at least exponential in $1/\varepsilon$. In this setting, Li and Shagrithaya [31] recently proved that random linear codes are almost surely list-recoverable with list size $L \leq (\ell/\varepsilon)^{O(\ell/\varepsilon)}$. By a reduction between code ensembles [30], the same upper bound also holds with high probability for Reed–Solomon codes over random evaluation sets. On the other hand, [31] proved that all (ρ, ℓ, L) -list-recoverable (from errors) linear codes over a large alphabet must satisfy $L \geq \ell^{\Omega(R/\varepsilon)}$, implying an exponential gap in the list-size between linear codes and plain random codes. A similar negative result was previously proven by Chen and Zhang [2] for Reed–Solomon codes and folded Reed–Solomon codes. Recently, Komech and Mosheiff [25] constructed a new ensemble of non-linear codes that achieve $L \leq O(\ell/\varepsilon)$ in list-recovery from errors. These are the only codes other than plain random codes known to achieve the list-recovery Elias bound.

Older works of Rudra and Wootters [36, 37], incomparable to [31], show (in our terms) that random linear codes achieve list-size $L \leq \ell^{\frac{1}{\varepsilon} \cdot \log^2(\ell/\varepsilon)}$ for list-recovery (from errors) in the large alphabet regime, but only under the guarantee that $\rho = 1 - \Omega(\varepsilon)$.

Other works concern the list-recoverability of folded Reed–Solomon codes, multiplicity codes, tensor codes, and variants of them (e.g., [27, 21, 39]). In particular, Tamo [39] shows that folded Reed–Solomon codes and multiplicity codes achieve list size $L \leq (\ell/\varepsilon)^{O(\frac{1+\log \ell}{\varepsilon})}$ (in the errors case) in the large alphabet regime.

In contrast, very little is known about list-recovery in the *small alphabet regime*, where q is sub-exponential in $1/\varepsilon$. To the best of our knowledge, the only positive result in this setting is the aforementioned $L = q^{O(\ell/\varepsilon)}$ for random linear codes [40]. On the other hand, we know that when q is a power of a small prime, random linear codes in this regime are very unlikely to be (α, ℓ, L) -list-recoverable from erasures with $\ell^{o(1/\varepsilon)}$ [17] (we state this as Theorem 1 below).

Provable separation between linear and nonlinear codes were also established in the setting of list decoding from erasures. For the special case of \mathbb{F}_2 , and aiming for an erasure decoding radius of $\alpha = 1 - \varepsilon$, Guruswami [11] showed that the output list size must satisfy

$L = \Omega(1/\varepsilon)$, as long as the rate is sufficiently non-vanishing. In contrast, plain random codes achieve $L = O(\log(1/\varepsilon))$, and we even have explicit constructions that nearly match plain random codes in some parameter regimes [1].

1.1 Our Contribution

We establish new upper bounds on the output list size for list-recovery of random linear codes near capacity in the small alphabet regime. To the best of our knowledge, these are the only known bounds for general list-recovery in this setting beyond the classical Zyablov–Pinsker argument. Notably, our results achieve list sizes with only linear dependence on $1/\varepsilon$, in contrast to the exponential dependence in previous bounds.

List-recovery from erasures over prime fields

For our first result, we consider (α, ℓ, L) -list-recovery from erasures. Recall that the capacity in this case is

$$R_{\text{erasures}}^*(\alpha, \ell, q) = (1 - \alpha)(1 - \log_q \ell).$$

As mentioned before, [17] showed that there is a price to pay for linearity over fields of small characteristic. More precisely, they proved the following.

► **Theorem 1** (informal; see [17, Theorem III.1]). *If ℓ divides $\text{char}(\mathbb{F}_q)$, then with high probability over the choice of the linear code, the output list size L cannot be taken smaller than $\ell^{\Omega(1/\varepsilon)}$.*

We show that this limitation disappears when considering a *prime* field size q . In this case, we can make the output list size $O_{\alpha, \ell, q}(1/\varepsilon)$.

► **Theorem 2** (list recovery from erasures over prime fields; see Theorem 25). *Given $1 \leq \ell \leq q$ with q prime and $\alpha \in [0, 1)$, there exists $C(\alpha, \ell, q) > 0$ such that the following holds. Let $\mathcal{C} \leq \mathbb{F}_q^n$ be a random linear code of rate $R_{\text{erasures}}^*(\alpha, \ell, q) - \varepsilon$ for some $\varepsilon > 0$. Then, \mathcal{C} is with high probability $(\alpha, \ell, \frac{C(\alpha, \ell, q)}{\varepsilon})$ -list-recoverable from erasures.*

While our focus is on the setting where α and q are constants, we determine effective bounds on $C(\alpha, \ell, q)$ even when α is a function of n , and q is slightly super-constant.¹

List-recovery from errors

Next, we consider the case of list-recovery from errors, where we recall the capacity is

$$R_{\text{errors}}^*(\rho, \ell, q) = 1 - h_{q, \ell}(\rho).$$

Here, contrary to what happens in the regime of large q -s, we do not observe any price to pay for linearity, at least in terms of the dependence on the gap-to-capacity ε .

► **Theorem 3** (list recovery from errors; see Theorem 31). *Given $1 \leq \ell \leq q$ with q a prime power and $\rho \in (0, 1 - \ell/q)$, there exists $C(\rho, \ell, q) > 0$ such that the following holds. Let $\mathcal{C} \leq \mathbb{F}_q^n$ be a random linear code of rate $R_{\text{errors}}^*(\rho, \ell, q) - \varepsilon$ for some $\varepsilon > 0$. Then, \mathcal{C} is with high probability $(\rho, \ell, \frac{C(\rho, \ell, q)}{\varepsilon})$ -list-recoverable from errors.*

¹ For a concrete example, when $\ell \leq (1 - \gamma)q$ for some constant γ , and α is bounded away from 1, we can take $C(\alpha, \ell, q) \leq q^{O(\ell \cdot \log q)}$ as long as $n \geq C(\alpha, \ell, q)$. We refer the reader to Theorem 25 for the precise bound.

Again, we determine effective bounds on the numerator $C(\rho, \ell, q)$, which can be found in Theorem 31.² For context, we recall that in the “large q regime” (say, $q \gg \ell^{1/\varepsilon}$), such a result is provably impossible. Specifically, when q is large, the list-recovery capacity is (essentially) the Singleton bound, namely, $R_{\text{errors}}^* \approx 1 - \rho$. It is known that at rate $1 - \rho - \varepsilon$, *any* linear code list-recoverable from errors must have $L \geq \ell^{\Omega(R/\varepsilon)}$ [31]. That is, the dependence of L on the gap-to-capacity *must be exponential*. But in our case, at least if ρ, ℓ and q are held constant, the dependence of L on ε is just $O(1/\varepsilon)$.

We conclude this section with some remarks.

► **Remark 4 (on the field insensitivity).** Both of our results show that there is no significant price to pay for linearity in list-recovery for certain parameter regimes. Our result in Theorem 3 on list-recovery from errors is insensitive to the field size. On the other hand, as discussed above, our result in Theorem 2 on list-recovery from erasures is (necessarily!) field sensitive. We provide some informal discussion on why this happens. First, note that we work with rates close to capacity, and the capacity in the erasures setting is larger for comparable α and ρ . Second, looking ahead (see Section 1.2 for more details), list-recovery from erasures depends on the “additive structure” of $T_1 \times \cdots \times T_n$, for arbitrary size- ℓ subsets $T_1, \dots, T_n \subseteq \mathbb{F}_q$. If the T_i ’s are *subspaces* of \mathbb{F}_q , then it is quite likely these form a bad configuration for list-recovery from erasures. In contrast, list-recovery from errors depends on the additive structure of the “puffed-up” combinatorial rectangles

$$B_\rho(T_1 \times T_2 \times \cdots \times T_n) = \{x \in \mathbb{F}_q^n : x_i \notin T_i \text{ for at most } \rho n \text{ choices of } i \in [n]\}.$$

Even if the T_i ’s are subspaces of \mathbb{F}_q , the “puffing up” operation kills any additive structure that could lead to a bad list-recovery configuration.

► **Remark 5 (on the dependence of L on the various parameters).** Recall that a code which is ε -close to capacity is said to achieve the Elias bound if $L = O(\ell/\varepsilon)$. Note that in both Theorem 2 and Theorem 3, the dependence of L on ε is as we would hope. However, the dependence on the other parameters (particularly ℓ and q) is exponentially worse than the $O(\ell)$ dependence. We leave it as a natural open problem to improve on this dependency.

► **Remark 6 (comparison to [40]).** As mentioned earlier, we are not aware of any prior arguments establishing non-trivial bounds on the list-size L for codes ε -close to capacity which do not require q to be large, other than what follows from the Zyablov–Pinsker argument [40] (given formally in [11]). Recall that this method guarantees list size $L = q^{O(\ell/\varepsilon)}$. In comparison, we obtain (roughly) $L = \frac{1}{\varepsilon} \cdot q^{O(\ell \log q)}$ in the case of erasures (over prime fields), and $L = \frac{1}{\varepsilon} \cdot q^{\ell \cdot \text{polylog}(q)}$ in the case of errors (over all fields, assuming ρ is not too close to 0 or $1 - \ell/q$). Thus, compared to [40], we obtain a smaller bound on L once, roughly, $\varepsilon < \frac{1}{\text{polylog}(q)}$. In particular, when ℓ and q are constants we achieve asymptotic improvement in L .

1.2 Technical Overview

At a conceptual level, our work reconsiders the approach of Guruswami, Håstad, and Kopparty [12], which allowed for an understanding of the list-decodability from errors of random linear codes over constant-sized alphabets, and adapts it to the case of list-recovery (either from erasures or errors). Recall that list-decoding from errors is the special case of list-recovery from errors with $\ell = 1$.

² Specifically, when ρ is bounded away from both 0 and $1 - \ell/q$, we can take $C(\rho, \ell, q) \leq q^{\ell \cdot \text{polylog}(q)}$. See Theorem 31 for the precise bound.

Using terminology that we define in this work, the first step in the argument of [12] is to argue that Hamming balls are nontrivially *mixing*. Specifically, fix any center $z \in \mathbb{F}_q^n$ and consider sampling twice uniformly and independently from the Hamming ball of radius ρ centered at z ; denote the two samples by X and X' . The authors argue that for some $\delta > 0$ and any $\alpha, \beta \in \mathbb{F}_q \setminus \{0\}$, it follows that for any other center $y \in \mathbb{F}_q^n$, $\Pr[\alpha X + \beta X' \in B_\rho(y)] \leq q^{-\delta n}$ for some $\delta = \delta(\rho, q) > 0$.³ Alternatively, we could say that for any $y \in \mathbb{F}_q^n$, we have $\Pr[\alpha X + \beta X' \in y + B_\rho(z)] \leq q^{-\delta n}$. From here, using some additional tools like *2-increasing chains* – whose existence they establish via a Ramsey-theoretic argument – they are able to show that random linear codes with ε gap-to-capacity are with high probability $(\rho, \frac{C(\rho, q)}{\varepsilon})$ -list-decodable from errors. In particular, their techniques promise the correct dependence on ε (although the dependence on the parameters q and ρ is quite poor).

We recommend viewing this “ δ -mixing” property in the following light. Any argument establishing that random linear codes have good list-decodability must somehow argue that random subspaces and Hamming balls don’t tend to “correlate” too much. In particular, it should not be the case that Hamming balls have noticeable “linear structure”, and in particular they should be “far” from being closed under addition.

In our work, we consider whether or not sets that are relevant for list-recovery, i.e., the sorts of sets that list-recoverable codes cannot intersect with too much, also have nontrivial mixing. Firstly, we crystallize in a definition what it means for any arbitrary set $T \subseteq \mathbb{F}_q^n$ to be δ -mixing, where $\delta > 0$: for any $\alpha, \beta \in \mathbb{F}_q \setminus \{0\}$ and $y \in \mathbb{F}_q^n$, if $X, X' \sim T$ – which denotes that X and X' are sampled independently and uniformly from T – then $\Pr[\alpha X + \beta X' \in y + T] \leq q^{-\delta n}$.

Now, for (α, ℓ, L) -list-recovery from erasures the relevant sets are combinatorial rectangles $T_1 \times T_2 \times \cdots \times T_n$ where for at least $(1 - \alpha)n$ values of $i \in [n]$ we have $|T_i| \leq \ell$. For (ρ, ℓ, L) -list-recovery from errors the sets are “puffed-up” combinatorial rectangles. Namely, for $T_1, T_2, \dots, T_n \subseteq \mathbb{F}_q$ each with $|T_i| \leq \ell$, we consider *list-recovery balls*

$$B_\rho(T_1 \times T_2 \times \cdots \times T_n) = \{x \in \mathbb{F}_q^n : x_i \notin T_i \text{ for at most } \rho n \text{ choices of } i \in [n]\}.$$

Following the argument of [12], once we establish that these sets are nontrivially mixing, we can obtain bounds on the list-size with the correct dependence on ε . Our task then boils down to understanding the mixingness of the sets relevant for list-recovery. We consider first the erasures case, and subsequently discuss the errors case.

List-recovery from erasures

Firstly, observe that for a combinatorial rectangle $T_1 \times \cdots \times T_n$, if each of the sets T_i is nontrivially mixing as a subset of \mathbb{F}_q (take the $n = 1$ case of the above definition), then $T_1 \times \cdots \times T_n$ is also nontrivially mixing (as a subset of \mathbb{F}_q^n). Hence, it suffices for us to consider whether or not subsets of \mathbb{F}_q mix. It is here that the dependence on the field size shows up.

Recall from the earlier discussion that [17] established that random linear codes over \mathbb{F}_{ℓ^t} (where ℓ is a prime power) are with high probability *not* (α, ℓ, L) -list-recoverable from erasures unless $L \geq \ell^{\Omega(1/\varepsilon)}$. Indeed, it is easy to find a subset $T \subseteq \mathbb{F}_{\ell^t}$ which is *not* δ -mixing for any $\delta > 0$: take $T = \mathbb{F}_\ell$ (or, more generally, any multiplicative coset $\gamma \cdot \mathbb{F}_\ell$ for $\gamma \in \mathbb{F}_{\ell^t} \setminus \{0\}$).

³ In fact for [12] it sufficed to only consider the $z = 0$ case, in which case one can always assume $\alpha = \beta = 1$. But their argument naturally generalizes to this case, and this is the notion of mixing that we require for our list-recovery results.

Since \mathbb{F}_ℓ is closed under addition, in particular we have that if X and X' are sampled independently and uniformly from T then $\Pr[X + X' \in T] = 1$, so T is *not* δ -mixing for any $\delta > 0$.

What went wrong in this example? The fact that \mathbb{F}_{ℓ^t} contains \mathbb{F}_ℓ as a subfield means that \mathbb{F}_{ℓ^t} contains non-trivial \mathbb{F}_ℓ -linear subspaces. Such subspaces naturally create “bad” input lists, and the argument of [17] establishes that indeed a random linear code is likely to contain many vectors from a combinatorial rectangle $T_1 \times \cdots \times T_n$ where at least a $(1 - \alpha)$ fraction of the T_i ’s are 1-dimensional \mathbb{F}_ℓ -subspaces of \mathbb{F}_{ℓ^t} .

If we insist that q be a *prime*, then \mathbb{F}_q *does not* have any non-trivial subspaces. However, in this case \mathbb{F}_q still contains some subsets with “additive structure;” for example, taking ℓ to be odd here for simplicity, centered intervals⁴ like $I = \{-\frac{\ell-1}{2}, -\frac{\ell-1}{2}+1, \dots, \frac{\ell-1}{2}-1, \frac{\ell-1}{2}\} \subseteq \mathbb{F}_q$ have the property that if one samples $X, X' \sim I$ independently, then $\Pr[X + X' \in I] \approx \frac{3}{4} + \frac{1}{4\ell^2}$ assuming $\ell \leq 2q/3$. But note that this is still non-trivially bounded away from 1! Remarkably, an argument of Lev [29] shows that this is the worst case over prime fields. More precisely, over all sets T of size ℓ , to maximize $\Pr[\alpha X + \beta X' \in \gamma + T]$ where $\alpha, \beta, \gamma \in \mathbb{F}_q$ with $\alpha, \beta \neq 0$, one should choose $T = I$ (the centered interval of length ℓ), $\alpha = \beta = 1$ and $\gamma = 0$. In our technical section we give an effective bound on $\Pr[X + X' \in I]$ for all $\ell < q$, which allows us to argue that combinatorial rectangles $T_1 \times \cdots \times T_n$ in which at least a $1 - \alpha$ fraction of the T_i ’s have size at most ℓ are nontrivially mixing, as desired.

List-recovery from errors

We now wish to establish that list-recovery balls are nontrivially mixing. Notably, unlike in the case of erasures, the argument here is *insensitive* to the base field. Let $T = T_1 \times \cdots \times T_n$, where each $|T_i| \leq \ell$. Let $X, X' \sim B_\rho(T)$; in this overview, we will sketch how one bounds $\Pr[X + X' \in B_\rho(T)]$ (the argument easily generalizes to allow for multipliers $\alpha, \beta \in \mathbb{F}_q \setminus \{0\}$ and a shift $y \in \mathbb{F}_q^n$).

Unlike in the case of combinatorial rectangles, it is *not* the case that $X = (X_1, \dots, X_n)$ and $X' = (X'_1, \dots, X'_n)$ have independent coordinates. For example, conditioned on X_1 lying in T_1 , then X_2 is less likely to lie in T_2 . However, these correlations are relatively minor, and we can essentially “pretend” that both X and X' are sampled as follows: for each $i \in [n]$, with probability $1 - \rho$ set the i -th coordinate to a uniformly random element of T_i , and otherwise set it to a uniformly random element of $\mathbb{F}_q \setminus T_i$, and these choices are made independently for each $i \in [n]$.⁵ We remark that a similar trick is implicit in [12], and made explicit in the context of rank-metric codes by Guruswami and Resch [18].

This new distribution is much more amenable to analysis. In particular, letting E_i be the indicator for the event $X_i + X'_i \in T_i$, then $X + X' \in B_\rho(T)$ iff $\sum_i E_i \geq (1 - \rho)n$. Thus, if we can argue $\Pr[E_i = 1] < 1 - \rho$ then a classic Chernoff-Hoeffding bound establishes $\Pr[\sum_i E_i \geq (1 - \rho)n]$ is exponentially small, implying the desired δ -mixing.

Bounding $\Pr[E_i = 1]$ is the most novel part of the analysis, and is done in Lemma 26. Recall that $X_i, X'_i \sim T_i$, and let also $Y_i, Y'_i \sim \mathbb{F}_q \setminus T_i$. We have

$$\begin{aligned} \Pr[E_i = 1] &= (1 - \rho)^2 \Pr[X_i + X'_i \in T_i] + 2\rho(1 - \rho) \Pr[X_i + Y'_i \in T_i] + \rho^2 \Pr[Y_i + Y'_i \in T_i] \\ &= \sum_{z_i \in T_i} (1 - \rho)^2 \Pr[X_i + X'_i = z_i] + 2\rho(1 - \rho) \Pr[X_i + Y'_i = z_i] + \rho^2 \Pr[Y_i + Y'_i = z_i]. \end{aligned} \tag{2}$$

⁴ Or, more generally, arithmetic progressions.

⁵ In fact for technical reasons we have to consider X_i lying in T_i with probability ω for $\omega \leq \rho$, and similarly X'_i lies in T_i with probability $\omega' \leq \rho$. But by a concentration argument one can easily establish that ω and ω' are with high probability very close to ρ .

As is standard, $\Pr[X_i + X'_i = z_i]$ is proportional to $1_{T_i} * 1_{T_i}(z_i)$, the convolution of the indicator functions for T_i , and similarly $\Pr[X_i + Y'_i = z_i]$ and $\Pr[Y_i + Y'_i = z_i]$ are proportional to $1_{T_i} * 1_{\mathbb{F}_q \setminus T_i}(z_i)$ and $1_{\mathbb{F}_q \setminus T_i} * 1_{\mathbb{F}_q \setminus T_i}(z_i)$, respectively. Using the simple identity $1_{\mathbb{F}_q \setminus T_i} = 1 - 1_{T_i}$, we can rewrite Equation (2) as

$$c \cdot 1_{T_i} * 1_{T_i}(z_i) + \text{other terms},$$

where c is a constant which we show to be positive assuming $\rho \in (0, 1 - \ell/q)$ (and indeed, if $\rho > 1 - \ell/q$ then it could be negative). Upon giving a trivial upper bound $1_{T_i} * 1_{T_i}(z_i)$ (which corresponds to the case that T_i does not mix at all, as we must do since we are making no assumptions on the field) and simplifying, we obtain the bound

$$\Pr[E_i = 1] \leq (1 - \rho)^2 + \rho^2 \cdot \frac{\ell}{q - \ell}.$$

To our satisfaction, we have that $(1 - \rho)^2 + \rho^2 \cdot \frac{\ell}{q - \ell} < 1 - \rho$ iff $\rho \in (0, 1 - \ell/q)$, which is precisely the range of decoding radius at which we can hope for positive rate list-recovery from errors in the first place! Thus, we have that $\Pr[\sum_i E_i \geq (1 - \rho)n]$ is exponentially small, establishing that list-recovery balls nontrivially mix.

1.3 Open Problems

Lastly, we leave here some directions for future research:

- In our results the dependency on ε is correct, but the dependency on ℓ and q is rather poor. Can we improve this dependency? Or, can we perhaps prove new lower bounds on L in terms of ℓ and q that apply when these parameters are not too big?
- [17] showed that over \mathbb{F}_q with $q = \ell^t$ (and hence small characteristic), a random linear code is with high probability *not* (α, ℓ, L) -list-recoverable from erasures unless $L \geq \ell^{\Omega(1/\varepsilon)}$. Can we show that this lower bound actually applies to *every* q -ary linear code?
- Many arguments for codes being list-recoverable from errors in fact establish the stronger property of *average-radius*-list-recovery, where now one instead shows that for any input lists $T_1, \dots, T_n \subseteq \mathbb{F}_q$ of size ℓ , given $L + 1$ codewords $c^{(1)}, \dots, c^{(L+1)}$ one has

$$\frac{1}{L + 1} \sum_{j=1}^{L+1} \sum_{i=1}^n 1_{\{c_i^{(j)} \notin T_i\}} > \rho n.$$

This in particular implies that there cannot be $L + 1$ codewords lying in a list-recovery ball of radius ρ . We believe our method should be able to establish this (slightly) stronger guarantee for random linear codes.

2 Preliminaries

2.1 Notation

We will often denote random variables and sets by uppercase Roman letters. The distinction will be clear from context. We write $[n] = \{1, \dots, n\}$ for any positive integer n . For a vector $v \in \mathbb{F}_q^n$ with \mathbb{F}_q the finite field of order q , we write $\text{Supp}(v) = \{i : v_i \neq 0\}$. We define the *weight* of v to be $\text{wt}(v) = |\text{Supp}(v)|$, and the *Hamming distance* between v and u is $d(u, v) = |\{i \in [n] : u_i \neq v_i\}| = |\text{Supp}(u - v)|$. For a collection of vectors $v_1, \dots, v_d \in \mathbb{F}_q^n$, we denote by $\text{Span}(v_1, \dots, v_d)$ the subspace of \mathbb{F}_q^n spanned by v_1, \dots, v_d .

We denote the *binary entropy function* by $h_2(\cdot)$, and recall that $h_2(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$. For a positive integer q we shorthand $\exp_q(x) = q^x$. Additionally, by default \log is the base-2 logarithm. We write 1_A for the indicator function of a set A , and write $1_{\{E\}}$ for the indicator random variable that equals 1 if and only if the event E holds.

2.2 The Random Code Model

For an alphabet size q , a *plain random code* \mathcal{C} of block length n and rate $R \in [0, 1]$ is obtained by including each $x \in [q]^n$ in \mathcal{C} with probability $q^{-(1-R)n}$, and these choices are made independently for each x . (Note then that such a code has size q^{Rn} in expectation, and by a Chernoff bound it follows that it has rate $R \pm o(1)$ with high probability.)

When q is a prime power, a *random linear code* \mathcal{C} of block length n and rate $R \in [0, 1]$ is obtained by sampling a uniformly random matrix $H \in \mathbb{F}_q^{(1-R)n \times n}$, and defining

$$\mathcal{C} = \{x \in \mathbb{F}_q^n : Hx = 0\}.$$

Note that H is full-rank with probability $1 - O(q^{-Rn})$, and therefore \mathcal{C} has rate R with high probability.

Given any subset $S \subseteq [q]^n$, if \mathcal{C} is a plain random code of rate R then $\Pr[S \subseteq \mathcal{C}] = q^{-(1-R)n|S|}$. When dealing with random linear codes, the probability that a set appears in the code is determined by the span of the set.

► **Proposition 7.** *Let $\mathcal{C} \leq \mathbb{F}_q^n$ be a random linear code of block length n and rate $R \in [0, 1]$, and let $x_1, \dots, x_b \in \mathbb{F}_q^n$. Then,*

$$\Pr[\forall i \in [b], x_i \in \mathcal{C}] = q^{-(1-R)n \dim(\text{Span}(x_1, \dots, x_b))}.$$

2.3 List-Recovery Notions

This section collects the basic notions of list-recovery we study.

List-recovery from erasures

We begin with the relevant definition.

► **Definition 8** (list recovery from erasures). *Let $\mathcal{C} \subseteq [q]^n$ be a q -ary code of block-length n . For an erasure radius $\alpha \in [0, 1)$ and input list size $1 \leq \ell \leq q$, we say that \mathcal{C} is (α, ℓ, L) -list-recoverable from erasures if for every $T_1, \dots, T_n \subseteq [q]$ such that $|T_i| \leq \ell$ for at least $(1 - \alpha)n$ of the i -s and $T_i = [q]$ for the remaining, it holds that*

$$|\mathcal{C} \cap (T_1 \times \dots \times T_n)| \leq L.$$

That is, in any combinatorial rectangle of which at least $(1 - \alpha)n$ of its side-lengths are at most ℓ (and the remainder can be as large as q), there are at most L codewords.

We will also consider list-recovery from errors. The concept of a list-recovery ball – which generalizes that of a Hamming ball – will be useful.

► **Definition 9** (list-recovery ball). *Let $q \in \mathbb{N}$ and let $T_1, \dots, T_n \subseteq [q]$. The list-recovery ball of radius ρ centered at $T_1 \times \dots \times T_n$ is*

$$B_\rho(T_1 \times \dots \times T_n) = \{x \in [q]^n : d(x, T_1 \times \dots \times T_n) \leq \rho n\}.$$

Above, we have extended the Hamming metric by setting

$$d(x, T_1 \times \dots \times T_n) = |\{i \in [n] : x_i \notin T_i\}|.$$

57:10 List-Recovery of Random Linear Codes over Small Fields

We now state the relevant capacity theorem for list-recovery from erasures. The proofs of the two implications are standard: the *possibility* result follows from analyzing the performance of a plain random code, while the *impossibility* result follows from a counting argument.

► **Theorem 10** (list-recovery from erasures capacity). *Let $1 \leq \ell \leq q$ and let $\alpha \in [0, 1)$. Fix $\varepsilon > 0$. For $n \in \mathbb{N}$ large enough, the following hold:*

- *There exists a code $\mathcal{C} \subseteq [q]^n$ of rate $1 - (1 - \alpha) \log_q \ell - \alpha - \varepsilon$ which is $(\alpha, \ell, \lceil \ell/\varepsilon \rceil)$ -list-recoverable from erasures.*
- *For any code $\mathcal{C} \subseteq [q]^n$ of rate $1 - (1 - \alpha) \log_q \ell - \alpha + \varepsilon$, there exist $T_1, \dots, T_n \subseteq [q]$ with $|T_i| \leq \ell$ for at least $(1 - \alpha)n$ values of $i \in [n]$ such that $|\mathcal{C} \cap (T_1 \times \dots \times T_n)| \geq q^{\varepsilon n - o(n)}$.*

We therefore say that the *capacity* for (α, ℓ, L) -list-recovery from erasures is $1 - \alpha - (1 - \alpha) \log_q \ell$. We will study what happens for codes of rate $1 - (1 - \alpha) \log_q \ell - \alpha - \varepsilon$ for some $\varepsilon > 0$, and determine the value of their output list-size L . We will be focused on the case where q is held constant, and the gap-to-capacity ε tends to 0.

List-recovery from errors

We now define what it means for a code to be list-recoverable from errors.

► **Definition 11** (list recovery from errors). *Let $\mathcal{C} \subseteq [q]^n$ be a q -ary code of block-length n . For a decoding radius $\rho \in (0, 1 - \ell/q)$ and input list size $1 \leq \ell \leq q$, we say that \mathcal{C} is (ρ, ℓ, L) -list-recoverable from errors if for every $T_1, \dots, T_n \subseteq [q]$ such that $|T_i| \leq \ell$ for all $i \in [n]$, it holds that*

$$|\mathcal{C} \cap B_\rho(T_1 \times \dots \times T_n)| \leq L.$$

That is, every list-recovery ball of radius ρ with side-lengths $\leq \ell$ contains at most L codewords.

We will need an estimate on the size of list-recovery balls. It makes use of the (q, ℓ) -entropy function, defined as follows:

$$h_{q,\ell}(x) = x \log_q \frac{q - \ell}{x} + (1 - x) \log_q \frac{q - \ell}{1 - x}. \quad (3)$$

An operational interpretation of this quantity is as the base- q entropy of a random variable which, with probability $1 - x$, samples a uniformly random element from a set of size ℓ , and with probability x samples a uniformly random element from the complement. Additionally, so long as $0 < x < 1 - \ell/q$ it holds that $0 < h_{q,\ell}(x) < 1$. Note that if $\ell = 1$ one recovers the q -entropy function, which we denote as h_q (i.e., if ℓ is omitted from the subscript, then it is by default 1).

We now state the relevant estimate.

► **Proposition 12** ([35, Proposition 2.4.11]). *Let $1 \leq \ell \leq q$ be integers and $\rho \in (0, 1 - \ell/q)$. Let $T_1, \dots, T_n \subseteq [q]$ with $|T_i| = \ell$ for all $i \in [n]$. Then,*

$$\frac{q^{nh_{q,\ell}(\rho)}}{\sqrt{2n}} \leq |B_\rho(T_1 \times \dots \times T_n)| \leq q^{nh_{q,\ell}(\rho)}.$$

This estimate drives the following capacity theorem.

► **Theorem 13** (list-recovery from errors capacity). *Let $1 \leq \ell \leq q$ and let $\rho \in (0, 1 - \ell/q)$. Fix $\varepsilon > 0$. For $n \in \mathbb{N}$ large enough, the following hold:*

- *There exists a code $\mathcal{C} \subseteq [q]^n$ of rate $1 - h_{q,\ell}(\rho) - \varepsilon$ which is $(\rho, \ell, \lceil \ell/\varepsilon \rceil)$ -list-recoverable from errors.*
- *For any code $\mathcal{C} \subseteq [q]^n$ of rate $1 - h_{q,\ell}(\rho) + \varepsilon$, there exists $T_1, \dots, T_n \subseteq [q]$ with $|T_i| \leq \ell$ for all $i \in [n]$ such that $|\mathcal{C} \cap B_\rho(T_1 \times \dots \times T_n)| \geq q^{\varepsilon n - o(n)}$.*

Thus, we will concern ourselves with codes of rate $1 - h_{q,\ell}(\rho) - \varepsilon$, and determine the output list size L for (ρ, ℓ, L) -list-recovery from errors. And, as in the list-recovery from erasures case, we will hold ρ, ℓ, q constants and consider the asymptotic behaviour of L as the gap-to-capacity $\varepsilon \rightarrow 0$.

We will also need the following lower bound on the difference $h_{q,\ell}(\rho) - h_{q,\ell}(\rho - \eta)$. See the full version of the paper [3] for the proof.

▷ **Claim 14.** For any integers $q > 0$ and $1 \leq \ell \leq q$, any $\rho \in (0, 1 - \ell/q]$, and any $\eta \in [0, \rho]$, we have

$$h_{q,\ell}(\rho) - h_{q,\ell}(\rho - \eta) \geq \eta \log_q \left(\frac{(q - \ell)(1 - \rho)}{\ell \cdot \rho} \right) \geq 0.$$

2.4 Increasing Chains

The following definition of an increasing chain was first introduced by Guruswami, Håstad, and Kopparty [12].

► **Definition 15** (c -increasing chain). *A sequence of vectors $v_1, \dots, v_d \in \mathbb{F}_q^\ell$ is said to be a c -increasing chain of length d if for all $j \in [d]$ we have*

$$\left| \text{Supp}(v_j) \setminus \left(\bigcup_{i=1}^{j-1} \text{Supp}(v_i) \right) \right| \geq c.$$

We require the following lemma on the existence of appropriately long increasing chains in an appropriate shift of an arbitrary subset $S \subseteq \mathbb{F}_q^\ell$.

► **Lemma 16** ([12, Lemma 6.3]). *For every prime power q , and all positive integers c, ℓ and $L \leq q^\ell$, the following holds. For every $S \subseteq \mathbb{F}_q^\ell$ with $|S| = L$, there is $w \in \mathbb{F}_q^\ell$ such that $S + w$ has a c -increasing chain of length at least $\frac{1}{c} \log_q \frac{L}{2} - (1 - \frac{1}{c}) \log_q((q - 1)\ell)$.*

2.5 Mixing Sets

In our analysis we need to understand the probability that the sum of two independent uniformly random samples X and X' from a set $T \subseteq \mathbb{F}_q$ lands in a shifted set $T + \gamma$, for an arbitrary shift $\gamma \in \mathbb{F}_q$ (and in fact a more general question of that form). We begin with the necessary definitions.

► **Definition 17** (mixing over \mathbb{F}_q). *For a prime power q , and $\delta \geq 0$, we say that $T \subseteq \mathbb{F}_q$ is δ -mixing, if for any $\alpha, \beta, \gamma \in \mathbb{F}_q$, where α and β are nonzero, it holds that*

$$\Pr_{X, X'}[\alpha X + \beta X' \in T + \gamma] \leq q^{-\delta},$$

where $X, X' \sim T$ are independent and uniformly distributed over T , and $T + \gamma = \{t + \gamma : t \in T\}$.

► **Definition 18** (mixing over \mathbb{F}_q^n). For $n \in \mathbb{N}$, a prime power q , and $\delta \geq 0$, we say that $T \subseteq \mathbb{F}_q^n$ is δ -mixing, if for any nonzero $\alpha, \beta \in \mathbb{F}_q$, and any $z \in \mathbb{F}_q^n$, it holds that

$$\Pr_{X, X'} [\alpha X + \beta X' \in T + z] \leq q^{-\delta n},$$

where $X, X' \sim T$ are independent and uniformly distributed over T .

► **Remark 19.** Note that a set mixes nontrivially when $\delta > 0$, and moreover, we will want $\delta > 0$ to not depend on n . However, in the case of list recovery from erasures, where for some i -s, $S_i = \mathbb{F}_q$, the case of $\delta = 0$ will be useful towards bounding the expected mixing of S_1, \dots, S_n .

The following connection then follows easily.

► **Remark 20.** Suppose that $T_1, \dots, T_n \subseteq \mathbb{F}_q$, and each T_i is δ_i -mixing. Then, $T = T_1 \times \dots \times T_n$ is δ -mixing, for $\delta = \mathbb{E}_{i \sim [n]} [\delta_i]$.

3 List-Recovery from Erasures over Prime Fields

In this section we establish the list-recoverability of random linear codes over prime fields. To achieve this, we must first understand the mixing properties of worst-case subsets of prime fields. Most of the proofs are deferred to the full version [3].

3.1 Worst-Case Mixing of Subsets of Prime Fields

Towards understanding mixing of subsets of prime fields, we leverage a general result of Lev [29] which characterizes worst-case T -s, when q is prime. Before we introduce it, we set up some relevant notation.

For a set $T \subseteq \mathbb{F}_q$ we let \tilde{T} denote a “centered interval” of length $|T|$. More precisely, \tilde{T} is the δ -centered interval associated with T if $\tilde{T} = [-\alpha, \alpha + \delta] \subseteq \mathbb{F}_q$ with $\alpha \in [0, \frac{q-1}{2}]$ and $\delta \in \{-1, 0, 1\}$ satisfying $|\tilde{T}| = 2\alpha + 1 + \delta = |T|$. Note that when $|T|$ is odd there is a unique centered interval \tilde{T} (because $\delta = 0$ necessarily), but when $|T|$ is even there are two centered intervals, corresponding to $\delta = \pm 1$.

► **Lemma 21** ([29, Theorem 1], adapted). Let $q \geq 3$ be prime and $A_1, \dots, A_k \subseteq \mathbb{F}_q$ be arbitrary sets with $\tilde{A}_1, \dots, \tilde{A}_k$ the associated δ_i -centered intervals. Then, if $|\delta_1 + \dots + \delta_k| \leq 1$, for any set $B \subseteq \mathbb{F}_q$ and some associated δ -centered interval \tilde{B} , we have

$$\Pr_{X_i \sim A_i} [X_1 + \dots + X_k \in B] \leq \Pr_{\tilde{X}_i \sim \tilde{A}_i} [\tilde{X}_1 + \dots + \tilde{X}_k \in \tilde{B}].$$

We can use Lemma 21 to prove the following.

► **Lemma 22.** Fix a prime $q \geq 3$. Let $T_1, T_2, T_3 \subseteq \mathbb{F}_q$ be arbitrary sets of size $\ell > 0$. Then,

$$\Pr_{X_1 \sim T_1, X_2 \sim T_2} [X_1 + X_2 \in T_3] \leq \begin{cases} \frac{3}{4} + \frac{1}{4\ell^2} + \frac{\max(0, 3\ell - 2q - 1) \cdot (3\ell - 2q + 1)}{4\ell^2}, & \text{if } \ell \text{ is odd,} \\ \frac{3}{4} + \frac{\max(0, 3\ell - 2q)^2}{4\ell^2}, & \text{if } \ell \text{ is even,} \end{cases}$$

and this is tight for all ℓ . In particular:

1. When $\ell \leq 2q/3$ we have

$$\Pr_{X_1 \sim T_1, X_2 \sim T_2} [X_1 + X_2 \in T_3] \leq \begin{cases} \frac{3}{4} + \frac{1}{4\ell^2}, & \text{if } \ell \text{ is odd,} \\ \frac{3}{4}, & \text{if } \ell \text{ is even.} \end{cases}$$

2. When $2q/3 < \ell \leq q-1$ we have

$$\Pr_{X_1 \sim T_1, X_2 \sim T_2} [X_1 + X_2 \in T_3] \leq \frac{q^2 - 3\ell(q-\ell)}{\ell^2}.$$

We can then record the following corollary.

- **Corollary 23.** *For a prime $q \geq 3$, any set $T \subseteq \mathbb{F}_q$ of size $\ell \leq q-1$,*
- *If $2 \leq \ell \leq 2q/3$, T is δ -mixing for $\delta \geq \log_q(16/13)$.*
 - *Otherwise, T is δ -mixing for $\delta \geq \log_q\left(\frac{\ell^2}{q^2 - 3\ell(q-\ell)}\right)$.*

3.2 List-Recovery from Erasures over Prime Fields via Mixing

In this section we adapt the technique in [12], together with the worst-case mixing result from Corollary 23, to establish list recovery from erasures over prime fields.

► **Lemma 24.** *Let $T \subseteq \mathbb{F}_q^n$ be δ -mixing. For $b \in \mathbb{N}$ and any $a > 0$ satisfying $n \geq q^{\frac{8a}{\delta}}$, the following holds. Let $X^{(1)}, \dots, X^{(b)}$ be sampled independently and uniformly at random from T . Then, we have that*

$$\Pr \left[\left| \text{Span} \left(X^{(1)}, \dots, X^{(b)} \right) \cap T \right| > C \cdot b \right] \leq q^{-an},$$

where $C = C(q, \delta, a) = q^{\frac{8a}{\delta}}$. In particular, when $T = T_1 \times \dots \times T_n$, and each T_i is δ_i -mixing, we get the same result as above for $\delta = \mathbb{E}_i[\delta_i]$.

Proof. Let E denote the bad event that we want to bound, namely $|\text{Span}(X^{(1)}, \dots, X^{(b)}) \cap T| > A$ for $A = b \cdot q^{\frac{8a}{\delta}}$. Note that E implies that there exists some set $S \subseteq \mathbb{F}_q^b$, $|S| = A+1$, such that $X_v \triangleq \sum_{i \in [b]} v_i X^{(i)} \in T$ for all $v \in S$.⁶ Hence, it suffices to bound the probability that such a set S exists.

Fix some $S \subseteq \mathbb{F}_q^b$ of size $A+1$. Applying Lemma 16 with $c = 2$ (and note that we can assume that $A+1 \leq q^b$), we know there exists $w \in \mathbb{F}_q^b$ such that $S + w$ has a 2-increasing chain of length $d = \frac{1}{2} \log_q \frac{A+1}{2} - \frac{1}{2} \log_q((q-1)b)$. That is, we have $v^{(1)}, \dots, v^{(d)} \in S$ such that for all $j \in [d]$,

$$\left| \text{Supp}(v^{(j)} + w) \setminus \left(\bigcup_{i=1}^{j-1} \text{Supp}(v^{(i)} + w) \right) \right| \geq 2.$$

Now, we can bound

$$\begin{aligned} \Pr[\forall v \in S, X_v \in T] &\leq \Pr[\forall j \in [d], X_{v^{(j)}} \in T] \\ &= \Pr[\forall j \in [d], X_{v^{(j)}} + X_w \in T + X_w] \\ &= \Pr[\forall j \in [d], X_{v^{(j)}+w} \in T + X_w]. \end{aligned} \tag{4}$$

Next, we bound Equation (4) by

$$\sum_{y \in \mathbb{F}_q^n} \Pr[\forall j \in [d], X_{v^{(j)}+w} \in T + y]. \tag{5}$$

⁶ Notice that if the X_i -s are not linearly independent, this can only decrease the probability that the intersection is large, so we can concentrate on the case that distinct v -s give rise to distinct X_v -s.

Towards bounding each term in the sum, observe that the increasing chain property tells us that for each $j \in [d]$, we can write $X_{v^{(j)}+w} = Y_{\text{past}}^{(j)} + Y_{\text{new}}^{(j)}$, where $Y_{\text{past}}^{(j)}$ contains X_k -s that participated in $\{X_{v^{(i)}+w}\}_{i < j}$, whereas $Y_{\text{new}}^{(j)}$ contains two new X_k -s. Now,

$$\Pr[X_{v^{(j)}+w} \in T+y \mid \forall i \in [j-1], X_{v^{(i)}+w} \in T+y] = \mathbb{E}_{z \sim Y_{\text{past}}^{(j)}} \left[\mathbf{1}(z) \cdot \Pr[Y_{\text{new}}^{(j)} \in T+y_z] \right], \quad (6)$$

where $\mathbf{1}(z)$ is an indicator for whether past X_k -s landed in $T+y$, and y_z is a fixed string that depends on the fixing of $Y_{\text{past}}^{(j)}$. Assume for simplicity that $Y_{\text{new}}^{(j)} = \alpha X^{(1)} + \beta X^{(2)}$, where $\alpha, \beta \in \mathbb{F}_q$ are nonzero. Then, using the fact that T is δ -mixing, each summand of Equation (5) can now be bounded by

$$\prod_{j \in [d]} \Pr[X_{v^{(j)}+w} \in T+y \mid \forall i \in [j-1], X_{v^{(i)}+w} \in T+y] \leq q^{-n\delta d},$$

and summing over all y -s gives us

$$\Pr[v \in S, X_v \in T] \leq q^n \cdot q^{-n\delta d} = q^{-(\delta d-1)n}.$$

Union-bounding over all S -s, we get

$$\Pr[E] \leq \binom{q^b}{A+1} q^{-(\delta d-1)n} \leq q^{b(A+1)-(\delta d-1)n}. \quad (7)$$

First, note that we set parameters so that $d \geq \frac{2a+1}{\delta}$. Indeed, we can set $d = \left\lceil \frac{1}{2} \log_q \left(\frac{A+1}{2b(q-1)} \right) \right\rceil$, and then need A to be at most, say, $4q \cdot q^{\frac{4a}{\delta}} \leq q^{\frac{8a}{\delta}}$. Under this choice of A , it also holds that $b(A+1) \leq \frac{\delta d-1}{2}n$, since n is large enough. Overall, Equation (7) gives $\Pr[E] \leq q^{-\frac{\delta d-1}{2}n} \leq q^{-an}$, as desired. The “In particular” part simply follows from Remark 20. \blacktriangleleft

We are now ready to give our list recovery result.

► **Theorem 25** (list recovery with erasures). *For any $n \in \mathbb{N}$, a prime q , an integer $\ell \leq q-1$, and $\alpha, \varepsilon \in (0, 1)$, the following holds. With probability at least $1 - q^{-n}$, a random linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate*

$$R = 1 - \alpha - (1 - \alpha) \log_q \ell - \varepsilon$$

is (α, ℓ, L) -list-recoverable from erasures, with

$$L = C_{q, \ell, \alpha} \cdot \frac{1}{\varepsilon},$$

provided that $n \geq L$. In particular, there exists a universal constant C such that:

- *When $\ell \leq \frac{2}{3}q$, we can take $C_{q, \ell, \alpha} \leq q^{C \log q \cdot ((1-\alpha)\ell+1)} \triangleq C_{q, \ell, \alpha}^{(0)}$, and,*
- *When $\ell = (1-\gamma)q$ for some $\gamma \in (0, 1/3)$, we can take $C_{q, \ell, \alpha} \leq \left(C_{q, \ell, \alpha}^{(0)} \right)^{\frac{(1-\gamma)^2}{1-3\gamma(1-\gamma)}}$.*

4 List-Recovery from Errors over Arbitrary Fields

We now turn to the case of list-recovery from errors. Unlike in the case of list-recovery from erasures, we will make no assumptions on the underlying field. We firstly show that list-recovery balls are non-trivially mixing. We subsequently sketch how, again using Lemma 22, we can conclude the desired list-recovery result. The proofs are deferred to the full version [3].

4.1 Mixture of List-Recovery Balls

► **Lemma 26.** *Suppose q is a prime power, $1 \leq \ell \leq q$ is an integer and $\omega_1, \omega_2 \in (0, 1 - \ell/q)$. Let $A, B, T \subseteq \mathbb{F}_q$ be of size ℓ , let $X_1 \sim \text{Unif}(A)$, $X_2 \sim \text{Unif}(B)$, $Y_1 \sim \text{Unif}(\mathbb{F}_q \setminus A)$ and $Y_2 \sim \text{Unif}(\mathbb{F}_q \setminus B)$ all independent. Then,*

$$(1 - \omega_1)(1 - \omega_2) \Pr[X_1 + X_2 \in T] + \omega_1(1 - \omega_2) \Pr[X_1 + Y_2 \in T] \quad (8)$$

$$+ \omega_2(1 - \omega_1) \Pr[Y_1 + X_2 \in T] + \omega_1\omega_2 \Pr[Y_1 + Y_2 \in T]$$

$$\leq (1 - \omega_1)(1 - \omega_2) + \omega_1\omega_2 \cdot \frac{\ell}{q - \ell}. \quad (9)$$

In the sequel, we will need upper and lower bounds on the RHS of Lemma 26 to $1 - \rho$. The following lemma establishes the required bounds.

► **Lemma 27.** *Suppose q is a prime power, $1 \leq \ell \leq q$ is an integer and $\rho \in (0, 1 - \ell/q)$. Suppose $0 \leq \eta \leq \rho$ and $\rho - \eta \leq \omega_1, \omega_2 \leq \rho$. Then, the following inequalities hold:*

$$1 - \rho > (1 - \rho)^2 + \rho^2 \cdot \frac{\ell}{q - \ell}, \quad (10)$$

$$(1 - \rho + \eta)^2 + (\rho - \eta)^2 \cdot \frac{\ell}{q - \ell} > (1 - \omega_1)(1 - \omega_2) + \omega_1\omega_2 \cdot \frac{\ell}{q - \ell}. \quad (11)$$

We can now state a lemma bounding the probability that a nontrivial linear combination of two uniform samples from a list-recovery ball lands in a shift of the list-recovery ball. Then, in Corollary 30, we show how to set parameters to turn this into a statement about the δ -mixing of a list-recovery ball.

► **Lemma 28.** *Let $n \in \mathbb{N}$, q a prime power, $1 \leq \ell \leq q$ an integer, and let $\rho \in (0, 1 - \ell/q)$. Let $T_1, \dots, T_n \subseteq \mathbb{F}_q$ be subsets, each of size ℓ . Fix $\eta > 0$ small enough so that*

$$1 - \rho > (1 - \rho + \eta)^2 + (\rho - \eta)^2 \cdot \frac{\ell}{q - \ell}.$$

Let $\alpha, \beta \in \mathbb{F}_q \setminus \{0\}$ and $y \in \mathbb{F}_q^n$, and let $X, X' \sim B_\rho(T_1 \times \dots \times T_n)$. Then,

$$\Pr[\alpha X + \beta X' \in y + B_\rho(T_1 \times \dots \times T_n)] \leq 2\sqrt{2n} \cdot \exp_q(n(h_{q,\ell}(\rho - \eta) - h_{q,\ell}(\rho))) \\ + 2n \cdot \exp_q\left(-nD_q\left(1 - \rho\|(1 - \rho + \eta)^2 + (\rho - \eta)^2 \frac{\ell}{q - \ell}\right)\right). \quad (12)$$

► **Remark 29.** Since Lemma 27 implies

$$1 - \rho > (1 - \rho)^2 + \rho^2 \cdot \frac{\ell}{q - \ell} \iff 0 < \rho < 1 - \frac{\ell}{q},$$

it follows that one can indeed choose η small enough to ensure

$$1 - \rho > (1 - \rho - \eta)^2 + (\rho + \eta)^2 \cdot \frac{\ell}{q - \ell}.$$

In Corollary 30 we show how to choose $\eta > 0$ to bound the two terms in Equation (12) by $q^{-\delta n}$ for a concrete $\delta > 0$ (which depends on ρ, ℓ, q).

► **Corollary 30.** *Let $n \in \mathbb{N}$, q a prime power, $1 \leq \ell < q$ an integer, and let $\rho \in (0, 1 - \ell/q)$. Let $T_1, \dots, T_n \subseteq \mathbb{F}_q$, each of size ℓ . The list-recovery ball $B_\rho(T_1 \times \dots \times T_n)$ is δ -mixing, for*

$$\delta = \log_q\left(\frac{(q - \ell)(1 - \rho)}{\rho\ell}\right) \cdot \frac{\rho^4(1 - \ell/q - \rho)^2}{16 \log q}$$

assuming $n \geq \left(\frac{\log q}{\rho(1 - \ell/q - \rho)}\right)^c$ for some universal constant c .

4.2 List Recovery from Errors

Having established that the list-recovery ball is δ -mixing, we will repeat roughly the same argument we had for list recovery with erasures, that uses Lemma 24.

► **Theorem 31** (list recovery with errors). *For every $n \in \mathbb{N}$, a prime power q , an integer $\ell < q$, $\rho \in (0, 1 - \ell/q)$, and $\varepsilon \in (0, 1)$, the following holds. With probability at least $1 - q^{-n}$, a random linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate*

$$R = 1 - h_{q,\ell}(\rho) - \varepsilon$$

is (ρ, ℓ, L) -list-recoverable from errors, with

$$L = C_{\rho,\ell,q} \cdot \frac{1}{\varepsilon},$$

provided that n is large enough. More concretely, there exists a universal constant C such that

$$C_{\rho,\ell,q} \leq q^{\ell \cdot \left(\frac{\log q}{\rho(1-\ell/q-\rho)} \right)^C}$$

provided that $n \geq \left(\frac{\log q}{\rho(1-\ell/q-\rho)} \right)^C$.

References

- 1 Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Near-optimal erasure list-decodable codes. In *35th Computational Complexity Conference (CCC)*, pages 1:1–1:27. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.CCC.2020.1.
- 2 Yeyuan Chen and Zihan Zhang. Explicit folded Reed-Solomon and multiplicity codes achieve relaxed generalized Singleton bounds, 2024. doi:10.48550/arXiv.2408.15925.
- 3 Dean Doron, Jonathan Mosheiff, Nicolas Resch, and João Ribeiro. List-recovery of random linear codes over small fields, 2025. doi:10.48550/arXiv.2505.05935.
- 4 Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly optimal pseudorandomness from hardness. *Journal of the ACM*, 69(6), November 2022. doi:10.1145/3555307.
- 5 Dean Doron and Mary Wootters. High-probability list-recovery, and applications to heavy hitters. In *49th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 55:1–55:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.ICALP.2022.55.
- 6 Peter Elias. List decoding for noisy channels. Technical Report Technical Report 335, Research Laboratory of Electronics, MIT, 1957. URL: <https://dspace.mit.edu/handle/1721.1/4484>.
- 7 Anna C. Gilbert, Yi Li, Ely Porat, and Martin J. Strauss. For-all sparse recovery in near-optimal time. *ACM Transactions on Algorithms*, 13(3):1–26, 2017. doi:10.1145/3039872.
- 8 Eitan Goldberg, Chong Shangguan, and Itzhak Tamo. Singleton-type bounds for list-decoding and list-recovery, and related results. In *International Symposium on Information Theory (ISIT)*, pages 2565–2570. IEEE, 2022. doi:10.1109/ISIT50566.2022.9834849.
- 9 Sivakanth Gopi, Swastik Kopparty, Rafael Oliveira, Noga Ron-Zewi, and Shubhangi Saraf. Locally testable and locally correctable codes approaching the Gilbert-Varshamov bound. *IEEE Transactions on Information Theory*, 64(8):5813–5831, 2018. doi:10.1109/TIT.2018.2809788.
- 10 Zeyu Guo, Ray Li, Chong Shangguan, Itzhak Tamo, and Mary Wootters. Improved list-decodability and list-recoverability of Reed-Solomon codes via tree packings. In *62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 708–719. IEEE, 2021. doi:10.1109/FOCS52979.2021.00074.
- 11 Venkatesan Guruswami. List decoding from erasures: Bounds and code constructions. *IEEE Transactions on Information Theory*, 49(11):2826–2833, 2003. doi:10.1109/TIT.2003.815776.

- 12 Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. *IEEE Transactions on Information Theory*, 57(2):718–725, 2011. Preliminary version at STOC 2010. doi:10.1109/TIT.2010.2095170.
- 13 Venkatesan Guruswami and Piotr Indyk. Near-optimal linear-time codes for unique decoding and new list-decodable codes over smaller alphabets. In *34th Annual Symposium on Theory of Computing (STOC)*, pages 812–821. ACM, 2002. doi:10.1145/509907.510023.
- 14 Venkatesan Guruswami and Piotr Indyk. Linear time encodable and list decodable codes. In *35th Annual Symposium on Theory of Computing (STOC)*, pages 126–135. ACM, 2003. doi:10.1145/780542.780562.
- 15 Venkatesan Guruswami and Piotr Indyk. Efficiently decodable codes meeting Gilbert-Varshamov bound for low rates. In *15th Annual Symposium on Discrete Algorithms (SODA)*, pages 756–757. SIAM, 2004. URL: <http://dl.acm.org/citation.cfm?id=982792.982907>.
- 16 Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005. doi:10.1109/TIT.2005.855587.
- 17 Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear codes. *IEEE Transactions on Information Theory*, 68(2):923–939, 2022. doi:10.1109/TIT.2021.3127126.
- 18 Venkatesan Guruswami and Nicolas Resch. On the list-decodability of random linear rank-metric codes. In *International Symposium on Information Theory (ISIT)*, pages 1505–1509. IEEE, 2018. doi:10.1109/ISIT.2018.8437698.
- 19 Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *Journal of the ACM*, 56(4):20:1–20:34, 2009. doi:10.1145/1538902.1538904.
- 20 Iftach Haitner, Yuval Ishai, Eran Omri, and Ronen Shaltiel. Parallel hashing via list recoverability. In *Advances in Cryptology – CRYPTO 2015*, pages 173–190. Springer Berlin Heidelberg, 2015. doi:10.1007/978-3-662-48000-7_9.
- 21 Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes and applications. *SIAM Journal on Computing*, 49(4):FOCS17–157, January 2020. doi:10.1137/17M116149X.
- 22 Justin Holmgren, Alex Lombardi, and Ron D. Rothblum. Fiat–Shamir via list-recoverable codes (or: parallel repetition of GMW is not zero-knowledge). In *53rd Annual Symposium on Theory of Computing (STOC)*, pages 750–760. ACM, 2021. doi:10.1145/3406325.3451116.
- 23 Piotr Indyk, Hung Q. Ngo, and Atri Rudra. Efficiently decodable non-adaptive group testing. In *21st Annual Symposium on Discrete Algorithms (SODA)*, pages 1126–1142. SIAM, 2010. doi:10.1137/1.9781611973075.91.
- 24 Itay Kaley and Amnon Ta-Shma. Unbalanced expanders from multiplicity codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 12:1–12:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.APPROX/RANDOM.2022.12.
- 25 Sergey Komech and Jonathan Mosheiff. Let’s have both! Optimal list-recoverability via alphabet permutation codes, 2025. doi:10.48550/arXiv.2502.05858.
- 26 Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *Journal of the ACM*, 64(2):1–42, 2017. doi:10.1145/3051093.
- 27 Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved decoding of folded Reed-Solomon and multiplicity codes. In *59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 212–223. IEEE, 2018. doi:10.1109/FOCS.2018.00029.
- 28 Kasper Green Larsen, Jelani Nelson, Huy L. Nguyen, and Mikkel Thorup. Heavy hitters via cluster-preserving clustering. In *57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 61–70. IEEE, 2016. doi:10.1109/FOCS.2016.16.

- 29 Vsevolod F. Lev. Linear equations over \mathbb{F}_p and moments of exponential sums. *Duke Mathematical Journal*, 107(2):239–263, 2001. doi:10.1215/S0012-7094-01-10722-9.
- 30 Matan Levi, Jonathan Mosheiff, and Nikhil Shagrithaya. Random Reed-Solomon codes and random linear codes are locally equivalent, 2024. arXiv:2406.02238.
- 31 Ray Li and Nikhil Shagrithaya. Near-optimal list-recovery of linear code families, 2025. doi:10.48550/arXiv.2502.13877.
- 32 Ben Lund and Aditya Potukuchi. On the list recoverability of randomly punctured codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 30:1–30:11. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.APPROX/RANDOM.2020.30.
- 33 Jonathan Mosheiff, Nicolas Resch, Kuo Shang, and Chen Yuan. Randomness-efficient constructions of capacity-achieving list-decodable codes, 2024. doi:10.48550/arXiv.2402.11533.
- 34 Hung Q. Ngo, Ely Porat, and Atri Rudra. Efficiently decodable error-correcting list disjoint matrices and applications. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 557–568. Springer, 2011.
- 35 Nicolas Resch. *List-Decodable Codes: (Randomized) Constructions and Applications*. PhD thesis, Carnegie Mellon University, 2020. URL: <http://reports-archive.adm.cs.cmu.edu/anon/2020/abstracts/20-113.html>.
- 36 Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *46th Annual Symposium on Theory of Computing (STOC)*, pages 764–773. ACM, 2014. doi:10.1145/2591796.2591797.
- 37 Atri Rudra and Mary Wootters. Average-radius list-recoverability of random linear codes. In *29th Annual Symposium on Discrete Algorithms (SODA)*, pages 644–662. SIAM, 2018. doi:10.1137/1.9781611975031.42.
- 38 Amnon Ta-Shma and David Zuckerman. Extractor codes. *IEEE Transactions on Information Theory*, 50(12):3015–3025, 2004. doi:10.1109/TIT.2004.838377.
- 39 Itzhak Tamo. Tighter list-size bounds for list-decoding and recovery of folded Reed-Solomon and multiplicity codes. *IEEE Transactions on Information Theory*, 70(12):8659–8668, 2024. doi:10.1109/TIT.2024.3402171.
- 40 Victor Vasilievich Zyablov and Mark Semenovitch Pinsker. List concatenated decoding. *Problemy Peredachi Informatsii*, 17(4):29–33, 1981.