Improved list-decodability of random linear binary codes

Ray Li* and Mary Wootters†

April 24, 2018

Abstract

There has been a great deal of work establishing that random linear codes are as list-decodable as uniformly random codes, in the sense that a random linear binary code of rate $1 - H(p) - \epsilon$ is $(p, O(1/\epsilon))$-list-decodable with high probability. In this work, we show that such codes are $(p, H(p)/\epsilon + 2)$-list-decodable with high probability, for any $p \in (0, 1/2)$ and $\epsilon > 0$. In addition to improving the constant in known list-size bounds, our argument—which is quite simple—works simultaneously for all values of $p$, while previous works obtaining $L = O(1/\epsilon)$ patched together different arguments to cover different parameter regimes.

Our approach is to strengthen an existential argument of (Guruswami, Håstad, Sudan and Zuckerman, IEEE Trans. IT, 2002) to hold with high probability. To complement our upper bound for random linear codes, we also improve an argument of (Guruswami, Narayanan, IEEE Trans. IT, 2014) to obtain an essentially tight lower bound of $1/\epsilon$ on the list size of uniformly random codes; this implies that random linear codes are in fact more list-decodable than uniformly random codes, in the sense that the list sizes are strictly smaller.

To demonstrate the applicability of these techniques, we use them to (a) obtain more information about the distribution of list sizes of random linear codes and (b) to prove a similar result for random linear rank-metric codes.

*Department of Computer Science, Stanford University. Research supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE - 1656518.
†Departments of Computer Science and Electrical Engineering, Stanford University.
1 Introduction

An error correcting code is a subset $C \subseteq \mathbb{F}_2^n$, which is ideally “spread out.” In this paper, we focus on one notion of “spread out” known as list decodability. We say that a code $C$ is $(p,L)$-list-decodable if any Hamming ball of radius $pn$ in $\mathbb{F}_2^n$ contains at most $L$ points of $C$: that is, if for all $x \in \mathbb{F}_2^n$, $|B(x,pn) \cap C| \leq L$, where $B(x,pn)$ is the Hamming ball of radius $pn$ centered at $x$. Since list decoding was introduced in the 1950’s [Eli57, Woz58], it has found applications beyond communication, for example in pseudorandomness [Vad12] and complexity theory [Sud00].

A classical result in list decoding is known as the list-decoding capacity theorem:

**Theorem 1.1** (List decoding capacity theorem). Let $p \in (0,1/2)$ and $\varepsilon > 0$.

1. There exist binary codes of rate $1 - H(p) - \varepsilon$ that are $(p, \lfloor 1/\varepsilon \rfloor)$-list decodable.

2. Any binary code of rate $1 - H(p) + \varepsilon$ that is $(p,L)$-list decodable up to distance $p$ must have $L \geq 2^{\Omega(n)}$.

Above, $H(p) = -(1-p)\log_2(1-p) - p\log_2(p)$ is the binary entropy function. We say that a family of binary codes with rate approaching $1 - H(p)$ which are $(p,L)$-list-decodable for $L = O(1)$ achieves list-decoding capacity.1

Theorem 1.1 is remarkable because it means that even when $pn$ is much larger than half the minimum distance of the code—the radius at which at most one codeword $c \in C$ lies in any Hamming ball—it still can be the case that only a constant number of $c \in C$ lie in any Hamming ball of radius $pn$. Because of this, there has been a great deal of work attempting to understand what codes achieve the bound in Theorem 1.1.

The existential part of Theorem 1.1 is proved by showing that a uniformly random subset of $\mathbb{F}_2^n$ is $(p,1/\varepsilon)$-list decodable with high probability. For a long time, uniformly random codes were the only example of binary codes known to come close to this bound, and today we still do not have many other options. There are explicit constructions of capacity-achieving list-decodable codes over large alphabets (either growing with $n$ or else large-but-constant) [DL12, GX12, GX13], but over binary alphabets we still do not have any explicit constructions; we refer the reader to the survey [Gur09] for an overview of progress in this area.

Because it is a major open problem to construct explicit binary codes of rate $1 - H(p) - \varepsilon$ with constant (or even $\mathrm{poly}(n)$) list-sizes, one natural line of work has been to study structured random approaches, in particular random linear codes. A random linear code $C \subseteq \mathbb{F}_2^n$ is simply a random subspace of $\mathbb{F}_2^n$, and the list-decodability of these codes has been well-studied [ZPS1, GHS02, GHK11, CGV13, Woo13, RW14, RW18]. There are several reasons to study the list-decodability of random linear codes. Not only is it a natural question in its own right as well as a natural stepping stone in the quest to obtain explicit binary list-decodable codes, but also the list-decodability of random linear codes is useful in other coding-theoretic applications. One example of this is in concatenated codes and related constructions [GI04, GR08a, HW15, HRZW17], where a random linear code is used as a short inner code. Here, the linearity is useful because (a) a linear code can be efficiently described; (b) it is sometimes desirable to obtain a linear code at the end of the day, hence all components of the construction must be linear; and (c) as in [HW15] sometimes the linearity is required for the construction to work.

To this end, the line of work mentioned above has aimed to establish that random linear codes are “as list-decodable” as uniformly random codes. That is, uniformly random codes are viewed

---

1Sometimes the phrase “achieves list-decoding capacity” is also used when $L = \mathrm{poly}(n)$; since this paper focuses on the exact constant in the $O(1)$ term however, we use it to mean that $L = O(1)$. 

---

1
(as is often the case in coding theory) as the optimal construction, and we try to approximate this optimality with random linear codes, despite the additional structure.

**Our contributions.** In this paper, we give an improved analysis of the list-decodability of random linear binary codes. More precisely, our contributions are as follows.

- **A unified analysis.** As we discuss more below, previous work on the list-decodability of random linear binary codes either work only in certain (non-overlapping) parameter regimes [GHK11, Woo13], or else get substantially sub-optimal bounds on the list-size [RW18]. Our argument obtains improved list size bounds over all these results and works in all parameter regimes.

  Our approach is surprisingly simple: we adapt an existential argument of Guruswami, Håstad, Sudan and Zuckerman [GHSZ02] to hold with high probability. Extending the argument in this way was asked as an open question in [GHSZ02] and had been open until now.

- **Improved list-size for random linear codes.** Not only does our result imply that random linear codes of rate $1 - H(p) - \varepsilon$ are $(p, L)$-list-decodable with list-size $L = O(1/\varepsilon)$, in fact we show that $L \leq H(p)/\varepsilon + 2$. In particular, the leading constant is small and—to the best of our knowledge—is the best known, even existentially, for any list-decodable code.

- **Tight list-size lower bound for uniformly random codes.** To complement our upper bound, we strengthen an argument of Guruswami and Narayanan [GN14] to show that a uniformly random binary code of rate $1 - H(p) - \varepsilon$ requires $L \geq (1 - \gamma)/\varepsilon$ for any constant $\gamma > 0$ and sufficiently small $\varepsilon$. In other words, the list size of $1/\varepsilon$ in Theorem 1.1 is tight even in the leading constant. Thus, random linear codes are, with high probability, list-decodable with smaller list sizes than completely random codes.\(^2\)

- **Finer-grained information about the combinatorial structure of random linear codes.** We extend our argument to obtain more information about the distribution of list sizes of random linear codes. More precisely, we obtain high-probability bounds on the number of points $x$ so that the list size at $x$, $L_C(x) := |B(x, pn) \cap C|$, is at least $\ell$.

- **Results for rank-metric codes.** Finally, we adapt our argument for random linear codes to apply to random linear rank-metric codes. As with standard (Hamming-metric) codes, recent work aimed to show that random linear rank-metric codes are nearly as list-decodable as uniformly random codes [Din15, GR17]. Our approach establishes that in fact, random linear binary rank-metric codes are more list-decodable than their uniformly random counterparts in certain parameter regimes, in the sense that the list sizes near capacity are strictly smaller. Along the way, we show that low-rate random linear binary rank-metric codes are list-decodable to capacity, answering a question of [GR17].

On the downside, we note that our arguments only work for binary codes and do not extend to larger alphabets; additionally, our positive results do not establish *average-radius* list-decodability, a stronger notion which was established in some of the previous works [CGV13, Woo13, RW18]. It would be very interesting to extend our results to these settings.

\(^2\)In retrospect, this may not be surprising: for example, it is well-known that random linear codes have better distance than completely random codes. However, the fact that we are able to prove this is surprising to the authors, since it requires taking advantage of the dependence between the codewords, rather than trying to get around it.
1.1 Outline of paper

After a brief overview of the notation in §1.2, we proceed in §2 with a survey of related work for both random linear codes and rank-metric codes, and we formally state our results in this context. In §3, we prove Theorem 2.4, which establishes our upper bound for random linear binary codes. In §4, we expand upon the ideas in the proof of Theorem 2.4 to prove Theorem 2.6, characterizing the list size distribution of random linear codes. In §5, we prove Theorem 2.11, which adapts our upper bound to random linear binary rank-metric codes. To round out the story, we must prove lower bounds on the list sizes of uniformly random codes for both standard and rank-metric codes. We state these lower bounds in Theorems 2.5 and 2.12. These proofs closely follow the approach in [GN14] and are proved in Appendices A and B, respectively.

1.2 Notation

Throughout most of the paper, we are interested in binary codes $C \subseteq \mathbb{F}_2^n$ of block length $n$. The dimension of a code $C$ is defined as $k = \log_2 |C|$, and the rate is the ratio $k/n$. We say that a binary code is linear if it forms a linear subspace of $\mathbb{F}_2^n$. We define a random linear binary code of rate $R$ to be the span of $k = Rn$ independently random vectors $b_1, \ldots, b_k \in \mathbb{F}_2^n$.

For two points $x, y \in \mathbb{F}_2^n$, we use $\Delta(x, y) = \sum_{i=1}^n \mathbb{I}[x_i = y_i]$ to denote the Hamming distance between $x$ and $y$, where, for an event $\mathcal{E}$, $\mathbb{I}[\mathcal{E}]$ is 1 if $\mathcal{E}$ occurs and 0 otherwise. For $x \in \mathbb{F}_2^n, r \in [0, n]$, we define the Hamming ball $B(x, r)$ of radius $r$ centered at $x$ to be $B(x, r) = \{y \in \mathbb{F}_2^n : \Delta(x, y) \leq r\}$, and the volume of $B(x, r)$ to be $\text{Vol}(n, r) := |B(0^n, r)| = \sum_{i=0}^{\lfloor r \rfloor} \binom{n}{i}$. We use the well known bound that, for any $p \in [0, 1]$, $\text{Vol}(n, pn) \leq 2^{H(p)n}$, where $H(p) = -(1 - p) \log_2(1 - p) - p \log_2(p)$ is the binary entropy function. One of our main technical results is about the distribution of list sizes of points $x \in \mathbb{F}_2^n$: given a code $C$ and $p \in (0, 1/2)$, we define the list size of a point $x \in \mathbb{F}_2^n$ to be $L_C(x) := |B(x, pn) \cap C|$.

For $\alpha > 0, \beta \in \mathbb{R}$, let $\exp_\alpha(\beta) := e^{\alpha \beta}$ and assume $\alpha = 1$ when it is omitted. For two sets $A, B \subseteq \mathbb{F}_2^n$, define the sumset $A + B = \{a + b : a \in A, b \in B\}$. When $b \in \mathbb{F}_2^n$, let $A + \{b\}$ denote $A + \{b\}$.

2 Previous Work and Our Results

In §2.1 below, we survey related work on the list-decodability of random linear binary codes, and state our results. In §2.2, we do the same for our results on the list-decodability of random linear binary rank-metric codes.

2.1 Random Linear Codes

The list-decodability of random linear binary codes has been well studied. Here we survey the results that are most relevant for this work. As this work focuses on binary codes, we focus this survey on results for binary codes, even though many of the works mentioned also apply to general $q$-ary codes. We additionally remark that, in contrast to the large alphabet setting [GR08b], capacity achieving binary codes have no known explicit constructions.

A modification of the proof of the list decoding capacity theorem shows that a random linear code of rate $1 - \frac{1}{\log q} - \varepsilon = (p, \exp(O(1/\varepsilon)))$-list decodable [ZP81]. However, whether or not random code

---

3 We note that this definition is slightly different than the standard definition, which is a uniformly random subspace of dimension $k$. However, our definition is easier to work with. Furthermore, since uniformly random $b_1, \ldots, b_k$ have rank strictly less than $k$ with probability at most $2^{-(n-k)}$, and since each dimension $k$ code is represented in the same number of ways, our results hold also for the more standard definition.
linear codes of this rate with list-sizes that do not depend exponentially on \( \epsilon \) remained open for decades: this question was explicitly asked in [Eli91].

A first step was given in the work of Guruswami, Håstad, Sudan and Zuckerman [GHSZ02], who proved via a beautiful potential-function-based-argument that there exist binary linear codes or rate \( 1 - H(p) - \epsilon \) which are \((p, 1/\epsilon)\)-list-decodable. However, this result did not hold with high probability. Our approach relies heavily on the approach of [GHSZ02], and we return to their argument in §3.

Over the next 15 years, a line of work [GHK11, CGV13, Woo13, RW14, RW15, RW18] has focused on the list-decodability (and related properties) of random linear codes, which should hold with high probability. The works most relevant to ours are [GHK11, Woo13], which together more or less settle the question. We state these results here for binary alphabets, although both works address larger alphabets as well.

The first result, of [GHK11], establishes a result for a constant \( p \), bounded away from 1/2.

**Theorem 2.1** (Theorem 2 of [GHK11]). Let \( p \in (0, 1/2) \). Then there exist constants \( C_p, \delta > 0 \) such that for all \( \epsilon > 0 \) and sufficiently large \( n \), for all \( R \leq 1 - H(p) - \epsilon \), if \( C \subseteq \mathbb{F}_2^n \) is a random linear code of rate \( R \), then \( C \) is \((p, C_p/\epsilon)\)-list decodable with probability at least \( 1 - 2^{-\delta n} \).

However, \( C_p \) is not small and tends to \( \infty \) as \( p \) approaches 1/2. The following result of [Woo13] fills in the gap when \( p \) is quite close to 1/2.

**Theorem 2.2** (Theorem 2 of [Woo13]). There exist constants \( C_1, C_2 \) so that for all sufficiently small \( \epsilon > 0 \) and sufficiently large \( n \), for \( p = 1/2 - C_1 \sqrt{\epsilon} \) and for all \( R \leq 1 - H(p) - \epsilon \), if \( C \subseteq \mathbb{F}_2^n \) is a random linear code of rate \( R \), then \( C \) is \((p, C_2/\epsilon)\)-list decodable with probability at least \( 1 - o(1) \).

The list-decoding capacity theorem implies that we cannot hope to take the rate \( R \) substantially larger than \( 1 - H(p) - \epsilon \) and obtain a constant list size. Moreover, the list size \( \Theta(1/\epsilon) \) is optimal for both random linear codes and uniformly random codes [Rud11, GN14]. More precisely, Guruswami and Narayanan show the following theorem (which we have specialized to binary codes).

**Theorem 2.3** (Theorem 20 of [GN14]). Let \( \epsilon > 0 \). A uniformly random binary code of rate \( 1 - H(p) - \epsilon \) is \((p, (1 - H(p))/\epsilon)\)-list decodable with probability at most \( \exp(-\Omega_{p, \epsilon}(n)) \).

We note that for general codes (not uniformly random or random linear) it is still unknown what the “correct” list size \( L \) is in terms of \( \epsilon \), although there are results in particular parameter regimes [Bli86, GV05] and for stronger notions of list-decodability [GN14].

### 2.1.1 Our results for random linear codes

We show that high probability a random linear binary code of rate \( 1 - H(p) - \epsilon \) is \((p, L)\)-list-decodable with \( L \sim H(p)/\epsilon \), while a uniformly random binary code of the same rate requires \( L \geq (1 - \gamma)/\epsilon \). More precisely, the upper bound is as follows (proved in §3).

**Theorem 2.4.** Let \( \epsilon > 0 \), \( p \in (0, 1/2) \), and \( R = 1 - H(p) - \epsilon \). Let \( C \subseteq \mathbb{F}_2^n \) be a random linear code of rate \( R \). Then with probability \( 1 - \exp(-\Omega_\epsilon(n)) \), the code \( C \) is \((p, H(p)/\epsilon + 2)\)-list decodable.

Theorem 2.4 improves upon the picture given by Theorems 2.1 and 2.2 in two ways. First, the leading constant on the list size, which is \( H(p) \), improves over both the constant \( C_p \) from

---

4In fact, in [GN14], Theorem 20 is stated with a list size of \((1 - H(p))/2\epsilon \) as a lower bound. However, the constant can be improved to \( 1 - H(p) \), because the factor of 2 is introduced to handle an additive constant term. Thus, for sufficiently large \( n \) their argument proves the stronger statement stated above.
Theorem 2.1 (which blows up as \( p \to 1/2 \)) and on the constant \( C_2 \) from Theorem 2.2 (which the authors do not see how to make less than 2). Moreover, when \( p \to 1/2 \), Theorem 2.4 improves on Theorem 2.2 in that it decouples \( p \) from \( \varepsilon \): in Theorem 2.2, we must take \( p = 1/2 - O(\sqrt{\varepsilon}) \) and \( R = 1 - \mathcal{H}(p) - \varepsilon \), while in Theorem 2.4, \( p \) and \( \varepsilon \) may be chosen independently. Thus, Theorem 2.4 offers the first true “list-decoding capacity theorem for binary linear codes,” in that it precisely mirrors the quantifiers in Theorem 1.1.

The list size of \( \mathcal{H}(p)/\varepsilon + 2 \) is smaller than the list size of \( 1/\varepsilon \) given by the classical list decoding capacity theorem for uniformly random codes. Further, the following negative result shows that the list size of \( 1/\varepsilon \) given by uniformly random binary codes in the list decoding capacity theorem is tight, even in the leading constant of 1.

**Theorem 2.5.** For any \( p \in (0, 1/2) \) and \( \varepsilon > 0 \), there exists a \( \gamma_{p, \varepsilon} = \exp(-\Omega_p(\frac{1}{\varepsilon})) \) and \( n_{p, \varepsilon} \in \mathbb{N} \) such that for all \( n \geq n_{p, \varepsilon} \), a random linear code \( \mathcal{C} \subseteq \mathbb{F}_2^n \) of rate \( R = 1 - \mathcal{H}(p) - \varepsilon \) is with probability \( 1 - \exp(-\Omega_{p, \varepsilon}(n)) \) not \((p, \frac{1}{\varepsilon})\)-list decodable.

The proof of Theorem 2.5 is obtained by tightening the second moment method proof of [GN14], and is given in Appendix A. Theorem 2.5, combined with Theorem 2.4, implies that, for all \( p \in (0, 1/2) \) and sufficiently small \( \varepsilon \), random linear codes of rate \( 1 - \mathcal{H}(p) - \varepsilon \) with high probability can be list decoded up to distance \( p \) with smaller list sizes than uniformly random codes. Perhaps surprisingly, the difference between the list size upper bound in Theorem 1.1 and the lower bound in Theorem 2.5 is bounded by 1 as \( \varepsilon \to 0 \), implying that the “correct” list size of a uniformly random code is tightly concentrated between \( [1/\varepsilon] \) for small \( \varepsilon \).

We are unaware of results in the literature that give even the existence of binary codes list decodable with list size better than \( H(p)/\varepsilon \). We remark that the Lovasz Local Lemma also gives the existence of \((p, H(p)/\varepsilon)\) list decodable codes, matching our high probability result for random linear codes. Though we suspect this argument is known, we are not aware of a published version, so we include a proof in Appendix C for completeness.

The techniques that we use to prove Theorem 2.4 can be refined to give more combinatorial information about random linear codes. It is our hope that such information will help in further derandomizing constructions of binary codes approaching list-decoding capacity. In §4, we prove the following structural result about random linear binary codes.

**Theorem 2.6.** Let \( \varepsilon, \gamma \in (0, 1) \) be constants, \( p \in (0, 1/2) \), \( L \) be a positive integer, and let \( R = 1 - \mathcal{H}(p) - \varepsilon \). Let \( \mathcal{C} \subseteq \mathbb{F}_2^n \) be a random linear code of rate \( R \). Then with probability \( 1 - \exp(-\Omega_{L}(\gamma \varepsilon n)) \), the code \( \mathcal{C} \) satisfies, for all \( 0 \leq \ell \leq L \),

\[
|\{x \in \mathbb{F}_2^n : L_{\mathcal{C}}(x) \geq \ell\}| \leq 2^n \cdot 2^{-n\ell(1-\gamma)}.
\]

To interpret this result, it is helpful to think of \( \gamma \) close to 0. Intuitively, it says that, with high probability over the choice of a random linear code \( \mathcal{C} \), the number of words \( x \in \mathbb{F}_2^n \) with “list size \( \ell \)” decays approximately exponentially as \( 2^{-n\ell} \). As we show in §4, Theorem 2.4 follows as a corollary of Theorem 2.6 (see Corollary 4.3), but as a warm-up to Theorem 2.6 we present a proof of Theorem 2.4 independent of Theorem 2.6 in §3.

**Remark 2.7.** Theorem 2.6 implies that, with high probability over the choice of the code, for any codeword \( c \in \mathcal{C}, \Pr_{x \in \mathcal{E}(c, pn)}[L_{\mathcal{C}}(x) = 1] \geq 1 - 2^{-n\varepsilon(1-\gamma)-o(1)} \). That is, “most” centers \( x \) within \( pn \) of a codeword \( c \in \mathcal{C} \) are not within \( pn \) of any other codeword \( c' \neq c \). This is in line with the conventional wisdom from the Shannon model: with high probability, random linear codes achieve capacity on the BSC, so for a random linear code, a random center \( x \) obtained by sending a codeword \( c \in \mathcal{C} \) through the binary symmetric channel BSC(\( p \)) has \( L_{\mathcal{C}}(x) \leq 1 \) with high probability. (See also [RU10]). Thus, Theorem 2.6 recovers this intuition for list size 1, and quantitatively extends it to list sizes larger than 1.
2.2 Rank-Metric Codes

As an application of our techniques for random linear codes, we turn our attention to rank metric codes. Rank metric codes, introduced by Del Gian in [Del78], are codes \( C \subseteq \mathbb{F}^{m \times n}_q \); that is, the codewords are \( m \times n \) matrices, where typically \( m \geq n \). The distance between two codewords \( X \) and \( Y \) is given by the rank of their difference: \( \Delta_R(X, Y) := \frac{1}{n} \text{rank}(X - Y) \), where \( \Delta_R \) is called the rank metric. We denote the rank ball by \( B_{q,R}(X, p) := \{ Y \in \mathbb{F}^{m \times n}_q : \Delta_R(X - Y) \leq pn \} \), and say that a rank metric code \( C \subseteq \mathbb{F}^{m \times n}_q \) is \((p, L)\)-list-decodable if \( |B_{q,R}(X, p) \cap C| \leq L \) for all \( X \in \mathbb{F}^{m \times n}_q \). The rate \( R \) of a rank metric code \( C \subseteq \mathbb{F}^{m \times n}_q \) is \( R := \log_q(|C|)/(mn) \).

Rank metric codes generalize standard (Hamming metric) codes, which are simply diagonal rank metric codes. The study of rank metric codes has been motivated by a wide range of applications, including magnetic storage [Rot91], cryptography [GPT91, Loi10, Loi17], space-time coding [LGB03, LK05], and network coding [KK08, SKK08], and distributed storage [SRV12, SRKV13].

The natural “list-decoding capacity” for rank metric codes is \( R = (1 - p)(1 - (n/m)p) \), which is the analog of the Gilbert-Varshamov bound [GY08]. It was shown in [Din15, GR17] that this is achievable by a uniformly random rank metric code.

Theorem 2.8 ([GR17], Proposition A.1.\(^5\)). Let \( \varepsilon > 0 \) and \( p \in (0, 1) \) and suppose that \( m, n \) are sufficiently large compared to \( 1/\varepsilon \). A uniformly random code \( C \subseteq \mathbb{F}^{m \times n}_q \) of rate \( R = (1 - p)(1 - bp) - \varepsilon \) is \((p, [1/\varepsilon])\)-list-decodable with probability at least \( 1 - O(q^{-\varepsilon mn}) \), where \( b = n/m \).

Moreover, it is shown in [Din15] that no linear code can beat this bound.

Theorem 2.9. Let \( b = \lim_{n \to \infty} \frac{n}{m} \) be a constant. Then for any \( R \in (0, 1) \) and \( p \in (0, 1) \), a \((p, L)\)-list-decodable \( \mathbb{F}_q \)-linear rank metric code \( C \subseteq \mathbb{F}^{m \times n}_q \) with rate \( R \) satisfies \( R \leq (1 - p)(1 - bp) \).

There has been a great deal of work aimed at establishing (or refuting) the list-decodability of explicit rank metric codes. It is shown in [WZ13] that Gabidulin codes [Gab85]—the rank-metric analog of Reed-Solomon codes—are not list-decodable to capacity, or even much beyond half their minimum distance. However, there have been works [GX13, GWX16] designing explicit codes meeting the lower bound of Theorem 2.9.

As with standard (Hamming-metric) codes, it is interesting to study the list-decodability of random linear rank-metric codes; we say that \( C \subseteq \mathbb{F}^{m \times n}_q \) is linear if it forms an \( \mathbb{F}_q \)-linear space. Following the approach of [ZP81] for Hamming metric codes, [Din15] shows that random linear rank metric codes of rate \( R = (1 - p)(1 - bp) - \varepsilon \) are \((p, \exp(O(1/\varepsilon)))\)-list-decodable, where as above \( b = n/m \). In a recent paper of Guruswami and Resch [GR17], this result was strengthened to give a list size of \( \exp(1/\varepsilon) \).

Theorem 2.10 ([GR17]). Let \( p \in (0, 1) \) and \( q \geq 2 \). There is some constant \( C_{p,q} \) so that the following holds. For all sufficiently large \( n \), with \( b = n/m \), a random linear rank metric code \( C \subseteq \mathbb{F}^{m \times n}_q \) of rate \( R = (1 - p)(1 - bp) - \varepsilon \) is \((p, C_{p,q}/\varepsilon)\)-list decodable with high probability.

The proof of Theorem 2.10 uses ideas from the approach of [GHK11] to prove Theorem 2.1. This is a beautiful argument, but as with the results of [GHK11], the result of [GR17] suffers from the fact that \( C_{p,q} \) tends to \( \infty \) as \( p \) approaches 1. It is shown in [GR17], Proposition A.2, that when \( p = 1 - \eta \), a uniformly random rank metric code of rate \( R = (\eta - \eta b + \eta^2 b)/2 \) is \((p, 4/(\eta - \eta b + \eta^2 b))\)-list-decodable, and that work poses the question of whether or not a random linear rank metric code can achieve this. Our results, described in the next section, show that the answer is “yes” for \( q = 2 \).

\(^5\) In [GR17], the result is stated with \( L = O(1/\varepsilon) \), but an inspection of the proof shows that we may take the leading constant to be 1.
2.2.1 Our Results for Rank Metric Codes

By applying the techniques in the proof of Theorem 2.4, we prove the following upper bound on the list size of random linear binary rank-metric codes.

**Theorem 2.11.** Let \( p \in (0, 1) \) and choose \( \varepsilon > 0 \). There is a constant \( C_\varepsilon \) so that the following holds. Let \( m \) and \( n \) be sufficiently large positive integers with \( n < m \) and let \( b = n/m \). A random linear rank metric code \( C \subseteq \mathbb{F}_2^{m \times n} \) of rate \( R = (1 - p)(1 - bp) - \varepsilon \) is \((p, \frac{p+bp-bp^2}{\varepsilon} + 2)\)-list decodable with probability at least \( 1 - \exp(-C_\varepsilon mn) \).

Notice that Theorem 2.11 works for all \( p \), improving upon Theorem 2.10. In particular, when \( p = 1 - \eta \), then setting \( \varepsilon = (1 - p)(1 - bp)/2 \) and applying Theorem 2.11 implies that a random linear binary rank metric code of rate \( R = (\eta - \eta b + \eta^2 b)/2 \) is \((p, L)\) list-decodable for \( L \leq \frac{2}{\eta - \eta b + \eta^2 b} \), answering the aforementioned open question of [GR17] in the affirmative.

We also prove a new lower bound on the list size of uniformly random rank-metric codes.

**Theorem 2.12.** Let \( \varepsilon > 0 \) and suppose \( m, n \) are sufficiently large so that \( b = n/m \). Let \( C \subseteq \mathbb{F}_q^{m \times n} \) be a uniformly random rank metric code of rate \( R = (1 - p)(1 - bp) - \varepsilon \). Then \( C \) is \((p, (1 - p)(1 - bp)/\varepsilon - 1)\)-list decodable with probability at most \( \exp(-\Omega_{p, \varepsilon}(n)) \).

Theorem 2.12 again uses the method of [GN14], and we prove it in Appendix B. Together, Theorems 2.11 and 2.12 show that for some values of \( p \), random linear binary rank metric codes have a strictly smaller list size than uniformly random rank metric codes with the same parameters. In particular, the upper bound of Theorem 2.11 is strictly smaller than the lower bound of Theorem 2.12 whenever \( p < \frac{1}{2b} \). For larger values of \( p \), we remark that the list size obtained by Theorem 2.11 is still strictly smaller than the \( 1/\varepsilon \) list size given by uniformly random codes in Theorem 2.8, even though in this case we don’t have a lower bound which proves that this is tight.

3 Simplified result for random linear binary codes

In this section, we prove Theorem 2.4, which we restate here.

**Theorem 3.1 (Theorem 2.4, restated).** Let \( \varepsilon > 0 \), \( p \in (0, 1/2) \), and \( R = 1 - H(p) - \varepsilon \). Let \( C \subseteq \mathbb{F}_2^n \) be a random linear code of rate \( R \). Then with probability \( 1 - \exp(-\Omega_{\varepsilon}(n)) \), the code \( C \) is \((p, H(p)/\varepsilon + 2)\)-list decodable.

Theorem 2.4 also follows from our more refined result, Theorem 2.6. However, since our techniques give a very simple proof of Theorem 2.4 on its own, we begin with just this simple proof. We start by reviewing the approach of [GHSZ02], which is the basis of our proof.

3.1 The approach of [GHSZ02]

Before anything was known about the list-decodability of a typical random linear code, Guruswami, Håstad, Sudan and Zuckerman [GHSZ02] proved the existence of binary linear codes of rate \( 1 - H(p) - \varepsilon \) that are \((p, 1/\varepsilon)\)-list decodable. Their result followed from a beautiful potential-function argument, which is the basis of our approach and which we describe here.

Let \( k := Rn = (1 - H(p) - \varepsilon)n \). We choose vectors \( b_1, \ldots, b_k \) one at a time, so that the code \( C_i := \text{span}(b_1, \ldots, b_i) \) remains “nice”: formally, so that a potential function \( S_C \) remains small. Once we have picked all \( k \) vectors, we set \( C = C_k \), and the fact that \( S_C \) is small implies list-decodability.
Recall that for a code $C$ and $x \in \mathbb{F}_2^n$, we set $L_C(x) = |\mathcal{B}(x, pn) \cap C|$. Define

$$\tilde{S}_C := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} 2^{\varepsilon n L_C(x)}.$$ 

It is not hard to show that for any vectors $b_1, \ldots, b_i \in \mathbb{F}_2^n$,

$$\mathbb{E}_{b_{i+1} \sim \mathbb{F}_2^n} \left[ \tilde{S}_{C_{i+1}}(0,b_{i+1}) | b_1, \ldots, b_i \right] \leq \tilde{S}_{C_i}^2 \quad \text{(1)}$$

That is, when a uniformly random vector $b_{i+1}$ is added to the basis $\{b_1, \ldots, b_i\}$, we expect the potential function not to grow too much. Hence, there exists a choice of vectors $b_1, \ldots, b_k$ so that $\tilde{S}_{C_{i+1}} \leq \tilde{S}_{C_i}^2$ for $i = 0, 1, \ldots, k-1$. \footnote{As a technical detail, one needs to be careful that $b_{i+1} \notin C_i$. One can guarantee $b_{i+1} \notin C_i$ by carefully examining the proof of (1), or use (1) to get a similar equation where we additionally condition $b_{i+1} \notin C_i$.}

As $C_0 = \{0\}$, we have $\tilde{S}_{C_0} \leq 1 + 2^{-n(1-H(p)-\varepsilon)}$. Setting $C = C_k = \text{span}(b_1, \ldots, b_k)$, we have

$$\tilde{S}_C \leq \tilde{S}_{C_0}^k \leq \left( 1 + 2^{-n(1-H(p)-\varepsilon)} \right)^{2^k} \leq \exp \left( 2^{k-n(1-H(p)-\varepsilon)} \right) \leq \varepsilon$$

by our choice of $k$. This implies that $\sum_x 2^{\varepsilon n L_C(x)} \leq \varepsilon \cdot 2^n$, and in particular, for all $x \in \mathbb{F}_2^n$, we have $2^{\varepsilon n L_C(x)} \leq \varepsilon \cdot 2^n$. Thus, for all $x$, $L_C(x) \leq \frac{1}{4} + o(1)$, as desired.

The approach of [GHSZ02] is extremely clever, but these ideas have not, to the best of our knowledge, been used in subsequent work on the list-decodability of random linear codes. One reason is that the crux of the argument, which is (1), holds in expectation, and it was not clear how to show that it holds with high probability; thus, the result remained existential, and other techniques were introduced to study typical random codes [GHK11, CGV13, Woo13, RW14, RW18].

### 3.2 Proof of Theorem 2.4

We improve the argument of [GHSZ02] in two ways. First, we show that in fact, (1) essentially holds with high probability over the choice of $b_{i+1}$, which allows us to use the approach sketched above for random linear codes. Second, we introduce one additional trick which takes advantage of the linearity of the code in order to reduce the constant in the list size from 1 to $H(p)$. Before diving into the details, we briefly describe the main ideas.

The first improvement follows from looking at the potential function in the right way. In this paragraph, all $o(1)$ terms are exponentially small in $n$. Our goal is $\tilde{S}_{C_k} \leq O(1)$. Write $\tilde{S}_{C_i} = 1 + \tilde{T}_{C_i}$. By above, $\tilde{T}_{C_0} = \tilde{S}_{C_0} - 1 = o(1)$. We show that with high probability, for all $i \leq k$, we have $\tilde{T}_{C_i} = o(1)$. In the [GHSZ02] argument we have

$$\mathbb{E} \tilde{S}_{C_{i+1}} \leq \tilde{S}_{C_i}^2 = (1 + \tilde{T}_{C_i})^2 = 1 + 2\tilde{T}_{C_i} (1 + o(1)),$$

and so $\mathbb{E} \tilde{T}_{C_{i+1}} = 2\tilde{T}_{C_i} (1 + o(1))$. One can show that, always, $2\tilde{T}_{C_i} \leq \tilde{T}_{C_{i+1}}$. By Markov’s inequality, $\tilde{T}_{C_{i+1}} = 2\tilde{T}_{C_i} (1 + o(1))$ with probability $1 - o(1)$, for appropriately chosen $o(1)$ terms. Union bounding over the $o(1)$ failure probabilities in the $k$ steps, we conclude that $\tilde{T}_{C_i}$ grows roughly as slowly as in the existential argument, giving the desired list decodability.

The second improvement follows from the linearity of the code. In the last step of the [GHSZ02] argument, we replace the summation “$\sum_x$” in $\sum_x 2^{\varepsilon n L_C(x)} \leq \varepsilon \cdot 2^n$ with a “$\forall x$.” We can save a bit because, by linearity, the contribution $2^{\varepsilon n L_C(x)}$ is the same for all $x$ in a coset $y + C$. 

8
Now we go through the details. It is convenient to change the definition of the potential function very slightly: losing the tilde, define, for a code \( C \subset \mathbb{F}_2^n \),
\[
A_C(x) := 2^{-\frac{\epsilon n L_C(x)}{1+\epsilon}} \quad \text{and} \quad S_C := \mathbb{E}_{x \sim \mathbb{F}_2^n} [A_C(x)].
\]
The term \( S_C \) differs from the term \( \tilde{S}_C \) above in that \( A_C(x) \) has an extra factor of \( \frac{1}{1+\epsilon} \) in the exponent. This is an extra “slack” term that helps guarantee a high probability result under the same parameters. However, this definition does not change how the potential function behaves. In particular, we still have the following lemma:

**Lemma 3.2** (Following [GHSZ02]). For all linear \( C \subseteq \mathbb{F}_2^n \) and all \( b \in \mathbb{F}_2^n \),
\[
\begin{align*}
L_{C+\{0,b\}}(x) & \leq L_C(x) + L_C(x+b) \\
A_{C+\{0,b\}}(x) & \leq A_C(x) \cdot A_C(x+b),
\end{align*}
\]
with equality if and only if \( b \notin C \).

**Proof.** To see (2), notice that
\[
L_{C+\{0,b\}}(x) = |\mathcal{B}(x, pn) \cap (C \cup (C + b))| \\
\leq |\mathcal{B}(x, pn) \cap C| + |\mathcal{B}(x, pn) \cap (C + b)| \\
= |\mathcal{B}(x, pn) \cap C| + |\mathcal{B}(x + b, pn) \cap C| \\
= L_C(x) + L_C(x+b),
\]
with equality in the second line if and only if \( b \notin C \). Inequality (3) follows as a consequence of (2), and this proves the lemma. \( \square \)

We additionally define
\[
B_C(x) := A_C(x) - 1 \quad \text{and} \quad T_C := S_C - 1.
\]
As noted above, it is helpful to think of \( T_C \) as a very small term; we would like to show—in accordance with (1)—that \( T_C \) approximately doubles each time we add a basis vector. Note that
\[
S_{\{0\}} = 1 + \left(2^{\frac{\epsilon n}{1+\epsilon}} - 1\right) \cdot \frac{\text{Vol}(n, pn)}{2^n} < 1 + 2^{\frac{\epsilon n}{1+\epsilon}} \cdot \frac{2^{H(p)n}}{2^n} = 1 + 2^{-n(1-H(p)-\frac{\epsilon}{1+\epsilon})}. \tag{4}
\]
With these definitions, we can prove the concentration result we need.

**Lemma 3.3.** Let \( p, \epsilon \), and \( R = 1 - H(p) - \epsilon \) be as in the statement of Theorem 2.4. Let \( C_{Rn} \subset \mathbb{F}_2^n \) be a random linear code of rate \( R \). Then with probability \( 1 - \exp(-\Omega_\epsilon(n)) \), the code \( C_{Rn} \) satisfies \( S_{C_{Rn}} \leq 2 \).

Before we prove Lemma 3.3 we show that this implies Theorem 2.4.

**Proof of Theorem 2.4 given Lemma 3.3.** We show that, for a binary linear code \( C \) of rate \( R = 1 - H(p) - \epsilon \), \( S_C \leq 2 \) implies \( (p, \frac{H(p)}{\epsilon} + 2) \)-list decodability. Suppose for sake of contradiction that a code \( C \) satisfies \( S_C \leq 2 \) and there exists \( x^* \in \mathbb{F}_2^n \) such that \( |\mathcal{B}(x^*, pn) \cap C| > H(p)/\epsilon + 2 \). For all \( x \in \mathbb{F}_2^n \) and \( c \in C \), we have
\[
|\mathcal{B}(x + c, pn) \cap C| = |\mathcal{B}(x, pn) \cap (C - c)| = |\mathcal{B}(x, pn) \cap C|,
\]
so \(|\mathcal{B}(x^* + c, pn) \cap \mathcal{C}| > H(p)/\varepsilon\) for all \(c \in \mathcal{C}\). If \(S_{\mathcal{C}} \leq 2\), then we have

\[
2^{n+1} \geq 2^n S_{\mathcal{C}} = \sum_{x \in \mathbb{F}_2^n} \exp \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot |\mathcal{B}(x, pn) \cap \mathcal{C}| \right) \\
\geq \sum_{c \in \mathcal{C}} \exp \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot |\mathcal{B}(x^* + c, pn) \cap \mathcal{C}| \right) \\
\geq \sum_{c \in \mathcal{C}} \exp \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot (H(p)/\varepsilon + 2) \right) \\
= |\mathcal{C}| \cdot \exp \left( n \cdot \frac{H(p) + 2\varepsilon}{1 + \varepsilon} \right) \\
= \exp \left( n \left( 1 + \frac{\varepsilon}{1 + \varepsilon}(1 - H(p) - \varepsilon) \right) \right).
\]

which is a contradiction for large enough \(n\). \(\Box\)

Finally, we prove Lemma 3.3.

**Proof of Lemma 3.3.** As in §3.1, let \(b_1, b_2, \ldots, b_k \in \mathbb{F}_2^n\) be independently and uniformly chosen, and let \(\mathcal{C}_i = \text{span}\{b_1, \ldots, b_i\}\).

**Lemma 3.4.** Suppose that \(\mathcal{C}\) is fixed and satisfies \(T_{\mathcal{C}} < 1\), so that \(S_{\mathcal{C}} < 2\). Then

\[
\Pr_{b \sim \mathbb{F}_2^n} \left[ S_{\mathcal{C} + \{0, b\}} > 1 + 2T_{\mathcal{C}} + T_{\mathcal{C}}^{1.5} \right] < T_{\mathcal{C}}^{0.5}.
\]

**Proof.** By Lemma 3.2, for all \(b\),

\[
S_{\mathcal{C} + \{0, b\}} = \mathbb{E}_x \left[ A_{\mathcal{C} + \{0, b\}}(x) \right] \\
\leq \mathbb{E}_x \left[ A_{\mathcal{C}}(x) A_{\mathcal{C}}(x + b) \right] \\
= \mathbb{E}_x \left[ (1 + B_{\mathcal{C}}(x)) \cdot (1 + B_{\mathcal{C}}(x + b)) \right] \\
= \mathbb{E}_x \left[ 1 + B_{\mathcal{C}}(x) + B_{\mathcal{C}}(x + b) + B_{\mathcal{C}}(x) B_{\mathcal{C}}(x + b) \right] \\
= 1 + 2T_{\mathcal{C}} + \mathbb{E}_x \left[ B_{\mathcal{C}}(x) B_{\mathcal{C}}(x + b) \right].
\]

Over the randomness of \(b\) and \(x\), we have \(x\) and \(x + b\) are statistically independent and uniform over \(\mathbb{F}_2^n\), so we have

\[
\mathbb{E}_{b, x} \left[ B_{\mathcal{C}}(x) B_{\mathcal{C}}(x + b) \right] = \mathbb{E}_{b, x} \left[ B_{\mathcal{C}}(x) \right] \cdot \mathbb{E}_{b, x} \left[ B_{\mathcal{C}}(x + b) \right] = T_{\mathcal{C}}^2.
\]

As \(B_{\mathcal{C}}\) is always nonnegative, we have, by Markov’s inequality,

\[
\Pr_b \left[ S_{\mathcal{C} + \{0, b\}} > 1 + 2T_{\mathcal{C}} + T_{\mathcal{C}}^{1.5} \right] \leq \Pr_b \left[ \mathbb{E}_x \left[ B_{\mathcal{C}}(x) B_{\mathcal{C}}(x + b) \right] > T_{\mathcal{C}}^{1.5} \right] < \frac{T_{\mathcal{C}}^2}{T_{\mathcal{C}}^{1.5}} = T_{\mathcal{C}}^{0.5}. \quad \Box
\]

Returning to the proof of Lemma 3.3, consider the sequence

\[
\delta_0 := 2^{-n(1 - H(p) - \varepsilon/2)} \\
\delta_i := 2\delta_{i-1} + \delta_{i-1}^{1.5}
\]

\[10\]
We can verify by induction that for \( i \leq n(1 - H(p) - \varepsilon) \), we have \( \delta_i < 2^{i+1}\delta_0 < 2^{-\frac{2n}{3}} \). To see this, notice that the base case is trivial, and if \( \delta_j < 2^{i+1}\delta_0 \) for \( j < i \), we have

\[
\delta_i = 2\delta_{i-1}(1 + \delta_i^{0.5}) = 2^i\delta_0 \prod_{j=0}^{i-1}(1 + \delta_j^{0.5}) \leq 2^i\delta_0 \cdot \exp \left( \sum_{j=0}^{i-1}\delta_j^{0.5} \right) < 2^{i+1}\delta_0.
\]

In the first two equalities, we applied the definitions of \( \delta_i \) and \( \delta_{i-1}, \ldots, \delta_1 \), respectively. In the first inequality, we used the estimate \( 1 + z \leq e^z \), and in the second we used the inductive hypothesis \( \delta_j < 2^{-\frac{2n}{3}} \) for \( j < i \). By this induction, we conclude that, if \( k = n(1 - H(p) - \varepsilon) \), then \( \delta_k < 2^{-\frac{2n}{3}} \).

Let \( b_1, \ldots, b_k \in \mathbb{F}_2^n \) be randomly chosen vectors, and let \( C_i = \text{span}(b_1, \ldots, b_i) \) with \( C_k = C \). By Lemma 3.4, conditioned on a fixed \( C_i \) satisfying \( T_{C_i} \leq \delta_i \), we have, with probability at least \( T_{C_i+1} > \delta_i+1 \). Furthermore, \( T_{C_0} \leq \delta_0 \) by the initial condition (4). Thus, with probability at least

\[
1 - (\delta_0^{0.5} + \delta_1^{0.5} + \cdots + \delta_k^{0.5}) > 1 - k2^{-\varepsilon^2n/2} \geq 1 - 2^{-\Omega(n)}
\]

we have \( T_{C_i} \leq \delta_i \) for all \( i \). In particular, \( T_C = T_{C_k} < \delta_k < 2^{-\frac{2n}{3}} \). Thus, \( S_C = 1 + T_C \leq 2 \) with probability \( 1 - \exp(-\Omega(n)) \), completing the proof of Lemma 3.3.

**Remark 3.5.** We do not see how to extend this proof to larger alphabets. If, for example, \( q = 3 \), then Lemma 3.4 would need to say \( \Pr[S_{C+\{0, b, 2b\}} > 1 + 3T_C + o(T_C)] < o(1) \). However, the same proof would fail to establish this, as we can no longer separate the expectation in (5); that is we cannot say

\[
\mathbb{E}_b \mathbb{E}_x [B_C(x)B_C(x+b)B_C(x+2b)] = \mathbb{E}_b \mathbb{E}_x [B_C(x)] \cdot \mathbb{E}_b \mathbb{E}_x [B_C(x+b)] \cdot \mathbb{E}_b \mathbb{E}_x [B_C(x+2b)] = T_C^3.
\]

4 **Characterizing the list size distribution**

In this section we establish Theorem 2.6. Define

\[
P_C^{(\ell)} := \frac{2^{-n}|\{x : L_C(x) = \ell\}|}{2^{-n}}, \quad P_C^{(\geq \ell)} := \sum_{i=\ell}^{\infty} P_C^{(i)}, \quad Q_C^{(\geq \ell)} := \sum_{i=\ell}^{\infty} i \cdot P_C^{(i)}
\]

The goal (per Theorem 2.6) is to bound \( P_C^{(\geq \ell)} \); in our argument, it is more convenient to work with \( Q_C^{(\geq \ell)} \), which is a proxy for \( P_C^{(\geq \ell)} \). We note a few useful properties of these definitions.

**Proposition 4.1.** Suppose that \( C \subseteq \mathbb{F}_2^n \) is a linear code. Then the following hold for \( \ell \geq 1 \).

1. \( Q_C^{(\geq \ell)} = 2^{-n} \cdot \sum_{x \in \mathbb{F}_2^n : L_C(x) \geq \ell} L_C(x) = \mathbb{E}_x [\mathbb{1} [L_C(x) \geq \ell] \cdot L_C(x)] \)
2. \( Q_C^{(\geq 1)} = 2^{-n} \cdot \text{Vol}(n, pn) \cdot 2^k \leq 2^{-n(1-H(p)) + k} \)
Proof. To see Item 1, notice that
\[
Q^{(\ell)}_C = 2^{-n} \sum_{i=\ell}^\infty i \cdot |\{x \mid L_C(x) = i\}|
\]
\[
= 2^{-n} \sum_{i=\ell}^\infty \sum_{x \in \mathbb{F}_2^p} i \cdot \mathbb{I}[L_C(x) = i]
\]
\[
= 2^{-n} \sum_{x : L_C(x) \geq \ell} L_C(x).
\]
To see Item 2, we begin with Item 1 and derive
\[
Q^{(\geq 1)}_C = 2^{-n} \sum_{x : L_C(x) \geq 1} L_C(x)
\]
\[
= 2^{-n} \sum_{x \in \mathbb{F}_2^p} \mathbb{I}[x + v \in C]
\]
\[
= 2^{-n} \sum_{v \in \mathbb{B}(x, pn)} \mathbb{I}[x + v \in C]
\]
\[
= 2^{-n} \sum_{v \in \mathbb{B}(x, pn)} |C|
\]
\[
= 2^{-n} \cdot \text{Vol}(n, pn) \cdot 2^k.
\]

Using these properties, we are now can state and prove Theorem 4.2, which implies Theorem 2.6.

**Theorem 4.2.** Fix \( L \geq 0 \). There exists a constant \( C_L \) depending only on \( L \) so that, for all \( \gamma \in (0,1) \) and sufficiently large \( n \), the following holds. Suppose that \( k \leq (1 - \gamma)n \). If \( C \subseteq \mathbb{F}_2^n \) is a random linear code of dimension \( k \), then with probability \( 1 - \exp(-C_L \gamma n) \), for all \( 1 \leq \ell \leq L \),
\[
Q^{(\geq \ell)}_C \leq \left( 2^{-n(1-H(p))} \cdot 2^k \right)^{\ell} \cdot 2^{\gamma^2 n}.
\]

Before we prove Theorem 4.2, we explain why it implies Theorem 2.6. By setting \( k = n(1 - H(p) - \varepsilon) \) and \( \gamma = \varepsilon \gamma/L \) in Theorem 4.2, we obtain that with high probability, for all \( 1 \leq \ell \leq L \),
\[
P^{(\geq \ell)}_C \leq Q^{(\geq \ell)}_C \leq 2^{-n\varepsilon \ell} \cdot 2^{\varepsilon^2 \ell^2 n} \leq 2^{-n \varepsilon (1 - \gamma')},
\]
which is Theorem 2.6. Theorem 4.2 also implies Theorem 2.4:

**Corollary 4.3** (Theorem 2.4). Let \( \varepsilon > 0 \), \( p \in (0,1/2) \), and \( R = 1 - H(p) - \varepsilon \). Let \( C \subseteq \mathbb{F}_2^n \) be a random linear code of rate \( R \). Then with probability \( 1 - \exp(-\Omega_\varepsilon(n)) \), the code \( C \) is \((p, H(p)/\varepsilon + 2)\)-list decodable.

**Proof.** Let \( L = H(p)/\varepsilon + 2 \) and choose \( \gamma \ll L^{-3} \), and let \( C \subseteq \mathbb{F}_2^n \) be a random linear code of dimension \( k = n(1 - H(p) - \varepsilon) \). Then \( k \leq (1 - \gamma)n \), so, by Theorem 4.2, with probability at least \( 1 - \exp(-C_L \gamma n) \), code \( C \) satisfies (6) with \( k = n(1 - H(p) - \varepsilon) \). In particular, choosing \( \ell = L \), this means that
\[
Q^{(\geq L)}_C \leq 2^{L(-n(1-H(p)) + k + \gamma L^2 n)} = 2^{-n L \varepsilon + \gamma L^2 n} < 2^{-n(1-R)}
\]
Suppose there exists \( x \in \mathbb{F}_2^n \) such that \( L_C(x) \geq L \). Then \( L_C(x + c) \geq L \) for all \( c \in C \), so that \( Q^{(\geq L)}_C > P^{(\ell)}_C \geq 2^{-n(1-R)} \). This is a contradiction, so there are no \( x \in \mathbb{F}_2^n \) such that \( L_C(x) \geq L \).
Proof of Theorem 4.2. The proof of Theorem 4.2 proceeds by induction on \( k \), with the inductive hypothesis that (6) holds for all \( 1 \leq \ell \leq L \). We begin with the base case by noting that (6) is satisfied for \( k = 0 \), for all \( 1 \leq \ell \leq L \). To see this, notice that for \( \ell = 1 \), we have \( Q_{\{0\}}^{(\geq \ell)} \leq 2^{-n(1-H(p))} \) by Proposition 4.1, Item 2, which satisfies (6) for \( k = 0, \ell = 1 \). For \( \ell \geq 2 \), \( Q_{\{0\}}^{(\geq \ell)} = 0 \), and so again (6) holds.

Now that we have established the base case of \( k = 0 \), we proceed by induction. Lemma 4.4 provides the inductive step; similar to the approach in §3, it shows that at every step the “expected behavior” holds with high probability.

**Lemma 4.4.** Let \( \gamma > 0 \), and suppose that \( \mathcal{C} \subseteq \mathbb{F}_2^n \) is a linear code of dimension \( k \leq (1-\gamma)n \) such that for all \( 1 \leq \ell \leq L \), we have

\[
Q_{\mathcal{C}}^{(\geq \ell)} \leq \left(2^{-n(1-H(p))} \cdot 2^k\right)^\ell \cdot 2^2\gamma^2n. \tag{7}
\]

Then, for a uniformly chosen \( b \in \mathbb{F}_2^n \), with probability at least \( 1 - 4 \cdot L^3 \cdot 2^{-\gamma n} \) over the choice of \( b \), we have, for all \( 1 \leq \ell \leq L \),

\[
Q_{\mathcal{C}+\{0,b\}}^{(\geq \ell)} \leq \left(2^{-n(1-H(p))} \cdot 2^{k+1}\right)^\ell \cdot 2^2\gamma^2n. \tag{8}
\]

**Proof.** By Proposition 4.1, Item 2, (8) always holds for \( \ell = 1 \), so suppose that \( \ell \geq 2 \). We have, for any \( b \in \mathbb{F}_2^n \),

\[
Q_{\mathcal{C}+\{0,b\}}^{(\geq \ell)} = \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}+\{0,b\}}(x) \geq \ell] \cdot L_{\mathcal{C}+\{0,b\}}(x) \right] \quad \text{(By Prop. 4.1, Item 1)} \\
\leq \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) + L_{\mathcal{C}}(x+b) \geq \ell] \cdot (L_{\mathcal{C}}(x) + L_{\mathcal{C}}(x+b)) \right] \quad \text{(By Lemma 3.2)} \\
= \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) + L_{\mathcal{C}}(x+b) \geq \ell] \cdot L_{\mathcal{C}}(x) + \mathbb{I}[L_{\mathcal{C}}(x) + L_{\mathcal{C}}(x+b) \geq \ell] \cdot L_{\mathcal{C}}(x+b) \right] \\
= 2 \cdot \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) \geq \ell] \cdot L_{\mathcal{C}}(x) \right] + \sum_{i=0}^{\ell-1} \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) = i] \cdot \mathbb{I}[L_{\mathcal{C}}(x+b) \geq \ell - i] \cdot L_{\mathcal{C}}(x+b) \right] \\
= 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) = i] \cdot \mathbb{I}[L_{\mathcal{C}}(x+b) \geq \ell - i] \right], \tag{9}
\]

where we have used Proposition 4.1, Item 1 in the final line. Since the only inequality above is an application of Lemma 3.2, which is an equality when \( b \notin \mathcal{C} \), the derivation above is an equality when \( b \notin \mathcal{C} \). Thus, when \( b \notin \mathcal{C} \), we have

\[
Q_{\mathcal{C}+\{0,b\}}^{(\geq \ell)} = 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot \mathbb{E}_x \left[ \mathbb{I}[L_{\mathcal{C}}(x) = i] \cdot \mathbb{I}[L_{\mathcal{C}}(x+b) \geq \ell - i] \right] \geq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)}. \tag{10}
\]

This fact (10) is useful later. For now, we move on with no assumption on \( b \). Taking expectations
on both sides of (9), we see

\[
\mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \right] \leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot \mathbf{E}_b \left[ \mathbb{I}[L_{\mathcal{C}}(x) = i] \cdot \mathbb{I}[L_{\mathcal{C}}(x + b) \geq \ell - i] \right]
\]

\[
= 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot P_{\mathcal{C}}^{(i)} \cdot P_{\mathcal{C}}^{(\geq \ell-i)}
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot Q_{\mathcal{C}}^{(\geq i)} \cdot Q_{\mathcal{C}}^{(\geq \ell-i)}
\]

Continuing, we bound

\[
\mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \right] \leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot \sum_{i=0}^{\ell-1} i \cdot Q_{\mathcal{C}}^{(\geq i)} \cdot Q_{\mathcal{C}}^{(\geq \ell-i)}
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L \cdot \sum_{i=1}^{\ell-1} Q_{\mathcal{C}}^{(\geq i)} \cdot Q_{\mathcal{C}}^{(\geq \ell-i)}
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L \cdot \sum_{i=1}^{\ell-1} \left( 2^{-(n(1-H(p)) \cdot 2^k)^i} \cdot 2 \gamma^{i^2 n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^{\ell-i} \cdot 2 \gamma^{(\ell-i)^2 n} \right) \quad \text{(By (7))}
\]

\[
= 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L \cdot \sum_{i=1}^{\ell-1} \left( 2^{-(n(1-H(p)) \cdot 2^k)^i} \cdot 2 \gamma^{i^2 n} \cdot 2^{-\gamma n} \right)
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L^2 \cdot 2^{-\gamma n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n}
\]

\[
\operatorname{(*)} \leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L^2 \cdot 2^{-\gamma n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n}
\]

The above derivation holds whether or not \( b \in \mathcal{C} \). Thus,

\[
\operatorname{(*)} \geq \mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \right] = \mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \mid b \notin \mathcal{C} \right] \Pr_b \left[ b \notin \mathcal{C} \right] + \mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \mid b \in \mathcal{C} \right] \Pr_b \left[ b \in \mathcal{C} \right]
\]

\[
\geq \mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \mid b \notin \mathcal{C} \right] \Pr_b \left[ b \notin \mathcal{C} \right]
\]

\[
= \mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \mid b \notin \mathcal{C} \right] \cdot \left( 1 - \frac{|C|}{2n} \right)
\]

and so

\[
\mathbf{E}_b \left[ Q_{\mathcal{C} + \{0, b\}}^{(\geq \ell)} \mid b \notin \mathcal{C} \right]
\]

\[
\leq \frac{2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L^2 \cdot 2^{-\gamma n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n}}{1 - \frac{|C|}{2n}} \quad \text{(11)}
\]

\[
\leq \left( 1 + \frac{2|C|}{2n} \right) \left( 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 2 \cdot L^2 \cdot 2^{-\gamma n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n} \right)
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n} \left( \frac{4|C|}{2n} + 3 \cdot L^2 \cdot 2^{-\gamma n} \right) \quad \text{(By (7))}
\]

\[
\leq 2 \cdot Q_{\mathcal{C}}^{(\geq \ell)} + 4 \cdot L^2 \cdot 2^{-\gamma n} \cdot \left( 2^{-n(1-H(p)) \cdot 2^k} \right)^\ell \cdot 2 \gamma^{\ell^2 n} \quad \text{(Using } |C| = 2^k \leq 2^{n(1-\gamma)})
\]
In line (11) above, we used the fact that $1/(1 - x) \leq 1 + 2x$ for all $0 \leq x \leq 1/2$, along with the fact that, since $k < n$, we have $|C|/2^n = 2^{k-n} \leq 1/2$. By (10), when $b \notin C$, the quantity $Q_{C+b}^{(\geq \ell)} - 2Q_{C}^{(\geq \ell)}$ is nonnegative. Hence, we may apply Markov’s inequality to obtain

$$
\Pr_{b} \left[ Q_{C+b}^{(\geq \ell)} - 2Q_{C}^{(\geq \ell)} \geq \left( 2^{-n(1-H(p))} \cdot 2^k \right) \ell \cdot 2^{-\gamma \ell^2 n} | b \notin C \right] \leq 4 \cdot L^2 \cdot 2^{-\gamma n}.
$$

When $b \in C$, we have $Q_{C+b}^{(\geq \ell)} = Q_{C}^{(\geq \ell)}$. Thus,

$$
\Pr_{b} \left[ Q_{C+b}^{(\geq \ell)} - 2Q_{C}^{(\geq \ell)} \geq \left( 2^{-n(1-H(p))} \cdot 2^k \right) \ell \cdot 2^{-\gamma \ell^2 n} | b \in C \right] = 0.
$$

Together these imply

$$
\Pr_{b} \left[ Q_{C+b}^{(\geq \ell)} - 2Q_{C}^{(\geq \ell)} \geq \left( 2^{-n(1-H(p))} \cdot 2^k \right) \ell \cdot 2^{-\gamma \ell^2 n} \right] \leq 4 \cdot L^2 \cdot 2^{-\gamma n}. \quad (12)
$$

Thus, with probability at least $1 - 4L^22^{-\gamma n}$, we have

$$
Q_{C+b}^{(\geq \ell)} \leq 2Q_{C}^{(\geq \ell)} + \left( 2^{-n(1-H(p))} \cdot 2^k \right) \ell \cdot 2^{-\gamma \ell^2 n}
\leq 3 \left( 2^{-n(1-H(p))} \cdot 2^k \right) \ell \cdot 2^{-\gamma \ell^2 n}
\leq \left( 2^{-n(1-H(p))} \cdot 2^{k+1} \right) \ell \cdot 2^{-\gamma \ell^2 n}
$$

where the first inequality is from (12), the second inequality is by the assumption (7), and the final inequality is because $\ell \geq 2$. This completes the proof of Lemma 4.4. □

Returning to the proof of Theorem 4.2, call a code $C$ of dimension $k$ good if (6) holds for all $1 \leq \ell \leq L$. Lemma 4.4 states that if $C$ is good, then $C + \{0, b\}$ fails to be good with probability at most $4L^32^{-\gamma n}$ over the choice of a uniformly random $b \in \mathbb{F}_2^s$.

Since we have already shown that $\{0\}$ is good at the beginning of the proof, it follows from the union bound that a random linear binary code $C = \text{span}(b_1, \ldots, b_k)$ fails to be good with probability at most $k \cdot 4L^32^{-\gamma n} = 2^{-\Omega(\gamma n)}$. This completes the proof of Theorem 4.2. □

5 List-decoding rank metric codes

Here, we prove Theorem 2.11, which is restated below.

**Theorem 5.1** (Theorem 2.11, restated). Let $p \in (0,1)$ and choose $\varepsilon > 0$. There is a constant $C_\varepsilon$ so that the following holds. Let $m$ and $n$ be sufficiently large positive integers with $n < m$ and let $b = n/m$. A random linear rank metric code $C \subseteq \mathbb{F}_2^{m \times n}$ of rate $R = (1 - p)(1 - bp) - \varepsilon$ is $(p, \frac{p + bp - bp^2}{\varepsilon} + 2)$-list decodable with probability at least $1 - \exp(-C_\varepsilon mn)$.

Let $k = R \cdot mn$. Let $Y_1, \ldots, Y_k \in \mathbb{F}_2^{m \times n}$ be i.i.d.uniformly random matrices, and let $C_k = \text{span}(Y_1, \ldots, Y_k)$ so that $C_k$ is a random linear rank metric code of rate $R = (1 - p)(1 - bp) - \varepsilon$. We overload the notation from §3 to be matrix-valued. For a rank metric code $C \subseteq \mathbb{F}_2^{m \times n}$ and
Given Lemma 5.3, we have
\[q^{mn(p + pb - p^2b)} \leq |B_R(X, pn)| \leq 4 \cdot q^{mn(p + pb - p^2b)}\]

Lemma 5.2 yields the identity
\[S(0) \leq 1 + 4 \cdot 2^{-mn(1 - p - pb + p^2b)},\]  \hspace{1cm} (13)

Following the proof of Theorem 2.4, we prove the following lemma which mirrors Lemma 3.3.

Lemma 5.3. Let \( p \in (0, 1), \gamma \in (0, 1) \). Let \( n < m \) be positive integers with \( b = n/m \). With probability \( 1 - \exp(-\Omega_\varepsilon(n)) \), a random linear rank metric code \( C \subseteq \mathbb{F}_q^{m \times n} \) of rate \( R = (1 - p)(1 - bp) - \varepsilon \) satisfies \( S_C \leq 2 \).

As before, the Theorem 2.11 follows:

Proof of Theorem 2.11 given Lemma 5.3. We show that \( S_C \leq 2 \) implies \((p, \frac{p + bp - bp^2}{\varepsilon} + 2)\)-list decodability. Suppose for sake of contradiction that a code \( C \) satisfies \( S_C \leq 2 \) and there exists \( X^* \in \mathbb{F}_q^{m \times n} \) such that \( |B_R(X^*, pn) \cap C| > \frac{p + bp - bp^2}{\varepsilon} + 2 \). We know that for all \( X \in \mathbb{F}_q^{m \times n} \) and \( C \subseteq C \), we have
\[|B_R(X + C, pn) \cap C| = |B_R(X, pn) \cap (C - C)| = |B_R(X, pn) \cap C|,\] so \( |B_R(X^* + C, pn) \cap C| > \frac{p + bp - bp^2}{\varepsilon} + 2 \) for all \( C \in C \).

If \( S_C \leq 2 \), then we have
\[2^{n+1} \geq 2^n S_C \geq \sum_{X \in \mathbb{F}_q^{m \times n}} \exp_2 \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot |B_R(X, pn) \cap C| \right)\]
\[\geq \sum_{C \subseteq C} \exp_2 \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot |B_R(X^* + C, pn) \cap C| \right)\]
\[\geq \sum_{C \subseteq C} \exp_2 \left( n \cdot \frac{\varepsilon}{1 + \varepsilon} \cdot \left( \frac{p + bp - bp^2}{\varepsilon} + 2 \right) \right)\]
\[= |C| \cdot \exp_2 \left( n \cdot \frac{p + bp - bp^2 + 2\varepsilon}{1 + \varepsilon} \right)\]
\[= \exp_2 \left( n \left( 1 + \frac{\varepsilon}{1 + \varepsilon} \right) \frac{(1 - p)(1 - bp) - \varepsilon}{1 + \varepsilon} \right),\]
which is a contradiction for large enough \( n \). \( \square \)
Before we prove Lemma 5.3, we state an analog to Lemmas 3.2.

**Lemma 5.4.** We have, for all linear \( C \) and all \( Y \)
\[
L_{C,\{0,Y\}}(X) \leq L_C(X) + L_C(X + Y)
\]
\[
A_{C,\{0,Y\}}(X) \leq A_C(X) \cdot A_C(X + Y),
\]
with equality if and only if \( Y \not\in C \).

The proof of Lemma 5.4 follows identically to that of Lemma 3.2.

**Proof of Lemma 5.4.** We begin with an analogue of Lemma 3.2.

**Lemma 5.5.** Suppose that \( C \) is fixed and satisfies \( T_C < 1 \), so that \( S_C < 2 \). Then
\[
\Pr_{Y \sim \mathbb{F}_2^{m \times n}} \left[ S_{C,\{0,Y\}} > 1 + 2T_C + T_C^{1.5} \right] < T_C^{0.5}.
\]

**Proof.** By Lemma 5.4, for all \( Y \),
\[
S_{C,\{0,Y\}} \leq \mathbb{E}_X \left[ A_{C,\{0,Y\}}(X) \right]
\leq \mathbb{E}_X \left[ A_C(X)A_C(X + Y) \right]
= \mathbb{E}_X \left[ (1 + B_C(X)) \cdot (1 + B_C(X + Y)) \right]
= \mathbb{E}_X \left[ 1 + B_C(X) + B_C(X + Y) + B_C(X)B_C(X + Y) \right]
= 1 + 2T_C + \mathbb{E}_X \left[ B_C(X)B_C(X + Y) \right].
\]

Over the randomness of \( Y \) and \( X \), we have \( X \) and \( X + Y \) are statistically independent and uniform over \( \mathbb{F}_2^{m \times n} \), so we have
\[
\mathbb{E}_Y \mathbb{E}_X \left[ B_C(X)B_C(X + Y) \right] = \mathbb{E}_Y \mathbb{E}_X \left[ B_C(X) \right] \cdot \mathbb{E}_Y \mathbb{E}_X \left[ B_C(X + Y) \right] = T_C^2.
\]

By Markov’s inequality,
\[
\Pr_Y \left[ S_{C,\{0,Y\}} > 1 + 2T_C + T_C^{1.5} \right] \leq \Pr_Y \left[ \mathbb{E}_X \left[ B_C(X)B_C(X + Y) \right] > T_C^{1.5} \right] < \frac{T_C^2}{T_C^{1.5}} = T_C^{0.5}.
\]

This proves the lemma. \( \square \)

As in the proof of Lemma 3.3, consider the sequence
\[
\delta_0 := \exp_2 \left( -mn \left( (1-p)(1-bp) - \frac{\varepsilon}{1+\varepsilon} \right) \right)
\]
\[
\delta_i := 2\delta_{i-1} + \delta_{i-1}^{1.5}
\]

Like in Lemma 3.3, we can verify by induction that for \( i \leq mn((1-p)(1-bp) - \varepsilon) \), we have \( \delta_i < 2^{i+1}\delta_0 < 2^{-\frac{2^i mn}{1+\varepsilon}} \). The base case is trivial, and if \( \delta_j < 2^{i+1}\delta_0 \) for \( j < i \), we have
\[
\delta_i = 2\delta_{i-1}(1 + \delta_{i-1}^{0.5}) = 2^i \delta_0 \prod_{j=0}^{i-1} (1 + \delta_j^{0.5}) \leq 2^i \delta_0 \cdot \exp \left( \sum_{j=0}^{i-1} \delta_j^{0.5} \right) < 2^{i+1}\delta_0.
\]
In the first inequality, we used the estimate $1 + z \leq e^z$, and in the second we used the inductive hypothesis $\delta_j < 2^{-\frac{z^2}{2}}$ for $j < i$. In particular, if $k = mn((1 - p)(1 - bp) - \varepsilon)$, then $\delta_k < 2^{-\frac{k^2}{2}}$.

Let $Y_1, \ldots, Y_k \in \mathbb{F}_2^{m \times n}$ be randomly chosen matrices, and let $C_i = \text{span}(Y_1, \ldots, Y_i)$ with $C_k = C$. By Lemma 5.5, conditioned on a fixed $C_i$ satisfying $T_{C_i} \leq \delta_i$, then with probability at most $T_{C_i}^{0.5}$, which is at most $\delta_i^{0.5}$, we have $T_{C_i+1} \geq \delta_{i+1}$. Furthermore, $T_{C_0} \leq \delta_0$ by (13). Thus, with probability at least

$$1 - \left(\delta_0^{0.5} + \delta_1^{0.5} + \cdots + \delta_k^{0.5}\right) > 1 - ke^{-e^2mn/2} \geq 1 - 2^{-\Omega(n)}$$

we have $T_{C_i} \leq \delta_i$ for all $i$. In particular, this implies $T_C = T_{C_k} < \delta_k < 2^{-\frac{k^2}{2}}$. Thus, we have $S_C = 1 + T_C \leq 2$ with probability $1 - \exp(-\Omega_e(n))$, completing the proof of Lemma 5.3.

6 Conclusion

In this work, we have given an improved analysis of the list-decodability of random linear binary codes. Our analysis works for all values of $p$, and also obtains improved bounds on the list size as the rate approaches list-decoding capacity. In particular, not only do our bounds improve on previous work for random linear codes, but they show that random linear codes are more list-decodable than completely random codes, in the sense that the list size is strictly smaller. Our techniques are quite simple, and strengthen an argument of [GHSZ02] to hold with high probability. In order to demonstrate the applicability of these techniques, we use them to (a) obtain more information about the distribution of list sizes of random linear codes and (b) to prove a similar result for random linear rank-metric codes, improving a recent result of [GR17].

We end with some open questions raised by our work.

1. With the exception of Theorem 2.12, our results—both our upper bounds and our lower bounds—hold only for binary alphabets. We conjecture that analogous results, and in particular list decoding random linear codes with list size $C/\varepsilon$ for $C < 1$, hold over larger alphabets.

2. We showed that random linear binary codes of rate $1 - H(p) - \varepsilon$ are with high probability $(p, L)$ list-decodable with $L \leq H(p)/\varepsilon$. The lower bounds of [Rud11, GN14] show that we must have $L \geq C/\varepsilon$, but the constant $C$ is much smaller than $H(p)$. Thus, we still do not know what the correct leading constant is for random linear codes.

3. Finally, there are currently no known explicit constructions of capacity-achieving binary list-decodable codes for general $p$. It is our hope that this work—which gives more information about the structure of linear codes which achieve list-decoding capacity—could lead to progress on this front. Given that we don’t know how to efficiently check if a given code is $(p, L)$-list-decodable, even an efficient Las Vegas construction (as opposed to a Monte Carlo construction) would be interesting.

Acknowledgements

We thank anonymous reviewers for helpful comments on a previous version of this manuscript.

References

[AS92] Noga Alon and Joel Spencer. The Probabilistic Method. John Wiley, 1992.
[Bli86] Volodia M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. Problems of Information Transmission, 22(1):719, 1986.

[CGV13] 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), pages 432–442. ACM-SIAM, 2013.

[Del78] Philippe Delsarte. Bilinear forms over a finite field, with applications to coding theory. Journal of Combinatorial Theory, Series A, 25(3):226–241, 1978.

[Din15] Yang Ding. On list-decodability of random rank metric codes and subspace codes. IEEE Trans. Information Theory, 61(1):51–59, 2015.

[DL12] Zeev Dvir and Shachar Lovett. Subspace evasive sets. In Proceedings of the Forty-Fourth annual ACM Symposium on Theory of Computing (STOC), pages 351–358. ACM, 2012.

[Eli57] Peter Elias. List decoding for noisy channels. Wescon Convention Record, Part 2, pages 94–104, 1957.

[Eli91] Peter Elias. Error-correcting codes for list decoding. IEEE Trans. Information Theory, 37(1):5–12, 1991.

[Gab85] Ernst Mukhamedovich Gabidulin. Theory of codes with maximum rank distance. Problemy Peredachi Informatsii, 21(1):3–16, 1985.

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

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

[GI04] Venkatesan Guruswami and Piotr Indyk. Efficiently decodable codes meeting gilbert-varshamov bound for low rates. In Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, pages 756–757. Society for Industrial and Applied Mathematics, 2004.

[GN14] Venkatesan Guruswami and Srivatsan Narayanan. Combinatorial limitations of average-radius list-decoding. IEEE Trans. Information Theory, 60(10):5827–5842, 2014.

[GPT91] Ernst M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a non-commutative ring and their applications in cryptology. In Proceedings of Advances in Cryptology - EUROCRYPT ‘91, Workshop on the Theory and Application of of Cryptographic Techniques, pages 482–489, 1991.

[GR08a] Venkatesan Guruswami and Atri Rudra. Concatenated codes can achieve list-decoding capacity. In Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008, pages 258–267, 2008.

[GR08b] Venkatesan Guruswami and Atri Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. IEEE Trans. Information Theory, 54(1):135–150, 2008.
[GR17] Venkatesan Guruswami and Nicolas Resch. On the list-decodability of random linear rank-metric codes. *arXiv preprint arXiv:1710.11516*, 2017.

[Gur09] Venkatesan Guruswami. List decoding of binary codes—a brief survey of some recent results. In *Proceedings of Coding and Cryptology, Second International Workshop (IWCC)*, pages 97–106, 2009.

[GV05] Venkatesan Guruswami and Salil Vadhan. A lower bound on list size for list decoding. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 318–329, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[GWX16] Venkatesan Guruswami, Carol Wang, and Chaoping Xing. Explicit list-decodable rank-metric and subspace codes via subspace designs. *IEEE Trans. Information Theory*, 62(5):2707–2718, 2016.

[GX12] 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 (STOC)*, pages 339–350. ACM, 2012.

[GX13] 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 (STOC)*, pages 843–852. ACM, 2013.

[GY08] Maximilien Gadouleau and Zhiyuan Yan. On the decoder error probability of bounded rank-distance decoders for maximum rank-distance codes. *IEEE Trans. Information Theory*, 54(7):3202–3206, 2008.

[HRZW17] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes & applications. In *58th Annual IEEE Symposium on Foundations of Computer Science*, 2017.

[HW15] Brett Hemenway and Mary Wootters. Linear-time list recovery of high-rate expander codes. In *International Colloquium on Automata, Languages, and Programming*, pages 701–712. Springer, 2015.

[KK08] Ralf Koetter and Frank R Kschischang. Coding for errors and erasures in random network coding. *IEEE Trans. Information Theory*, 54(8):3579–3591, 2008.

[LGB03] P. Lusina, Ernst M. Gabidulin, and Martin Bossert. Maximum rank distance codes as space-time codes. *IEEE Trans. Information Theory*, 49(10):2757–2760, 2003.

[LK05] Hsiao-feng Lu and P. Vijay Kumar. A unified construction of space-time codes with optimal rate-diversity tradeoff. *IEEE Trans. Information Theory*, 51(5):1709–1730, 2005.

[Loi10] Pierre Loidreau. Designing a rank metric based mceliece cryptosystem. In *Proceedings of the Post-Quantum Cryptography, Third International Workshop on Post-Quantum Cryptography, (PQCrypto)*, pages 142–152, 2010.
[Loi17] Pierre Loidreau. A new rank metric codes based encryption scheme. In Proceedings of the Post-Quantum Cryptography, 8th International Workshop on Post-Quantum Cryptography, (PQCrypto), pages 3–17, 2017.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. The theory of error-correcting codes. Elsevier, 1977.

[Rot91] Ron M. Roth. Maximum-rank array codes and their application to crisscross error correction. IEEE Trans. Information Theory, 37(2):328–336, 1991.

[RU10] Atri Rudra and Steve Uurtamo. Two theorems on list decoding. In Maria Serna, Ronen Shaltiel, Klaus Jansen, and José Rolim, editors, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pages 696–709, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[Rud11] Atri Rudra. Limits to list decoding of random codes. IEEE Trans. Information Theory, 57(3):1398–1408, 2011.

[RW14] 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 (STOC), pages 764–773. ACM, 2014.

[RW15] Atri Rudra and Mary Wootters. It’ll probably work out: Improved list-decoding through random operations. In Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS), pages 287–296. ACM, 2015.

[RW18] Atri Rudra and Mary Wootters. Average-radius list-recovery of random linear codes. In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). ACM-SIAM, 2018.

[SKK08] Danilo Silva, Frank R Kschischang, and Ralf Koetter. A rank-metric approach to error control in random network coding. IEEE Trans. Information Theory, 54(9):3951–3967, 2008.

[SRKV13] Natalia Silberstein, Ankit Singh Rawat, O Ozan Koyluoglu, and Sriram Vishwanath. Optimal locally repairable codes via rank-metric codes. In Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), pages 1819–1823. IEEE, 2013.

[SRV12] Natalia Silberstein, Ankit Singh Rawat, and Sriram Vishwanath. Error resilience in distributed storage via rank-metric codes. In Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 1150–1157. IEEE, 2012.

[Sud00] Madhu Sudan. List decoding: algorithms and applications. SIGACT News, 31(1):16–27, 2000.

[Vad12] Salil P. Vadhan. Pseudorandomness. Foundations and Trends in Theoretical Computer Science, 7(1-3):1–336, 2012.

[Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In Proceedings of the Forty-Fifth Symposium on Theory of Computing Conference (STOC), pages 853–860, 2013.
A Lower bound on the list size of random codes

In this appendix, we prove Theorem 2.5, which states that, for any $\gamma > 0$ and $\varepsilon$ small enough, the smallest list size $L$ such that a completely random code is with high probability $(p, L)$-list decodable, satisfies $L \in [(1 - \gamma)/\varepsilon, 1/\varepsilon]$. In other words, the list size of $1/\varepsilon$ given by the classical list decoding capacity theorem is tight in the leading constant factor. We prove this by tightening the second moment argument of Guruswami and Narayanan [GN14]. We follow the exact same proof outline, differing only in how we bound one expression. This improvement appears as Lemma A.5. In both Appendices A and B, we use the following useful fact.

**Fact A.1.** Let $Z$ be a random variable that takes only nonnegative values. Then

$$\Pr[Z = 0] \leq \frac{\Var Z}{\mathbb{E}[Z]^2}.$$  

For this appendix, we assume that $pn$ is an integer. We need a few bounds on Hamming balls and their intersections. We first use the following classical estimate on the volume of the Hamming ball (see, for example, [MS77][pp. 308-310]).

**Lemma A.2.** Suppose that $1 \leq pn \leq n/2$. Then

$$\frac{2^{H(p)n}}{\sqrt{8pn(1 - p)}} \leq \binom{n}{pn} \leq \Vol(n, pn) \leq 2^{H(p)n}.$$  

We also need a bound on the volume of the intersections of two Hamming balls whose centers are distance $d$ apart, which we denote by

$$\Vol(n, pn; d) := |B(0, pn) \cap B(1^d0^{n-d}, pn)|.$$  

Lemma A.3 below bounds $\Vol(n, pn; d)$ above.

**Lemma A.3.** Suppose $0 < p < 1/2$. Let

$$\alpha_p := \frac{1}{2} \log_2 \frac{1}{4p(1 - p)}.$$  

There exists a constant $C_p > 0$ depending only on $p$ such that, for sufficiently large $n$, for all $1 \leq d \leq n$, we have

$$\frac{\Vol(n, pn; d)}{\Vol(n, pn)} \leq 2^{-\alpha_p d} \cdot C_p \sqrt{n}. \quad (14)$$
Proof. Let $w = 1^d0^{n-d}$ be a string of weight $d$, so that we wish to upper bound the number of points $x$ in $D := \mathcal{B}(0, pn) \cap \mathcal{B}(w, pm)$.

First, notice that if $pn < d/2$, then $\mathcal{B}(0, pn) \cap \mathcal{B}(w, pm) = \emptyset$, so (14) holds. Thus, assume that $pn \geq d/2$. Suppose that $x$ restricted to the first $d$ coordinates has weight $i$, and that $x$ restricted to the last $n - d$ coordinates has weight $j$. Such an $x$ is in $D$ if and only if $(d - i) + j \leq pn$ and $i + j \leq pn$. For a fixed $i, j$, the number of such $x$ is exactly $\binom{d}{i}\binom{n-d}{j}$. Thus,

$$\text{Vol}(n, pn; d) = \sum_{i=0}^{d} \sum_{j=0}^{\text{max}(i, d-i)} \binom{d}{i} \binom{n-d}{j}$$

$$\leq \sum_{i=0}^{d} \binom{d}{i} \sum_{j=0}^{pn-d/2} \binom{n-d}{j}$$

$$= \sum_{i=0}^{d} \binom{d}{i} \text{Vol}(n-d, np - (d/2))$$

$$= 2^d \cdot \text{Vol}(n-d, np - (d/2))$$

$$\leq \exp\left( d + H_2\left( \frac{pn-d/2}{n-d} \right) \cdot (n-d) \right),$$

using Lemma A.2 in the final line. Define

$$f(d) := d + H_2\left( \frac{pn-d/2}{n-d} \right) \cdot (n-d).$$

When $d = 0$, (14) holds for sufficiently large $n$. Thus, we show that the derivative $f'(d)$ is sufficiently negative that (14) continues to hold as $d$ increases.

**Claim A.4.** We have $f'(d) \leq -\alpha_p$.

**Proof.** We compute

$$f'(d) = 1 + \frac{1}{2} \log_2 \frac{(pn-d/2)(n-pn-d/2)}{(n-d)^2}.$$

It can be verified that

$$\frac{(pn-d/2)(n-pn-d/2)}{(n-d)^2} \leq p(1-p).$$

Indeed, the above is equivalent to $p(1-p) \leq 1/4$, which is always true. Since the left-hand side is always positive (using the assumption $p < 1/2$ and $d/2 \leq pm$), we may take logs of both sides and conclude that

$$f'(d) \leq 1 + \frac{1}{2} \log_2 p(1-p) = -\alpha_p.$$

Returning to the proof of Lemma A.3, as $f(0) = H(p)n$, we have $f(d) \leq H(p)n - \alpha_p d$ for all $d \geq 0$, and in particular all integers $d \geq 0$. Taking $C_p = \sqrt{8p(1-p)}$, and applying Lemma A.2, we conclude

$$\frac{\text{Vol}(n, pn; d)}{\text{Vol}(n, pn)} \leq \frac{\exp_2(H(p)n - \alpha_p d)}{\exp_2(H(p)n) / \sqrt{8np(1-p)}} \leq 2^{-\alpha_p d} \cdot C_p \sqrt{n}.$$
In Lemma A.5, we record computation used in the proof of Theorem 2.5; this computation improves over the bounding in [GN14] and helps us obtain a better overall list size lower bound.

**Lemma A.5.** Let $0 < p < 1/2$, $1 \leq \ell \leq L$ be integers and $\mu := 2^{-n}\text{Vol}(n, pn)$. Let

$$a, b, x_1, \ldots, x_\ell, x_{\ell+1}, \ldots, x_L, y_{\ell+1}, \ldots, y_L$$

be chosen independently and uniformly at random from $\mathbb{F}_2^n$ (so there are $2 + 2L - \ell$ points total). Let $\mathcal{E}$ be the event

$$\mathcal{E} = \left( \bigwedge_{i=1}^\ell [\Delta(a, x_i) \leq pn] \land [\Delta(b, x_i) \leq pn] \right) \land \left( \bigwedge_{i=\ell+1}^L [\Delta(a, x_i) \leq pn] \land [\Delta(b, y_i) \leq pn] \right).$$

Then

$$\Pr[\mathcal{E}] \leq \min \left( \mu^{2L-\ell+1}, 2^{-n} \cdot \mu^{2L-\ell} \cdot C_p^L \cdot n^{L/2} \cdot (1 + 2^{-\alpha_p \ell})^n \right).$$

**Proof.** First we show the probability of $\mathcal{E}$ is at most $\mu^{2L-\ell+1}$. This is the argument that was presented in [GN14]. For the event $\mathcal{E}$ to occur, we need the following necessary (but not necessarily sufficient) conditions: (1) $\Delta(a, x_1) \leq pn$ and $\Delta(b, x_1) \leq pn$, (2) $\Delta(a, x_i) \leq pn$ for $i = 2, \ldots, L$, and (3) $\Delta(b, y_i) \leq pn$ for $i = \ell + 1, \ldots, L$. Notice that each of (1), (2), (3) are independent. Using the randomness of $a$ and $b$, (1) happens with probability $\mu^2$. Conditioned on the positions of $a$ and $b$, (2) and (3) occur with probability $\mu^{L-1}$ and $\mu^{L-\ell}$, respectively. Thus, we have $\Pr[\mathcal{E}] \leq \mu^{2L-\ell+1}$.

For the other bound, we first condition on the locations of $a$ and $b$. Note that, for $0 \leq d \leq n$, $a$ and $b$ have Hamming distance exactly $d$ with probability $2^{-n} \binom{n}{d}$. For $i = 1, \ldots, \ell$, conditioned on the positions of $a$ and $b$ being Hamming distance $d$ apart, the probability that $\Delta(a, x_i) \leq pn$ and $\Delta(b, x_i) \leq pn$ is exactly $2^{-n}\text{Vol}(n, pn; d)$. For $i = \ell + 1, \ldots, L$, the probability that $\Delta(a, x_i) \leq pn$ and the probability that $\Delta(b, y_i) \leq pn$ are each exactly $\mu$. We thus can write, using Lemma A.3,

$$\Pr[\mathcal{E}] = \sum_{d=0}^n \Pr[\Delta(a, b) = d] \cdot \left( \frac{\text{Vol}(n, pn; d)}{2^n} \right)^\ell \cdot \mu^{2L-2\ell}$$

$$= \sum_{d=0}^n 2^{-n} \binom{n}{d} \cdot \left( \frac{\text{Vol}(n, pn; d)}{\text{Vol}(n, pn)} \cdot \frac{\text{Vol}(n, pn)}{2^n} \right)^\ell \cdot \mu^{2L-2\ell}$$

$$= \sum_{d=0}^n 2^{-n} \binom{n}{d} \cdot \left( \frac{\text{Vol}(n, pn; d)}{\text{Vol}(n, pn)} \right)^\ell \cdot \mu^{2L-\ell}$$

$$= 2^{-n} \mu^{2L-\ell} \sum_{d=0}^n \binom{n}{d} \cdot \left( \frac{\text{Vol}(n, pn; d)}{\text{Vol}(n, pn)} \right)^\ell$$

$$\leq 2^{-n} \mu^{2L-\ell} \cdot C_p^L \cdot n^{L/2} \cdot (1 + 2^{-\alpha_p \ell})^n$$

as desired. \(\Box\)

We now prove our theorem.

**Theorem A.6** (Theorem 2.5, restated). For any $p \in (0, 1/2)$ and $\varepsilon > 0$, there exists a $\gamma_{p, \varepsilon} = \exp(-\Omega_p(\varepsilon))$ and $n_{p, \varepsilon} \in \mathbb{N}$ such that for all $n \geq n_{p, \varepsilon}$, a random linear code $C \subseteq \mathbb{F}_2^n$ of rate $R = 1 - H(p) - \varepsilon$ is with probability $1 - \exp(-\Omega_{p, \varepsilon}(n))$ not $(p, \frac{1-\gamma_{p, \varepsilon}}{\varepsilon})$-list decodable.
Proof. We follow the outline [GN14]. Let \( \alpha_p \) be as in Lemma A.3. With hindsight, let

\[
\ell_{p, \varepsilon} := \frac{1 - H(p)}{2\varepsilon}, \quad \gamma_{p, \varepsilon} := \frac{1}{2\ln 2} \cdot 2^{-\alpha_p \ell_{p, \varepsilon}} = \exp \left( -\Omega_p \left( \frac{1}{\varepsilon} \right) \right),
\]

Let \( C \subseteq \mathbb{F}_2^n \) be a uniformly random code of rate \( R = 1 - H(p) - \varepsilon \). We think of each codeword \( c \in C \) as the encoding of a distinct message \( x \in \mathbb{F}_2^{Rn} \). Choose \( L = \lfloor (1 - \gamma_{p, \varepsilon})/\varepsilon \rfloor \). We show that, if \( \varepsilon \) is sufficiently small, \( C \) is with high probability not \((p, L)\)-list decodable.

Let \( \mu = 2^{-n} \text{Vol}(n, pn) \). For any list decoding center \( a \in \mathbb{F}_2^n \) and any ordered list of \( L \) messages \( X = (x_1, \ldots, x_L) \in (\mathbb{F}_2^{Rn})^L \), let \( \mathbb{I}(a, X) \) be the indicator random variable for the event that the encodings of \( x_1, \ldots, x_L \) all fall in \( B(a, pn) \). Let \( W = \sum_{a, X} \mathbb{I}(a, X) \). Note that \( C \) is \((p, L)\)-list decodable if and only if \( W = 0 \).

Just as in [GN14], we can compute \( \mathbb{E}[W] \geq \frac{1}{2} \mu L 2^n 2^{RnL} \), and we have

\[
\text{Var} W = \sum_{X, Y \in a, b} \left( \mathbb{E}_C[\mathbb{I}(a, X) \mathbb{I}(b, Y)] - \mathbb{E}_C[\mathbb{I}(a, X)] \cdot \mathbb{E}_C[\mathbb{I}(b, Y)] \right)
\]

\[
= \sum_{X \cap Y \neq \emptyset} a, b \sum_{X \cap Y \neq \emptyset} a, b \left( \mathbb{E}_C[\mathbb{I}(a, X) \mathbb{I}(b, Y)] - \mathbb{E}_C[\mathbb{I}(a, X)] \cdot \mathbb{E}_C[\mathbb{I}(b, Y)] \right)
\]

\[
\leq \sum_{X \cap Y \neq \emptyset} a, b \mathbb{E}_C[\mathbb{I}(a, X) \mathbb{I}(b, Y)]
\]

\[
= \sum_{\ell=1}^L \sum_{|X \cap Y| = \ell} a, b \mathbb{E}_C[\mathbb{I}(a, X) \mathbb{I}(b, Y)]
\]

\[
= \sum_{\ell=1}^L 2^{2n} \mathbb{Pr}_{a, b, C}(\mathbb{I}(a, X) \text{ and } \mathbb{I}(b, Y)).
\]

Suppose \( X \) and \( Y \) are \( L \)-tuples such that \(|X \cap Y| = \ell\), where the intersection treats \( X \) and \( Y \) as sets. Suppose that the elements of \( X \) are \( x_1, \ldots, x_L \) and the elements of \( Y \) are \( x_1, \ldots, x_{\ell}, y_{\ell+1}, \ldots, y_L \). In this notation, the event \( \mathbb{I}(a, X) \text{ and } \mathbb{I}(b, Y) \) is exactly event \( \mathcal{E} \) in Lemma A.5. Hence,

\[
\text{Var} W \leq \sum_{\ell=1}^L 2^{2n} \mathbb{Pr}_{a, b, C}(\mathbb{I}(a, X) \text{ and } \mathbb{I}(b, Y))
\]

\[
\leq \sum_{\ell=1}^L 2^{2n} \cdot \min \left( \mu^{2L-\ell+1}, 2^{-n} \mu^{2L-\ell} \cdot C_p^L \cdot n^{L/2} \cdot (1 + 2^{-\alpha_p \ell})^n \right).
\]

As \( \alpha_p > 0 \), by choice of \( \gamma_{p, \varepsilon} \), for all \( \ell \geq \ell_{p, \varepsilon} \), we have

\[
\frac{1 + 2^{-\alpha_p \ell}}{2} \leq 2^{-(1-2\gamma_{p, \varepsilon})}.
\]

We also bound the number of \( X, Y \) such that \(|X \cap Y| = \ell\) by \( L^{2L} 2^{Rn(2L-\ell)} \). Indeed, there are at
most $2^{R_n(2L-\ell)}$ ways to choose $X, Y$ as sets and at most $L^{2L}$ ways to order the sets. Thus,

$$\frac{\text{Var} W}{\mathbb{E}[W]^2} \leq \frac{4}{\mu^{2L}2^{2n}2^{R_nL}} \sum_{\ell=1}^{L} \sum_{|X \cap Y| = \ell} 2^{2n} \cdot \min\left(\mu^{2L-\ell+1}, 2^{-n} \cdot \mu^{2L-\ell} \cdot C_p^L \cdot n^{L/2} \cdot (1 + 2^{-\alpha_p \ell})^n\right)$$

$$\leq \frac{4}{\mu^{2L}2^{2n}2^{R_nL}} \sum_{\ell=1}^{L} \sum_{|X \cap Y| = \ell} 2^{2n} \cdot \mu^{2L-\ell+1} + \frac{4}{\mu^{2L}2^{2n}2^{R_nL}} \sum_{\ell=\ell_{p,\varepsilon}}^{L} \sum_{|X \cap Y| = \ell} 2^{2n} \cdot \mu^{2L-\ell} \cdot C_p^L \cdot n^{L/2} \cdot \left(\frac{1 + 2^{-\alpha_p \ell}}{2}\right)^n$$

$$\leq \sum_{\ell=1}^{4L(2R_n \mu)^{-\ell}} \mu + \sum_{\ell=\ell_{p,\varepsilon}}^{L} 4L(2R_n \mu)^{-\ell} \cdot C_p^L \cdot n^{L/2} \cdot 2^{-(1-2\gamma_{p,\varepsilon})n}$$

$$\leq 4L^{2L+1} \cdot 2^{\ell_{p,\varepsilon}n} \cdot 2^{-(1-H(p))n} + 4L^{2L+1} \cdot C_p^L \cdot n^{L/2} \cdot 2^{\varepsilon n L -(1-2\gamma_{p,\varepsilon})n}$$

By choice of $\ell_{p,\varepsilon} = \frac{1-H(p)}{2\varepsilon}$, the first term in the last line is $\exp(-\Omega_{p,\varepsilon}(n))$, and by choice of $L = \lceil \frac{1}{\varepsilon \gamma_{p,\varepsilon}} \rceil$, the second term in the last line is also $\exp(-\Omega_{p,\varepsilon}(n))$. Invoking Fact A.1,

$$\Pr[W = 0] \leq \frac{\text{Var} W}{\mathbb{E}[W]^2} \leq \exp(-\Omega_{p,\varepsilon}(n)).$$

Since $C$ is $(p, L)$-list-decodable if and only if $W = 0$, we conclude, for $n$ sufficiently large, that $C$ is with probability $1 - \exp(-\Omega_{p,\varepsilon}(n))$ not $(p, L)$-list decodable. \hfill \Box

## B Lower bound on the list size of random rank metric codes

In this section, we prove Theorem 2.12.

**Theorem B.1** (Theorem 2.12 restated). Let $\varepsilon > 0$ and suppose $m, n$ are sufficiently large so that $b = n/m$. Let $C \subseteq \mathbb{F}_q^{m \times n}$ be a uniformly random rank metric code of rate $R = (1-p)(1-bp) - \varepsilon$. Then $C$ is $(p, (1-p)(1-bp)/\varepsilon - 1)$-list decodable with probability at most $\exp(-\Omega_{p,\varepsilon}(n))$.

The proof is a second moment argument and is nearly identical to the one in [GN14]; we work out the details here for completeness.

**Proof.** Let $\varepsilon > 0$ and fix $L = (1-p)(1-bp)/\varepsilon - 1$. Let $C \subseteq \mathbb{F}_q^{m \times n}$ be a uniformly random rank metric code. We think of each $c \in C$ as the encoding of a message $x \in \mathbb{F}_q^{R_m n}$. For any center $A \in \mathbb{F}_q^{m \times n}$ and any ordered list of $L$ distinct messages $X := (x_1, x_2, \ldots, x_L) \in (\mathbb{F}_q^{R_m n})^L$, let $\mathbb{I}(A, X)$ be the indicator random variable for the event that, for all $i$, the encodings of $x_1, \ldots, x_L$ are all contained in $B_R(A, pn)$. Define

$$W := \sum_{A, X} \mathbb{I}(A, X).$$

26
Note that $\mathcal{C}$ is $(p, L)$-list decodable if and only if $W = 0$.

Let $\mu = q^{-m n}[\mathcal{B}_{q,R}(0, pn)]$. By Lemma 5.2, we have

$$\mu q^{Rmn} < 4q^{-\varepsilon mn}.\tag{15}$$

The event that the encoding of $x_i$ falls inside $\mathcal{B}_{q,R}(A, pn)$ is exactly $\mu$, and these events are statistically independent, so we have $E[\mathbb{I}(a, X)] = \mu^L$. Assuming $k \geq L + 1$, the number of pairs $(A, X)$ is at least $\frac{1}{2}q^{RmnL} \cdot q^{mn}$, as the number of ordered $L$-lists $X$ of $L$ distinct messages is

$$q^{Rmn} \cdot (q^{Rmn} - 1) \cdots (q^{mn} - L + 1) \geq q^{RmnL} \left(1 - \frac{L}{2q^{Rmn}}\right) \geq q^{RmnL} \left(1 - \frac{q^L}{q^{Rmn}}\right) \geq \frac{1}{2}q^{RmnL}.$$ 

Thus, by linearity of expectation, we have

$$E[W] \geq \frac{1}{2} \mu^L q^{mn} q^{RmnL}.$$

Following the method in [GN14], we now bound the variance of $W$ above. For two lists of messages $X$ and $Y$, we let $|X \cap Y|$ denote the size of their intersection when we view them as sets. If $X$ and $Y$ are disjoint, then the events $\mathbb{I}(A, X), \mathbb{I}(B, Y)$ are independent for any pair of centers $A, B \in \mathbb{F}_q^{m \times n}$. Thus,

$$\text{Var} W = \sum_{X, Y} \sum_{A, B} \left( E_{\mathcal{C}}[\mathbb{I}(A, X)\mathbb{I}(B, Y)] - E_{\mathcal{C}}[\mathbb{I}(A, X)] \cdot E_{\mathcal{C}}[\mathbb{I}(B, Y)] \right)$$

$$= \sum_{X \cap Y \neq \emptyset} \sum_{A, B} \left( E_{\mathcal{C}}[\mathbb{I}(A, X)\mathbb{I}(B, Y)] - E_{\mathcal{C}}[\mathbb{I}(A, X)] \cdot E_{\mathcal{C}}[\mathbb{I}(B, Y)] \right)$$

$$\leq \sum_{X \cap Y \neq \emptyset} \sum_{A, B} E_{\mathcal{C}}[\mathbb{I}(A, X)\mathbb{I}(B, Y)]$$

$$= \sum_{\ell=1}^{L} \sum_{|X \cap Y| = \ell} \sum_{A, B} q^{2mn} \Pr_{A, B, C}[\mathbb{I}(A, X) \text{ and } \mathbb{I}(B, Y)]$$

where, in the last line, $A, B$ are picked uniformly at random from $\mathbb{F}_q^{m \times n}$.

Fix $0 < l \leq L$ and a pair $(X, Y)$ of $L$-tuples of matrices such that $|X \cap Y| = \ell$, and fix an arbitrary $z \in X \cap Y$, which is guaranteed to exist as $X, Y$ intersect. For the event $\mathbb{I}(A, X) = \mathbb{I}(B, Y) = 1$ to happen, all of the following must occur:

1. $A$ and $B$ are both within rank-metric distance $pn$ of the encoding of $z$.
2. For each $x \in X \setminus \{z\}$, the encoding of $x$ falls inside $B_{q,R}(A, pn)$.
3. For each $y \in Y \setminus X$, the encoding of $y$ falls inside $B_{q,R}(B, pn)$.

The first event occurs with probability $\mu^2$, and conditioned on the choices of $A$ and $B$, the second and third events occur with probabilities $\mu^{L-1}$ and $\mu^{L-\ell}$ and are independent given $A$ and $B$. Thus, the probability all the events occur is $\mu^{2L-\ell+1}$, so the probability that $\mathbb{I}(A, X) = \mathbb{I}(B, Y) = 1$ is at
most $\mu^{2L-\ell+1}$. Finally, noting that the number of pairs $(X, Y)$ with $|X \cap Y| = L$ is bounded by $L^{2L} \cdot q^{Rmn(2L-\ell)}$, we conclude

$$\text{Var} W \leq \sum_{\ell=1}^{L} L^{2L} \cdot q^{Rmn(2L-\ell)} \cdot q^{2mn} \mu^{2L-\ell+1}.$$  

Using Fact A.1 along with (15), we have

$$\Pr[W = 0] \leq \frac{\text{Var} W}{\text{E}[W]^2} \leq \sum_{\ell=1}^{L} 4L^{2L} \cdot q^{Rmn} \mu^{-\ell} \mu \leq 4L^{2L+1} \cdot q^{Rmn} \cdot 4q^{-mn(1-p)(1-\delta p)}.$$  

This quantity is $\exp_q(-\Omega_{p,\epsilon}(n))$ for our choice of $L$. Recalling that $C$ is $(p, L)$-list-decodable if and only of $W = 0$, we conclude that with high probability, $C$ is not $(p, L)$-list-decodable, as desired. $\square$

\section{Existence of rate $1 - H(p) - \epsilon$ codes achieving list size $\frac{H(p)}{\epsilon}$}

In this section, we use the Symmetric Lovasz Local Lemma (see, e.g., [AS92]).

\textbf{Lemma C.1} (Lovasz Local Lemma). Let $d$ be a positive integer and $p \in (0, 1)$. Let $A_1, \ldots, A_N$ be events and $G = (V, E)$ be a graph with vertices $V = [N]$ such that each vertex has degree at most $d$, and, for all $i \in [N]$, the event $A_i$ is independent of the events $\{A_j : j \in [N], j \neq i, ij \notin E\}$. Suppose also that for all $i$, $\Pr[A_i] \leq p$. If $ep(d + 1) < 1$, then

$$\Pr \left[ \bigwedge_{i=1}^{N} \overline{A_i} \right] \geq \left(1 - \frac{1}{d+1}\right)^N > 0,$$

i.e. with positive probability, none of the events $A_i$ hold.

We prove the following.

\textbf{Theorem C.2.} Let $p \in [0, 1/2)$ and $\epsilon > 0$. There exist binary codes of rate $R = 1 - H(p) - \epsilon$ that are $(p, \frac{H(p)}{\epsilon} + 1)$-list decodable.

\textbf{Remark C.3.} Theorem C.2 does not contradict Theorem 2.5, our high probability lower bound, as Theorem C.2 gives an existential result, rather than a high probability result.

\textit{Proof.} Let $M = 2^{Rn+1}$. Let $L = \lceil H(p)/\epsilon \rceil + 1 > \frac{H(p)}{\epsilon}$. Take a random code $C$ where $c_1, \ldots, c_M$ are chosen from $\mathbb{F}_2^n$ independently and uniformly at random. For $y \in \mathbb{F}_2^n$ and $I = \{i_1, \ldots, i_L\}$, let $A(y, I)$ denote the event that $c_i \in \mathcal{B}(y, pn)$ for $j = 0, \ldots, L$. There are $2^nM^{L+1}$ such events. We have that $C$ is $(p, L)$-list decodable if and only if none of the events $A(y, I)$ occur.

Define a graph $G$ on the events $A(y, I)$ where $A(y, I), A(z, J)$ are connected by an edge if $I \cap J \neq \emptyset$. Note that, as the $c_i$'s are all independent and each $A(y, I)$ depends only on $c_i$ such that $i \in I$, we have $A(y, I)$ is independent of all $A(z, J)$ satisfying $I \cap J = \emptyset$. In other words, $A(y, I)$ is independent from all nonneighbor events in $G$. For each $I \subseteq [M]$ with $|I| = L + 1$, there are at most $(L + 1) \cdot M^L$ choices of $J$ such that $I \cap J \neq \emptyset$. Thus, each $A(y, I)$ has degree at most $d := 2^n \cdot (L + 1) \cdot M^L$. On the other hand, we can compute

$$p := \Pr[A(y, I)] = \left(\frac{\text{Vol}(n, pn)}{2^n}\right)^{L+1} \leq 2^{-n(L+1)(1-H(p))}.$$

28
We thus have
\[
e \cdot p(d + 1) \leq 3 \cdot 2^{-n(L+1)(1-H(p))} \cdot (L + 1) \cdot 2^n \cdot 2^{RnL+L} \\
< 6L \cdot 2^{-n(L+1)(1-H(p))} \cdot 2^n \cdot 2^{(1-H(p) - \varepsilon)nL+L} \\
= 6L \cdot 2^{nH(p) - \varepsilon nL + L} < 1,
\]
for our choice of \( L \) and sufficiently large \( n \). Thus, by Lemma C.1, we have
\[
\Pr \left[ \bigwedge_{y,I} A(y, I) \right] \geq \left( 1 - \frac{1}{d+1} \right)^{2^n M^{L+1}} \\
> \exp \left( - \frac{2^n M^{L+1}}{d} \right) \\
= \exp \left( - \frac{M}{L + 1} \right) > 0.
\]
For the second inequality, we used the estimate \( 1 - \frac{1}{d+1} \geq e^{-\frac{1}{d}} \). At this point, we are essentially done, but it is possible in our random choice that, for example, all \( c_i \) are the same point. We check that this probability is sufficiently small compared to the probability of list decodability.

There are \( 2^{nM} \) ways to choose \( c_1, \ldots, c_M \). The probability that \( |C| \leq M/2 \) is at most,
\[
\Pr \left[ |C| \leq \frac{M}{2} \right] \leq \frac{\left( 2^{nM}/2 \right) \cdot (M/2)^M}{2^{nM/2}} \leq \frac{2^{nM/2} (M^2/2)^{2nM/2}}{(M/2)^2} \cdot \frac{(M/2)^M}{2^{nM}} \\
= \frac{e^{M/2} (M/2)^{M/2}}{2^{nM/2}} \cdot \frac{(M/2)^M}{2^n} \leq \left( \frac{eM/2}{2^n} \right)^M \\
< e^{-\frac{M}{d+1}} < \Pr \left[ \bigwedge_{y,I} A(y, I) \right]
\]
where the second inequality was bounded by Stirling’s approximation.

Thus, with positive probability, we have both (1) none of the events \( A(y, I) \) hold, i.e. \( C \) is \( (p, L) \)-list decodable, and (2) \( C \) has rate at least \( R = 1 - H(p) - \varepsilon \), so in particular there exists some \( C \) with the desired list decodability and rate. \( \Box \)