# LIST-DECODING CAPACITY IMPLIES CAPACITY ON THE Q-ARY SYMMETRIC CHANNEL

FRANCISCO PERNICE, OSCAR SPRUMONT, AND MARY WOOTTERS

Abstract. It is known that the Shannon capacity of the q-ary symmetric channel (qSC) is the same as the list-decoding capacity of an adversarial channel, raising the question of whether there is a formal (and black-box) connection between the two. We show that there is: Any linear code $C \subseteq \mathbb{F}_q^n$ that has minimum distance $d_{\min} = \omega(q^3)$ and achieves list-decoding capacity also achieves capacity on the qSC.

## 1. Introduction

A linear code of length $n$ over a finite field $\mathbb{F}_q$ is a $\mathbb{F}_q$-linear subspace $C \subseteq \mathbb{F}_q^n$. In coding theory, we think of a code $C$ as the set of encodings of possible messages. If a sender wants to send a message to a receiver over a noisy channel, they choose the corresponding element $c \in C$ (called a *codeword*), and transmit $c$ over the channel. A receiver sees a noisy version of the codeword, $\tilde{c}$, and must recover $c$. Two primary goals in designing such a code $C$ are (a) low redundancy, meaning that $n$, the length of a codeword $c \in C$, is not too much larger than $\log_q |C|$, the length of a message encoded with the code; and (b) tolerance of as many errors as possible.

The requirement of low redundancy is quantified by the *rate* of the code: The rate $R$ of a code $C \subseteq \mathbb{F}_q^n$ is defined as $R = \frac{\log_q |C|}{n}$. The rate $R$ is always between 0 and 1, and the larger it is, the lower the redundancy of the code.

The requirement on error tolerance depends on the channel model. In this paper, we focus on two well-studied channel models, both parameterized by $p \in (0, 1)$. The first model is the $q$-ary symmetric channel[1] qSC$_p$. The second model is an adversarial channel that may corrupt up to a $p$-fraction of the symbols sent in a worst-case way. Thus, in each model, we are concerned with the best possible trade-off between $R$ and $p$.

In the case of the qSC$_p$, the best trade-off between $R$ and $p$ is well-understood. The best rate at which reliable communication possible on the qSC$_p$—known as the *Shannon capacity* of the channel—is $R = 1 - h_q(p)$, where

$$h_q(p) := (1-p)\log_q \frac{1}{1-p} + p \log_q \frac{q-1}{p}$$

is the $q$-ary entropy function [Sha48]. A family of codes that approaches this trade-off is said to *achieve capacity* on the qSC$_p$ (see Definition 1).

[1] In the qSC$_p$, each $q$-ary symbol is corrupted independently with probability $p$; if a symbol is corrupted, it is replaced by a uniformly random different symbol in $\mathbb{F}_q$.

As one would expect, the best possible trade-off between $R$ and $p$ is worse in the adversarial case than in the stochastic case.[2] However, if one relaxes the definition of "reliable decoding," one can do better. More precisely, we consider the notion of *list-decoding*, a classical notion introduced by Elias and Wozencraft [Eli57, Woz58]. In list-decoding, the decoder's goal is no longer to return only the transmitted codeword $c \in C$, but rather a short list of possible codewords that is guaranteed to include $c$. Formally, $C$ is *$(p, L)$-list-decodable* if for all $w \in \mathbb{F}_q^n$,

$$|\{c \in C \,:\, d(w, c) \leq pn\}| \leq L.$$

It is known (see, e.g., [GRS23], Theorem 7.4.1) that the best [3] possible trade-off between $R$ and $p$ in the list-decoding setting is also $1 - h_q(p)$, exactly the same as the Shannon capacity of the qSC$_p$! If a family of list-decodable codes approaches this trade-off, we say that it *achieves list-decoding capacity* (see Definition 2).

**Our question.** This state of affairs raises a natural question: Since the capacity of the qSC$_p$ is the same as the list-decoding capacity, is there a formal connection between these two notions? In the list-decoding literature, it is common to introduce list-decoding as a "bridge" between the Shannon and Hamming models, in that list-decoding allows one to reach the Shannon capacity of a channel, even under a worst-case (Hamming) model of errors.[4] But can this be made formal?

In this paper we show that there is a formal connection, at least in one direction. That is, we show that *any* list-decodable code $C$ with good enough minimum distance[5] achieves capacity on the qSC$_p$. (As we note below, the converse is not true.) In the next section, we describe our results in more detail.

## 1.1. **Our Results.** Our main result is the following theorem.

**Theorem 1.** *Let $p \in (0, 1)$. Let $\{C_n \subseteq \mathbb{F}_q^n\}$ be a family of linear codes that achieves list-decoding capacity on the adversarial channel that introduces a $p$-fraction of corruptions. If $d_{\min}(C_n) = \omega\left(\frac{q^3}{(1-p)^2}\right)$, then $\{C_n\}$ achieves capacity on the qSC$_p$.*

Theorem 1 follows from a more general statement about $(p, L)$-list-decodability, which we present in Theorem 15. Next, we state a slightly weaker but more easily digestible version of that result.

[2] In fact, for small $q$, the best possible trade-off between $R$ and $p$ is still unknown in the adversarial model.

[3] In this informal discussion, we have not mentioned the size $L$ of the list the decoder is allowed to output. Slightly more formally, the *list-decoding capacity theorem* states that there are codes of rate $R = 1 - h_q(p) - \epsilon$ that are $(p, L)$-list-decodable where $L$ depends on $\epsilon$ but not on $n$; and that conversely any code of rate bounded above $1 - h_q(p)$ cannot be list-decodable with any list size $L$ that is sub-exponential in $n$.

[4] For example, the chapter on list-decoding in the textbook [GRS23] is titled "Bridging the Gap Between Shannon and Hamming: List Decoding."

[5] The *minimum distance* of a code $C$ is $d_{\min}(C) = \min_{c \neq c' \in C} d(c, c')$, where $d(\cdot, \cdot)$ denotes Hamming distance.

**Theorem 2.** *Let $C \subseteq \mathbb{F}_q^n$ be a linear, $(p, L)$-list-decodable code with minimum distance $d_{\min}(C) \geq \frac{q^3}{(1-p)^2} \ln^4(nL)$ for sufficiently large $n$. Then $C$ can be used for reliable communication on the $qSC_{p'}$, where $p' = p - \frac{7}{\ln n}$.*

See Section 5 for the proof of Theorems 1 and 2.

**Remark 1** (Requirement on the field size)**.** *As stated, the hypothesis of Theorem 2 requires that $d_{\min}(C)$ be at least $q^3$, which in particular implies that $q < n^{1/3}$. This rules out codes like Folded Reed-Solomon codes, where $q$ is a large polynomial in $n$. However, for known constructions of list-decodable codes over growing alphabets, the conclusion of Theorem 2 already follows from previous work, and so our contribution is to focus on the case where $q \geq 2$ is fixed.*

*In more detail, it is known that any linear code with distance at least $pn$ over a sufficiently large alphabet admits reliable communication over the $qSC_{p'}$ for some $p' = p - o(1)$* [RU10]*; and it is also known that any linear $(p, L)$-list-decodable code $C \subseteq \mathbb{F}_q^n$ with $q \geq L$ has distance at least $pn$* [GST21, Proposition 6.5]*. Thus, any linear $(p, L)$-list-decodable code $C \subseteq \mathbb{F}_q^n$ with $q = \omega(1)$ and $L = O(1)$ has distance at least $pn$ by* [GST21]*, and therefore admits reliable communication on the $qSC$ by* [RU10]*.*

We note that the converse of Theorem 1 (qSC capacity implying list-decoding capacity) does not hold in general. For instance, Reed-Muller codes achieve capacity over the $\mathrm{BSC}_p$ [RP24, AS23] but have $2^{\Omega(n)}$ codewords of weight $\leq pn$, for every constant $p$ (see, e.g., [ASSY23]). We leave it as an intriguing open problem to identify sufficient conditions under which the converse holds.

We show in Appendix B that the requirement that the minimum distance be large cannot be avoided entirely. More precisely, we show that there exist codes $C \subseteq \mathbb{F}_q^n$ with constant minimum distance that achieve list decoding capacity but do not achieve capacity on the $q$-ary symmetric channel.

Finally, we give a simple proof that Theorem 2 holds for erasures, without the linearity and alphabet size requirements.

**Theorem 3.** *Let $C \subseteq \mathbb{Z}_q^n$ be a $(p, L)$-list-decodable code with minimum distance $d_{\min}(C) = \omega(\log L)$. Then $C$ can be used for reliable communication on the $q$-ary erasure channel, $qEC_{p'}$, where $p' = p - o(1)$.*

We prove Theorem 3 in Appendix A.

**Remark 2** (New results for capacity-achieving codes on the $\mathrm{qSC}_p$?)**.** *It is natural to ask whether or not our results imply new constructions of capacity-achieving codes on the $qSC_p$. The answer is yes, in that there are known capacity-achieving list-decodable codes over constant-sized alphabets, even with efficient algorithms, to which our theorem applies and which to the best of our knowledge had not previously been known to achieve capacity on the qSC (e.g.* [HRW17, KRSW18, GR22]*). However, it seems unlikely that these constructions—which are designed to solve the (as we show here) strictly harder problem of list-decoding—would yield more practical results on the qSC than polar codes* [Ari09]*. Thus, our main contribution is to establish a formal connection between these two problems.*

1.2. **Overview of Techniques.** Before we discuss our techniques, we set up a bit more notation. For a transmitted codeword $c \in C$, we write the received corrupted codeword as

$$\tilde{c} = c + z,$$

where $z \in \mathbb{F}_q^n$ is an error vector. Thus, in the $\mathrm{qSC}_p$, each coordinate of $z_i$ is independent, and is equal to 0 with probability $1 - p$ and is uniformly random in $\mathbb{F}_q^*$ with probability $p$. In the adversarial channel, $z$ is an arbitrary vector of weight at most $pn$.

Now we can give a high-level overview of our techniques. Let $D^* : \mathbb{F}_q^n \to C$ be a maximum-likelihood decoder on the $\mathrm{qSC}_p$.[6] As $C$ is linear, we can pick $D^*$ so that its success depends only on the error vector $z$, and we may assume without loss of generality that the transmitted codeword was $c = \vec{0}$. Define the function

$$f(z) := \begin{cases} 1 & \text{if } D^*(z) = \vec{0}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The expectation of $f$ is exactly the probability that $D^*$ outputs the correct codeword. Showing that $C$ allows reliable communication on $\mathrm{qSC}_{p'}$ for some $p' = p - o(1)$ is thus equivalent to showing that

$$\mathop{\mathbb{E}}_{z \sim p'}[f(z)] = 1 - o(1). \tag{2}$$

In the equation above, we use $z \sim p'$ to mean that, independently for each coordinate $i \in [n]$,

$$z_i = \begin{cases} 0 & \text{with probability } 1 - p' \\ j \in \{1, 2, \ldots, q-1\} & \text{with probability } \frac{p'}{q-1} \end{cases}.$$

One key observation is that if $C$ is $(p, L)$-list-decodable, then we must have

$$\mathop{\mathbb{E}}_{z \sim p'}[f(z)] \geq \frac{1}{L} - o(1) \tag{3}$$

for, say, $p' = p - n^{-1/4}$. Indeed, consider the following decoder $D$: upon receiving some corrupted codeword $m \in \mathbb{F}_q^n$, find all the codewords $c \in C$ such that $\mathrm{wt}(c + m) \leq pn$, and output one such codeword uniformly at random. Since $C$ is $(p, L)$-list decodable, there can never be more than $L$ codewords $c \in C$ satisfying $\mathrm{wt}(c + m) \leq pn$. Thus as long as the error string $z$ has weight smaller than $pn$, the decoder $D$ will succeed in outputting the correct codeword with probability at least $\frac{1}{L}$. But if $z \sim p'$, then $z$ has weight smaller than $pn$ with high probability. Thus, on $z \sim p'$ our decoder $D$ outputs the correct codeword with probability at least $\frac{1}{L} - o(1)$. Since the max-likelihood decoder $D^*$ is optimal, it must perform at least as well as the decoder $D$, and we thus get (3).

In order to deduce (2) from (3), it will then suffice to show that the function $g(p') = \mathbb{E}_{z \sim p'}[f(z)]$ has a sharp transition as a function of $p'$. This was proven for $q = 2$ in [Zém93, TZ00]. Thus, combined with their result, the above argument immediately implies Theorem 2 for binary codes. Our main technical contribution is to generalize

[6] That is, given a corrupted codeword $\tilde{c} = c + z$, the decoder $D^*$ returns a closest codeword $D^*(\tilde{c})$ to $\tilde{c}$.

their arguments to larger field sizes.[7] Formally, our goal will be to bound the derivative of $\mathbb{E}[f]$ by

$$\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}[f(z)] \le -\omega(1)\cdot \mathop{\mathbb{E}}_{z\sim p}[f(z)]\big(1-\mathop{\mathbb{E}}_{z\sim p}[f(z)]\big). \tag{4}$$

Margulis [Mar74] and Russo [Rus82] pioneered the use of such inequalities for proving that the expectation of certain Boolean functions transitions quickly from $1-o(1)$ to $o(1)$. The point is that whenever $\mathbb{E}[f]$ is away from 0 and away from 1, we have $\mathbb{E}[f](1-\mathbb{E}[f])$ away from 0, and thus the $\omega(1)$ term in (4) ensures that in this regime the derivative of $\mathbb{E}[f]$ is large.

Now for any monotone function $f:\mathbb{F}_q^n\to\{0,1\}$, we can bound the derivative of the expectation as

$$\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}[f(z)] \le -\frac{1}{q-1}\mathop{\mathbb{E}}_{z\sim p}[h_f(z)], \tag{5}$$

where we define the quantity

$$h_f(z) := \begin{cases} \Big|\big\{i\in[n] : z_i = 0 \text{ and } \exists a\ne 0 \text{ s.t. } f(za_i)=0\big\}\Big| & \text{if } f(z)=1,\\ 0 & \text{otherwise.}\end{cases}$$

For $q=2$, the inequality (5) first appeared in [Mar74, Rus82] and is called Russo's Lemma; the generalization to larger $q$ is stated as Lemma 9. To make use of (5), we prove in Section 3 the following isoperimetric inequality, which is a generalization of a bound proven by Talagrand for $q=2$:

$$\mathop{\mathbb{E}}_{z\sim p}\big[h_f(z)\big] \ge \frac{1-p}{2}\sqrt{\Delta_f}\cdot \mathop{\mathbb{E}}_{z\sim p}[f(z)]\big(1-\mathop{\mathbb{E}}_{z\sim p}[f(z)]\big), \tag{6}$$

where we denoted the minimum positive value of $h_f$ by

$$\Delta_f := \min_{x\in\mathbb{F}_q^n : h_f(x)\ne 0}\{h_f(x)\}.$$

To obtain our desired inequality (4) from the bounds (5) and (6), it will thus suffice to show that our specific function $f$ (the indicator function of a successful decoding, defined in (1)) satisfies $\Delta_f = \omega(1)$. That is, we want to show that if $x\in\mathbb{F}_q^n$ is some error string that leads to correct decoding, and if one of the neighbors of $x$ leads to incorrect decoding, then there must be many such "bad" neighbors of $x$.

Intuitively, this is because in order for $x$ and one of its neighbors to be mapped to different codewords, it must be the case that $x$ is about halfway between the transmitted codeword (as above, let us assume that the transmitted codeword is the all-0 codeword), and some other codeword $c$. Formally, it must be the case that

$$d(x,0)\le d(x,c)\le d(x,0)+2,$$

where $d(a,b)$ denotes the Hamming distance between vectors $a$ and $b$. For simplicity, in this section assume that $d(x,c)=d(x,0)+1$. Then for any coordinate $i\in[n]$ where

[7] As discussed in Section 1.3, such a generalization was asserted in [KCC10] in a different context, but as discussed in Appendix C, we believe that their proof is missing key elements.

$x_i = 0$ and $c_i \neq 0$, the neighbor $x'$ of $x$ obtained by setting the $i^{\text{th}}$ coordinate to $c_i$ is closer to $c$ than to 0. Each such neighbor $x'$ thus satisfies $f(x') = 0$, and we have

$$h_f(x) \geq \left|\left\{i \in [n] : x_i = 0, c_i \neq 0\right\}\right|.$$

The exact number of such coordinates $i \in [n]$ can be bounded in terms of the distance between 0 and $c$, which itself is bounded by the minimum distance of the code. See Section 4 for more details.

Once we obtain a lower bound on $\Delta_f$, our desired inequality (4) follows from equations (5) and (6).

### 1.3. Related Work.

Both capacity-achieving codes on the qSC and capacity-achieving list-decodable codes are extremely well-studied, with lines of work going back to the 1950's or earlier. In this section, we mention a few of the most relevant works.

**Capacity-Achieving Codes on the qSC$_p$.** As noted above, the capacity of the qSC$_p$ has been known since Shannon's seminal work in the 1940's [Sha48]. Shannon already observed that random codes achieve capacity, but it was not until the 1960's that explicit constructions were obtained with Forney's *concatenated codes* [For66]. These codes have poor performance in terms of the gap to capacity, which has been more recently improved by Arıkan with polar codes [Ari09]; polar codes also have efficient decoding algorithms. Even more recently, Reed-Muller codes have also been shown to achieve capacity on the binary symmetric channel [ASW15, KKM$^+$16, RP24, AS23]. Ensembles of LDPC codes are also known to achieve capacity under maximum-likelihood decoding [Gal62], or even under efficient algorithms [LMS$^+$97, KRU13].

**List decoding.** The notion of list decoding was first introduced by Elias [Eli57] and Wozencraft [Woz58], and since then several ensembles of codes have been shown to achieve list-decoding capacity. Here, we survey these codes and also comment on how Theorem 1 applies to them.

The original work of Elias and Wozencraft showed that uniformly random codes achieve list-decoding capacity with high probability, and a more recent line of work has established the same for uniformly random linear codes [ZP81, GHK10, Woo13, LW21, GLM$^+$22]. Both random codes and random linear codes are known to achieve capacity on the qSC with high probability, so our results do not imply anything new for these ensembles.

The first explicit constructions were obtained by Guruswami and Rudra, who showed in [GR08] that *folded Reed-Solomon codes* achieve list decoding capacity. This breakthrough was followed by a line of work on explicit constructions, both improving the list size for folded RS codes (e.g., [KRSW18, Tam24, Sri24, CZ24]), and by extending these results to other families of codes, including multiplicity codes [GW13, Kop15]. Going back to randomized constructions (though with more structure than random linear codes), a recent line of work has also established that randomly punctured Reed-Solomon codes [BGM23, GZ23, AGL24, BST24] are also capacity-achieving list-decodable codes. As discussed in Remark 1, our resuls to not apply to these codes (as they are over large alphabets), but for those with constant list-sizes, the fact that they achieve capacity on the qSC follows from known results.

Finally, we turn to constructions of capacity-achieving list-decodable codes with constant alphabet sizes. These include Gallager's ensemble of LDPC codes [MRR$^+$20]; concatenated codes [GR10]; and constructions based on algebraic geometry (AG) codes, for example [GX13, HRW17, GR22, GX22, BDGZ23]. Some of these codes—for example, [HRW17, GR22]—come with efficient list-decoding algorithms and were also not already known to achieve capacity on the qSC$_p$. Combined with our results, this implies that not only do these codes achieve capacity on the qSC$_p$, but they also have efficient decoding algorithms.

**Threshold Phenomena.** As mentioned in Section 1.2, our main technical contribution is a threshold result for the success probability of decoding on the qSC$_p$. Our techniques build on a long line of work, which we briefly mention here.

Margulis [Mar74], Russo [Rus82] and Talagrand [Tal93] showed that the expectation of any monotone[8] Boolean function $f$ sharply transitions from $\mathbb{E}[f] \geq 1 - o(1)$ to $\mathbb{E}[f] \leq o(1)$, as long as no $z \in f^{-1}(1)$ has a small, nonzero number of neighbors in $f^{-1}(0)$. This fact has already seen many applications in coding theory [Zém93, TZ00, TZ04, KKM$^+$16, KCP16]. In particular, Tillich and Zémor proved in [Zém93, TZ00] that the decoding probability of any binary linear code with large minimum distance undergoes a sharp transition. Our main technical contribution is a generalization of these results to larger alphabets (see Sections 3 and 4, and in particular Theorem 14).

We note that an attempt was already made in [KCC10] to generalize the results of Tillich and Zémor to larger alphabets, but unfortunately the proof in [KCC10] seems to be missing key elements and does not seem to be correct as written. We explain in Appendix C why we think that a new proof of this generalization is needed.

## 2. Conventions and Preliminaries

2.1. **Finite fields.** We will work with the finite field $\mathbb{F}_q$ over $q$ elements. Given any vector $z \in \mathbb{F}_q^n$, we denote its Hamming weight by

$$\mathrm{wt}(z) := \{i \in [n] : z_i \neq 0\}.$$

Given two vectors $y, z \in \mathbb{F}_q^n$, we denote their Hamming distance by

$$d(y, z) := \mathrm{wt}(z - y).$$

2.2. **Probability Theory.** For any probability distribution $\mathcal{D}$ over a set $X$, we will use the notation $x \sim \mathcal{D}$ to denote a random element $x \in X$ sampled according to $\mathcal{D}$. The main probability distribution that we will use throughout this paper is the distribution of a *p-noisy string* over $\mathbb{F}_q^n$, which we define as follows. For $p \in [0, 1]$, we use

$$z \sim p$$

[8] See Section 2.3 for a formal definition of monotonicity.

to denote a $p$-noisy string $z \in \mathbb{F}_q^n$, meaning that for each $i \in \{1, \ldots, n\}$, $i^{\text{th}}$ entry is independently

$$z_i = \begin{cases} 0 & \text{with probability } 1-p, \\ j & \text{with probability } \frac{p}{q-1}, \text{ for all } j \in \{1, 2, \ldots, q-1\}. \end{cases}$$

We will also need Hoeffding's inequality (see for example, [BLM13]).

**Lemma 4** (Hoeffding's Inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $[0,1]$. Then for any $t > 0$, we have*

$$\Pr\Big[\sum_{i=1}^n X_i > \sum_{i=1}^n \mathbb{E}[X_i] + t\Big] \leq e^{-\frac{2t^2}{n}}$$

*and*

$$\Pr\Big[\sum_{i=1}^n X_i < \sum_{i=1}^n \mathbb{E}[X_i] - t\Big] \leq e^{-\frac{2t^2}{n}}.$$

2.3. **Monotonicity.** A property that will play a key role in our analysis is monotonicity. We say that a function $f : \mathbb{F}_q^n \to \{0,1\}$ is *monotone decreasing* if for any index $i \in [n]$, any point $z \in \mathbb{F}_q^n$, and any $a \in \{1, 2, \ldots, q-1\}$, we have

$$f(z0_i) \leq f(za_i).$$

The expectation of any monotone decreasing function is decreasing (see for e.g. [BLM13], page 280).

**Fact 5.** *For any monotone decreasing function $f : \mathbb{F}_q^n \to \{0,1\}$, the function*

$$g(p) := \mathop{\mathbb{E}}_{z \sim p}[f(z)]$$

*is decreasing.*

We derive some useful properties of monotone functions in Section 3.

2.4. **Coding and Decoding.** A linear code of length $n$ over the field $\mathbb{F}_q$ is a subspace $C \subseteq \mathbb{F}_q^n$. A deterministic decoder for a code $C \subseteq \mathbb{F}_q^n$ is a function $D : \mathbb{F}_q^n \to C$, and a randomized decoder is a probability distribution over the set of deterministic decoders. We say that the code $C \subseteq \mathbb{F}_q^n$ *admits reliable communication* on the qSC$_p$ if there exists a randomized decoder $D$ such that for all $c \in C$, we have

$$\Pr\big[D(c+z) = c\big] \geq 1 - o(1),$$

where the probability is taken over both the randomized decoder $D$ and the random variable $z \sim p$. A *maximum-likelihood* decoder $D^*$ over the qSC is a deterministic decoder that, upon seeing a corrupted message $m$, returns a codeword $c \in C$ closest to $m$. We say that $D^*$ is *symmetric* if it breaks ties in a symmetric way, meaning that if $D^*(z) = 0$ then $D^*(z+c) = c$. It is well-known that any symmetric maximum-likelihood decoder is optimal for decoding random errors (see, e.g. [MS77], page 8).

**Fact 6.** *For any code $C \subseteq \mathbb{F}_q^n$, any corresponding symmetric maximum-likelihood decoder $D^*$ is optimal. That is, for any $p \in (0,1)$ and any decoder $D$, we have*

$$\min_{c \in C} \left\{ \Pr_{z \sim p} \left[ D(c+z) = c \right] \right\} \leq \min_{c \in C} \left\{ \Pr_{z \sim p} \left[ D^*(c+z) = c \right] \right\}.$$

2.5. **Communication Channels.** We will be interested in the performance of codes under both stochastic and adversarial noise.

The model we will use for stochastic noise is the q-ary Symmetric Channel, which is characterized by a parameter $p \in (0,1)$. When an element $c$ of the code $C \subseteq \mathbb{F}_q^n$ is sent through the q-ary Symmetric Channel $\mathrm{qSC}_p$, each of its $n$ entries is corrupted independently at random with probability $p$. If corrupted, the $i^{\text{th}}$ entry $c_i$ is replaced by a uniformly random element in $\mathbb{F}_q \setminus \{c_i\}$. It is well-known [Sha48] that a uniformly random code $C \subseteq \mathbb{F}_q^n$ of rate

$$\mathrm{rate}(C) = 1 - h_q(p) - o(1)$$

will with high probability admit reliable communication on the $\mathrm{qS}C_p$, where $h_q$ denotes the q-ary entropy function

$$h_q(p) := (1-p) \log_q \frac{1}{1-p} + p \log_q \frac{q-1}{p}.$$

On the other hand, no code $C \subseteq \mathbb{F}_q^n$ of rate $1 - h_q(p) + o(1)$ can admit reliable communication on the $\mathrm{qS}C_p$. Thus we have the following definition of qSC-capacity.

**Definition 1.** *A sequence of codes $\{C_n \subseteq \mathbb{F}_q^n\}$ of rate $1 - h_q(p)$ achieves capacity over the $qSC_p$ if there exists a function $\epsilon(n) = o(1)$ such that each $C_n$ satisfies*

$$\Pr_{z \sim p - \epsilon_n} [D(c+z) = c] \geq 1 - \epsilon_n.$$

In the adversarial noise model, the location of the errors is decided by some adversary rather than being stochastic. We say that the code $C \subseteq \mathbb{F}_q^n$ is $(p, L)$-*list decodable* if for any $z \in \mathbb{F}_q^n$, the ball of radius $pn$ around $z$ contains at most $pn$ codewords $c \in C$. A uniformly random code $C \subseteq \mathbb{F}_q^n$ of rate

$$\mathrm{rate}(C) = 1 - h_q(p) - \epsilon \tag{7}$$

is $(p, O(\frac{1}{\epsilon}))$-list decodable with high probability (see for e.g. [GRS23], Theorem 7.4.1). On the other hand, no code $C \subseteq \mathbb{F}_q^n$ of rate larger than $1 - h_q(p) + \epsilon$ is $(p, L)$-list decodable for any $L < q^{\Omega(\epsilon n)}$. Thus we obtain the following definition of list decoding capacity.

**Definition 2.** *A sequence of codes $\{C_n \subseteq \mathbb{F}_q^n\}$ of rate $1 - h_q(p)$ achieves list decoding capacity if for every function $L(n) = \omega(1)$, there exists a function $\epsilon(n) = o(1)$ such that each $C_n$ is $(p - \epsilon(n), L(n))$-list decodable.*

We note for example that by equation (7), uniformly random codes achieve list decoding capacity with $\epsilon(n) = O(\frac{1}{L(n)})$. But in general, we allow for worse dependence of $\epsilon$ on $L$: we only require that $\epsilon$ go to 0 as $n$ goes to infinity.

## 3. An isoperimetric inequality over finite fields

In this section, we generalize an $\mathbb{F}_2$-result of Talagrand to finite fields of all sizes. For any function $f : \mathbb{F}_q^n \to \{0,1\}$ and any point $z \in \mathbb{F}_q^n$, we define the quantity

$$h_f(z) := \begin{cases} \left|\{i \in [n] : z_i = 0 \text{ and } \exists a \neq 0 \text{ s.t. } f(za_i) = 0\}\right| & \text{if } f(z) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Our goal will be to relate the quantities $\mathbb{E}[h_f]$ and $\mathbb{E}[f]$. Theorem 7 below was proven for the special case of $q = 2$ by Talagrand in [Tal93]; we extend that result to arbitrary field sizes.

**Theorem 7.** *For any monotone decreasing function $f : \mathbb{F}_q^n \to \mathbb{F}_2$ and any noise parameter $p \in [0,1]$, we have*

$$\mathbb{E}\left[\sqrt{h_f}\right] \geq \frac{1-p}{2} \cdot \mathbb{E}[f]\big(1 - \mathbb{E}[f]\big),$$

*where all the expectations are taken over the $q$-ary $p$-noisy distribution.*

*Proof.* We proceed by induction on $n$. For the base case $n = 1$, either we have $f(0) = f(a)$ for all $a \in \mathbb{F}_q$, in which case the right-hand side is 0 and the inequality holds trivially; or we have $f(0) = 1$ and $f(a) = 0$ for some $a \in \{1, 2, \dots, q-1\}$, in which case we get

$$\begin{aligned} \mathbb{E}\left[\sqrt{h_f}\right] &= 1 - p \\ &\geq \frac{1-p}{2} \cdot \mathbb{E}[f]\big(1 - \mathbb{E}[f]\big). \end{aligned}$$

We thus turn to the induction step. Suppose the desired statement holds for $n - 1$, and consider some function $f : \mathbb{F}_q^n \to \mathbb{F}_2$. We may assume that

$$\mathbb{E}\left[\sqrt{h_f}\right] \leq \frac{1-p}{2}, \tag{8}$$

as otherwise the desired claim is trivial. For each $a \in \mathbb{F}_q$, define the following function of $n - 1$ variables.

$$f_a(x) := f(xa_n).$$

For convenience, we will denote the expectation of each of these functions by $E_a := \mathbb{E}[f_a]$, where again the expectations are taken over the $p$-noisy distribution. By definition, we have

$$\mathbb{E}[f] = (1-p)E_0 + \frac{p}{q-1} \sum_{a \in \mathbb{F}_q \setminus \{0\}} E_a, \tag{9}$$

and thus

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &= \Big((1-p)E_0 + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big)\Big(1-(1-p)E_0 - \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big) \\ &= (1-p)E_0\Big(1-E_0+pE_0-\frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a)\Big) \\ &\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big(1-E_a+(1-\frac{p}{q-1})E_a-(1-p)E_0-\frac{p}{q-1}\sum_{b\in\mathbb{F}_q\setminus\{0,a\}} E_b\Big).\end{aligned}$$

Extracting from the expression above the terms corresponding to the variance of each $f_a$, we get

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &= (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a(1-E_a) \\ &\quad + (1-p)pE_0\Big(E_0-\frac{1}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big) \\ &\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big((1-\frac{p}{q-1})E_a-(1-p)E_0\Big) \\ &\quad - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q\setminus\{0\}\\ b\in\mathbb{F}_q\setminus\{0,a\}}} E_aE_b. \qquad (10)\end{aligned}$$

The first line in the equation above is the sum of the individual variances. We will now want to bound the contribution of the other terms. For this it will be useful to replace each factor of $1-p$ by a factor of $1-\frac{p}{q-1}$, so that we can complete the square. That is, we write the summands in the two middle lines of (10) as

$$\begin{aligned}(1-p)pE_0\Big(E_0-\frac{1}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big) &= (1-\frac{p}{q-1})\frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_0\big(E_0-E_a\big) \\ &\quad - (1-\frac{1}{q-1})\frac{p^2}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_0(E_0-E_a)\end{aligned}$$

and

$$\begin{aligned}\frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big((1-\frac{p}{q-1})E_a-(1-p)E_0\Big) &= \frac{p}{q-1}(1-\frac{p}{q-1})\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a\Big(E_a-E_0\Big) \\ &\quad + \sum_{a\in\mathbb{F}_q\setminus\{0\}} \frac{p^2}{q-1}(1-\frac{1}{q-1})E_aE_0.\end{aligned}$$

Combining the two equations above with (10), we get

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &= (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a(1-E_a)\\ &\quad + (1-\frac{p}{q-1})\frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\big(E_0-E_a\big)^2 - \frac{p^2(q-2)}{(q-1)^2}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_0(E_0-E_a)\\ &\quad + \sum_{b\in\mathbb{F}_q\setminus\{0\}}\frac{p^2(q-2)}{(q-1)^2}E_bE_0 - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q\setminus\{0\}\\ b\in\mathbb{F}_q\setminus\{0,a\}}} E_aE_b.\end{aligned}$$

Adding an artificial summation to the fourth and fifth sums above, we get

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &= (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a(1-E_a)\\ &\quad + (1-\frac{p}{q-1})\frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\big(E_0-E_a\big)^2\\ &\quad - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q\setminus\{0\}\\ b\in\mathbb{F}_q\setminus\{0,a\}}}\Big(E_0(E_0-E_a) - E_bE_0 + E_aE_b\Big).\end{aligned}$$

Defining the quantity

$$E_{\min} := \mathop{\mathbb{E}}_{x\sim p^{n-1}}\Big[\min_{a\in\mathbb{F}_q\setminus\{0\}}\big\{f(xa_n)\big\}\Big],$$

we can then bound the variance of $f$ by

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &\le (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a(1-E_a) + (1-\frac{p}{q-1})p\big(E_0-E_{\min}\big)^2\\ &\quad - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q\setminus\{0\}\\ b\in\mathbb{F}_q\setminus\{0,a\}}}(E_0-E_b)(E_0-E_a)\\ &\le (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}} E_a(1-E_a) + (1-\frac{p}{q-1})p\big(E_0-E_{\min}\big)^2. \qquad (11)\end{aligned}$$

Now that we have obtained a convenient expression for the right-hand side of our theorem's inequality, we turn to bounding the left-hand side. We note that since $f$ is monotone, we must have

$$\begin{aligned}\mathop{\mathbb{E}}_{z\sim p^n}\Big[\sqrt{h_f(z)}\Big] &= (1-p)\mathop{\mathbb{E}}_{x\sim p^{n-1}}\Big[\sqrt{h_{f_0}(x) + \mathbb{1}\big\{f(x0_n)=1, f(xa_n)=0 \text{ for some } a\in\mathbb{F}_q\setminus\{0\}\big\}}\Big]\\ &\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\mathop{\mathbb{E}}_{x\sim p^{n-1}}\Big[\sqrt{h_{f_a}(x)}\Big].\end{aligned}$$

Defining the function $f^-(x) := f(x0_n) - \min_{a\in\mathbb{F}_q\setminus\{0\}}\big\{f(xa_n)\big\}$, we then get

(12)
$$\mathop{\mathbb{E}}_{z\sim p^n}\left[\sqrt{h_f(z)}\right] = (1-p)\mathop{\mathbb{E}}_{x\sim p^{n-1}}\left[\sqrt{h_{f_0}(x)+f^-(x)}\right] + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\mathop{\mathbb{E}}_{x\sim p^{n-1}}\left[\sqrt{h_{f_a}(x)}\right].$$

But applying the Cauchy-Schwarz inequality $\mathbb{E}[\sqrt{gh}]^2 \le \mathbb{E}[g]\,\mathbb{E}[h]$ and the equality $(a+b)(a-b) = a^2 - b^2$, we have

$$\begin{aligned}\mathbb{E}[f^-]^2 &= \mathbb{E}\left[\big(\sqrt{h_{f_0}+f^-}-\sqrt{h_{f_0}}\big)^{\frac{1}{2}}\big(\sqrt{h_{f_0}+f^-}+\sqrt{h_{f_0}}\big)^{\frac{1}{2}}\right]^2\\ &\le \mathbb{E}\left[\sqrt{h_{f_0}+f^-}-\sqrt{h_{f_0}}\right]\mathbb{E}\left[\sqrt{h_{f_0}+f^-}+\sqrt{h_{f_0}}\right],\end{aligned}$$

where in the first line we used the fact that $f^-$ takes values in $\{0,1\}$, and thus $\sqrt{f^-(x)} = f^-(x)$ for all $x \in \mathbb{F}_2^{n-1}$. We can now bound the expected square root of $h_{f_0} + f^-$ by

$$\begin{aligned}\mathbb{E}\left[\sqrt{h_{f_0}+f^-}\right] &\ge \mathbb{E}\left[\sqrt{h_{f_0}}\right] + \frac{\mathbb{E}[f^-]^2}{\mathbb{E}\left[\sqrt{h_{f_0}+f^-}+\sqrt{h_{f_0}}\right]}\\ &\ge \mathbb{E}\left[\sqrt{h_{f_0}}\right] + \frac{\mathbb{E}[f^-]^2}{\mathbb{E}\left[f^-\right]+2\,\mathbb{E}\left[\sqrt{h_{f_0}}\right]},\end{aligned}$$

where in the last line we used the fact that $\sqrt{a^2+b^2} \le \sqrt{(a+b)^2} = a+b$, with $a = \sqrt{h_{f_0}}$ and $b = \sqrt{f^-}$. Combining the inequality above with equation (12), we get

$$\begin{aligned}\mathbb{E}\left[\sqrt{h_f}\right] &\ge (1-p)\,\mathbb{E}\left[\sqrt{h_{f_0}}\right] + (1-p)\frac{(E_0-E_{\min})^2}{\mathbb{E}\left[f^-\right]+2\,\mathbb{E}\left[\sqrt{h_{f_0}}\right]} + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\mathbb{E}\left[\sqrt{h_{f_a}}\right]\\ &\ge (1-p)\,\mathbb{E}\left[\sqrt{h_{f_0}}\right] + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\mathbb{E}\left[\sqrt{h_{f_a}}\right] + \frac{1-p}{2}(E_0-E_{\min})^2,\end{aligned}$$

where in the last line we used assumption (8) and equation (12) to get $\mathbb{E}\left[\sqrt{h_{f_0}}\right] \le \frac{1}{2}$. Applying our induction hypothesis to the functions $\{f_a\}_{a\in\mathbb{F}_q}$, we then have

$$\mathbb{E}\left[\sqrt{h_f}\right] \ge \frac{(1-p)^2}{2}E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q\setminus\{0\}}\frac{1-p}{2}E_a(1-E_a) + \frac{1-p}{2}(E_0-E_{\min})^2.$$

Combining this with equation (11), we indeed get

$$\mathbb{E}\left[\sqrt{h_f}\right] \ge \frac{1-p}{2}\cdot\mathbb{E}[f]\big(1-\mathbb{E}[f]\big).$$

□

We now denote the minimum non-zero value of $h_f(x)$ by

$$\Delta_f := \min\big\{h_f(z) : z \in \mathbb{F}_q^n \text{ such that } h_f(z) \neq 0\big\}.$$

The expectation of $h_f$ can then be bounded as follows.

**Theorem 8.** *For any monotone decreasing function $f : \mathbb{F}_q^n \to \mathbb{F}_2$ and any noise parameter $p \in [0,1]$, we have*

$$\mathbb{E}\left[h_f\right] \geq \frac{1-p}{2}\sqrt{\Delta_f} \cdot \mathbb{E}[f]\big(1 - \mathbb{E}[f]\big),$$

*where all the expectations are taken with respect to the q-ary p-noisy distribution.*

*Proof.* By Theorem 7 and the Cauchy-Schwarz inequality $\mathbb{E}[\sqrt{g_1 g_2}]^2 \leq \mathbb{E}[g_1]\,\mathbb{E}[g_2]$, we have

$$\begin{aligned}\frac{1-p}{2} \cdot \mathbb{E}[f]\big(1 - \mathbb{E}[f]\big) &\leq \mathbb{E}\left[\sqrt{h_f}\right] \\ &\leq \sqrt{\mathbb{E}\left[h_f\right] \Pr_x\left[h_f(x) \neq 0\right]}.\end{aligned}$$

By definition of $\Delta_f$, we then get

$$\begin{aligned}\frac{1-p}{2} \cdot \mathbb{E}[f]\big(1 - \mathbb{E}[f]\big) &\leq \sqrt{\mathbb{E}\left[h_f\right] \cdot \frac{\mathbb{E}\left[h_f\right]}{\Delta_f}} \\ &= \frac{1}{\sqrt{\Delta_f}}\,\mathbb{E}\left[h_f\right].\end{aligned}$$

□

## 4. Sharp Transition of the Decoding Error Probability

In this section, we will show that the decoding error probability of a linear code $C \subseteq \mathbb{F}_q^n$ over the channel $\mathrm{qS}C_p$ transitions rapidly from 0 to 1 (as a function of $p$). For $q = 2$, this was proven by Tillich and Zémor in [Zém93, TZ00]. We follow a similar approach, and generalize their results to arbitrary field size $q$. The first building block of our argument is Russo's Lemma, which for $q = 2$ first appeared in [Mar74, Rus82]. Over arbitrary field size $q$ and for our particular definition of monotonicity, it generalizes as follows.

**Lemma 9.** *Let $f : \mathbb{F}_q^n \to \{0,1\}$ be a monotone decreasing function. Then we have*

$$\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}[f(z)] \leq -\frac{1}{q-1}\mathop{\mathbb{E}}_{z\sim p}[h_f(z)].$$

*Proof.* We think of the parameter $p$ as a vector $(p_1, p_2, \ldots, p_n)$ with $p_i = p$ for all $i \in [n]$. By definition, we have

$$\begin{aligned}\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}\big[f(z)\big] &= \sum_{i\in[n]} \frac{d}{dp_i}\mathop{\mathbb{E}}_{z\sim p}\big[f(z)\big] \\ &= \sum_{i\in[n]} \frac{d}{dp_i}\mathop{\mathbb{E}}_{z\sim p}\Big[(1-p_i)f(z0_i) + p_i \mathop{\mathbb{E}}_{a\in\{1,2,\ldots,q-1\}}\big[f(za_i)\big]\Big] \\ &= \sum_{i\in[n]} \mathop{\mathbb{E}}_{\substack{z\sim p\\ a\neq 0}}\big[-f(z0_i) + f(za_i)\big].\end{aligned}$$

Since $f$ is monotone decreasing, we must always have $f(z0_i) \geq f(za_i)$, and thus

$$\frac{d}{dp} \mathop{\mathbb{E}}_{z\sim p}\left[f(z)\right] = -\sum_{i\in[n]} \mathop{\mathbb{E}}_{\substack{z\sim p\\ a\neq 0}}\left[\left|f(z0_i) - f(za_i)\right|\right].$$

Considering only the vectors $z$ with $z_i = 0$, we then get

$$\begin{aligned}\frac{d}{dp} \mathop{\mathbb{E}}_{z\sim p}\left[f(z)\right] &\leq -\sum_{i\in[n]} \mathop{\mathbb{E}}_{z\sim p}\left[\mathbb{1}\{z_i = 0\} \cdot \mathop{\mathbb{E}}_{a\neq 0}\left[|f(z) - f(za_i)|\right]\right]\\ &\leq -\sum_{i\in[n]} \mathop{\mathbb{E}}_{z\sim p}\left[\mathbb{1}\{z_i = 0\} \cdot \frac{1}{q-1}\mathbb{1}\left\{\exists a \neq 0 \text{ s.t. } |f(z) - f(za_i)| = 1\right\}\right]\\ &= -\frac{1}{q-1} \mathop{\mathbb{E}}_{z\sim p}[h_f(z)].\end{aligned}$$

□

From Lemma 9, it is clear that for any monotone function $f : \mathbb{F}_q^n \to \mathbb{F}_2$, any lower bound on $\mathbb{E}[h_f]$ will yield an upper bound on the width of the threshold of $\mathbb{E}[f]$. We thus turn to proving bounds on $\mathbb{E}[h_f]$, for $f$ the indicator function of a successful decoding. For this we will need the following helpful lemma. Recall that for any vectors $a, b \in \mathbb{F}_q^n$, we denote by $d(a, b)$ the Hamming weight of $a - b$.

**Lemma 10.** *Suppose $z \in \mathbb{F}_q^n$ and $c \in C$ satisfy*

$$d(z, 0) \leq d(z, c).$$

*Then we must have*

$$\left|\mathrm{supp}(c) \setminus \mathrm{supp}(z)\right| \geq \frac{d_{\min}(C)}{q} - d(z, c) + \min_{c'\in C}\{d(z, c')\}.$$

*Proof.* For notational simplicity, we define the set

$$S := \mathrm{supp}(c) \setminus \mathrm{supp}(z)$$

and the slack quantity

$$\nu := d(z, c) - \min_{c'\in C}\{d(z, c')\}$$

Our goal is to show that $|S| \geq \frac{d_{\min}}{q} - \nu$. We first note that since $d(z, c) \geq d(z, 0)$, we must have

$$|S| \geq \left|\{i \in \mathrm{supp}(c) \cap \mathrm{supp}(z) : c_i = z_i\}\right|. \tag{13}$$

We also note that

$$\left|\{i \in \mathrm{supp}(z) \cap \mathrm{supp}(c) : c_i = z_i\}\right| \geq \frac{1}{q-1}\left|\mathrm{supp}(c) \cap \mathrm{supp}(z)\right| - \nu. \tag{14}$$

This is because for all $\alpha \in \{1, 2, \ldots, q-1\}$, we have

$$d(z, \alpha c) = \left|\mathrm{supp}(z) \setminus \mathrm{supp}(c)\right| + \left|\mathrm{supp}(c) \setminus \mathrm{supp}(z)\right| + \left|\{i \in \mathrm{supp}(z) \cap \mathrm{supp}(c) : \alpha c_i \neq z_i\}\right|,$$

while by averaging there must be some $\alpha \in \{1, 2, \ldots, q-1\}$ such that

$$\big|\{i \in \operatorname{supp}(z) \cap \operatorname{supp}(c) : \alpha c_i = z_i\}\big| \geq \frac{1}{q-1}\big|\operatorname{supp}(z) \cap \operatorname{supp}(c)\big|.$$

Since $d(z, c) \leq d(z, \alpha c) + \nu$ for every codeword $\alpha c$, we then get equation (14). With respect to the minimum distance of our code $C$, this gives us

$$\begin{aligned} d_{\min} &\leq \operatorname{wt}(c) \\ &= \big|\operatorname{supp}(c) \setminus \operatorname{supp}(z)\big| + \big|\operatorname{supp}(c) \cap \operatorname{supp}(z)\big| \\ &\leq |S| + (q-1)(|S| + \nu) \\ &\leq q|S| + q\nu, \end{aligned}$$

where in the third line we used equations (13) and (14). □

We are now ready to prove our bound on $\mathbb{E}[h_f]$, for $f$ the indicator function of a successful decoding. Consider the following total order $\prec$ on $\mathbb{F}_q^n$. If $\operatorname{wt}(a) < \operatorname{wt}(b)$, then $a \prec b$. If $\operatorname{wt}(a) = \operatorname{wt}(b)$ and the support of $a$ comes after the support of $b$ in the lexicographic order, then $a \prec b$. For completeness' sake (this last point will not appear in our analysis), if $a$ and $b$ have the same support and $a$ comes after $b$ in the full lexicographic order (i.e. the lexicographic order with order $0 < 1 < 2 < \ldots < q-1$ over $\mathbb{F}_q$), then we say $a \prec b$. Consider the max-likelihood decoder $D^* : \mathbb{F}_q^n \to C$ defined by

$$D^*(z) := \min_{c \in C}\{z - c\}, \tag{15}$$

where the comparisons between vectors are taken with respect to the total order $\prec$ . For each codeword $c \in C$, we define the decoding region of $c$ as follows.

$$\Omega_c := \{z \in \mathbb{F}_q^n : D^*(z) = c\}.$$

**Claim 11.** *For all $c \in C$, we have*

$$\Pr_{z \sim p}[D^*(z + c) = c] = \Pr_{z \sim p}[z \in \Omega_0].$$

*Proof.* It is clear that

$$\Pr_{z \sim p}[D^*(z) = 0] = \Pr_{z \sim p}[z \in \Omega_0].$$

Thus it will suffice to show that for any codeword $c \in C$, the map $z \mapsto z + c$ is a bijection between $\Omega_0$ and $\Omega_c$. But this is indeed the case, as by linearity of $C$ we have

$$\begin{aligned} z \in \Omega_0 &\iff z \prec z - c' \text{ for all } c' \in C \\ &\iff z + c - c \prec z + c - c' \text{ for all } c' \in C \\ &\iff z + c \in \Omega_c. \end{aligned}$$

□

For simplicity, when looking at the 0 codeword we will drop the subscript and write

$$\Omega := \Omega_0.$$

We will also abuse notation and write $\Delta_\Omega$ and $h_\Omega$ to mean $\Delta_{\mathbb{1}_\Omega}$ and $h_{\mathbb{1}_\Omega}$ respectively.

**Lemma 12.** *Consider any linear code $C \subseteq \mathbb{F}_q^n$. Its corresponding decoding region $\Omega$ satisfies*

$$\Delta_\Omega \geq \frac{d_{\min}}{q} - 3,$$

*where $d_{\min}$ is the minimum distance of $C$.*

*Proof.* Consider any $z \in \mathbb{F}_q^n$ with $h_\Omega(z) \neq 0$. By definition, the following two conditions must hold.

(i) $D^*(z) = 0$,

(ii) There exist a codeword $c \in C$ and a coordinate $i \in [n]$ such that $D^*(zc_i) = c$.

Our goal will be to show that there are at least $\frac{d_{\min}}{q} - 3$ choices of coordinates $i$ where $z_i = 0$ and point (ii) above holds. We note that points (i) and (ii) imply that

$$d(z, 0) \leq d(z, c) \leq d(z, 0) + 2. \tag{16}$$

By Lemma 10, we must then have

$$\left|\mathrm{supp}(c) \setminus \mathrm{supp}(z)\right| \geq \frac{d_{\min}}{q} - 2. \tag{17}$$

We now consider two separate cases, depending on the weight of $z - c$.

**Case 1:** $d(z, c) \in \left\{d(z, 0), d(z, 0) + 1\right\}$. Then for every $j \in \mathrm{supp}(c) \setminus \mathrm{supp}(z)$, we have

$$\begin{aligned} d(zc_j, c) &= d(z, c) - 1 \\ &\leq d(z, 0) \\ &= d(zc_j, 0) - 1, \end{aligned}$$

and thus $zc_j \notin \Omega$. By equation 17, we thus have $h_\Omega(z) \geq \frac{d_{\min}}{q} - 2$.

**Case 2:** $d(z, c) = d(z, 0) + 2$. Then for every $j \in \mathrm{supp}(c) \setminus \mathrm{supp}(z)$, we have

$$\begin{aligned} d(zc_j, c) &= d(z, c) - 1 \\ &= d(z, 0) + 1 \\ &= d(zc_j, 0). \end{aligned} \tag{18}$$

We want to show that for all but one choices of $j \in \mathrm{supp}(c) \setminus \mathrm{supp}(z)$, we have

$$zc_j - c \prec zc_j,$$

or equivalently that the support of $zc_j - c$ comes after the support of $zc_j$ in the lexicographic order. We note that by point (ii) above, there exists a coordinate $i \in [n]$ such that $\mathrm{supp}(zc_i - c)$ comes after $\mathrm{supp}(zc_i)$ in the lexicographic order. But this means that $\mathrm{supp}(z - c)$ must come after $\mathrm{supp}(z)$ in the lexicographic order. Define the coordinate

$$\begin{aligned} j^* &:= \min\left\{j \in \mathrm{supp}(z - c) \setminus \mathrm{supp}(z)\right\} \\ &= \min\left\{j \in \mathrm{supp}(c) \setminus \mathrm{supp}(z)\right\}, \end{aligned}$$

where the minimum is taken over the standard order $1 < 2 < 3 \ldots < n$. Then for any coordinate $j > j^*$, $\mathrm{supp}(zc_j - c)$ must come after $\mathrm{supp}(zc_j)$ in the lexicographic order. Combining this with equation (18), we get that for every coordinate $j \in \mathrm{supp}(c) \backslash \mathrm{supp}(z)$, $j \neq j^*$, we have

$$zc_j - c \prec zc_j.$$

By equation (17), there are at least $\frac{d_{\min}}{q} - 3$ such coordinates. Thus we indeed have

$$h_\Omega(z) \geq \frac{d_{\min}}{q} - 3.$$

□

Combining our results from Sections 3 and 4, we get the following bound on the derivative of the decoding success probability.

**Lemma 13.** *Consider any linear code $C \subseteq \mathbb{F}_q^n$ with minimum distance $d_{\min} \geq 4q$, and any noise parameter $p \in [0, 1]$. The decoding region $\Omega$ for the code $C$ satisfies*

$$\frac{d}{dp} \Pr[z \in \Omega] \leq -\frac{1-p}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}} \Pr[z \in \Omega]\Big(1 - \Pr[z \in \Omega]\Big),$$

*where all the probabilities are taken with respect to the q-ary p-biased distribution $z \sim p$.*

*Proof.* By definition, the decoding region $\Omega$ is monotone decreasing. By Lemma 9 and Theorem 8, we then get

$$\begin{aligned}\frac{d}{dp} \Pr_{z\sim p}[z \in \Omega] &\leq -\frac{1}{q-1} \mathop{\mathbb{E}}_{z\sim p}[h_\Omega(z)] \\ &\leq -\frac{1-p}{2(q-1)} \sqrt{\Delta_\Omega} \Pr_{z\sim p}[z \in \Omega]\Big(1 - \Pr_{z\sim p}[z \in \Omega]\Big).\end{aligned}$$

Applying Lemma 12, we must thus indeed have

$$\begin{aligned}\frac{d}{dp} \Pr[z \in \Omega] &\leq -\frac{1-p}{2(q-1)} \sqrt{\frac{d_{\min}}{q} - 3} \Pr[z \in \Omega]\Big(1 - \Pr[z \in \Omega]\Big) \\ &\leq -\frac{1-p}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}} \Pr[z \in \Omega]\Big(1 - \Pr[z \in \Omega]\Big).\end{aligned}$$

□

We are now ready to prove our sharp transition result. For convenience, given a fixed code $C$, we denote the probability of a decoding success by

$$g(p) := \Pr_{z\sim p}[z \in \Omega].$$

The theorem below shows that the function $g$ transitions very rapidly from 1 to 0. In spirit, it says that for any noise parameters $p_0 < p_1$ that aren't extremely close to each other, either $g(p_0) \approx 1$ or $g(p_1) \approx 0$.

**Theorem 14.** *Consider any linear code $C \subseteq \mathbb{F}_q^n$ with minimum distance $d_{\min} \geq 4q$, and any noise parameters $0 \leq p_0 \leq p_1 \leq 1$. Then we have*

$$g(p_1)\Big(1 - g(p_0)\Big) \leq e^{-\frac{1-p_1}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}}(p_1 - p_0)},$$

*where $d_{\min}$ denotes the minimum distance of the code $C$.*

*Proof.* Define the function

$$G(p) := \ln \frac{g(p)}{1 - g(p)}.$$

Then by Lemma 13, we have

$$\begin{aligned} \frac{dG}{dp} &= \frac{1}{g(p)\Big(1 - g(p)\Big)} \cdot \frac{dg}{dp} \\ &\leq -\frac{1-p}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}}. \end{aligned}$$

By the fundamental theorem of calculus, we then have

$$\begin{aligned} G(p_0) - G(p_1) &= -\int_{p_0}^{p_1} \frac{dG}{dp} dp \\ &\geq \frac{1-p_1}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}}\Big(p_1 - p_0\Big). \end{aligned}$$

By definition of $G$, we thus get

$$\begin{aligned} g(p_1)\Big(1 - g(p_0)\Big) &\leq \frac{g(p_1)}{1 - g(p_1)} \cdot \frac{1 - g(p_0)}{g(p_0)} \\ &= e^{G(p_1) - G(p_0)} \\ &\leq e^{-\frac{1-p_1}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}}(p_1 - p_0)}. \end{aligned}$$

□

## 5. Proof of Main Results

In this section, we use our results from Section 4 to prove Theorems 1 and 2. We first prove a generalization of Theorem 2 below. Taking $\delta = \frac{4q^{3/2}}{(1-p)\sqrt{d_{\min}}} \ln(nL)$ in the following theorem gives Theorem 2.

**Theorem 15.** *Let $C \subseteq \mathbb{F}_q^n$ be a linear, $(p, L)$-list decodable code with minimum distance $d_{\min} \geq 4q$. Then for any $\delta > 0$ and any $c \in C$, we have*

$$\Pr_{z \sim p - n^{-\frac{1}{4}} - \delta}\Big[D^*(c + z) = c\Big] \geq 1 - 2Le^{-\frac{1-p}{4}\frac{\sqrt{d_{\min}}}{q^{3/2}}\delta}.$$

*Proof.* Define the following decoder $D : \mathbb{F}_q^n \to C$. Upon seeing a message $m \in \mathbb{F}_q^n$, the decoder $D$ finds all codewords $c \in C$ that satisfy $\mathrm{wt}(m - c) \leq pn$, and outputs one of them uniformly at random. The probability of success of this decoder under errors of probability $p - n^{-\frac{1}{4}}$ is bounded by

$$\begin{aligned}\Pr_{z \sim p - n^{-\frac{1}{4}}}[D(c+z) = c] &\geq \Pr_{z \sim p - n^{-\frac{1}{4}}}[\mathrm{wt}(z) \leq pn] \Pr_{z \sim p - n^{-\frac{1}{4}}}[D(c+z) = c \big| \mathrm{wt}(z) \leq pn] \\ &\geq (1 - e^{-2\sqrt{n}}) \cdot \frac{1}{L} \\ &\geq \frac{1}{2L},\end{aligned}$$

where in the second inequality we used Hoeffding's inequality (Lemma 4) for the first term, and the fact that $C$ is $(p, L)$-decodable for the second term. Now by Fact 6, the max-likelihood decoder $D^*$ can only have a better decoding probability than $D$, so we have

$$\Pr_{z \sim p - n^{-\frac{1}{4}}}[D^*(c+z) = c] \geq \frac{1}{2L}. \tag{19}$$

By Theorem 14 and Claim 11, we then get

$$\Pr_{z \sim p - n^{-\frac{1}{4}} - \delta}[D^*(c+z) = c] \geq 1 - 2Le^{-\frac{1-p}{4} \cdot \frac{\sqrt{d_{\min}}}{q^{3/2}} \delta}.$$

□

We then turn to proving Theorem 1 from Theorem 2.

**Theorem 1.** *Let $\{C_n \subseteq \mathbb{F}_q^n\}$ be a family of linear codes with rate $1 - h_q(p)$, and suppose $\{C_n\}$ achieves list-decoding capacity. If $d_{\min}(C_n) = \omega\Big(\frac{q^3}{(1-p)^2}\Big)$, then $\{C_n\}$ achieves capacity over the $q$-ary symmetric channel.*

*Proof.* By Definition 2, there exists a function $\epsilon(n) = o(1)$ such that each $C_n$ is $\Big(p - \epsilon_n, \frac{(1-p)^2 d_{\min}}{q^3}\Big)$-list decodable. Applying Theorem 15 with $\delta = \Big(\frac{q^3}{(1-p)^2 d_{\min}}\Big)^{\frac{1}{4}} = o(1)$, we then get

$$\begin{aligned}\Pr_{z \sim p - \epsilon_n - n^{-\frac{1}{4}} - \delta}\Big[D^*(c+z) = c\Big] &\geq 1 - \frac{2(1-p)^2 d_{\min}}{q^3} \cdot e^{-\frac{1}{4}\left(\frac{(1-p)^2 d_{\min}}{q^3}\right)^{\frac{1}{4}}} \\ &\geq 1 - o(1).\end{aligned}$$

□

## Acknowledgments

We thank Hervé Chabanne, Sivakanth Gopi, Anup Rao and Gilles Zémor for useful discussions. The work of Francisco Pernice was supported in part by an MIT Jacobs Presidential Fellowship. The work of Oscar Sprumont was supported in part by NSF CCF-2131899, NSF CCF-1813135 and Anna Karlin's Bill and Melinda Gates Endowed

Chair. The work of Mary Wootters was supported in part by NSF CCF-2231157 and CCF-2133154.

## Appendix A. Erasure Channel

In this section, we recall and prove Theorem 3.

**Theorem 3.** *Let $C \subseteq \mathbb{Z}_q^n$ be a $(p, L)$-list decodable code with minimum distance $\omega(\log L)$. Then $C$ admits reliable communication on the $qSC_{p'}$ for $p' = p - \frac{\log n}{\sqrt{n}}$.*

*Proof.* Fix any arbitrary sent codeword $c \in C$. For any erasure pattern $z \in \{0,1\}^n$, we define the set of codewords that could be mistaken for $c$ as

$$S(z) := \Big\{ c' \in C : c|_{\{i\in[n]:z_i=0\}} = c'|_{\{i\in[n]:z_i=0\}} \Big\}.$$

Our goal will be to show that with high probability over the choice of $z$, $c$ is the only element in $S(z)$. We first note that by Hoeffding's inequality (Lemma 4), we have

$$\begin{aligned} \Pr_{z\sim p-\frac{\log n}{\sqrt{n}}}[\mathrm{wt}(z) > pn] &< e^{-2\log^2 n} \\ &= o(1). \end{aligned} \tag{20}$$

We also note that by our assumption on the minimal distance of $C$, the probability that any $c' \in C$ be in $S(z)$ can be bounded by

$$\begin{aligned} \Pr_{z\sim p-\frac{\log n}{\sqrt{n}}}[c' \in S(z)] &= p^{\mathrm{wt}(c+c')} \\ &\leq p^{-\omega(\log nL)} \\ &\leq o(\frac{1}{L}). \end{aligned} \tag{21}$$

But since $C$ is $(p, L)$-list decodable, there are at most $L$ codewords $c' \in C$ satisfying $\mathrm{wt}(c + c') \leq pn$. Combining equations (20) and 21) and applying the union bound, we thus get

$$\begin{aligned} \Pr_{z\sim p-\frac{\log n}{\sqrt{n}}}\big[|S(z)| > 1\big] &\leq \Pr_{z\sim p-\frac{\log n}{\sqrt{n}}}[\mathrm{wt}(z) > pn] + \sum_{c'\in C:\mathrm{wt}(c+c')\leq pn} \Pr_{z\sim p-\frac{\log n}{\sqrt{n}}}[c' \in S(z)] \\ &\leq o(1) + L \cdot o(\frac{1}{L}) \\ &\leq o(1). \end{aligned}$$

□

## Appendix B. Necessity of distance condition

We construct a linear code $C \subseteq \mathbb{F}_2^n$ of small distance which achieves list-decoding capacity but not q-SC$_p$ capacity. We start with a linear code $C' \subseteq \mathbb{F}_2^n$ which is list-decoding capacity achieving with list size $L' = L_n'$. Then there is $\epsilon' = \epsilon_n' = o(1)$ so that $C'$ is $(p-\epsilon_n', L_n')$-list decodable. Consider the code $C = \mathrm{span}\{e_1, C'\}$, where $(e_1)_i = 1$ if $i = 1$ and $(e_1)_i = 0$ otherwise. Fix a $z \in \mathbb{F}_2^n$, and suppose $C' \cap B_{(p-2\epsilon')n+1}(z) = \{c_1, \ldots, c_t\}$.

Since $(p-2\epsilon')n+1 < (p-\epsilon')n$ for all $n$ large enough, we have $t \leq L$ for all $n$ large enough. But then $C \cap B_{(p-2\epsilon')n}(z) \subseteq \{c_1, \ldots, c_t, c_1+e_1, \ldots, c_t+e_1\}$. Hence, setting $\epsilon = 2\epsilon'$ and $L = 2L'$, we conclude that $C$ also achieves list-decoding capacity.

However, it's clear that $C$ cannot achieve q-SC capacity. Indeed, if we send some $c \in C$ and the first bit gets corrupted, we can no longer distinguish between $c$ and $c+e_1$. In other words, the probability of decoding error is bounded below by approximately $p$.

## Appendix C. Discussion of the Proofs in [KCC10]

In this section, we discuss the proof of [KCC10] for the claim that the decoding success probability of any $q$-ary linear code with large minimum distance transitions rapidly from $1-o(1)$ to $o(1)$. As far as we can tell, the following two issues suggest that their proof may not be complete. We thank Hervé Chabanne for useful discussions on this subject.

The first issue is that the arguments of [KCC10] rely on the following definition of monotonicity: a function $f : \mathbb{F}_q^n \to \{0,1\}$ is deemed *monotone* if whenever the support of $x \in \mathbb{F}_q^n$ is a subset of the support of $y \in \mathbb{F}_q^n$ and $f(x) = 1$, then we must also have $f(y) = 1$. This definition allows for an easier adaptation of $\mathbb{F}_2$ isoperimetric inequalities, and the authors of [KCC10] prove sharp thresholds results for functions that are monotone in this sense. They then claim that the non-decoding region of a linear code $C \subseteq \mathbb{F}_q^n$ satisfies this version of monotonicity. Unfortunately, this is not true. For instance, consider the case where $q = 3$ and our code is the span of the all-1 vector. Then the error string $x$ that has a 1 in the first $\frac{n}{2}+1$ coordinates and 0 everywhere else leads to a decoding failure, while the error string $y$ that has a 1 in the first $\frac{n}{4}+1$ coordinates, a 2 in the next $\frac{n}{4}+1$ coordinates, and a 0 everywhere else leads to a decoding success.

The second issue in [KCC10] has to do with the neighborhood of boundary points. A critical part of their argument is their claim that any vector that is in the non-decoding region $U_0 := \left\{x \in \mathbb{F}_q^n : \exists c \in C \setminus \{0\} : d(x,c) \leq d(x,0)\right\}$ and has at least one neighbor outside of $U_0$ must have at least $\frac{d}{2}$ such neighbors, where $d$ is the minimum distance of the code. This is true for $q = 2$, but it does not hold for larger alphabets. For example, suppose we are working with a field $\mathbb{F}_q$ for some $q \geq n$, and suppose our code is again the span of the all-1 vector. Now suppose that the error string $x$ is $x = (0,1,1,2,3,4,\ldots,n-3,n-2)$. This vector is in the non-decoding region $U_0$, as it is closer to the vector $(1,1,1,...,1)$ than to the 0 vector. It also has a neighbor outside of $U_0$, for instance the vector $(0,0,1,2,3,4,5,...,n-4,n-3,n-2)$. However it only has 2 such neighbors: the ones obtained by setting the second coordinate or the third coordinate of $x$ to 0. Setting any other coordinate to 0 would yield a vector that is still in $U_0$, as it is equally far from the 0 vector and the all-1 vector.

## References

[AGL24] Omar Alrabiah, Venkatesan Guruswami, and Ray Li. Randomly punctured reed-solomon codes achieve list-decoding capacity over linear-sized fields. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory* *of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1458–1469. ACM, 2024.

[Ari09] Erdal Arikan. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory*, 55(7):3051–3073, 2009.

[AS23] Emmanuel Abbe and Colin Sandon. A proof that reed-muller codes achieve shannon capacity on symmetric channels. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 177–193. IEEE, 2023.

[ASSY23] Emmanuel Abbe, Ori Sberlo, Amir Shpilka, and Min Ye. Reed-muller codes. *Foundations and Trends in Communications and Information Theory*, 20(1–2):1–156, 2023.

[ASW15] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed-muller codes for random erasures and errors. *IEEE Trans. Inf. Theory*, 61(10):5229–5252, 2015.

[BDGZ23] Joshua Brakensiek, Manik Dhar, Sivakanth Gopi, and Zihan Zhang. AG codes achieve list decoding capacity over contant-sized fields. *CoRR*, abs/2310.12898, 2023.

[BGM23] Joshua Brakensiek, Sivakanth Gopi, and Visu Makam. Generic reed-solomon codes achieve list-decoding capacity. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1488–1501. ACM, 2023.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[BST24] Amit Berman, Yaron Shany, and Itzhak Tamo. Explicit subcodes of reed-solomon codes that efficiently achieve list decoding capacity. *CoRR*, abs/2401.15034, 2024.

[CZ24] Yeyuan Chen and Zihan Zhang. Explicit folded reed-solomon and multiplicity codes achieve relaxed generalized singleton bound. arXiv preprint, 2024, 2024.

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

[For66] George D. Forney. Concatenated codes. *MIT Press*, 1966.

[Gal62] Robert G. Gallager. Low-density parity-check codes. *IRE Trans. Inf. Theory*, 8(1):21–28, 1962.

[GHK10] Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 409–416. ACM, 2010.

[GLM$^+$22] Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear codes. *IEEE Trans. Inf. Theory*, 68(2):923–939, 2022.

[GR08] Venkatesan Guruswami and Atri Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Trans. Inf. Theory*, 54(1):135–150, 2008.

[GR10] Venkatesan Guruswami and Atri Rudra. The existence of concatenated codes list-decodable up to the hamming bound. *IEEE transactions on information theory*, 56(10):5195–5206, 2010.

[GR22] Zeyu Guo and Noga Ron-Zewi. Efficient list-decoding with constant alphabet and list sizes. *IEEE Trans. Inf. Theory*, 68(3):1663–1682, 2022.

[GRS23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/web-coding-book.pdf`, 2023.

[GST21] Eitan Goldberg, Chong Shangguan, and Itzhak Tamo. Singleton-type bounds for list-decoding and list-recovery, and related results. *CoRR*, abs/2112.05592, 2021.

[GW13] Venkatesan Guruswami and Carol Wang. Linear-algebraic list decoding for variants of reed-solomon codes. *IEEE Trans. Inf. Theory*, 59(6):3257–3268, 2013.

[GX13] Venkatesan Guruswami and Chaoping Xing. List decoding reed-solomon, algebraic-geometric, and gabidulin subcodes up to the singleton bound. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 843–852. ACM, 2013.

[GX22] Venkatesan Guruswami and Chaoping Xing. Optimal rate list decoding over bounded alphabets using algebraic-geometric codes. *J. ACM*, 69(2):10:1–10:48, 2022.

[GZ23] Zeyu Guo and Zihan Zhang. Randomly punctured reed-solomon codes achieve the list decoding capacity over polynomial-size alphabets. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 164–176. IEEE, 2023.

[HRW17] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes & applications. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 204–215. IEEE Computer Society, 2017.

[KCC10] Bruno Kindarji, Gérard D. Cohen, and Hervé Chabanne. On the threshold of maximum-distance separable codes. In *IEEE International Symposium on Information Theory, ISIT 2010, June 13-18, 2010, Austin, Texas, USA, Proceedings*, pages 1163–1167. IEEE, 2010.

[KCP16] Santhosh Kumar, A. Robert Calderbank, and Henry D. Pfister. Beyond double transitivity: Capacity-achieving cyclic codes on erasure channels. In *2016 IEEE Information Theory Workshop, ITW 2016, Cambridge, United Kingdom, September 11-14, 2016*, pages 241–245. IEEE, 2016.

[KKM$^+$16] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Sasoglu, and Rüdiger L. Urbanke. Reed-muller codes achieve capacity on erasure channels. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 658–669. ACM, 2016.

[Kop15] Swastik Kopparty. List-decoding multiplicity codes. *Theory Comput.*, 11:149–182, 2015.

[KRSW18] Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved decoding of folded reed-solomon and multiplicity codes. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 212–223. IEEE Computer Society, 2018.

[KRU13] Shrinivas Kudekar, Tom Richardson, and Rüdiger L. Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Trans. Inf. Theory*, 59(12):7761–7813, 2013.

[LMS$^+$97] Michael Luby, Michael Mitzenmacher, Mohammad Amin Shokrollahi, Daniel A. Spielman, and Volker Stemann. Practical loss-resilient codes. In Frank Thomson Leighton and Peter W. Shor, editors, *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 150–159. ACM, 1997.

[LW21] Ray Li and Mary Wootters. Improved list-decodability of random linear binary codes. *IEEE Trans. Inf. Theory*, 67(3):1522–1536, 2021.

[Mar74] Grigory A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problems of Information Transmission*, 10(2):101–108, 1974.

[MRR$^+$20] Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. LDPC codes achieve list decoding capacity. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 458–469. IEEE, 2020.

[MS77] Florence MacWilliams and Neil Sloane. *The theory of error correcting codes.* North-Holland Publishing Company, 1977.

[RP24] Galen Reeves and Henry D. Pfister. Reed-muller codes on BMS channels achieve vanishing bit-error probability for all rates below capacity. *IEEE Trans. Inf. Theory*, 70(2):920–949, 2024.

[RU10] Atri Rudra and Steve Uurtamo. Two theorems on list decoding. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, pages 696–709. Springer, 2010.

[Rus82] Lucio Russo. An approximate zero-one law. *Probability Theory and Related Fields*, 61(1):129–139, 1982.

[Sha48] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.

[Sri24] Shashank Srivastava. Improved list size for folded reed-solomon codes. *arXiv preprint arXiv:2410.09031*, 2024.

[Tal93] Michel Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric and Functional Analysis*, 3(3):295–314, 1993.

[Tam24] Itzhak Tamo. Tighter list-size bounds for list-decoding and recovery of folded reed-solomon and multiplicity codes. *IEEE Transactions on Information Theory*, pages 1–1, 2024.

[TZ00] Jean-Pierre Tillich and Gilles Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Comb. Probab. Comput.*, 9(5):465–479, 2000.

[TZ04] Jean-Pierre Tillich and Gilles Zémor. The gaussian isoperimetric inequality and decoding error probabilities for the gaussian channel. *IEEE Trans. Inf. Theory*, 50(2):328–331, 2004.

[Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 853–860. ACM, 2013.

[Woz58] John M. Wozencraft. List decoding. *Quarterly Progress Report, Research Laboratory of Electronics*, pages 90–95, 1958.

[Zém93] Gilles Zémor. Threshold effects in codes. In Gérard D. Cohen, Simon Litsyn, Antoine Lobstein, and Gilles Zémor, editors, *Algebraic Coding, First French-Israeli Workshop, Paris, France, July 19-21, 1993, Proceedings*, volume 781 of *Lecture Notes in Computer Science*, pages 278–286. Springer, 1993.

[ZP81] Victor Vasilievich Zyablov and Mark Semenovich Pinsker. List concatenated decoding. *Problemy Peredachi Informatsii*, 17(4):29–33, 1981.

Massachusetts Institute of Technology
*Email address*: fpernice@mit.edu

School of Computer Science, University of Washington
*Email address*: osprum@cs.washington.edu

Stanford
*Email address*: marykw@stanford.edu