Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 139 (2021)

Punctured Low-Bias Codes Behave Like Random Linear Codes

Venkatesan Guruswami
University of California, Berkeley
venkatg@berkeley.edu

Jonathan Mosheiff
Carnegie Mellon University
jmosheif@cs.cmu.edu

Abstract

Random linear codes are a workhorse in coding theory, and are used to show the existence of codes with the best known or even near-optimal trade-offs in many noise models. However, they have little structure besides linearity, and are not amenable to tractable error-correction algorithms.

In this work, we prove a general derandomization result applicable to random linear codes. Namely, in settings where the coding-theoretic property of interest is "local" (in the sense of forbidding certain bad configurations involving few vectors—code distance and list-decodability being notable examples), one can replace random linear codes (RLCs) with a significantly derandomized variant with essentially no loss in parameters. Specifically, instead of randomly sampling coordinates of the (long) Hadamard code (which is an equivalent way to describe RLCs), one can randomly sample coordinates of any code with low bias. Over large alphabets, the low bias requirement can be weakened to just large distance. Furthermore, large distance suffices even with a small alphabet in order to match the current best known bounds for RLC list-decodability.

In particular, by virtue of our result, all current (and future) achievability bounds for list-decodability of random linear codes extend *automatically* to random puncturings of any low-bias (or large alphabet) "mother" code. We also show that our punctured codes emulate the behavior of RLCs on stochastic channels, thus giving a derandomization of RLCs in the context of achieving Shannon capacity as well. Thus, we have a randomness-efficient way to sample codes achieving capacity in both worst-case and stochastic settings that can further inherit algebraic or other algorithmically useful structural properties of the mother code.

Research supported in part by NSF grants CCF-1814603 and CCF-1908125, and a Simons Investigator Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ISSN 1433-8092

Contents

1 Introduction

Random linear codes (RLCs) are ubiquitous in coding theory, serving as a fundamental building block in the construction of codes since the works of Shannon. RLCs are extensively studied, and known to enjoy excellent combinatorial properties. In particular, they achieve Shannon capacity, the Gilbert-Varshamov rate vs. distance trade-off, and are list-decodable up to capacity.

An RLC of length n and rate[1] $0 < R < 1$ over alphabet $\mathbb{F}_q$ is the row span of a matrix A, sampled uniformly from $\mathbb{F}_q^{k\times n}$, where $k = Rn$. Here, $\mathbb{F}_q$ denotes the finite field of order q. Equivalently, one can think of A as a matrix with n random independent columns, each sampled uniformly from $\mathbb{F}_q^k$.

In this work we study derandomizations of RLCs. Specifically, we consider a code $\mathcal{C}$ generated by a matrix whose columns are independently sampled from some distribution μ, where the support of μ is a much smaller set than $\mathbb{F}_q^k$ (possibly exponentially so). We are able to prove that, under fairly modest and general assumptions on μ, this random code $\mathcal{C}$ is similar to an RLC with respect to *local properties*, a notion which will soon be explained. A special case of this result is that $\mathcal{C}$ is very likely to achieve *list-decoding capacity*, since RLCs are known to do so (in fact, the convergence of RLCs to list-decoding capacity has been extensively studied). Independently of this result, we also show that, similarly to an RLC, $\mathcal{C}$ is likely to achieve capacity with regard to every memoryless additive-noise channel.

To describe our result more formally, we turn to the notion of punctured codes. Puncturing is a basic operation by which new codes are constructed from existing ones. A puncturing of a code $\mathcal{D} \subseteq \mathbb{F}_q^m$ is a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ (where we usually think of m as being much large than n) whose coordinates are taken from those of $\mathcal{D}$. More precisely, $\mathcal{C}$ is a puncturing of $\mathcal{D}$ if $\mathcal{C} = \{(x_{i_1}, \dots, x_{i_n}) \mid x = (x_1 \dots x_m) \in \mathcal{D}\}$, for some integers $i_1, \dots, i_n \in [m]$. We sometimes refer to $\mathcal{D}$ as the *mother code*. When $i_1, \dots, i_n$ are sampled uniformly and independently from $[m]$, we say that $\mathcal{C}$ is a random n-puncturing of $\mathcal{D}$. Puncturing generally increases the rate of a code, and we often take $\mathcal{D}$ to be a code of rate approaching 0, while the rate of $\mathcal{C}$ is a constant in $(0, 1)$.

An equivalent way to describe an RLC of length n and rate R is as a random puncturing of the *Hadamard code* of length q^k (where $k = Rn$). The Hadamard code has very poor rate. Consequently, to obtain a punctured code of length n and constant rate from a Hadamard mother code, the mother code must be taken to have length exponential in n. Our motivating question is whether, in this construction, one can replace the Hadamard code by a shorter mother code and still obtain a punctured code with the excellent combinatorial properties of an RLC. In Theorem A, below, we give this question a strong positive answer.

Of the many special properties of a Hadamard code, the property crucial for that code's excellent performance as a mother code turns out be its optimal bias. Focusing for simplicity on the case $q = 2$, the bias of a binary code $\mathcal{D}$ of length m is defined as

$$\max\left\{\frac{|\text{number of ones in } x - \text{number of zeros in } x|}{m} \mid x \in \mathcal{D} \setminus \{0\}\right\} \ .$$

The bias of a Hadamard code is the smallest possible, namely, zero. The following informal result can thus be seen as an extension of the statement "*A random puncturing of a Hadamard code is an RLC*".

[1] More accurately, R is the design rate of the code. The actual rate, $\frac{\operatorname{rank} A}{n}$, equals R if and only if A is of full rank—an event which holds with very high probability. See Lemma 2.3 for more details.

Theorem A (Main result about puncturing of low-bias codes)**.** *Let $\mathcal{D}$ be a linear code with small bias and let $\mathcal{C}$ be a random puncturing of $\mathcal{D}$. Then, $\mathcal{C}$ is likely to have any* monotone-decreasing local property *that is typically satisfied by an RLC of similar rate.*

Theorem B, below, gives broad conditions under which the hypothesis of Theorem A can be further relaxed, from requiring low bias to just large distance of the mother code $\mathcal{D}$.

Informally, a code property $\mathcal{P}$ is monotone-decreasing and local if, whenever a code $\mathcal{C}$ does not satisfy $\mathcal{P}$, there exists a small *"bad set"* of codewords in $\mathcal{C}$ that bears witness to this fact. By monotone-decreasing we mean that adding codewords to the code can only make the property harder to satisfy. A code that has the same monotone-decreasing local properties as an RLC is said to be locally similar to an RLC. Local properties of codes were originally introduced in [MRRSW20], with the motivation of studying the list-decodability of Gallagher's LDPC codes. We turn to explain this connection.

List-decodability. In the model of list-decoding, the goal is to decode beyond the unique-decoding radius. A code is said to be (combinatorially) list-decodable up to radius ρ ($0 < \rho < 1 - \frac{1}{q}$) if every Hamming ball of radius ρn in $\mathbb{F}_q^n$ has intersection of size at most L (where L is small, say constant in n) with the code. Being list-decodable is a monotone-decreasing local property. Indeed, to show that a code is *not* list-decodable, it suffices to point to a set of $L + 1$ codewords that all reside within the same radius ρ Hamming ball. A list-decodable code, when accompanied by a decoding algorithm, will allow the correction of a ρ fraction of errors up to some bounded ambiguity in the worst-case. We refer the reader to [Gur06] for a detailed discussion of the motivation, usefulness, and potential of the list-decoding model.

The list-decodability of RLCs has been extensively studied in previous works [ZP81; GHK11; CGV13; Woo13; RW14b; RW18; LW21; GLMRSW20]. The paper [ZP81] already establishes that RLCs are list-decodable up to capacity. Namely, for any fixed R and ρ satisfying $R < 1 - h_q(\rho)$, an [2] RLC of rate R is almost surely list-decodable up to radius ρ with constant list-size $L = L(R, \rho, q)$. The focus of the later works is pinpointing the exact dependence of the list-size L on the other parameters.

Since list-decodability is a monotone-decreasing local property, the aforementioned results about list-deocdability of an RLC also apply to any code that is locally similar to an RLC. Therefore, Theorem A yields a powerful reduction, allowing us to apply these results about RLCs to the punctured code $\mathcal{C}$. Moreover, any positive RLC list-decoding bound discovered in the future would also immediately apply to $\mathcal{C}$. The latter may be relevant since there are still some gaps in our knowledge of RLC list-decodability, especially in the low-rate non-binary regime.

In addition to list-decodability, Theorem A also applies to other local properties, such as list-recoverability, and its special case, the perfect hashing property. Hence, Theorem A immediately yields a positive result (see Section 6.3) about the list-recoverability of $\mathcal{C}$, via reduction to established results about the list-recoverability of an RLC.

Theorem A and RLC derandomization. Perhaps more importantly than its specialization to any concrete local property, Theorem A is a statement about the robustness of the mechanism by which an RLC is generated: The theorem says that it is possible to choose a code $\mathcal{D}$ that is radically different from a Hadamard code, randomly puncture it, and end up with a code that, in a local view, has the same desirable properties as an RLC.

[2]The q-ary entropy function h_q is formally defined in Section 4.5

Theorem A allows us great flexibility in choosing the mother code $\mathcal{D}$. While the only structural property of an RLC is its linearity, the punctured code $\mathcal{C}$ of Theorem A can be made to have additional structure via certain choices of the mother code. For example, $\mathcal{D}$ can be taken to be a dual-BCH code, namely, a code in which every codeword encodes a low-degree polynomial over $\mathbb{F}_{2^\ell}$ ($\ell \in \mathbb{N}$) by the trace of its evaluations over $\mathbb{F}_{2^\ell}$. In a random puncturing $\mathcal{C}$ of $\mathcal{D}$, the codewords correspond to evaluations over a random subset of $\mathbb{F}_{2^\ell}$. It is well-known (see [GR11]) that dual-BCH codes have small bias, so Theorem A applies. Hence, this code $\mathcal{C}$ enjoys both an algebraic structure and local similarity to an RLC.

Enforcing a structure on $\mathcal{C}$ has potential algorithmic advantages. For example, recall that the local similarity of $\mathcal{C}$ to an RLC only guarantees with high probability the combinatorial property of list-decodability, but not the existence of an efficient list-decoding algorithm for $\mathcal{C}$. Indeed, for the RLC ensemble itself, it is very likely that no efficient list-decoding algorithm exists for any constant radius ρ. Hopefully, by choosing a suitable structure for $\mathcal{C}$, one may be obtain a code which is not only combinatorially list-decodable up to capacity, but also amenable to efficient list-decoding algorithms.

Another application of Theorem A is a direct derandomization of RLCs. Utilizing constructions such as [ABNNR92; Ta-17], $\mathcal{D}$ can be taken to be an explicit low-bias linear code of length $O(n)$ (where n is the desired length of $\mathcal{C}$). With such a short mother code, only $O(n)$ random bits are needed to construct the punctured code $\mathcal{C}$, which is locally similar to an RLC. This is in contrast to the $O(n^2)$ random bits needed to construct an actual RLC of the same length. While methods to sample linear codes with $O(n)$ randomness were known in some settings, the approach and analysis was tailored to the specific setting (e.g, via random Toeplitz matrices for the Gilbert-Varshamov trade-off). In contrast, our approach applies uniformly for all local properties. For list-decoding, this gives the first linear randomness method to sample codes that achieve the trade-offs of RLCs.

The idea of relating the local properties of a more structured code $\mathcal{C}$ to those of an RLC already figures prominently in the previously mentioned [MRRSW20], where a *Gallagher LDPC Code* is cast in the role of $\mathcal{C}$. While our methods in the present work are very different, our proof of Theorem A does use the framework of [MRRSW20], as well as the *RLC Threshold Theorem* [MRRSW20, Thm. 2.8] proven there.

Replacing low bias by large distance. A linear code of low bias necessarily has large minimal distance. For example, in the binary case, the normalized Hamming weight of a codeword $x \in \mathcal{D}\setminus\{0\}$ is at least $\frac{1-\mathrm{bias}(\mathcal{D})}{2}$. Theorem B extends Theorem A by considering the case in which the mother code $\mathcal{D}$ has large minimal distance, but not necessarily low bias. For large enough alphabet size q, large minimal distance of $\mathcal{D}$ is enough to guarantee the conclusion of Theorem A. For small q, while we do not have such a general result, we are able to use specific characteristics of the the property of list-decodability to prove that $\mathcal{C}$ is, with high probability, list-decodable up to capacity.

Theorem B (Main result about puncturing of large-distance codes)**.** *Let $\mathcal{D}$ be a linear code whose minimal distance is near $1-\frac{1}{q}$, and let $\mathcal{C}$ be a random puncturing of $\mathcal{D}$ of rate R. Then:*

1. *Let $\mathcal{P}$ be a monotone-decreasing local property that is likely satisfied by an RLC of rate $R - 2\log_q 2$. Then, $\mathcal{C}$ is likely to satisfy $\mathcal{P}$.*

2. *Regardless of its alphabet size, $\mathcal{C}$ is likely to be list-decodable and list-recoverable up to capacity (similarly to an RLC of rate R).*

Achieving Shannon capacity with punctured codes. Theorems A and B are most relevant when considering the performance of $\mathcal{C}$ as an error correcting code in the *worst-case error* model. To complete the picture, we also prove the following result, dealing with the *random error* model. It is well-known that RLCs achieve capacity with regard to any memoryless additive noise channel. The following informal theorem generalizes this statement to our punctured code $\mathcal{C}$.

Theorem C (Puncturing of large-distance codes in stochastic channels)**.** *Let $\mathcal{D}$ be a linear code whose minimal distance is near $1-\frac{1}{q}$, and let $\mathcal{C}$ be a random puncturing of $\mathcal{D}$ with rate R. Let $\mathcal{N}$ be a memoryless additive noise channel with capacity at least $R+\varepsilon$. Then, it is possible to reliably communicate across $\mathcal{N}$ using $\mathcal{C}$.*

1.1 Previous work

Randomly punctured codes have recently gotten a lot of attention, motivated by the study of Reed-Solomon (RS) codes [RW14a; ST20; GLSTW20; FKS20; GST21]. The RS code of dimension $1 \le k \le q$ over the set $S \subseteq \mathbb{F}_q$ is defined as

$$\mathrm{RS}_{\mathbb{F}_q}(S;k) = \{(f(\alpha_1),\dots,f(\alpha_n)) \mid f \in \mathbb{F}_q[x],\ \deg(f) < k\} \enspace .$$

The length of the code is $n = |S|$. A classical algorithm [GS99] can efficiently list-decode an RS code up to the Johnson radius $1-\sqrt{R}-o(1)$.

An important open question is whether efficient list-decoding of some RS codes is possible all the way up to the capacity radius $1-R-o(1)$. A necessary condition for such an algorithmic result is to have RS codes which are *combinatorially* list-decodable up to capacity, and such codes are yet unknown. In fact, even the existence of RS codes that are combinatorially decodable beyond the Johnson bound was only recently proven [ST20].

Our freedom in constructing an RS code lies mainly in the choice of the evaluation set S. A natural choice is to take S to be a uniformly random subset of $\mathbb{F}_q$ of the desired size n. When S is sampled this way, the code $\mathrm{RS}_{\mathbb{F}_q}(S;k)$ is essentially a random puncturing of the full RS code $\mathrm{RS}_{\mathbb{F}_q}(\mathbb{F}_q;k)$. All of the aforementioned works [RW14a; ST20; GLSTW20; FKS20; GST21] take this viewpoint. In particular, in [RW14a; FKS20; GST21], the main results about list-decodability of RS codes are all immediate special cases of more general results about randomly punctured codes. Our Theorems A and B are similar in spirit to the latter. We note that our work is the *first* in this line to yield punctured codes that achieve list-decoding capacity, and we do so *for every choice of rate.* The previous works showed trade-offs that were bounded away from list-decoding capacity for all rates.

1.2 On an error in a previous version of this paper regarding RS codes

The full RS code $\mathrm{RS}_{\mathbb{F}_q}(\mathbb{F}_q;k)$ for small k has near-optimal distance, and thus it would seem that one could apply Theorem B to it, proving that an RS code over a random evaluation set is likely to be locally-similar to an RLC, and, in particular, list-decodable up to capacity. Unfortunately, the *large distance* requirement of the theorem becomes stricter as the alphabet q grows, and impossible to achieve whenever q is larger than n (see Theorem 4). An earlier version of the current paper erroneously claimed to overcome this difficulty by passing to trace codes over a smaller alphabet, and use it to deduce the list-decodability of certain RS codes. Later, prompted by a question from Zeyu Guo, we have found an error in the proof and retracted this claim.

2 Main Results

Before stating our main results, we formally define some of the relevant notions.

Definition 2.1 (Random puncturing)**.** *Fix some prime power q. Let $m, n \in \mathbb{N}$. An $(m \to n)$* puncturing map *is a function $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ of the form $\varphi(u = (u_1, \dots, u_m)) = (u_{i_1}, u_{i_2}, \cdots, u_{i_n})$ for some $i_1, \dots, i_n \in [m]$. If $i_1, \dots, i_n$ are sampled i.i.d. and uniformly from $[m]$, we say that φ is a* random $(m \to n)$ puncturing map*.*

A random n-puncturing *of a code $\mathcal{D} \subseteq \mathbb{F}_q^m$ is a random code $\mathcal{C} = \varphi(\mathcal{D}) = \{\varphi(u) \mid u \in \mathcal{D}\}$, where $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ is a random puncturing map. The* design rate *of $\mathcal{C}$ is $\frac{\log_q |\mathcal{D}|}{n}$.*

Definition 2.2. *Let $\mathcal{D} \subseteq \mathbb{F}_q^m$, where q is a power of some prime p, be a linear code and let $\eta > 0$.*

1. *If every $u \in \mathcal{D} \setminus \{0\}$ has $\mathrm{wt}(u) \geq \frac{(q-1)(1-\eta)}{q}$, we say that $\mathcal{D}$ has η-*optimal distance*. Here, $\mathrm{wt}(u) = \frac{|\{i \in [m] | u_i \neq 0\}|}{m}$ denotes the* normalized Hamming weight *of $u \in \mathbb{F}_q^m$.*
2. *A vector $u \in \mathbb{F}_q^m$ is said to be η-*biased *if $\left|\sum_{i=1}^m \omega^{\mathrm{tr}(a \cdot u_i)}\right| \leq m\eta$ for all $a \in \mathbb{F}_q^*$. Here, $\omega = e^{\frac{2\pi i}{p}}$ and $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ is the field trace map (see Section 4.4). The code $\mathcal{D}$ is said to be η-*biased *if every $u \in \mathcal{D} \setminus \{0\}$ is η-biased.*

As shown in Lemma 4.13, an η-biased code also has η-optimal distance, so the former is a stronger notion. For intuition, note that in the binary case η-bias implies $\frac{1-\eta}{2} \leq \mathrm{wt}(u) \leq \frac{1+\eta}{2}$ for any $u \in \mathcal{D} \setminus \{0\}$, whereas η-optimal distance only implies the lower bound on $\mathrm{wt}(u)$.

It may be simpler for the reader to focus on the case where q is a prime, i.e., $q = p$. In this case, tr is merely the identity map.

If $\mathcal{C}$ is a random n-puncturing of a code $\mathcal{D}$, the rate of $\mathcal{C}$ is clearly bounded from above by its design rate. The following lemma shows that when $\mathcal{D}$ is of almost optimal distance, these two terms are very likely to coincide. In light of this lemma, we blur the distinction between design rate and actual rate.

Lemma 2.3 (Actual rate equals design rate whp)**.** *Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code of η-optimal distance, and let $\mathcal{C}$ be a length-n random puncturing of $\mathcal{D}$, of design rate $R \leq 1 - \log_q(1 + \eta q) - \varepsilon$. Then, with probability at least $1 - q^{-n\varepsilon}$, the rate of $\mathcal{C}$ is equal to its design rate.*

Proof. The rate of $\mathcal{C}$ is smaller than R if and only if there exists a non-zero word $u \in \mathcal{D}$ such that only coordinates $i \in [m]$ for which $u_i = 0$ are sampled for inclusion in $\mathcal{C}$. For a given u, this happens with probability

$$(1 - \mathrm{wt}(u))^n \leq \left(\frac{1}{q} + \frac{q-1}{q}\eta\right)^n \leq \left(\frac{1}{q} + \eta\right)^n = q^{-n(1-\log(1+q\eta))} .$$

The claim follows by a union bound over the non-zero words of $\mathcal{D}$, of which there are $q^{Rn} - 1$, and the assumed upper bound on R. □

Definition 2.4 (clustered sets and list-decodability)**.** *Fix $\rho \in [0, 1]$. A set of vectors $W \subseteq \mathbb{F}_q^n$ is called ρ-*clustered *if there exists some $z \in \mathbb{F}_q^n$ such that $\mathrm{wt}(u - z) \leq \rho$ for each $u \in W$. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is said to be (ρ, L)-*list-decodable *if it does not contain any ρ-clustered set of codewords of size $L + 1$.*

2.1 A framework for studying properties of codes

In order to formulate our results, we need to recall some of the framework for studying *local and row-symmetric properties of linear*[3] *codes*, established in [MRRSW20; GMRSW21].[4]

A property $\mathcal{P}$ of length-n linear codes over $\mathbb{F}_q$ is a collection of linear codes in $\mathbb{F}_q^n$. A linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ such that $\mathcal{C} \in \mathcal{P}$ is said to satisfy $\mathcal{P}$. If $\mathcal{P}$ is upwards closed with regard to containment, it is said to be monotone-increasing.[5]

Definition 2.5 (Local and row-symmetric properties)**.** *Let $\mathcal{P}$ be a monotone-increasing property of linear codes in $\mathbb{F}_q^n$. We define the following notions.*

1. *Fix $b \in \mathbb{N}$. Suppose that 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 $\mathbb{F}_q^n$ with $|B| \le b$, and such that*

$$\mathcal{C} \text{ satisfies } \mathcal{P} \quad \Longleftrightarrow \quad \exists B \in \mathcal{B}_\mathcal{P} \ \ B \subseteq \mathcal{C} \ .$$

   Then, we say that $\mathcal{P}$ is a b-local property.

2. *Suppose that, whenever a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ satisfies $\mathcal{P}$ and π is a permutation on $\{1, \dots, n\}$, the code $\{\pi x \mid x \in \mathcal{C}\}$ also satisfies $\mathcal{P}$. We then say that $\mathcal{P}$ is row-symmetric*[6]*. Here, πx is the vector obtained by permuting the coordinates of x according to π.*

The following is immediate from the definition of (ρ, L)-list-decodability.

Observation 2.6. *Let q be a prime power and $n \in \mathbb{N}$. Fix $\rho \in (0,1)$, $L \in \mathbb{N}$, and let $\mathcal{P}$ be the monotone-increasing property consisting of codes in $\mathbb{F}_q^n$ that are **not** (ρ, L)-list-decodable. Then, $\mathcal{P}$ is $(L+1)$-local and row-symmetric.*

Let $\mathcal{P}$ be a monotone-increasing property over $\mathbb{F}_q^n$. Suppose that $\mathcal{P}$ is nonempty, namely, that it is satisfied by the complete code $\mathbb{F}_q^n$. We denote its threshold by

$$\mathrm{RLC}(\mathcal{P}) = \min \left\{ R \in [0,1] \mid \Pr\left[C_{\mathrm{RLC}}^{n,q}(R) \text{ satisfies } \mathcal{P} \right] \ge \frac{1}{2} \right\} \ ,$$

where $C_{\mathrm{RLC}}^{n,q}(R)$ is a random linear code of rate R in $\mathbb{F}_q^n$.

This terminology is motivated by the following theorem, which states that the probability of an RLC of rate R satisfying a local, row-symmetric and monotone-increasing property $\mathcal{P}$, as a function of R, rapidly climbs near the threshold from $o(1)$ to $1 - o(1)$.

[3] This framework makes sense for linear as well as non-linear codes. In this work we restrict ourselves to the linear case.

[4] The notion of a *local property* from [MRRSW20] was later refined and split into two parts in [GMRSW21], where it appears as a *row-symmetric* and *local* property. We follow the latter convention.

[5] While Section 1 discussed monotoned-decreasing properties, it will henceforth be more convenient to deal with monotone-increasing properties. Note that the negation of a monotone-decreasing property is monotone-increasing. Hence, the statement “the code $\mathcal{C}$ satisfies every monotone-decreasing local property typically satisfied by an RLC” is equivalent to “every monotone-increasing local property typically *not* satisfied by an RLC is also *not* satisfied by $\mathcal{C}$.

[6] The reason for this terminology will be made clear in Observation 6.2.

Theorem 2.7 (Thresholds for local and row-symmetric properties [MRRSW20, Thm. 2.8])**.** [7] *Let $\mathcal{P} \subseteq \mathbb{F}_q^n$ be a random linear code of radius R and Let $\mathcal{P}$ be a monotone-increasing, b-local and row-symmetric property over $\mathbb{F}_q^n$, where $\frac{n}{\log_q n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. The following now holds for every $\varepsilon > 0$.*

1. *If $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon$ then*
$$\Pr\left[\mathcal{C} \textit{ satisfies } \mathcal{P}\right] \leq q^{-(\varepsilon - o_{n\to\infty}(1))n} \quad .$$

2. *If $R \geq \mathrm{RLC}(\mathcal{P}) + \varepsilon$ then*
$$\Pr\left[\mathcal{C} \textit{ satisfies } \mathcal{P}\right] \geq 1 - q^{-(\varepsilon - o_{n\to\infty}(1))n} \quad .$$

2.2 Theorem A: Randomly punctured low-bias codes

Theorem 1, below, is a formal statement of Theorem A.

Theorem 1 (Puncturings of low-bias linear codes are locally similar to random linear codes)**.** *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 $\frac{n}{\log n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon$ for some $\varepsilon > 0$. Suppose that $\mathcal{D}$ is $\left(\frac{\varepsilon b \ln q}{q^b}\right)$-biased. Then,*
$$\Pr\left[\mathcal{C} \textit{ satisfies } \mathcal{P}\right] \leq q^{-(\varepsilon - o_{n\to\infty}(1))n} \quad .$$

2.3 Applications of Theorem 1 for list-decodability

In our discussion of list-decoding capacity in Section 1 we treated the rate R as fixed, and the list-decoding capacity radius ρ as a function of R and the field size. Henceforth we will prefer to think of R as depending on some fixed ρ.

The *List-Decoding Capacity Theorem* [GRS, Thm. 7.4.1] states that the *optimal rate* for radius ρ list-decoding over the field $\mathbb{F}_q$ is $R^* = 1 - h_q(\rho)$, where
$$h_q(\rho) = -\rho \log_q \rho - (1-\rho)\log_q(1-\rho) + \rho \log_q(q-1) \tag{1}$$
is the q-ary entropy function (see Section 4.5). In other words, there exist infinite families of codes of rate $R^* - \varepsilon$ that are list-decodable up to radius ρ with list-size independent of the block length n, but no such families exist for rate $R^* + \varepsilon$.

As mentioned in Section 1, the list-decodability of RLCs has been extensively studied. Positive and negative results of this sort can be stated, respectively, as lower and upper bounds on $\mathrm{RLC}\left(\mathrm{LD}_{\rho,L,q}\right)$, where $\mathrm{LD}_{\rho,L,q}$ is the monotone-increasing property of a code over $\mathbb{F}_q$ *not* being (ρ, L)-list-decodable. For example, the main result of [GHK11] can be stated as
$$\mathrm{RLC}\left(\mathrm{LD}_{\rho,L,q}\right) \geq 1 - h_q(\rho) - O_{\rho,q}\left(\frac{1}{L}\right) \tag{2}$$

[7] The theorem as stated in [MRRSW20] deals only with the regime of constant q and b. The current statement, which allows q and b to depend on n, follows by inspecting the proof in [MRRSW20].

for all ρ, q and L. In the binary regime, the main result of [LW21] together with a negative result from [GLMRSW20] yield the very precise bound

$$1 - h_2(\rho) \cdot \frac{L-1}{L-2} \le \mathrm{RLC}\left(\mathrm{LD}_{\rho,L,2}\right) \le 1 - h_2(\rho)\frac{L+1}{L} \tag{3}$$

for all ρ and $L \ge 3$. By Observation 2.6, $\mathrm{LD}_{\rho,L,q}$ is $(L+1)$-local and row-symmetric, so Theorem 1 applies to it. Plugging in Eqs. (2) and (3) yields the following corollary.

Corollary 2 (Puncturings of certain linear codes are locally similar to random linear codes)**.** *Fix a prime power q, $\rho \in \left(0, 1 - \frac{1}{q}\right)$, $L \in \mathbb{N}$ and $\varepsilon > 0$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be an $\left(\frac{\varepsilon(L+1)\ln q}{q^{L+1}}\right)$-biased linear code . Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate R, where $\frac{n}{\log n} \ge \omega_{n\to\infty}\left(q^{2(L+1)}\right)$. Then,*

1. *If $R < 1 - h_q(\rho) - \frac{C_{\rho,q}}{L} - \varepsilon$ then $\mathcal{C}$ is (ρ, L)-list-decodable with probability $1 - q^{-(\varepsilon - o_{n\to\infty}(1))n}$. Here, $C_{\rho,q}$ is a constant depending on ρ and q.*
2. *If $q = 2$, $L \ge 3$ and $R < 1 - h_2(\rho) \cdot \frac{L-1}{L-2} - \varepsilon$ then $\mathcal{C}$ is (ρ, L)-list-decodable with probability $1 - 2^{-(\varepsilon - o_{n\to\infty}(1))n}$.*

Other positive results about RLC list-decodability (e.g., [Woo13]) can be similarly used to obtain bounds on the list-decodability of randomly punctured codes.

List-recoverability is another property of interest to which Theorem 1 applies, similarly allowing us to reduce from known results about RLCs. See Section 6.3 for more details.

2.4 Derandomization of RLCs

As discussed in Section 1, Theorem 1 can be invoked to derandomize RLCs by casting a code of short block length and low bias in the role of the mother code $\mathcal{D}$. One result that can be achieved via this method is the following theorem. For simplicity, we focus on the binary case.

Theorem 3 (Codes locally similar to an RLC with linear randomness)**.** *There exists a randomized algorithm that, given $b \in \mathbb{N}$, $\varepsilon > 0$, $R^* \in [\varepsilon, 1]$ and $n \in \mathbb{N}$, where $\frac{n}{\log_2 n} \ge \omega_{n\to\infty}\left(2^{2b}\right)$ and $n \ge \omega_{n\to\infty}(1/\varepsilon)$, samples a generating matrix for a linear code $\mathcal{C} \subseteq \mathbb{F}_2^n$ of rate $R = R^* - \varepsilon$ such that*

$$\Pr\left[\mathcal{C} \text{ satisfies some property } \mathcal{P} \in \mathcal{K}\right] \le 2^{-\Omega(\varepsilon n)}. \tag{4}$$

Here, $\mathcal{K}$ is the family of all monotone-increasing, b-local and row-symmetric properties $\mathcal{P}$ over $\mathbb{F}_2^n$ for which the threshold $\mathrm{RLC}(\mathcal{P})$ is at least R^*. This algorithm uses $O\left(n\left(b + \log_2 \frac{1}{\varepsilon}\right)\right)$ random bits, and works in time polynomial in n.

2.5 Theorem B: Randomly punctured codes of near-optimal distance

Theorems 4 and 5 are detailed version of, respectively, the two parts of Theorem B. These theorems extend Theorem 1, in certain scenarios, to the case where the mother code $\mathcal{D}$ has near-optimal distance but not necessarily low bias. Theorem 4 states that the conclusion of Theorem 1 is still valid, provided that the alphabet q is large enough, and the property $\mathcal{P}$ is scalar-invariant—a new definition given here.

Definition 2.8 (Scalar invariant property)**.** *Let $\mathcal{P}$ be a property of codes in $\mathbb{F}_q^n$. Suppose that, for every code $\mathcal{C}$ satisfying $\mathcal{P}$ and for every diagonal full-rank matrix $\Lambda \in \mathbb{F}_q^{n\times n}$, the code $\Lambda\mathcal{C} := \{\Lambda u \mid u \in \mathcal{C}\}$ also satisfies $\mathcal{P}$. Then, $\mathcal{P}$ is said to be* scalar-invariant.

It is not hard to see that (ρ, L)-list-decodability (as well as its negation, the property of *not* being (ρ, L)-list-decodable) is a scalar-invariant property for all ρ and L. The same holds for list-recoverability (see Section 6.3).

Theorem 4 (Puncturings of near-optimal distance linear codes with large alphabet are locally similar to random linear codes)**.** *Let q be a prime power, and let $\mathcal{P}$ be a monotone-increasing, row-symmetric, b-local and scalar-invariant property over $\mathbb{F}_q^n$, where $\frac{n}{\log n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code of q^{-b}-optimal distance. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon - 2\log_q 2$ for some $\varepsilon > 0$. Then,*

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

Theorem 5 deals specifically with list-decodability rather than a general code property. The theorem exploits certain characteristics of the proof of [GHK11] (see Eq. (2)), and shows that this specific result about list-decodability of RLCs can be applied to our randomly-punctured code $\mathcal{C}$ as long as the mother code has near-optimal distance, regardless of the alphabet size. The relevant characteristics of [GHK11] are discussed in Remark 6.14.

Theorem 5 (A puncturing of a near-optimal distance code is whp list-decodable up to capacity)**.** *Fix a prime power q. Let $L, n \in \mathbb{N}$ and $0 < \rho < \frac{q-1}{q}$, such that $\frac{n}{\log_q n} \geq \omega_{n\to\infty}\left(q^{L+1}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code with η-optimal distance, where $\eta = q^{-L+1}$. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate R, where $R \leq 1 - h_q(\rho) - \frac{K}{L}$ for some constant $K = K_{\rho,q}$. Then,*

$$\Pr\left[\mathcal{C} \text{ is } (\rho, L)\text{-list-decodable}\right] \geq 1 - q^{-\Omega(n)} \quad .$$

Furthermore, one can take

$$K_{\rho,q} \leq \exp\left(O\left(\frac{(\log q)^2}{\min\left\{(1-1/q-\rho)^2, \rho\right\}}\right)\right) \tag{5}$$

and, in particular, $K_{\rho,q} \leq \mathrm{poly}(q)$ whenever ρ is bounded away from 0 and $1 - \frac{1}{q}$.

2.6 Theorem C: Randomly punctured codes in the stochastic error model

Definition 2.9 (Additive noise channel)**.** *Let ν be a distribution over $\mathbb{F}_q$. The ν-*memoryless additive noise channel *with distribution ν takes as input a vector $x \in \mathbb{F}_q^n$ and outputs the vector $x + z$, where $z \in \mathbb{F}_q^n$ has entries independently sampled from ν.*

For $z \in \mathbb{F}_q^n$, we write $\nu(z) = \prod_{i=1}^n \nu(z_i)$ for the probability of z under the product distribution ν^n.

The capacity of the ν-memoryless additive noise channel is $1 - H_q(\nu)$. Here, $H_q(\nu)$ is the base-q entropy

$$H_q(\nu) = -\sum_{x\in\mathrm{supp}(\nu)} \nu(x)\log_q \nu(x) \quad .$$

The maximum likelihood decoder under uniform prior (MLDU) for a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ with regard to the ν-memoryless additive noise channel receives a word $y \in \mathbb{F}_q^n$ and returns a codeword $x \in \mathcal{C}$ for which $\nu(y-x)$ is maximal. In other words, x maximizes the likelihood of the channel outputting y given input x.

It is well-known that an RLC with MLDU decoding achieves capacity with regard to any memoryless additive noise channel. Theorem C, stated here formally as Theorem 6, generalizes this fact to random puncturings of low-bias codes.

Theorem 6 (A puncturing of a low-bias code achieves capacity with regard to memoryless additive noise)**.** *Fix a prime power q, a distribution ν over $\mathbb{F}_q$ and $0 < \varepsilon < 1$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be an $\frac{\varepsilon}{3(q-1)}$-biased linear code. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq 1 - H_q(\nu) - \varepsilon$. Then, with probability $1 - q^{-\Omega_\nu(\varepsilon n)}$, it holds for all $x \in \mathcal{C}$ that*

$$\Pr_{z \sim \nu^n}[\text{the MLDU outputs } x \text{ on input } x+z] \geq 1 - 2q^{-c_\nu \varepsilon^2 n} \quad , \tag{6}$$

for some $c_\nu > 0$ that depends only on ν.

2.7 Organization

The rest of the paper is organized as follows. In Section 3 we survey our techniques by sketching a proof for a weaker version of Corollary 2. Section 4 establishes some general definitions and lemmas used in the main proofs. In Section 5 we prove Theorem 5 about list-decodability of randomly punctured codes, based on the result of [GHK11]. In Section 6 we give more details on the code property framework, and prove Theorems 1 and 4 about punctured codes that are locally-similar to an RLC. In Section 7 we prove Theorem 3, dealing with derandomization of RLCs. Finally, Theorem 6 about punctured codes in the stochastic error model is proved in Section 8.

3 Technical overview

For the sake of exposition, we begin by proving Theorem 7—a weaker version of Corollary 2. Theorem 7 showcases the techniques by which we prove our main results in a simplified setting. Rather than reducing from state of the art results about RLC list-decodability, Theorem 7 is proven directly, resulting in worse bounds on the list-size. For simplicity, we restrict ourselves to the binary regime.

Theorem 7. *Let $\rho \in \left(0, \frac{1}{2}\right)$ and $L \in \mathbb{N}$. Then, there exist $\eta(L) > 0$ and $\varepsilon(L) > 0$ with $\varepsilon(L) \xrightarrow[L\to\infty]{} 0$, such that the following holds. Let $\mathcal{D} \subseteq \mathbb{F}_2^m$ be a linear η-biased code, and let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq 1 - h_2(\rho) - \varepsilon$. Then $\mathcal{C}$ is (ρ, L)-list-decodable with high probability as $n \to \infty$.*

Proof sketch. Let $\varphi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be the random puncturing map by which $\mathcal{C}$ is generated from $\mathcal{D}$. Write $b = \lceil \log_2(L+1) \rceil$. Now any set of $L+1$ vectors in $\mathbb{F}_2^n$ must contain a subset of b linearly-independent vectors. In particular, for $\mathcal{C}$ to contain a ρ-clustered set of size $L+1$, it must contain a ρ-clustered set of b linearly-independent vectors (this argument originated in [ZP81]). Thus, the probability, taken over the random puncturing φ, that $\mathcal{C}$ is **not** (ρ, L)-list-decodable is at most

$$\begin{aligned}
&\Pr\left[\exists v_1, \ldots, v_b \in \mathcal{C} \text{ which are } \rho\text{-clustered and linearly-independent}\right] \\
&= \Pr\left[\exists u_1, \ldots, u_b \in \mathcal{D} \text{ s.t. } \varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered and linearly-independent}\right] \\
&\le \Pr\left[\exists u_1, \ldots, u_b \in \mathcal{D} \text{ which are linearly independent, s.t. } \varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered}\right] \\
&\le \sum_{\substack{u_1,\ldots u_b \in \mathcal{D} \\ \text{linearly independent}}} \Pr\left[\varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered}\right] \quad ,
\end{aligned}$$

where the penultimate inequality is because linear-independence of $u_1, \ldots, u_b$ is a necessary condition for linear-independence of $\varphi(u_1), \ldots, \varphi(u_b)$. The sum on the right hand side has at most $|\mathcal{D}|^b = 2^{bRn}$ terms, so it suffices to show that

$$\Pr\left[\varphi(u_1), \ldots, \varphi(u_b) \text{ are } \rho\text{-clustered}\right] \le 2^{-bRn-\omega(1)} \tag{7}$$

whenever $u_1, \ldots, u_b \in \mathcal{D}$ are linearly independent.

Let $B \in \mathbb{F}_2^{m\times b}$ be the matrix whose columns are $u_1, \ldots, u_b$, and let σ denote the distribution, over $\mathbb{F}_2^b$, of a uniformly random row of B. Let $A \in \mathbb{F}_2^{n\times b}$ be the matrix whose columns are $\varphi(u_1), \ldots, \varphi(u_b)$. A crucial observation is that A is a random matrix whose rows are sampled independently from σ. At this point, if σ were the uniform distribution over $\mathbb{F}_2^b$, we would be done. Indeed, σ being uniform means that the columns of A, call them $c_1, c_2, \ldots, c_b$, are sampled independently and uniformly from $\mathbb{F}_2^n$. This establishes Eq. (7) since

$$\begin{aligned}
\Pr_{c_1,\ldots,c_b\sim \mathsf{U}(\mathbb{F}_2^n)}\left[c_1, \ldots, c_b \text{ are } \rho\text{-clustered}\right] &\le \sum_{z\in\mathbb{F}_2^n} \sum_{y_1,\ldots,y_b\in B(z,\rho n)} \Pr_{c_1,\ldots,c_b\sim \mathsf{U}(\mathbb{F}_2^n)}\left[\bigwedge_{i=1}^{b}(c_i = y_i)\right] \\
&= \sum_{z\in\mathbb{F}_2^n} \sum_{y_1,\ldots,y_b\in B(z,\rho n)} (2^{-b})^n \qquad (8)\\
&\le \sum_{z\in\mathbb{F}_2^n} 2^{bh_2(\rho)n}(2^{-b})^n = 2^{n(bh_2(\rho)-b+1)} \\
&\le 2^{-bRn-n} \quad , \qquad (9)
\end{aligned}$$

where $B(z, \rho n)$ denotes the Hamming ball of radius ρn around z, and the last inequality Eq. (9) holds for, say, $\varepsilon = \frac{2}{b}$. Note that $\varepsilon \le O\left(\frac{1}{\log L}\right)$.

We now use a certain formulation of the *Vazirani XOR-Lemma* (see, e.g., [Gol11]) to show that σ is in fact arbitrarily close to the uniform distribution over $\mathbb{F}_2^b$. This allows us to finish the theorem by extending the above argument from uniform σ to almost-uniform σ.

Lemma 3.1 (Vazirani XOR-Lemma)**.** *Let σ be a distribution over $\mathbb{F}_2^b$ such that for every $y \in \mathbb{F}_q^b \setminus \{0\}$, we have $\frac{1-\eta}{2} \le \Pr_{x\sim\sigma}\left[\langle x, y\rangle = 1\right] \le \frac{1+\eta}{2}$. Then, σ is $\left(2^b \cdot \eta\right)$-close in* total-variation distance *to the uniform distribution over $\mathbb{F}_2^b$.*

In our case, $\Pr_{x\sim\sigma}\left[\langle x, y\rangle = 1\right] = \mathrm{wt}(By)$. Since the columns of B belong to $\mathcal{D}$ and are linearly-independent, By is a non-zero codeword of $\mathcal{D}$. Our assumption about $\mathcal{D}$ having small bias means that $\mathrm{wt}(By)$ is very close to $\frac{1}{2}$, so the hypothesis of Lemma 3.1 is satisfied. Thus, in the above calculation the rows of A are sampled i.i.d from a distribution $\sigma \sim \mathbb{F}_2^b$ which has statistical distance at most $2^b\eta$ from uniform. Therefore, we can replace the 2^{-b} term in Eq. (8) by an upper bound $(2^{-b}+2^b\eta)$. By taking η small enough, say at most 2^{-2b}, the bound in Eq. (9) remains valid by slightly adjusting parameters (e.g., taking $\varepsilon = \frac{3}{b}$). $\square$

4 Preliminaries

4.1 General notation

We denote the uniform distribution over a finite nonempty set S by $\mathsf{U}(S)$.

For $a, b \in \mathbb{R}$, we denote $\exp_a(b) = a^b$.

The constants implied by asymptotic notation are universal unless stated otherwise. To indicate that the hidden constant may depend on, e.g., for the parameter p, we write "$O_p(\cdot)$".

If $A \in \mathbb{F}_q^{m\times b}$ and $\mathcal{C} \subseteq \mathbb{F}_q^m$, we write $A \subseteq \mathcal{C}$ to mean that each column of A is a codeword in $\mathcal{C}$. Given a puncturing map $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$, let $\varphi(A)$ denote the matrix obtained from A by applying φ to each column.

4.2 A characterization of list-decodable linear codes

Recall the notion of a ρ-clustered set (Definition 2.4.)

Definition 4.1. *Fix $\rho \in [0,1]$, $L \in \mathbb{N}$. A matrix $A \in \mathbb{F}_q^{n\times b}$ $(1 \le b \le L)$ with* $\operatorname{rank} A = b$ *is* (ρ, L)-span-clustered *if the column-span of A contains a ρ-clustered set of size L.*

Note that for a linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ we have

$$\mathcal{C} \text{ is } \textit{not } (\rho, L)\text{-list-decodable} \iff \exists A \in \mathbb{F}_q^{n\times b} \text{such that } A \text{ is } (\rho, L+1)\text{-span-clustered and } A \subseteq \mathcal{C} \ .$$

Furthermore, we can always take b to be in the range $[\log_q(L+1), L+1]$. Indeed, a matrix of rank smaller than $\log_q(L+1)$ cannot be $(\rho,(L+1))$-span-clustered since its span has cardinality smaller than $L+1$. On the other hand, a rank larger than $L+1$ is never needed since, given a ρ-clustered set $W \subseteq \mathcal{C}$ with $|W| = L+1$, one can take A to be a matrix whose columns are a maximal linearly independent subset of W.

4.3 The scalar-multiplied code $\Lambda\mathcal{C}$ and scalar-expanded code $\mathcal{D}^*$

Let

$$\Gamma_n = \left\{\Lambda \in \mathbb{F}_q^{n\times n} \mid \Lambda \text{ is diagonal and of full-rank}\right\} \ .$$

Recall that, for $\Lambda \in \Gamma_n$ and a code $\mathcal{C} \subseteq \mathbb{F}_q^n$, we denote $\Lambda\mathcal{C} = \{\Lambda u \mid u \in \mathcal{C}\}$ (Definition 2.8). If $\mathcal{P}$ is a scalar-invariant property (such as list-decodability), the question of whether a code $\mathcal{C}$ satisfies $\mathcal{P}$ reduces to that of any code of the form $\Lambda\mathcal{C}$. To take advantage of this reduction, we shall focus on the random code $\Lambda\mathcal{C}$ where $\Lambda \sim \mathsf{U}(\Gamma_n)$. If $\mathcal{C}$ is a random puncturing of some code $\mathcal{D}$, we can realize the code $\Lambda\mathcal{C}$ as a puncturing of the code $\mathcal{D}^*$, which we now define.

Definition 4.2. *Given $u \in \mathbb{F}_q^m$, let $u^* \in \mathbb{F}_q^{m(q-1)}$ denote the vector*

$$u^* = \bigodot\nolimits_{a\in\mathbb{F}_q^*} (au) \ ,$$

where $\bigodot$ stands for concatenation of vectors. Given a matrix $B \in \mathbb{F}_q^{m\times b}$ with columns $a_1, \ldots, a_b$, let $B^* \in \mathbb{F}_q^{m(q-1)\times b}$ be the matrix whose columns are $a_1^*, \ldots, a_b^*$. Denote $\mathcal{D}^* = \{u^* \mid u \in \mathcal{D}\} \subseteq \mathbb{F}_q^{m(q-1)}$.

Observation 4.3. *The code $\Lambda\mathcal{C}$, where $\Lambda \sim \mathsf{U}(\Gamma_n)$ and $\mathcal{C}$ is a random n-puncturing of $\mathcal{D}$, is distributed identically to a random n-puncturing of $\mathcal{D}^*$.*

4.4 Fourier transform

We recall the following elementary facts about the Fourier transform[8] of a function $f : \mathbb{F}_q^b \to \mathbb{C}$.

Definition 4.4 (Fourier (and inverse Fourier) transform)**.** *Suppose that $q = p^r$ for some prime p, and let $\omega = e^{\frac{2\pi i}{p}}$. Let $b \in \mathbb{N}$ and let $f : \mathbb{F}_q^b \to \mathbb{C}$. Then $\widehat{f} : \mathbb{F}_q^b \to \mathbb{C}$ is defined by*

$$\widehat{f}(y) = \sum_{x\in\mathbb{F}_q^b} f(x)\cdot\overline{\chi_y(x)}, \quad \textit{where} \quad \chi_y(x) = \omega^{\mathrm{tr}\langle x,y\rangle} \ .$$

Here, $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ *stands for the field trace function* $\mathrm{tr}(x) = \sum_{i=0}^{r-1} x^{p^i}$. *We also have the Fourier inversion formula:*

$$f(x) = q^{-b}\sum_{y\in\mathbb{F}_q^b} \widehat{f}(y)\chi_y(x) \ .$$

Fact 4.5 (Parseval's identity)**.** *Let $f, g : \mathbb{F}_q^b \to \mathbb{C}$. Then,*

$$\sum_{x\in\mathbb{F}_q^b} f(x)\overline{g(x)} = \mathbb{E}_{y\sim\mathsf{U}(\mathbb{F}_q^b)}\left[\widehat{f}(y)\overline{\widehat{g}(y)}\right] \ .$$

In particular, $\sum_{x\in\mathbb{F}_q^b} |f(x)|^2 = \mathbb{E}_{y\sim\mathsf{U}(\mathbb{F}_q^b)}\left[\left|\widehat{f}(y)\right|^2\right]$.

4.5 Entropy and KL-divergence

Given $x \in [0, 1]$, we write

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

for the base-q entropy of a random variable over $\{0, \ldots, q-1\}$, which takes 0 with probability $1-x$ and each $i \in \{1, \ldots, q-1\}$ with probability $\frac{x}{q-1}$.

The q-ary Kullback-Leibler Divergence of two distributions τ, σ over a finite set S is

$$D_{\mathrm{KL}_q}(\tau \parallel \sigma) = \sum_{s\in S}\tau(s)\log_q\frac{\tau(s)}{\sigma(s)} \ .$$

[8]Our convention is to use counting norm for f and expectation norm for $\hat{f}$.

4.6 The empirical distribution of the rows of a matrix

Definition 4.6. *Given a vector $a \in \mathbb{F}_q^n$ we define its* empirical distribution Emp_a *over* $\mathbb{F}_q$ *by*

$$\mathsf{Emp}_a(x) = \Pr_{i \in [n]}[a_i = x] \quad .$$

More generally, given $A \in \mathbb{F}_q^{n \times b}$, let Emp_A denote its empirical row distribution*, that is, the distribution over $\mathbb{F}_q^b$ defined by*

$$\mathsf{Emp}_A(x) = \Pr_{i \in [n]}[A_i = x] \quad ,$$

where A_i denotes the i'th row of A.

Fact 4.7 ([CT06, Thm. 11.1.4])**.** *Let $X \in \mathbb{F}_q^{n \times b}$ have rows sampled identically and independently from some distribution σ over $\mathbb{F}_q^b$. Then, for any distribution τ over $\mathbb{F}_q^b$,*

$$\Pr[\mathsf{Emp}_A = \tau] \leq q^{-D_{\mathrm{KL}_q}(\tau \| \sigma) \cdot n} \quad .$$

Definition 4.8. *Let τ be a distribution over $\mathbb{F}_q^b$. We denote $\dim(\tau) = \dim \operatorname{supp}(\tau)$. If $\dim(\tau) = b$, we say that τ is a* full-rank *distribution.*

Definition 4.9 (Matrix of a particular distribution)**.** *Let τ be a distribution over $\mathbb{F}_q^b$ (where $b \in \mathbb{N}$). For $n \in \mathbb{N}$, we denote*

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

A distribution τ over $\mathbb{F}_q^b$ is said to be n-feasible if $\tau(x) \cdot n$ is an integer for all $x \in \mathbb{F}_q^b$. Observe that any n-feasible distribution over $\mathbb{F}_q^b$ corresponds to a partition of n identical balls into q^b buckets. The bound below thus follows immediately.

Fact 4.10. *The number of n-feasible distributions over $\mathbb{F}_q^b$ is at most $(n+1)^{q^b}$.*

Clearly, n-feasibility of τ is a necessary condition for $\mathcal{M}_{n,\tau}$ to be nonempty. When this condition holds, $|\mathcal{M}_{n,\tau}|$ is equal to the multinomial coefficient $\frac{n!}{\prod_{x \in \mathbb{F}_q^b} (\tau(x) n)!}$. By standard bounds on multinomial coefficients, we have

$$n^{-O(q^b)} \cdot q^{n \cdot H_q(\tau)} \leq |\mathcal{M}_{n,\tau}| \leq q^{n \cdot H_q(\tau)} \quad . \tag{10}$$

4.6.1 The Fourier transform of an empirical distribution

We record several useful properties of the function $\widehat{\mathsf{Emp}_A}$ for a given matrix A. The following is immediate.

Fact 4.11. *A vector $u \in \mathbb{F}_q^n$ is η-biased ($\eta > 0$) if and only if $\left|\widehat{\mathsf{Emp}_u}(a)\right| \leq \eta$ for all $a \in \mathbb{F}_q^*$.*

The following identity shows that the Fourier transform of Emp_A (where $A \in \mathbb{F}_q^{n \times b}$) is in fact composed of the Fourier transforms of Emp_{Ay} over $y \in \mathbb{F}_q^b$. Let $a \in \mathbb{F}_q$. Then,

$$\widehat{\mathsf{Emp}_A}(ay) = \sum_{x \in \mathbb{F}_q^b} \mathsf{Emp}_A(x) \omega^{-\mathrm{tr}(a\langle x,y \rangle)} = \mathbb{E}_{x \sim \mathsf{Emp}_A}\left[\omega^{-\mathrm{tr}(a\langle x,y \rangle)}\right] = \mathbb{E}_{z \sim \mathsf{Emp}_{Ay}}\left[\omega^{-\mathrm{tr}(az)}\right] = \widehat{\mathsf{Emp}_{Ay}}(a) \quad . \tag{11}$$

By Fact 4.5, the normalized Hamming Weight of a vector $u \in \mathbb{F}_q^n$ can be conveniently expressed in terms of the Fourier transform of Emp_u.

$$\mathrm{wt}(u) = \sum_{x\in\mathbb{F}_q} \mathbf{1}_{x\neq 0}\cdot \mathsf{Emp}_u(x) = \frac{q-1}{q}\cdot\widehat{\mathsf{Emp}_u}(0) - \frac{1}{q}\cdot\sum_{a\in\mathbb{F}_q^*}\widehat{\mathsf{Emp}_u}(a) = \frac{q-1}{q} - \frac{1}{q}\cdot\sum_{a\in\mathbb{F}_q^*}\widehat{\mathsf{Emp}_u}(a). \tag{12}$$

This yields the following relation between bias and weight.

Lemma 4.12. *Let $u \in \mathbb{F}_q^n$ be η-biased for some $\eta > 0$. Then*

$$\frac{q-1}{q}(1-\eta) \le \mathrm{wt}(u) \le \frac{q-1}{q}(1+\eta)\ .$$

Proof. By Eq. (12) and Fact 4.11,

$$\left|\mathrm{wt}(u) - \frac{q-1}{q}\right| = \left|\frac{1}{q}\cdot\sum_{a\in\mathbb{F}_q^*}\widehat{\mathsf{Emp}_u}(a)\right| \le \frac{q-1}{q}\cdot\eta\ . \quad \square$$

We have the following immediate conclusion.

Lemma 4.13. *For any $\eta \ge 0$, an η-biased code also has η-optimal distance.*

5 A random puncturing of a near-optimal-distance code is likely to be list-decodable

Our goal in this section is to prove Theorem 5 on the list-decodability of random puncturings of any mother code of sufficiently high distance.

5.1 GHK list-decodability bound for random linear codes revisited

The main result of [GHK11] gives bounds on the list-size for list-decoding of RLCs up to capacity. Here, we go deeper and slightly reformulate[9] the main technical claim of that paper.

Theorem 5.1 ([GHK11, Thm. 6.1]). *Let q be a prime power and let $\rho \in (0, 1-1/q)$. Then, there is a constant $K' = K'_{\rho,q} \ge 1$ such that, for all $b, L \in \mathbb{N}$, we have*

$$\left|\left\{A \in \mathbb{F}_q^{n\times b} \mid A \textit{ is } (\rho, L+1)\textit{-span-clustered}\right\}\right| \le q^{(bh_q(\rho)-4)\cdot n}$$

whenever $L \ge K'\cdot b$ and n is large enough, and

$$\left|\left\{A \in \mathbb{F}_q^{n\times b} \mid A \textit{ is } (\rho, L+1)\textit{-span-clustered}\right\}\right| \le q^{(bh_q(\rho)+1)\cdot n} \tag{13}$$

in general.

Furthermore, one can take

$$K' \le \exp\left(O\left(\frac{(\log_2 q)^2}{\min\{(1-1/q-\rho)^2, \rho\}}\right)\right)\ . \tag{14}$$

[9] See Remark 5.2 for the differences in our formulation.

Remark 5.2. *There are several differences between our formulation of the theorem and the one that appears in [GHK11]. We list and justify them here.*

(i) The random vectors $X_1, \ldots, X_\ell$ from the original formulation have become the columns of the matrix A, and we changed the name ℓ to b.

(ii) The original statement of [GHK11, Thm. 6.1] only deals with matrices whose span contains a large set clustered around 0. In our statement we already apply the reduction to a ball with arbitrary center, which appears in [GHK11, Thm. 2.1].

(iii) Eq. (13) is a rather naive bound, originally derived as part of the proof of [GHK11, Thm. 2.1].

(iv) The asymptotic statement about $K'_{\rho,q}$ comes from inspecting the proof in [GHK11]. Specifically, in the notation of that paper, [GHK11, Lem. 6.3] yields a 2-increasing chain of length $d = \Omega(\log_q K')$ whenever $L \geq K' \cdot b$. The exponent in the q-ary analog of [GHK11, Lem 4.1] satisfies $\delta_p = \Theta\left(\frac{\min\left\{\rho, \left(1-\frac{1}{q}-\rho\right)^2\right\}}{\log_2 q}\right)$. Finally, the requirement in [GHK11, Thm. 6.1] is that K' be large enough so that $d \cdot \delta_p \geq \Omega(1)$.

It will be convenient to formulate a corollary from Theorem 5.1 in terms of the row-distributions of certain matrices.

Definition 5.3. *Fix a prime power q. Let $b, n \in N$ and let τ be an n-feasible distribution over $\mathbb{F}_q^b$. If a matrix $A \in \mathcal{M}_{n,\tau}$ is $(\rho, L+1)$-span-clustered, we say that τ is $(\rho, L+1)$*-span-clustered (with regard to n).

Remark 5.4. *Observe that the notion of τ being $(\rho, L+1)$-span clustered is well defined, and in particular does not depend on the choice of A in Definition 5.3. In other words, either every matrix in $\mathcal{M}_{n,\tau}$ is $(\rho, L+1)$-span-clustered, or none of them are. Indeed, suppose that $A \in \mathbb{F}_q^{n\times b}$ is $(\rho, L+1)$-span-clustered with regard to some center $z \in \mathbb{F}_q^n$, and let B be a matrix obtained from A by permuting the rows of the latter according to some permutation π over $[n]$. Then, B is $(\rho, L+1)$-span-clustered with regard to the center vector resulting from applying π to z.*

The above observation is a result of the fact that containing a $(\rho, L+1)$ -span-clustered matrix is a row-symmetric *code property (see Definition 2.5).*

Corollary 5.5. *In the setting of Theorem 5.1, every $(\rho, L+1)$-span-clustered (with regard to n), n-feasible distribution τ over $\mathbb{F}_q^b$ satisfies*

$$H_q(\tau) \leq b \cdot \left(h_q(\rho) + \frac{5K'_{\rho,q}}{L}\right) - 3 \ .$$

for every b and n such that $\frac{n}{\log_q n} \geq \omega\left(q^{L+1}\right)$.

Proof. By Remark 5.4, $\mathcal{M}_{n,\tau} \subseteq \left\{A \in \mathbb{F}_q^{n\times b} \mid A \text{ is } (\rho, L+1)\text{-span-clustered}\right\}$. Thus, Eq. (10) and Theorem 5.1 yield the following:

If $L \geq K'_{\rho,q} \cdot b$**:**

$$H_q(\tau) \leq \log_q |\mathcal{M}_{n,\tau}| + O\left(\frac{q^b \cdot \log_q n}{n}\right) \leq bh_q(\rho) - 4 + O\left(\frac{q^b \cdot \log_q n}{n}\right)$$

If $L < K'_{\rho,q} \cdot b$**:**

$$\begin{aligned} H_q(\tau) &\leq \log_q |\mathcal{M}_{n,\tau}| + O\left(\frac{q^b \cdot \log_q n}{n}\right) \leq bh_q(\rho) + 1 + O\left(\frac{q^b \cdot \log_q n}{n}\right) \\ &\leq bh_q(\rho) + \frac{5bK'_{\rho,q}}{L} - 4 + O\left(\frac{q^b \cdot \log_q n}{n}\right) \ . \end{aligned}$$

The claim now follows from our assumption that $\frac{n}{\log_q n} \geq \omega\left(q^{L+1}\right)$. $\square$

5.2 Proof of Theorem 5

We now turn to proving Theorem 5, restated here.

Theorem 5 (A puncturing of a near-optimal distance code is whp list-decodable up to capacity)**.** *Fix a prime power q. Let $L, n \in \mathbb{N}$ and $0 < \rho < \frac{q-1}{q}$, such that $\frac{n}{\log_q n} \geq \omega_{n\to\infty}\left(q^{L+1}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code with η-optimal distance, where $\eta = q^{-L+1}$. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate R, where $R \leq 1 - h_q(\rho) - \frac{K}{L}$ for some constant $K = K_{\rho,q}$. Then,*

$$\Pr\left[\mathcal{C} \text{ is } (\rho, L)\text{-list-decodable}\right] \geq 1 - q^{-\Omega(n)} \ .$$

Furthermore, one can take

$$K_{\rho,q} \leq \exp\left(O\left(\frac{(\log q)^2}{\min\left\{(1-1/q-\rho)^2, \rho\right\}}\right)\right) \tag{5}$$

and, in particular, $K_{\rho,q} \leq \mathrm{poly}(q)$ whenever ρ is bounded away from 0 and $1 - \frac{1}{q}$.

Before proving the theorem, we compare it to several known results about list-decodability of RLCs. By the List-Decoding Capacity Theorem, Theorem 5 achieves the optimal trade-off between q, ρ and R. We thus turn to discuss the secondary trade-off, which involves the former three parameters and the *list-size* L. As mentioned in Section 2.5, Theorem 5 is derived by reduction to the result of [GHK11] on list-decodability of RLCs. The main theorem of [GHK11] states that a RLC of rate $R = 1 - h_q(\rho) - \frac{K'_{\rho,q}}{L}$ is with high probability is (ρ, L)-list-decodable, where $K'_{\rho,q} \leq \exp\left(O\left(\frac{(\log q)^2}{\min\{(1-1/q-\rho)^2,\rho\}}\right)\right)$ is proportional to the constant $K_{\rho,q}$ that appears in Theorem 5. Denoting the *gap-to-capacity of the rate* by $\varepsilon = 1 - h_q(\rho) - R$, [GHK11] shows that an RLC of rate R is almost surely (ρ, L)-list-decodable with $L \approx \frac{K'_{\rho,q}}{\varepsilon}$. In Theorem 5, we have $\varepsilon = \frac{K_{\rho,q}}{L}$, so $L = \frac{K_{\rho,q}}{\varepsilon} = O\left(\frac{K'_{\rho,q}}{\varepsilon}\right)$. Thus, we can informally state Theorem 5 as “A random puncturing of a code of near-optimal distance is very likely to be list-decodable up to capacity, with a similar list-size trade-off to that guaranteed by [GHK11] for RLCs.”

The list-size L guaranteed by Theorem 5 inherits some desirable properties from [GHK11]: it is constant in terms of n, and has linear dependence on $\frac{1}{\varepsilon}$, which is tight for RLCs [GN14, Thm. 16]. As for the dependence on q and ρ, we get good list-size bounds when q is not too large and ρ is bounded away from 0 and $1-\frac{1}{q}$, but, unfortunately, the constant $K_{\rho,q}$ grows exponentially as $\rho \to 1-\frac{1}{q}$. In comparison with [GHK11], other works on RLC list-decodability are more specialized, and give tighter upper bounds on the list-size in specific regimes. Notably, [Woo13] does well when ρ is large and ε is of similar magnitude to R, and [LW21] gives an extremely tight upper bound (see [GLMRSW20]) on the list-size for every ρ and ε, when $q = 2$.

We note that, while Theorem 5 only achieves the analogue of [GHK11] for randomly punctured codes, Theorems 1 and 4, with their somewhat stronger hypotheses, achieve (in particular) analogues of all known *and future* positive results about list-decodability of RLCs. The obstacle to concluding such a broad result solely from the hypothesis of Theorem 5 is discussed in Remark 6.14.

Our proof of Theorem 5 relies on the following lemma, whose proof we defer to Section 5.3.

Lemma 5.6 (Puncturings of large-distance codes are locally similar in expectation to random linear codes)**.** *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 linear code of η-optimal distance ($\eta \geq 0$). Let $\Lambda \sim \mathsf{U}(\Gamma_n)$ and, independently, let φ be a random ($m \to n$) puncturing map. Denote $R = \frac{\log_q |\mathcal{D}|}{n}$. Then,*

$$\mathbb{E}\left[|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \Lambda \cdot \varphi(\mathcal{D})\}|\right] \leq \exp_q\Big(n \cdot \Big(H_q(\tau) - (1-R)b + \log_q\Big(1+\eta q^b\Big) + \log_q 2\Big)\Big)$$

Remark 5.7. *Lemma 5.6 bounds the expected number of τ-distributed matrices in the code $\Lambda \cdot \varphi(\mathcal{D})$. The lemma says that this number is not much larger than the expected number of τ-distributed matrices in a random linear code of similar rate. Indeed, for a given matrix A, the probability of A being contained in the random linear code $C_{\mathrm{RLC}}^{n,q}(R)$ is $q^{-n(1-R)\cdot \operatorname{rank}(A)}$. Thus, by Eq. (10),*

$$\mathbb{E}\left[\left|\left\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq C_{\mathrm{RLC}}^{n,q}(R)\right\}\right|\right] = |\mathcal{M}_{n,\tau}| \cdot q^{n(R-1)\cdot b} \approx q^{n(H_q(\tau)-(1-R)b)} \ .$$

Using Lemma 5.6, we conclude Theorem 5 from Theorem 5.1.

Proof of Theorem 5. Take $K_{\rho,q} = 5K'$, where $K'_{\rho,q}$ is as in Theorem 5.1. We need to show that

$$\Pr\left[\mathcal{C} \text{ is not } (\rho, L)\text{-list-decodable}\right] \leq q^{-\Omega(\varepsilon n)}.$$

Since being not (ρ, L)-list-decodable is a scalar-invariant property (see Definition 2.8 and Section 4.3), it suffices to show instead that

$$\Pr\left[\Lambda\mathcal{C} \text{ is not } (\rho, L)\text{-list-decodable}\right] \leq q^{-\Omega(\varepsilon n)} \ , \tag{15}$$

where the matrix Λ is sampled uniformly from Γ_n.

Now, if $\Lambda\mathcal{C}$ is *not* (ρ, L)-list-decodable, then $\Lambda\mathcal{C}$ contains some $(\rho, L+1)$-span-clustered matrix $A \in \mathbb{F}_q^{n\times b}$ for some b, $\log_q(L+1) \leq b \leq L+1$. Hence,

$$\begin{aligned}
&\Pr\left[\Lambda\mathcal{C} \text{ is not } (\rho, L)\text{-list-decodable}\right] \\
&\leq \sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \Pr\left[\exists A \in \mathbb{F}_q^{n\times b} \text{ s.t. } A \text{ is } (\rho, L+1)\text{-span-clustered and } A \subseteq \Lambda\mathcal{C}\right] \\
&\leq \sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \mathbb{E}\left[\left|\left\{A \in \mathbb{F}_q^{n\times b} \mid A \text{ is } (\rho, L+1)\text{-span-clustered and } A \subseteq \Lambda\mathcal{C}\right\}\right|\right] \ .
\end{aligned}$$

By Remark 5.4, we can write

$$\left\{A \in \mathbb{F}_q^{n\times b} \mid A \text{ is } (\rho, L+1)\text{-span-clustered}\right\} = \bigcup_{\tau\in T_b} \mathcal{M}_{n,\tau}$$

where T_b is a set of n-feasible distributions over $\mathbb{F}_q^b$. Therefore, by Lemma 5.6 and our assumption that $\eta \le q^{-L+1} \le q^{-b}$, the probability that $\Lambda\mathcal{C}$ is not (ρ, L)-list-decodable is at most

$$\begin{aligned}
&\sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \sum_{\tau\in T_b} \mathbb{E}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \Lambda\mathcal{C}\}\right|\right] \\
&\le \sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \sum_{\tau\in T_b} \exp_q\Big(n\cdot\Big(H_q(\tau) - (1-R)b + \log_q(1+\eta q^b) + \log_q 2\Big)\Big) \\
&\le \sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \sum_{\tau\in T_b} \exp_q(n\cdot(H_q(\tau) - (1-R)b + 2)) \ .
\end{aligned}$$

By Corollary 5.5, each term of the inner sum is at most q^{-n}. Therefore, by Fact 4.10,

$$\Pr\left[\Lambda\mathcal{C} \text{ is not } (\rho, L)\text{-list-decodable}\right] \le \sum_{b=\lceil \log_q(L+1)\rceil}^{L+1} \sum_{\tau\in T_b} q^{-n} \le q^{-n}\sum_{b=1}^{L+1}(n+1)^{q^b} \le q^{-n}(L+1)(n+1)^{q^{L+1}} \ ,$$

and the theorem follows due to our assumption that $\frac{n}{\log_q n} \ge \omega\left(q^{L+1}\right)$. $\square$

5.3 Proof of Lemma 5.6

Lemma 5.6 follows from Lemmas 5.8 to 5.10, stated and proven below. The proof of Lemma 5.6 is then completed at the end of this section.

Lemma 5.8 is a variation of the *Vazirani XOR-Lemma* (see [Gol11], and Lemma 3.1 for a special case). Given a distribution σ over $\mathbb{F}_q^b$, the XOR-Lemma relates the *total-variation distance* of σ from the uniform distribution over $\mathbb{F}_q^b$, to the maximum of $|\widehat{\sigma}(y)|$ over all $y \ne 0$. In Lemma 5.8, rather than taking a maximum, we consider the ℓ_1 norm of $\widehat{\sigma}$, which yields a tighter bound when only a small number of entries of $\widehat{\sigma}$ are large in absolute value.

Lemma 5.8. *Fix a prime power q, and $b \in \mathbb{N}$. Let σ be a distribution over $\mathbb{F}_q^L$ and let $f : \mathbb{F}_q^b \to \mathbb{R}$ be a non-negative function. Then,*

$$\mathbb{E}_{x\sim\sigma}\left[f(x)\right] \le \left(\sum_{y\in\mathbb{F}_q^b} |\widehat{\sigma}(y)|\right)\cdot \mathbb{E}_{x\sim \mathsf{U}(\mathbb{F}_q^b)}\left[f(x)\right] \ .$$

Proof. We have

$$\sigma(x) = q^{-b}\sum_{y\in\mathbb{F}_q^b} \widehat{\sigma}(y)\omega^{-\mathrm{tr}(\langle x,y\rangle)} \le q^{-b}\sum_{y\in\mathbb{F}_q^b} |\widehat{\sigma}(y)|$$

for all $x \in \mathbb{F}_q^b$. So

$$\mathbb{E}_{x\sim\sigma}[f(x)] = \sum_{x\in\mathbb{F}_q^b} \sigma(x)f(x) \le q^{-b}\left(\sum_{y\in\mathbb{F}_q^b} |\widehat{\sigma}(y)|\right)\cdot\left(\sum_{x\in\mathbb{F}_q^b} f(x)\right) = \left(\sum_{y\in\mathbb{F}_q^b} |\widehat{\sigma}(y)|\right)\cdot\mathbb{E}_{x\sim\mathsf{U}(\mathbb{F}_q^b)}[f(x)] \enspace . \qquad \square$$

We next bound the expectation of an arbitrary non-negative test function over the empirical row-distribution of a given matrix B, assuming that the column-span of B has good bias or distance. The bias based bound is an immediate application of Lemma 5.8. The weight based bound requires an additional trick, and only yields a result relating to the row-distribution of B^* rather than B itself (recall Definition 4.2 for a reminder about B^*). One reason for the difference between the two cases is that under the weight-based hypothesis we have an upper bound only on the entries of the Fourier transform (Eq. (18)), rather than on their absolute value.

Lemma 5.9. *Let $B \in \mathbb{F}_q^{m\times b}$ have $\operatorname{rank} B = b$, and let $f : \mathbb{F}_q^b \to \mathbb{R}$ be a non-negative function. Then, the following holds for all $\eta \ge 0$:*

1. *Suppose that the column-span of B (as a code in $\mathbb{F}_q^m$) is η-biased. Then,*

$$\mathbb{E}_{x\sim\mathsf{Emp}_B}[f(x)] \le (1+q^b\eta)\cdot\mathbb{E}_{x\sim\mathsf{U}(\mathbb{F}_q^b)}[f(x)] \enspace .$$

2. *Suppose that the column-span of B has η-optimal distance. Then,*

$$\mathbb{E}_{x\sim\mathsf{Emp}_{B^*}}[f(x)] \le 2(1+q^b\eta)\cdot\mathbb{E}_{x\sim\mathsf{U}(\mathbb{F}_q^b)}[f(x)] \enspace .$$

Proof. We first prove Item 1. Hence, by Lemma 5.8 it suffices to show that

$$\sum_{y\in\mathbb{F}_q^b} \left|\widehat{\mathsf{Emp}_B}(y)\right| \le 1 + q^b\eta \enspace .$$

By Eq. (11), the above is equivalent to

$$\sum_{y\in\mathbb{F}_q^b} \left|\widehat{\mathsf{Emp}_{By}}(1)\right| \le 1 + q^b\eta \enspace . \tag{16}$$

For $y = 0$ we have $\widehat{\mathsf{Emp}_{By}}(1) = \widehat{\mathsf{Emp}_0}(1) = 1$. For any $y \in \mathbb{F}_q^b \setminus \{0\}$, since B has full column-rank, By is a non-zero codeword of $\mathcal{D}$. By hypothesis, By is η-biased, so Fact 4.11 yields $\left|\widehat{\mathsf{Emp}_{By}}(1)\right| \le \eta$, establishing Eq. (16).

We now turn to Item 2. Let σ denote the distribution, over $\mathbb{F}_q^b$, of the random variable $a \cdot x$, where $a \sim \mathsf{U}(\mathbb{F}_q^*)$ and x is independently sampled from Emp_B. By Lemma 5.8, to prove Item 2 it suffices to show that

$$\sum_{y\in\mathbb{F}_q^b} |\widehat{\sigma}(y)| \le 2\left(1+q^b\eta\right) \enspace . \tag{17}$$

By Eq. (12) followed by Eq. (11),

$$\mathrm{wt}(By) = \frac{q-1}{q} - \frac{1}{q}\cdot\sum_{a\in\mathbb{F}_q^*}\widehat{\mathsf{Emp}_{By}}(a) = \frac{q-1}{q} - \frac{1}{q}\cdot\sum_{a\in\mathbb{F}_q^*}\widehat{\mathsf{Emp}_B}(ay) = \frac{q-1}{q}\cdot(1-\widehat{\sigma}(y)) \enspace ,$$

so $\widehat{\sigma}(y) = 1 - \frac{q}{q-1} \cdot \mathrm{wt}(By)$.

In particular, if $y \neq 0$ then By is a non-zero element in the column-span of B. Hence, by hypothesis,

$$\widehat{\sigma}(y) = 1 - \frac{q}{q-1} \cdot \mathrm{wt}(By) \leq \eta \ . \tag{18}$$

Let $P = \{y \in \mathbb{F}_q^b \mid \widehat{\sigma}(y) \geq 0\}$ and $N = \mathbb{F}_q^b \setminus P$. By Eq. (18),

$$\sum_{y \in P \setminus \{0\}} \widehat{\sigma}(y) \leq q^b \eta \ .$$

Note that $\widehat{\sigma}(0) = \sum_{x \in \mathbb{F}_q^b} \sigma(x) = 1$, and thus,

$$\sum_{y \in P} \widehat{\sigma}(y) = 1 + \sum_{y \in P \setminus \{0\}} \widehat{\sigma}(y) \leq 1 + q^b \eta \ .$$

Consequently,

$$0 \leq q^b \cdot \sigma(0) = \sum_{y \in \mathbb{F}_q^b} \widehat{\sigma}(y) = \sum_{y \in P} \widehat{\sigma}(y) + \sum_{y \in N} \widehat{\sigma}(y) \leq 1 + q^b \eta + \sum_{y \in \mathbb{N}} \widehat{\sigma}(y)$$

and so,

$$\sum_{y \in N} |\widehat{\sigma}(y)| = -\sum_{y \in N} \widehat{\sigma}(y) \leq 1 + q^b \eta \ .$$

Eq. (17) now follows since

$$\sum_{y \in \mathbb{F}_q^b} |\widehat{\sigma}(y)| = \sum_{y \in P} |\widehat{\sigma}(y)| + \sum_{y \in N} |\widehat{\sigma}(y)| \leq 2(1 + q^b \eta) \ . \quad \square$$

Lemma 5.10 bounds the probability of a random puncturing of a given matrix B having a certain empirical distribution τ. Due to the concavity argument in Eq. (20), this lemma gives tighter bounds when Emp_B is close to the uniform distribution over $\mathbb{F}_q^b$. Notably, as Lemma 5.9 shows, good bias or similar properties of the column-span of B ensure that Emp_B is indeed close to uniform.

Lemma 5.10. *Fix some distribution τ over $\mathbb{F}_q^b$. Let $B \in \mathbb{F}_q^{m \times b}$ have* $\operatorname{rank} B = b$. *Let $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ be a random puncturing map. Then,*

$$\Pr\left[\varphi(B) \in \mathcal{M}_{n,\tau}\right] \leq \exp_q\!\left(n \left(\log_q \mathbb{E}_{x \sim \mathsf{Emp}_B}\left[\tau(x)\right] + H_q(\tau)\right)\right) \ .$$

Proof. By Fact 4.7,

$$\Pr\left[\varphi(B) \in \mathcal{M}_{n,\tau}\right] = \Pr\left[\mathsf{Emp}_{\varphi(B)} = \tau\right] \leq q^{-n \cdot D_{\mathrm{KL}_q}(\tau \| \mathsf{Emp}_B)} \ . \tag{19}$$

By concavity of log,

$$\begin{aligned} D_{\mathrm{KL}_q}\left(\tau \parallel \mathsf{Emp}_B\right) &= \sum_{x \in \mathbb{F}_q^b} \tau(x) \log_q \frac{\tau(x)}{\mathsf{Emp}_B(x)} = -H_q(\tau) - \sum_{x \in \mathbb{F}_q^b} \tau(x) \log_q \mathsf{Emp}_B(x) \\ &\geq -H_q(\tau) - \log_q \mathbb{E}_{x \sim \mathsf{Emp}_B}\left[\tau(x)\right] \ . \end{aligned} \tag{20}$$

The claim follows from Eq. (19) and Eq. (20). $\square$

Proof of Lemma 5.6. By Observation 4.3, $\Lambda \cdot \varphi(\mathcal{D})$ is distributed identically to $\varphi^*(\mathcal{D}^*)$, where φ^* is a random $((q-1)m \to n)$ puncturing map. Thus,

$$\begin{aligned}\mathbb{E}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \Lambda \cdot \varphi(\mathcal{D})\}\right|\right] &= \mathbb{E}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \varphi^*(\mathcal{D}^*)\}\right|\right] \\ &\leq \mathbb{E}\left[\left|\left\{B \in \mathbb{F}_q^{m\times b} \mid B \subseteq \mathcal{D} \text{ and } \varphi^*(B^*) \in \mathcal{M}_{n,\tau}\right\}\right|\right] \ . \end{aligned} \tag{21}$$

We proceed to bound the expectation of the right-hand side.

Suppose that $\varphi^*(B^*) \in \mathcal{M}_{n,\tau}$. Because τ is of full-rank, we have $\operatorname{rank} B = \operatorname{rank} B^* \geq \operatorname{rank} \varphi^*(B^*) = b$, so $\operatorname{rank} B = b$.

Let $B \in \mathbb{F}_q^{m\times b}$ such that $\operatorname{rank} B = b$ and $B \subseteq \mathcal{D}$. Since the column-span of B is contained in $\mathcal{D}$, it is of η-optimal distance. Hence, by Item 2 of Lemma 5.9,

$$\mathbb{E}_{x\sim \mathsf{Emp}_{B^*}}\left[\tau(x)\right] \leq \Upsilon \ , \tag{22}$$

where $\Upsilon = 2q^{-b}\left(1 + q^b\eta\right)$. Lemma 5.10 yields

$$\begin{aligned}\mathbb{E}\left[\left|\left\{B \in \mathbb{F}_q^{m\times b} \mid B \subseteq \mathcal{D} \text{ and } \varphi^*(B^*) \in \mathcal{M}_{n,\tau}\right\}\right|\right] &= \sum_{\substack{B\in\mathbb{F}_q^{m\times b}\\ B\subseteq\mathcal{D}\\ \operatorname{rank} B = b}} \Pr_{\Lambda,\varphi}\left[\varphi^*(B^*) \in \mathcal{M}_{n,\tau}\right] \\ &\leq q^{bRn} \cdot \exp_q\!\left(n\left(\log_q \Upsilon + H_q(\tau)\right)\right) \ , \end{aligned} \tag{23}$$

and the claim follows from Eqs. (21), (22) and (23). □

6 Random puncturings of certain codes are locally-similar to RLCs

In this section we prove Theorems 1 and 4, restated below.

Theorem 1 (Puncturings of low-bias linear codes are locally similar to random linear codes)**.** *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 $\frac{n}{\log n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon$ for some $\varepsilon > 0$. Suppose that $\mathcal{D}$ is $\left(\frac{\varepsilon b \ln q}{q^b}\right)$-biased. Then,*

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

Theorem 4 (Puncturings of near-optimal distance linear codes with large alphabet are locally similar to random linear codes)**.** *Let q be a prime power, and let $\mathcal{P}$ be a monotone-increasing, row-symmetric, b-local and scalar-invariant property over $\mathbb{F}_q^n$, where $\frac{n}{\log n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be a linear code of q^{-b}-optimal distance. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \leq \mathrm{RLC}(\mathcal{P}) - \varepsilon - 2\log_q 2$ for some $\varepsilon > 0$. Then,*

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

6.1 Monotone-increasing properties and minimal sets

A monotone-increasing property $\mathcal{P}$ of codes has a unique minimal-set $\mathcal{M}_{\mathcal{P}}$, namely, a matrix $A \subseteq \mathbb{F}_q^{n\times b}$ ($b \in \mathbb{N}$) of full column-rank belongs to $\mathcal{M}_{\mathcal{P}}$ if the code consisting of the column-span of A satisfies $\mathcal{P}$, but no proper linear subspace of that code does so.

Example 6.1 (Minimal set for list-decodability)**.** *Fix a prime power q, $n, L \in \mathbb{N}$ and $\rho \in [0,1]$. Consider the monotone-increasing property $\mathcal{P}$ consisting of all linear codes in $\mathbb{F}_q^n$ that are **not** (ρ, L) -list-decodable. Then,*

$$\mathcal{M}_{\mathcal{P}} = \left\{ A \subseteq \mathbb{F}_q^{n\times b} \mid b \in \mathbb{N}, \quad A \textit{ is } (\rho, L+1)\textit{-span-clustered} \right\} .$$

We can reformulate the notions of local, row-symmetric and scalar-invariant monotone-increasing properties in terms of the minimal set $\mathcal{M}_{\mathcal{P}}$ (see Definitions 2.5 and 2.8).

Observation 6.2 (Local, row-symmetric and scalar-invariant properties in terms of $\mathcal{M}_{\mathcal{P}}$)**.** *Let $\mathcal{P}$ be a monotone-increasing property of linear codes in $\mathbb{F}_q^n$.*

1. *Fix $b \in \mathbb{N}$. Then, $\mathcal{P}$ is b-local if and only every matrix in $\mathcal{M}_{\mathcal{P}}$ has at most b columns.*
2. *The property $\mathcal{P}$ is row-symmetric if and only if, for each $A \in \mathcal{M}_{\mathcal{P}}$, it holds that every matrix obtained by permuting the rows of A also belongs to $\mathcal{M}_{\mathcal{P}}$.*
3. *The property $\mathcal{P}$ is scalar-invariant if, for each $A \in \mathcal{M}_{\mathcal{P}}$ and every full-rank diagonal matrix $\Lambda \in \mathbb{F}_q^{n\times n}$ it holds that $\Lambda A \in \mathcal{M}_{\mathcal{P}}$.*

6.2 Row-symmetric b-local properties in terms of distributions over $\mathbb{F}_q^b$

Thresholds for row-symmetric and local properties can be characterized in terms of empirical distributions of certain matrices. We recall this connection.

Fact 6.3 ([GMRSW21, Fact 2.15])**.** *Let $\mathcal{P}$ be a monotone-increasing, b-local, row-symmetric property over $\mathbb{F}_q^n$. Then, there exists a set $\mathcal{T}_{\mathcal{P}}$ of distributions over $\mathbb{F}_q^b$ such that $|\mathcal{T}_{\mathcal{P}}| \le (n+1)^{q^b}$, and $\mathcal{M}_{\mathcal{P}} = \bigcup_{\tau\in\mathcal{T}_{\mathcal{P}}} \mathcal{M}_{n,\tau}$.*

Definition 6.4 (Implied distribution [MRRSW20, Def. 2.6])**.** *Let τ be a distribution over $\mathbb{F}_q^b$ and let $D \in \mathbb{F}_q^{b\times a}$ such that $\operatorname{rank} D = a$ for some $a \le b$. The distribution (over $\mathbb{F}_q^a$) of the random vector xD, where $x \sim \tau$ (note that x is a row vector), is said to be τ-implied. We denote the set of τ-implied distributions by $\mathcal{I}_\tau$.*

The motivation for Definition 6.4 is the following observation, which follows immediately from the linearity of the code.

Observation 6.5. *Let τ be a distribution over $\mathbb{F}_q^b$, and let $\tau' \in \mathcal{I}_\tau$. Then, any linear code containing a matrix in $\mathcal{M}_{n,\tau}$ must also contain some matrix in $\mathcal{M}_{n,\tau'}$.*

We now have the following characterization of the threshold.

Theorem 6.6 ([MRRSW20, Thm. 2.8]). [10] *Let $\mathcal{P}$ be a monotone-increasing, b-local, row-symmetric property over $\mathbb{F}_q^n$, and let $\mathcal{T}_{\mathcal{P}}$ be as in Fact 6.3. Then,*

$$\mathrm{RLC}(\mathcal{P}) = \min_{\tau\in\mathcal{T}_{\mathcal{P}}} \max_{\tau'\in\mathcal{I}_\tau} \left(1 - \frac{H_q(\tau')}{\dim(\tau')}\right) \pm \frac{2q^{2b}\log_q n}{n} \ .$$

Below, we demonstrate Theorem 6.6 via the property of list-recoverability. The motivation is two-fold. First, list-recoverability itself is a property of significant interest. Secondly, we will use Theorem 6.6 in the proof of Theorems 1 and 4, and the special case of list-recoverability will help familiarize the reader with this tool.

6.3 List-recoverability as a property of codes

List-recovery is an important generalization of list-decoding, where the decoder is given not one, but a subset of ℓ symbols per position, and the goal is to list all codewords which "miss" at most a fraction ρ of these subsets. We formally define the notion of (combinatorial) list-recoverability.

Definition 6.7. *Fix $1 \le \ell \le q$ and let $\rho \in (0, 1 - \ell/q)$. The set W is said to be (ρ,ℓ)-*recovery-clustered *if there exist sets $Z_1, \dots, Z_n \subseteq \mathbb{F}_q$, each of which is of size at most ℓ, such that $|\{i \in [n] \mid u_i \notin Z_i\}| \le \rho n$ for all $u \in W$. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is called (ρ,ℓ,L)*-list-recoverable *if it does not contain any (ρ,ℓ)-recovery-clustered set of size $L+1$.*

The following is immediate:

Observation 6.8. *Fix a prime power q, $1 \le \ell < q$, $\rho \in (0, 1-\ell/q)$ and $L \in \mathbb{N}$. Let $\mathcal{P}$ denote the monotone-increasing linear-code property of codes in $\mathbb{F}_q^n$ that are* not *(ρ,ℓ,L)-list-recoverable. Then, $\mathcal{P}$ is row-symmetric, $(L+1)$-local and scalar-invariant.*

Say that a matrix $A \in \mathbb{F}_q^{n\times b}$ is (ρ,ℓ,L)-recovery-span-clustered *if the column span of A contains a (ρ,ℓ)-clustered set of size $L+1$. Then, $\mathcal{M}_{\mathcal{P}}$ is contained in the set of all (ρ,ℓ,L)-recovery-span-clustered matrices.*

Note that list-recovery generalizes list-decodability (Definition 2.4), i.e., a set W is ρ-clustered if and only if it is $(\rho,1)$-recovery-clustered. Likewise, a code is (ρ,L)-list-decodable if and only if it is $(\rho,1,L)$-list-recoverable.

The *List-Recovery Capacity Theorem* [Res20, Thm. 2.4.12] gives the threshold rate for list-recoverability as $R^* = 1 - h_{q,\ell}(\rho)$, where $h_{q,\ell}(\rho) = \rho\log_q\left(\frac{q-\ell}{\rho}\right) + (1-\rho)\log_q\left(\frac{\ell}{1-\rho}\right)$. Namely, for every $\varepsilon > 0$ there exists a family of $(\rho,\ell,O_{\rho,\ell,\varepsilon}(1))$-list-recoverable codes of rate at least $R^* - \varepsilon$ but, on the other hand, every (ρ,ℓ,L)-list-recoverable family of codes of rate $\ge R^* + \varepsilon$ has L exponentially large in εn.

We now use Theorem 6.6 to prove that RLCs achieve list-recovery capacity[11].

Proposition 6.9 (RLCs achieve list-recovery capacity). *For any fixed q, ρ, ℓ and L, we have*

$$\mathrm{RLC}^{n,q}(\rho,\ell,L) \ge 1 - h_{q,\ell}(\rho) - \frac{\ell}{\log_q L} - o_{n\to\infty}(1) \ .$$

[10] The precise error term does not appear in the statement of this theorem in [MRRSW20], but follows by inspecting the proof there.

[11] Better lower bounds on $\mathrm{RLC}^{n,q}(\rho,\ell,L)$ are known. See, e.g., [RW18].

Proposition 6.9 means that RLCs get to within ε of the capacity rate for list-recovery with list-size $L \approx q^{\frac{\ell}{\varepsilon}}$.

To prove Proposition 6.9 we need the following claim.

Claim 6.10. *Let $B \in \mathbb{F}_q^{n\times b}$ be a matrix whose columns form a (ρ,ℓ)-recovery-clustered set. Denote $\tau = \mathsf{Emp}_B$. Then, $H_q(\tau) \le b \cdot h_q(\ell,\rho) + \ell$, where $h_{q,\ell}(\rho) = \rho \log_q\left(\frac{q-\ell}{\rho}\right) + (1-\rho)\log_q\left(\frac{\ell}{1-\rho}\right)$.*

Proof. Let $Z_1,\dots,Z_n$ be subsets of $\mathbb{F}_q$, each of size ℓ, such that for all $j \in [b]$, we have $|\{i \in [n] \mid B_{i,j} \notin Z_i\}| \le \rho n$. Let i be sampled uniformly from $[n]$. We now have

$$H_q(\tau) = H_q(B_i) \le H_q(B_i, Z_i) = H_q(Z_i) + H_q(B_i \mid Z_i) \le H_q(Z_i) + \sum_{j=1}^{b} H_q(B_{i,j} \mid Z_i).$$

The number of different options for Z_i is $\binom{q}{\ell}$ so $H_q(Z_i) \le \log_q(\binom{q}{\ell}) \le \ell$. Let ρ'_j ($j \in [b]$) denote the probability that $B_{i,j} \notin Z_i$, and note that $\rho'_j \le \rho$. Then,

$$H_q(B_{i,j} \mid Z_i) \le h_{q,\ell}(\rho'_j) \le h_{q,\ell}(\rho).$$

Consequently, $H_q(\tau) \le b \cdot h_{q,\ell}(\rho) + \ell$, establishing the claim. $\square$

Proof of Proposition 6.9. Let $\mathcal{P}$ denote the property consisting of codes over $\mathbb{F}_q^n$ that are **not** (ρ,ℓ,L)-list-recoverable. Let $\tau \in \mathcal{T}_{\mathcal{P}}$ and let $A \in \mathcal{M}_{n,\tau}$ be a matrix in $\mathbb{F}_q^{n\times a}$ ($a \in \mathbb{N}$). By Observation 6.8, A is $(\rho,\ell,L+1)$-recovery-span-clustered. Let W be a (ρ,ℓ)-recovery-clustered set of size $L+1$, contained in the column-span of A. Note that W must contain a linearly-independent subset U of size $b := \lceil \log_q |W| \rceil = \lceil \log_q(L+1) \rceil$. Let $D \in \mathbb{F}_q^{a\times b}$ such that $B := AD$ is the matrix whose columns are the elements of U, and note that U is also (ρ,ℓ)-recovery-clustered. Let $\tau' = \mathsf{Emp}_B$. By Claim 6.10, $H_q(\tau') \le b \cdot h_{q,\ell}(\rho) + \ell$. Furthermore, we can express τ' as the distribution of the random vector xD, where $x \sim \tau$. Consequently, $\tau' \in \mathcal{I}_\tau$. Therefore,

$$\max_{\tau'' \in \mathcal{I}_\tau} \left(1 - \frac{H_q(\tau'')}{\dim(\tau')}\right) \ge 1 - \frac{H_q(\tau')}{b} \ge 1 - h_{q,\ell}(\rho) - \frac{\ell}{b}.$$

The claim now follows by Theorem 6.6. $\square$

We note that the above derivation of Proposition 6.9 could also be achieved via more standard arguments, which do not require Theorem 6.6. The actual power Theorem 6.6 is that it enables reductions from other random code models to the RLC model, as demonstrated in the proof of Theorems 1 and 4, via Lemma 6.12. This sort of argument involves an application of Theorem 6.6 in *its less intuitive direction*: rather than starting from an upper bound on $H_q(\tau)$ for some set of distributions and using Theorem 6.6 to obtain a lower bound on RLC($\mathcal{P}$), we start from some known lower bound on RLC($\mathcal{P}$) and use the theorem to get an upper bound on the entropy of certain "bad distributions." The latter entropy bound is then typically used in a union-bound argument to obtain a lower bound on the threshold rate for some non-RLC model. This type of argument was used in [MRRSW20] to prove that *Gallagher LDPC codes* are as list-decodable (and list-recoverable) as RLCs.

Remark 6.11 (Average versions of list-decodability and list-recoverability). *Average-radius list-decodability is a stronger property where we demand that for every $L+1$ codewords their average distance to any center exceeds ρ (as opposed to maximum distance for list-decodability). A code not being (ρ, L)-average-radius list-decodable is also an $(L+1)$-local, row-symmetric and scalar-invariant property.*

For list-recovery, we can define a stronger variant where in Definition 6.7 we allow input sets Z_i such that the average size $|Z_i|$ over all $i \in [n]$ is at most ℓ. A violation of this stronger property is also a local, row-symmetric and scalar-invariant property.

The generality of our framework thus means that we can get results for these variants also automatically. We note that certain results for list-decodability for RLCs, e.g., [GHK11; LW21], do not extend to average-radius list-decoding (or list-recovery).

6.4 Proof of Theorems 1 and 4

We are now ready to prove Theorems 1 and 4. Our proofs take advantage of Theorem 6.6 via the following lemma.

Lemma 6.12 (A generic reduction to random linear codes). *Let $n \in \mathbb{N}$, q a prime power and $b \in \mathbb{N}$ such that $\frac{n}{\log_q n} \geq \omega_{n\to\infty}\left(q^{2b}\right)$. Let $\mathcal{C} \in \mathbb{F}_q^n$ 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 $B \in \mathbb{F}_q^{Rn\times a}$ with* $\operatorname{rank} B = a$, *we have*

$$\mathbb{E}_{\mathcal{C}}\left[|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \mathcal{C}\}|\right] \leq q^{(H_q(\tau) - a(1-R) + a\varepsilon)n} \quad , \tag{24}$$

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

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

Proof. Let $\tau \in \mathcal{T}_{\mathcal{P}}$. By Theorem 6.6, there is some distribution $\tau' \in \mathcal{T}_{\mathcal{P}}$ over $\mathbb{F}_q^a$ (where $1 \leq a \leq b$) such that

$$\frac{H_q(\tau')}{a} \leq 1 - \mathrm{RLC}(\mathcal{P}) + o_{n\to\infty}(1) \quad .$$

Now, by Observation 6.5, followed by Markov's bound,

$$\begin{aligned}
\Pr\left[\exists A \in \mathcal{M}_{n,\tau} \quad A \subseteq \mathcal{C}\right] &\leq \Pr\left[\exists A \in \mathcal{M}_{n,\tau'} \quad A \subseteq \mathcal{C}\right] \leq \mathbb{E}\left[\left|\left\{A \in \mathcal{M}_{n,\tau'} \mid A \subseteq \mathcal{C}\right\}\right|\right] \\
&\leq \exp_q\!\left(\left(H_q(\tau') - a(1-R) + a\varepsilon\right) n\right) \\
&\leq \exp_q\!\left(an\left(R - \mathrm{RLC}(\mathcal{P}) + \varepsilon + o_{n\to\infty}(1)\right)\right) \\
&\leq \exp_q\!\left(-na(\varepsilon - o_{n\to\infty}(1))\right) \quad .
\end{aligned}$$

Therefore, by Fact 6.3,

$$\begin{aligned}
\Pr\left[\mathcal{C} \text{ satisfies } \mathcal{P}\right] &\leq \sum_{\tau \in \mathcal{T}_{\mathcal{P}}} \Pr\left[\exists A \in \mathcal{M}_{n,\tau} \quad A \subseteq \mathcal{C}\right] \leq |\mathcal{T}_{\mathcal{P}}|\, q^{-n(\varepsilon - o_{n\to\infty}(1))} \\
&\leq (n+1)^{q^b} q^{-n(\varepsilon - o_{n\to\infty}(1))} \leq q^{-n(\varepsilon - o_{n\to\infty}(1))} \quad .
\end{aligned}$$

□

We turn to proving the theorems.

Proof of Theorem 4. Let τ be a distribution over $\mathbb{F}_q^a$, where $a \le b$. By Lemma 5.6,

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \Lambda\mathcal{C}\}\right|\right] &\le \exp_q\!\big(n \cdot \big(H_q(\tau) - (1-R)a + \log_q\left(1+\eta q^a\right) + \log_q 2\big)\big) \\ &\le \exp_q\!\big(n \cdot \big(H_q(\tau) - (1-R)a + 2\log_q 2\big)\big) \\ &\le \exp_q(n \cdot (H_q(\tau) - (1-R)a + \varepsilon)) \\ &\le \exp_q(n \cdot (H_q(\tau) - (1-R)a + a\varepsilon)) \ ,\end{aligned}$$

where $\Lambda \sim \mathsf{U}(\Gamma_n)$. By Lemma 6.12, $\Lambda\mathcal{C}$ satisfies $\mathcal{P}$ with probability at most $q^{-n(\varepsilon - o_{n\to\infty}(1))}$. Since $\mathcal{P}$ is scalar-invariant, the same holds for $\mathcal{C}$. □

Theorem 1 requires the following lemma, which is analogous to Lemma 5.6.

Lemma 6.13 (Puncturings of low-bias codes are locally similar to random linear codes)**.** *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 an η-biased linear code ($\eta \ge 0$). Let φ be a random $(m \to n)$ puncturing map. Denote $R = \frac{\log_q |\mathcal{D}|}{n}$. Then,*

$$\mathbb{E}_{\mathcal{C}}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \mathcal{C}\}\right|\right] \le \exp_q\!\left(n \cdot \left(H_q(\tau) - (1-R)b + \log_q\left(1 + \eta q^b\right)\right)\right)$$

Proof. Let τ be a full-rank distribution over $\mathbb{F}_q^b$. Item 1 of Lemma 5.9 yields B

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

for all $B \in \mathbb{F}_q^{m\times b}$ such that $\operatorname{rank} B = b$ and $B \subseteq \mathcal{D}$. By Lemma 5.10,

$$\Pr\left[\varphi(B) \in \mathcal{M}_{n,\tau}\right] \le \exp_q\!\left(n\left(-b + H_q(\tau) + \log_q\left(1 + q^b\eta\right)\right)\right) \ .$$

The claim now follows by the union bound over the $\le q^{Rnb}$ choices of . □

Proof of Theorem 1. Let τ be a distribution over $\mathbb{F}_q^a$ with $a \le b$. Lemma 6.13 yields

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \mathcal{C}\}\right|\right] &\le \exp_q\!\big(n \cdot \big(H_q(\tau) - (1-R)a + \log_q\left(1+\eta q^a\right)\big)\big) \\ &\le \exp_q\!\left(n \cdot \left(H_q(\tau) - (1-R)a + \frac{\eta q^a}{\ln q}\right)\right) \\ &= \exp_q\!\left(n \cdot \left(H_q(\tau) - (1-R)a + \frac{\varepsilon b}{q^{b-a}}\right)\right) \\ &\le \exp_q(n \cdot (H_q(\tau) - (1-R)a + a\varepsilon)) \ ,\end{aligned}$$

which implies the claim by virtue of Lemma 6.12. □

Remark 6.14 (On the conditions in Theorems 1 and 4, and comparison to Theorem 5)**.** *In the above proof of Theorems 1 and 4, as in the proof of Theorem 5, the core of the proof is obtaining an upper bound on terms of the form $\mathbb{E}\left[\left|\{A \in \mathcal{M}_{n,\tau} \mid A \subseteq \Lambda\mathcal{C}\}\right|\right]$ for certain distributions τ, where $\mathcal{C}$ is a random puncturing of a mother code $\mathcal{D}$.*

When our assumption about $\mathcal{D}$ is that of near-optimal distance, we bound this expectation via Lemma 5.6, which includes a bothersome $\log_q 2$ term. This term needs to be bounded from above by $a\varepsilon$. One way to overcome this term is to take q large enough to make $\log_q 2$ negligibly small, as we do in Theorem 4. In Theorem 5 where we use [GHK11], the problem is handled differently: Corollary 5.5 provides us with a "slack" that dominates the $\log_q 2$ term whenever a is small, whereas for large a the $a\varepsilon$ upper bound is not too restrictive. Finally, in Theorem 1 we avoid the bothersome term altogether via Lemma 6.13, due to our assumption that $\mathcal{D}$ has small bias.

7 Derandomization of RLCs

In this section we prove Theorem 3, restated here.

Theorem 3 (Codes locally similar to an RLC with linear randomness)**.** *There exists a randomized algorithm that, given $b \in \mathbb{N}$, $\varepsilon > 0$, $R^* \in [\varepsilon, 1]$ and $n \in \mathbb{N}$, where $\frac{n}{\log_2 n} \geq \omega_{n\to\infty}\left(2^{2b}\right)$ and $n \geq \omega_{n\to\infty}(1/\varepsilon)$, samples a generating matrix for a linear code $\mathcal{C} \subseteq \mathbb{F}_2^n$ of rate $R = R^* - \varepsilon$ such that*

$$\Pr\left[\mathcal{C} \text{ satisfies some property } \mathcal{P} \in \mathcal{K}\right] \leq 2^{-\Omega(\varepsilon n)}. \tag{4}$$

Here, $\mathcal{K}$ is the family of all monotone-increasing, b-local and row-symmetric properties $\mathcal{P}$ over $\mathbb{F}_2^n$ for which the threshold $\mathrm{RLC}(\mathcal{P})$ is at least R^*. This algorithm uses $O\left(n\left(b + \log_2 \frac{1}{\varepsilon}\right)\right)$ random bits, and works in time polynomial in n.

Proof. Fix a property $\mathcal{P} \in \mathcal{K}$. Fix $\eta = \frac{\varepsilon b \ln 2}{2^b}$. Let $\mathcal{D} \subseteq \mathbb{F}_2^m$ be an η-biased linear code of dimension Rn, where $m \leq O(n \cdot \eta^{-c})$ for some universal $c \geq 2$. Explicit constructions of such a code $\mathcal{D}$ are given in [ABNNR92; Ta-17]. We also assume that

$$m \geq \frac{n}{1 - 2^{-\frac{\varepsilon}{2}}} \ , \tag{25}$$

noting that $\frac{m}{n}$ can be taken to be as large as desired.

Sample a random increasing sequence of n integers $1 \leq i_1 < i_2 < \cdots < i_n \leq m$ uniformly from among all such sequences. Note that such a sequence can be encoded by

$$\log_2 \binom{m}{n} + O(1) \leq n\left(\log_2 \frac{m}{n} + O(1)\right) \leq O\left(n\left(b + \log_2 \frac{1}{\varepsilon}\right)\right)$$

random bits, whose decoding can be done in $\mathrm{poly}(m)$ time.

Let $\mathcal{C}$ be the code defined by the random sequence $i_1, \ldots, i_n$ via $\mathcal{C} = \{(u_{i_1} \ldots u_{i_n}) \mid u \in \mathcal{D}\} \subseteq \mathbb{F}_2^n$. Clearly, a generating matrix for $\mathcal{C}$ can be obtained from that of $\mathcal{D}$ in $\mathrm{poly}(m) = \mathrm{poly}(n)$ time. Hence, to prove the theorem it suffices to show that $\mathcal{C}$ satisfies Eq. (4). Let $\mathcal{C}' \subseteq \mathbb{F}_2^n$ be a random n-puncturing of $\mathcal{D}$. Let T be the event that $\mathcal{C}'$ satisfies $\mathcal{P}$. Let J denote the event that no coordinate of $\mathcal{D}$ is sampled more than once for inclusion in $\mathcal{C}'$. Note that $\Pr[J] \geq \left(1 - \frac{n}{m}\right)^n$. By Theorem 1, $\Pr[T] \leq 2^{-(\varepsilon - o(1))n}$. Thus,

$$\Pr[T \mid J] \leq \frac{\Pr[T]}{\Pr[J]} \leq \exp_2\left(\left(-\varepsilon - \log_2\left(1 - \frac{n}{m}\right) + o(1)\right)n\right) \leq 2^{-\Omega(\varepsilon n)} \ ,$$

where the last inequality follows from Eq. (25).

By row-symmetry, $\mathcal{P}$ is invariant to coordinate permutations of $\mathcal{C}$. Observe that a uniformly random coordinate permutation of $\mathcal{C}$ yields a code distributed identically to the distribution of $\mathcal{C}'$ conditioned on the event J. Therefore,

$$\Pr[\mathcal{C} \text{ satisfies } \mathcal{P}] = \Pr[T \mid J] \le 2^{-\Omega(\varepsilon n)} \tag{26}$$

for every $\mathcal{P} \in \mathcal{K}$.

It remains to show that Eq. (26) implies Eq. (4). Let $\mathcal{K}' = (\mathcal{P} \in \mathcal{K} \mid |\mathcal{T}_\mathcal{P}| = 1)$ (recall Fact 6.3 for the definition of $\mathcal{T}_\mathcal{P}$). Observe that a necessary condition for the event in Eq. (4) is that $\mathcal{C}$ satisfies some property in $\mathcal{K}'$. Indeed, suppose that $\mathcal{C}$ satisfies a property $\mathcal{P} \in \mathcal{K}$ and let $\tau \in \mathcal{T}_\mathcal{P}$ such that $\mathcal{C}$ contains a matrix in $\mathcal{M}_{n,\tau}$. Let $\mathcal{P}'$ denote the b-local, row-symmetric and monotone-increasing property for which $\mathcal{T}_{\mathcal{P}'} = \{\tau\}$. Clearly, $\mathcal{C}$ satisfies $\mathcal{P}'$. Since $\mathcal{P}'$ implies $\mathcal{P}$, we have $\mathrm{RLC}(\mathcal{P}') \ge \mathrm{RLC}(\mathcal{P}) \ge R^*$ and so $\mathcal{P}' \in \mathcal{K}'$. Thus, to prove the theorem, it suffices to show that

$$\Pr\left[\mathcal{C} \text{ satisfies some property } \mathcal{P}' \in \mathcal{K}'\right] \le 2^{-\Omega(\varepsilon n)} \ . \tag{27}$$

Now, by Fact 4.10, $|\mathcal{K}'| \le (n+1)^{2^b} \le 2^{o(n)}$. Thus, Eq. (27) follows from Eq. (26) by a union bound on $\mathcal{K}'$, noting that $\mathcal{K}' \subseteq \mathcal{K}$. $\square$

8 Random puncturings of low-bias codes achieve capacity versus memoryless additive noise

Here we prove Theorem 6.

Theorem 6 (A puncturing of a low-bias code achieves capacity with regard to memoryless additive noise)**.** *Fix a prime power q, a distribution ν over $\mathbb{F}_q$ and $0 < \varepsilon < 1$. Let $\mathcal{D} \subseteq \mathbb{F}_q^m$ be an $\frac{\varepsilon}{3(q-1)}$-biased linear code. Let $\mathcal{C}$ be a random n-puncturing of $\mathcal{D}$ of design rate $R \le 1 - H_q(\nu) - \varepsilon$. Then, with probability $1 - q^{-\Omega_\nu(\varepsilon n)}$, it holds for all $x \in \mathcal{C}$ that*

$$\Pr_{z \sim \nu^n}[\text{the MLDU outputs } x \text{ on input } x+z] \ge 1 - 2q^{-c_\nu \varepsilon^2 n} \ , \tag{6}$$

for some $c_\nu > 0$ that depends only on ν.

We require the following lemma.

Lemma 8.1. *Let $C \subseteq \mathbb{F}_q^n$ be a random rate R puncturing of an η-biased linear code. Fix $x \in \mathbb{F}_q^n \setminus \{0\}$. Then,*

$$\Pr_{\mathcal{C}}[x \in \mathcal{C}] \le q^{n\cdot(-1+R+(q-1)\eta)} \ .$$

Proof. Denote the mother code by $\mathcal{D} \subseteq \mathbb{F}_q^m$ and let $\varphi : \mathbb{F}_q^m \to \mathbb{F}_q^n$ be the puncturing map. By the union bound,

$$\begin{aligned}
\Pr[x \in \mathcal{C}] &\le \sum_{u \in \mathcal{D}} \Pr[\varphi(u) = x] = \sum_{u \in \mathcal{D} \setminus \{0\}} \prod_{i=1}^{n} \Pr[\varphi_i(u) = x_i] \\
&= \sum_{u \in \mathcal{D} \setminus \{0\}} \prod_{i=1}^{n} \frac{|\{j \in [m] \mid u_j = x_i\}|}{m} \\
&\le \sum_{u \in \mathcal{D} \setminus \{0\}} \prod_{i=1}^{n} \left(1 - \frac{q-1}{q}(1-\eta)\right) && \text{(by Lemma 4.12)} \\
&= \sum_{u \in \mathcal{D} \setminus \{0\}} \left(\frac{1+(q-1)\eta}{q}\right)^{n} \\
&\le q^{Rn} \left(\frac{1+(q-1)\eta}{q}\right)^{n} \le q^{n \cdot (-1+R+(q-1)\eta)} \ .
\end{aligned}$$

Proof of Theorem 6. Let J denote the event that our code $\mathcal{C}$ is MLDU-decodable with error probability at most $2q^{-c_\nu \varepsilon^2 n}$ with regard to the ν-memoryless additive noise channel (where $c_\nu > 0$ shall be chosen later). In other words, J means that Eq. (6) holds for all $x \in \mathcal{C}$. By linearity of $\mathcal{C}$, a sufficient condition for the latter is that Eq. (6) holds for $x = 0$. By the definition of the MLDU, the codeword 0 with noise vector z is decoded correctly whenever $\nu(z) > \nu(z-x)$ for all $x \in \mathcal{C} \setminus \{0\}$. Thus,

$$\text{The code } \mathcal{C} \text{ satisfies } J \quad \text{if} \quad \Pr_{z \sim \nu^n}[\forall x \in \mathcal{C} \setminus \{0\} \ \ \nu(z) > \nu(z-x)] \ge 1 - 2q^{-c_\nu \varepsilon^2 n} \ . \tag{28}$$

Let $z \sim \nu^n$ and let M_z denote the event that z belongs to an $\frac{\varepsilon}{3}$-typical-set, namely, $q^{-H_q(\nu)n - \frac{\varepsilon n}{3}} \le \nu(z) \le q^{-H_q(\nu)n + \frac{\varepsilon n}{3}}$. It is well known (e.g., [YM12]) that

$$\Pr_{z \sim \nu^n}[M_z] \ge 1 - q^{-c'_\nu \varepsilon^2 n} \ ,$$

for some positive constant c'_ν.

Denote $E_z = \left\{x \in \mathbb{F}_q^n \setminus \{0\} \mid \nu(z) \le \nu(z-x)\right\}$. Since $\sum_{x \in \mathbb{F}_q^n} \nu(z-x) = \sum_{x \in \mathbb{F}_q^n} \nu(x) = 1$, we have $|E_z| \le \frac{1}{\nu(z)}$. Now, for a fixed code $\mathcal{C}$,

$$\begin{aligned}
\Pr_{z \sim \nu^n}[\exists x \in \mathcal{C} \setminus \{0\} \ \ \nu(z) \le \nu(z-x)] &= \Pr_{z \sim \nu^n}[E_z \cap \mathcal{C} \neq \emptyset] \\
&\le \Pr_{z \sim \nu^n}[E_z \cap \mathcal{C} \neq \emptyset \mid M_z] + \Pr_{z \sim \nu^n}\left[\overline{M_z}\right] \\
&\le \Pr_{z \sim \nu^n}[E_z \cap \mathcal{C} \neq \emptyset \mid M_z] + q^{-c'_\nu \varepsilon^2 n} \ .
\end{aligned} \tag{29}$$

Let $\eta = \frac{\varepsilon}{3(q-1)}$. By Lemma 8.1, for any z such that M_z holds, we have

$$
\begin{aligned}
\Pr_{\mathcal{C}}\left[E_z \cap \mathcal{C} \neq \emptyset\right] &\leq \sum_{x \in E_z} \Pr_{\mathcal{C}}\left[x \in \mathcal{C}\right] \\
&\leq \sum_{x \in E_z} q^{n\cdot(-1+R+(q-1)\eta)} \leq |E_z| \cdot q^{n\cdot(-1+R+(q-1)\eta)} \\
&\leq \frac{1}{\nu(z)} \cdot q^{n\cdot(-1+R+(q-1)\eta)} \\
&\leq q^{n\cdot\left(-1+R+(q-1)\eta+H_q(\nu)+\frac{\varepsilon}{3}\right)} \\
&\leq q^{-\frac{\varepsilon n}{3}} \quad .
\end{aligned} \tag{30}
$$

Now,

$$
\begin{aligned}
&\Pr_{\mathcal{C}}\left[\mathcal{C} \text{ does not satisfy } J\right] \\
&\leq \Pr_{\mathcal{C}}\left[\Pr_{z\sim\nu^n}\left[\forall x \in \mathcal{C}\setminus\{0\} \quad \nu(z) > \nu(z-x)\right] < 1 - 2q^{-c_\nu \varepsilon^2 n}\right] && \text{(by Eq. (28))} \\
&= \Pr_{\mathcal{C}}\left[\Pr_{z\sim\nu^n}\left[\exists x \in \mathcal{C}\setminus\{0\} \quad \nu(z) \leq \nu(z-x)\right] \geq 2q^{-c_\nu \varepsilon^2 n}\right] \\
&\leq \Pr_{\mathcal{C}}\left[\Pr_{z\sim\nu^n}\left[E_z \cap \mathcal{C} \neq \emptyset \mid M_z\right] \geq 2q^{-c_\nu \varepsilon^2 n} - q^{-c'_\nu \varepsilon^2 n}\right] && \text{(by Eq. (29))} \\
&\leq \frac{\mathbb{E}_{\mathcal{C}}\left[\Pr_{z\sim\nu^n}\left[E_z \cap \mathcal{C} \neq \emptyset \mid M_z\right]\right]}{2q^{-c_\nu \varepsilon^2 n} - q^{-c'_\nu \varepsilon^2 n}} && \text{(by Markov's inequality)} \\
&\leq \frac{q^{-\frac{\varepsilon n}{3}}}{2q^{-c_\nu \varepsilon^2 n} - q^{-c'_\nu \varepsilon^2 n}} \quad . && \text{(by Eq. (30))}
\end{aligned}
$$

Taking $c_\nu = \min\left\{c_{\nu'}, \frac{1}{4}\right\}$ makes the right-hand side $q^{-\Omega_\nu(\varepsilon n)}$, finishing the proof. □

Acknowledgments

The authors thank João Ribeiro for pointing out an improvement to Theorem 3, and Zeyu Guo for finding an error in a previous version of this paper.

References

[ABNNR92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. "Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs". In: *IEEE Trans. Inf. Theory* 38.2 (1992), pp. 509–516. DOI: `10.1109/18.119713`. URL: `https://doi.org/10.1109/18.119713`.

[CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. "Restricted Isometry of Fourier Matrices and List Decodability of Random Linear Codes". In: *SIAM J. Comput.* 42.5 (2013), pp. 1888–1914. DOI: `10.1137/120896773`. URL: `https://doi.org/10.1137/120896773`.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)* Wiley, 2006. ISBN: 978-0-471-24195-9. URL: http://www.elementsofinformationtheory.com/.

[FKS20] Asaf Ferber, Matthew Kwan, and Lisa Sauermann. "List-decodability with large radius for Reed-Solomon codes". In: *CoRR* abs/2012.10584 (2020). To appear in FOCS 2021. arXiv: 2012.10584. URL: https://arxiv.org/abs/2012.10584.

[GHK11] Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. "On the List-Decodability of Random Linear Codes". In: *IEEE Trans. Inf. Theory* 57.2 (2011), pp. 718–725. DOI: 10.1109/TIT.2010.2095170. URL: https://doi.org/10.1109/TIT.2010.2095170.

[GLMRSW20] 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: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*. Ed. by Jaroslaw Byrka and Raghu Meka. Vol. 176. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, 9:1–9:21. DOI: 10.4230/LIPIcs.APPROX/RANDOM.2020.9. URL: https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2020.9.

[GLSTW20] Zeyu Guo, Ray Li, Chong Shangguan, Itzhak Tamo, and Mary Wootters. "Improved List-Decodability of Reed-Solomon Codes via Tree Packings". In: *CoRR* abs/2011.04453 (2020). To appear in FOCS 2021. arXiv: 2011.04453. URL: https://arxiv.org/abs/2011.04453.

[GMRSW21] Venkatesan Guruswami, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. "Sharp Threshold Rates for Random Codes". In: *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference*. Ed. by James R. Lee. Vol. 185. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, 5:1–5:20. DOI: 10.4230/LIPIcs.ITCS.2021.5. URL: https://doi.org/10.4230/LIPIcs.ITCS.2021.5.

[GN14] Venkatesan Guruswami and Srivatsan Narayanan. "Combinatorial Limitations of Average-Radius List-Decoding". In: *IEEE Trans. Inf. Theory* 60.10 (2014), pp. 5827–5842. DOI: 10.1109/TIT.2014.2343224. URL: https://doi.org/10.1109/TIT.2014.2343224.

[Gol11] Oded Goldreich. "Three XOR-Lemmas - An Exposition". In: *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*. Ed. by Oded Goldreich. Vol. 6650. Lecture Notes in Computer Science. Springer, 2011, pp. 248–272. DOI: 10.1007/978-3-642-22670-0_22. URL: https://doi.org/10.1007/978-3-642-22670-0%5C_22.

[GR11] Venkatesan Guruswami and Atri Rudra. "Soft Decoding, Dual BCH Codes, and Better List-Decodable varepsilon-Biased Codes". In: *IEEE Trans. Inf. Theory* 57.2 (2011), pp. 705–717. DOI: 10.1109/TIT.2010.2095193. URL: https://doi.org/10.1109/TIT.2010.2095193.

[GRS] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential coding theory.* Draft available at http://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/.

[GS99] Venkatesan Guruswami and Madhu Sudan. "Improved decoding of Reed-Solomon and algebraic-geometry codes". In: *IEEE Trans. Inf. Theory* 45.6 (1999), pp. 1757–1767. DOI: 10.1109/18.782097. URL: https://doi.org/10.1109/18.782097.

[GST21] Eitan Goldberg, Chong Shangguan, and Itzhak Tamo. "List-decoding and list-recovery of Reed-Solomon codes beyond the Johnson radius for any rate". In: *CoRR* abs/2105.14754 (2021). arXiv: 2105.14754. URL: https://arxiv.org/abs/2105.14754.

[Gur06] Venkatesan Guruswami. "Algorithmic Results in List Decoding". In: *Found. Trends Theor. Comput. Sci.* 2.2 (2006). DOI: 10.1561/0400000007. URL: https://doi.org/10.1561/0400000007.

[LW21] Ray Li and Mary Wootters. "Improved List-Decodability of Random Linear Binary Codes". In: *IEEE Trans. Inf. Theory* 67.3 (2021), pp. 1522–1536. DOI: 10.1109/TIT.2020.3041650. URL: https://doi.org/10.1109/TIT.2020.3041650.

[MRRSW20] Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. "LDPC Codes Achieve List Decoding Capacity". In: *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020.* IEEE, 2020, pp. 458–469. DOI: 10.1109/FOCS46700.2020.00050. URL: https://doi.org/10.1109/FOCS46700.2020.00050.

[Res20] Nicolas Resch. "List-Decodable Codes: (Randomized) Constructions and Applications". PhD thesis. 2020. URL: http://reports-archive.adm.cs.cmu.edu/anon/2020/CMU-CS-20-113.pdf.

[RW14a] Atri Rudra and Mary Wootters. "Every list-decodable code for high noise has abundant near-optimal rate puncturings". In: *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*. Ed. by David B. Shmoys. ACM, 2014, pp. 764–773. DOI: 10.1145/2591796.2591797. URL: https://doi.org/10.1145/2591796.2591797.

[RW14b] Atri Rudra and Mary Wootters. "It'll probably work out: improved list-decoding through random operations". In: *Electronic Colloquium on Computational Complexity (ECCC)* 21 (2014), p. 104. URL: http://eccc.hpi-web.de/report/2014/104.

[RW18] Atri Rudra and Mary Wootters. "Average-radius list-recovery of random linear codes". In: *Proceedings of the 2018 ACM-SIAM Symposium on Discrete Algorithms, SODA*. 2018.

[ST20] Chong Shangguan and Itzhak Tamo. "Combinatorial list-decoding of Reed-Solomon codes beyond the Johnson radius". In: *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020.* Ed. by Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy. ACM, 2020, pp. 538–551. DOI: 10.1145/3357713.3384295. URL: https://doi.org/10.1145/3357713.3384295.

[Ta-17] Amnon Ta-Shma. “Explicit, almost optimal, epsilon-balanced codes”. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*. Ed. by Hamed Hatami, Pierre McKenzie, and Valerie King. ACM, 2017, pp. 238–251. DOI: 10.1145/3055399.3055408. URL: https://doi.org/10.1145/3055399.3055408.

[Woo13] Mary Wootters. “On the list decodability of random linear codes with large error rates”. In: *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*. 2013, pp. 853–860. DOI: 10.1145/2488608.2488716. URL: http://doi.acm.org/10.1145/2488608.2488716.

[YM12] En-Hui Yang and Jin Meng. “Non-asymptotic equipartition properties for independent and identically distributed sources”. In: *2012 Information Theory and Applications Workshop*. 2012, pp. 39–46. DOI: 10.1109/ITA.2012.6181822.

[ZP81] Victor Vasilievich Zyablov and Mark Semenovich Pinsker. “List Concatenated Decoding”. In: *Problemy Peredachi Informatsii* 17.4 (1981), pp. 29–33.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il