Bounds for List-Decoding and List-Recovery of Random Linear Codes

Venkatesan Guruswami
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
venkatg@cs.cmu.edu

Ray Li
Department of Computer Science, Stanford University, CA, USA
rayyli@stanford.edu

Jonathan Mosheiff
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
jmosheif@cs.cmu.edu

Nicolas Resch
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
nresch@cs.cmu.edu

Shashwat Silas
Department of Computer Science, Stanford University, CA, USA
silas@stanford.edu

Mary Wootters
Department of Computer Science, Stanford University, CA, USA
marykw@stanford.edu

Abstract

A family of error-correcting codes is list-decodable from error fraction p if, for every code in the family, the number of codewords in any Hamming ball of fractional radius p is less than some integer L that is independent of the code length. It is said to be list-recoverable for input list size ℓ if for every sufficiently large subset of codewords (of size L or more), there is a coordinate where the codewords take more than ℓ values. The parameter L is said to be the "list size" in either case. The capacity, i.e., the largest possible rate for these notions as the list size $L \to \infty$, is known to be $1 - h_q(p)$ for list-decoding, and $1 - \log_q \ell$ for list-recovery, where q is the alphabet size of the code family.

In this work, we study the list size of *random linear codes* for both list-decoding and list-recovery as the rate approaches capacity. We show the following claims hold with high probability over the choice of the code (below q is the alphabet size, and $\varepsilon > 0$ is the gap to capacity).

- A random linear code of rate $1 - \log_q(\ell) - \varepsilon$ requires list size $L \geq \ell^{\Omega(1/\varepsilon)}$ for list-recovery from input list size ℓ. This is surprisingly in contrast to completely random codes, where $L = O(\ell/\varepsilon)$ suffices w.h.p.
- A random linear code of rate $1 - h_q(p) - \varepsilon$ requires list size $L \geq \lfloor h_q(p)/\varepsilon + 0.99 \rfloor$ for list-decoding from error fraction p, when ε is sufficiently small.
- A random *binary* linear code of rate $1 - h_2(p) - \varepsilon$ is list-decodable from *average* error fraction p with list size with $L \leq \lfloor h_2(p)/\varepsilon \rfloor + 2$. (The average error version measures the average Hamming distance of the codewords from the center of the Hamming ball, instead of the maximum distance as in list-decoding.)

The second and third results together precisely pin down the list sizes for binary random linear codes for both list-decoding and average-radius list-decoding to three possible values.

Our lower bounds follow by exhibiting an explicit subset of codewords so that this subset – or some symbol-wise permutation of it – lies in a random linear code with high probability. This uses a recent characterization of (Mosheiff, Resch, Ron-Zewi, Silas, Wootters, 2019) of configurations of codewords that are contained in random linear codes. Our upper bound follows from a refinement of the techniques of (Guruswami, Håstad, Sudan, Zuckerman, 2002) and strengthens a previous result of (Li, Wootters, 2018), which applied to list-decoding rather than average-radius list-decoding.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020).
Editors: Jarosław Byrka and Raghu Meka; Article No. 9; pp. 9:1–9:21

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

2012 ACM Subject Classification Mathematics of computing → Coding theory

Keywords and phrases list-decoding, list-recovery, random linear codes, coding theory

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.9

Category RANDOM

Related Version A full version of the paper is available at `https://arxiv.org/abs/2004.13247`.

Funding *Venkatesan Guruswami*: Partially funded by NSF grants CCF-1563742 and CCF-1814603.
Ray Li: Partially funded by NSF-CAREER grant CCF-1844628, NSF-BSF grant CCF-1814629, a Sloan Research Fellowship and NSF-GRFP grant DGE-1656518.
Jonathan Mosheiff: Partially funded by NSF grants CCF-1563742 and CCF-1814603.
Nicolas Resch: Partially funded by NSF grants CCF-1563742, CCF-1814603, CCF-1527110, CCF-1618280, CCF-1814603, CCF-1910588, NSF CAREER award CCF-1750808 and a Sloan Research Fellowship.
Shashwat Silas: Partially funded by NSF-CAREER grant CCF-1844628, NSF-BSF grant CCF-1814629, a Sloan Research Fellowship and a Google Graduate Fellowship.
Mary Wootters: Partially funded by NSF-CAREER grant CCF-1844628, NSF-BSF grant CCF-1814629 and a Sloan Research Fellowship.

1 Introduction

In coding theory, one is interested in the combinatorial properties of sets $\mathcal{C} \subseteq \mathbb{F}_q^n$.[1] Such a set $\mathcal{C}$ is called a *code* of *length* n over the *alphabet* $\mathbb{F}_q$, and the elements $c \in \mathcal{C}$ are called *codewords*.

List-decoding, introduced by Elias and Wozencraft in the 1950's [10, 44], is such a combinatorial property. For $p \in [0,1]$ and integer $L \geq 1$, we say that a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is (p, L)-*list-decodable* if, for all $z \in \mathbb{F}_q^n$,

$$|\{c \in \mathcal{C} : \delta(c, z) \leq p\}| < L,$$

where $\delta(x, y) = \frac{1}{n}|\{i : x_i \neq y_i\}|$ denotes relative Hamming distance. That is, $\mathcal{C}$ is list-decodable if not too many codewords of $\mathcal{C}$ live in any small enough Hamming ball. In this paper, we are interested in the trade-offs between p, L, and the rate of the code $\mathcal{C}$. The *rate* R of $\mathcal{C}$ is defined as $R = \frac{\log_q |\mathcal{C}|}{n}$. The rate lies in the interval $[0, 1]$, and larger is better.

Variations of list-decoding

In this work, we consider standard list-decoding along with two variations.

The first variation is a strengthening of list-decoding known as *average-radius list-decoding*. A code $\mathcal{C}$ is (p, L)-*average-radius list-decodable* if for any set $\Lambda \subseteq \mathcal{C}$ of size L and $z \in \mathbb{F}_q^n$,

$$\frac{1}{L}\sum_{c \in \Lambda} \delta(c, z) \geq p.$$

[1] Here and throughout the paper, $\mathbb{F}_q$ denotes the finite field with q elements. In this we work only consider linear codes, so we always assume that the alphabet is a finite field.

It is not hard to see that (p, L)-average-radius list-decodability implies (p, L)-list-decodability, and this stronger formulation has led to stronger lower bounds than are achievable otherwise [22]. In addition to stronger lower bounds, average-radius list-decoding – essentially replacing a maximum with an average in the definition of list-decoding – is a natural concept, and it has helped establish connections between list-decoding and compressed sensing [7].

The second variation, known as *list-recovery,* is a version where the "noise" is replaced by uncertainty about each symbol of the received word z. Formally, we say that a code $\mathcal{C}$ is (ℓ, L)-*list-recoverable* if for any sets $S_1, \ldots, S_n \subseteq \mathbb{F}_q$ with $|S_i| \leq \ell$ for all i,

$$|\{c \in \mathcal{C} : c_i \in S_i \, \forall i\}| < L.$$

List-recovery was originally used as a stepping-stone to list-decoding and unique-decoding (e.g., [17, 18, 19, 20]) but it has since become a useful primitive in its own right, with applications beyond coding theory [31, 38, 13, 28, 8].

Pinning down the output list size

We are motivated by the problem of pinning down the output list size L for (average-radius) list-decoding and for list-recovery. For all three of these problems, given q and p (respectively, q and ℓ), there exists an *optimal rate*, denoted R^*. Namely, R^* is the largest rate so that, for any $\varepsilon > 0$, there are q-ary codes of rate $R^* - \varepsilon$ and arbitrarily large length, which are (p, L)-(average-weight)-list-decodable (resp. (ℓ, L)-list-recoverable), for some $L(q, p, \varepsilon)$ (resp. $L(q, \ell, \varepsilon)$). Importantly, L must not depend on the length of the code. The *list-decoding capacity theorem* gives the dependence of R^* on q and p (resp. q and ℓ): $R^* = 1 - h_q(p)$ for both standard and average-radius list-decoding, [11, 45] and $R^* = 1 - \log_q(\ell)$ for list-recovery (e.g., [42]).

We are interested in the trade-off between the list size L, the parameters p, q, ℓ of the problem, and this gap ε; we refer to ε as the *gap to capacity.* Pinning down the list size L is an important problem. For example, for many of the algorithmic applications within coding theory, the list size represents a bottleneck on the running time of an algorithm that must check each item in the list before pruning it down [20, 9, 26, 27, 21]. For applications in pseudorandomness, for example to expanders or extractors, the list size corresponds to the expansion or to the amount of entropy in the input, respectively, and it is of interest to precisely pin down these quantities.

We make progress on pinning down the output list sizes for the case of *random linear codes.* A random linear code is a uniformly random subspace of $\mathbb{F}_q^n$ of certain dimension. The list-decodability of random linear codes has been well studied for many reasons [45, 15, 7, 43, 40, 42, 34]. First, it is a natural mathematical question that studies the interplay between two fundamental notions in $\mathbb{F}_q^n$: subspaces and Hamming balls. Second, there are constructions of codes which use random linear codes (and their list-decodability) as a building block [20, 24, 30, 29], and improvements in the parameters of random linear codes will lead to improvements in these constructions as well. Third, random linear codes can be seen as one way to partially derandomize completely random codes; this is especially motivating in the binary (or fixed alphabet) case, where we do not know of any explicit constructions of optimally list-decodable codes, linear or otherwise.

1.1 Contributions

Our main results are improved bounds on the list size of random linear codes. We defer the formal theorem statements until after we have set up notation, but we informally summarize our results here. Below, we consider codes of *rate* $R^* - \varepsilon$, where as above we use R^* to denote best achievable rate for each particular problem.

(1) **Lower bound on the list size for list-recovery of random linear codes.** We show that if a random linear code of rate $R^* - \varepsilon$ is list-recoverable with high probability with input list sizes ℓ and output list size L, then we must have $L = \ell^{\Omega(1/\varepsilon)}$. This is in contrast to completely random codes, for which the output list size is $L = O(\ell/\varepsilon)$ with high probability.
This gap between random linear codes and completely random codes demonstrates that in some sense zero-error list-recovery behaves more like *erasure-list-decoding* [14] than it does like list-decoding with errors. Such a gap is present between general and linear codes in erasure list-decoding, but as we see below, there is no such gap for list-decoding from errors.
Our result extends to the setting of list-recovery with erasures as well. The formal theorem statement and proof can be found in Section 3.

(2) **Better lower bounds on the list size for list-decoding random linear codes.** We show that if a q-ary random linear code of rate $R^* - \varepsilon$ is list-decodable with high probability up to radius p with an output list size of L, then we must have $L \geq \left\lfloor \frac{h_q(p)}{\varepsilon} + 0.99 \right\rfloor$. By [34], this result is tight for list-decoding of binary random linear codes up to a small additive factor. As an immediate corollary, $L \geq \left\lfloor \frac{h_q(p)}{\varepsilon} + 0.99 \right\rfloor$ for *average-radius* list-decoding of random linear codes as well, and, as we will see below, this is also tight for binary random linear codes, up to a small additive factor. We conjecture that the leading constant $h_q(p)$ is also correct for $q > 2$.
Previous work [22] has established that $L = \Omega(1/\varepsilon)$, but to the best of our knowledge this is the first work that pins down the leading constant. In particular, [22] shows that, in the situation above, we have $L \geq c_{p,q}/\varepsilon$, where $c_{p,q}$ is a constant that goes to zero as p goes to $1 - 1/q$. In contrast, we show below that the leading constant is at least $h_q(p)$, which goes to 1 as p goes to $1 - 1/q$.
The formal theorem statement and proof can be found in Section 4.

(3) **Completely pinning down the list size for average-radius list-decoding of binary random linear codes.** We prove a new upper bound on the *average-radius* list-decodability of binary random linear codes, which matches our lower bound, even up to the leading constant. More precisely, we show that with high probability, a random binary linear code of rate $R^* - \varepsilon$ is *average-radius* list-decodable up to radius p with $L \leq \lfloor h_2(p)/\varepsilon \rfloor + 2$.
Such a bound was known for standard list-decoding [34], but our upper bound holds even for the stronger notion of average-radius list-decoding, and improves the additive constant by 1.[2] In particular, this shows that for both list-decoding and average-radius list-decoding of binary random linear codes, the best possible L is concentrated on at most three values: $\lfloor h_q(p)/\varepsilon \rfloor + 2$, $\lfloor h_q(p)/\varepsilon \rfloor + 1$ and $\lfloor h(p)/\varepsilon + 0.99 \rfloor$. This tight concentration demonstrates the sharpness of our upper and lower bound techniques.
The formal theorem statement and proof can be found in Section A.

1.2 Overview of techniques

In this section, we give a brief overview of our techniques.

Lower bounds

To illustrate the techniques for our lower bounds, we warm up with a back-of-the-envelope calculation which suggests why the "right" answer for our result (1) above is $\ell^{\Omega(1/\varepsilon)}$.

[2] Under our definition of list-decoding, [34] show (p, L) list-decodability with $L = \lfloor h(p)/\varepsilon \rfloor + 3$.

Consider a random linear code $\mathcal{C} \subset \mathbb{F}_q^n$, of rate $R = 1 - \log_q(\ell) - \varepsilon$, where $\varepsilon \in (0, \frac{1}{2})$. That is, $\mathcal{C}$ is the kernel[3] of a uniformly random matrix sampled from $\mathbb{F}_q^{(1-R)n \times n}$. Suppose that ℓ is a prime power, and $q = \ell^t$ for some $t \geq 2$. Thus, $\mathbb{F}_\ell$ is a sub-field of $\mathbb{F}_q$. Let D be an integer slightly smaller than $\frac{1}{2\varepsilon}$ and let $L = \ell^D \geq \ell^{\Omega(1/\varepsilon)}$. We claim that $\mathcal{C}$ is unlikely to be (ℓ, L)-list-recoverable.

Given a matrix $M \in \mathbb{F}_q^{n \times D}$, we write $M \subseteq \mathcal{C}$ ("$\mathcal{C}$ contains M") to mean that each of the columns of M is a codeword in $\mathcal{C}$.

Let $\mathcal{M}$ denote the set of all full-rank matrices $M \in \mathbb{F}_q^{n \times D}$ that have the following property: for every row M_i of M there exists some $x_i \in \mathbb{F}_q^*$ such that all entries of M_i belong to the set $x_i \cdot \mathbb{F}_\ell$.

We will show that $\mathcal{M}$ is *bad* and *abundant*. By *bad* we mean that a linear code containing a matrix from $\mathcal{M}$ cannot be (ℓ, L)-list-recoverable. We say that $\mathcal{M}$ is *abundant* (for the rate R) if a random linear code of rate R is likely to contain at least one matrix from $\mathcal{M}$. Clearly, the combination of these properties means that $\mathcal{C}$ is unlikely to be (ℓ, L)-list-recoverable.

We first prove that $\mathcal{M}$ is bad. Assume that $\mathcal{C}$ contains some matrix $M \in \mathcal{M}$. By linearity of the code, $\mathcal{C}$ also contains every vector of the form Mu, for $u \in \mathbb{F}_q^D$. In particular, consider the set of vectors $B := \left\{Mu \mid u \in \mathbb{F}_\ell^D\right\} \subseteq \mathcal{C}$. Observe that $\mathcal{C}$ cannot be (ℓ, L)-list-recoverable, since B is a "bad list" for list-recoverability with these parameters: First, since M has full-rank, B is of cardinality $\ell^D = L$. Now, given $i \in [n]$, we need to show that there exists a subset $S_i \subseteq \mathbb{F}_q$ with $|S_i| = \ell$, such that $v_i \in S_i$ for all $v \in B$. We take S_i to be the set $x_i \cdot \mathbb{F}_\ell$, which contains all entries of the row M_i. For $j \in [D]$, write $M_{i,j} = x_i \cdot w_j$ ($w_j \in \mathbb{F}_\ell$), and let $v = Mu$ for some $u \in \mathbb{F}_\ell^D$. Then

$$v_i = \sum_{j=1}^{D} M_{i,j} u_j = x_i \cdot \left(\sum_{j=1}^{D} w_j \cdot u_j \right) \in x_i \cdot \mathbb{F}_\ell,$$

and we conclude that $\mathcal{M}$ is bad.

Showing that $\mathcal{M}$ is abundant is harder, and at this stage we only provide some intuition for this fact. Let us compute the expected number of matrices $M \in \mathcal{M}$ that are contained in $\mathcal{C}$. First, we estimate the cardinality of $\mathcal{M}$. One may generate a matrix in $\mathcal{M}$ by choosing each of its rows in an essentially independent fashion.[4] Choosing a row amounts to choosing one of $\frac{q-1}{\ell-1}$ sets of the form $x \cdot \mathbb{F}_\ell$ ($x \in \mathbb{F}_q^*$) and then taking each entry to be an element of that set. Accounting for multiple counting of the all-zero row, the number of possible rows is thus $\frac{q-1}{\ell-1} \cdot (\ell^D - 1) + 1$, which we approximate as $q \cdot \ell^{D-1}$. Thus, $|\mathcal{M}| \approx (q \cdot \ell^{D-1})^n$. Next, it is not hard to see that a random linear code contains a given matrix of rank r with probability $q^{-(1-R) \cdot r \cdot n}$. Consequently, for $M \in \mathcal{M}$, we have $\Pr[M \subseteq \mathcal{C}] = q^{-(1-R)Dn}$. Therefore,

$$\begin{aligned} \mathbb{E}\left|\{M \in \mathcal{M} : M \subseteq \mathcal{C}\}\right| &= |\mathcal{M}| q^{-(1-R)Dn} \approx \left(q \cdot \ell^{D-1}\right)^n \cdot q^{-(1-R)Dn} \\ &= \left(\ell^{-1} \cdot q^{1-\varepsilon D}\right)^n = \ell^{(-1+t(1-\varepsilon D))n}, \end{aligned}$$

where, the penultimate equality is due to substituting $1 - \log_q(\ell) - \varepsilon$ for R. Finally, since $t \geq 2$ and $D < \frac{1}{2\varepsilon}$, the right-hand side of the above is $\ell^{\Omega(n)}$. Thus, in expectation, $\mathcal{C}$ contains many "bad" lists for list-recovery.

---

[3] This is one of several natural models for a random linear code. Another possible model is taking a uniformly random subspace of dimension Rn. It is not hard to see that the total variation distance between these distributions is exponentially small. In particular, our model yields a code of dimension exactly Rn with probability $1 - \exp(-\Omega(n))$.

[4] We say "essentially" since the resulting matrix might not have full rank, but this happens with negligibly small probability.

Of course, this back-of-the-envelope calculation does *not* yield the result advertised above. It might be the case that, even though the expected number of $M \in \mathcal{M}$ so that $M \subseteq \mathcal{C}$ is large, the probability that such an M exists is still small. In fact, as [37] shows, there are simple examples where this does happen. Thus, proving that $\mathcal{M}$ is abundant requires more work.

A standard approach to show that $\mathcal{M}$ is abundant would be via the second-moment method. Recently, [37] gave a general theorem which encompasses second-moment calculations in this context. In particular, they showed that there is essentially only one reason that a set $\mathcal{M}$ might not be abundant: there exists some matrix $A \in \mathbb{F}_q^{D \times D'}$, such that the set $\{MA \mid M \in \mathcal{M}\}$ is small. If this occurs, we say that $\mathcal{M}$ is *implicitly rare*.[5] They used this result to study the list-decodability of random Low-Density Parity-Check codes, but we can use their result to do our second moment calculation. We show that our example of $\mathcal{M}$ above[6] is not implicitly rare, by showing that there is no such linear map A. This establishes that the back-of-the-envelope calculation is in fact correct. Appealing to the machinery of [37], rather than applying the second moment method from scratch, allows us to get tighter constants with slightly less work, and gives a more principled approach to our lower bounds; indeed, our result (2) follows the same outline.

The intuition for our second result (2) is similar: we give an example of a class $\mathcal{M}$ which is *bad* for list-decoding and *abundant.* We define $\mathcal{M}$ as follows: Let $u \in \mathbb{F}_q^D$ be a random vector with independent $\text{Bernoulli}_q(p)$ entries, namely, each entry is 0 with probability $1-p$, and chosen uniformly from $\mathbb{F}_q^*$ with probability p. Let x be uniformly sampled from $\mathbb{F}_q$. Let τ denote the distribution (over $\mathbb{F}_q^D$) of the random vector $u + x \cdot 1_D$. Finally, define $\mathcal{M}$ to be the set of all matrices $M \in \mathbb{F}_q^{n \times D}$, such that a uniformly sampled row of M has the distribution τ. As before, we show that $\mathcal{M}$ is abundant by showing that $\mathcal{M}$ is not implicitly rare and using the result of [37].

Upper bounds

Our argument for our upper bound result (3) closely follows that of [34], which itself builds on the argument of [16]. The argument imagines building the random linear code one dimension at a time and uses a potential function to show that, so long as we do not add too many dimensions, no ball intersects the code too much. We now provide an informal overview of our approach, specifically comparing and contrasting it with the arguments of Guruswami, Håstad, Sudan and Zuckerman [16]; and Li and Wootters [34].

Let $R = 1 - h(\rho) - \varepsilon$ and put $k := Rn$ (which we assume for exposition is an integer). Note that sampling a random linear code of rate R is the same as sampling $b_1, \ldots, b_k \in \mathbb{F}_2^n$ independently and uniformly at random and outputting $\text{span}\{b_1, \ldots, b_k\}$. Consider the "intermediate" codes $\mathcal{C}_i = \text{span}\{b_1, \ldots, b_i\}$; [34] (following [16]) define a potential function $S_{\mathcal{C}_i}$ and endeavor to show that $S_{\mathcal{C}_i}$ does not grow too quickly. The work [16] demonstrated that this holds in expectation; the work [34] improved their argument to show that it holds with high probability. In both cases the potential function is such that it is easy to show that, so long as $S_{\mathcal{C}}$ is $O(1)$, the code $\mathcal{C}$ is list-decodable.

The potential function in these works keeps track of the radius p list-size at each vector $x \in \mathbb{F}_2^n$, that is, the cardinalities $|\{c \in \mathcal{C}_i : \delta(x, c) \leq p\}|$ for $i = 1, \ldots, k$, and shows that so long as i is not too large all these cardinalities remain at most L. For average-radius list-decoding, we instead keep track of a sort of "weighted" list size, where codewords that

[5] The term "implicitly rare" is used by the first version of [37], available at `https://arxiv.org/abs/1909.06430v1`.

[6] More precisely, we study an example similar to this one; the example above was slightly tweaked to simplify the exposition for this back-of-the-envelope explanation.

are very close x are weighted more heavily. We can reuse much of the analysis from [34] to demonstrate that on the k-th step the potential function is still bounded by a constant (in fact, it is at most 2). The real novelty in our argument is a demonstration that, assuming this potential function is small, the code is indeed (p, L)-average-radius list-decodable. This step is more involved than the argument in [16, 34] to establish (p, L)-list-decodability.

1.3 Related work

We now highlight some related work. In what follows, ε is always the "gap-to-capacity", i.e., if the capacity for a particular problem is R^*, then the result concerns codes of rate $R^* - \varepsilon$.

Lower bounds for list sizes of arbitrary codes

It is known that a typical (i.e., uniformly random) list-decodable code of rate $R^* - \varepsilon$ has list size $L = \Theta(1/\varepsilon)$, and a natural question to ask is whether *every* code requires a list of size $L = \Omega(1/\varepsilon)$. Blinovsky ([5, 4]) showed that lists of size $\Omega_p(\log(1/\varepsilon))$ are necessary for list-decoding a code of rate $R^* - \varepsilon$. Later, Guruswami and Vadhan [25] considered the high-noise regime where $p = 1 - 1/q - \eta$ and showed that lists of size $\Omega_q(1/\eta^2)$ are necessary. Finally, Guruswami and Narayanan [22] showed that for *average-radius* list-decoding, the list size must be $\Omega_p(1/\sqrt{\varepsilon})$.

Existing lower bounds for random linear codes

For the special case of random linear codes, Guruswami and Narayanan [22] showed that lists of size $c_{p,q}/\varepsilon$ are necessary. The constant $c_{p,q}$ is not explicitly computed (and in fact relies on a constant from [15] which we discuss below), but one can deduce from the proof that if p tends to $1 - 1/q$ then $c_{p,q}$ will tend to 0. Their lower bound follows from a second moment method argument, i.e., they consider a certain random variable X whose positivity is equivalent to the failure of a random linear code to be list-decodable, and then show that $\mathrm{Var}(X) = o(\mathbb{E}[X])^2$. In this sense our approach is similar to theirs, because we rely on results from [37] which themselves are proved using a second moment method. However, we are able to get stronger results (in the sense that our leading constant does not decay as $p \to 1 - 1/q$, and moreover is optimal for binary codes). One of the reasons may be the notion of "implicit rareness" from [37], which provides a useful characterization of the lists contained in a random linear code.

The work [22] also established lower bounds on list-decoding random linear codes from *erasures*. While we do not discuss list-decoding from erasures in this work (except in the sense that erasure list-recovery is a generalization of list-decoding from erasures), this result is relevant to our work because [22] established an exponential lower bound of the form $L \geq \exp(\Omega(1/\varepsilon))$, in contrast to the list size $O(1/\varepsilon)$ that is attained by uniformly random codes. Thus, our results suggest that (zero-error) list-recovery behaves more like list-decoding from erasures than from errors, at least with respect to the list size of random (linear) codes.

Existing upper bounds for random linear codes

We now turn our attention to upper bounds on list sizes for random linear codes. A long line of works [45, 16, 15, 7, 43, 40, 42, 34] has studied this problem, and we highlight the most relevant results now. Zyablov and Pinsker [45] showed that random linear codes of rate $R^* - \varepsilon$ have lists of size at most $q^{1/\varepsilon}$.[7] Guruswami, Håstad, Sudan and Zuckerman [16] first

[7] For list-recovery with input lists of size ℓ, the argument of [45] shows that the list size is at most $q^{\ell/\varepsilon}$. Furthermore, their results for list-decoding also apply to average-radius list-decoding.

showed the existence of capacity-achieving binary linear codes with lists of size $O(1/\varepsilon)$. Li and Wootters [34] revisited their techniques and showed that in fact random linear codes of rate $R^* - \varepsilon$ have lists of size $O(1/\varepsilon)$ with high probability; moreover they computed the constant coefficient in the big-Oh notation. However, neither of these results apply to either average-radius list-decoding or to list-recovery. As discussed above in Section 1.2, our new upper bound is the result of an improvement of the techniques of [34], which extends their result to average-radius list-decoding.

As for larger alphabets, Guruswami, Håstad and Kopparty [15] showed that there exists a constant $C_{p,q}$ for which random linear codes are $(p, C_{p,q}/\varepsilon)$-list-decodable with high probability. Unfortunately, if p tends to $1 - 1/q$ then this constant tends to infinity. To address this, an ongoing line of works [7, 43, 41, 42] has studied the list-decodability of random linear codes in the "high-noise regime" where p is close to $1 - 1/q$; these results also apply to average-radius list-decodability. These results imply that for binary random linear codes, when $p = 1 - 1/q - \Theta(\sqrt{\varepsilon})$, random linear codes with rate $R^* - \varepsilon$ are average-radius list-decodable with list sizes $O(1/\varepsilon)$. However, the constant hiding in the big-Oh is not correct (in particular, the authors do not see how to make it smaller than 2). Moreover, these results only hold in a particular parameter regime for p and ε, and degrade as the alphabet size grows.

As for list-recovery, a result by Rudra and Wootters [42] guarantees that random linear codes with rate $R^* - \varepsilon$ over sufficiently large alphabets $\mathbb{F}_q$ have lists of sizes at most $(q\ell)^{O(\log(\ell)/\varepsilon)}$. To the best of our knowledge, no lower bounds were known.

Relevant results for other ensembles of codes

Lastly, we discuss some other results concerning other code ensembles. First of all, recent work of [37] shows that a random code from Gallager's ensemble of LDPC codes [12] achieves list-decoding capacity with high probability. More generally, they show that random LDPC codes have similar combinatorial properties to random linear codes, including list-decoding, average-radius list-decoding, and list-recovery. As part of their approach, they develop techniques to characterize the lists that appear in a random linear code with high probability, which we utilize for our work.

Finally, we note that there are no known explicit constructions of list-decodable codes of rate $R^* - \varepsilon$ which achieve a list size even of $O(1/\varepsilon)$. Over large alphabets, the best explicit constructions of capacity-achieving list-decodable or list-recoverable codes have list sizes at least $(1/\varepsilon)^{\Omega(1/\varepsilon)}$ (e.g., [33, 32]). Further, if one insists on *binary* codes, or even codes over alphabets of size independent of ε, we do not know of *any* explicit constructions of list-decodable codes with rate approaching R^*.

Two-point concentration

We showed that the optimal list size L of a random linear code is concentrated on at most three values for both list-decoding and average-radius list-decoding: $\lfloor h(p)/\varepsilon \rfloor + 2$, $\lfloor h(p)/\varepsilon \rfloor + 1$, and, if the value is different, $\lfloor h(p)/\varepsilon + 0.99 \rfloor$.

In [34, Theorem 2.5], it was also shown that the optimal list size of a *completely* random binary code is concentrated on two or three values for list-decoding. This type of concentration is also well studied in graph theory, where it is known that in Erdős-Rényi graphs, a number of graph parameters are concentrated on two values. Examples include the clique number (size of the largest clique) [36, 6], the chromatic number [35, 2, 1], and the diameter [39].

1.4 Discussion and open problems

In this work, we have made progress on pinning down the output list sizes for (average-radius) list-decoding and list-recovery for random linear codes. Before we continue with the technical portion of the paper, we highlight some open questions and future directions.

- We showed that random linear codes of rate $1 - h_q(p) - \varepsilon$ are not (p, L)-list-decodable for $L \sim \frac{h_q(p)}{\varepsilon}$. We conjecture this lower bound is tight, i.e. that random linear codes of rate $1 - h_q(p) - \varepsilon$ are (p, L)-(average-radius) list-decodable for $L = \frac{h_q(p)}{\varepsilon}(1 + o(1))$, where the $o(1) \to 0$ as $\varepsilon \to 0$. Our Theorem 14 (and earlier in [34] for list-decoding) shows it is true for $q = 2$, and we conjecture this is true for larger q.
- Our results show that list-decoding and average-radius list-decoding have essentially the same output list sizes over binary alphabets, for random linear codes. It would be interesting to extend this to larger alphabets, or even to more general families of codes. This is especially interesting given that there is an exponential gap in the best known lower bounds (on the list-size for arbitrary codes) between list-decoding and average-radius list-decoding for general codes.
- We have used different techniques for our upper and lower bounds. However, we think it is an interesting direction to use the characterization of [37] – which we used to prove our lower bounds – to prove upper bounds as well. This would entail showing that every sufficiently bad list is implicitly rare.
- Finally, we note that our lower bounds for list-recovery rely on the field $\mathbb{F}_q$ being an extension field (that is, $q = p^t$ for some $t > 1$). It is an interesting question whether or not an exponential lower bound also holds over prime fields. We note that other lower bounds on list-decoding and list-recovery for Reed-Solomon codes also apply only to extension fields [23, 3]; perhaps all of these bounds taken together are evidence that better list-decodability may be possible in general over prime fields.

1.5 Organization

In Section 2 we set up notation and formally state the results of [37] that we build on for our lower bounds. In Section 3 we state and sketch the proof of our lower bound on list-recovery of random linear codes. In Section 4 we state and sketch the proof of our lower bound on the list-decodability of random linear codes. In Appendix A we state and prove our upper bound on the list-decodability of random linear codes.

2 Preliminaries

In this section, we set notation and introduce the notions and results from [37] that we need for our lower bounds.

2.1 Notation

Unless otherwise specified, all logarithms are base 2. We use the notation $\exp(x)$ to mean e^x. For an integer a, we define $[a] := \{1, \ldots, a\}$. For a vector $x \in \mathbb{F}_q^A$ and $I \subset [A]$, we use $x_I \in \mathbb{F}_q^{|I|}$ to denote the vector $(x_i)_{i \in I}$ with coordinates from I in increasing order. We use $\mathbf{1}_D$ to denote the all ones vector of length D.

We use several notions from information theory. Define the q-ary entropy $h_q : [0,1] \to [0,1]$ by

$$h_q(x) \stackrel{\text{def}}{=} x \log_q(q-1) - x \log_q x - (1-x)\log_q(1-x) \tag{1}$$

We assume $q = 2$ if q is omitted from the subscript.

For a random variable X with domain $\mathcal{X}$, we use $H(X)$ to denote the *entropy* of X:

$$H(X) = -\sum_{x \in \mathcal{X}} \Pr_X(x) \log(\Pr_X(x)).$$

For a probability distribution τ, we may also use $H(\tau)$ to denote the entropy of a random variable with distribution τ.

Let X be a random variable supported on $\mathcal{X}$ and Y be a random variable supported on $\mathcal{Y}$. We define the *conditional entropy* of Y given X as

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)}.$$

It is easy to check that $H(X) - H(X|Y) = H(Y) - H(Y|X)$ and we call this the *mutual information* $I(X;Y)$:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

For random variables X, Y, Z, we define the *conditional mutual information* $I(X;Y|Z)$ by

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(Y|Z) - H(Y|X,Z)$$

Conditional entropy, mutual information, and conditional mutual information satisfy the *data processing inequality:* for any function f supported on the domain of Y, we have

$$H(X|f(Y)) \geq H(X|Y) \quad \text{and} \quad I(X;Y) \geq I(X;f(Y)) \quad \text{and} \quad I(X;Y|Z) \geq I(X;f(Y)|Z).$$

We also use *Fano's inequality*, which states that if X is a random variable supported on $\mathcal{X}$ and Y is a random variable supported on $\mathcal{Y}$, and if $f : \mathcal{Y} \to \mathcal{X}$ is a function and $p_{err} = \Pr_{X,Y}[f(Y) \neq X]$

$$H(X|Y) \leq h(p_{err}) + p_{err} \cdot \log(|\mathcal{X}| - 1)$$

We define

$$H_q(X) \stackrel{\text{def}}{=} \frac{H(X)}{\log q}, \qquad I_q(X;Y) = \frac{I(X;Y)}{\log q}.$$

and similarly for conditional entropy and conditional mutual information.

For a distribution τ on $\mathbb{F}_q^L$ and a matrix $A \in \mathbb{F}_q^{L' \times L}$, we define the distribution $A\tau$ on $\mathbb{F}_q^{L'}$ in the natural way by

$$\Pr_{A\tau}(x) = \sum_{\{y \in \mathbb{F}_q^L \,:\, Ay = x\}} \Pr_\tau(y),$$

namely, $A\tau$ is the distribution of the random vector Ay, where $y \sim \tau$.

We have defined list-decoding, average-radius list-decoding, and list-recovery in the introduction. We will in fact consider a more general version of list-recovery, which also tolerates erasures:

▶ **Definition 1** (List-recovery from erasures)**.** *A code $C \subset \mathbb{F}_q^n$ is (α, ℓ, L)*-list-recoverable from erasures *if the following holds. Let $S_1, \ldots, S_n \subset \mathbb{F}_q$ be lists so that $|S_i| \leq \ell$ for at least αn values of i. Then*

$$|\{c \in \mathcal{C} \,:\, \forall i \in [n], c_i \in S_i\}| < L.$$

We take $\alpha = 1$ if it is omitted.

2.2 Tools from [37]

As discussed in Section 1.2, for our lower bounds we use tools from the recent work [37]. We work with matrices $M \in \mathbb{F}_q^{n \times L}$ ($L \in \mathbb{N}$), where we view the columns of M as potential codewords in $\mathcal{C}$. We use the notation "$M \subseteq \mathcal{C}$" to mean that the columns of M are all contained in $\mathcal{C}$.

We group together sets of such matrices M according to their row distribution.

▶ **Definition 2** (τ_M, $\dim(\tau)$, $\mathcal{M}_{n,\tau}$)**.** *Given a matrix $M \in \mathbb{F}_q^{n \times L}$, the empirical row distribution defined by the rows of M over $\mathbb{F}_q^L$ is called the type τ_M of M. That is, τ_M is the distribution so that for $v \in \mathbb{F}_q^L$,*

$$\Pr_{\tau_M}(v) = \frac{|\{i \,:\, \textit{the } i\textit{'th row of } M \textit{ is equal to } v\}|}{n}.$$

For a distribution τ on $\mathbb{F}_q^L$, we use $\dim(\tau)$ to refer to $\dim(\mathrm{span}(\mathrm{supp}(\tau)))$. We use $\mathcal{M}_{n,\tau}$ to refer to the set of all matrices in $\mathbb{F}_q^{n \times L}$ which have empirical row distribution τ.

▶ Remark 3. We remark that for some distributions τ over $\mathbb{F}_q^L$, the set $\mathcal{M}_{n,\tau}$ may be empty due to $n \cdot \Pr_\tau(v)$ not being an integer. For such τ we can define $\mathcal{M}_{n,\tau}$ to consist of matrices M with either $\lfloor n \cdot \Pr_\tau(v) \rfloor$ or $\lceil n \cdot \Pr_\tau(v) \rceil$ copies of v. This has a negligible effect on the analysis as we always take n to be sufficiently large compared to other parameters, so for clarity of exposition we ignore this technicality.

Given $M \in \mathcal{M}_{n,\tau}$, note that $\mathcal{M}_{n,\tau}$ consists exactly of those matrices obtained by permuting the rows of M. In particular, since the *random linear code* model is invariant to such permutations, all of the matrices in $\mathcal{M}_{n,\tau}$ have the same probability of being contained in $\mathcal{C}$.

As discussed in Section 1.2, we prove a lower bound by exhibiting a distributions τ over $\mathbb{F}_q^L$ such that the corresponding set $\mathcal{M}_{n,\tau}$ is both *bad* and *abundant*. When $\mathcal{M}_{n,\tau}$ satisfies these properties, we say that τ itself is, respectively, *bad* and *abundant*.

The work [37] characterizes which distributions τ satisfy the *abundance* property, namely, which classes $\mathcal{M}_{n,\tau}$ are likely to have at least one of their elements appear (as a matrix) in a random linear code of a given rate. To motivate the definition below, suppose that the distribution τ has low entropy: $H_q(\tau) < \gamma \cdot \dim(\tau)$ for some $\gamma \in (0,1)$. This implies that the class $\mathcal{M}_{n,\tau}$ is not too big: more precisely, it is not hard to see that $|\mathcal{M}_{n,\tau}| \leq q^{H_q(\tau) \cdot n} \leq q^{\gamma \dim(\tau) n}$. Using a calculation like we did in Section 1.2, we see that, since $\mathcal{M}_{n,\tau}$ is not very large, it is unlikely for a random linear code of rate less than $1 - \gamma$ to contain a matrix from $\mathcal{M}_{n,\tau}$.

However, this is not the only reason that $\mathcal{M}_{n,\tau}$ might be unlikely to appear in a random linear code. As is shown in [37], it could also be because a random output of τ, subject to some linear transformation (perhaps to a space of smaller dimension), has low entropy. We call such distributions *implicitly rare:*

▶ **Definition 4** (γ-implicitly rare)**.** *We say that a distribution τ over $\mathbb{F}_q^L$ is* γ-implicitly rare *if there exists a full-rank linear transformation $A : \mathbb{F}_q^L \to \mathbb{F}_q^{L'}$ where $L' \leq L$ such that*

$$H_q(A\tau) < \gamma \cdot \dim(A\tau)$$

Observe that by taking A to be the identity map, we recover the case where τ itself has low entropy. Furthermore, note that every matrix in $\mathcal{M}_{n,A\tau}$ has all of its columns contained in the column-span of some matrix in $\mathcal{M}_{n,\tau}$. This implies that if no matrix in $\mathcal{M}_{n,A\tau}$ lies in a code, then no matrix in $\mathcal{M}_{n,\tau}$ lies in the code. Thus, abundance of the distribution $A\tau$ implies abundance of τ.

For an illustrative example of an implicitly rare distribution, we refer the reader to [37, Example 2.5]. Specifically, the example provides a case where for some full-rank matrix A, we have $H_q(A\tau)/\dim(A\tau) > H_q(\tau)/\dim(\tau)$.

Essentially, [37] shows that a row distribution τ is likely to appear in a random linear code (namely, τ satisfies the abundance property) if and only if it is *not* implicitly rare. The following theorem follows from Lemma 2.7 in [37].[8]

▶ **Theorem 5** (Follows from Lemma 2.7 in [37]). *Let $R \in (0,1)$ and fix $\eta > 0$. Let τ be a $(1-R-\eta)$-implicitly rare distribution over $\mathbb{F}_q^L$ ($L \in \mathbb{N}$), and let $\mathcal{C}$ be a random linear code of rate R. Then*

$$\Pr[\exists M \in \mathcal{M}_{n,\tau} : M \subseteq \mathcal{C}] \leq q^{-\eta n}$$

Conversely, suppose that τ is not $(1-R+\eta)$-implicitly rare. Then

$$\Pr[\exists M \in \mathcal{M}_{n,\tau} : M \subseteq \mathcal{C}] \geq 1 - n^{O_{L,q}(1)} \cdot q^{-\eta n} \ .$$

The first part of the theorem follows from a natural first-moment method argument, while the second part follows from the analogous second-moment argument. We emphasize that it is important that we allow arbitrary full-rank linear transformations $A : \mathbb{F}_q^L \to \mathbb{F}_q^{L'}$ in Definition 4: if we only allowed A to be the identity map, the second part of the theorem would be false.

3 Lower bounds for list-recovery with erasures

Our main result in this section is the following.

▶ **Theorem 6.** *Fix $0 \leq \rho < 1$. Fix a prime power $\ell \geq 2$ and an integer $t \geq 2$, and let $q = \ell^t$. Fix $0 < \varepsilon \leq \frac{1-\rho}{20t}$ and let $L = \ell^{\lceil \frac{1-\rho}{20\varepsilon} \rceil}$. For $n \in \mathbb{N}$, let $\mathcal{C} \subseteq \mathbb{F}_q^n$ denote a random linear code of rate $R := 1 - (\rho + (1-\rho)\log_q(\ell)) - \varepsilon$. Then the probability of $\mathcal{C}$ being $(1-\rho, \ell, L)$-erasure list-recoverable is at most $q^{-\Omega(n)}$.*

We will prove Theorem 6 below, after we build up the necessary building blocks. As discussed in Sections 1.2 and 2, to prove Theorem 6 we seek a distribution τ that is both is *bad* and *abundant*. That is, $\mathcal{C}$ should likely contains some matrix from $\mathcal{M}_{n,\tau}$, and the corresponding codewords should yield a counterexample to the list-recoverability of $\mathcal{C}$. We will describe our choice of τ in Definition 7; we will show that it is bad in Proposition 8; and finally we will show that it is not implicitly rare (and hence abundant by Theorem 5) in Lemma 9.

Our construction of the distribution τ follows similar lines to that in Section 1.2.

▶ **Definition 7** (The bad distribution τ for list-recovery lower bounds). *Fix ρ, ℓ, t, q as in Theorem 6. Let $D \geq t$ be a positive integer. Let $\mathbb{F}_\ell$ be a subfield of $\mathbb{F}_q$, where $q = \ell^t$ and $t \geq 2$. Let $\alpha_1, \ldots, \alpha_{(q-1)/(\ell-1)}$ be a set so that $\alpha_i \mathbb{F}_\ell^*$ are disjoint cosets of $\mathbb{F}_\ell^*$ partitioning $\mathbb{F}_q^*$. Let $L = \ell^D$. Let $G \in \mathbb{F}_\ell^{L \times D}$ be the matrix whose rows are all of the distinct elements of $\mathbb{F}_\ell^D$.*

Let σ be the distribution that with probability $1-\rho$ returns $\alpha_i u$ for (i,u) uniform in $\left\{1, \ldots, \frac{q-1}{\ell-1}\right\} \times \mathbb{F}_\ell^D$; and with probability ρ returns a uniformly random element of $\mathbb{F}_q^D$.

Let τ be the distribution given by Gv for $v \sim \sigma$.

[8] This is also given as Theorem 2.2 in the first version of [37], available at `https://arxiv.org/abs/1909.06430v1`.

To motivate this construction, consider first the $\rho = 0$ case. Now consider a matrix $M \in \mathbb{F}_q^{n \times L}$ that has row distribution given by τ. If we ignore the coefficients α_i, the columns of M span a D-dimensional subspace of $\mathbb{F}_\ell^n$. In particular, they are *bad,* in the sense that each coordinate of these codewords are contained in a list of size ℓ (namely, $\mathbb{F}_\ell$). Moreover, as soon as any D linearly independent columns of M are contained in $\mathcal{C}$, all of the columns of M are contained in $\mathcal{C}$; this suggests that it's relatively likely (compared to, say, a random matrix in $\mathbb{F}_q^{L \times n}$) that $M \subseteq \mathcal{C}$. These properties don't change when we multiply by the coefficients α_i: each coordinate is now contained in some list $\alpha_i \mathbb{F}_\ell$ rather than $\mathbb{F}_\ell$ (notice that the fact that the α_i are coset representatives means that all of these possible lists are disjoint, other than zero), and it's still just as likely that $M \subseteq \mathcal{C}$. However, by throwing these multiples α_i into the mix, we have increased the size of $\mathcal{M}_{n,\tau}$, making τ more *abundant.* In particular, note that, over all choices of (i, u), the value $\alpha_i u$ is distinct except when $u = 0$. Thus, τ has entropy close to the entropy of the uniform distribution on (i, u), so $H_q(\tau) \approx \log_q(q\ell^{D-1}) \approx D(\log_q(\ell) + \frac{1}{D})$. Using a similar idea, we can estimate the entropy of $A\tau$ for all matrices A, showing that τ is *not* $\log_q(\ell) + \frac{1}{10D}$ implicitly rare, implying that it is abundant.

To generalize to the $\rho > 0$ case, the construction essentially "frees" a ρ fraction of the coordinates relative to the $\rho = 0$ case. This further increases the size of $\mathcal{M}_{n,\tau}$ (making τ even more abundant), while still maintaining the badness property for list-recovery with a ρ fraction of erasures.

▶ **Proposition 8** (τ is bad). *Let τ be as in Definition 7. Let $\mathcal{C} \subseteq \mathbb{F}_q^n$, and let $M \in \mathcal{M}_{n,\tau}$. If $M \subseteq \mathcal{C}$, then $\mathcal{C}$ is not $(1-\rho, \ell, L)$-list-recoverable.*

Proof. Suppose that $M \subseteq \mathcal{C}$. Let $w_1, w_2, \ldots, w_n \in \mathbb{F}_q^L$ be the rows of M. It suffices to show that there are input lists $S_1, \ldots, S_n$ so that $w_j \in S_j^L$ for all $j \in [n]$, and so that for at least $(1-\rho)n$ values of $j \in [n]$, we have $|S_j| \leq \ell$. Recall that each row w_j of M is of the form Gv_j where a $(1-\rho)$ fraction of the v_j are of the form $\alpha_{i_j} \cdot u_j$ for $(i_j, u_j) \in [(q-1)/(\ell-1)] \times \mathbb{F}_\ell^D$, and a ρ fraction of the v_j are arbitrary vectors in $\mathbb{F}_q^D$.[9]

In the first case, set $S_j = \alpha_{i_j} \cdot \mathbb{F}_\ell$. Because the elements of G are all in $\mathbb{F}_\ell$, all the coordinates of $w_j = Gv_j = \alpha_{i_j} G u_j$ lie in S_j. Moreover by definition $|S_j| \leq \ell$. In the second case, set $S_j = \mathbb{F}_q$. By definition all the coordinates of $w_j \in \mathbb{F}_q^L$ lie in $S_j = \mathbb{F}_q$.

This completes the proof. ◀

Next, we claim that τ is not implicitly rare, which will imply that τ is abundant. Due to space constraints, the proof is omitted.

▶ **Lemma 9** (τ is abundant). *Let τ be as in Definition 7. Then τ is* not

$$\left(\rho + (1-\rho)\log_q(\ell) + \frac{1-\rho}{10D}\right)\text{-implicitly rare.}$$

We now show how to use Proposition 8 and Lemma 9 to prove Theorem 6.

Proof of Theorem 6. Let $0 < \varepsilon \leq \frac{1-\rho}{20t}$. Let τ be as in Definition 7, choosing $D = \lceil \frac{1-\rho}{20\varepsilon} \rceil$. By our choice of ε, we indeed have $D \geq t$. Lemma 9 shows that τ is *not* $(\rho + (1-\rho)\log_q(\ell) + \frac{1-\rho}{10D})$-implicitly rare. By choice of D, we have $\frac{1-\rho}{10D} > \varepsilon$. From Theorem 5 with $\eta = \frac{1-\rho}{10D} - \varepsilon$, we see that for any sufficiently large n, a random code of rate

[9] As per Remark 3, we may ignore the rounding issue that ρn may not be an integer. This is without loss of generality, as we may replace τ with a very similar distribution so that a $\lceil \rho n \rceil$ fraction of the v_j are arbitrary in $\mathbb{F}_q^D$, and adjust all parameters by a term that is $o(1)$ as $n \to \infty$.

$$\left(1-\left(\rho+(1-\rho)\log_q(\ell)+\frac{1-\rho}{10D}\right)\right)+\eta=1-\left(\rho+(1-\rho)\log_q(\ell)\right)-\varepsilon$$

contains ℓ^D codewords given by a matrix $M \in \mathcal{M}_{n,\tau}$ with probability at least $1-q^{\Omega(\varepsilon n)}$. By Proposition 8, if this occurs, then $\mathcal{C}$ is not $(1-\rho,\ell,L)$-list-recoverable. ◀

4 Lower bounds for list-decoding with errors

Our main theorem in this section is the following.

▶ **Theorem 10.** *Fix a prime power q, fix $p \in (0, 1-\frac{1}{q})$, and fix $\delta \in (0,1)$. There exists $\varepsilon_{p,q,\delta} > 0$ such that for all $\varepsilon \in (0, \varepsilon_{p,q,\delta})$ and n sufficiently large, a random linear code in $\mathbb{F}_q^n$ of rate $1-h_q(p)-\varepsilon$ is not $\left(p, \left\lfloor \frac{h_q(p)}{\varepsilon} - \delta \right\rfloor\right)$-list-decodable with probability $1-q^{-\Omega(n)}$.*

Our proof of Theorem 10 below follows the same outline as the proof of Theorem 6 above. We first define a bad distribution τ in Definition 11; then we will show that it is bad in Proposition 12; then we will show that it is not implicitly rare (and hence abundant by Theorem 5) in Lemma 13. Finally we will prove Theorem 10 from these pieces.

Below, we let $\text{Bernoulli}_q(p)$ be the distribution that returns $0 \in \mathbb{F}_q$ with probability $1-p$ and any other element of $\mathbb{F}_q$ with probability $\frac{p}{q-1}$.

▶ **Definition 11** (The bad distribution τ for list-decoding lower bounds)**.** *Let $p \in (0, 1-\frac{1}{q})$ and $\delta > 0$. Choose $L > 0$. Define the distribution τ on $\mathbb{F}_q^L$ as the distribution of the random vector $u + \alpha \mathbf{1}_L$, where $u \sim \text{Bernoulli}_q(p)^L$, and α is sampled independently and uniformly from $\mathbb{F}_q$.*

First, we observe that τ is indeed bad, in the sense that it provides a counter-example to list-decodability.

▶ **Proposition 12** (τ is bad)**.** *Let τ be as in Definition 11. Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ and let $M \in \mathcal{M}_{n,\tau}$. If $M \subseteq \mathcal{C}$, then $\mathcal{C}$ is not (p,L)-list-decodable.*

Proof. Let $M \in \mathcal{M}_{n,\tau}$. We want to show that the columns of M all lie in a single ball of radius pn.

By definition of τ and $\mathcal{M}_{n,\tau}$, we may write the j-th row of M as $u^{(j)} + \alpha_j \mathbf{1}_L$, so that the empirical distribution of the pairs $(u^{(j)}, \alpha_j)_{1 \le j \le n}$ is $\text{Bernoulli}_q(p)^L \times \text{Uniform}(\mathbb{F}_q)$.[10]

For any $i \in [L]$, the number of $j \in [n]$ such that $M_{i,j} = u_i^{(j)} + \alpha_j \neq \alpha_j$ is exactly the number of times $u_i^{(j)} \neq 0$, which is pn, since $u_i^{(j)}$ is distributed as $\text{Bernoulli}_q(p)$. Thus, each column $M_{i,*}$ of M has distance at most pn from the word $(\alpha_1, \ldots, \alpha_n)$, so that any code containing M has L codewords in a ball of radius pn and hence is not (p,L)-list-decodable. ◀

Next, we claim that τ is appropriately implicitly rare for large enough L. Due to space constraints, the proof is omitted.

▶ **Lemma 13.** *Let $p \in (0, 1-\frac{1}{q})$ and let $\delta > 0$. There exists $L_{p,q,\delta}$ such that, for $L \ge L_{p,q,\delta}$, the distribution τ given in Definition 11 is not $\left(h_q(p) + \frac{h_q(p)}{L+\delta}\right)$-implicitly rare.*

[10] This is without loss of generality: if not, as per Remark 3, we can associate pairs with rows so that the empirical distribution is close to $\text{Bernoulli}_q(p)^L \times \text{Uniform}(\mathbb{F}_q)$ up to an additive factors that are $o(1)$ as $n \to \infty$. After adjusting parameters, this has a negligible effect on the analysis and final result.

We now show how to use Proposition 12 and Lemma 13 to prove Theorem 10.

Proof of Theorem 10. Let $L_{p,q,\delta/2}$ be as in Lemma 13 and choose $\varepsilon_{p,q,\delta} \stackrel{\text{def}}{=} \frac{h_q(p)}{L_{p,q,\delta/2}+1}$. Fix $\varepsilon < \varepsilon_{p,q,\delta}$. Let $L = \left\lfloor \frac{h_q(p)}{\varepsilon} - \delta \right\rfloor$. Let τ be as in Definition 11 with this choice of L. By Lemma 13, as $L \geq L_{p,q,\delta/2}$, τ is not $\left(h_q(p) + \frac{h_q(p)}{L+\delta/2}\right)$-implicitly rare. Thus, as $\varepsilon \leq \frac{h_q(p)}{L+\delta} < \frac{h_q(p)}{L+\delta/2}$, there is some constant $c_{p,q,\varepsilon} > 0$ so that τ is not $(h_q(p) + \varepsilon + c_{p,q,\varepsilon})$-implicitly rare.

Then Theorem 5 with $\eta = c_{p,q,\varepsilon}$ tells us that, for n sufficiently large, a random linear code of rate $1 - (h_q(p) + \varepsilon + c_{p,q,\varepsilon}) + c_{p,q,\varepsilon} = 1 - h_q(p) - \varepsilon$ contains L codewords given by some matrix $M \in \mathcal{M}_{n,\tau}$ with probability at least $1 - q^{-\Omega_{p,q,\varepsilon}(n)}$.

Finally, Proposition 12 implies that $\mathcal{C}$ is not (p, L)-list-decodable. Our choice of L proves the theorem. ◀

References

1 Dimitris Achlioptas and Assaf Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162(3):1335–1351, 2005.

2 Noga Alon and Michael Krivelevich. The concentration of the chromatic number of random graphs. *Combinatorica*, 17(3):303–313, 1997.

3 Eli Ben-Sasson, Swastik Kopparty, and Jaikumar Radhakrishnan. Subspace polynomials and limits to list decoding of reed–solomon codes. *IEEE Transactions on Information Theory*, 56(1):113–120, 2009.

4 Vladimir M Blinovsky. Code bounds for multiple packings over a nonbinary finite alphabet. *Problems of Information Transmission*, 41(1):23–32, 2005.

5 Volodia M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Problems of Information Transmission*, 22(1):7–19, 1986.

6 Béla Bollobás and Paul Erdös. Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 80, pages 419–427. Cambridge University Press, 1976.

7 Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 432–442, 2013. doi:10.1137/1.9781611973105.31.

8 Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly optimal pseudorandomness from hardness. Technical report, ECCC preprint TR19-099, 2019.

9 Zeev Dvir and Shachar Lovett. Subspace evasive sets. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 351–358. ACM, 2012.

10 Peter Elias. List decoding for noisy channels. *Wescon Convention Record, Part 2*, pages 94–104, 1957.

11 Peter Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37(1):5–12, 1991.

12 Robert G. Gallager. Low-density parity-check codes. *IRE Trans. Information Theory*, 8(1):21–28, 1962. doi:10.1109/TIT.1962.1057683.

13 Anna C Gilbert, Hung Q Ngo, Ely Porat, Atri Rudra, and Martin J Strauss. l2/l2-foreach sparse recovery with low risk. In *International Colloquium on Automata, Languages, and Programming*, pages 461–472. Springer, 2013.

14 Venkatesan Guruswami. List decoding from erasures: Bounds and code constructions. *IEEE Transactions on Information Theory*, 49(11):2826–2833, 2003.

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

16 Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Trans. Information Theory*, 48(5):1021–1034, 2002. doi:10.1109/18.995539.

17 Venkatesan Guruswami and Piotr Indyk. Expander-based constructions of efficiently decodable codes. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 658–667, 2001. doi:10.1109/SFCS.2001.959942.

18 Venkatesan Guruswami and Piotr Indyk. Near-optimal linear-time codes for unique decoding and new list-decodable codes over smaller alphabets. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 812–821, 2002.

19 Venkatesan Guruswami and Piotr Indyk. Linear time encodable and list decodable codes. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 126–135, 2003.

20 Venkatesan Guruswami and Piotr Indyk. Efficiently decodable codes meeting gilbert-varshamov bound for low rates. In *SODA*, volume 4, pages 756–757. Citeseer, 2004.

21 Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016.

22 Venkatesan Guruswami and Srivatsan Narayanan. Combinatorial limitations of average-radius list-decoding. *IEEE Trans. Information Theory*, 60(10):5827–5842, 2014. doi:10.1109/TIT.2014.2343224.

23 Venkatesan Guruswami and Atri Rudra. Limits to list decoding reed-solomon codes. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 602–609, 2005.

24 Venkatesan Guruswami and Atri Rudra. Concatenated codes can achieve list-decoding capacity. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(054), 2008. URL: http://eccc.hpi-web.de/eccc-reports/2008/TR08-054/index.html.

25 Venkatesan Guruswami and Salil P. Vadhan. A lower bound on list size for list decoding. In *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th InternationalWorkshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22–24, 2005, Proceedings*, pages 318–329, 2005. doi:10.1007/11538462_27.

26 Venkatesan Guruswami and Chaoping Xing. Folded codes from function field towers and improved optimal rate list decoding. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 339–350. ACM, 2012.

27 Venkatesan Guruswami and Chaoping Xing. List decoding Reed-Solomon, Algebraic-Geometric, and Gabidulin subcodes up to the Singleton bound. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 843–852. ACM, 2013.

28 Iftach Haitner, Yuval Ishai, Eran Omri, and Ronen Shaltiel. Parallel hashing via list recoverability. In *Annual Cryptology Conference*, pages 173–190. Springer, 2015.

29 Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes & applications. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 204–215. IEEE, 2017.

30 Brett Hemenway and Mary Wootters. Linear-time list recovery of high-rate expander codes. *Information and Computation*, 261:202–218, 2018.

31 Piotr Indyk, Hung Q Ngo, and Atri Rudra. Efficiently decodable non-adaptive group testing. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1126–1142. SIAM, 2010.

32 Swastik Kopparty, Nicolas Resch, Noga Ron-Zewi, Shubhangi Saraf, and Shashwat Silas. On list recovery of high-rate tensor codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 2019. URL: https://eccc.weizmann.ac.il/report/2019/080/.

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

34 Ray Li and Mary Wootters. Improved list-decodability of random linear binary codes. *arXiv preprint*, 2018. arXiv:1801.07839.

35 Tomasz Luczak. A note on the sharp concentration of the chromatic number of random graphs. *Combinatorica*, 11(3):295–297, 1991.

36 David W Matula. Employee party problem. In *Notices of the American Mathematical Society*, volume 19, pages A382–A382. Amer Mathematical Soc 201 Charles St, Providence, RI 02940-2213, 1972.

37 Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. Ldpc codes achieve list decoding capacity. *arXiv preprint*, 2019. arXiv:1909.06430.

38 Hung Q Ngo, Ely Porat, and Atri Rudra. Efficiently decodable error-correcting list disjunct matrices and applications. In *International Colloquium on Automata, Languages, and Programming*, pages 557–568. Springer, 2011.

39 Oliver Riordan and Nicholas Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19(5-6):835–926, 2010.

40 Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 764–773. ACM, 2014.

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

42 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.

43 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*, pages 853–860, 2013. doi:10.1145/2488608.2488716.

44 Jack Wozencraft. List decoding. *Quarter Progress Report*, 48:90–95, 1958.

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

A Upper bounds for average-radius list-decoding over $\mathbb{F}_2$

In this section we prove the following theorem. Recall that we abbreviate $h(p) = h_2(p)$.

▶ **Theorem 14.** *Let $n \in \mathbb{N}$. Let $p \in (0, \frac{1}{2})$ and $R = 1 - h(p) - \varepsilon$, where $0 < \varepsilon < 1 - h(p)$. Let $L = \lfloor \frac{h(p)}{\varepsilon} + 2 \rfloor$. Then, a random linear code $\mathcal{C} \leq \mathbb{F}_2^n$ of rate R is (p, L)-average-radius list-decodable with probability $1 - 2^{-\Omega_{p,\varepsilon}(n)}$.*

Recall from the introduction that, following the techniques in [16] and [34], we imagine sampling independent and uniform vectors $b_1, \ldots, b_k$ and constructing the "intermediate" random linear codes $\mathcal{C}_i = \text{span}\{b_1, \ldots, b_i\}$. A potential function based argument is used to show that, with high probability, each of these intermediate codes is indeed (p, L)-average-radius list-decodable; in particular, this is true for $\mathcal{C}_k$.

Before discussing our potential function, we first briefly review the techniques of [16] and [34]; in particular, we describe the potential function they use. First, for a code $\mathcal{C}$ and a vector $x \in \mathbb{F}_2^n$, we define

$$L_{\mathcal{C}}(x) := |\{c \in \mathcal{C} : \delta(x, c) \leq p\}| \ .$$

In [16], the authors define

$$S_{\mathcal{C}} := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} 2^{\varepsilon n L_{\mathcal{C}}(x)}$$

and observe that, for any $b_1, \ldots, b_i \in \mathbb{F}_2^n$,

$$\mathop{\mathbb{E}}_{b_{i+1} \sim \mathbb{F}_2^n} \left[S_{\mathcal{C}_i + \{0, b_{i+1}\}} \right] = S_{\mathcal{C}_i}^2 \ ,$$

where we recall $\mathcal{C}_i = \mathrm{span}\{b_1, \ldots, b_i\}$.[11] That is, the potential function squares in expectation, so the probabilistic method guarantees that we can choose some b_{i+1} for which $S_{\mathcal{C}_{i+1}} \le S_{\mathcal{C}_i}^2$. Thus, for some choice of $b_1, \ldots, b_k$, one has $S_{\mathcal{C}_k} \le (S_{\{0\}})^{2^k}$.

In [34], the definition of $S_{\mathcal{C}}$ is slightly modified:

$$S_{\mathcal{C}} := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} 2^{\frac{\varepsilon n L_{\mathcal{C}}(x)}{1+\varepsilon}} \ .$$

This little bit of extra room allows to show that, in fact, with high probability over the choice of b_{i+1}, $S_{\mathcal{C}_i + \{0, b_{i+1}\}} \le S_{\mathcal{C}_i}^2$. By a union bound, it follows that with high probability, $S_{\mathcal{C}_k} \le (S_{\{0\}})^{2^k}$.

In either case, to conclude the proof, one observes the bound[12] $S_{\{0\}} \le 1 + 2^{-n(1-h(p)-\varepsilon)}$ and then uses

$$S_{\mathcal{C}_k} \le (S_{\{0\}})^{2^k} \le (1 + 2^{-n(1-h(p)-\varepsilon)})^{2^k} \le \exp 2^{k-n(1-h(p)-\varepsilon)} \le O(1)$$

for k chosen as above.

A.1 Alterations for average-radius list-decoding

While this argument analyzes the (absolute-radius) list-decodability of random linear codes very effectively, it is not immediately clear how to generalize the argument to study average-radius list-decodability. We now introduce the additional ideas we need to derive Theorem 14. We will fix a threshold parameter $\lambda \in (p, \frac{1}{2})$ for which $h(\lambda) < 1 - R = h(p) + \varepsilon$, to be determined later, and define

$$\eta \stackrel{\mathrm{def}}{=} 1 - R - h(\lambda).$$

We define the function $M_{R,\lambda} : [0,1] \to \mathbb{R}$ by

$$M_{R,\lambda}(\gamma) := \begin{cases} 1 - R - h(\gamma) & \text{if } \gamma < \lambda \\ 0 & \text{if } \gamma \ge \lambda \end{cases} .$$

▶ Remark 15. One can think of this quantity as a sort of "normalized entropy change" up to the threshold λ. Recalling that $1 - R = h(p) + \varepsilon$, if $\gamma < \lambda$, then

$$M_{R,\lambda}(\gamma) \approx \frac{1}{n} \left(h(p) - h(\lambda) \right) \approx \log \left(\frac{|B^n(p)|}{|B^n(\gamma)|} \right),$$

where $B^n(p)$ denotes the Hamming ball in $\mathbb{F}_2^n$ of radius p. Hence, $M_{R,\lambda}(\gamma)$ is something like a normalized "surprise" an observer would experience if they are expecting a random vector of weight $\le p$ and see a vector of weight $\le \gamma$.

[11] Here and throughout, for two subsets $A, B \subseteq \mathbb{F}_2^n$, we denote $A + B = \{a + b : a \in A, b \in B$. Thus, $\mathcal{C}_i + \{0, b_{i+1}\} = \mathcal{C}_{i+1}$.

[12] Actually, for the potential function in [34], one has $S_{\{0\}} \le 1 + 2^{-n(1-h(p)-\frac{\varepsilon}{1+\varepsilon})}$, but this difference does not matter for the conclusion.

For a linear code $\mathcal{C} \leq \mathbb{F}_2^n$ and $x \in \mathbb{F}_2^n$ we define

$$L_{\mathcal{C},R,\lambda}(x) := \sum_{y \in \mathcal{C}} M_{R,\lambda}(\delta(x,y)).$$

This is intuitively the "smoothed-out" list-size of x, where nearby codewords are weighted more heavily than far away codewords, and the weighting is given by the "entropy change" implied by the distance from x to y.

Next, we define

$$A_{\mathcal{C},R,\lambda}(x) := 2^{\frac{nL_{\mathcal{C},R,\lambda}(x)}{1+\eta}} \quad \text{and} \quad S_{\mathcal{C},R,\lambda} := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} A_{\mathcal{C},R,\lambda}(x).$$

The quantity $S_{\mathcal{C},R,\lambda}$ is the potential function we analyze.

A.2 Proof of Theorem 14

In this subsection we prove Theorem 14. The quantities R and λ (and hence $\eta = 1-R-h(\lambda)$) will be fixed throughout – although the precise value of λ will be determined later – and so we will suppress their dependence and simply write $M(x)$, $L_{\mathcal{C}}(x)$, $A_{\mathcal{C}}(x)$ and $S_{\mathcal{C}}$.

First, we observe that the following analog of [34, Lemma 3.2] holds. The proof is a simple adaptation of theirs (which in turn follows [16]).

▶ **Lemma 16.** *For all $\mathcal{C} \leq \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^n$,*

$$L_{\mathcal{C}+\{0,b\}}(x) \leq L_{\mathcal{C}}(x) + L_{\mathcal{C}}(x+b), \tag{2}$$

$$A_{\mathcal{C}+\{0,b\}}(x) \leq A_{\mathcal{C}}(x) \cdot A_{\mathcal{C}}(x+b). \tag{3}$$

Moreover, equality holds if and only if $b \notin \mathcal{C}$.

Next, we bound $S_{\{0\}}$. We have

$$S_{\{0\}} \leq 1 + 2^{-n} \sum_{\substack{x \in \mathbb{F}_2^n \\ \mathrm{wt}(x) \leq \lambda}} 2^{\frac{n \cdot (1-R-h(\mathrm{wt}(x)))}{1+\eta}} \leq 1 + \sum_{i=0}^{\lfloor \lambda n \rfloor} 2^{-n\left(1-h(i/n)-\frac{h(\lambda)+\eta-h(i/n)}{1+\eta}\right)}.$$

As this sum is dominated by its last term, we deduce

$$S_{\{0\}} \leq 1 + (\lambda n) 2^{-n\left(1-h(\lambda)-\frac{\eta}{1+\eta}\right)}. \tag{4}$$

From here, we can combine Lemma 16 and (4) to deduce

▶ **Lemma 17.** *Let $p \in (0, \frac{1}{2})$ and $R = 1 - h(p) - \varepsilon$ for $0 < \varepsilon < 1 - h(p)$. Let $\mathcal{C}_{Rn} \leq \mathbb{F}_2^n$ be a random linear code of rate R. Then $S_{\mathcal{C}_{Rn}} \leq 2$ with probability at least $1 - \exp(-\Omega_\eta(n))$.*

The proof of this lemma is completely analogous to that of [34, Lemma 3.3]. One only needs to be careful about the growth rate of $S_{\mathcal{C}}$. In particular, this proof crucially uses that η is positive. We again choose vectors $b_1, \ldots, b_{Rn}$ independently and uniformly at random. If $\mathcal{C}_i = \mathrm{span}\{b_1, \ldots, b_i\}$, we need "in expectation" that $S_{\mathcal{C}_i} \leq 1 + 2^{-\Omega(n)}$ for all i for the error bounds to succeed. As we expect the $o(1)$ term to roughly double, we need $2^{Rn} \cdot (S_{(0)} - 1) = 2^{-n(\eta - \frac{\eta}{1+\eta})} \leq 2^{-\Omega_\eta(n)}$.

Thus, in order to conclude Theorem 14, we are simply required to demonstrate that $S_{\mathcal{C}} \leq 2$ implies that $\mathcal{C}$ is (p, L)-average-radius list-decodable: this is the crux of our contribution. The main lemma we require is the following.

▶ **Lemma 18.** *Let $\mathcal{C} \leq \mathbb{F}_2^n$ be a linear code of rate R such that $S_{\mathcal{C}} \leq 2$. Then, for all $x \in \mathbb{F}_2^n$ and $D \subseteq \mathcal{C} \cap B(x, \lambda)$, it holds that*

$$\sum_{y \in D} h(\delta(x,y)) \geq (|D| - 1 - \eta)(1-R) - \frac{1+\eta}{n}.$$

Proof. First, observe that for any $x \in \mathbb{F}_2^n$,

$$L_{\mathcal{C}}(x) \geq \sum_{y \in D} ((1-R) - h(\delta(x,y))) = |D|(1-R) - \sum_{y \in D} h(\delta(x,y))$$

$$\text{so} \qquad \log A_{\mathcal{C}}(x) \geq n \frac{|D|(1-R) - \sum_{y \in D} h(\delta(x,y))}{1+\eta}. \tag{5}$$

Next, as $\delta(x,y) = \delta(x+z, y+z)$ for any $z \in \mathbb{F}_2^n$, we have, for any $x \in \mathbb{F}_2^n$ and $c \in \mathcal{C}$, that $L_{\mathcal{C}}(x) = L_{\mathcal{C}}(x+c)$ and hence $A_{\mathcal{C}}(x) = A_{\mathcal{C}}(x+c)$. Thus, $\max_{x \in \mathbb{F}_2^n} A_{\mathcal{C}}(x)$ is attained at at least $|\mathcal{C}|$ different values of x, so

$$S_{\mathcal{C}} = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} A_{\mathcal{C}}(x) \geq \frac{1}{2^n} \cdot |\mathcal{C}| \cdot \max_{x \in \mathbb{F}_2^n} A_{\mathcal{C}}(x) = 2^{-(1-R)n} \cdot \max_{x \in \mathbb{F}_2^n} A_{\mathcal{C}}(x).$$

Combining this with (5), we have, for any $x \in \mathbb{F}_2^n$,

$$\begin{aligned} 1 &\geq \log S_{\mathcal{C}} \geq -(1-R)n + \log\left(A_{\mathcal{C}}(x)\right) \\ &\geq n \cdot \left(-(1-R) + \frac{|D|(1-R) - \sum_{y \in D} h(\delta(x,y))}{1+\eta} \right) \\ &= n \cdot \frac{(|D| - 1 - \eta)(1-R) - \sum_{y \in D} h(\delta(x,y))}{1+\eta}. \end{aligned}$$

Rearranging yields the lemma. ◀

We may now conclude Theorem 14.

Proof of Theorem 14. Since $L = \lfloor \frac{h(p)}{\varepsilon} + 2 \rfloor > \frac{h(p)}{\varepsilon} + 1 = \frac{1-R}{\varepsilon}$, there exists $\eta > 0$ small enough so that for all sufficiently large n

$$L > \frac{1 - R + \eta + \frac{1+\eta}{n}}{\varepsilon - \eta}. \tag{6}$$

Thus, we define λ so that η (which we defined as $\eta = 1 - R - h(\lambda)$) satisfies (6). Let $\mathcal{C}$ be a random linear code of rate R. Due to Lemma 17, the conclusion of Lemma 18, holds with probability $1 - 2^{-\Omega_\eta(n)}$ for $\mathcal{C}$. It remains to show that, assuming n is sufficiently large, any code $\mathcal{C}$ satisfying the conclusion of Lemma 18 is (p, L)-average-radius list-decodable.

Let $x \in \mathbb{F}_2^n$ and $\Lambda \subseteq \mathcal{C}$ such that $|\Lambda| = L$; our goal is to show that, for all such x and Λ,

$$\frac{1}{L} \sum_{y \in \Lambda} \delta(x,y) > p. \tag{7}$$

Let

$$D = \{y \in \Lambda : \delta(x,y) \leq \lambda\}$$

and define

$$h^*(\alpha) = \begin{cases} h(\alpha) & \text{if } \alpha \leq \frac{1}{2} \\ 1 & \text{if } \alpha > \frac{1}{2} \end{cases}.$$

Now,

$$\sum_{y\in\Lambda} h^*(\delta(x,y)) \geq \sum_{y\in D} h(\delta(x,y)) + (L-|D|)h(\lambda) \tag{8}$$

$$\geq (|D|-1-\eta)(1-R) + (L-|D|)(1-R-\eta) - \frac{1+\eta}{n} \tag{9}$$

$$= (1-R)\cdot(L-1) - \eta\cdot(1-R) - \eta\cdot(L-|D|) - \frac{1+\eta}{n}$$

$$\geq (1-R)\cdot(L-1) - \eta\cdot(L+1) - \frac{1+\eta}{n}$$

$$= (1-R)L - (1-R) - \eta\cdot(L+1) - \frac{1+\eta}{n}$$

$$= Lh(p) - (1-R) - (L+1)\eta + L\varepsilon - \frac{1+\eta}{n} \tag{10}$$

$$= Lh(p) - (1-R+\eta) + L(\varepsilon-\eta) - \frac{1+\eta}{n}$$

$$> Lh(p). \tag{11}$$

Here, Inequality (8) holds because $h^*(\alpha) > h(\lambda)$ for all $\alpha > \lambda$; Inequality (9) is the conclusion of Lemma 18; Equality (10) follows from the fact that $R = 1 - h(p) - \varepsilon$; and Inequality (11) follows from (6). Thus, we deduce

$$\frac{1}{L}\sum_{y\in\Lambda} h^*(\delta(x,y)) > h(p). \tag{12}$$

Since h^* is concave,

$$h^*\left(\frac{1}{L}\sum_{y\in\Lambda}(\delta(x,y))\right) \geq \frac{1}{L}\sum_{y\in\Lambda} h^*(\delta(x,y)),$$

and so (7) follows from (12), the monotonicity of h^* and the fact that $h^*(p) = h(p)$. ◀