Pseudorandom Linear Codes Are List-Decodable to Capacity

Aaron (Louie) Putterman
Harvard University, Cambrdige, MA, USA

Edward Pyne
Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We introduce a novel family of expander-based error correcting codes. These codes can be sampled with randomness linear in the block-length, and achieve list decoding capacity (among other local properties). Our expander-based codes can be made starting from any family of sufficiently low-bias codes, and as a consequence, we give the first construction of a family of algebraic codes that can be sampled with linear randomness and achieve list-decoding capacity. We achieve this by introducing the notion of a *pseudorandom* puncturing of a code, where we select n indices of a base code $C \subset \mathbb{F}_q^m$ in a correlated fashion. Concretely, whereas a random linear code (i.e. a truly random puncturing of the Hadamard code) requires $O(n \log(m))$ random bits to sample, we sample a pseudorandom linear code with $O(n + \log(m))$ random bits by instantiating our pseudorandom puncturing as a length n random walk on an exapnder graph on $[m]$. In particular, we extend a result of Guruswami and Mosheiff (FOCS 2022) and show that a pseudorandom puncturing of a small-bias code satisfies the same local properties as a random linear code with high probability. As a further application of our techniques, we also show that pseudorandom puncturings of Reed-Solomon codes are list-recoverable beyond the Johnson bound, extending a result of Lund and Potukuchi (RANDOM 2020). We do this by instead analyzing properties of codes with large distance, and show that pseudorandom puncturings still work well in this regime.

2012 ACM Subject Classification Theory of computation → Error-correcting codes; Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Derandomization, error-correcting codes

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.90

Funding *Aaron (Louie) Putterman*: Supported in party by the Simons Investigator Fellowship of Boaz Barak, NSF grant DMS-2134157, DARPA grant W911NF2010021, and DOE grant DE-SC0022199. Supported in part by the Simons Investigator Award of Salil Vadhan and Simons Investigator Award of Madhu Sudan and NSF Award CCF 2152413.
Edward Pyne: Supported by an Akamai Presidential Fellowship.

Acknowledgements We would like to thank Madhu Sudan for helpful conversations.

1 Introduction

A central topic of interest in coding theory is that of list decodability. We seek an encoding function $E : \{0,1\}^k \to \{0,1\}^n$, such that for any possible received codeword r, there are only a few encoded messages that are close to r. More formally, we say that an encoding function is (ρ, L) *list-decodable* if for any possible received message $\hat{z} \in \{0,1\}^n$, there are at most L codewords in $\{0,1\}^k$ whose encodings are within hamming distance L of $\hat{z}$. This notion of list decodability provably allows for a notion of error-correction beyond the unique-decoding radius of $n/4$. As such, many years of research have focused on the construction of codes with better list-decoding properties. In this work, we suggest a new tool that can be used in such codes, by combining expander graphs with error-correcting codes in a novel manner.

Random linear codes (RLCs) are a fundamental tool in coding theory and specifically the construction of list-decodable codes because of their many favorable combinatorial properties. RLCs attain near-optimal distance and list decodability with high probability, while still maintaining efficient encoding. Indeed, a RLC is simply the image of a uniformly random generating matrix $E \in \{0,1\}^{n\times k}$, mapping messages of length k to codewords of length n as an $\mathbb{F}_2$-linear map. However, the main drawback of RLCs is that there is no known algorithm for efficient decoding or list decoding. This is often attributed to a lack of structure in the codes, resulting from the fact that each codeword $E(x)$ is chosen uniformly at random.

While there existed codes that achieve some favorable properties of RLCs using much less randomness (for instance the Toeplitz codes which achieve a near-optimal rate-distance tradeoff using $O(n)$ random bits), there did not exist $O(n)$-randomness constructions of codes achieving list-decoding capacity until the work of Guruswami and Moshieff [14]. To explicate their result, we recall that for a binary code $\mathcal{C} \subset \mathbb{F}_2^n$, the *bias* is defined as

$$\max_{c\in\mathcal{C}} \frac{|2|c| - n|}{n}.$$

Starting from the observation that an RLC is equivalent to a random puncturing of a Hadamard code, the work of [14] showed that taking a random puncturing of any code of sufficiently low-bias is "locally similar" to an RLC. Local similarity in this context refers to *local properties* of a code, which includes characterizations like list-decodability. Local properties are characterized by *not* containing certain sets of a small number of bad codewords.

Using this result, they were able to show that for every n, one can sample a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ generated with $O(n)$ bits of randomness with rate R that is list-decodable to distance $1 - R - \varepsilon$. This follows by using sufficiently low-bias codes of length $n' = O(n)$, and correspondingly choosing a random subset of size n of the indices. Choosing such a subset requires $\log \binom{n'}{n} = O(n)$ random bits. However, this construction has three potential drawbacks:

1. The "mother" codes of sufficiently low-bias can be quite complex constructions [1, 21].
2. The subsets of indices that are chosen are not structured, and thus unlikely to be amenable to efficient decoding. This is in contrast to the highly successful paradigm of decoding codes with graph-theoretic constructions [1, 6, 9, 12, 17, 20].
3. The random puncturing procedure can be implemented using $O(n)$ bits of randomness only when the mother code itself is of length $n' = O(n)$ (and hence is not randomness efficient for an arbitrary base code).

We aim to make progress on these drawbacks by introducing a more randomness efficient and structured way of performing puncturings and providing a new lens with which we can analyze the success events of puncturing procedures. We are particularly excited to introduce a new way to use expanders to create error-correcting codes given the rich history of decoding algorithms for these codes.

As an aside, we note that there have been several recent works studying random puncturings. In particular, a series of works [2, 10] has proved that random puncturings of Reed-Solomon codes achieve list-decoding capacity over linear-sized alphabets. The work we present here works for even constant size alphabets, provided the bias of the code is sufficiently small.

1.1 Our Contributions

We introduce the notion of a *pseudorandom* puncturing, and show codes generated in this fashion exhibit several desirable properties exhibited by truly random puncturings. In a truly random puncturing, we choose n indices i.i.d. uniformly from the $[m]$ indices of the mother code, and preserve these indices.

We begin by describing a special case of our main result:[1]

▶ **Theorem 1** (Optimal list-decoding with linear randomness). *For every $\rho \in (0, 1/2)$, $\varepsilon > 0$, there is an explicit family of binary linear codes of rate $R \geq 1 - h_2(\rho) - \varepsilon$ that can be sampled with $O(n/\varepsilon)$ random bits and are, with high probability, (ρ, L) list decodable for $L := O(1/\varepsilon)$.*

We now explain the framework for achieving this result. One way to view a truly random puncturing is to take a length n random walk over the complete graph on m vertices. Letting the sequence of visited vertices be $I = (i_1, \ldots, i_n)$, we define the puncturing by retaining the code indices in I. Performing such a random walk will then necessitate $n \cdot \log m$ random bits (exactly equivalent to choosing n indices in $[m]$ in the naive fashion).

However, this view of puncturing allows the application of derandomization techniques. We introduce an expander walk puncturing, by replacing the complete graph on m vertices with a sufficiently-expanding d-regular expander on m vertices. Now we perform a length n random walk on this expander, which only requires $O(\log m + n \cdot \log d)$ random bits, and (as before) let the punctured code be defined by the indices of the walk. Even when taking the mother code to be a Hadamard code of length $2^{O(n)}$, we can sample a puncturing with $O(n)$ random bits by having d a constant independent of n. Our primary contribution is showing such a walk is "close enough" to a truly random walk such that we can still conclude the punctured code has the desired properties.

Our results directly extend [14] (who in turn used the framework developed by [13,19]), in that we show even an *expander walk* puncturing of every sufficiently small bias code achieves list-decoding capacity.

▶ **Theorem 2** (More general case, informally). *Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear, sufficiently low-bias code. Let $\mathcal{C}$ be a λ-expander walk puncturing of $\mathcal{D}$ for sufficiently small λ. Then, $\mathcal{C}$ is likely to have every* monotone-decreasing, local property *that is typically satisfied by an RLC of similar rate. In fact, for every sufficiently low-bias mother code, we can pseudorandomly subsample the indices of our puncturing with randomness $O(nb)$, where b is the locality of the property, even when the block length m of $\mathcal{D}$ is exponential in n.*

As a consequence, we can now construct pseudorandom linear codes (with generator matrix G) list-decodable to capacity that are sampled with $O(n)$ random bits such that the rows of G come from an arbitrary low-bias mother code, and the columns of G are a pseudorandom subset.

In particular, as noted by [14], we can take the mother code $\mathcal{D}$ to be a dual-BCH code, where every codeword encodes a low-degree polynomial over $\mathbb{F}_{2^\ell}$ by the trace of its evaluations over $\mathbb{F}_{2^\ell}$. In the setting of [14], a random puncturing of $\mathcal{D}$ corresponds to codewords being evaluations over a *random* subset of $\mathbb{F}_{2^\ell}$. However, by instead taking an expander walk puncturing, the codewords are now evaluations over more constrained subsets of $\mathbb{F}_{2^\ell}$. In fact, we decrease the randomness required in this construction from $\Omega(n \log n)$ to $O(n)$, while still preserving its list-decodability. The work of [14] requires $\log(\binom{\text{poly}(n/\varepsilon)}{n}) = \Omega(n \log(n))$ bits of randomness to choose a subset of n coordinates from a BCH code, as low-bias BCH codes have a super-linear block length of $\text{poly}(n/\varepsilon)$ [15]. With our expander walk puncturing, we achieve $O(n)$ randomness even in this regime. We view it as an interesting open question to find pseudorandom puncturings with sufficient algebraic structure such that they may make decoding in this scheme tractable.

[1] We remark that [14] achieved this specific result by a different proof.

We remark that our analysis techniques using expander graphs have already found themselves applicable in subsequent works in coding theory. The work of [4] showed that *random shortening* (a different technique from random puncturing) can also be shown to work in a pseudorandom manner, using exactly the same hitting-set style argument. This leads to randomness efficient yet still non-trivial random shortening.

We further illustrate the flexibility of the pseudorandom puncturing approach in two regimes. First, we observe (Theorem 35) that pseudorandom linear codes achieve capacity against the memoryless additive channel, extending the analogous result of [14].

Finally, we apply our techniques in a different regime: we partially derandomize the result of Lund and Potukuchi [18], who show that random puncturings of Reed-Solomon codes can be list-recovered beyond the "Johnson bound."

▶ **Definition 3** (Zero-Error List Recoverability). *Let $\mathcal{C} \in \mathbb{F}_q^n$ be a code. We say $\mathcal{C}$ is (ℓ, L)* ***zero-error list recoverable*** *if for every collection of sets $A_1, \ldots, A_n$ with $|A_i| \leq \ell$ for all i, we have $|\{c \in \mathcal{C} : c \in A_1 \times \ldots \times A_n\}| \leq L$.*

We show that expander walk puncturings of Reed Solomon codes are zero-error list recoverable beyond the Johnson bound:

▶ **Theorem 4** (Zero-Error List Recovery of Reed Solomon Codes). *Given a prime power q and $\varepsilon \geq 1/\sqrt{q}$, there are Reed-Solomon codes of length n and rate $\Omega(\varepsilon/\log q)$ that can be sampled with $O(n)$ random bits that are $(\varepsilon^{-2}, O(\varepsilon^{-2}))$-zero error list recoverable whp.*

For comparison, the Johnson bound (for list recovery) states that a code $\mathcal{C}$ over $\mathbb{F}_q^n$ with distance at least $n(1-\varepsilon)$ is (ρ, ℓ, L) list-recoverable for any $\ell \leq \frac{(1-\rho)^2}{\varepsilon}$ and $L = \frac{\ell}{(1-\rho)^2 - \varepsilon\ell}$. As remarked in [18], this roughly translates to saying that any code over alphabet $\mathbb{F}_q$ of relative distance $1 - 1/q - \varepsilon$ is $(\rho, \ell, O(\ell))$ list-recoverable for $\ell = O(1/\varepsilon)$, and $\rho \leq 1 - \sqrt{2\varepsilon\ell}$. A natural question (and one that exceeds the capabilities of the Johnson bound) is whether one is able to construct q-ary Reed-Solomon codes that are (ρ, ℓ, L) list-recoverable for $\rho = 0, \ell = \omega(1/\varepsilon), L = \text{poly}(\ell)$. We show that this is indeed possible, even using an expander walk puncturing puncturing, by achieving $\ell = 1/\varepsilon^2, L = O(\ell)$.

We do this by analyzing how expander walk puncturings work for codes with near-maximal distance. We note that [14] analyzed random puncturings in the case of large distance as well, though their more structured analysis does not carry over to the regime of pseudorandom puncturings.

1.2 Proof Techniques and Comparison to [14]

As mentioned previously, for a truly random puncturing $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$, one perspective of such a function is that we take a length n random walk on a complete graph on the vertex set $[m]$. By reading off the vertices that this truly random walk visits $(u_1, \ldots u_n)$, and using these as the indices that we preserve from our m-dimensional vector, we can exactly model the action of a truly random puncturing. This model is easy to analyze in the sense that the n steps of the random walk are all independent. That is, the next vertex to be visited has no dependence on the current vertex that the walk is at, since each vertex is connected to all other vertices. A priori, it is not clear that replacing this complete graph with an expander should allow us to preserve all the desirable properties of our random puncturing. To this end, one of our main contributions is to characterize the "success" events of a random puncturing (i.e. in this case a success is being list-decodable) as a type of hitting constraint. We then invoke known results showing that the probability of satisfying this hitting constraint is approximately preserved even under an expander random walk. This allows us to improve and build upon the result of [14].

1.3 Overview

In Section 2 we recall concentration results for expander random walks, local properties of codes, and notation related to distributions on rows of a matrix. In Section 3 we prove a weaker version of Theorem 1 with exponential list sizes to introduce our proof strategy. In Section 4 we prove Theorem 1. In Section 5 we prove random puncturings achieve capacity in the memoryless channel. In Section 6 we prove Theorem 4, and in Appendix B we conclude a small derandomization of a result constructing unbalanced expanders.

2 Preliminaries

We first introduce concepts required for the proofs.

We recall the definitions of q-ary entropy and KL-divergence.

▶ **Definition 5** (q-ary entropy). *For $x \in [0,1]$ the q-ary entropy is defined to be*

$$h_q(x) = -x\log_q(x) - (1-x)\log_q(1-x) + x\log_q(q-1).$$

▶ **Definition 6** (q-ary KL divergence). *The q-ary KL divergence of two distributions τ, σ over a set S is defined as*

$$D_{KL_q}(\tau\|\sigma) = \sum_{s\in S} \tau(s)\log_q \frac{\tau(s)}{\sigma(s)}.$$

2.1 Properties of Expander Walks

We recall some useful statements of properties of expander random walks. We reference the excellent survey of Hoory, Linial, and Wigderson [16].

First, we reintroduce the definition of an expander, and that ones exist with good properties. Our results are not sensitive to the precise degree-expansion tradeoff, except in optimizing the constant factor on the number of bits required to sample.

▶ **Definition 7** (Expander graphs [16]). *We say a graph (G, V) is an (m, d, λ)-expander if G is d-regular on m vertices, and satisfies $|\lambda_2(G)|, |\lambda_m(G)| \leq \lambda d$. The notation $\lambda_i(G)$ refers to the ith eigenvalue of the adjacency matrix of G.*

▶ **Theorem 8** (Existence of near-optimal expanders [3, 22]). *For a fixed $d \in \mathbb{N}, \lambda$, there exist (strongly) explicit constructions of (m, d, λ)-expanders for all m large enough if $\lambda \geq \frac{1}{d^{0.49}}$.*

We remark that our notion of explicitness is that, given a vertex $v \in [m]$ and a neighbor $i \in [d]$, we can compute $\Gamma(v, i)$ in time $\text{poly}(\log m)$.

▶ Remark 9. To take an n-step random walk on an m vertex, degree d graph takes $\log(m) + n \cdot O(\log d)$ random bits. Using known degree-expansion trade-offs (Theorem 8) for the existence of strongly explicit expanders, (m, d, λ)-expanders exist for $\lambda \geq \frac{1}{d^{0.49}}$. So, for any given λ, one can choose $d = \left(\frac{1}{\lambda}\right)^{1/0.49}$, and then perform a random walk on an $(m, \left(\frac{1}{\lambda}\right)^{1/0.49}, \lambda)$ expander using $\log(m) + cn \cdot \log(1/\lambda) + O(n)$ random bits suffices for some constant $c = \frac{1}{0.49}$.

We will also make use of the non-equal expander hitting set lemma, which states that a random walk on an expander lies inside a sequence of sets with probability approximately the product of the sets densities. Our analysis relies on the ability of the sets to differ at each timestep.

▶ **Theorem 10** (Non-equal expander hitting-set lemma, [16], Theorem 3.11)**.** *Let $B_1, B_2, \dots B_t$ be vertex sets of densities $\beta_1, \dots \beta_t$ in an (m, d, λ)-graph G. Let $X_1, \dots X_n$ be an n-step random walk on G. Then,*

$$\Pr[\forall i \in [n], X_i \in B_i] \leq \prod_{i=1}^{n-1} \left(\sqrt{\beta_i \beta_{i+1}} + \lambda\right) \leq \left(\max_i \beta_i + \lambda\right)^{n-1}.$$

Additionally, we will require the expander Chernoff bound [7].

▶ **Theorem 11** (Expander Chernoff bound [7])**.** *Let G be an (m, d, λ) regular graph. Let $B \subset [m]$ be a set with density $\mu := |B|/m$. Let $X_1, \dots, X_n$ be an n-step random walk on G, initialized at a random vertex. Then,*

$$\Pr\left[\left|\sum_{i=1}^{n} \mathbb{I}[X_i \in B] - n\mu\right| > n\varepsilon\right] \leq 2e^{-\Omega((1-\lambda)n\varepsilon^2)}.$$

2.2 Expander Walk Puncturing

Of primary importance in this paper will be following instantiation of a pseudorandom puncturing:

▶ **Definition 12** (λ-expander-walk puncturing)**.** *Given a prime power q and $m, n \in \mathbb{N}$, a $(m \to n)$ **λ-expander-walk puncturing map (with puncturing graph G)** $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ is a random function obtained by taking an expander $G = ([m], E)$ satisfying $\lambda(G) \leq \lambda$ and taking a length n random walk. Letting the vertex labels of the walk be $(i_1, \dots, i_n)$, we define the map by*

$$\varphi(u = (u_1, \dots, u_m)) = (u_{i_1}, \dots, u_{i_n}).$$

*For $j \in [n]$ let $\varphi_j = i_j$ be the jth index of the map. Given a code $\mathcal{D} \subset \mathbb{F}_q^m$, which we call the **mother code**, we say $\mathcal{C}$ is a **λ-expander-walk puncturing** of D if*

$$\mathcal{C} := \varphi(\mathcal{D}) = \{\varphi(u) : u \in \mathcal{D}\}.$$

*The **design rate** of $\mathcal{C}$ is $R = \log_q |\mathcal{D}|/n$.*

We note that our λ-expander-walk puncturing map places no constraints on the expander beyond its spectral gap.

[14] show that the rate of a random puncturing (of a small-bias code) is equal to the design rate with high probability. We extend this result to λ-expander-walk puncturing puncturings, subject to mild constraints on the parameter λ.

▶ **Lemma 13** (Actual rate equals design rate with high probability)**.** *Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code of distance at least $(1 - 1/q - \eta)$, and let $\mathcal{C}$ be a length n λ-expander-walk puncturing of $\mathcal{D}$, of design rate $R \leq 1 - \log_q(1 + \eta q + \lambda q) - \varepsilon$. Then, with probability at least $1 - q^{-\varepsilon n}$, the rate of $\mathcal{C}$ is equal to its design rate.*

Proof. The event that the rate is less than the design rate occurs if there is some nonzero codeword $u \in \mathcal{D}$ such that $\varphi(u) = 0$. Fixing $u \in \mathcal{D}$, let $T \subset [m]$ be the coordinates on which u is zero. We have $\frac{|T|}{m} \leq \frac{1}{q} + \eta$ by assumption on distance. Then

$$\Pr[\varphi(u) = 0] = \Pr[\varphi_1 \in T \wedge \dots \wedge \varphi_n \in T] \leq \left(\frac{1}{q} + \eta + \lambda\right)^n = q^{-n(1 - \log_q(1 + q\eta + q\lambda))}$$

where the first inequality comes from Theorem 10. Then a union bound over the q^{Rn} codewords completes the proof. ◀

2.3 Properties of Codes

As in [14] and [19], we will be proving a result that generalizes to a wide class of *properties* of codes.

Our results will rely on the distance and bias of codes.

▶ **Definition 14** (Bias and distance). *Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code.*

1. *We say that $\mathcal{D}$ has η-**optimal distance** if the weight of every codeword is bounded below by $(1-1/q)(1-\eta)$. That is,*

$$\forall c \in \mathcal{D}, wt(c) \geq (1-1/q)(1-\eta).$$

2. *We say that $\mathcal{D}$ is η-**biased** if for every non-zero codeword $c \in \mathcal{D}$, for every $a \in \mathbb{F}_q^*$:*

$$\left|\sum_{i=1}^{m} \omega^{\mathrm{tr}(a \cdot c_i)}\right| \leq m\eta.$$

Here, we use that $\omega = e^{2\pi i/p}$, (where q is a power of a prime p), and $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ is defined as

$$\mathrm{tr}(x) = \sum_{i=0}^{r-1} x^{p^i},$$

where $r = \log_p q$.

We remark that Vazirani's ZOR lemma implies that small bias implies the row distribution is close to uniform:

▶ **Lemma 15** (Vazirani's XOR lemma [8,14]). *Let σ be an η-biased distribution over $\mathbb{F}_q^b$. Then, σ is $(q^b \cdot \eta)$-close in total-variation distance to the uniform distribution over $\mathbb{F}_q^b$.*

▶ Remark 16. A code that is η-biased has η-optimal distance, and most of our analysis uses only this property (though we use results of [14] which rely on the bias condition).

Now, we will first introduce a few specific examples of local properties, and then the more general definition for which our result will ultimately apply.

We first define ρ-clustered. We note that $\mathrm{wt}(x)$ is the normalized Hamming weight of x, and $B(z, \alpha)$ is the Hamming ball of weight α centered at z.

▶ **Definition 17** (ρ-clustered [14]). *Fix $\rho \in [0,1]$. We say that a set of vectors $W \subseteq \mathbb{F}_q^n$ is ρ-**clustered** if there exists a $z \in \mathbb{F}_q^n$ such that $wt(w-z) \leq \rho$ (equivalently, $w \in B(z, \rho n)$) for all $w \in W$.*

We recall the folklore observation that this definition gives a clean characterization of list decodability:

▶ **Observation 18.** *A code $C \subseteq \mathbb{F}_q^n$ is (ρ, L)-list decodable if and only if it does not contain a ρ-clustered set of codewords of size $L+1$.*

Both list-decodability and list-recoverability are special cases of properties of codes [14,19].

▶ **Definition 19** (Properties of a code). *A property $\mathcal{P}$ of length n linear codes over $\mathbb{F}_q$ is a collection of linear codes in $\mathbb{F}_q^n$. For such a code $\mathcal{C}$, if $\mathcal{C} \in \mathcal{P}$, then we say that $\mathcal{C}$ satisfies property $\mathcal{P}$. A property $\mathcal{P}$ is said to be monotone-increasing if $\mathcal{P}$ is upwards closed with respect to containment.*

▶ **Definition 20** (Local and row-symmetric properties). *Let $\mathcal{P}$ be a monotone-increasing property of linear codes in $\mathbb{F}_q^n$.*

1. *If, for a fixed $b \in \mathbb{N}$, there exists a family $\mathcal{B}_\mathcal{P}$ of sets of words, such that every $B \in \mathcal{B}_\mathcal{P}$ is a subset of $\mathbb{F}_q^n$, $|B| \leq b$, and*

$$\mathcal{C} \text{ satisfies } \mathcal{P} \iff \exists B \in \mathcal{B}_\mathcal{P} : B \subseteq \mathcal{C},$$

   then we say $\mathcal{P}$ is a b-local property.
2. *If, whenever a code $\mathcal{C}$ satisfies $\mathcal{P}$ and π is a permutation on $\{1, \dots n\}$, the code $\{\pi x | x \in \mathcal{C}\}$ also satisfies $\mathcal{P}$, then we say that $\mathcal{P}$ is row-symmetric. πx in this notation refers to permuting the entries of a vector of length n according to the permutation π.*

Note that the property of being *not* (ρ, L) list-decodable is a L-local row-symmetric property. We will use this in our result.

▶ **Definition 21** (Threshold of a property). *For $\mathcal{P}$ over $\mathbb{F}_q^n$, we will let*

$$\mathrm{RLC}(\mathcal{P}) = \min\left\{R \in [0,1] \middle| \Pr\left[\textit{RLC of length } n,\ \textit{rate } R,\ \textit{domain } \mathbb{F}_q \textit{ satisfies } \mathcal{P}\right] \geq 1/2\right\}.$$

This definition is motivated by the following theorem which was proved in [19].

▶ **Theorem 22** (Sharp threshold behavior [19]). *Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a random linear code of rate R and let $\mathcal{P}$ be a monotone-increasing, b-local, and row-symmetric property over $\mathbb{F}_q^n$, where $\frac{n}{\log_q n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Then, for every $\varepsilon > 0$, the following hold:*

1. *If $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon$*

$$\Pr[\mathcal{C} \textit{ satisfies } \mathcal{P}] \leq q^{-n(\varepsilon - o_{n\to\infty}(1))}.$$

2. *If $R \geq \mathrm{RLC}(\mathcal{P}) + \varepsilon$*

$$\Pr[\mathcal{C} \textit{ satisfies } \mathcal{P}] \geq 1 - q^{-n(\varepsilon - o_{n\to\infty}(1))}.$$

Because of Theorem 22, it suffices to show that the local behavior of a pseudorandom puncturing is similar to that of a random linear code. From there, we can invoke this result about thresholds to conclude whether or not a property $\mathcal{P}$ is satisfied with high probability.

2.4 Empirical Distributions

In order to eventually prove a tight bound on list sizes, we will need the notion of empirical distributions (types) from [5] [14].

▶ **Definition 23** (Empirical Distribution). *For a vector $a \in \mathbb{F}_q^n$, the* empirical distribution Emp_a *assigns probability $\forall x \in \mathbb{F}_q$:*

$$\mathrm{Emp}_a(x) = \frac{\textit{number of instances of } x \textit{ in } a}{n}.$$

This extends to a matrix $A \in \mathbb{F}_q^{n\times b}$ by defining $\forall x \in \mathbb{F}_q^b$

$$\mathrm{Emp}_A(x) = \frac{\textit{number of instances of } x \textit{ in rows of } A}{n}.$$

Note in this second case, Emp_A is a distribution over vectors $\in \mathbb{F}_q^b$.

For convenience, we will also introduce the set of matrices for a distribution.

▶ **Definition 24** (Matrices of a distribution). *Let τ be a distribution over $\mathbb{F}_q^b$. For $n \in \mathbb{N}$,*

$$\mathcal{M}_{n,\tau} = \left\{A \in \mathbb{F}_q^{n\times b} | \mathrm{Emp}_A = \tau\right\}.$$

Lastly, we will consider all sequences of samples that lead to a specific empirical distribution.

▶ **Definition 25** (Type class). *The* type class *of a distribution τ over $\mathbb{F}_q^b$ (denoted $T(\tau)$) is the set of all sequences $(x_i)_{i=1}^n$ in $\left(\mathbb{F}_q^b\right)^n$ such that for the matrix*

$$X = \begin{bmatrix} - & x_1 & - \\ \vdots & \vdots & \vdots \\ - & x_n & - \end{bmatrix},$$

we have that $\mathrm{Emp}_X = \tau$.

Lastly, we define a full-rank distribution.

▶ **Definition 26.** *For a distribution τ over $\mathbb{F}_q^b$, we say τ is a full-rank distribution if for every $x_i \in \mathbb{F}_q^b$ in the support of τ, if we write these x_i as the rows of a matrix:*

$$\begin{bmatrix} - & x_1 & - \\ - & x_2 & - \\ \vdots & \vdots & \vdots \\ - & x_m & - \end{bmatrix},$$

then this matrix is full rank.

▶ **Theorem 27** ([5, 14]). *Let $X \in \mathbb{F}_q^{n\times b}$ have rows identically and indpendently sampled from some distribution σ over $\mathbb{F}_q^b$. Then, for any distribution τ over $\mathbb{F}_q^b$,*

$$\Pr[\mathrm{Emp}_X = \tau] \leq q^{-D_{KL_q}(\tau\|\sigma)n}.$$

3 List Decodability of Pseudorandom Linear Codes

In this section we give an outline of our main proof technique. For simplicity, we do not attain optimal list-size, and consider only the regime of list-decoding (as opposed to more general local properties). Our proof closely follows that of Theorem 7 of [14]. We state our initial result:

▶ **Theorem 28.** *Let $\rho \in (0, 1/2)$. Then suppose the following bounds hold as $L \to \infty$:*

$$\eta(L) \leq L^{-4}, \quad \lambda(L) \leq L^{-2}, \quad \varepsilon \geq 4/\log(L).$$

Let $\mathcal{D} \subset \mathbb{F}_2^m$ be an arbitrary linear η-biased code, and let $\mathcal{C} \subset \mathbb{F}_2^n$ be a λ-expander-walk n-puncturing of $\mathcal{D}$ of design rate $R \leq 1 - h_2(\rho) - \varepsilon$. Then $\mathcal{C}$ is (ρ, L)-list-decodable and has actual rate R with high probability as $n \to \infty$ and requires $O(n/\varepsilon)$ random bits to construct.

Proof. Let φ be an $(m \to n)$ λ-expander-walk puncturing, for λ to be chosen later, and let $\mathcal{C} := \varphi(\mathcal{D})$. Let $b := \lceil \log(L + 1) \rceil$, and note that with this choice we have

$$\eta \leq 2^{-2b}, \quad \lambda \leq 2^{-b}, \quad \varepsilon \geq 4/b.$$

Recall that $\mathcal{C}$ fails to be list decodeable if there exist $L+1$ codewords that are ρ-clustered (Observation 18). Recalling an argument first used in [23], a necessary condition for this is for $\mathcal{C}$ to contain b linearly independent (L.I.) ρ-clustered codewords. Thus:

$$\begin{aligned}\Pr[\mathcal{C} \text{ fails to be } (\rho, L)\text{-list decodable}] &\leq \Pr[\exists v_1, \ldots, v_b \in \mathcal{C} \text{ that are L.I. and } \rho\text{-clustered}] \\ &\leq \sum_{\substack{u_1,\ldots,u_b \in \mathcal{D} \\ \text{lin. indep.}}} \Pr[\varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered}].\end{aligned}$$

Here, we have used the substitution that $v_i = \varphi(u_i)$, where $u_i \in \mathcal{D}$. Because $v_1, \ldots v_b$ are linearly independent, this means that $u_1, \ldots u_b$ must also be linearly independent. Note that this sum is over at most $|\mathcal{D}|^b \leq 2^{bRn}$ terms. Now fix arbitrary, linearly independent $u_1, \ldots, u_b \in \mathcal{D}$ and let

$$B := \begin{bmatrix} | & & | \\ u_1 & \ldots & u_b \\ | & & | \end{bmatrix} \in \mathbb{F}_2^{m \times b}, \qquad A := \begin{bmatrix} | & & | \\ \varphi(u_1) & \ldots & \varphi(u_b) \\ | & & | \end{bmatrix} \in \mathbb{F}_2^{n \times b}$$

where A is a random matrix defined in terms of the puncturing φ. We now note

$$\begin{aligned}\Pr[\varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered}] &= \Pr[\exists z, y_1, \ldots, y_b \in B(z, \rho n) \text{ s.t. } \forall i, y_i = \varphi(u_i)] \\ &\leq \sum_{z \in \mathbb{F}_2^n} \sum_{y_1,\ldots,y_b \in B(z,\rho n)} \Pr[\forall i, \varphi(u_i) = y_i]\end{aligned}$$

Now fix arbitrary z and $y_1, \ldots, y_b \in B(z, \rho n)$ (of which there are at most $2^{n + bh_2(\rho)n}$). Define the matrix

$$Y := \begin{bmatrix} | & & | \\ y_1 & \ldots & y_b \\ | & & | \end{bmatrix}.$$

Finally, for $\sigma \in \mathbb{F}_2^b$ let $T_\sigma \subseteq [m]$ be defined as

$$T_\sigma := \{j \in [m] : B_j = \sigma\},$$

where B_j is the jth row of the matrix B. In words, each set T_σ is the set of indices j such that the jth row of B equals σ.

We now argue that $\tau_\sigma \approx 2^{-b}$ for every σ. Note that this would hold exactly if the rows of B_j were uniformly distributed over $\{0,1\}^b$. We first claim that the row distribution is low-bias. For every $y \in \{0,1\}^b$, observe that

$$\Pr_{j \leftarrow U_m}[\langle B_j, y \rangle = 1] = \mathrm{wt}(By).$$

As the u_i are linearly independent codewords, we have that By is a nonzero codeword, and hence $\mathrm{wt}(By) \in 1/2 \pm \eta$. Then applying Lemma 15 and our choice of η we obtain

$$\tau_\sigma \leq 2^{-b} + 2^b \eta \leq 2^{-b+1}$$

Then we have

$$\begin{aligned}\Pr[\forall i, \varphi(u_i) = y_i] &= \Pr[\varphi_1 \in T_{Y_1}, \ldots, \varphi_n \in T_{Y_n}] \\ &\leq \left(\max_\sigma \tau_\sigma + \lambda\right)^{n-1} && \text{(Theorem 10)} \\ &\leq \left(2^{-b+2}\right)^n \cdot 2^{b-2},\end{aligned}$$

where the first equality follows from the fact that enforcing the columns of A to be the same as the columns of Y is the same as requiring the rows of A to be the same as the rows of Y. Now, for each row, the probability that these rows are equal is the probability that the ith coordinate selected in the expander random walk puncturing (denoted by φ_i) corresponds to a row from the matrix B which equals Y_i, i.e., that φ_i is in T_{Y_i}. The final line follows from $\lambda \leq 2^{-b}$ and $\tau_i \leq 2^{-b+1}$ and changing the product to be over n terms. Thus the entire expression is bounded as

$$\begin{aligned}\Pr[\mathcal{C} \text{ fails to be } (\rho, L)\text{-list decodable}] &\leq 2^{bRn} \cdot 2^n \cdot 2^{bh_2(\rho)n} \cdot (2^{-b+2})^n \cdot 2^{b-2} \\ &\leq 2^{b(1-h_2(\rho)-4/b)n+3n+bh_2(\rho)n-bn+b-2} \\ &= 2^{-n+b-2} \to 0,\end{aligned}$$

where in the second line we use $\varepsilon \geq 4/b$.

Randomness Complexity. Note that this construction only requires $O(\log m + n \log d)$ random bits, where d is the degree of the expander graph. As our expansion constraint is $\lambda \leq 2^{-b}$, and by Theorem 8, this can be done such that $O(\log m + n \log d) = O(nb)$. Finally, because the expanders we are using are *strongly* explicit, we can perform this random walk in time $\text{poly}(n, \log(2^{Rn})) = \text{poly}(n)$.

Rate. Note that the design rate of $\mathcal{C}$ is bounded by $1 - h_2(\rho) - \varepsilon$. Applying Lemma 13 to $\mathcal{D}$ with $\varepsilon = \varepsilon$, we have that the actual rate of $\mathcal{C}$ is equal to this design rate as long as

$$1 - h_2(\rho) - \varepsilon \leq 1 - \log(1 + 2^{-b+1}) - \varepsilon$$

i.e. $\log(1 + 2^{-b+1}) \leq h_2(\rho)$, which occurs for sufficiently large L as ρ is constant. ◀

In the simplest case, we can take the mother code to the Hadamard code mapping messages of length Rn to codewords of length 2^{Rn}. The generator matrix for this Hadamard code is $\in \mathbb{F}_2^{2^{Rn} \times Rn}$. Choosing the starting vertex for the expander random walk in this case takes Rn bits of randomness, and for every subsequent step, the amount of randomness required depends only on the degree of the expander. Ultimately, the pseudorandom puncturing results in a generator matrix of size $n \times Rn$. For a desired rate $1 - H(\rho) - \varepsilon$, we take $b = 4/\varepsilon$. Correspondingly, we need $\lambda \leq 2^{-4/\varepsilon}$, which forces $\log d = \Omega(1/\varepsilon)$. Thus, we pay for the randomness linearly in $1/\varepsilon$.

However, because we set $\varepsilon = \frac{4}{b}$, we get that $b = \frac{4}{\varepsilon}$, meaning that the list size $L = 2^{\Omega(1/\varepsilon)}$, which is far from optimal. In the next section, we give a more careful argument that achieves optimal list sizes.

4 Pseudorandom Puncturings Preserve Local Properties

In this section, we give an analogue of the more detailed analysis presented in [14] for our instantiation of a pseudorandom puncturings.

In particular, we will show the following, and use it to conclude Theorem 1:

▶ **Theorem 29.** *Let q be a prime power, and let $\mathcal{P}$ be a monotone-increasing, row-symmetric and b-local property over $\mathbb{F}_q^n$, where $\frac{n}{\log n} \geq \omega_{n\to\infty}(q^{2b})$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code. Let $\mathcal{C}$ be a $\lambda = \frac{\varepsilon \ln q}{8q^b}$-expander-walk puncturing of $\mathcal{D}$ of design rate $R \leq RLC(\mathcal{P}) - \varepsilon$ for some $\varepsilon > 0$. Suppose that $\mathcal{D}$ is $\eta = \frac{\varepsilon b \ln q}{4q^{2b+1}}$-biased. Then,*

$$\Pr[\mathcal{C} \textit{ satisfies } \mathcal{P}] \leq q^{(-\varepsilon + o_{n\to\infty}(1))n}.$$

At a high level, our proof has the following form:

1. First, fix an η-biased code $\mathcal{D}$, a distribution τ over $\mathbb{F}_q^b$, and a set of b linearly independent columns in $\mathcal{D}$, which we denote $(\mathcal{D})_{\text{res}}$. We show that if we sample rows of $(\mathcal{D})_{\text{res}}$ via an expander-walk puncturing, we can upper bound the probability of our sampled rows having the same marginal probabilities as τ. This bound will be in terms of q, the KL divergence between τ and the distribution produced by a *truly* random puncturing, and some error terms. That is, we will show (for specific conditions):

$$\Pr[\text{Emp}_X = \tau] \leq q^{n\left(-D_{\text{KL}_q}(\tau\|\sigma)+\log_q\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)+o_n(1)\right)},$$

   where Emp_X is the empirical distribution (under the pseudorandom puncturing) of the rows sampled from $(\mathcal{D})_{\text{res}}$, and σ is the empirical distribution of the rows of $(\mathcal{D})_{\text{res}}$.
2. Next, we invoke results from [14] which characterize codes satisfying local properties in terms of the number of submatrices contained in the code that have a specific row distribution. This result is independent of the puncturing procedure, and shows that it suffices to prove

$$\mathbb{E}_{\mathcal{C}}\left[|\{X \subseteq \mathcal{C} | X \in \mathcal{M}_{n,\tau}\}|\right] \leq q^{(H_q(\tau)-a(1-R)+a\varepsilon)n},$$

   where $\mathcal{M}_{n,\tau}$ is from Definition 24, and a is the "rank" of the distribution τ (i.e. the rank of the matrix whose rows are from the support of τ).
3. Finally, we use item (1) to prove the bound from item (2) and conclude our proof. That is, we will use the fact that a matrix $X \subseteq \mathcal{C}$ is in $\mathcal{M}_{n,\tau}$ only if $\text{Emp}_X = \tau$. As we have strong bounds on this event from item (1), we can invoke a union bound and prove the desired result.

As in [14] we can then invoke Theorem 22. Because we will show that the local behavior of our instantiation of a pseudorandom puncturing is similar to that of a random linear code, we can use the tight characterization of the local behavior of RLCs to conclude our result.

4.1 Analysis

First, we prove the following lemma (which is a modification of a statement from [5]):

▶ **Lemma 30.** *Let $D \in \mathbb{F}_q^{m\times b}$ be b linearly independent codewords from an η-biased code of length m, where $\eta < q^{-2b}/4$. Further, let σ be Emp_D. Suppose that we sample rows of D in accordance with a length n λ-expander random walk over vertex set $[m]$, and place these as the rows in a matrix $X \in \mathbb{F}_q^{n\times b}$. Then for every distribution τ over $\mathbb{F}_q^b$,*

$$\Pr[\text{Emp}_X = \tau] \leq q^{n\left(-D_{KL_q}(\tau\|\sigma)+\log_q\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)+o_n(1)\right)}.$$

Proof. We let $T(\tau)$ denote the type class of τ (see Definition 25). In this context, we will let $P \in T(\tau)$ denote a specific sequence of samples in $\left(\mathbb{F}_q^b\right)^n$, such that the marginals are τ. We will let P_i be an element in $\mathbb{F}_q^b$ corresponding to the ith sample of this sequence. From the perspective of the expander random walk, we will let B_i denote the set of all vertices of the expander (i.e. indices from $[m]$) such that the corresponding sample (corresponding row of D) is P_i. We will let β_i denote the density of B_i. We will let $X_1, \ldots X_n$ denote the random walk over $[m]$, the rows of the mother code.

Then we have:

$$\begin{aligned}
\Pr[\mathrm{Emp}_X = \tau] &= \sum_{P \in T(\tau)} \Pr[\wedge_{i=1}^{n} X_i \in B_i] \\
&= \sum_{P \in T(\tau)} \prod_{i=1}^{n} \Pr[X_i \in B_i | \forall j < i, X_j \in B_j] \\
&\leq \sum_{P \in T(\tau)} \prod_{i=1}^{n-1} \left(\sqrt{\beta_i \beta_{i+1}} + \lambda \right) \\
&\leq \sum_{P \in T(\tau)} \prod_{i=1}^{n-1} \left(\beta_i \cdot q^{\log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \right) \\
&= \sum_{P \in T(\tau)} \prod_{i=1}^{n-1} q^{\log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \sigma(P_i)
\end{aligned}$$

where the last inequality comes from the fact that (letting $\beta = \min_i \beta_i$):

$$\prod_i (\sqrt{\beta_i \beta_{i+1}} + \lambda) \leq \prod_i \beta_i (1 + \beta^{-1}\lambda) \leq \left(\prod_i \beta_i \cdot q^{\log_q (1+\beta^{-1}\lambda)} \right).$$

Then, because the mother code is η-biased and we have a selection of linearly independent codewords, we get that $\beta \geq q^{-b} \cdot \left(1 - q^{2b} \cdot \eta\right)$ by Vazirani's XOR Lemma [8]. So, $\beta^{-1}\lambda \leq \frac{\lambda q^b}{(1-q^{2b}\eta)}$. Now we can bound $\Pr[\mathrm{Emp}_X = \tau]$. We see that

$$\begin{aligned}
\Pr[\mathrm{Emp}_X = \tau] &\leq \sum_{P \in T(\tau)} \prod_{i=1}^{n-1} q^{\log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \sigma(P_i) \\
&\leq \frac{q^b}{1 - q^{2b}\eta} \sum_{P \in T(\tau)} \prod_{i=1}^{n} q^{\log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \sigma(P_i) \\
&= \frac{q^b}{1 - q^{2b}\eta} q^{n \log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \cdot \sum_{P \in T(\tau)} \prod_{i=1}^{n} \sigma(P_i) \\
&= \frac{q^b}{1 - q^{2b}\eta} q^{n \log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right)} \cdot q^{-D_{\mathrm{KL}_q}(\tau \| \sigma) n}
\end{aligned}$$

where the last equality holds from the fact that $\sum_{P \in T(\tau)} \prod_{i=1}^{n} \sigma(P_i) = q^{-D_{\mathrm{KL}_q}(\tau \| \sigma) n}$ (Theorem 27). The second inequality comes from upper bounding $\frac{1}{\sigma(P_i)}$, so we can extend the product to n terms. By our choice of η, the leading term $\frac{q^b}{1-q^{2b}\eta}$ is $O(q^b)$, and is thus $q^{n \cdot o_n(1)}$. ◀

Using Lemma 30, we can now prove the following key lemma:

▶ **Lemma 31.** *Fix a distribution τ over $\mathbb{F}_q^b$. Let $B \in F_q^{m \times b}$ have* rank $B = b$ *and its column span be η-biased. Let $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ be a λ-expander-walk puncturing. Then,*

$$\Pr[\varphi(B) \in \mathcal{M}_{n,\tau}] \leq q^{n \left(\log_q \mathbb{E}_{x \sim \mathrm{Emp}_B} [\tau(x)] + H_q(\tau) + \log_q \left(1 + \frac{\lambda q^b}{(1-q^{2b}\eta)} \right) + o_n(1) \right)}.$$

Proof. We have that

$$\Pr[\varphi(B) \in \mathcal{M}_{n,\tau}] = \Pr[\mathrm{Emp}_{\varphi(B)} = \tau] \leq q^{n\cdot\left(-D_{\mathrm{KL}_q}(\tau\|\sigma)+\log_q\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)+o_n(1)\right)},$$

by Lemma 30, where $\sigma = \mathrm{Emp}_B$. From here, as in [14], we attain the stated bound by using the concavity of log and the definition of D_{KL_q}. ◀

We then recall two lemmas from [14], with no modification, which we use in the proof:

▶ **Lemma 32** (Lemma 5.9 [14]). *Let $B \in \mathbb{F}_q^{m\times b}$ have* rank $B = b$*, and let $f : \mathbb{F}_q^b \to \mathbb{R}$ be a non-negative function. Suppose that the column span of B is η-biased, for some $\eta \geq 0$. Then,*

$$\mathbb{E}_{x\sim\mathrm{Emp}_B}[f(x)] \leq \left(1+q^b\eta\right)\cdot\mathbb{E}_{x\sim U(\mathbb{F}_q^b)}[f(x)].$$

▶ **Lemma 33** (Lemma 6.12 [14]). *Let $n \in \mathbb{N}$, q a prime power and $b \in \mathbb{N}$ such that $\frac{n}{\log_q n} \geq \omega_{n\to\infty}(q^{2b})$. Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a linear code of rate $R \in [0,1]$, sampled at random from some ensemble. Suppose that, for every $1 \leq a \leq b$, every distribution τ over $\mathbb{F}_q^a$ and every matrix $A \in \mathbb{F}_q^{n\times a}$ with* rank $A = a$*, we have*

$$\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} | A \subseteq \mathcal{C}\}|\right] \leq q^{(H_q(\tau)-a(1-R)+a\varepsilon)n},$$

for some fixed $\varepsilon > 0$. Then, for every row-symmetric and b-local property $\mathcal{P}$ over $\mathbb{F}_q^n$ such that $R \leq RLC(\mathcal{P}) - 2\varepsilon$, it holds that

$$\Pr_{\mathcal{C}}[\mathcal{C} \text{ satisfies } \mathcal{P}] \leq q^{-n(\varepsilon-o_{n\to\infty}(1))}.$$

Lastly, we require one final lemma before we can conclude the final result:

▶ **Lemma 34.** *Fix $b \in \mathbb{N}$, and a full-rank distribution τ over $\mathbb{F}_q^b$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a η-biased linear code. Let φ be a λ-pseudorandom $(m \to n)$ puncturing map. Let $R = \frac{\log_q|\mathcal{D}|}{n}$. Then,*

$$\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} | A \subseteq \mathcal{C}\}|\right] \leq q^{n\left(H_q(\tau)-(1-R)b+\log_q\left[\left(1+\eta q^b\right)\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)\right]+o_n(1)\right)}.$$

Proof. Let τ be a full-rank distribution over $\mathbb{F}_q^b$. From Lemma 32, we have that

$$\mathbb{E}_{x\sim\mathrm{Emp}_B}[\tau(x)] \leq q^{-b}\left(1+\eta q^b\right),$$

for all $B \in \mathbb{F}_q^{m\times b}$ such that rank $B = b$ and $B \subseteq \mathcal{D}$. From Lemma 31,

$$\Pr[\varphi(B) \in \mathcal{M}_{n,\tau}] \leq q^{n\left(\log_q \mathbb{E}_{x\sim\mathrm{Emp}_B}[\tau(x)]+H_q(\tau)+\log_q\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)+o_n(1)\right)}.$$

Plugging in for $\mathbb{E}_{x\sim\mathrm{Emp}_B}[\tau(x)]$, we get that

$$\begin{aligned}\Pr[\varphi(B) \in \mathcal{M}_{n,\tau}] &\leq q^{n\left(\log_q\left(q^{-b}\cdot\left(1+\eta q^b\right)\right)+H_q(\tau)+\log_q\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)+o_n(1)\right)} \\ &= q^{n\left(-b+H_q(\tau)+\log_q\left[\left(1+\eta q^b\right)\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)\right]+o_n(1)\right)}.\end{aligned}$$

Now, by taking a union bound over the at most q^{Rnb} choices of B, we get that

$$\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} | A \subseteq \mathcal{C}\}|\right] \leq q^{n\left(H_q(\tau)-(1-R)b+\log_q\left[\left(1+\eta q^b\right)\left(1+\frac{\lambda q^b}{(1-q^{2b}\eta)}\right)\right]+o_n(1)\right)}.$$ ◀

Now, we can prove the main theorem of this section, by showing that for specific choices of λ, η, the bound from the previous lemma satisfies the conditions of Lemma 33.

Proof of Theorem 29. Let ε be given. Let τ be a distribution over $\mathbb{F}_q^a$ with $a \leq b$. Note that WLOG we can assume that τ is full-rank, as otherwise, we can project τ to a full-rank distribution τ' over $\mathbb{F}_q^r$ for $r \leq a \leq b$. That is, there exists a matrix $P \in \mathbb{F}_q^{r \times a}$ such that $P\tau = \tau'$. From Lemma 34, we have that

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} | A \subseteq \mathcal{C}\}|\right] &\leq q^{n\left(H_q(\tau)-(1-R)a+\log_q\left[(1+\eta q^a)\left(1+\frac{\lambda q^a}{(1-q^{2a}\eta)}\right)\right]+o_n(1)\right)} \\ &\leq q^{n\left(H_q(\tau)-(1-R)a+\frac{\eta q^a}{\ln q}+\frac{\lambda q^a}{\ln q\cdot(1-q^{2a}\eta)}+\frac{\eta\lambda q^{2a}}{(1-q^{2a}\eta)\ln q}+o_n(1)\right)}.\end{aligned}$$

Now, note that by our substitutions $\eta = \frac{\varepsilon \ln q}{4q^{2b+1}}$ and $\lambda = \frac{\varepsilon \ln q}{8q^b}$,

$$\begin{aligned}\frac{\eta q^a}{\ln q} &= \frac{\varepsilon b \ln q}{4q^{2b-a}\cdot q \cdot \ln q} \\ &\leq \frac{\varepsilon b}{4q^{2b-a}\cdot q} \\ &\leq \varepsilon a/4.\end{aligned}$$

Additionally,

$$\begin{aligned}\frac{\lambda q^a}{\ln q\cdot(1-q^{2a}\eta)} &= \frac{\varepsilon}{8q^{b-a}\cdot(1-q^{2a}\cdot\frac{\varepsilon\ln q}{4q^{2b+1}})} \\ &= \frac{\varepsilon}{8q^{b-a}\cdot(1-\frac{\varepsilon\ln q}{4q^{2b-2a+1}})} \\ &\leq \frac{\varepsilon}{8q^{b-a}\cdot(1-\frac{\varepsilon}{4q^{2b-2a}})} \\ &\leq \frac{\varepsilon}{8q^{b-a}(1-\varepsilon/4)} \leq \frac{\varepsilon}{4}.\end{aligned}$$

Lastly, by combining the above two results,

$$\frac{\eta\lambda q^{2a}}{(1-q^{2a}\eta)\ln q} \leq \ln q \cdot \frac{\varepsilon b}{4q^{2b-a}\cdot q}\cdot\frac{\varepsilon}{4} \leq \frac{\varepsilon^2 b}{16q^{2b-a}},$$

where we have taken advantage of the fact that our expression is $\ln q$ multiplied by the two terms we have already bounded before. Thus, all three expressions are bounded by $a \cdot \frac{\varepsilon}{4}$. As a result,

$$\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} | A \subseteq \mathcal{C}\}|\right] \leq q^{n\left(H_q(\tau)-(1-R)a+\frac{3}{4}\cdot a\varepsilon+o_n(1)\right)}.$$

We can then invoke Lemma 33 to conclude our result for sufficiently large n. ◀

Now, by noting that list-decoding is a $O(1/\varepsilon)$-local property in our specific setting, we can conclude Theorem 1 by using Theorem 29 and [11]. That is:

Proof of Theorem 1. Suppose we fix $\rho \in (0, 1/2)$. Then, there exists a constant $\alpha > 0$ such that the threshold for an RLC being $(\rho, \alpha/\varepsilon)$ list-decodable is $1 - H(\rho) - \varepsilon$ [11]. We let $\mathcal{P}$ denote the property of being $(\rho, \alpha/\varepsilon)$-list decodable. This means that $b = O(1/\varepsilon)$. Now, let $\mathcal{D}$ be a mother-code over $\mathbb{F}_q^m$ which is $\eta = \frac{\varepsilon b \ln q}{4q^{2b+1}}$-biased (and note that the Hadamard code satisfies this property). From Theorem 29, we know that for a $\lambda = \frac{\varepsilon \ln q}{8q^b}$-expander-walk puncturing of design rate $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon$ of $\mathcal{D}$, for every $\varepsilon > 0$:

$$\Pr[\mathcal{C} \text{ satisfies } \mathcal{P}] \leq q^{(-\varepsilon+o_{n\to\infty}(1))n}.$$

Thus, we can choose the design rate of the code to be $\mathrm{RLC}(\mathcal{P}) - \varepsilon = 1 - H(\rho) - 2\varepsilon$, and with high probability, our code will still be $(\rho, \alpha/\varepsilon)$ list-decodable.

By our choice in parameters for λ, we require the degree of the graph to be $O\left(\frac{q^{2b}}{\varepsilon^2(\ln q)^2}\right)$. This means that every step will require $O(b \log q + \log 1/\varepsilon)$ random bits. Using the fact that $b = O(1/\varepsilon)$, this means that every step in the expander random walk requires $O((\log q)/\varepsilon)$ random bits. Because q will be chosen to be a constant, this requires $O(1/\varepsilon)$ random bits per step. Initializing the random walk requires $\log m$ random bits, where m is the length of the mother code $\mathcal{D}$. Fortunately, $m \leq q^n$, so $\log m \leq n \log q$ (every generating matrix of length longer than q^n will have duplicate rows). As such the total amount of randomness required is $O(n + n \cdot 1/\varepsilon) = O(n/\varepsilon)$, as we desire. ◀

We can conclude the statement of Theorem 2 almost identically, where we instead treat b as a parameter, instead of substituting $O(1/\varepsilon)$.

5 Random Noise Tolerance of PRLCs

We show pseudorandom linear codes achieve capacity against the memoryless additive channel. Our proof follows directly from the argument of [14], except we argue that a *pseudorandom* puncturing approximately preserves the probability that a random vector lies in the code. In the context of this channel, we use the MLDU (maximum likelihood decoder under uniform prior). Upon receiving a corrupted codeword $z \in \mathbb{F}_q^n$, this decoder returns the codeword x that maximizes

$$\Pr[\text{receive } z | \text{original codeword was } x].$$

▶ **Theorem 35.** *Given a prime power q and a distribution X over $\mathbb{F}_q$ and $\varepsilon \in (0,1)$, let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be an $\varepsilon/8q$-biased linear code and let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a $\varepsilon/8q$-expander-walk puncturing of $\mathcal{D}$ with design rate $R \leq 1 - H_q(X) - \varepsilon$. Then there is a constant $c_X > 0$ such that with probability $1 - q^{c_X \cdot \varepsilon n}$, for every $x \in \mathcal{C}$ we have*

$$\Pr_{z \leftarrow X^n}[\textit{MLDU decodes } x + z \textit{ to } x] \geq 1 - 2q^{c_X \cdot \varepsilon^2 n}.$$

Inspecting the proof of [14, Theorem 6] gives the following:

▶ Remark 36. Let q be a prime power, ν a distribution over $\mathbb{F}_q$, $\varepsilon \in (0,1)$, and $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a probabilistically constructed linear code of design rate $R \leq 1 - H_q(\nu) - \varepsilon$ such that for every non-zero $x \in \mathbb{F}_q^n$

$$\Pr_{\mathcal{C}}[x \in \mathcal{C}] \leq q^{n \cdot (-1+R+\varepsilon/4)}.$$

Then, with probability $1 - q^{-\Omega_\nu(\varepsilon n)}$, it holds for all $x \in \mathcal{C}$ that

$$\Pr_{z \sim \nu^n}[\text{the MLDU outputs } x \text{ on input } x + z] \geq 1 - 2q^{-c_\nu \varepsilon^2 n}.$$

We will prove the following lemma, from which we can then immediately conclude Theorem 35 by using Remark 36.

▶ **Lemma 37.** *Let $\mathcal{C} \subset \mathbb{F}_q^n$ be a λ-expander-walk puncturing of a η-biased linear code $\mathcal{D} \subset \mathbb{F}_q^m$. For every $x \in \mathbb{F}_q^n \setminus \{0\}$, we have*

$$\Pr[x \in \mathcal{C}] \leq q^{n(R+\eta q+\lambda q-1)}.$$

Proof. Fix an arbitrary nonzero x and fix arbitrary $u \in \mathcal{D}$. For $i \in [q]$ let

$$T_i := \{j \in [m] : u_j = i\}.$$

and let $\tau_i := T_i/m$. By the bias of $\mathcal{D}$, we have $\tau_i \leq \frac{1}{q} + \frac{q-1}{q}\eta$ for every i.

Thus we have

$$\begin{aligned}\Pr[x \in \mathcal{C}] &= \Pr[\varphi_1 \in T_{x_1} \wedge \ldots \wedge \varphi_n \in T_{x_n}] \\ &\leq \left(\max_i \tau_i + \lambda\right)^n && \text{(Theorem 10)} \\ &\leq q^{n(\eta q+\lambda q-1)}\end{aligned}$$

and then a union bound over the q^{Rn} codewords completes the proof. ◀

Proof of Theorem 35. Let $\lambda = \eta = \frac{\varepsilon}{8q}$. The bound from the previous lemma then states that

$$\Pr[x \in \mathcal{C}] \leq q^{n(R+\varepsilon/8+\varepsilon/8-1)} = q^{n(R+\varepsilon/4-1)}.$$

We can then invoke Remark 36. ◀

6 Pseudorandom Puncturings of Large Distance Codes

We next show that our instantiation of pseudorandom puncturings of large distance codes are list recoverable beyond the Johnson Bound. We show this for the specific case of zero-error list-recovery, as it simplifies the exposition.

▶ **Theorem 38.** *Fix $\alpha \in (0, 1]$. Let $\mathcal{D} \subset \mathbb{F}_q^m$ be a linear code with distance at least $m(1 - q^{-1} - \varepsilon^2)$. Let φ be a 1/4-expander walk $(m \to n)$ puncturing with $n = O(\log|\mathcal{D}|/\varepsilon)$. Then $\varphi(\mathcal{D})$ has rate $\Omega\left(\frac{\varepsilon}{\log q}\right)$ and is $(\ell, \ell(1+\alpha))$-zero error list recoverable with high probability, assuming:*

$$1/\sqrt{q} \leq \varepsilon \leq \min(c, \alpha/4), \quad \ell \leq \alpha/4\varepsilon^2.$$

We state the main theorem that allows us to establish this, which is analogous to Theorem 3.1 of [18].

▶ **Theorem 39.** *Given $\alpha \in (0, 1)$, let $q, m, d, \ell, n \in \mathbb{N}$. Given a code $\mathcal{D} \subset \mathbb{F}_q^m$ of distance at least $1 - mq^{-1} - d$. Suppose that*

$$d \geq mq^{-1}, \quad 4\alpha^{-1} \leq \ell \leq \alpha m/1600d, \quad n = \Omega\left(\sqrt{\ell/\alpha}\log|D|\right), \quad m > n.$$

Then the probability that $C := \varphi(D)$ (where φ is a 1/4-expander-walk puncturing) is $(\ell, (1+\alpha)\ell)$-zero error list recoverable is at least $1 - \exp(-\sigma n/100)$.

Our proof differs from that of [18] in two ways: our puncturing is pseudorandom, rather than truly random, and we argue about puncturings produced with replacement (which is natural in the setting of expander random walks which may revisit vertices).

We first introduce some notation that will be used in the proof.

▶ **Definition 40.** *For an arbitrary code $\mathcal{C} \subset \mathbb{F}_q^m$, let*

$$T(\mathcal{C}) = \{i \in [m] \mid \exists c_1 \neq c_2 \in \mathcal{C}, c_1[i] = c_2[i]\}.$$

We first argue that, given an index set $\varphi \in [m]^n$ such that $\varphi(\mathcal{D})$ is not list recoverable with the claimed parameters, there is a small subcode that fails to be.

▶ **Lemma 41.** *Let $\varphi \in [m]^n$ be such that $\varphi(\mathcal{D})$ fails to be $(\ell, \ell(1+\alpha))$-zero error list recoverable. Then there is a subcode $\mathcal{C}' \subset \mathcal{D}$ such that:*

- $|\mathcal{C}'| \leq 10\sqrt{\ell/\gamma}$
- $|\{i \in [n] : \varphi_i \in T(\mathcal{C}')\}| \geq n/8$

The proof of this lemma closely follows Theorem 3.1 in [18], and as such we defer it to the appendix. We furthermore require a concentration bound for the number of bad indices selected by the puncturing map, which is a simple consequence of the expander Chernoff bound.

▶ **Lemma 42.** *Let $B \subset [m]$ be a bad set of indices satisfying $\beta := |B|/m \leq 1/16$, and let φ be a $1/4$-expander-walk puncturing. Then*

$$\Pr\left[|\{i \in [n] : \varphi_i \in B\}| \geq \frac{n}{8}\right] \leq \exp(-\Omega(n)).$$

This follows from the expander Chernoff bound, as stated in Theorem 11. We can then prove Theorem 39.

Proof of Theorem 39. Let X be the indicator that $\varphi(\mathcal{D})$ fails to be $(\ell, \ell(1+\alpha))$-zero error list recoverable. Then

$$\begin{aligned}
\mathbb{E}[X] &\leq \sum_{\mathcal{C}' \subset \varphi(\mathcal{D}) : |\mathcal{C}'| \leq 10\sqrt{\ell/\gamma}} \Pr[|\{i : \varphi_i \in T(\mathcal{C}')\}| \geq \sigma n/4] && \text{(Lemma 41)} \\
&\leq \sum_{\mathcal{C}' \subset \varphi(\mathcal{D}) : |\mathcal{C}'| \leq 10\sqrt{\ell/\gamma}} \exp(-\Omega(n)) && \text{(Lemma 42)} \\
&\leq \exp\left(10\sqrt{\ell/\gamma}\log|\mathcal{D}| - \Omega(n)\right) \\
&\leq \exp(-\Omega(n))
\end{aligned}$$

where the second line follows by observing

$$|T(\mathcal{C}')| \leq d|\mathcal{C}'|^2 \leq 100d\ell/\gamma \leq \frac{m}{16}$$

where the first inequality follows from the distance of the code and the third follows from our bound on ℓ, and so $T(\mathcal{C}') \subset [m]$ satisfies the properties of Lemma 42. ◀

Proof of Theorem 38. Let $n := \lceil c'\varepsilon^{-1}\log|\mathcal{D}|\rceil$ for come constant $c' > 0$ such that the construct of Theorem 39 is satisfied (for parameters to be chosen later). We first show that the actual rate of $\varphi(\mathcal{D})$ is equal to the design rate with high probability. Analogously to the proof of Lemma 13, note that this event is equivalent to there existing $u \in \mathcal{D}$ such that $\varphi(u) = 0$. Fixing arbitrary $u \in \mathcal{D}$, let $T \subset [m]$ be the coordinates on which u is zero. Then $|T|/m \leq 1/q + \varepsilon^2 \leq .6$ by the absolute constraint on ε and that $q \geq 2$. Then

$$\Pr[\varphi(u) = 0] = (.6 + \lambda)^n = 2^{-\Omega(n)}$$

Thus the probability that all such codewords are not mapped to all zero indices is $|\mathcal{D}|2^{-n} \leq \exp(-\Omega(n))$, so with high probability the rate of $\varphi(\mathcal{D})$ is equal to the design rate of $\Omega(\varepsilon/\log q)$. Finally, choose $\ell = \alpha/4\varepsilon^2$ and $d = \lfloor m\varepsilon^2 \rfloor$ and applying Theorem 39 completes the proof. ◀

References

1 N. Alon, J. Bruck, J. Naor, M. Naor, and R. M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.

2 Omar Alrabiah, Venkatesan Guruswami, and Ray Li. Randomly punctured reed-solomon codes achieve list-decoding capacity over linear-sized fields. *CoRR*, abs/2304.09445, 2023. doi:10.48550/arXiv.2304.09445.

3 Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-ramanujan graphs using the zig-zag product. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 325–334. ACM, 2008. doi:10.1145/1374376.1374424.

4 Xue Chen, Kuan Cheng, Xin Li, and Songtao Mao. Random shortening of linear codes and application. *ECCC*, 2023(128), 2023.

5 Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2001. doi:10.1002/0471200611.

6 Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick. Good quantum LDPC codes with linear time decoders. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 905–918. ACM, 2023. doi:10.1145/3564246.3585101.

7 D. Gillman. A chernoff bound for random walks on expander graphs. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 680–691, 1993. doi:10.1109/SFCS.1993.366819.

8 Oded Goldreich. Three xor-lemmas – an exposition. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation – In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, volume 6650 of *Lecture Notes in Computer Science*, pages 248–272. Springer, 2011. doi:10.1007/978-3-642-22670-0_22.

9 Shouzhen Gu, Christopher A. Pattison, and Eugene Tang. An efficient decoder for a linear distance quantum LDPC code. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 919–932. ACM, 2023. doi:10.1145/3564246.3585169.

10 Zeyu Guo and Zihan Zhang. Randomly punctured reed-solomon codes achieve the list decoding capacity over polynomial-size alphabets. *CoRR*, abs/2304.01403, 2023. doi:10.48550/arXiv.2304.01403.

11 Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. *IEEE Trans. Inf. Theory*, 57(2):718–725, 2011. doi:10.1109/TIT.2010.2095170.

12 Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Trans. Inf. Theory*, 51(10):3393–3400, 2005. doi:10.1109/TIT.2005.855587.

13 Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear 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 9:1–9:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPICS.APPROX/RANDOM.2020.9.

14 Venkatesan Guruswami and Jonathan Mosheiff. Punctured low-bias codes behave like random linear codes. *CoRR*, abs/2109.11725, 2021. arXiv:2109.11725.

15 Venkatesan Guruswami and Atri Rudra. Soft decoding, dual BCH codes, and better list-decodable varepsilon-biased codes. *IEEE Trans. Inf. Theory*, 57(2):705–717, 2011. doi:10.1109/TIT.2010.2095193.

16 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43:439–561, 2006.

17 Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of ta-shma's codes via splittable regularity. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1527–1536. ACM, 2021. doi:10.1145/3406325.3451126.

18 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. doi:10.4230/LIPIcs.APPROX/RANDOM.2020.30.

19 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.

20 Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996. doi:10.1109/18.556667.

21 Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 238–251, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3055399.3055408.

22 Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. doi:10.1561/0400000010.

23 Victor Vasilievich Zyablov and Mark Semenovich Pinsker. List concatenated decoding. *Problemy Peredachi Informatsii*, 17(4):29–33, 1981.

A Omitted Proofs

Proof of Lemma 41. By assumption, for $i \in [n]$ there are subsets $A_i \subseteq [q]$ such that $|A_i| \leq \ell$ and, letting

$$\mathrm{BAD} := \{c \in \mathcal{D} : \varphi(c) \in \prod_{i \in [n]} A_i\},$$

we have $|\mathrm{BAD}| \geq \ell(1+\alpha)$. Let $\mathcal{C}' \subset \mathrm{BAD}$ be defined by randomly including each element of BAD with probability

$$p = \sqrt{\frac{2\ell}{\gamma}} \cdot \frac{(1+\alpha)}{|\mathrm{BAD}|}.$$

Note that

$$\mathbb{E}[|\mathcal{C}'|] = p|\mathrm{BAD}| \leq \sqrt{8\ell/\gamma}.$$

Now note that for every i, there are at least $\alpha\ell/2$ pairs $\{c_1, c_2\} \in \mathrm{BAD}$ with $c_1[\varphi_i] = c_2[\varphi_i]$, which holds as $|A_i| \leq \ell$ and $|\mathrm{BAD}| \geq (1+\alpha)\ell$. Thus, for every $i \in [n]$,

$$\Pr[\varphi_i \notin T(\mathcal{C}')] = (1-p^2)^{\alpha\ell/2} < 1/2$$

Thus,

$$\mathbb{E}[|\{i : \varphi_i \in T(\mathcal{C}')\}|] \geq n/4.$$

Finally, define the random variable

$$Y = \sum_{i \in [n]} \mathbb{I}[\varphi_i \in T(\mathcal{C}')] - \frac{n}{8} \frac{|\mathcal{C}'|}{\mathbb{E}[|\mathcal{C}'|]}.$$

We have $\mathbb{E}[Y] \geq n/4$, and hence there exists $\mathcal{C}'$ such that Y achieves its expectation, which can only occur when

$$|\{i : \varphi_i \in T(\mathcal{C}')\}| \geq n/4$$

and $|\mathcal{C}'| \leq 10\sqrt{\ell/\gamma}$, so we conclude. ◀

B Relationship of Zero-Error List-Recoverability Bounds to Unbalanced Expanders

We will require the definition of an unbalanced expander.

▶ **Definition 43** (Unbalanced Expander). *A (k, d, ε)-regular unbalanced expander is a bipartite graph on vertex set $V = L \cup R$, $|L| \geq |R|$, where the degree of every vertex in L is d, and for every $S \subseteq L$ such that $|S| = k$, we have that $|N(S)| \geq D|S|(1-\varepsilon)$.*

Further, we will require a procedure that turns a code into a graph.

▶ **Definition 44** (Bipartite Graph of a Code). *For a code $\mathcal{C} \subseteq [q]^n$, we denote by $G(\mathcal{C})$ the bipartite graph with vertex set $\mathcal{C} \cup ([n] \times [q])$. For an arbitrary $c = (c_1, \ldots c_n) \in \mathcal{C}$, we associate it with the neighbors $\{(1, c_1), \ldots, (n, c_n)\}$.*

In [18], the authors show the following result, by relating zero-error list-recoverability to expansion of a graph G:

▶ **Theorem 45** ([18]). *Let q, n be sufficiently large integers and $\alpha \in (0, 1), \varepsilon > q^{-1/2}$ be real numbers. For every code $\mathcal{D} \subseteq [q]^m$ with relative distance $1 - 1/q - \varepsilon^2$, there is a subset $S \subseteq [m]$ such that $|S| = O(\varepsilon m \log q)$ such that $G(\mathcal{D}_S)$ is a $\left(\alpha\varepsilon^{-2}, |S|, \alpha\right)$-unbalanced expander.*

[18] instantiate this result for degree d Reed-Solomon codes, with $m = q$, $\varepsilon = (d/q)^{-1/2}$. Thus, $n = \widetilde{O}(\sqrt{q})$. Note that in this setting, we are only guaranteed the existence among $\binom{n^2}{n} = 2^{O(n \log n)}$ possible choices for the punctured set. With our construction, we can again use degree d Reed-Solomon Codes, and recover that there exists such an unbalanced-expander among only $2^{O(n)}$ possible choices for the punctured set.