Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 17 (2012)

Combinatorial limitations of average-radius list-decoding

Venkatesan Guruswami* Srivatsan Narayanan*

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

April 2013

Abstract

We study certain combinatorial aspects of list-decoding, motivated by the exponential gap between the known upper bound (of $O(1/\gamma)$) and lower bound (of $\Omega_p(\log(1/\gamma))$) for the list-size needed to list decode up to error fraction p with rate γ away from capacity, i.e., $1 - h(p) - \gamma$ (here $p \in (0, \frac{1}{2})$ and $\gamma > 0$). Our main result is the following:

- We prove that in any binary code $C \subseteq \{0,1\}^n$ of rate $1 - h(p) - \gamma$, there must exist a set $\mathcal{L} \subset C$ of $\Omega_p(1/\sqrt{\gamma})$ codewords such that the average distance of the points in $\mathcal{L}$ from their centroid is at most pn. In other words, there must exist $\Omega_p(1/\sqrt{\gamma})$ codewords with low "average radius."

  The standard notion of list-decoding corresponds to working with the *maximum* distance of a collection of codewords from a center instead of *average* distance. The average-radius form is in itself quite natural; for instance, the classical Johnson bound in fact implies average-radius list-decodability.

The remaining results concern the standard notion of list-decoding, and help clarify the current state of affairs regarding combinatorial bounds for list-decoding:

- We give a short simple proof, over all fixed alphabets, of the above-mentioned $\Omega_p(\log(1/\gamma))$ lower bound. Earlier, this bound followed from a complicated, more general result of Blinovsky.
- We show that one *cannot* improve the $\Omega_p(\log(1/\gamma))$ lower bound via techniques based on identifying the zero-rate regime for list-decoding of constant-weight codes (this is a typical approach for negative results in coding theory, including the $\Omega_p(\log(1/\gamma))$ list-size lower bound). On a positive note, our $\Omega_p(1/\sqrt{\gamma})$ lower bound for average-radius list-decoding circumvents this barrier.
- We exhibit a "reverse connection" between the existence of constant-weight and general codes for list-decoding, showing that the best possible list-size, as a function of the gap γ of the rate to the capacity limit, is the same up to constant factors for both constant-weight codes (with weight bounded away from p by a constant) and general codes.
- We give simple second moment based proofs that w.h.p. a list-size of $\Omega_p(1/\gamma)$ is needed for list-decoding *random* codes from errors as well as erasures. For *random linear* codes, the corresponding list-size bounds are $\Omega_p(1/\gamma)$ for errors and $\exp(\Omega_p(1/\gamma))$ for erasures.

*Research supported in part by NSF grants CCF 0953155 and CCF 0963975. Email: `guruswami@cmu.edu`, `srivatsa@cs.cmu.edu`

ISSN 1433-8092

1 Introduction

The list-decoding problem for an error-correcting code $C \subseteq \Sigma^n$ consists of finding the set of all codewords of C with Hamming distance at most pn from an input string $y \in \Sigma^n$. Though it was originally introduced in early work of Elias and Wozencraft [6, 15] in the context of estimating the decoding error probability for random error models, recently the main interest in list-decoding has been for adversarial error models. List decoding enables correcting up to a factor two more worst-case errors compared to algorithms that are always restricted to output a unique answer, and this potential has even been realized algorithmically [10, 8].

In this work, we are interested in some fundamental combinatorial questions concerning list-decoding, which highlight the important tradeoffs in this model. Fix $p \in (0, \frac{1}{2})$ and a positive integer L. We say that a binary code $C \subseteq \{0,1\}^n$ is (p, L) list-decodable if every Hamming ball of radius pn has *less than* L codewords. Here, p corresponds to the error-fraction and L to the list-size needed by the error-correction algorithm. Note that (p, L) list-decodability imposes a sparsity requirement on the distribution of codewords in the Hamming space. A natural combinatorial question that arises in this context is to place bounds on the largest size of a code meeting this requirement. In particular, an outstanding open question is to characterize the maximum rate (defined to be the limiting ratio $\frac{1}{n}\log|C|$ as $n \to \infty$) of a (p, L) list-decodable code.

By a simple volume packing argument, it can be shown that a (p, L) list-decodable code has rate at most $1 - h(p) + o(1)$. (Throughout, for $z \in [0, \frac{1}{2}]$, we use $h(z)$ to denote the binary entropy function at z.) Indeed, picking a random center x, the Hamming ball $\mathbf{B}(x, pn)$ contains at least $|C| \cdot \binom{n}{pn} 2^{-n}$ codewords in expectation. Bounding this by $(L-1)$, we get the claim. On the positive side, in the limit of large L, the rate of a (p, L) list-decodable code approaches the optimal $1 - h(p)$. More precisely, for any $\gamma > 0$, there exists a $(p, 1/\gamma)$ list-decodable code of rate at least $1 - h(p) - \gamma$. In fact, a random code of rate $1 - h(p) - \gamma$ is $(p, 1/\gamma)$ list-decodable w.h.p. [16, 7], and a similar result holds for random linear codes (with list-size $O_p(1/\gamma)$) [9]. In other words, a dense random packing of $2^{(1-h(p)-\gamma)n}$ Hamming balls of radius pn (and therefore volume $\approx 2^{h(p)n}$ each) is "near-perfect" w.h.p. in the sense that no point is covered by more than $O_p(1/\gamma)$ balls.

The determination of the best asymptotic code rate of binary (p, L) list-decodable codes as p, L are held fixed and the block length grows is wide open for every choice of $p \in (0, \frac{1}{2})$ and integer $L \geqslant 1$. However, we *do* know that for each fixed $p \in (0, \frac{1}{2})$, this rate approaches $1 - h(p)$ in the limit as $L \to \infty$. To understand this rate of convergence as a function of list-size L, following [9], let us define $L_{p,\gamma}$ to be the minimum integer L such that there exist (p, L) list-decodable codes of rate $1 - h(p) - \gamma$ for infinitely many block lengths n (the quantity γ is the "gap" to "list-decoding capacity"). In [1], Blinovsky showed that a (p, L) list-decodable code has rate at most $1 - h(p) - 2^{-\Theta_p(L)}$. In particular, this implies that for any finite L, a (p, L) list-decodable code has rate strictly below the optimal $1 - h(p)$. Stated in terms of $L_{p,\gamma}$, his result implies the corollary $L_{p,\gamma} \geqslant \Omega_p(\log(1/\gamma))$ for rates γ-close to capacity. We provide a short and simple proof of this corollary in Section 4. Our proof works almost as easily over non-binary alphabets. (Blinovsky's subsequent proof for the non-binary case in [3, 4] involved substantial technical effort. However, his results also give non-trivial bounds for every finite L, as opposed to just the growth rate of $L_{p,\gamma}$.)

Observe the exponential gap (in terms of the dependence on γ) between the $O(1/\gamma)$ upper bound and $\Omega_p(\log(1/\gamma))$ lower bounds on the quantity $L_{p,\gamma}$. Despite being a basic and fundamental question about sphere packings in the Hamming space and its direct relevance to list-decoding, there has been no progress on narrowing this asymptotic gap in the 25 years since the works of Zyablov-Pinsker [16] and Blinovsky [1]. This is the motivating challenge driving this work.

1.1 Prior work on list-size lower bounds

We now discuss some lower bounds (besides Blinovsky's general lower bound) on list-size that have been obtained in restricted cases.

Rudra shows that the $O_p(1/\gamma)$ bound obtained via the probabilistic method for random codes is, in fact, tight up to constant factors [14]. Formally, there exists $L = \Omega_p(1/\gamma)$ such that a random code of rate $1 - h(p) - \gamma$ is *not* (p, L) list-decodable w.h.p. His proof uses near-capacity-achieving codes for the binary symmetric channel, the existence of which is promised by Shannon's theorem, followed by a second moment argument. We give a simpler proof of this result via a more direct use of the second moment method. This has the advantage that it works uniformly for random general as well as random linear codes, and for channels that introduce errors as well as erasures.

Guruswami and Vadhan [12] consider the problem of list-size bounds when the channel may corrupt close to half the bits, that is, when $p = \frac{1}{2} - \varepsilon$, and more generally $p = 1 - 1/q - \varepsilon$ for codes over an alphabet of size q. (Note that decoding is impossible if the channel could corrupt up to a half fraction of bits.) They show that there exists $c > 0$ such that for all $\varepsilon > 0$ and all block lengths n, any $(\frac{1}{2} - \varepsilon, c/\varepsilon^2)$ list-decodable code contains $O_\varepsilon(1)$ codewords. For p bounded away from $\frac{1}{2}$ (or $1 - 1/q$ in the q-ary case), their methods do not yield any nontrivial list-size lower bound as a function of gap γ to list-decoding capacity.

1.2 Our main results

We have already mentioned our new proof of the $\Omega(\log(1/\gamma))$ list-size lower bound for list-decoding general codes, and the asymptotically optimal list-size lower bound for random (and random linear) codes.

Our main result concerns an average-radius variant of list-decoding. This variant was implicitly used in [1, 12] en route their list-size lower bounds for standard list-decoding. In this work, we formally abstract this notion: a code is (p, L) *average-radius list-decodable* if for every L codewords, the *average* distance of their centroid from the L codewords exceeds pn. Note that this is a stronger requirement than (p, L) list-decodability where only the *maximum* distance from any center point to the L codewords must exceed pn.

We are able to prove nearly tight bounds on the achievable rate of a (p, L) average-radius list-decodable code. To state our result formally, denote by $L^{\mathsf{avg}}_{p,\gamma}$ the minimum L such that there exists a (p, L) average-radius list-decodable code family of rate $1 - h(p) - \gamma$. A simple random coding argument shows that a random code of $1 - h(p) - \gamma$ is $(p, 1/\gamma)$ average-radius list-decodable (matching the list-decodability of random codes). That is, $L^{\mathsf{avg}}_{p,\gamma} \leqslant 1/\gamma$. Our main technical result is a lower bound on the list-size that is polynomially related to the upper bound, namely $L^{\mathsf{avg}}_{p,\gamma} \geqslant \Omega_p(\gamma^{-1/2})$.

We remark that the classical Johnson bound in coding theory in fact proves the average-radius list-decodability of codes with good minimum distance — namely, a binary code of relative distance δ is $(J(\delta - \delta/L), L)$ average-radius list-decodable, where $J(z) = (1 - \sqrt{1 - 2z})/2$ for $z \in [0, \frac{1}{2}]$. (This follows from a direct inspection of the proof of the Johnson bound [11].) Also, one can show that if a binary code is $(\frac{1}{2} - 2^i\varepsilon, O(1/(2^{2i}\varepsilon^2))$ list-decodable for all $i = 0, 1, 2, \ldots$, then it is also $(\frac{1}{2} - 2\varepsilon, O(1/\varepsilon^2))$ average-radius list-decodable [5]. This shows that at least in the high noise regime, there is some reduction between these notions. Further, a suitable soft version of average-radius list-decodability can be used to construct matrices with a certain restricted isometry property [5]. For these reasons, we feel that average-radius list-decodability is a natural notion to study, even beyond treating it as a vehicle to understand (standard) list-decoding. In fact, somewhat surprisingly, one of our *constructions* of traditional list-decodable codes with a strong weight requirement on the codewords proceeds naturally via average-radius list-decodability; see Theorem 18 and the discussion following it for details.

1.3 Our other results

We also prove several other results that clarify the landscape of combinatorial limitations of list-decodable codes. Many results showing rate limitations in coding theory proceed via a typical approach in which they pass to a constant weight $\lambda \in (p, \frac{1}{2}]$; i.e., they restrict the codewords to be of weight exactly λn. They show that under this restriction, a code with the stated properties must have a constant number of codewords (that is, asymptotically *zero rate*). Mapping this bound back to the unrestricted setting one gets a rate upper bound of $1 - h(\lambda)$ for the original problem. For instance, the Elias-Bassalygo bound for rate R vs. relative distance δ is of this nature (here λ is picked to be the Johnson radius for list-decoding for codes of relative distance δ).

The above is also the approach taken in Blinovsky's work [1] as well as that of [12]. We show that such an approach does not and *cannot* give any bound better than Blinovsky's $\Omega_p(\log(1/\gamma))$ bound for $L_{p,\gamma}$. More precisely, for any $\lambda \geqslant p + 2^{-b_p L}$ for some $b_p > 0$, we show that there exists a (p, L) (average-radius) list-decodable code of rate $\Omega_{p,L}(1)$. Thus in order to improve the lower bound, we *must* be able to handle codes of strictly positive rate, and cannot deduce the bound by pinning down the zero-rate regime of constant-weight codes. This perhaps points to why improvements to Blinovsky's bounds have been difficult. On a positive note, we remark that we *are* able to effect such a proof for average-radius list-decoding (some details follow next).

To describe the method underlying our list-size lower bound for average-radius list-decoding, it is convenient to express the statement as an upper bound on rate in terms of list-size L. Note that a list-size lower bound of $L \geqslant \Omega_p(1/\sqrt{\gamma})$ for (p, L) average-radius list-decodable codes of rate $1 - h(p) - \gamma$ amounts to proving an upper bound of $1 - h(p) - \Omega_p(1/L^2)$ on the rate of (p, L) average-radius list-decodable codes. Our proof of such an upper bound proceeds by first showing a rate upper bound of $h(\lambda) - h(p) - a_p/L^2$ for such codes when the codewords are all restricted to all have weight λn, for a suitable choice of λ, namely $\lambda = p + a'_p/L$. To map this bound back to the original setting (with no weight restrictions on codewords), one simply notes that every (p, L) average-radius list-decodable code of rate R contains as a subcode, the translate of a constant λn-weight code of rate $R-(1-h(\lambda))$. (The second step uses a well-known argument.)

Generally speaking, by passing to a constant-weight subcode, one can translate combinatorial results on limitations of constant-weight codes to results showing limitations for the case of general codes. But this leaves open the possibility that the problem of showing limitations of constant-weight codes may be harder than the corresponding problem for general codes, or worse still, have a different answer making it impossible to solve the problem for general codes via the methodology of passing to constant-weight codes. We show that for the problem of list-decoding this is fortunately not the case, and there is, in fact, a "reverse connection" of the following form: A rate upper bound of $1 - h(p) - \gamma$ for (p, L) list-decodable codes implies a rate upper bound of $h(\lambda) - h(p) - \left(\frac{\lambda - p}{\frac{1}{2} - p}\right)\gamma$ for (p, L) list-decodable codes whose codewords must all have Hamming weight λn. A similar claim holds also for average-radius list-decodability, though we don't state it formally.

1.4 Our proof techniques

Our proofs in this paper employ variants of the standard probabilistic method. We show an extremely simple probabilistic argument that yields a $\Omega_p(\log(1/\gamma))$ bound on the list-size of a standard list-decodable code; we emphasize that this is qualitatively the tightest known bound in this regime.

For the "average-radius list-decoding" problem that we introduce, we are able to improve this list-size bound to $\Omega_p(1/\sqrt{\gamma})$. The proof is based on the idea that instead of picking the "bad list-decoding center" x uniformly at random, one can try to pick it randomly very close to a designated codeword c^*, and this still gives similar guarantees on the number of near-by codewords. Now since the quantity of interest is the average radius, this close-by codeword gives enough savings for us. In order to estimate the probability that a typical codeword c belongs to the list around x, we write this probability explicitly as a function of the Hamming distance between c^* and c, which is then lower bounded using properties of hypergeometric distributions and Taylor approximations for the binary entropy function.

For limitations of list-decoding random codes, we define a random variable W that counts the number of "violations" of the list-decoding property of the code. We then show that W has a exponentially large mean, around which it is concentrated w.h.p. This yields that the code cannot be list-decodable with high probability, for suitable values of rate and list-size parameters.

1.5 Organization

We define some useful notation and the formal notion of average-radius list-decodability in Section 2. Our main list-size lower bound for average-radius list-decoding appears in Section 3. We give our short proof of Blinovsky's lower bounds for binary and general alphabets in Section 4. Our results about the zero-error rate regime for constant-weight codes and the reverse connection between list-decoding bounds for general codes and constant-weight codes appear in Section 5. Finally, our list-size lower bounds for random codes are stated in Section 6; for reasons of space, the proofs for these bounds appear in the appendix.

2 Preliminaries and notation

2.1 List decoding

We recall some standard terminology regarding error-correcting codes.

Let $[n]$ denote the index set $\{1, 2, \ldots, n\}$. For $q \geqslant 2$, let $[q]$ denote the set $\{0, 1, \ldots, q-1\}$. A *q-ary code* refers to any subset $C \subseteq [q]^n$, where n is the *blocklength* of C. We will mainly focus on the special case of binary codes corresponding to $q = 2$. The rate $R = R(C)$ is defined to be $\frac{\log|C|}{n \log q}$. For $x \in [q]^n$ and $S \subseteq [n]$, the restriction of x to the coordinates in S is denoted $x|_S$. Let $\mathrm{Supp}(x) := \{i \in [n] \ : \ x_i \neq 0\}$. A *subcode* of C is a subset C' of C. We say that C is a *constant-weight code* with weight $w \in [0, n]$, if all its codewords have weight exactly w. (Such codes are studied in Section 5.)

For $x, y \in [q]^n$, define the *Hamming distance* between x and y, denoted $d(x, y)$, to be the number of coordinates in which x and y differ. The *(Hamming) weight* of x, denoted $\mathrm{wt}(x)$, is $d(\mathbf{0}, x)$, where $\mathbf{0}$ is the vector in $[q]^n$ with zeroes in all coordinates. The *(Hamming) ball* of radius r centered at x, denoted $\mathbf{B}(x, r)$, is the set $\{y \in [q]^n \ : \ d(x, y) \leqslant r\}$. In this paper, we also need the following notions of distance of a (small) "list" $\mathcal{L}$ of vectors from a "center" x:

Definition 1. *Given a center $x \in [q]^n$ and a nonempty list $\mathcal{L} \subseteq [q]^n$, define the maximum and average distances of $\mathcal{L}$ from x respectively by:*

$$D_{\mathit{max}}(x, \mathcal{L}) := \max\{d(x, c) \ : \ c \in \mathcal{L}\}, \text{ and}$$
$$D_{\mathit{avg}}(x, \mathcal{L}) := \mathbf{E}_{c \in \mathcal{L}} \left[d(x, c) \right] = \frac{1}{|\mathcal{L}|} \sum_{c \in \mathcal{L}} d(x, c).$$

It is well-known (cf., e.g., Lemma 5 in [12]) that the average-radius of a list is minimized by the *coordinatewise majority* (or *centroid*) of the list:

Fact 2. *Let $\mathcal{L} = \{c_1, c_2, \ldots, c_L\} \subseteq \{0,1\}^n$ be an arbitrary list of codewords, and let $a \in \{0,1\}^n$ be its* centroid*; that is, for any coordinate j, the j^{th} entry of a is the majority of the corresponding entries of*

$c_1, c_2, \ldots, c_L$ *(breaking ties arbitrarily). Then*

$$D_{\textsf{avg}}(a, \mathcal{L}) = \min_{a' \in \{0,1\}^n} D_{\textsf{avg}}(a', \mathcal{L}).$$

Next, we formalize the error recovery capability of the code using list-decoding.

Definition 3. *Fix $0 < p < \frac{1}{2}$ and a positive integer L. Let C be a q-ary code with blocklength n.*

1. *C is said to be (p, L)* list-decodable *if for all $x \in [q]^n$, the ball $\mathbf{B}(x, pn)$ contains at most $L - 1$ codewords of C. Equivalently, for any x and any list $\mathcal{L} \subseteq C$ of size at least L, we have $D_{\textsf{max}}(x, \mathcal{L}) > pn$.*
2. *C is said to be (p, L)* average-radius list-decodable *if for any center x and any L-tuple $\mathcal{L}$ of codewords, we have $D_{\textsf{avg}}(x, \mathcal{L}) > pn$.*

For constant-weight codes, it is convenient to augment the above notation with the weight parameter:

Definition 4. *Let p, L, q, n, C be as in Definition 3, and let $0 < \lambda \leqslant \frac{1}{2}$. C is said to be $(\lambda; p, L)$* (average-radius) list-decodable *if C is (p, L) (average-radius) list-decodable, and every codeword in C has weight exactly λn.*

We remark that the list-decodability property is standard in literature. Moreover, while the notion of average-radius list-decodability is formally introduced by this paper, it is already implicit in [1, 2, 12]. The following proposition asserts that this is a syntactically stronger property than standard list-decodability:

Proposition 5. *If C is (p, L) average-radius list-decodable, then C is (p, L) list-decodable.*

Proof: The claim follows from the observation that the maximum distance of a list from a center x always dominates its average distance from x. $\square$

In particular, any limitation we establish for list-decodable codes also carries over for average-radius list-decodable codes.

Following (and extending) the notation in [9], we make the following definitions to quantify the tradeoffs in the different parameters of a code: the rate R, the error-correction radius p, the list-size L, and the weight λ of the codewords (for "constant weight" codes). Further, for general codes (without the constant-weight restriction), it is usually more convenient to replace the rate R by the parameter $\gamma := 1 - h(p) - R$; this measures the "gap" to the "limiting rate" or the "capacity" of $1 - h(p)$ for $(p, O(1))$ list-decodable codes.

Fix $p, \lambda \in (0, \frac{1}{2}]$ such that $p < \lambda$, $0 \leqslant R \leqslant 1$, and a positive integer L.

Definition 6.

1. *Say that the triple $(p, L; R)$ is* achievable for list-decodable codes *if there exist (p, L) list-decodable codes of rate R for infinitely many lengths n.*

   Define $R_{p,L}$ to be the supremum over R such that $(p, L; R)$ is achievable for list-decodable codes, and define $\gamma_{p,L} := 1 - h(p) - R_{p,L}$. Similarly, define $L_{p,\gamma}$ to be the least integer L such that $(p, L; 1 - h(p) - \gamma)$ is achievable.
2. *(***For constant weight codes.***) Say that the 4-tuple $(\lambda; p, L; R)$ is achievable if there exists $(\lambda; p, L)$ list-decodable codes of rate R. Define $R_{p,L}(\lambda)$ to be the supremum rate R for which the 4-tuple $(\lambda; p, L; R)$ is achievable.*

We can also define analogous quantities for average-radius list-decoding (denoted by a superscript $\textsf{avg}$), but to prevent notational clutter, we will not explicitly do so. Throughout this paper, p is treated as a fixed constant in $(0, \frac{1}{2})$, and we will not attempt to optimize the dependence of our bounds on p.

2.2 Standard distributions and functions

In this paper, we use 'log' for logarithms to base 2 and 'ln' for natural logarithms. Also, to avoid cumbersome notation, we often denote b^z by $\exp_b(z)$. Standard asymptotic notations (O, o, and Ω) is employed throughout this paper; we sometimes subscript this notation by a parameter (typically p) to mean that the hidden constant could depend arbitrarily on the parameter.

Our proofs make a heavy of *hypergeometric distributions*, which we review here for the sake of completeness as well as to set the notation. Suppose a set contains n objects, exactly $m < n$ of which are *m*arked, and suppose we *s*ample $s < n$ objects uniformly at random from the set *without replacement*. Let the random variable T count the number of marked objects in the sample; then T follows the hypergeometric distribution with parameters (n, m, s). A simple counting argument shows that, for $t \leqslant \min\{m, s\}$,

$$\mathbf{Pr}[T = t] = \frac{\binom{m}{t}\binom{n-m}{s-t}}{\binom{n}{s}}.$$

We will denote the above expression by $f(n, m, s, t)$. By convention, $f(n, m, s, t)$ is set to 0 if $n < \max\{m, s\}$ or $t > \min\{m, s\}$.

Our proofs rely on the following two properties of hypergeometric random variables. While these claims are standard, we have included a proof in Appendix A.1 for completeness.

Fact 7 (Interchange property)**.** *For all integers n, m, s with $n \geqslant \max\{m, s\}$, the hypergeometric distribution with parameters (n, m, s) is identical to that with parameters (n, s, m). That is, for all t, we have $f(n, m, s, t) = f(n, s, m, t)$.*

Fact 8. *Suppose n, m, m', s are integers such that $m \geqslant m'$ and $n \geqslant \max\{m, s\}$. Then the hypergeometric distribution with parameters (n, m, s)* stochastically dominates *the hypergeometric distribution with parameters (n, m', s). That is, for all τ, we have*

$$\sum_{t=\tau}^{\infty} f(n, m, s, t) \geqslant \sum_{t=\tau}^{\infty} f(n, m', s, t).$$

Throughout this paper, we are especially concerned with the asymptotic behaviour of binomial coefficients, which is characterized in terms of the binary entropy function, defined as $h(z) := -z \log z - (1 - z) \log(1 - z)$. We will use the following standard estimate without proof.

Fact 9. *Fix $z \in (0, 1)$, and suppose $n \to \infty$ such that zn is an integer. Then*

$$\exp_2(h(z)n - o(n)) \leqslant \binom{n}{zn} \leqslant \sum_{i=0}^{zn} \binom{n}{i} \leqslant \exp_2(h(z)n).$$

3 Bounds for average-radius list-decodability

In this section, we prove that the largest asymptotic rate of (p, L) average-radius list-decodable binary codes is bounded by

$$1 - h(p) - \frac{1}{L} - o(1) \leqslant R_{p,L} \leqslant 1 - h(p) - \frac{a_p}{L^2} + o(1),$$

where a_p is a constant depending only on p. (Here p is a fixed constant bounded away from 0 and $\frac{1}{2}$.) Note that the corresponding upper and lower bounds on $\gamma := 1 - h(p) - R$ are polynomially related, ignoring the dependence on p.

We first state the rate lower bound.

Theorem 10. *Fix $p \in (0, \frac{1}{2})$ and a positive integer L. Then, for all $\varepsilon > 0$ and all sufficiently large lengths n, there exists a (p, L) average-radius list-decodable code of rate at least $1 - h(p) - 1/L - \varepsilon$.*

Proof: We will show that a random code of the desired rate is (p, L) average-radius list-decodable w.h.p. Consider a random code $C : \{0,1\}^{Rn} \to \{0,1\}^n$ of rate $R := 1-h(p)-1/L-\varepsilon$; i.e., for each $x \in \{0,1\}^{Rn}$, we pick $C(x)$ independently and uniformly at random from $\{0,1\}^n$. For any $a \in \{0,1\}^n$ and any distinct L-tuple $\{x_1, \ldots, x_L\} \subseteq \{0,1\}^{Rn}$, we are interested in bounding the probability of the event that $D \leqslant Lpn$, where $D := \sum_{i=1}^{L} d(a, C(x_i))$.

To estimate this probability, let A be the $\{0,1\}$-string of length Ln obtained by concatenating a repeatedly L times. Similarly, let Y be the $\{0,1\}$-string obtained by concatenating $C(x_1), \ldots, C(x_L)$. In this notation, note that D is simply the Hamming distance between A and Y. Now, Y is distributed uniformly at random in $\{0,1\}^{Ln}$ independently of the choice of A, hence the probability that $D \leqslant pLn$ is at most $\exp_2((h(p)-1)Ln)$ (Fact 9).

Finally, by a union bound over the choice of a and $\{x_1, \ldots, x_L\}$, the code fails to be (p, L) average-radius list-decodable with probability at most

$$2^n \binom{2^{Rn}}{L} \cdot \exp_2((h(p)-1)Ln) \leqslant \exp_2(n + (R + h(p) - 1)Ln) = \exp_2(-\varepsilon Ln),$$

for the given choice of R, establishing the claim. □

We now show an upper bound of $1 - h(p) - a_p/L^2$ on the rate of a (p, L) average-radius list-decodable code. As stated in the Introduction, the main idea behind the construction is that instead of picking the "bad list decoding center" x uniformly at random, we pick it randomly *very close to a designated codeword c^** (which itself is a uniformly random element from C). Now as long as we are guaranteed to find a list of $L-1$ other codewords near the center, we can include c^* in our list to lower its average radius.

However formalizing the above intuition into a proof is nontrivial, since our restriction of the center x to be very close to c^* introduces statistical dependencies while analyzing the number of codewords near x. We are able to control these dependencies, but this requires some heavy calculations involving hypergeometric distributions and the entropy function.

We are now ready to state our main result establishing a rate upper bound for (p, L) average-radius list-decodable codes. In fact, the bulk of the work is to show an analogous upper bound for the special case of a constant-weight code C, i.e., all codewords have weight exactly λn, for some $\lambda \in (p, \frac{1}{2})$. We can then map this bound for general codes using a standard argument (given in Lemma 12).

Theorem 11 (Main theorem)**.** *Fix $p \in (0, \frac{1}{2})$, and let L be a sufficiently large positive integer. Then there exist $a_p, a'_p > 0$ (depending only on p) such that the following holds (for sufficiently large lengths n):*

1. *If C is a (p, L) average-radius list-decodable code, then C has rate at most $1 - h(p) - a_p/L^2 + o(1)$.*
2. *For $\lambda := p + a'_p/L$, if C is a $(\lambda; p, L)$ average-radius list-decodable code, then C has rate at most $h(\lambda) - h(p) - a_p/L^2 + o(1)$.*

As already mentioned in Section 1.3, the second claim readily implies the first via the following well-known argument (a partial converse to this statement for list-decoding will be given in Section 5):

Lemma 12. *Let $\lambda \in (p, \frac{1}{2}]$ be such that λn is an integer. If C is a (p, L) average-radius list-decodable code of rate $R = 1 - h(p) - \gamma$, then there exists a $(\lambda; p, L)$ average-radius list-decodable code of rate at least $h(\lambda) - h(p) - \gamma - o(1)$.*

Proof: For a random center x, the expected number of codewords $c \in C$ with $d(x, c) = \lambda n$ is exactly $|C| \cdot \binom{n}{\lambda n} 2^{-n}$. For the assumed value of rate R, using Fact 9, this is at least

$$\exp_2((h(\lambda) - h(p) - \gamma - o(1))n).$$

Then there exists an x such that the subcode $C' \subseteq C$ consisting of all codewords at a distance λn from x has rate at least $h(\lambda) - h(p) - \gamma - o(1)$. The claim follows by translating C' by $-x$. $\square$

Before we proceed to the proof of the first part of Theorem 11, we will establish the following folklore result, whose proof illustrates our idea in a simple case.

Lemma 13 (A warm-up lemma)**.** *Fix p, λ so that $p < \lambda \leqslant \frac{1}{2}$. Then, if C is a $(\lambda; p, L)$ list-decodable code, then C has rate at most $h(\lambda) - h(p) + o(1)$.*

Proof: The main idea behind the proof is that a random center of a *particular weight* (carefully chosen) is close to a large number of codewords in expectation. Pick a random subset $S \subseteq [n]$ of coordinates of size αn, with $\alpha := (\lambda - p)/(1 - 2p)$, and let $\overline{S} := [n] \setminus S$. (The motivation for this choice of α will be clear shortly.) Define the center x be the *indicator vector* of S; i.e., $\mathrm{Supp}(x) = S$.

Consider the set $\mathcal{L}$ of codewords $c \in C$ such that $\mathrm{wt}(c|_S) \geqslant (1-p)\alpha n$; this is our candidate bad list of codewords. Then each $c \in \mathcal{L}$ is close to c:

$$d(x, c) = (\alpha n - \mathrm{wt}(c|_S)) + \mathrm{wt}(c|_{\overline{S}}) \leqslant \alpha p n + (\lambda - \alpha(1-p))n = (\lambda - \alpha(1-2p))n,$$

which equals pn for the given choice of α. Hence the size of $\mathcal{L}$ is a lower bound on the list-size of the code.

We complete the proof by computing $\mathbf{E}\,|\mathcal{L}|$. For any fixed $c \in C$, the random variable $\mathrm{wt}(c|_S)$ follows the hypergeometric distribution with parameters $(n, \lambda n, \alpha n)$, which is identical to the hypergeometric distribution with parameters $(n, \alpha n, \lambda n)$ (see Fact 7). Hence the probability that c is included in the list $\mathcal{L}$ is at least

$$f(n, \alpha n, \lambda n, \alpha(1-p)n) := \frac{\binom{\alpha n}{(1-p)\alpha n}\binom{(1-\alpha)n}{(\lambda - \alpha(1-p))n}}{\binom{n}{\lambda n}} = \frac{\binom{\alpha n}{p\alpha n}\binom{(1-\alpha)n}{p(1-\alpha)n}}{\binom{n}{\lambda n}},$$

where the second step uses the identity $\lambda - (1-p)\alpha = p(1-\alpha)$, which holds for our particular choice of α. As $n \to \infty$, this is equal to

$$\exp_2\left(\alpha n h(p) + (1-\alpha) n h(p) - h(\lambda) n - o(n)\right) = \exp_2((h(p) - h(\lambda) - o(1))n).$$

Thus, by linearity of expectations, the expected size of $\mathcal{L}$ is at least $|C| \cdot \exp_2((h(p) - h(\lambda) - o(1))n)$. On the other hand, the (p, L) list-decodability of C says that $|\mathcal{L}| \leqslant L$ (with probability 1). Comparing these lower and upper bounds on $\mathbf{E}\,|\mathcal{L}|$ yields the claim. $\square$

Proof of Theorem 11 (part 2): At a high level, we proceed as in the proof of Lemma 13, but in addition to the bad list $\mathcal{L}$ of codewords, we will a special codeword $c^* \in C$ such that $d(x, c^*)$ is *much smaller than the codewords in* $\mathcal{L}$. Then defining $\mathcal{L}^*$ to consist of c^* and $(L-1)$ other codewords from $\mathcal{L}$, we see that the *average* distance of $\mathcal{L}^*$ is much smaller than before, thus enabling us to obtain an improved rate bound.

We now provide the details. Pick a uniformly random codeword $c^* \in C$. Let $S \subseteq [n]$ be a random subset of $\mathrm{Supp}(c^*)$ of size βn, where the parameter β is chosen appropriately later[1] (this plays the role of α in Lemma 13). Also, let x be the indicator vector of S.

As before, consider the set $\mathcal{L}$ of codewords $c \in C$ such that $\mathrm{wt}(c|_S) \geqslant (1-p)|S|$. For a fixed $c \in C$, the random variable $\mathrm{wt}(c|_S)$ follows the hypergeometric distribution with parameters $(\lambda n, (\lambda - \delta)n, \beta n)$, where $\delta = \delta(c^*, c)$ is defined by $d(c^*, c) := 2\delta n$. (Observe that the normalization ensures that $0 \leqslant \delta \leqslant \lambda$ for all pairs $c^*, c \in C$.) To see this, notice that we are sampling βn coordinates from $\mathrm{Supp}(c^*)$ without replacement, and that $\mathrm{wt}(c|_S)$ simply counts the number of coordinates picked from $\mathrm{Supp}(c^*) \cap \mathrm{Supp}(c)$

[1] At this point, the reader might find it useful to think of both $\lambda - p$ and β as $\Theta(1/L)$; roughly speaking, this setting translates to a rate upper bound of $h(\lambda) - h(p) - \Omega(\beta/L)$.

(the size of this intersection is exactly $(\lambda-\delta)n$). Thus, conditioned on c^*, the probability that a fixed $c \in C$ is included in $\mathcal{L}$ is

$$Q(\delta) := \sum_{w=(1-p)\beta n}^{\beta n} f(\lambda n, (\lambda-\delta)n, \beta n, w). \tag{1}$$

By linearity of expectations, and taking expectations over c^*, the expected size of $\mathcal{L}$ can be written as

$$\mathbf{E}_{c^* \in C}\left[\sum_{c\in C} Q(\delta(c^*, c))\right] = |C| \cdot \mathbf{E}\, Q(\delta(c^*, c)), \tag{2}$$

where *both* c^* and c are picked uniformly at random from C. The following lemma provides a lower bound on this expectation.

Lemma 14. *For $A_1 := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$ and $A_2 = \frac{2}{p^2}$, we have*

$$\mathbf{E}\, Q(\delta(c^*, c)) \geqslant \exp_2\left(-(A_1\beta + A_2\beta^2 + o(1))n\right),$$

where the expectation is taken over pairs c^*, c of codewords.

Remark. In the above estimate, the coefficient A_1 is tight for all values of p and λ (assuming $\beta \to 0$ keeping p and λ fixed), but A_2 can be improved significantly. For our purposes, it suffices that A_2 depends on p alone, and not on λ or β. $\square$

Proof: By a standard application of the Cauchy-Schwarz inequality, we can show that $\mathbf{E}\,\delta \leqslant \lambda(1-\lambda)$. To see this, let f_j denote the fraction of codewords of C that have 1 in the jth coordinate. The weight constraint on the codewords implies that $\sum_{j=1}^{n} f_j = \lambda n$. Therefore,

$$\begin{aligned}\mathbf{E}_{c^*,c}\ [d(c^*, c)] &= \sum_{j=1}^{n} 2f_j(1-f_j) = 2\sum_{j=1}^{n} f_j - 2\sum_{j=1}^{n} f_j^2 \\ &\leqslant 2\sum_{j=1}^{n} f_j - \frac{2}{n}\left(\sum_{j=1}^{n} f_j\right)^2 = 2\lambda n - 2\lambda^2 n,\end{aligned}$$

and so, $\mathbf{E}\,\delta \leqslant \lambda(1-\lambda)$. Now, by Markov's inequality, the probability that $\delta \leqslant \lambda(1-\lambda) + 1/n$ is at least $1 - \frac{\lambda(1-\lambda)}{\lambda(1-\lambda)+\frac{1}{n}} \geqslant \frac{1}{n}$.

Moreover, using Fact 8 (with $\tau := \beta(1-p)$), we know that $Q(\delta)$ is a monotonically decreasing function of δ. Therefore,

$$\begin{aligned}\mathbf{E}\, Q(\delta(c^*, c)) &\geqslant \frac{1}{n} \cdot Q(\lambda(1-\lambda) + o(1)) \\ &\geqslant \frac{1}{n} \cdot f\left(\lambda n, (\lambda^2 - o(1))n, \beta n, \beta(1-p)n\right).\end{aligned}$$

The rest of the proof consists of lower bounding the right hand side. As $n \to \infty$, using Fact 9, we get $\mathbf{E}\, Q(\delta) \geqslant \exp_2(\varepsilon n - o(n))$, where

$$\varepsilon := \lambda^2 \cdot h\left(\frac{(1-p)\beta}{\lambda^2}\right) + \lambda(1-\lambda) \cdot h\left(\frac{p\beta}{\lambda(1-\lambda)}\right) - \lambda \cdot h\left(\frac{\beta}{\lambda}\right).$$

We are interested in lower bounding the exponent ε, and we do this by bounding each of the above entropy terms individually using Fact 24 (see Appendix A.2), and canceling common terms. We just mention the final bound ignoring the intermediate steps:

$$\varepsilon \geqslant \beta \left((1-p) \log \frac{\lambda^2}{1-p} + p \log \frac{\lambda(1-\lambda)}{p} - \log \lambda \right) - \beta^2 (\log e) \left(\frac{(1-p)^2}{\lambda^2} + \frac{p^2}{\lambda(1-\lambda)} \right).$$

Noting that

$$(1-p) \log \frac{\lambda^2}{1-p} + p \log \frac{\lambda(1-\lambda)}{p} - \log \lambda = (1-p) \log \frac{\lambda}{1-p} + p \log \frac{1-\lambda}{p} = -A_1,$$

and

$$(\log e) \left(\frac{(1-p)^2}{\lambda^2} + \frac{p^2}{\lambda(1-\lambda)} \right) \leqslant (\log e) \left(\frac{1}{p^2} + 1 \right) \leqslant \frac{2}{p^2},$$

we get the claim. $\square$

We now return to the proof of Theorem 11. From (2) and Lemma 14, if the code C has rate at least $A_1\beta + A_2\beta^2 + o(1)$ (for a suitable $o(1)$ term), the list $\mathcal{L}$ has size at least L in expectation. Fix some choice of c^* and S such that $|\mathcal{L}| \geqslant L$. Let $\mathcal{L}^*$ be any list containing c^* and $L-1$ other codewords from $\mathcal{L}$; we are interested in $D_{\mathsf{avg}}(x, \mathcal{L}^*)$. Clearly, $d(x, c^*) = (\lambda - \beta)n$. On the other hand, for $c \in \mathcal{L}^* \setminus \{c^*\}$, we can bound its distance from x as: $d(x, c) \leqslant \beta p n + (\lambda - \beta(1-p))n = (\lambda - \beta(1-2p))n$, where the two terms are respectively the contribution by S and $[n] \setminus S$. Averaging these L distances, we get that

$$D_{\mathsf{avg}}(x, \mathcal{L}^*) \leqslant (\lambda - \beta(1 - 2p + 2p/L))\, n.$$

Now, we pick β so that this expression is at most pn; i.e., set

$$\beta := \frac{\lambda - p}{1 - 2p + 2p/L}. \tag{3}$$

(Compare with the choice of α in Lemma 13.) For this choice of β, the list $\mathcal{L}^*$ violates the average-radius list-decodability property of C.

Thus the rate of a (p, L) average-radius list-decodable code is upper bounded by $R \leqslant A_1\beta + A_2\beta^2 + o(1)$, where β is given by (3). Further technical manipulations brings this to the following more convenient form: If $L > \frac{2p}{1-2p}$, then

$$R \leqslant (h(\lambda) - h(p)) - \frac{B_1(\lambda - p)}{L} + B_2(\lambda - p)^2 + o(1).$$

for $B_1 := \frac{1}{2}p$ and $B_2 := \frac{3}{p^2(1-2p)^2}$; see Lemma 26 in Appendix A.2 for a proof. Note that the second term dominates the third whenever $\lambda - p$ is small enough. In particular, for

$$\lambda := p + \frac{B_1}{2B_2 L} = p + \frac{p^3(1-2p)^2}{12L},$$

the rate is upper bounded by

$$R \leqslant h(\lambda) - h(p) - \frac{B_1^2}{4B_2 L^2} + o(1) = h(\lambda) - h(p) - \frac{p^4(1-2p)^2}{48L^2} + o(1).$$

$\square$

4 Bounds for (standard) list-decodability

In this section, we consider the rate vs. list-size tradeoff for the traditional list-decodability notion. For the special case when the fraction of errors is close to $\frac{1}{2}$, [12] showed that any code family of growing size correcting up to $\frac{1}{2} - \varepsilon$ fraction of errors must have a list-size $\Omega(1/\varepsilon^2)$, which is optimal up to constant factors. When p is bounded away from $1/2$, Blinovsky [1, 3] gives the best known bounds on the rate of a (p, L) list-decodable code. His results imply (see [14] for the calculations) that any (p, L) list-decodable code of rate $1 - h(p) - \gamma$ has list-size L at least $\Omega_p(\log(1/\gamma))$. We give a short and simple proof of this latter claim in this section.

Theorem 15 ([1, 3])**.** *1. Suppose C is $(\lambda; p, L)$ list-decodable code with $\lambda = p + \frac{1}{2}p^L$. Then $|C| \leqslant 2L^2/p$, independent of its blocklength n. (In particular, the rate approaches 0 as $n \to \infty$.)*

2. Any (p, L) list-decodable code has rate at most $1 - h(p) - \Omega_p(p^L)$.

Proof:

1. For the sake of contradiction, assume that $|C| > 2L^2/p$. Pick a random L-tuple of codewords (without replacement) $\mathcal{L} = \{c_1, c_2, \ldots, c_L\}$, and let S be the set of indices $i \in [n]$ such that each $c_j \in \mathcal{L}$ has 1 in the ith coordinate. Define x to be the indicator vector of S. Note that $d(x, c_j) = \mathrm{wt}(c_j) - \mathrm{wt}(x) = \lambda n - |S|$. So $D_{\mathsf{max}}(x, \mathcal{L})$ is also $\lambda n - |S|$, and hence, $\mathbf{E}\, D_{\mathsf{max}}(x, \mathcal{L}) = \lambda n - \mathbf{E}\,|S|$. Thus to obtain a contradiction, it suffices to show that $\mathbf{E}\,|S| \geqslant (\lambda - p)n = \frac{1}{2}p^L n$.

   Let $M := |C|$ be the total number of codewords of C, and let M_i be the number of codewords of C with 1 in the i^{th} position. Then the probability that $i \in S$ is equal to $g(M_i)/\binom{M}{L}$, where the function $g : \mathbb{R}^{\geqslant 0} \to \mathbb{R}^{\geqslant 0}$ is defined by $g(z) := \binom{\max\{z, L-1\}}{L}$. By standard closure properties of convex functions, g is convex on $\mathbb{R}$. (Specifically, $z \mapsto \max\{z, L-1\}$ is convex over $\mathbb{R}$, and restricted to its image, namely, the interval $[L-1, \infty)$, the function $z \mapsto \binom{z}{L}$ is convex. Hence their composition, namely g, is convex as well.)

   We are now ready to bound $\mathbf{E}\,|S|$:

$$\frac{1}{n}\mathbf{E}\,|S| \overset{(a)}{=} \frac{1}{\binom{M}{L}} \cdot \frac{1}{n}\sum_{i=1}^{n} g(M_i) \overset{(b)}{\geqslant} \frac{1}{\binom{M}{L}} \cdot g\left(\frac{1}{n}\sum_{i=1}^{n} M_i\right) = \frac{g(\lambda M)}{\binom{M}{L}} \overset{(c)}{=} \frac{\binom{\lambda M}{L}}{\binom{M}{L}}.$$

   Here we have used (a) the linearity of expectations, (b) Jensen's inequality, and (c) the fact that $\lambda M \geqslant 2L^2 \geqslant L - 1$. We complete the proof using a straightforward approximation of the binomial coefficients.

$$\frac{1}{n}\mathbf{E}\,|S| \geqslant \frac{(\lambda M - L)^L}{M^L} = \lambda^L \left(1 - \frac{L}{\lambda M}\right)^L \geqslant \lambda^L \left(1 - \frac{L^2}{\lambda M}\right) \geqslant \frac{1}{2}\lambda^L \geqslant \frac{1}{2}p^L.$$

2. By Lemma 12, the rate of a *general* (p, L) list-decodable code is upper bounded by $1 - h\left(p + \frac{1}{2}p^L\right) + o(1)$, which, by Fact 23 (see Section A in the Appendix), is at most $1 - h(p) - \frac{1}{4}(1 - 2p) \cdot p^L + o(1)$.

□

The above method can be adapted for q-ary codes with an additional trick:

Theorem 16. *1. Suppose C is a q-ary $(\lambda; p, L)$ list-decodable code with $\lambda = p + \frac{1}{2L}p^L$. Then $|C| \leqslant 2L^2/\lambda$.*

2. Suppose C is a q-ary (p, L) list-decodable code. Then there exists a constant $b = b_{p,q} > 0$ such that the rate of C is at most $1 - h_q(p) - \Omega_{q,p}\left(\frac{1}{L}p^L\right)$.

Our proof of this theorem uses the following lemma due to Erdös (see Section 2.1 of [13] for a reference.) This result was implicitly established in our proof of Theorem 15, so we will omit a formal proof.

Lemma 17 (Erdös 1964)**.** *Suppose $\mathcal{A}$ is a set system over the ground set $[n]$, such that each $A \in \mathcal{A}$ has size at least λn. Then, if $|\mathcal{A}| \geqslant 2L^2/\lambda$, then there exist distinct $A_1, A_2, \ldots, A_L$ in $\mathcal{A}$ such that $\bigcap_{i=1}^{L} A_i$ has size at least $\frac{1}{2}n\lambda^L$.*

Proof of Theorem 16:

1. Towards a contradiction, assume $|C| \geqslant 2L^2/\lambda$. Consider the set system $\mathcal{A} = \{\mathrm{Supp}(c) \ : \ c \in C\}$. By Lemma 17, there exists an L-tuple $\{c_1, c_2, \ldots, c_L\}$ of codewords such that the intersection of their support, say S, has size at least $\frac{1}{2}n\lambda^L \geqslant \frac{1}{2}np^L$. Arbitrarily partition the coordinates in S into L parts, say $S_1, \ldots, S_L$ so that each S_j has size at least $\frac{1}{2L}p^L n$.

   Now consider a center x such that x agrees with c_j on all coordinates $i \in S_j$; for $i \notin S$, set x_i to be zero. Then, clearly, $d(x, c_j) \leqslant \mathrm{wt}(c_j) - |S_j| = \lambda n - \frac{1}{2L}p^L \cdot n = pn$. Thus the list $\{c_1, \ldots, c_L\}$ contradicts the (p, L) list-decodability of C.

2. From a q-ary generalization of Lemma 12 (proof omitted), the rate of a (p, L) q-ary list-decodable code is at least $1 - h_q\left(p + \frac{1}{2L}p^L\right)$. For L large enough, this is at most $1 - h_q(p) - \Omega_{q,p}\left(\frac{1}{2L}p^L\right)$, which implies the claim.

$\square$

5 Constant-weight vs. General codes

In this section, we will understand the rate vs. list-size trade-offs for constant-weight codes, that is, codes with every codeword having weight λn, where $\lambda \in (p, \frac{1}{2}]$ is a parameter. (Setting $\lambda = \frac{1}{2}$ roughly corresponds to arbitrary codes having no weight restrictions.) As observed earlier, a typical approach in coding theory to establish rate upper bounds is to study the problem under the above constant-weight restriction. One then proceeds to show a strong negative result of the flavor that a code with the stated properties must have a constant size (and in particular *zero* rate). For instance, the first part of Theorem 15 above is of this form. Finally, mapping this bound to arbitrary codes, one obtains a rate upper bound of $1 - h(\lambda)$ for the original problem. (Note that Lemma 12 provides a particular formal example of the last step.)

In particular, Blinovsky's rate upper bound (Theorem 15) of $1 - h(p) - 2^{-O(L)}$ for (p, L) list-decodable codes follows this approach.[2] More precisely, he proves that, under the weight-λ restriction, such code must have zero rate for all $\lambda \leqslant p + 2^{-b_p L}$ for some $b_p < \infty$. One may then imagine improving the rate upper bound to $1 - h(p) - L^{-O(1)}$ *simply by* establishing the latter result for correspondingly higher values of λ (i.e., up to $p + L^{-O(1)}$). We show that this approach cannot work by establishing that (average-radius) list-decodable codes of positive (though possibly small) rates exist as long as $\lambda - p \geqslant 2^{-O(L)}$. Thus Blinovsky's result identifies the correct *zero-rate regime* for the list-decoding problem; in particular, his bound is also the best possible if we restrict ourselves to this approach. In this context, it is also worth noting that for average-radius list-decodable codes, Theorem 11 already provides a better rate upper bound than what the zero-rate regime indicates, thus suggesting that the "zero-rate regime barrier" is not an inherent obstacle, but more a limitation of the current proof techniques.

In the opposite direction, we show that the task of establishing rate upper bounds for constant weight codes is not significantly harder than the general problem. Formally, we state that that if the "gap to list-decoding capacity" for general codes is γ, then the gap to capacity for weight-λn codes is *at least* $\left(\frac{\lambda - p}{\frac{1}{2} - p}\right)\gamma$.

[2] For notational ease, we suppress the dependence on p in the O and Ω notations in this informal discussion.

Stated differently, if our goal is to establish a $L^{-O(1)}$ lower bound on the gap γ, then we do not lose by first passing to a suitable λ (that is not too close to p).

5.1 Zero-rate regime

Theorem 18. *Fix $p \in (0, \frac{1}{2})$, and set $b = b_p := \frac{1}{2}\left(\frac{1}{2} - p\right)^2$. Then for $L \geqslant \frac{1}{2b}\log\left(\frac{32}{b}\right)$ and all sufficiently large n, there exists a $(\lambda; p, L)$ average-radius list-decodable code of rate at least $R - o(1)$, with $\lambda \leqslant p + 5e^{-bL}$ and $R := \min\{e^{-2bL}, e^{-bL}/(6L)\} = \Omega_{p,L}(1)$.*

Proof: The basic idea of the proof is that a random code is (p, L) average-radius list-decodable, even if the codewords are biased to have weight close to pn. We then use expurgation to ensure that all codewords have the same weight. We now provide the details. Set $\varepsilon := e^{-bL}$ and $\lambda' := p + 4\varepsilon$; verify that for the assumed values of L, we have $\frac{1}{2} - \lambda' \geqslant \frac{1}{2}\left(\frac{1}{2} - p\right)$. Choose a random code $C : \{0,1\}^{Rn} \to \{0,1\}^n$ in the following way. For each $x \in \{0,1\}^{Rn}$, each coordinate of $C(x)$ chosen independently to be 1 with probability λ' (and 0 with the complementary probability).

Firstly, for a fixed $x \in \{0,1\}^{Rn}$, by Chernoff bound, its encoding $C(x)$ has weight in the interval $(\lambda' \pm \varepsilon)n$ with probability at least $1 - 2\exp(-2\varepsilon^2 n) \geqslant 1 - \exp_2(-2\varepsilon^2 n + o(n))$. By union bound, this holds for all x with probability at least $1 - \exp_2(Rn - 2\varepsilon^2 n + o(n))$.

Next, we consider the event that C is (p, L) average-radius list-decodable. Specifically, we require that for every L-tuple of messages $X := \{x_1, \ldots, x_L\} \subseteq \{0,1\}^{Rn}$ and every center $a \in \{0,1\}^n$, the encodings of the x_is are pn-far from a on average. It is easy to bound the probability of this event for a *fixed* pair (a, X), and naively, we might hope to achieve this for *all* such pairs by a simple union bound. However, this does not quite work, since the union bound over a contributes a 2^n factor loss to the probability and results in a trivial bound. To get around this issue, we note that for any list of messages X, it suffices to control the above event for a specific choice of a, namely, an arbitrary *centroid* of the encodings of $x_1, \ldots, x_L$; we then finish the argument by a union bound over all X. Since the centroid minimizes the average distance of a center to a given list (see Fact 2), the code is now guaranteed to be (p, L) average-radius list-decodable.

Fix an L-list $X := \{x_1, \ldots, x_L\}$ of messages, let a denote the centroid of their encodings. For a fixed $j \in [n]$, by Chernoff bound, the probability that the j^{th} entry of a is 1 is at most $\exp_2\left(-2\left(\frac{1}{2} - \lambda'\right)^2 L\right)$, which is at most

$$\exp_2\left(-\frac{1}{2}\left(\frac{1}{2} - p\right)^2 L\right) = \exp(-bL) = \varepsilon.$$

Moreover, the entries of a in the n coordinates are all independent, and hence, by another application of the Chernoff bound (in the multiplicative form), the weight of a is at most $2\varepsilon n$, except with probability at most $\exp_2(-\varepsilon n/3)$. Assuming that this event holds, for each $x \in X$,

$$d(a, x) \geqslant \mathrm{wt}(x) - \mathrm{wt}(a) \geqslant (\lambda' - \varepsilon)n - 2\varepsilon n > (\lambda' - 4\varepsilon)n =: pn,$$

and hence the average distance of X from a is also more than pn. Finally, by a union bound over X, we can conclude that the code is (p, L) average-radius list-decodable, except with probability $\exp_2(RLn - \varepsilon n/3)$.

Thus, for $R = \min\{\varepsilon^2, \varepsilon/(6L)\}$, with probability $1 - o(1)$, the random code C satisfies the following:

- Each codeword in C has weight at most $(\lambda' + \varepsilon)n$. Note that $\lambda' + \varepsilon = p + 5\varepsilon = p + 5e^{-bL}$.
- C is (p, L) average-radius list-decodable.

Fix any C with the above properties. This satisfies all our requirements, except that its codewords could have varying weights. Fortunately, however, this is easily fixed, since, by the pigeonhole principle, C contains a constant-weight subcode C' of size at least $|C|/(n+1)$, and hence of rate $R - o(1)$. Now, if w_0 denotes the weight of the codewords of C', then note that $w_0 \leqslant (p + 5e^{-bL})n$, establishing the claim with $\lambda := w_0/n$. $\square$

Note that the *statement* of Theorem 18 also yields as a corollary $(\lambda; p, L)$ list-decodable codes of positive rate with λ exponentially close to p, since standard list-decodability is only a weaker requirement. However, interestingly, the above *proof* does not work directly because we do not have a simple analogue of Fact 2 identifying the best center that minimizes the maximum radius of a list. Indeed, the authors are not aware of any proof of this result except going through average-radius list-decodability.

5.2 A reverse connection between constant-weight and arbitrary codes

Lemma 19. *Fix p, λ such that $0 < p < \lambda \leqslant \frac{1}{2}$. Then in the notation of Definition 6, if $\gamma := 1 - h(p) - R_{p,L}$, then*

$$h(\lambda) - h(p) - \gamma \leqslant R_{p,L}(\lambda) \leqslant h(\lambda) - h(p) - \left(\frac{\lambda - p}{\frac{1}{2} - p}\right)\gamma.$$

Proof: The left inequality is essentially the content of Lemma 12; we show the second inequality here. The manipulations in this proof are of a similar flavor to those in Lemma 13, but the exact details are different.

Suppose C is a $(\lambda; p, L)$ list-decodable code of blocklength n and rate R, such that each codeword in C has weight exactly λn. Pick a random subset $S \subseteq [n]$ of coordinates of size $\alpha_2 n$, with $\alpha_2 := (\lambda - p)/(\frac{1}{2} - p)$, and let $\overline{S} := [n] \setminus S$. (Interestingly, our setting of α_2 differs from the parameter α employed in the proof of Lemma 13 only by a factor of 2. The motivation for this choice of α_2 will become clear shortly.) Consider the subcode C' consisting of codewords $c \in C$ such that $\mathrm{wt}(c|_S) \geqslant \alpha_2 n/2$. For our choice of α_2, one can verify that if $c \in C'$, then c has weight at most $p(1-\alpha_2)n = p|\overline{S}|$ when restricted to $\overline{S}$ (this is the motivation behind our choice of α_2).

Consider the restriction of C' to the coordinates in S, $C'|_S := \{c|_S \ : \ c \in C'\}$. Our key observation is that $C'|_S$, as a code of blocklength $\alpha_2 n$, is (p, L) list-decodable. Suppose not. Then there exists a center $x' \in \{0,1\}^S$ and a size-L list $\mathcal{L} \subseteq C$ such that $d(x', c|_S) \leqslant p\alpha_2 n$ for all $c \in \mathcal{L}$. Now, extend x' to $x \in \{0,1\}^n$ such that x agrees with x' on (the coordinates in) S and is zero on the remaining coordinates. Then $\mathcal{L}$ violates the (p, L) list-decodability of C, since for every $c \in \mathcal{L}$,

$$d(x, c) = d(x', c|_S) + \mathrm{wt}(c|_{\overline{S}}) \leqslant p\alpha_2 n + p(1-\alpha_2)n = pn.$$

Therefore, $C'|_S$ must be (p, L) list-decodable, and hence, by the hypothesis of the lemma, its size is at most $\exp_2((1 - h(p) - \gamma + o(1))\alpha_2 n)$ with probability 1. (It is crucial for this proof that the blocklength of C' is $\alpha_2 n$, which is significantly smaller than n.)

Now, for a fixed $c \in C$, the random variable $\mathrm{wt}(c|_S)$ follows the hypergeometric distribution with parameters $(n, \lambda n, \alpha_2 n)$, which is identical to the hypergeometric distribution with parameters $(n, \alpha_2 n, \lambda n)$. Hence, the probability that c is included in C' is at least

$$\begin{aligned} f(n, \alpha_2 n, \lambda n, \alpha_2 n/2) &= \frac{\binom{\alpha_2 n}{\alpha_2 n/2}\binom{(1-\alpha_2)n}{(\lambda - \alpha_2/2)n}}{\binom{n}{\lambda n}} \\ &\overset{(*)}{=} \frac{\binom{\alpha_2 n}{\alpha_2 n/2}\binom{(1-\alpha_2)n}{p(1-\alpha_2)n}}{\binom{n}{\lambda n}} \\ &\geqslant \exp_2\left(\alpha_2 n + h(p)(1-\alpha_2)n - h(\lambda)n - o(n)\right). \end{aligned}$$

In the step marked $(*)$, we have used the the identity $\lambda - \alpha_2/2 = p(1-\alpha_2)$, which holds for our particular choice of α_2. Summing this over all $c \in C$, the expected size of $C'|_S$ is at least

$$\exp_2\left(Rn + \alpha_2 n + h(p)(1-\alpha_2)n - h(\lambda)n - o(n)\right).$$

Finally, comparing the upper and lower bound on the expected size of $C'|_S$, we get

$$R + \alpha_2 + (1-\alpha_2)h(p) - h(\lambda) - o(1) \leqslant (1 - h(p) - \gamma)\alpha_2 + o(1),$$

which can be rearranged to give the desired bound $R \leqslant h(\lambda) - h(p) - \alpha_2\gamma + o(1)$. □

6 List-size bounds for random codes

In this section, we establish optimal (up to constant factors) bounds on the list-size of random codes, both general as well as linear.[3] Results of this vein were already shown by Rudra for the errors case [14], based on the large near-disjoint packings of Hamming balls implied by Shannon's capacity theorems. Here we give a direct proof based on the second moment method.[4] In addition, our proofs extend easily to give list-size bounds for *list-decodable codes for erasure channels* as well.

Throughout this section and Appendix B, we work with random q-ary codes – both general and linear. A *random q-ary code* (for $q \geqslant 2$) is simply a random map $C : [q]^k \to [q]^n$ where, for each $x \in [q]^k$, its image $C(x)$ is picked *independently and uniformly at random* from $[q]^n$. On the other hand, a *q-ary random linear code* is a random linear map $C : \mathbf{F}_q^k \to \mathbf{F}_q^n$ obtained in the following way. We fix an arbitrary *basis* (typically, but not necessarily, the standard basis) for the vector space $\mathbf{F}_q^k$, and the *encoding of the basis vectors* is chosen independently and uniformly at random from $\mathbf{F}_q^n$; the encoding map C naturally extends for all messages in $\mathbf{F}_q^k$ via linearity.

6.1 Proof idea

Our results proceed directly via the second moment method. Towards this goal, we define a random variable W that counts the number of *witnesses* (i.e., a bad list of codewords together with the center) that *certify the violation of the (p, L) list-decodability property*. Thus the code is (p, L) list-decodable if and only if $W = 0$. We then show that (a) W has large expectation (i.e., $\mathbf{E}\, W$ is exponential in n), but (b) its variance is relatively small (i.e., $\mathbf{Var}\, W/(\mathbf{E}\, W)^2$ is exponentially small in n). Then using the Chebyshev inequality (Fact 28), we can conclude that $W > 0$, except with an exponentially small probability, which is what we set out to show.

As a particular example, consider the case of random general codes under errors. In this case, the "potential violations" of the list-decoding property are indexed by pairs (a, X), where $a \in \{0,1\}^n$ is an arbitrary center, and X is an arbitrary distinct L-tuple of messages $\{x_1, x_2, \ldots, x_L\} \subseteq \{0,1\}^k$. We thus define the indicator random variable $\mathbb{I}(a, X)$ for the event that $d(a, C(x)) \leqslant pn$ for all $x \in X$, and let $W := \sum_{a,X} \mathbb{I}(a, X)$. The mean and variance estimates for W follow by standard calculations. See the formal proofs for details.

[3]In contrast to Sections 3–5, our results on random codes are stated as bounds on the list-size in terms of the rate. Recall that a rate upper bound of $1 - h_q(p) - \Omega_{q,p}(1/L)$ corresponds to a list-size bound of $\Omega_{q,p}(1/\gamma)$ for codes of rate $1 - h_q(p) - \gamma$.

[4]We remark that the argument in [14] is also based on the second moment method, but applied to a more complicated random variable.

6.2 Error list-decodability bounds

We state our bounds for standard list-decodable codes (under errors), deferring the complete proofs to Appendices B.1 and B.2.

Theorem 20. *Fix $q \geqslant 2$, $0 < p < 1 - 1/q$ and $\gamma > 0$.*

1. *A random q-ary code of rate $1 - h_q(p) - \gamma$ is $\left(p, \frac{1-h_q(p)}{2\gamma}\right)$ list-decodable with probability at most $\exp_q\left(-\Omega_{p,\gamma}(n)\right)$.*
2. *A random q-ary* linear *code of rate $1 - h_q(p) - \gamma$ is $\left(p, \frac{\delta_{q,p}}{2\gamma}\right)$ list-decodable with probability at most $\exp_q\left(-\Omega_{p,\gamma}(n)\right)$. Here, $\delta_{q,p}$ is a constant depending on only q and p.*

6.3 Erasure list-decodability bounds

The technique outlined in Section 6.1 also extends to give list-size bounds for random q-ary codes under the *erasure model*, which we now review briefly. In this model, the output alphabet is the usual alphabet $[q]$ augmented with a special *erasure symbol* '$?$'. For a string $a \in ([q] \cup \{?\})^n$, define $\mathrm{Supp}^*(a)$ to be the set of indices i such that $a_i \neq ?$. Also, we say that $a, b \in ([q] \cup \{?\})^n$ *agree* with each other if $a_i = b_i$ for all $i \in \mathrm{Supp}^*(a) \cap \mathrm{Supp}^*(b)$.

Definition 21. *A code $C \subseteq [q]^n$ is said to be (p, L)* erasure list-decodable *if for all $a \in ([q] \cup \{?\})^n$ satisfying $|\mathrm{Supp}^*(a)| = (1-p)n$, at most $L - 1$ codewords of C (treated as strings over $([q] \cup \{?\})$) agree with a.*

We are now ready to state our bounds for random (general and linear) codes under erasures. Note the exponential gap between the list-sizes of linear and general random codes under erasures.

Theorem 22. *Fix $q \geqslant 2$, $0 < p < 1$ and $\gamma > 0$.*

1. *A random q-ary code of rate $1 - p - \gamma$ is $\left(p, \frac{1-p}{2\gamma}\right)$ erasure list-decodable with probability at most $\exp_q\left(-\Omega_{p,\gamma}(n)\right)$.*
2. *Let q be a prime power. A random q-ary* linear *code of rate $1 - p - \gamma$ is $\left(p, \frac{1}{q} \cdot \exp_2\left(\frac{p(1-p)}{16\gamma}\right)\right)$ erasure list-decodable with probability at most $\exp_2\left(-\Omega_p(n)\right)$.*

The proofs for the above two bounds appear respectively in Appendices B.3 and B.4.

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. 1, 2, 3, 5, 11

[2] V. M. Blinovsky. *Asymptotic Combinatorial Coding Theory*. Kluwer Academic Publishers, Boston, 1997. 5

[3] V. M. Blinovsky. Code bounds for multiple packings over a nonbinary finite alphabet. *Problems of Information Transmission*, 41(1):23–32, 2005. 1, 11

[4] V. M. Blinovsky. On the convexity of one coding-theory function. *Problems of Information Transmission*, 44(1):34–39, 2008. 1

[5] M. Cheraghchi and V. Guruswami. Restricted isometry via list decoding. Work in progress, 2012. 2

[6] P. Elias. List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957. 1

[7] P. Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37:5–12, 1991. 1

[8] V. Guruswami. Linear-algebraic list decoding of folded Reed-Solomon codes. In *Proceedings of the 26th IEEE Conference on Computational Complexity*, pages 77–85, June 2011. 1

[9] V. Guruswami, J. Håstad, and S. Kopparty. On the list-decodability of random linear codes. *IEEE Transactions on Information Theory*, 57(2):718–725, 2011. 1, 5, 23

[10] V. Guruswami and A. Rudra. Explicit codes achieving list decoding capacity: Error-correction up to the Singleton bound. *IEEE Transactions on Information Theory*, 54(1):135–150, January 2008. 1

[11] V. Guruswami and M. Sudan. Extensions to the Johnson bound. *Unpublished manuscript*, 2001. Available at `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.9405`. 2

[12] V. Guruswami and S. P. Vadhan. A lower bound on list size for list decoding. *IEEE Transactions on Information Theory*, 56(11):5681–5688, 2010. 2, 3, 4, 5, 11

[13] S. Jukna. *Extremal Combinatorics: with applications in Computer Science*. Springer, 2001. 12

[14] A. Rudra. Limits to list decoding of random codes. *IEEE Transactions on Information Theory*, 57(3):1398–1408, 2011. 2, 11, 15

[15] J. M. Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958. 1

[16] 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. 1

A Technical results on standard functions

A.1 Properties of hypergeometric distributions

Proof of Fact 7: We consider a modification of the experiment in the definition of the hypergeometric distribution. Consider a set of n distinguishable objects that are marked by two players, Alice and Bob. Alice picks m objects uniformly at random and marks it 'A'. Simultaneously, Bob picks s objects uniformly at random and marks it 'B'. Moreover, the choices of Alice and Bob are *independent* of each other. We claim that the number of objects T marked by *both* Alice and Bob follows the hypergeometric distribution with parameters (n, m, s). Indeed, *conditioned* on the subset A of objects selected by Alice, the number of objects *from* A that are picked by Bob follows the hypergeometric distribution with parameters (n, m, s) (independent of A); we now obtain the claim by unconditioning on A.

Note that the above experiment is symmetric w.r.t. Alice and Bob, and hence the same argument shows that T follows the hypergeometric distribution with parameters (n, s, m) as well. The lemma now follows. $\square$

Proof of Fact 8: Consider an urn containing n balls, of which exactly m' are black, $m - m'$ are green, and the remaining are white. Sample s balls from the urn without replacement. Then, the number B of black balls picked follows the hypergeometric distribution with parameters (n, m', s), whereas the number N of nonwhite (i.e., black or green) balls picked follows the hypergeometric distribution with parameters (n, m, s). Since, for any outcome, it holds that $N \geqslant B$, the probability that $N \geqslant \tau$ is at least that of the event that $B \geqslant \tau$, which is what we wanted to show. $\square$

Remark. The joint random variable (B, N) is a *stochastic coupling* between the two hypergeometric distributions. $\square$

A.2 Properties of the binary entropy function

In this section, we will prove some standard properties of the binary entropy function used in this paper.

Fact 23. *For any p, λ such that $0 < p < \lambda \leqslant \frac{1}{2}$, we have*

$$h(\lambda) - h(p) \geqslant \frac{1}{2}(1 - 2p) \cdot (\lambda - p).$$

Proof: We begin with the identity

$$h(\lambda) - h(p) = \int_p^\lambda h'(z)\, dz = (\log e) \int_p^\lambda \ln\left(\frac{1-z}{z}\right)\, dz.$$

For $u \geqslant 1$, we have $\ln u \geqslant \frac{u-1}{u}$, which implies that for $0 < z \leqslant \frac{1}{2}$,

$$\ln\left(\frac{1-z}{z}\right) \geqslant \frac{\frac{1-z}{z} - 1}{\frac{1-z}{z}} = \frac{1-2z}{1-z} \geqslant (1 - 2z).$$

Therefore,

$$h(\lambda) - h(p) \geqslant (\log e) \int_p^\lambda (1 - 2z)\, dz = (\log e)(1 - \lambda - p)(\lambda - p) \geqslant (\log e)\left(\frac{1}{2} - p\right)(\lambda - p),$$

which establishes the claim.

$\square$

Fact 24. *For all $z \in (0, 1)$, we have $z \log(1/z) + (\log e)(z - z^2) \leqslant h(z) \leqslant z \log(1/z) + (\log e)z$.*

Proof: After expanding the definition of $h(\cdot)$, the above inequality reduces to

$$z - z^2 \leqslant -(1 - z)\ln(1 - z) \leqslant z.$$

We can equivalently write this as

$$\ln(1 - z) \leqslant -z, \text{ and } \ln\left(1 + \frac{z}{1-z}\right) \leqslant \frac{z}{1-z},$$

both of which are special cases of the standard inequality $\ln(1+z) \leqslant z$ valid for all real z. $\square$

Next, we show how to massage the rate upper bound given in Theorem 11 in Section 3 into a more convenient form. For the remainder of the section, we set $A_1 := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$, and $A_2 := \frac{2}{p^2}$.

Lemma 25.

$$A_1 \leqslant (1-2p) \cdot \frac{h(\lambda) - h(p)}{\lambda - p} + \frac{5}{p}(\lambda - p).$$

Proof: We begin with

$$\begin{aligned} A_1 &= (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right) \\ &\leqslant (1-p)\log\left(\frac{1-p}{p}\right) + p\log\left(\frac{p}{1-\lambda}\right) \\ &= (1-2p)\log\left(\frac{1-p}{p}\right) + p\log\left(\frac{1-p}{1-\lambda}\right) \\ &= (1-2p)h'(p) + p\log\left(\frac{1-p}{1-\lambda}\right). \end{aligned}$$

To complete the proof, we bound each term separately. First,

$$\begin{aligned} h'(p) = h'(\lambda) - \int_p^\lambda h''(z)\,dz &= h'(\lambda) + \int_p^\lambda \frac{\log e}{z(1-z)}\,dz \\ &\leqslant h'(\lambda) + \int_p^\lambda \frac{4}{z}\,dz \leqslant h'(\lambda) + \frac{4(\lambda - p)}{p}. \end{aligned}$$

Also, by the concavity of h, $h(\lambda) - h(p) \geqslant h'(\lambda)(\lambda - p)$, so $h'(p) \leqslant \frac{h(\lambda)-h(p)}{\lambda - p} + \frac{4(\lambda-p)}{p}$. On the other hand, applying the inequality $\ln z \leqslant z - 1$ with $z = \frac{1-p}{1-\lambda}$, we get

$$\log\left(\frac{1-p}{1-\lambda}\right) \leqslant (\log e)\frac{\lambda - p}{1-\lambda} \leqslant 4(\lambda - p) \leqslant \frac{\lambda - p}{p^2},$$

since $p < \frac{1}{2}$ and $e < 4$. Plugging these in the upper bound for A_1 gives the claim. $\square$

Lemma 26. *Fix $\varepsilon \in \left(0, \frac{1-2p}{2p}\right)$, and set $\beta := (\lambda - p)/(1 - 2p + 2p\varepsilon)$. Then*

$$A_1\beta + A_2\beta^2 \leqslant h(\lambda) - h(p) - B_1\varepsilon(\lambda - p) + B_2(\lambda - p)^2$$

for $B_1 := \frac{1}{2}p$ and $B_2 := \frac{3}{p^2(1-2p)^2}$. (Note that B_1 and B_2 are independent of λ and ε.)

Proof: From Lemma 25, we have

$$\begin{aligned} A_1\beta &\leqslant \left[(1-2p) \cdot \frac{h(\lambda) - h(p)}{\lambda - p} + \frac{5(\lambda - p)}{p}\right] \cdot \frac{\lambda - p}{1 - 2p + 2p\varepsilon} \\ &\leqslant \frac{1-2p}{1-2p+2p\varepsilon} \cdot (h(\lambda) - h(p)) + \frac{5(\lambda - p)^2}{p(1-2p)}. \end{aligned}$$

Assuming $0 < \varepsilon < \frac{1-2p}{2p}$, we can upper bound this by

$$\begin{aligned} A_1\beta &\leqslant \frac{1-2p-p\varepsilon}{1-2p}\cdot(h(\lambda)-h(p)) + \frac{5(\lambda-p)^2}{p(1-2p)} \\ &= h(\lambda)-h(p) - \frac{h(\lambda)-h(p)}{1-2p}\cdot p\varepsilon + \frac{5(\lambda-p)^2}{p(1-2p)} \\ &\leqslant h(\lambda)-h(p) - \frac{p\varepsilon(\lambda-p)}{2} + \frac{5(\lambda-p)^2}{p(1-2p)} \end{aligned}$$

using Fact 23. Also, $A_2\beta^2 \leqslant \frac{2(\lambda-p)^2}{p^2(1-2p)^2}$. Thus,

$$\begin{aligned} A_1\beta + A_2\beta^2 &\leqslant h(\lambda)-h(p) - \frac{p\varepsilon(\lambda-p)}{2} + \frac{5(\lambda-p)^2}{p(1-2p)} + \frac{2(\lambda-p)^2}{p^2(1-2p)^2} \\ &\leqslant h(\lambda)-h(p) - \frac{p\varepsilon(\lambda-p)}{2} + \frac{3(\lambda-p)^2}{p^2(1-2p)^2}. \end{aligned}$$

□

B List-decoding bounds for random codes

Throughout this section, we fix the parameters q, p, and n. For $a \in [q]^n$, let $\mathbf{B}_q(a, pn)$ be the q-ary Hamming ball with center a and radius pn. Let μ denote the fraction of points of $[q]^n$ that are inside a Hamming ball of radius pn; i.e., $\mu = |\mathbf{B}_q(a, pn)|/q^n$ for an arbitrary $a \in [q]^n$. We need the following estimate on μ (this generalizes Fact 9 for larger alphabet sizes):

Fact 27. *As $n \to \infty$, $\exp_q((h_q(p) - 1 - o(1))n) \leqslant \mu \leqslant \exp_q((h_q(p) - 1)n)$.*

We also need the following simple corollary of Chebyshev's inequality:

Fact 28. *Let W be a nonnegative random variable. Then, $W = 0$ with probability at most $\frac{\mathbf{Var}\, W}{(\mathbf{E}\, W)^2}$.*

B.1 Proof of part 1 of Theorem 20 (random general codes under errors)

Consider a random code $C : [q]^k \to [q]^n$, where $k := (1 - h_q(p) - \gamma)n$. Fix a positive integer L, to be specified later. For any center $a \in [q]^n$, and any (ordered) list of L *messages* $X := (x_1, x_2, \ldots, x_L) \subseteq [q]^k$, let $\mathbb{I}(a, X)$ be the indicator random variable for the event that the encoding of x falls inside the ball $\mathbf{B}_q(a, pn)$ for all $x \in X$. Moreover, define $W := \sum_{a,X} \mathbb{I}(a, X)$. Clearly, the code C is (p, L) list-decodable if and only if $W > 0$.

For a fixed center a and a fixed message x, the event that the encoding of x falls inside $\mathbf{B}_q(a, pn)$ occurs with probability μ; since the encodings of distinct messages are statistically independent, $\mathbf{Pr}\, \mathbb{I}(a, X) = \mu^L$. Also, assuming $k \geqslant L + 1$, the number of possible (a, X) pairs is at least $\frac{1}{2}q^{kL} \cdot q^n$, since the number of ordered L-lists X of distinct messages is

$$q^k(q^k-1)\cdots(q^k-L+1) \geqslant q^{kL}\left(1 - \sum_{i=0}^{L-1}\frac{i}{q^k}\right) = q^{kL}\left(1 - \frac{\binom{L}{2}}{q^k}\right) \geqslant q^{kL}\left(1 - \frac{2^L}{2^k}\right) \geqslant \frac{1}{2}q^{kL}.$$

Therefore, by linearity of expectations, $\mathbf{E}\, W \geqslant \frac{1}{2}\mu^L q^n q^{kL}$.

We now upper bound the variance of W. For two lists of messages X and Y, define the *intersection parameter* $l = l(X,Y) := |X \cap Y|$. If X and Y are disjoint (equivalently, if $l(X,Y) = 0$), then the events $\mathbb{I}(a,X)$ and $\mathbb{I}(b,Y)$ are independent for any pair of centers a, b. Therefore,

$$\begin{aligned}
\mathbf{Var}\, W &= \sum_{X,Y} \sum_{a,b} (\mathbf{E}\,[\mathbb{I}(a,X)\mathbb{I}(b,Y)] - \mathbf{E}\,[\mathbb{I}(a,X)] \cdot \mathbf{E}\,[\mathbb{I}(b,Y)]) \\
&= \sum_{X \cap Y \neq \emptyset} \sum_{a,b} (\mathbf{E}\,[\mathbb{I}(a,X)\mathbb{I}(b,Y)] - \mathbf{E}\,[\mathbb{I}(a,X)] \cdot \mathbf{E}\,[\mathbb{I}(b,Y)]) \\
&\leqslant \sum_{X \cap Y \neq \emptyset} \sum_{a,b} \mathbf{E}\,[\mathbb{I}(a,X)\mathbb{I}(b,Y)] \\
&= \sum_{X \cap Y \neq \emptyset} \sum_{a,b} \mathbf{Pr}\,[\mathbb{I}(a,X) = 1 \text{ and } \mathbb{I}(b,Y) = 1]. \\
&= \sum_{l=1}^{L} \sum_{|X \cap Y| = l} \sum_{a,b} \mathbf{Pr}\,[\mathbb{I}(a,X) = 1 \text{ and } \mathbb{I}(b,Y) = 1].
\end{aligned}$$

For convenience, we convert the inner summation into an expectation by randomizing over the centers a, b:

$$\mathbf{Var}\, W \leqslant q^{2n} \sum_{l=1}^{L} \sum_{|X \cap Y| = l} \mathbf{Pr}_{a,b,C}\,[\mathbb{I}(a,X) = 1 \text{ and } \mathbb{I}(b,Y) = 1]. \tag{4}$$

Here, in addition to the randomness of the code, the centers a and b are picked uniformly at random from $[q]^n$.

Fix $0 < l \leqslant L$, and a pair (X,Y) such that $|X \cap Y| = l$. Fix an arbitrary $z \in X \cap Y$; such a z is guaranteed to exist since X and Y intersect. Now, the event $\mathcal{E}$ that $\mathbb{I}(a,X) = \mathbb{I}(b,Y) = 1$ can be equivalently expressed as the conjunction of the events

- Both a, b fall inside $\mathbf{B}_q(C(z), pn)$;
- For each $x \in X \setminus z$, the encoding of x falls inside $\mathbf{B}_q(a, pn)$; and
- For $y \in Y \setminus X$, the encoding of y falls inside $\mathbf{B}_q(b, pn)$.

The first event occurs with probability μ^2, and conditioned on the choice of a and b, the second and third events occur with probabilities μ^{L-1} and μ^{L-l} respectively (and they are independent given a and b). Therefore the probability of $\mathcal{E}$ is μ^{2L-l+1}. Finally, by an easy counting, the number of pairs (X,Y) with $|X \cap Y| = l$ is at most $L^{2L} q^{k(2L-l)}$. Thus, we can bound the variance of W as

$$\mathbf{Var}\, W \leqslant q^{2n} \sum_{l=1}^{L} L^{2L} q^{k(2L-l)} \mu^{2L-l+1}.$$

Dividing by $(\mathbf{E}\, W)^2$, we get

$$\frac{\mathbf{Var}\, W}{(\mathbf{E}\, W)^2} \leqslant \sum_{l=1}^{L} 4L^{2L} (q^k \mu)^{-l} \mu.$$

For our choice of parameters, we have $q^k\mu = q^{-\gamma n}$, and hence

$$\frac{\mathbf{Var}\, W}{(\mathbf{E}\, W)^2} \leqslant \sum_{l=1}^{L} L^{4L} q^{\gamma l n}\mu \leqslant L^{4L+1} q^{\gamma L n} q^{-(1-h_q(p))n}.$$

This quantity is $\exp_q\left(-\Omega_{p,\gamma}(n)\right)$ for $L := \frac{1-h_q(p)}{2\gamma}$, and hence we are done by Fact 28.

B.2 Proof of part 2 of Theorem 20 (random linear codes under errors)

We follow the same outline as in Appendix B.1, so we will only highlight the differences. Let C be a random *linear* code of blocklength n and *dimension* $k = (1 - h_q(p) - \gamma)n$. We consider pairs (a, X) as before, but we now allow only *linearly independent list of messages* X. Moreover, the definition of W is unchanged, except that we sum over only the admissible X. Finally, we modify the definition of the parameter l to take linearity into account. For a pair of lists X and Y (each of which is linearly independent), we define $l = l(X, Y) := \dim(\mathrm{Span}(X) \cap \mathrm{Span}(Y))$ (where, for any set Z of message vectors, $\mathrm{Span}(Z)$ denotes its linear span). Note that $l = 0$ if and only if they X and Y are linearly independent of each other.

For any *linearly independent set* X, the encodings of vectors in X are statistically independent, and hence $\mathbf{E}\,\mathbb{I}(a, X) = \mu^L$. Once again, the number of linearly independent lists X is again at least $\frac{1}{2}q^{kL}$; indeed, the number of such lists is

$$(q^k - 1)(q^k - q)\cdots(q^k - q^{L-1}) \geqslant q^{kL}\left(1 - \sum_{i=0}^{L-1} q^{i-k}\right) \geqslant q^{kL}\left(1 - q^{L-k}\right) \geqslant \frac{1}{2}q^{kL}.$$

Therefore, as before, $\mathbf{E}\, W \geqslant \frac{1}{2}(q^k\mu)^L q^n$.

As before, the events $\mathbb{I}(a, X)$ and $\mathbb{I}(b, Y)$ are statistically independent whenever X and Y are linearly independent, i.e., $l = 0$. Therefore, as before, we can bound the variance of W by

$$\mathbf{Var}\, W \leqslant q^{2n} \sum_{l=1}^{L} \sum_{l(X,Y)=l} \mathbf{Pr}_{a,b,C}\left[\mathcal{E}\right],$$

where $\mathcal{E}$ is the event that $\mathbb{I}(a, X) = 1$ and $\mathbb{I}(b, Y) = 1$. Now, fix an l such that $1 \leqslant l \leqslant L$, and fix a pair X, Y such that $\dim(\mathrm{Span}\ X \cap \mathrm{Span}\ Y) = l$. Then, Y can be partitioned as $Y = Y_0 \cup Y_1$, with (a) $|Y_0| = l$ and $|Y_1| = L - l$, (b) X is linearly independent from Y_1, and (c) $Y_0 \subseteq \mathrm{Span}(X \cup\ Y_1)$. Fix an arbitrary $y_0 \in Y_0$. Then, by the span condition, we can write $y_0 = \sum_{u \in X \cup Y_1} \theta_u \cdot u$ for some set of *scalars* $\{\theta_u\}_{u \in X \cup Y_1}$. Note that it is possible that y_0 lies in the span of X. But, since Y is an independent set, y_0 *cannot* be written as a linear combination of vectors from Y_1 alone; in particular, there exists some $u \in X$ with $\theta_u \neq 0$.

In order to upper bound the probability of $\mathcal{E}$, we estimate the probability that $C(y_0) \in \mathbf{B}_q(b, pn)$, after *conditioning* on the subevent $\mathcal{E}'$ that $C(u) \in \mathbf{B}_q(a, pn)$ for all $u \in X$, and $C(u) \in \mathbf{B}_q(b, pn)$ for all $u \in Y_1$. (It is easy to check that the latter event occurs with probability $\mu^{|X \cup Y_1|} = \mu^{2L-l}$.)

At this point, it is convenient to re-center the vectors in $X \cup Y_1$ as follows: For $u \in X$, define $C'(u) := C(u) - a$, and for $u \in Y_1$, define $C'(u) := C(u) - b$. After conditioning on $\mathcal{E}'$, the random variables $C'(u)$ (for $u \in X \cup Y_1$) are i.i.d. and are uniformly distributed inside the ball $\mathbf{B}_q(\mathbf{0}, pn)$; furthermore, they are also independent of the choice of a and b. In terms of these new random variables, we can write

$$C(y_0) - b = \sum_{u \in X \cup Y_1} \theta_u \cdot C'(u) + \left(\sum_{u \in X} \theta_u\right) a + \left(\sum_{u \in Y_1} \theta_u - 1\right) b.$$

We claim that conditioned on $\mathcal{E}'$, $C(y_0) - b \in \mathbf{B}_q(\mathbf{0}, pn)$ occurs with probability at most $q^{-\Omega(n)}$. We discuss two cases:

1. Suppose $\sum_{u \in X} \theta_u \neq 0$, *or* $\sum_{u \in Y_1} \theta_u \neq 1$. Then, conditioned on the choice of $C'(u)$s, the random variable $C(y_0) - b$ is distributed uniformly at random inside $\mathbf{F}_q^n$ and hence falls inside $\mathbf{B}_q(\mathbf{0}, pn)$ with probability μ.

2. Suppose that $\sum_{u \in X} \theta_u = 0$, *and* $\sum_{u \in Y_1} \theta_u = 1$. In this case, we have

$$C(y_0) - b = \sum_{u \in X \cup Y_1 \,:\, \theta_u \neq 0} \theta_u \cdot C'(u). \tag{5}$$

Thus, if $m := |\{u \;:\; \theta_u \neq 0\}|$, then $C(y_0) - b$ is a sum of m *points sampled independently and uniformly from the ball* $\mathbf{B}_q(\mathbf{0}, pn)$. Also, as observed earlier, there exists some $u \in X$ such that $\theta_u \neq 0$; moreover, since $\sum_{u \in X} \theta_u = 0$, there are at least *two* u's in X with $\theta_u \neq 0$; i.e., $m \geqslant 2$. We now bound the probability of $\mathcal{E}$ conditioned on $\mathcal{E}'$ using the following fact:

Lemma 29. *For every $q \geqslant 2$ and every $p \in (0, \frac{1}{2})$, there exists $\delta = \delta_{q,p}$ such that the following holds all large enough integers n. If $m \geqslant 2$, and if $v_1, v_2, \ldots, v_m$ are m independent and uniformly random samples from $\mathbf{B}_q(\mathbf{0}, pn)$, then the probability that $v_1 + v_2 + \cdots + v_m \in \mathbf{B}_q(\mathbf{0}, pn)$ is bounded by $n^{O(m)} \cdot q^{-\delta n}$.*

We skip a formal proof of this lemma. A special case of this statement corresponding to $m = q = 2$ can be found in [9] (see Lemma 7), and the proof given there generalizes to give our claim with syntactic modifications.

We now return to the proof of Theorem 20. Since $m \leqslant 2L = O_{n\to\infty}(1)$, Lemma 29 implies that, conditioned on $\mathcal{E}'$, the stated event $\mathcal{E}$ also occurs with probability at most $q^{-\delta n + O(L \log n)} = q^{-\delta n + o(n)}$. (Without loss of generality, we may choose δ small enough so that this bound is larger than μ.)

Therefore, the conditional probability of $\mathcal{E}$ is at most the maximum of the two cases, namely $\exp_q(-\delta_{q,p} n + o(n))$. To complete the variance bound, we need an estimate on the number of pairs (X, Y) such that $l(X, Y) = l$. Partition Y as $Y_0 \cup Y_1$ as before. Now, $X \cup Y_1$ can be picked in at most $q^{k(2L-l)}$ ways. Also, for each $y \in Y_0$, we can write y as a linear combination of vectors in $X \cup Y_1$ in at most $q^{2L-l} \leqslant q^{2L}$ ways. Thus the total number of pairs (X, Y) with $l(X, Y) = l$ is at most $q^{2Ll} \cdot q^{k(2L-l)}$. Thus, the variance can be bounded as

$$\begin{aligned}
\mathbf{Var}\, W &\leqslant q^{2n} \sum_{l=1}^{L} q^{2Ll} \cdot q^{k(2L-l)} \mu^{2L-l} q^{-\delta n + o(n)} \\
&\leqslant \sum_{l=1}^{L} 4(\mathbf{E}\, W)^2 \cdot q^{2Ll} \left(q^k \mu\right)^{-l} q^{-an + o(n)} \\
&\leqslant 4(\mathbf{E}\, W)^2 \cdot \sum_{l=1}^{L} q^{2Ll} q^{\gamma ln - an + o(n)} \\
&\leqslant 4L q^{2L^2} q^{\gamma Ln - an + o(n)} \cdot (\mathbf{E}\, W)^2.
\end{aligned}$$

Therefore, as before, the probability that $W = 0$ is also at most $\exp_q(\gamma Ln - an + o(n))$. Thus, setting $L := \delta/(2\gamma)$, the claim follows.

B.3 Proof of part 1 of Theorem 22 (random general codes under erasures)

Consider a random code $C : [q]^k \to [q]^n$, where $k = (1-p-\gamma)n$. Let $\mathcal{A}$ be the set of potential inputs to the decoding algorithm, that is, $\mathcal{A} := \{a \in ([q] \cup \{?\})^n \ : \ |\operatorname{Supp}^*(a)| = (1-p)n\}$. We modify the definition of W as follows. For every $a \in \mathcal{A}$ and ordered L-list X of messages, define $\mathbb{I}(a, X)$ to be the indicator random variable for the event that, for all $x \in X$, the encoding $C(x)$ of x agrees with a; finally, in the definition of W, we consider only (a, X) pairs of the above form. As in the errors case, the code is (p, L) erasure list-decodable if and only if $W = 0$.

For every $a \in \mathcal{A}$ and $x \in [q]^k$, the encoding of x agrees with a with probability $q^{-(1-p)n}$, and hence by independence, the probability that $\mathbb{I}(a, X) = 1$ is $\exp_q(-(1-p)Ln)$. Therefore,

$$\mathbf{E}\, W \geqslant q^{-(1-p)Ln} \cdot \binom{n}{np} q^{(1-p)n} \cdot \frac{1}{2} q^{kL},$$

where the second factor is the number of possible a, and the third term is a lower bound on the number of X's. Moreover, proceeding as before, we can bound the variance of W by

$$\mathbf{Var}\, W \leqslant \sum_{l=1}^{L} \sum_{|X \cap Y| = l} \sum_{a,b} \mathbf{Pr}\ [\mathcal{E}]\,, \tag{6}$$

where $\mathcal{E}$ is the event that $\mathbb{I}(a, X) = \mathbb{I}(b, Y) = 1$.

Now, fix an arbitrary pair (X, Y) with $|X \cap Y| = l > 0$. Observe that the event $\mathcal{E}$ implies that both a and b agree with the encoding $C(u)$ of some $u \in X \cap Y$ (indeed, such a u is guaranteed to exist). Since $C(u)$ is a string over $[q]$ (i.e., it does not contain any '?"s), it follows that a and b must themselves agree with each other. Moreover, the event $\mathcal{E}$ requires that (a) $C(x)$ agrees with a for all $x \in X \smallsetminus Y$, (b) $C(y)$ agrees with b for all $y \in Y \smallsetminus X$, and (c) $C(z)$ agrees with *both* a and b for $z \in X \cap Y$. Therefore, the probability of $\mathcal{E}$ is at most

$$\exp_q(-|S||X \smallsetminus Y| - |T||Y \smallsetminus X| - |S \cup T||X \cap Y|) = \exp_q(-2(1-p)(L-l)n - |S \cup T| l)\,,$$

where $S := \operatorname{Supp}^*(a)$ and $T := \operatorname{Supp}^*(b)$. Now, for a given pair (S, T), the number of pairs of centers (a, b) such that (a) $\operatorname{Supp}^*(a) = S$, (b) $\operatorname{Supp}^*(b) = T$, and (c) a and b agree with each other (i.e., $a|_{S \cap T} = b|_{S \cap T}$), is equal to $q^{|S \cup T|}$. Thus, the inner summation in (6),

$$\begin{aligned}
\sum_{a,b} \mathbf{Pr}\ [\mathbb{I}(a, X) = 1 \text{ and } \mathbb{I}(b, Y) = 1] &= \sum_{S,T} \exp_q(-2(1-p)(L-l)n - |S \cup T|(l-1)) \\
&\leqslant \binom{n}{pn}^2 \exp_q(-2(1-p)(L-l)n - (1-p)n(l-1)) \\
&= \binom{n}{pn}^2 q^{-(1-p)(2L-l-1)n}.
\end{aligned}$$

Finally, plugging in this estimate in (6),

$$\begin{aligned}
\mathbf{Var}\, W &\leqslant \sum_{l=1}^{L} L^{2L} q^{k(2L-l)} \cdot q^{-(1-p)(2L-l-1)} \binom{n}{np}^2 \\
&= (\mathbf{E}\, W)^2 \cdot \sum_{l=1}^{L} 4L^{2L} \cdot q^{((1-p)n-k)l-(1-p)n} \\
&\leqslant (\mathbf{E}\, W)^2 \cdot 4L^{2L+1} \cdot q^{\gamma n L - (1-p)n}.
\end{aligned}$$

Thus, for $L := \frac{1-p}{2\gamma}$, the variance of W is $o((\mathbf{E}\, W)^2)$, and hence we are done.

B.4 Proof of part 2 of Theorem 22 (random linear codes under erasures)

We first note that if a linear code contains a list of L *linearly independent* codewords agreeing with some $a \in \mathcal{A}$, then its list-size is at least q^{L-1}. Indeed, if $c_1, c_2, \ldots, c_L$ are codewords agreeing with a, then, in fact, so does every *'affine' linear combination* of the codewords; i.e., every vector of the form $\theta_1 c_1 + \cdots + \theta_L c_L$ where the θ_i are scalars satisfying $\theta_1 + \theta_2 + \cdots + \theta_L = 1$. Note that the number of such linear combinations is exactly q^{L-1}.

Consider a random linear code C of blocklength n and dimension $k = (1-p-\gamma)n$. Recall that $\mathcal{A}$ is the set of strings a over $\mathbf{F}_q \cup \{?\}$ such that $|\operatorname{Supp}^*(a)| = (1-p)n$. For $a \in \mathcal{A}$ and any linearly independent L-list X of messages, define $\mathbb{I}(a, X)$ to be the indicator random variable for the event that $C(x)$ agrees with a for all $x \in X$, and let $W := \sum_{a,X} \mathbb{I}(a, X)$.

For fixed a and X, it is easy to see that $\mathbf{E}\, \mathbb{I}(a, X) = \exp_q(-(1-p)Ln)$, and therefore (as in Appendix B.3),

$$\mathbf{E}\, W \geqslant q^{-(1-p)Ln} \cdot \binom{n}{np} q^{(1-p)n} \cdot \frac{1}{2} q^{kL}.$$

For a pair of lists X and Y (each of which is linearly independent), define $l = l(X, Y) := \dim(\operatorname{Span}(X) \cap \operatorname{Span}(Y))$. It is easy to check that if $l = 0$ (i.e., X and Y are linearly independent), the random variables $\mathbb{I}(a, X)$ and $\mathbb{I}(b, Y)$ are statistically independent. Therefore, we can bound the variance of W by

$$\mathbf{Var}\, W \leqslant \sum_{l=1}^{L} \sum_{l(X,Y)=l} \sum_{a,b} \mathbf{Pr}\,[\mathcal{E}],$$

where $\mathcal{E}$ is the event that $\mathbb{I}(a, X) = 1$ and $\mathbb{I}(b, Y) = 1$.

Fix a pair X, Y such that $\dim(\operatorname{Span}\ X \cap \operatorname{Span}\ Y) = l > 0$. As in Subsection B.2, we partition Y as $Y_0 \cup Y_1$, where (a) $|Y_0| = l$ and $|Y_1| = L - l$, (b) X is linearly independent from Y_1, and (c) $Y_0 \subseteq \operatorname{Span}(X \cup Y_1)$. Moreover, pick $y_0 \in Y_0$ arbitrarily, so that $y_0 = \sum_{u \in X \cup Y_1} \theta_u \cdot u$ for some scalars $\{\theta_u\}_{u \in X \cup Y_1}$. Note that $\theta_x \neq 0$ for at least one $x \in X$.

Now, fix a pair of strings $a, b \in \mathcal{A}$, and let $S := \operatorname{Supp}^*(a)$ and $T := \operatorname{Supp}^*(b)$. We are interested in the probability of $\mathcal{E}$ for this choice of a and b. (Note that for general codes, this event implies that the strings a and b had to agree with each other; this is not so for linear codes.) For any $x \in X$, conditioned on the event that $C(x)|_S = a|_S$, the random variable $C(x)|_{T \smallsetminus S}$ is uniformly distributed over $\mathbf{F}_q^{T \smallsetminus S}$. Since $y_0 = \sum_{x \in X} \theta_x \cdot x + \sum_{y \in Y_1} \theta_y \cdot y$ (with $\theta_x \neq 0$ for some $x \in X$), it follows that $C(y_0)|_{T \smallsetminus S}$ is also uniformly distributed over $\mathbf{F}_q^{T \smallsetminus S}$. Hence, conditioned on the event that $C(x)$ agrees with a for all $x \in X$ and $C(y)$ agrees with b for all $y \in Y_1$, the probability that $C(y_0)$ agrees with b is at most $q^{-|T \smallsetminus S|}$. Hence,

$$\begin{aligned}
\sum_{a,b} \mathbf{Pr}\,[\mathcal{E}] &\leqslant \sum_{S,T} q^{-(1-p)n|X \cup Y_1|} q^{-|T \smallsetminus S|} \cdot q^{|S|+|T|}, \\
&= q^{-(1-p)n(2L-l)} q^{2(1-p)n} \cdot \sum_{S,T} q^{-|T \smallsetminus S|}, \\
&\leqslant q^{-(1-p)n(2L-l)} q^{2(1-p)n} \cdot \sum_{S,T} 2^{-|T \smallsetminus S|}, \\
&= q^{-(1-p)n(2L-l)} q^{2(1-p)n} \binom{n}{np}^2 \mathbf{E}_{S,T} \left[2^{-|T \smallsetminus S|}\right].
\end{aligned}$$

Here, the expectation is over $S, T \subseteq [n]$ of size $(1-p)n$, chosen independently and uniformly at random.

By Lemma 30 below, this quantity can be bounded by

$$q^{-(1-p)n(2L-l)} q^{2(1-p)n} \binom{n}{np}^2 \cdot \exp_2\left(-\frac{1}{8}p(1-p)n + o(n)\right).$$

Plugging this in our upper bound for the variance, we have

$$\begin{aligned}
\mathbf{Var}\, W &\leqslant \sum_{l=1}^{L} q^{2Ll} q^{k(2L-l)} \cdot q^{-(1-p)n(2L-l)} q^{2(1-p)n} \binom{n}{np}^2 \cdot 2^{-\frac{1}{8}p(1-p)n+o(n)} \\
&\leqslant 4(\mathbf{E}\, W)^2 \sum_{l=1}^{L} q^{2Ll} q^{((1-p)n-k)l} \cdot 2^{-\frac{1}{8}p(1-p)n+o(n)} \\
&\leqslant 4Lq^{2L^2} q^{\gamma nL} 2^{-\frac{1}{8}p(1-p)n+o(n)} \cdot (\mathbf{E}\, W)^2
\end{aligned}$$

Thus, for

$$L := \frac{p(1-p)}{16\gamma \log q},$$

this ratio is $o(1)$. Thus, the code contains a bad list of L *linearly independent* messages w.h.p.; this implies that its list-size is at least q^{L-1}.

Lemma 30. *If S, T are independently and uniformly random subsets of $[n]$ of size $(1-p)n$, then*

$$\mathbf{E}_{S,T}\left[2^{-|T\setminus S|}\right] \leqslant \exp_2\left(-\frac{p(1-p)n}{8} + o(n)\right).$$

Proof: We prove this by thresholding on $|T \setminus S|$. It can be easily checked that the random variable $|T \setminus S|$ has the hypergeometric distribution with parameters $(n, pn, (1-p)n)$, and hence its mean is $p(1-p)n$. Hence, since hypergeometric random variables are concentrated around their mean, we expect that $|T \setminus S| \geqslant \frac{1}{8}p(1-p)$, except with an exponentially small probability.

We now justify the above intuition by explicit calculations. For any t, the probability that $|T \setminus S| = t$ is equal to

$$f(n, (1-p)n, pn, t) := \frac{\binom{(1-p)n}{t}\binom{pn}{pn-t}}{\binom{n}{pn}} = \frac{\binom{(1-p)n}{t}\binom{pn}{t}}{\binom{n}{pn}}.$$

For $t \leqslant \frac{1}{8}p(1-p)n$, this can be upper bounded by $2^{\varepsilon n + o(n)}$, where

$$\varepsilon := (1-p)h\left(\frac{p}{8}\right) + ph\left(\frac{1-p}{8}\right) - h(p).$$

We are interested in upper bounding the exponent ε. We will assume that $p \leqslant 1/2$; the argument in the $p \geqslant 1/2$ case is symmetric (by replacing p by $1-p$). By concavity of $h(\cdot)$,

$$\varepsilon \leqslant h\left((1-p) \cdot \frac{p}{8} + p \cdot \frac{1-p}{8}\right) - h(p) \leqslant h(p/4) - h(p).$$

By Fact 24,

$$\varepsilon \leqslant \left[\frac{p}{4}\log\left(\frac{4}{p}\right) - p\log\left(\frac{1}{p}\right)\right] + (\log e)\left[\frac{p}{4} - (p - p^2)\right].$$

For $0 < p \leqslant 1/2$, the first term is negative, and hence $\varepsilon \leqslant (\log e)\left(p^2 - \frac{3p}{4}\right) \leqslant -\frac{1}{4}p \log e \leqslant -\frac{p(1-p)}{4}$.

Thus, summing over all $t \leqslant \frac{1}{8}p(1-p)n$, the event $|T \smallsetminus S| \leqslant \frac{1}{8}p(1-p)n$ occurs with probability at most $\exp_2\left(-\frac{p(1-p)n}{4} + o(n)\right)$. Hence, the desired expectation is bounded as

$$\mathbf{E}\left[2^{-|T \smallsetminus S|}\right] \leqslant \mathbf{Pr}\left[|T \smallsetminus S| \leqslant \frac{1}{8}p(1-p)n\right] \cdot 1 + \exp_2\left(-\frac{1}{8}p(1-p)n\right),$$

establishing the claim. $\square$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de