Limits to List Decoding Random Codes

Atri Rudra*

Department of Computer Science and Engineering,
University at Buffalo, The State University of New York,
Buffalo, NY, 14620.
atri@cse.buffalo.edu

Abstract

It has been known since [Zyablov and Pinsker 1982] that a random q-ary code of rate $1 - H_q(\rho) - \varepsilon$ (where $0 < \rho < 1 - 1/q$, $\varepsilon > 0$ and $H_q(\cdot)$ is the q-ary entropy function) with high probability is a $(\rho, 1/\varepsilon)$-list decodable code. (That is, every Hamming ball of radius at most ρn has at most $1/\varepsilon$ codewords in it.) In this paper we prove the "converse" result. In particular, we prove that for *every* $0 < \rho < 1 - 1/q$, a random code of rate $1 - H_q(\rho) - \varepsilon$, with high probability, is not a (ρ, L)-list decodable code for any $L \leqslant \frac{c}{\varepsilon}$, where c is a constant that depends only on ρ and q. We also prove a similar lower bound for random linear codes.

Previously, such a tight lower bound on the list size was only known for the case when $\rho \geqslant 1 - 1/q - O(\sqrt{\varepsilon})$ [Guruswami and Vadhan, 2005]. For binary codes a lower bound is known for all $0 < \rho < 1/2$, though the lower bound is asymptotically weaker than our bound [Blinovsky, 1986]. These results however are not subsumed by ours as they hold for arbitrary codes of rate $1 - H_q(\rho) - \varepsilon$.

*Research supported in part by startup funds from University at Buffalo.

1 Introduction

One of the central questions in the theory of error-correcting codes (henceforth just codes) is to determine the optimal (combinatorial) tradeoff between the amount of redundancy used in a code and the amount of errors it can tolerate during transmission over a noisy channel. The first result in this vein is the seminal work of Shannon that precisely determined this tradeoff for a class of stochastic channels [10]. In this paper, we will look at the harsher adversarial noise model pioneered by Hamming [7], where we model the channel as an adversary. That is, other than a bound on the total number of errors, the channel can arbitrarily corrupt the transmitted message.

Under the adversarial noise model, it is well known that for the same amount of redundancy, lesser number of errors can be corrected than stochastic noise models (by almost a factor of two). This result follows from a simple argument that exploits the requirement that one always has to recover the transmitted message from the received transmission. However, if one relaxes the strict constraint of uniquely outputting the transmitted message to allow a list of messages to be output (with the guarantee that the transmitted message is in the list), then it can be shown that the optimal tradeoff between the amount of redundancy and the amount of correctable adversarial errors coincides with the tradeoff for certain stochastic noise models. This relaxed notion of outputting a list of possible transmitted messages, called *list decoding*, was put forth by Elias [2] and Wozencraft [13] in the late 1950's.

We point out that in the notion of list decoding, the size of the output list of messages is a crucial quantity. In particular, one can always "successfully" list decode by outputting the list of all possible messages, in which case the problem becomes trivial. Thus, the concept of list decoding is only interesting when the output list is constrained to be "small." This paper deals with quantifying the "smallness" of the *list size*.

Before we state our results more precisely, we quickly set up some notation (see Section 2 for more details). A code introduces redundancy by mapping a message to a (longer) *codeword*. The redundancy of a code is measured by its *rate*, which is the ratio of the the number of information symbols in the message to that in the codeword. Thus, for a code with *encoding* function $E : \Sigma^k \rightarrow \Sigma^n$, the rate equals k/n. The *block length* of the code equals n, and Σ is its alphabet. A code with an alphabet size of $q = |\Sigma|$ is called a q-ary code. The goal in *decoding* is to find, given a noisy received word, the actual transmitted codeword. We will generally talk about the fraction of errors that can be successfully decoded from and denote it by ρ. A code is called (ρ, L)-list decodable if for every received word there are at most L codewords that differ from the received word in at most ρ fraction of positions.

Zyablov and Pinsker established the following optimal tradeoff between the rate and ρ for list decoding [14]. First, they showed that there exists q-ary $(\rho, 1/\varepsilon)$-list decodable codes of rate $1 - H_q(\rho) - \varepsilon$ for any $\varepsilon > 0$ (where $H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$ is the q-ary entropy function). Second, they showed that any q-ary (ρ, L)-list decodable of rate $1 - H_q(\rho) + \varepsilon$ needs L to be exponentially large in the block length of the code. Thus, the quantity $1 - H_q(\rho)$ exactly captures the optimal tradeoff between rate and ρ for list decoding (with small lists). This quantity also matches the "capacity" of the so called q-ary Symmetric channel.[1]

[1] In particular, in the q-ary symmetric channel, the alphabet size is q and the following random noise process acts independently on every transmitted symbol: every symbol remains unchanged with probability $1 - \rho$ and is changed to each of the $q - 1$ possibilities with probability of $\rho/(q-1)$. It can be shown that the capacity of this channel is $1 - H_q(\rho)$, that is, there exists code of rate $1 - H_q(\rho) - \varepsilon$ which allows for reliable communication over the q-ary symmetric channel with all but an exponentially small probability. On the other hand, for every code of

However, the result of Zyablov and Pinsker does not pin-point the optimal value of L for any (ρ, L)-list decodable code with rate $1 - H_q(\rho) - \varepsilon$. In particular, their results do not seem to imply any *lower bound* on L for such codes. This leads to the following natural question (which was also the starting point of our work):

Question 1. *Do q-ary (ρ, L)-list decodable codes of rate $1 - H_q(\rho) - \varepsilon$ need the list size L to be at least $\Omega(1/\varepsilon)$?*

We now digress briefly to talk about algorithmic aspects and its implications for list decoding. For most applications of list decoding, the combinatorial guarantee of good list decodability must be backed up with efficient polynomial time list decoding algorithms. Note that this imposes an a priori requirement that the list size needs to be at most some polynomial in the block length of the code. Good list decodable codes with efficient list decoding algorithms have found many applications in theoretical computer science in general and complexity theory in particular (see for example the survey by Sudan [11] and Guruswami's thesis [4, Chap. 12]). Such codes also have potential applications in traditional coding theory applications domains such as communication (cf. [9, Chap. 1]). One interesting contrast in these applications are the regimes of ρ that they operate in. The applications in complexity theory require ρ to be very close to $1 - 1/q$, while in the communication setting, ρ being closer to zero is the more interesting setting. Thus, the entire spectrum of ρ merits study.

We now return to Question 1. For *binary* (ρ, L)-list decodable codes with rate $1 - H(\rho) - \varepsilon$ Blinovsky provides some partial answers [1]. In particular, for $\rho = 1/2 - \sqrt{\varepsilon}$, a tight bound on L of $\Omega(1/\varepsilon)$ is shown in [1]. For smaller ρ (in particular for constant ρ independent of ε), the result in [1] implies a lower bound of $\Omega(\log(1/\varepsilon))$ on L.[2] Thus, Blinovsky's result implies a positive resolution of Question 1 only for binary codes for $\rho = 1/2 - \sqrt{\varepsilon}$. This result was extended to q-ary codes by Guruswami and Vadhan [6]. In particular, they show that any q-ary $(1 - 1/q - \sqrt{\varepsilon}, L)$-list decodable code with *any* constant rate needs $L = \Omega(1/\varepsilon)$. The result in [6] is proved by a different and simpler proof than the one used in [1].

Unfortunately, it is not clear how to strengthen the proof of Blinovsky to answer Question 1 in the affirmative for binary codes. Further, the proof of Guruswami and Vadhan crucially uses the fact that ρ is very close to $1 - 1/q$. Given this, answering Question 1 in its full generality seems to be tricky. However, if we scale down our ambitions, as a special case of Question 1 one can ask if the answer for random codes is positive (say with high probability). That is

Question 2. *Do* random *q-ary (ρ, L)-list decodable codes of rate $1 - H_q(\rho) - \varepsilon$ with high probability need the list size L to be at least $\Omega(1/\varepsilon)$?*

The above is a natural question given the result that random codes with high probability need the list size to be at most $1/\varepsilon$ [14].[3] In particular, Question 2 asks if the analysis of [14] is tight.

In this paper, we answer Question 2 affirmatively. Our main result states that a random q-ary (ρ, L)-list decodable code of rate $1 - H_q(\rho) - \varepsilon$ with high probability needs L to be $\Omega((1 - H_q(\rho))/\varepsilon)$, which is tight for any constant $\rho < 1 - 1/q$. In fact, our results also hold if we restrict our attention

rate $1 - H_q(\rho) + \varepsilon$, there is a constant probability that every decoding procedure will not be able to recover the transmitted message.

[2] These bounds are implicit in [1]. See Appendix A for why the result of Blinovsky implies the lower bounds claimed here.

[3] This question was first put to us by Bobby Kleinberg.

to random *linear* codes.[4] We remark that our results are somewhat incomparable with those of [1, 6]. On the one hand, our results give tight lower bounds for the range of values of ρ and q for which the previous works do not work. On the other hand, the previous works are more general as their lower bounds work for *arbitrary* codes.

Proof Overview We now will briefly sketch the main ideas in our proof. First we review the proof of the "positive" result of [14], as it is the starting point for our proof. In particular, we will argue that for a random code of rate $1 - H_q(\rho) - 1/L$, the probability that some received word $\mathbf{y}$ has some $L+1$ set of codewords within a relative Hamming distance of ρ is exponentially small in the block length of the code. Note that this implies that with high probability the code is (ρ, L)-list decodable. Let k and n denote the dimension and block length of the code (note that $k/n = 1 - H_q(\rho) - 1/L$). To bound the probability of a "bad" event, the proof first shows that for a fixed set of $L+1$ messages and a received word $\mathbf{y}$, the probability that all the corresponding $L+1$ codewords lie within a Hamming ball of radius ρn centered around $\mathbf{y}$ is at most $q^{-n(L+1)(1-H_q(\rho))}$. Thus, the probability that some set of $L+1$ codewords lie within the Hamming ball is (by the union bound) at most

$$\binom{q^k}{L+1} \cdot q^{-n(L+1)(1-H_q(\rho))} \leqslant q^{-n(L+1)(1-H_q(\rho)-k/n)} \leqslant q^{-n(1+1/L)}.$$

Again by the union bound, the probability that the bad event happens for some received word is at most $q^{-n/L}$, as required.

Let us recast the calculation above in the following manner. Consider a random code of rate $1 - H_q(\rho) - \varepsilon$. Then the expected number of received words that have some L codewords in a Hamming ball of radius ρn around it is at most $q^{-nL(1-H_q(\rho)-k/n-1/L)}$. As $k/n = 1 - H_q(\rho) - \varepsilon$, if we pick $L = \left\lceil \frac{1}{2\varepsilon} \right\rceil$ then the expected number of received words with L codewords in a Hamming ball of radius ρn is $q^{\varepsilon n L}$. Now this is encouraging news if we want to prove a *lower bound* on L. Unfortunately, the bound on expectation above is an upper bound. However, if somehow the corresponding events were disjoint then we will be in good shape as for disjoint events the union bound is tight.

The main idea of this paper is to make the relevant events in the paragraph above disjoint. In particular, for now assume that $\rho < 1/2(1 - 1/q)$ and consider a code Y of constant rate with distance 2ρ (by the Gilbert-Varshamov bound [3, 12] such codes exist). Let $\mathbf{y}_1 \neq \mathbf{y}_2$ be codewords in this code. Now the events that a fixed set of L codewords lie inside the Hamming balls of (relative) radius ρ around $\mathbf{y}_1$ and $\mathbf{y}_2$ are *disjoint*. By doing the calculations a bit carefully one can show that this implies that in expectation, exponentially many $\mathbf{y} \in Y$ have some L set of codewords within relative Hamming distance of ρ. Thus, for some code, there exists some received word for which the output list size needs to be at least L. To convert this into a high probability event, we bound the variance of these events (again the notion of disjointness discussed above helps) and then appeal to Chebyschev's inequality.

The drawback of the approach above is that one can only prove the required tight lower bound on the list size for $\rho < 1/2(1 - 1/q)$. To push up ρ close to $1 - 1/q$, we will need the following idea. Instead of carving the space of received words into (exponentially many) Hamming balls of relative radius 2ρ, we will carve the space into exponentially many disjoint clusters with the

[4]The story for random linear codes is a bit different from that of general codes, as for random linear codes only an upper bound of $q^{O(1/\varepsilon)}$ on L is known.

following properties. Every vector in a cluster is within a relative Hamming distance of ρ from the cluster center. Further, the size of every cluster is very close to the volume of a Hamming ball of relative radius ρ. It turns out that this approximation is good enough for the proof idea in the previous paragraph to go through. Such a carving with high probability can be obtained from a random code of rate close to $1 - H_q(\rho)$ (the cluster centers are the codewords in this code). Interestingly, the proof of this claim is implicit in Shannon's original work.

Organization In Section 2 we will set up some preliminaries including the proof of the existence of the special kind of carving mentioned in the paragraph above. We prove our main result in Section 3 and conclude with some open questions in Section 4.

2 Preliminaries

For an integer $m \geqslant 1$, we will use $[m]$ to denote the set $\{1, \ldots, m\}$.

We will now quickly review the basic concepts from coding theory that will be needed for this work. A code of *dimension* k and *block length* n over an alphabet Σ is a subset of Σ^n of size $|\Sigma|^k$. By abuse of notation we will also think of a code C as a map from elements in Σ^k to their corresponding codeword in Σ^n. The *rate* of such a code equals k/n. Each vector in C is called a codeword. Throughout this paper, we will use $q \geqslant 2$ to denote the size of the alphabet of a code. We will denote by $\mathbb{F}_q$ the field with q elements. A code C over $\mathbb{F}_q$ is called a linear code if C is a linear subspace of $\mathbb{F}_q^n$. In this case the dimension of the code coincides with the dimension of C as a vector space over $\mathbb{F}_q$.

We will use boldface letters to denote vectors and $\mathbf{0}$ will denote the all zeros-vector. The Hamming distance between two vectors $\mathbf{u}, \mathbf{v} \in \Sigma^n$, denoted by $\Delta(\mathbf{u}, \mathbf{v})$ is the number of places they differ in. The volume of a Hamming ball of radius d is defined as follows:

$$\mathrm{Vol}_q(\mathbf{u}, d) = |\{\mathbf{v} | \Delta(\mathbf{u}, \mathbf{v}) \leqslant d\}| .$$

We will use the following well known bound (cf. [8]):

$$q^{H_q(\rho)n - o(n)} \leqslant \mathrm{Vol}_q(\mathbf{y}, \rho n) \leqslant q^{H_q(\rho)n}, \tag{1}$$

for every $\mathbf{y} \in [q]^n$ and $0 < \rho < 1 - 1/q$.

The (minimum) distance of a code C is the minimum Hamming distance between any two pairs of distinct codewords from C. The relative distance is the ratio of the distance to the block length.

We will need the following notion of a carving of a vector space.

Definition 1. *Let $n \geqslant 1$ and $q \geqslant 2$ be integers and let $0 < \rho < 1 - 1/q$ and $\gamma \geqslant 0$ be reals. Then $\mathcal{P} = (\mathcal{H}, \mathcal{B})$ is a* (ρ, γ)-carving *of $[q]^n$ if the following hold:*

(a) $\mathcal{H} \subseteq [q]^n$ and $\mathcal{B} : \mathcal{H} \to 2^{[q]^n}$.

(b) For every $\mathbf{x} \neq \mathbf{y} \in \mathcal{H}$, $\mathcal{B}(\mathbf{x}) \cap \mathcal{B}(\mathbf{y}) = \emptyset$.

(c) For every $\mathbf{y} \in \mathcal{H}$ and $\mathbf{x} \in \mathcal{B}(\mathbf{y})$, $\Delta(\mathbf{y}, \mathbf{x}) \leqslant \rho n$.

(d) For every $\mathbf{y} \in \mathcal{H}$, $\mathrm{Vol}_q(\mathbf{0}, \rho n) \geqslant |\mathcal{B}(\mathbf{y})| \geqslant (1 - q^{-\gamma n})\mathrm{Vol}_q(\mathbf{0}, \rho n)$.

The size *of $\mathcal{P}$ is $|\mathcal{H}|$.*

In our proof we will need a (ρ, γ)-carving of $[q]^n$ of size $q^{\Omega(n)}$.

2.1 Existence of (ρ, γ)-carvings

To get a feel for these kinds of carvings, let us first consider the special case when $0 < \rho < 1/2(1-1/q)$ and $\gamma = \infty$. Let $\mathcal{P} = (\mathcal{H}, \mathcal{B})$ be a (ρ, γ)-carving of $[q]^n$. Note that since $\gamma = \infty$, then by definition, $\mathcal{B}$ maps each element in $\mathcal{H}$ to Hamming balls of radius ρn around them. Thus, if we pick $\mathcal{H}$ to be a q-ary code of distance $2\rho n + 2$, then $\mathcal{P}$ does satisfy the conditions of Definition 1. By the Gilbert-Varshamov bound, we know that there exists $\mathcal{H}$ with $|\mathcal{H}| \geqslant q^{(1-H_q(2\rho)-\varepsilon)n}$ for any $\varepsilon > 0$. Unfortunately, the limitation of $\rho < 1/2(1-1/q)$ is unsatisfactory. Next we show how we can remove this constraint (at the expense of having a smaller γ).

Lemma 1. *Let $q \geqslant 2$ be an integer and let $0 < \rho < 1 - 1/q$ and $\gamma > 0$ be reals. Then for large enough n, there exist a (ρ, γ)-carving $\mathcal{P} = (\mathcal{H}, \mathcal{B})$ of $[q]^n$ of size at least $q^{(1-H_q(\rho)-2\gamma)n}$.*

We remark that the proof of Lemma 1 follows from Shannon's proof of the capacity of the q-ary Symmetric Channel (with cross-over probability ρ) [10]. In particular, picking $\mathcal{H}$ to be a random code of rate slightly less than $1 - H_q(\rho)$ satisfies the required property with high probability. As a corollary, this implies that for random codes, for *most* error patterns, list decoding up to a radius of ρ will output at most *one* codeword. The connection to Shannon's proof has been made before (cf. [5, 9]). For the sake of completeness, we now present the proof.

Proof of Lemma 1 Let $k = \lfloor (1 - H_q(\rho) - 3\gamma/2)n \rfloor$ and define a code $C : [q]^k \to [q]^n$ as follows. For every $\mathbf{m} \in [q]^k$, pick $C(\mathbf{m})$ to be a uniformly chosen random vector from $[q]^n$. Further, the choices for different $\mathbf{m}$ are independent. Define the set $S_{\mathbf{m}}$ as follows:

$$S_{\mathbf{m}} = \{\mathbf{y} \in [q]^n | \Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n \text{ and } \forall \mathbf{m}' \neq \mathbf{m} \in [q]^k,\ \Delta(\mathbf{y}, C(\mathbf{m})) < \Delta(\mathbf{y}, C(\mathbf{m}'))\}.$$

Define $T_{\mathbf{m}} = \{\mathbf{y} \in [q]^n | \Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n \text{ and } \mathbf{y} \notin S_{\mathbf{m}}\}$. Note that $|T_{\mathbf{m}}|$ is a random variable and we will prove an upper bound on its expectation. Consider the following sequence of inequalities:

$$\begin{aligned}
\mathop{\mathbf{E}}_{C}[|T_{\mathbf{m}}|] &\leqslant \sum_{\substack{\mathbf{y} \in [q]^n \\ \Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n}} \sum_{\mathbf{m}' \neq \mathbf{m} \in [q]^k} \Pr[\Delta(\mathbf{y}, C(\mathbf{m}')) \leqslant \Delta(\mathbf{y}, C(\mathbf{m}))] && (2) \\
&\leqslant \sum_{\substack{\mathbf{y} \in [q]^n \\ \Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n}} \sum_{\mathbf{m}' \neq \mathbf{m} \in [q]^k} \Pr[\Delta(\mathbf{y}, C(\mathbf{m}')) \leqslant \rho n] && (3) \\
&\leqslant \sum_{\substack{\mathbf{y} \in [q]^n \\ \Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n}} q^k \cdot \left(\frac{\mathrm{Vol}_q(\mathbf{y}, \rho n)}{q^n} \right) && (4) \\
&\leqslant q^{(1-H_q(\rho)-\frac{3\gamma}{2}-1+H_q(\rho))n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n) && (5) \\
&= q^{-\frac{3\gamma}{2}n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n) && (6)
\end{aligned}$$

In the above, (2) follows from the definition of $T_{\mathbf{m}}$, the fact that the random choices for $C(\mathbf{m})$ and $C(\mathbf{m}')$ are independent and the union bound. (3) follows from the fact that $\Delta(\mathbf{y}, C(\mathbf{m})) \leqslant \rho n$. (4) follows from the fact that $C(\mathbf{m}')$ is a random vector in $[q]^n$ (and by bounding the number of choices for $\mathbf{m}'$ to be q^k). (5) follows from the upper bound in (1), the choice of k and the fact that the volume of a Hamming ball is translation invariant.

We are almost done except for the possibility that (6) need not hold simultaneously for every $\mathbf{m} \in [q]^k$. To remedy this we will remove half the number of points in C. In particular, note that since (6) holds for arbitrary $\mathbf{m}$,

$$\mathop{\mathbf{E}}_{C}\mathop{\mathbf{E}}_{\mathbf{m}}[|T_{\mathbf{m}}|] = \mathop{\mathbf{E}}_{\mathbf{m}}\mathop{\mathbf{E}}_{C}[|T_{\mathbf{m}}|] \leqslant q^{-\frac{3\gamma}{2}n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n).$$

Thus, by the probabilistic method, there exists a C such that $\mathop{\mathbf{E}}_{\mathbf{m}}[|T_{\mathbf{m}}|] \leqslant q^{-\frac{3\gamma}{2}n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n)$. For the rest of the proof, let C denote such a code. Define $\mathcal{H}$ to be the set of codewords $C(\mathbf{m})$ for which $|T_{\mathbf{m}}|$ is in the smaller half among all $\mathbf{m} \in [q]^k$. By a Markov argument, this implies that for every $C(\mathbf{m}) \in \mathcal{H}$, $|T_{\mathbf{m}}| \leqslant 2 \cdot q^{-\frac{3\gamma}{2}n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n) \leqslant q^{-\gamma n} \cdot \mathrm{Vol}_q(\mathbf{0}, \rho n)$ (where the last inequality is true for large enough n). For every $\mathbf{y} \in \mathcal{H}$ such that $\mathbf{y} = C(\mathbf{m})$, define $\mathcal{B}(\mathbf{y}) = S_{\mathbf{m}}$. Finally, note that for large enough n, $|\mathcal{H}| = q^k/2 \geqslant q^{(1-H_q(\rho)-2\gamma)n}$. It can be easily verified that $(\mathcal{H}, \mathcal{B})$ satisfies all the conditions of Definition 1. ■

3 Lower Bound for General Random Codes

We start with the main technical result of the paper.

Lemma 2. *Let $q \geqslant 2$ be an integer and let $0 < \rho < 1 - 1/q$ and $\varepsilon > 0$ be reals. If $\mathcal{P} = (\mathcal{H}, \mathcal{B})$ is a (ρ, ε)-carving of $[q]^n$ of size $q^{\alpha n}$ for some $\alpha > 0$ then the following holds for every large enough integer n. A random code of rate $1 - H_q(\rho) - \varepsilon$, with high probability, has at least α/ε codewords in a Hamming ball of radius ρn centered at some vector in $\mathcal{H}$.*

Lemmas 1 and 2 imply our main result.

Theorem 3 (Main Result)**.** *Let $q \geqslant 2$ be an integer and $0 < \rho < 1 - 1/q$ and $\varepsilon > 0$ be reals. Then a random code of rate $1 - H_q(\rho) - \varepsilon$, with high probability, is not (ρ, L) list-decodable for any $L \leqslant c_{q,\rho}/\varepsilon$, where $c_{q,\rho} > 0$ is a real number that depends only on q and ρ.*

In the rest of the section, we will prove Lemma 2.

Define $L = \left\lfloor \frac{\alpha}{3\varepsilon} \right\rfloor$ and $k = \lfloor (1 - H_q(\rho) - \varepsilon)n \rfloor$. Let $\mathcal{M}$ be the set of L tuples of distinct messages from $[q]^k$. Note that $|\mathcal{M}| = \binom{q^k}{L}$. Let C be a random q-ary code of dimension k and block length n (i.e. each message is assigned a random independent vector from $[q]^n$).

We now define a few indicator variables that will be needed in the proof. For any $\mathbf{m} = (\mathbf{m}_1, \ldots, \mathbf{m}_L) \in \mathcal{M}$ and $\mathbf{y} \in [q]^n$, define the indicator variable $X_{\mathbf{m},\mathbf{y}}$ as follows.[5]

$$X_{\mathbf{m},\mathbf{y}} = 1 \text{ if and only if } \{C(\mathbf{m}_1), \ldots, C(\mathbf{m}_L)\} \subseteq \mathcal{B}(\mathbf{y}).$$

Note that if $X_{\mathbf{m},\mathbf{y}} = 1$ then $\mathbf{m}$ and $\mathbf{y}$ form a "witness" to the fact that the code C needs to have an output list size of at least L. We also define a related indicator random variable:

$$Y_{\mathbf{m}} = 1 \text{ if and only if } \sum_{\mathbf{y} \in \mathcal{H}} X_{\mathbf{m},\mathbf{y}} \geqslant 1.$$

[5] All the indicator variables should also depend on C but we suppress the dependence to make the expressions simpler.

Finally define the following random variable:

$$Z = \sum_{\mathbf{m} \in \mathcal{M}} Y_{\mathbf{m}}.$$

Note that to prove Lemma 2 it suffices to show that with high probability $Z \geqslant 1$. To this end, we will first bound the expectation and variance of Z and then invoke Chebyschev's inequality to obtain the required high probability guarantee.

We begin by proving a lower bound on $\mathbf{E}[Z]$. As the codewords in C are chosen uniformly and independently at random from $[q]^n$, for every $\mathbf{m} \in \mathcal{M}$ and $\mathbf{y} \in \mathcal{H}$, we have:

$$\Pr[X_{\mathbf{m},\mathbf{y}} = 1] = \left(\frac{|\mathcal{B}(\mathbf{y})|}{q^n}\right)^L.$$

By property (d) in Definition 1, this implies that

$$(1 - q^{-\varepsilon n})^L \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L \leqslant \Pr[X_{\mathbf{m},\mathbf{y}} = 1] \leqslant \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L \tag{7}$$

By property (b) in Definition 1, it follows that for any $\mathbf{m} \in \mathcal{M}$ and $\mathbf{y}_1 \neq \mathbf{y}_2 \in \mathcal{H}$, the events $X_{\mathbf{m},\mathbf{y}_1} = 1$ and $X_{\mathbf{m},\mathbf{y}_2} = 1$ are *disjoint* events. This along with the lower bound in (7) implies the following for every $\mathbf{m} \in \mathcal{M}$:

$$\begin{aligned}
\mathbf{E}[Y_{\mathbf{m}}] &= \sum_{\mathbf{y} \in \mathcal{H}} \Pr[X_{\mathbf{m},\mathbf{y}} = 1] \\
&\geqslant |\mathcal{H}|(1 - q^{-\varepsilon n})^L \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L \\
&\geqslant |\mathcal{H}|(1 - 2Lq^{-\varepsilon n}) \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L, \qquad (8)
\end{aligned}$$

where the last inequality follows for large enough n. Using the upper bound in (7) in the above, we also get the following bound:

$$\mathbf{E}\left[Y_{\mathbf{m}}\right] \leqslant |\mathcal{H}| \cdot \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L \tag{9}$$

By linearity of expectation, we have

$$\begin{aligned}
\mathbf{E}[Z] &= \sum_{\mathbf{m} \in \mathcal{M}} \mathbf{E}[Y_{\mathbf{m}}] \\
&\geqslant |\mathcal{M}| \cdot |\mathcal{H}| \cdot (1 - 2Lq^{-\varepsilon n}) \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L && (10) \\
&\geqslant \binom{q^k}{L} \cdot \frac{q^{\alpha n}}{2} \cdot \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n}\right)^L && (11) \\
&\geqslant \frac{q^{kL + \alpha n}}{2L^L} \cdot q^{-nL(1 - H_q(\rho)) - o(n)} && (12) \\
&\geqslant q^{nL(1 - H_q(\rho) - \varepsilon + \frac{\alpha}{L} - 1 + H_q(\rho)) - o(n)} && (13)
\end{aligned}$$

$$\geqslant q^{2\varepsilon nL - o(n)}. \tag{14}$$

In the above (10) follows from (8) while (11) follows from the fact that for large enough n, $2Lq^{-\varepsilon n} \leqslant 1/2$ (and plugging in the values of $|\mathcal{M}|$ and $|\mathcal{H}|$). (12) follows from the lower bound in (1) and the lower bound $\binom{a}{b} \geqslant (a/b)^b$. (13) and (14) follow from the choice of k and L (and by absorbing the "extra" terms like $2L^L$ into the $o(n)$ term). Using (9) in the above instead of (8) gives us the following bound

$$\mathbf{E}\left[Z\right] \leqslant q^{2\varepsilon nL}. \tag{15}$$

Next, we bound the variance of Z. As a first step in that direction, we will upper bound $\mathbf{E}[Z^2]$. By definition, we have

$$\mathbf{E}\left[Z^2\right] = \sum_{\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}} \Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1]. \tag{16}$$

By abuse of notation, for every $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$, define $\mathbf{m}^1 \cap \mathbf{m}^2$ to be the set of vectors from $[q]^k$ that occur in both the tuples $\mathbf{m}^1$ and $\mathbf{m}^2$. Similarly, $\mathbf{m}^1 \cup \mathbf{m}^2$ will denote the set of vectors from $[q]^k$ that occur in $\mathbf{m}^1$ or $\mathbf{m}^2$. With this notation in place, let us rewrite the summation in (16):

$$\mathbf{E}\left[Z^2\right] = \sum_{i=1}^{L} \sum_{\substack{\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M} \\ |\mathbf{m}^1 \cap \mathbf{m}^2| = i}} \Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] + \sum_{\substack{\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M} \\ \mathbf{m}^1 \cap \mathbf{m}^2 = \emptyset}} \Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] \tag{17}$$

We will bound the two summations in (17) using the following observation. Note that if $\mathbf{m}^1 \cap \mathbf{m}^2 \neq \emptyset$ then for every $\mathbf{y}_1 \neq \mathbf{y}_2 \in \mathcal{H}$, *both* $X_{\mathbf{m}^1, \mathbf{y}_1}$ and $X_{\mathbf{m}^2, \mathbf{y}_2}$ cannot be 1 *simultaneously* (this follows from the definition of $X_{(\cdot,\cdot)}$ and property (b) in Definition 1).

We will now bound the first summation in (17). Fix $1 \leqslant i \leqslant L$ and $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$ such that $|\mathbf{m}^1 \cap \mathbf{m}^2| = i$. By the definition of the indicator variable $Y_{(\cdot)}$,

$$\begin{aligned}
\Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] &= \sum_{\mathbf{y}_1 \in \mathcal{H}} \sum_{\mathbf{y}_2 \in \mathcal{H}} \Pr[X_{\mathbf{m}^1, \mathbf{y}_1} = 1 \text{ and } X_{\mathbf{m}^2, \mathbf{y}_2} = 1] \\
&= \sum_{\mathbf{y} \in \mathcal{H}} \Pr[X_{\mathbf{m}^1, \mathbf{y}} = 1 \text{ and } X_{\mathbf{m}^2, \mathbf{y}} = 1] && (18) \\
&= \sum_{\mathbf{y} \in \mathcal{H}} \left(\frac{\mathrm{Vol}_q(\mathbf{y}, \rho n)}{q^n} \right)^{2L-i} && (19) \\
&= |\mathcal{H}| \cdot \left(\frac{\mathrm{Vol}_q(\mathbf{0}, \rho n)}{q^n} \right)^{2L-i} && (20)
\end{aligned}$$

In the above, (18) follows from the discussion in the paragraph above. (19) follows from the fact that every message in $\mathbf{m}^1 \cup \mathbf{m}^2$ is mapped to an independent random vector in $[q]^n$ by our choice of C (note that $|\mathbf{m}^1 \cup \mathbf{m}^2| = 2L - i$). Finally, (20) follows from the fact that the volume of a Hamming ball is translation invariant.

Now the number of tuples $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$ such that $|\mathbf{m}^1 \cap \mathbf{m}^2| = i$ is upper bounded by $\binom{L}{i}^2 q^{k(2L-i)} \leqslant L^{2L} q^{k(2L-i)}$. Thus, by (20) we get

$$\sum_{i=1}^{L} \sum_{\substack{\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M} \\ |\mathbf{m}^1 \cap \mathbf{m}^2| = i}} \Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] \leqslant L^{2L} |\mathcal{H}| \sum_{i=1}^{L} \left(q^{k-n} \mathrm{Vol}_q(\mathbf{0}, \rho n) \right)^{2L-i}$$

$$\leqslant L^{2L}|\mathcal{H}|\sum_{i=1}^{L} q^{(2L-i)n(1-H_q(\rho)-\varepsilon-1+H_q(\rho))} \tag{21}$$

$$\leqslant L^{2L}|\mathcal{H}|\sum_{i=1}^{L} q^{-\varepsilon nL} \tag{22}$$

$$= L^{2L+1}\cdot q^{nL\left(\frac{\alpha}{L}-\varepsilon\right)} \tag{23}$$

$$\leqslant L^{2L+1}\cdot q^{2\varepsilon nL} \tag{24}$$

In the above (21) follows from our choice of k and the upper bound in (1). (22) follows from the fact that $i \leqslant L$ while (23) follows from the size of $\mathcal{H}$. Finally, (24) follows from the choice of L.

We now proceed to upper bound the second summation in (17). Fix $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$ such that $\mathbf{m}^1 \cap \mathbf{m}^2 = \emptyset$. By the definition of $Y_{(\cdot)}$,

$$\Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] = \sum_{\mathbf{y}_1,\mathbf{y}_2\in\mathcal{H}} \Pr[X_{\mathbf{m}^1,\mathbf{y}_1} = 1 \text{ and } X_{\mathbf{m}^2,\mathbf{y}_2} = 1]$$

$$= \sum_{\mathbf{y}_1,\mathbf{y}_2\in\mathcal{H}} \left(\frac{\mathrm{Vol}_q(\mathbf{y},\rho n)}{q^n}\right)^{2L} \tag{25}$$

$$= (|\mathcal{H}|)^2\cdot\left(\frac{\mathrm{Vol}_q(\mathbf{0},\rho n)}{q^n}\right)^{2L} \tag{26}$$

In the above (25) follows from the fact that the messages in $\mathbf{m}^1 \cup \mathbf{m}^2$ are assigned random independent codewords from $[q]^n$. Since the number of tuples $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$ such that $\mathbf{m}^1 \cap \mathbf{m}^2 = \emptyset$ is upper bounded by $|\mathcal{M}|^2$, by (26) we have

$$\sum_{\substack{\mathbf{m}^1,\mathbf{m}^2\in\mathcal{M}\\ \mathbf{m}^1\cap\mathbf{m}^2=\emptyset}} \Pr[Y_{\mathbf{m}^1} = 1 \text{ and } Y_{\mathbf{m}^2} = 1] \leqslant (|\mathcal{M}|\cdot|\mathcal{H}|)^2\left(\frac{\mathrm{Vol}_q(\mathbf{0},\rho n)}{q^n}\right)^{2L}$$

$$\leqslant \left(\frac{\mathbf{E}[Z]}{1-2Lq^{-\varepsilon n}}\right)^2 \tag{27}$$

$$\leqslant (1+8Lq^{-\varepsilon n})\cdot(\mathbf{E}[Z])^2. \tag{28}$$

In the above (27) follows from (10) while (28) is true for large enough n.

We are finally ready to bound the variance of Z.

$$\sigma^2[Z] = \mathbf{E}[Z^2] - (\mathbf{E}[Z])^2$$

$$\leqslant L^{2L+1}\cdot q^{2\varepsilon nL} + 8Lq^{-\varepsilon n}(\mathbf{E}[Z])^2 \tag{29}$$

$$\leqslant L^{2L+1}\cdot q^{2\varepsilon nL} + 8Lq^{4\varepsilon nL-\varepsilon n} \tag{30}$$

$$\leqslant q^{4\varepsilon nL-\varepsilon n+o(n)} \tag{31}$$

In the above, (29) follows from (17), (24) and (28). (30) follows from (15). (31) follows from absorbing the multiplicative constants in the $o(n)$ term.

Recall that we set out to prove that $\Pr[Z \geqslant 1]$ is large. Indeed,

$$\Pr[Z < 1] \leqslant \Pr[|Z - \mathbf{E}[Z]| > \mathbf{E}[Z]/2] \tag{32}$$

$$\leqslant \frac{4\sigma^2[Z]}{(\mathbf{E}[Z])^2} \tag{33}$$

$$\leqslant \frac{4q^{4\varepsilon nL - \varepsilon n + o(n)}}{q^{4\varepsilon nL - o(n)}} \tag{34}$$

$$\leqslant q^{-\varepsilon n/2}, \tag{35}$$

as desired. In the above (32) follows from the fact that for large enough n, $\mathbf{E}[Z] \geqslant 2$ (the latter fact follows from (14)). (33) follows from the Chebyschev's inequality. (34) follows from (31) and (14). (35) is true for large enough n. The proof of Lemma 2 is now complete.

3.1 Lower Bound for Random Linear Codes

The following result analogous to Theorem 3 holds for linear codes.

Theorem 4 (Linear Codes)**.** *Let $q \geqslant 2$ be an integer and $0 < \rho < 1 - 1/q$ and $\varepsilon > 0$ be reals. Then a random linear code of rate $1 - H_q(\rho) - \varepsilon$, with high probability, is not (ρ, L) list-decodable for any $L \leqslant c_{q,\rho}/\varepsilon$, where $c_{q,\rho} > 0$ is a real number that depends only on q and ρ.*

The proof of Theorem 4 is very similar to that of Theorem 3, so here we will just sketch how the proof of Theorem 3 needs to be modified. First, for a q-ary random linear code code C (of dimension k and block length n) it is not true that $C(\mathbf{m}_1)$ and $C(\mathbf{m}_2)$ are random *independent* vectors in $\mathbb{F}_q^n$. However, if $\mathbf{m}_1$ and $\mathbf{m}_2$ are linearly independent vectors over $\mathbb{F}_q$, then $C(\mathbf{m}_1)$ and $C(\mathbf{m}_2)$ are both random independent vectors from $\mathbb{F}_q^n$. Thus, we choose $\mathcal{M}$ to be the set of L tuples from $\mathbb{F}_q^k$ that are all linearly independent. It is well known that $\binom{q^k}{L} \geqslant |\mathcal{M}| \geqslant (1 - q^{-n+L})\binom{q^k}{L} \geqslant \frac{1}{2}\binom{q^k}{L}$. Recall in the proof of Theorem 3, we had $|\mathcal{M}| = \binom{q^k}{L}$. However, this change in size only changes the constants in the calculations.

The second place where the proof differs from that of Theorem 3 is the interpretation of $\mathbf{m}^1 \cap \mathbf{m}^2$ for any $\mathbf{m}^1, \mathbf{m}^2 \in \mathcal{M}$ (the interpretation for $\mathbf{m}^1 \cup \mathbf{m}^2$ remains the same): $\mathbf{m}^1 \cap \mathbf{m}^2$ denotes the smallest set such that the vectors in $(\mathbf{m}^1 \cup \mathbf{m}^2) \setminus (\mathbf{m}^1 \cap \mathbf{m}^2)$ are independent over $\mathbb{F}_q$. The rest of the calculations however, syntactically remain the same.

4 Open Questions

In this work we proved that a random q-ary (ρ, L) list decodable code of rate $1 - H_q(\rho) - \varepsilon$ needs L to be at least $\Omega((1 - H_q(\rho))/\varepsilon)$ with high probability. It would be nice if we can prove a lower bound of the form $L \geqslant c_q/\varepsilon$, where c_q is an absolute constant that only depends on q. The obvious open question is to resolve Question 1. We conjecture that the answer should be positive.

Acknowledgments

Thanks to Bobby Kleinberg for asking whether lower bounds on list sizes for list decoding random codes are known. We thank Venkatesan Guruswami, Madhu Sudan and Santosh Vempala for helpful discussions on related topics.

References

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

[2] P. Elias. List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957.

[3] E. N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31:504–522, 1952.

[4] V. Guruswami. *List decoding of error-correcting codes*. Number 3282 in Lecture Notes in Computer Science. Springer, 2004.

[5] V. Guruswami. *Algorithmic Results in List Decoding*, volume 2 (Issue 2) of *Foundations and Trends in Theoretical Computer Science (FnT-TCS)*. NOW publishers, 2007.

[6] V. Guruswami and S. Vadhan. A lower bound on list size for list decoding. In *Proceedings of the 9th International Workshop on Randomization and Computation (RANDOM)*, pages 318–329, 2005.

[7] R. W. Hamming. Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, 29:147–160, April 1950.

[8] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier/North-Holland, Amsterdam, 1981.

[9] A. Rudra. *List Decoding and Property Testing of Error Correcting Codes*. PhD thesis, University of Washington, 2007.

[10] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[11] M. Sudan. List decoding: Algorithms and applications. *SIGACT News*, 31:16–27, 2000.

[12] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akadamii Nauk*, 117:739–741, 1957.

[13] J. M. Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958.

[14] V. V. Zyablov and M. S. Pinsker. List cascade decoding. *Problems of Information Transmission*, 17(4):29–34, 1981 (in Russian); pp. 236-240 (in English), 1982.

A Blinovsky's Lower bound

Blinovsky proved the following lower bound for (ρ, L)-list decodable binary codes.

Theorem 5 ([1]). *Let $L \geqslant 1$ be an integer and $0 \leqslant \rho \leqslant 1/2$ be a real. Then every (ρ, L)-list decodable binary code has rate at most*

$$1 - H(\lambda),$$

where λ ($0 \leqslant \lambda \leqslant 1/2$) is related to ρ as follows (below $\ell = \lceil L/2 \rceil$):

$$\rho = \sum_{i=1}^{\ell} \binom{2i-2}{i-1} \frac{(\lambda(1-\lambda))^i}{i}. \tag{36}$$

The expression in (36) is a bit complicated and there does not seem to be an clean expression for the lower bound on L when the code has rate $1 - H(\rho) - \varepsilon$. We will now approximate (36) to obtain a range within which the "true" lower bound has to lie. In particular, we will show that (36) implies

$$\rho + f(\ell, \rho) \leqslant \lambda \leqslant \rho + g(\ell, \rho), \tag{37}$$

where the functions $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ will be specified later. For small enough $\gamma > 0$, it is known that

$$c_1 \gamma^2 \leqslant H(\rho + \gamma) - H(\rho) \leqslant c_2 \gamma \log\left(\frac{1-\rho}{\rho}\right), \tag{38}$$

where c_1 and c_2 are absolute constants. A proof of the lower bound can be found, e.g., in [9, Chap. 2]. The upper bound follows from the facts that $(H(x))' = \log((1-x)/x)$ and $H(\cdot)$ is strictly concave.

Combining (37) and (38) we have

$$c_1 (f(\ell, \rho))^2 \leqslant H(\lambda) - H(\rho) \leqslant c_2 \log((1-\rho)/\rho) g(\ell, \rho).$$

Since we are interested in a code of rate $1 - H(\rho) - \varepsilon$, equating the rate with $1 - H(\lambda)$ and using the inequalities above, we get that ℓ ($= \lceil L/2 \rceil$) must satisfy the following two conditions:

$$f(\ell, \rho) \leqslant \sqrt{\frac{\varepsilon}{c_1}} \quad \text{and} \quad g(\ell, \rho) \geqslant \frac{\varepsilon}{c_2 \log((1-\rho)/\rho)}. \tag{39}$$

We will now slightly modify the calculations from [4, Sec. 5.3.3] to obtain the functions $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$. The calculations as stated in [4] obtain an expression for $f(\ell, 1/2 - \gamma)$ for small enough γ.

It is known (cf. [4]) that for any $0 \leqslant y \leqslant 1/2$,

$$\sum_{i=0}^{\infty} \binom{2i-2}{i-1} \frac{(y(1-y))^i}{i} = y.$$

Along with (36), this implies that

$$\lambda = \rho + \sum_{i=\ell+1}^{\infty} \binom{2i-2}{i-2} \frac{(\lambda(1-\lambda))^i}{i}. \tag{40}$$

Note that this in particular, implies that $\lambda \geqslant p$. We further claim that $\lambda \leqslant p + \sqrt{\varepsilon/c_1}$. Indeed, if $\lambda > p + \sqrt{\varepsilon/c_1}$ then by (38) and the fact that $H(\cdot)$ is an increasing function, we get $1 - H(\lambda) <$

$1 - H(\rho) - \varepsilon$. This by Theorem 5 would imply that there does not exists a (ρ, L)-list decodable code of rate $1 - H(\rho) - \varepsilon$ for any $L \geqslant 1$, which is a contradiction. Thus,

$$\rho \leqslant \lambda \leqslant p + \sqrt{\frac{\varepsilon}{c_1}} \tag{41}$$

Using the fact that if $i = \ell + 1 + j$,

$$\frac{\binom{2i-2}{i-1}\frac{1}{i}}{\binom{2\ell}{\ell}\frac{1}{\ell+1}} = 2^j \prod_{s=\ell+1}^{\ell+j} \frac{2s-1}{s+1},$$

we get

$$\left(\frac{2(2\ell+1)}{\ell+2}\right)^j \leqslant \frac{\binom{2i-2}{i-1}\frac{1}{i}}{\binom{2\ell}{\ell}\frac{1}{\ell+1}} \leqslant 4^j.$$

Using the above in (37), we get

$$\binom{2\ell}{\ell}\frac{(\lambda(1-\lambda))^{\ell+1}}{\ell+1}\sum_{j=0}^{\infty}(\lambda(1-\lambda))^j\left(\frac{2(2\ell+1)}{\ell+2}\right)^j \leqslant \lambda - \rho \leqslant \binom{2\ell}{\ell}\frac{(\lambda(1-\lambda))^{\ell+1}}{\ell+1}\sum_{j=0}^{\infty}(4\lambda(1-\lambda))^j,$$

which along with the fact that $\rho(1-\rho) \leqslant \lambda \leqslant (\rho + \sqrt{\varepsilon/c_1})(1 - \rho - \sqrt{\varepsilon/c_1})$ (which in turn follows from (41)) implies that we can choose

$$f(\ell, \rho) = \frac{1}{\ell+1}\binom{2\ell}{\ell}\frac{(\rho(1-\rho))^{\ell+1}}{1 - 4\rho(1-\rho) + 6\rho(1-\rho)/(\ell+2)} \tag{42}$$

and

$$g(\ell, \rho) = \frac{1}{\ell+1}\binom{2\ell}{\ell}\frac{\left((\rho + \sqrt{\varepsilon/c_1})(1 - \rho - \sqrt{\varepsilon/c_1})\right)^{\ell+1}}{1 - 4\rho(1-\rho) - 4(1-2\rho)\sqrt{\varepsilon/c_1} + 4\varepsilon/c_1)}. \tag{43}$$

Using the fact that $\frac{1}{\ell+1}\binom{2\ell}{\ell} = \Theta(4^\ell/\ell^{3/2})$, for constant ρ (and ε being small enough), the RHS of both (42) and (43) are $\Theta\left(\frac{(4\rho(1-\rho))^\ell}{\ell^{3/2}(1-4\rho(1-\rho))}\right)$.[6] This along with (39) implies the following:

Corollary 6. *Let $0 < \rho < 1/2$ be a constant and $\varepsilon > 0$ be a small enough real. Then every (ρ, L)-list decodable code of rate $1 - H(\rho) - \varepsilon$ must satisfy $L = \Omega(\log(1/\varepsilon))$. Further, Theorem 5 cannot imply an asymptotically stronger lower bound on L.*

[6] For $\rho = 1/2 - \sqrt{\varepsilon}$, (42) implies a tight lower bound of $L = \Omega(1/\varepsilon)$.