Distance and Symmetry: Two Pillars of a Good Code

Oscar Sprumont

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:
Anup Rao, Chair
Paul Beame
Thomas Rothvoss

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Distance and Symmetry: Two Pillars of a Good Code

Oscar Sprumont

Chair of the Supervisory Committee:
Anup Rao
Computer Science & Engineering

This thesis explores the benefits of distance and symmetry in the design of error correcting codes. Our main measure of distance will be the *minimum distance* of a code, i.e., the smallest number of coordinates any two codewords may differ in. Our main criterion for symmetry will be *transitivity*, i.e., the requirement that any two coordinates be interchangeable. We will also consider generalizations of these two notions, for instance *double transitivity* (the requirement that any two pairs of coordinates be interchangeable) and *generalized distances* (the minimum number of nontrivial coordinates in any subcode of a certain size). We argue that codes with large distances and high symmetry present desirable properties for communication on noisy channels. Concretely, we prove the following results:

1. Any linear code that achieves list decoding capacity and has superconstant minimum distance also achieves capacity on the symmetric channel.

2. For any linear code $C \subseteq \{0,1\}^N$ with large enough generalized distances, the bit-decoding and block-decoding thresholds of C on the erasure channel are asymptotically equal.

3. Any transitive linear code $C \subseteq \mathbb{F}_q^N$ contains at most $q^{h_q(\alpha)\cdot \dim C}$ codewords of weight αN. This upper bound is tight, as evidenced by repetition codes.

4. For every doubly transitive code $C \subseteq \{0,1\}^N$, there is a range of noises $p \in [0, \frac{1}{2}]$ for which C achieves the information-theoretic optimal trade-off between rate and list decoding size.

5. The canonical example of linear codes with large distances and high symmetries is the family of Reed-Muller codes. We show that for appropriate choices of Reed-Muller codes C_1, C_2 and C_3, the tensor code $C_1 \otimes C_2 \otimes C_3$ achieves capacity on the symmetric channel with quasilinear decoding time.

ACKNOWLEDGMENTS

I thank Anup Rao for teaching me most of what I know about mathematics, coding theory, and theoretical computer science. Your ability to process complex proofs and concepts, boil them down to their essence, and communicate them in the clearest way possible is truly exceptional and has been incredibly valuable to my development as a researcher. I could not have asked for a better advisor.

I thank Emmanuel Abbé for hosting me at EPFL from September 2025 to March 2026. I learned a lot from my time with him and am very grateful for the opportunities he provided me. I thank Paul Beame and Thomas Rothvoss for their useful feedback on this thesis, and Anna Karlin for all of her help and support over the years. I also thank Alexander Barg, Paul Beame, and Henry Pfister for insightful career and publication advice.

I thank all my coauthors - Emmanuel Abbé, Francisco Pernice, Henry Pfister, Anup Rao, Colin Sandon, Mary Wootters, and Gilles Zémor - for sharing so many of their beautiful ideas with me. I also thank Siddharth Iyer and Michael Whitmeyer for always being willing to teach me new concepts or listen to me ramble on about math.

In addition to Michael and Siddharth, I have met many wonderful people in Seattle who I am lucky to call friends. I thank Emil Azadian, Mirte Boot, Beatriz Cuevas, and Elke Tukker for making my first year so enjoyable and immediately helping me feel at home. I thank Sujay Balebail, Nicole Baram, Romain Camilleri, Kesav Krishnan, and Maria Kuruvilla for being my Covid bubble and keeping me sane (and in fact, happy!) during somewhat stressful times. I am grateful to them - and to Flora Hollifield, Nathan Klein, David Martel, Pascal Sturmfels, and Emily Vo - for all the trips and dinner parties over the years. I cherish all the little moments we shared and hope there will be many more in the future.

I thank the UW CSE soccer community - Mars Gao, Tyler Han, Jason Hoffman, Sam Hurst, Chandler Peterson, Pascal Sturmfels, Gian Marco Visani, Lancelot Wathieu, Michael Whitmeyer, Chris Yin, and Zhihan Zhang - for their friendship and for the opportunity to release any of the stress that comes with caring about one's work. I also thank the board games and mafia communities for all the fun and entertainment they provided.

I thank my Montreal friends for being a constant source of joy and happiness. In particular, I am grateful to Peter John Daoud, Su Yu Ding, Allen Liu, and Yi Fei Liu for visiting me in Seattle, and to Alex Bergeron-Harris and Filip Kevin Phanalasy for hosting me whenever I would come back to Montreal.

I thank my parents, Véronique Le Gallo and Yves Sprumont, and my brother Adrien Sprumont for their love, warmth, and unconditional support. I do not know what I would do without you.

I thank Nicole Baram for her love and kindness, for always being ready to make fun of me or herself, and for cheering me up whenever needed. I am glad to be pack with you.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

The field of coding theory is centered around the following fundamental question: if two parties are communicating in an environment that partially corrupts their messages, how can they ensure that no valuable information is lost? As one might expect, the correct approach is to send messages containing enough redundant information that any corrupted parts can be recovered without much ambiguity.

Such sets of potential messages are called *error correcting codes*. The code you use most often in your day-to-day life is perhaps the English language. In English, errors are correctable because words with different meanings tend to sound very different. Thus when a toddler says "Vee doggo doggo hepy," you might be able to understand that they mean "The dog is happy," even though they did not pronounce a single word of that sentence correctly. Things would of course be much harder if the English language contained 200 two-syllable words sounding somewhere between "dog" and "doggo," or if "heppi" was a synonym of "asleep."

The exact same principle applies to *binary codes* $C \subseteq \{0,1\}^N$, the computer equivalent of human languages. If any two codewords $c, c' \in C$ are far away from each other - in the sense that c and c' differ in many coordinates - then a few errors here and there are very unlikely to cause any confusion. The exact definition of what constitutes a "confusion" will vary depending on our noise model and our decoding framework, but we will worry about that in due time (that is, Sections 2.2 and 2.3). One aspect that *will* remain fairly constant throughout our discussion however is that for a fixed length N, codes of larger sizes tend to be more vulnerable to corruptions than codes of smaller sizes. Intuitively, this is simply because every codeword you add to your set C is one more string the receiver needs to be

able to distinguish from the codeword you sent them.

On the other hand, for obvious reasons - computers, just like humans, do not enjoy listening to a 10 minute explanation that should have taken 30 seconds - we want to minimize the amount of redundant information we transmit. The main challenge is then to design codes that both

(a) introduce as little redundancy as possible; and

(b) tolerate as many errors as possible.

This thesis will be concerned with trying to identify global and structural properties of a code $C \subseteq \{0,1\}^N$ that ensure certain guarantees regarding points (a) and (b) above. In particular, we will rely heavily on the concepts of *distance* and *symmetry*. We loosely use the term *distance* to refer to the minimum number of coordinates we need in order to differentiate the sent message $c \in C$ from any subset of codewords $S \subseteq C$ of a certain size; we loosely use the term *symmetry* to refer to the requirement that every coordinate - or sometimes, group of coordinates - be in some sense "equivalent."[1] Our goal will be to show that distance and symmetry can be leveraged to reliably communicate over noisy channels.

1.1 Overview of Main Contributions

A q-ary code of length N is simply a subset $C \subseteq \mathbb{F}_q^N$, where $\mathbb{F}_q$ is the finite field with q elements. As mentioned in the previous section, our goal is to design codes with low redundancy and high tolerance to noise. The redundancy of a code $C \subseteq \mathbb{F}_q^N$ is quantified by its *rate*, which we define as

$$R := \frac{\log_q |C|}{N} \in [0,1].$$

Codes of larger rates have lower redundancy, as they contain more messages (and thus convey more information) than codes of lower rates. The tolerance to noise is dependent on

[1] See Section 2.7 for a more formal definition of the notions of distance and symmetry.

the communication channel used by the two parties. We will consider the following channels, each parametrized by some constant $p \in [0, 1]$.

(i) In the adversarial model, the adversary is allowed to corrupt up to pN coordinates. They get to choose which entries to corrupt and which elements of $\mathbb{F}_q$ to replace them with.

(ii) In the stochastic model, each coordinate is corrupted independently at random with probability p. The two forms of corruptions we will consider are random erasures (each corrupted entry is replaced by an erasure symbol) and random errors (each corrupted entry c_i is replaced by a uniformly random element in $\mathbb{F}_q \setminus c_i$).

In the stochastic model, the optimal tradeoff between redundancy and noise tolerance is known (see Section 2.2). The largest rate at which one can recover from p-noisy erasures is

$$R = 1 - p,$$

whereas for p-noisy errors it is

$$R = 1 - h_q(p), \tag{1.1}$$

where h_q denotes the q-ary entropy function

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

Any sequence of codes that achieves the optimal tradeoff between rate and noise tolerance for a specific channel is said to *achieve capacity* on that channel.

In the adversarial channel, the optimal tradeoff between redundancy and noise tolerance is still unknown[2]. However, in the list decoding framework - where instead of outputting one codeword, the receiver only needs to output a list of L codewords with the guarantee that

[2]It is unknown for constant values of q (although there are known lower and upper bounds, see Section 2.2.2). For q tending to infinity, we know that the optimal tradeoff is $R = 1 - p$.

the sent message is in that list - the optimal tradeoff is well understood. As long as the list size L is superconstant, there are codes of rate

$$R = 1 - h_q(p) \tag{1.2}$$

that can list-decode pN adversarial errors.

We will be interested in leveraging the structural properties of certain linear codes to obtain good decoding performance on noisy channels, in both the unique decoding and the list decoding frameworks. The first property we will rely on is the concept of *distance*. We define the minimum distance of a code C to be the minimum Hamming distance[3] between any two codewords $c, c' \in C$. We define the r^{th} generalized distance of a linear code C to be the minimum support size[4] of any r-dimensional subcode of C.

The second property we will be interested in is *symmetry*, and in particular *transitivity*. We say that a code $C \subseteq \mathbb{F}_q^N$ is transitive if for any two indices $i, j \in [N]$, there exists a permutation $\pi : [N] \to [N]$ such that $\pi(i) = j$ and such that rearranging the coordinates of any codeword $c \in C$ according to π gives you another codeword. We say that a code $C \subseteq \mathbb{F}_q^N$ is *doubly transitive* if for any pairs of indices $(i, k), (j, l) \in [N]$ with $i \neq k$ and $j \neq l$, there exists a permutation $\pi : [N] \to [N]$ such that $\pi(i) = j, \pi(k) = l$, and rearranging the coordinates of any codeword $c \in C$ according to π gives you another codeword. This dissertation makes the following contributions:

Leveraging Distance

- The reader may have noticed that the capacity of the symmetric channel under unique decoding is exactly the same as the capacity of the adversarial channel under list decoding (see (1.1) and (1.2)). This is not a coincidence. We show that any linear code $C \subseteq \mathbb{F}_q^N$ that achieves list decoding capacity and has superconstant minimum

[3]The Hamming distance between two points $x, y \in \mathbb{F}_q^N$ is the number of coordinates $i \in [N]$ where $x_i \neq y_i$.

[4]The support of a set $S \subseteq \mathbb{F}_q^N$ is the set of coordinates $i \in [N]$ such that there exists some $c \in S$ with $c_i \neq 0$.

distance also achieves capacity on the symmetric channel [58]. As a consequence of this result, we obtain new capacity-achieving codes for the symmetric channel, notably multiplicity codes and algebraic geometry codes.

- Fix any linear code $C \subseteq \mathbb{F}_2^N$ and define $p_{\text{bit}} \in [0,1]$ to be the smallest erasure probability at which we can still reliably decode any one coordinate of our sent codeword $c \in C$. Define $p_{\text{full}} \in [0,1]$ to be the smallest erasure probability at which we can reliably decode the entire sent codeword c. We show that as long as C has large enough generalized distances, p_{bit} and p_{full} are asymptotically equal [59]. This gives a new proof of the celebrated result [49, 50] that Reed-Muller codes achieve capacity on the erasure channel.

Leveraging Symmetry

- We prove upper bounds on the weight distribution of transitive and doubly transitive linear codes.

  1. We show that for any $\alpha \in (0,1)$, any transitive linear code $C \subseteq \mathbb{F}_q^N$ has at most $q^{h_q(\alpha) \cdot \dim C}$ codewords $c \in C$ with αN nonzero entries [61]. This upper bound is tight, since it is attained by repetition codes.
  2. We give a simpler proof of the fact that for every doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of large enough rate, there exists an interval $I \subseteq [0, N]$ of linear size centered around $\frac{N}{2}$ such that the weight distribution of C on I is the same as the binomial distribution. This fact was originally proven in [64].

- We give list decoding bounds for decoding transitive and doubly transitive linear codes from random errors [61]. In some regimes, our bounds for doubly transitive codes achieve the information-theoretic optimal trade-off between the rate of the code and the size of the list needed for decoding.

- For any rate $R \in (0, 1)$, we show that one can combine three Reed-Muller codes (formally, by taking their tensor product) to obtain a code C of rate R that is decodable to capacity in quasilinear time [2].

Chapter 2

NOTATION, CONVENTIONS AND NECESSARY BACKGROUND

We denote the set of all non-negative integers by

$$\mathbb{N} := \{0, 1, 2, 3, \dots\}.$$

For any positive integer m, we define the set $[m] := \{1, 2, \dots, m\}$. For any $m, d \in \mathbb{N}$, we also use the notation

$$[m \pm d] := \Big\{\lceil m - d \rceil, \lceil m - d \rceil + 1, \dots, \lfloor m + d \rfloor\Big\}.$$

For any non-negative integers $d < m$, we define the binomial coefficient

$$\binom{m}{d} := \frac{m!}{d!(m-d)!},$$

where $m! = \prod_{i=1}^{m} i$ denotes the factorial function. For convenience, we will use the shorthand notation

$$\binom{m}{\leq d} := \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{d}$$

for any integer $d < m$ and

$$\binom{m}{S} := \sum_{s \in S} \binom{m}{s}$$

for any subset $S \subseteq \{0, 1, \dots, m\}$. We denote by $\mathbb{F}_q$ the finite field with q elements. For any $x \in \mathbb{F}_q^m$, we write

$$\mathrm{wt}(x) := \Big|\big\{j \in [m] : x_j \neq 0\big\}\Big|$$

to denote the Hamming weight of x. For any $x \in \mathbb{F}_q^N$, any $i \in [N]$ and any $a \in \mathbb{F}_q$, we define $x^{i \mapsto a} \in \mathbb{F}_q^N$ to be the vector with coordinates

$$[x^{i \mapsto a}]_j = \begin{cases} a & \text{if } j = i, \\ x_j & \text{otherwise.} \end{cases}$$

We say that a function $f : \mathbb{F}_q^n \to \{0, 1\}$ is *monotone increasing* if for any index $i \in [n]$, any point $x \in \mathbb{F}_q^n$, and any $a \in \{1, 2, \ldots, q-1\}$, we have

$$f(x^{i \mapsto 0}) \leq f(x^{i \mapsto a}). \tag{2.1}$$

We say that $f : \mathbb{F}_q^n \to \{0, 1\}$ is *monotone decreasing* if $1 - f$ is monotone increasing.

2.1 Linear Codes

An N-bit q-ary code is a subset $C \subseteq \mathbb{F}_q^N$. We call each element $c \in C$ a *codeword* of C. Whenever C is a subspace of $\mathbb{F}_q^N$, we say that C is a *linear* code. Any linear code $C \subseteq \mathbb{F}_q^N$ can be represented by its generator matrix, which is a $N \times \dim\ \mathrm{C}$ matrix G whose columns form a basis of C. The matrix G generates all codewords of C in the sense that

$$C = \{Gv : v \in \mathbb{F}_q^{\dim\ C}\}.$$

Another useful way to describe a linear code $C \subseteq \mathbb{F}_q^N$ is via its parity-check matrix, which is an $N \times (N - \dim\ \mathrm{C})$ matrix H whose columns span the orthogonal complement $C^\perp$ of C[1]. The linear code C can then be expressed as

$$C = \{c \in \mathbb{F}_q^N : c^\intercal H = \vec{0}\}.$$

The *tensor product* $C = C_1 \otimes C_2$ of two codes $C_1 \in \mathbb{F}_q^{N_1}$ and $C_2 \in \mathbb{F}_q^{N_2}$ is a code of length $N = N_1 \cdot N_2$ and dimension $\dim C_1 \cdot \dim C_2$. Its generator matrix is the tensor product of the generator matrices of C_1 and C_2. Equivalently, C is the set of all matrices $A \in \mathbb{F}_q^{N_2 \times N_1}$ where each row of A is a codeword of C_1 and each column of A is a codeword of C_2.

[1] We call $C^\perp$ the *dual code* of C.

As mentioned in the introduction, we think of a code $C \subseteq \mathbb{F}_q^N$ as the set of all possible messages that one party may want to transmit to another. The sender chooses some codeword $c \in C$ and transmits it over a communication channel[2]. The receiver then sees a corrupted version of this message and must recover the original codeword correctly. We typically want to pick a subset C that is as large as possible, since a code C of small size would be "wasting" a large number of coordinates. On the other hand, we cannot afford to choose a code C of too large a size, since every codeword we add to C is one more string the original message could be confused with. For instance, if the two parties were to choose the code $C = \mathbb{F}_q^N$, then any error introduced by the channel would make it impossible for the receiver to recover the original codeword. The main challenge is then to design codes that convey as much information as possible while still tolerating as many errors as possible.

Definition 1. *The rate R of a code $C \subseteq \mathbb{F}_q^N$ is defined to be*

$$R := \log_q \frac{|C|}{q^N}.$$

The rate of a code is a measure of the efficiency with which the code conveys information. For $C \subseteq \mathbb{F}_2^N$, if a sender transmits a uniformly random codeword $c \in C$, then on average each bit of c conveys R bits of information to the receiver.

2.2 Coding Channels and Noise Models

When the sender transmits a codeword $c \in C$ through a *communication channel*, the sent codeword gets corrupted. We define the notion of communication channel below.

Definition 2. *A communication channel is a function*

$$f : 2^{\mathbb{F}_q^N} \times \mathbb{F}_q^N \to \mathcal{P}(\Sigma^N),$$

where Σ is any finite alphabet and $\mathcal{P}(\Sigma^N)$ is any probability distribution over Σ^N.

---

[2]See Definition 2.

We call any element of $\mathbb{F}_q$ or Σ a *symbol*. We think of the first input to f as the code $C \subseteq \mathbb{F}_q^N$ that the two parties will use. We think of the second input to f as the codeword $c \in C$ that the sender wants to communicate to the receiver. The output $f(C, c)$ is then the probability distribution over all corrupted messages that the receiver might see. We note that the probability distribution $f(C, c)$ can sometimes be independent of C (see e.g. the two stochastic channels described in Section 2.2.1) or concentrated on a single point (see e.g. the adversarial channel described in Section 2.2.2). One of the most important properties of any communication channel is its capacity.

Definition 3. *The* capacity *of a communication channel* $f : 2^{\mathbb{F}_q^N} \times \mathbb{F}_q^N \to \mathcal{P}(\Sigma^N)$ *is the smallest rate of any code* $C \subseteq \mathbb{F}_q^N$ *that admits reliable communication over* f.

By "admits reliable communication over f," we mean that there exists a decoder $D : \Sigma^N \to C$ such that for any codeword $c \in C$, we have $\Pr\left[D\big(f(C, c)\big) = c\right] \geq 1 - o(1)$. We define the capacity of a *sequence* of channels $\{f_N\}$ to be the infimum limit of the capacity of f_N as N tends to infinity. Similarly, we define the rate of a sequence of codes $\{C_N\}$ to be the infimum limit of the rate of C_N as N tends to infinity.

2.2.1 Stochastic Channels

In every stochastic channel that we will discuss in this thesis, the coordinates are corrupted independently at random with some probability p. We will consider the following two different forms of noise.

(i) The q-ary erasure channel (qEC_p): every corrupted coordinate is replaced by an erasure symbol.

(ii) The q-ary symmetric channel (qSC_p): every corrupted coordinate c_i is replaced by a uniformly random element in $\mathbb{F}_q \setminus c_i$.

The Symmetric Channel

When a sender sends some codeword $c \in C$ through the channel qSC_p, a p-noisy error string $z \in \mathbb{F}_q^N$ (which we will call the *error pattern*) is sampled and the receiver receives the vector $y \in \mathbb{F}_q^N$ with coordinates

$$y_i = c_i + z_i.$$

The receiver's goal is then to recover the sent codeword c. The capacity of the symmetric channel qSC_p is known to be $1 - h_q(p)$.

Theorem 1 ([69])**.** *The capacity of the q-ary symmetric channel qSC_p is $1 - h_q(p)$.*

Proof. We will first show that any code $C \subseteq \mathbb{F}_q^N$ of rate $1 - h_q(p) + \Omega(1)$ is not resilient to p-noisy errors. For this, we note that on the qSC_p, with high probability the received message will be a distance $pN \pm o(N)$ away from the sent codeword c. In order for the decoder to succeed with probability greater than $\frac{3}{4}$, there must then be at least one weight $p' = p \pm o(1)$ for which the majority of strings $x \in \mathbb{F}_q^N$ with $d(c, x) = p'N$ get decoded to c. By the volume bound (Lemma 17), this means that for each codeword $c \in C$, there must be at least $q^{h_q(p)N - o(N)}$ distinct strings $x \in \mathbb{F}_q^N$ that get decoded to c. Since there are only q^N strings in $\mathbb{F}_q^N$, we must then have $|C| \cdot q^{h_q(p)N - o(N)} \leq q^N$, and thus $|C| \leq q^{(1-h_q(p))N + o(N)}$.

For the second part of the proof, we will show that a uniformly random codeword c of a uniformly random code $C \subseteq \mathbb{F}_q^N$ of size $|C| = q^{(1-h_q(p))N - N^{9/10}}$ is resilient to p-noisy errors with high probability. This would prove our claim, since we can then take C' to be the code C minus all the codewords of C that are not resilient to p-noisy errors. Thus take a uniformly random codeword c of our uniformly random code C, and apply p-noisy errors to c. The received message x is a uniformly random point in $\mathbb{F}_q^N$, and by the Chernoff bound we have $d(c, x) \leq pN + N^{3/4}$ with high probability. Now x leads to a decoding error if and only if there exists a different codeword $c' \in C$ such that $d(x, c') < d(x, c)$. But by the volume bound (Lemma 17) and the subadditivity of entropy (Lemma 16) there are only $q^{h_q(p + N^{-1/4})N} \leq q^{h_q(p)N + N^{4/5}}$ such potential strings c'. By the union bound, the probability that any such c' is in our code C is then bounded by $|C| \cdot \frac{q^{h_q(p)N + N^{4/5}}}{q^N} \leq o(1)$. □

The Erasure Channel

When a sender sends some codeword $c \in C$ through the erasure channel qEC_p, a p-noisy string $z \in \{0,1\}^N$ (which we will call the *erasure pattern*) is sampled and the receiver receives a vector $y \in \{0,1,\ldots,q-1,*\}^N$ with coordinates

$$y_i = \begin{cases} c_i & \text{if } z_i = 0, \\ * & \text{otherwise.} \end{cases}$$

A notion that will be important in our analysis is the notion of *covered codeword*. For any erasure pattern $z \in \{0,1\}^N$, we say that a codeword $c \in C$ is *covered* by z if $z_i = 1$ for all coordinates $i \in [N]$ where $c_i \neq 0$. We denote this by

$$z \succ c.$$

We note that if the erasure pattern z covers some codeword $c \neq \vec{0}$, then it is impossible for the receiver to distinguish between a sent message $c' \in C$ and the codeword $c + c'$. We denote the subcode of C covered by an erasure pattern z by

$$S_C(z) := \Big\{ c \in C : z \succ c \Big\}. \tag{2.2}$$

The capacity of the erasure channel qEC_p is known to be $1-p$.

Theorem 2. *The capacity of the q-ary erasure channel qEC_p is $1-p$.*

Since the proof is almost identical to the proof of Theorem 1, we defer it to Section A.1. Theorems 1 and 2 motivate the following capacity definitions.

Definition 4. *A sequence of codes $\{C_N \subseteq \mathbb{F}_q^N\}$ of rate $1-p$* achieves capacity over the erasure channel *if with high probability, it can recover from random errors of rate $p - o(1)$.*

Definition 5. *A sequence of codes $\{C_N \subseteq \mathbb{F}_q^N\}$ of rate $1 - h_q(p)$* achieves capacity over the symmetric channel *if with high probability, it can recover from random errors of rate $p - o(1)$.*

Several families of codes are now known to achieve capacity over both the erasure and error channels. In his seminal paper [69], Shannon established that uniformly random codes achieve capacity. The first explicit families of codes to provably achieve capacity were only obtained decades later: Forney introduced concatenated codes in [21], Gallagher designed LDPC codes in [22] (see also [54, 51]), and Berrou constructed turbo codes in [12]. More recently, Arikan showed that polar codes - which have both a deterministic construction and efficient encoding and decoding algorithms - also achieve capacity [7]. This gave renewed attention to the closely related Reed-Muller codes, which were shown to achieve capacity over both the symmetric and erasure channels in [4, 49, 62, 1].

Binary Memoryless Symmetric Channels

Binary Memoryless Symmetric (BMS) channels are a generalization of the binary Erasure Channel and the binary Symmetric Channel. Denoting by $x \in \{0,1\}^N$ the input to the channel and by $y \in \Sigma^N$ the output to the channel (where Σ must be a subset of $\mathbb{R}$), a BMS channel is any communication channel with the following properties:

- The random variable y_i only depends on x_i.
- The probability that $y_i \in [a,b]$ conditioned on $x_i = 1$ is equal to the probability that $y_i \in [-b,-a]$ conditioned on $x_i = 0$.

The binary symmetric channel and the binary erasure channel are both BMS channels. Another common BMS channel is the binary additive white Gaussian noise channel, where independent Gaussian noise is added to each coordinate.

2.2.2 Adversarial Channels

In an adversarial channel with parameter $p \in [0, \frac{1}{2}]$, the adversary is allowed to corrupt up to pN coordinates. The adversary gets to choose which coordinates to corrupt, and which elements of $\mathbb{F}_q$ to replace them with.

Claim 3. *A code $C \subseteq \mathbb{F}_q^N$ is resilient to pN adversarial errors if and only if its minimum distance*[3] *satisfies $d_{\min}(C) > 2pN$.*

Proof. Suppose two codewords $c, c' \in C$ satisfy $d(c, c') \leq 2pN$. Then there exists a string $x \in \mathbb{F}_q^N$ such that $d(x, c) \leq pN$ and $d(x, c') \leq pN$. If the adversary corrupts both c and c' to x, the receiver cannot distinguish c from c'. □

Note 4. *The reader may be curious as to why we did not also consider adversarial erasures as a standalone channel. The reason is that a code $C \subseteq \mathbb{F}_q^N$ is resilient to adversarial erasures of rate p if and only if its minimum distance satisfies $d_{\min}(C) > pN$. By Claim 3, this means that the adversarial erasure channel is simply a reparametrization of the adversarial error channel.*

Somewhat surprisingly, the exact capacity of the adversarial channel is still unknown. There are however some well-known bounds. We start with two upper bounds.

Lemma 5 (The Hamming bound [37])**.** *Any code $C \subseteq \mathbb{F}_q^N$ that can decode pN adversarial errors must have rate*

$$R \leq 1 - h_q(p) + o(1).$$

Proof. We have already seen in Theorem 1 that there exists some $\delta = o(1)$ such that any code of rate $R \geq 1 - h_q(p) + \delta$ cannot reliably decode p-noisy errors. □

Lemma 6 (The Singleton bound [70])**.** *Any code $C \subseteq \mathbb{F}_q^N$ that can decode pN adversarial errors must have rate*

$$R \leq 1 - 2p + o(1).$$

Proof. By Claim 3, it will be sufficient to prove that

$$RN + d_{\min}(C) \leq N + 1. \tag{2.3}$$

[3] see Definition 9

But there are q^{RN} codewords in C, whereas there are only q^{RN-1} possible evaluations on the first $RN-1$ coordinates. Thus there must exist two codewords $c, c' \in C$ that have the exact same evaluation on the first $RN-1$ coordinates. These two codewords are a distance $\leq N - RN + 1$ away from each other, which immediately gives us equation (2.3).

□

We note that for $q = 2$, the Hamming bound is strictly stronger than the Singleton bound. For $q > 2$, the Hamming bound is stronger for smaller values of p but weaker for larger values of p. We also state a well-known lower bound on the rate of the largest code that can decode pn adversarial errors.

Lemma 7 (The Gilbert-Varshamov bound [25, 78])**.** *There exists a code $C \subseteq \mathbb{F}_q^N$ that can decode pN adversarial errors and has rate*

$$R \geq 1 - h_q(2p).$$

Proof. We greedily construct a code C with the desired properties. Initialize our set of "available points" $S \subseteq \mathbb{F}_q^N$ to be $S = \mathbb{F}_q^N$. Pick any point $c \in S$ and add it to C. Remove from S all the points $x \in \mathbb{F}_q^N$ satisfying $d(x, c) \leq 2pN$. Then pick any other point c' in our updated set S and add it to the code C. Remove from S all points y satisfying $d(y, c') \leq 2pN$. Keep arbitrarily picking points in S and removing the ball of radius $2pN$ around them, until the set S is empty. By construction, the minimum distance of the code C you obtain is at most $2pN$. Thus by Lemma 3, C can decode pN adversarial errors. On the other hand, by Lemma 17, every time a point $c \in \mathbb{F}_q^N$ gets added to C, at most $q^{h_q(2p)N}$ points get removed from S. Thus by the time S is empty, C will contain at least $q^{(1-h_q(2p))N}$ codewords. □

Note that $\lim_{q\to\infty} h_q(2p) = 2p$ (see Lemma 18). Thus by Theorems 6 and 7, in the setting where the alphabet size q is allowed to grow to infinity with the length N of the code, the capacity of the symmetric channel under adversarial errors is $R = 1 - 2p$; meanwhile by Theorems 1 and 18, the capacity of the symmetric channel under random errors in this

setting is $R = 1 - p$. So over large alphabets, one can reliably decode twice as many random errors as adversarial ones.

2.3 List Decoding

As we have seen in Sections 2.2.1 and 2.2.2, the best possible trade-off between the code rate R and the noise parameter p is worse in the adversarial case than it is in the stochastic case. We will now consider a notion that bridges these two worlds: list decoding. Introduced by Elias [20] and Wozencraft [81], list decoding can be thought of as a relaxation of what we consider a successful 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. A code that can always list decode from pN adversarial errors with list size L is called (p, L)-list decodable.

Definition 6. *A code $C \subseteq \mathbb{F}_q^N$ is* (p, L)-list-decodable *if for all $x \in \mathbb{F}_q^N$,*

$$\left|\{c \in C \,:\, d(x, c) \leq pN\}\right| \leq L.$$

The best possible trade-off between R, p and L in the list-decoding setting is well understood.

Theorem 8 ([20])**.** *Consider any $q \geq 2$, any $0 \leq p \leq 1 - \frac{1}{q}$ and any $\varepsilon > 0$ small enough. Then we have:*

1. *For every $R \leq 1 - h_q(p) - \varepsilon$, there exists a sequence of $\left(p, \frac{1}{\varepsilon}\right)$-list decodable codes $\{C_N\}$ of rate R.*

2. *For every $R \geq 1 - h_q(p) + \varepsilon$, there exists no (p, L)-list decodable sequence of codes $\{C_N\}$ of rate R with $L \leq q^{o(\varepsilon N)}$.*

Proof. We first prove the first statement. Consider q^{RN} independent uniformly random points $c^1, \ldots, c^{q^{RN}} \in \mathbb{F}_q^N$ and let $C = \{c^1, \ldots, c^{q^{RN}}\}$. Note that with high probability, C will

have size $|C| \geq q^{RN-o(N)}$. Now fix any $x \in \mathbb{F}_q^N$. By the volume bound (Lemma 17), for any choice of $\frac{1}{\varepsilon}$ codewords $c^{i_1}, c^{i_2}, \ldots, c^{i_{1/\varepsilon}}$, the probability that all these codewords lie in the ball of radius pN around x is bounded by

$$\Pr\left[c^{i_1}, c^{i_2}, \ldots, c^{i_{1/\varepsilon}} \in B(x, pN)\right] \leq q^{-\frac{1}{\varepsilon}(1-h_q(p))N}.$$

By union bound, the probability that there exists some point $x \in \mathbb{F}_q^N$ for which $\left|C \cap B(x, pN)\right| \geq \frac{1}{\varepsilon}$ is then bounded by

$$\begin{aligned}
\Pr\left[\text{C is not } (p, 1/\varepsilon) \text{ - list decodable}\right] &\leq q^N \binom{q^{RN}}{1/\varepsilon} q^{-\frac{1}{\varepsilon}(1-h_q(p))N} \\
&< q^{N+\frac{RN}{\varepsilon}} \cdot q^{-\frac{N}{\varepsilon}+\frac{h_q(p)N}{\varepsilon}} \\
&\leq q^{N+\frac{N-Nh_q(p)-\varepsilon N}{\varepsilon}} \cdot q^{-\frac{N}{\varepsilon}+\frac{h_q(p)N}{\varepsilon}} \\
&= 1.
\end{aligned}$$

By the probabilistic method, there must thus exist a code $C \subseteq \mathbb{F}_q^N$ of rate $R - o(1)$ that is $(p, \frac{1}{\varepsilon})$-list decodable. We now prove the second statement. Fix any code $C \subseteq \mathbb{F}_q^N$ of rate $R - o(1)$. Then by the volume bound (Lemma 17), for every $c \in C$ we have that for a uniformly random point $x \in \mathbb{F}_q^N$,

$$\Pr_{x \in \mathbb{F}_q^N}\left[c \in B(x, pn)\right] \geq q^{-(1-h_q(p))N-o(N)}.$$

Thus for a uniformly random point $x \in \mathbb{F}_q^N$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \in \mathbb{F}_q^N}\left[\left|C \cap B(x, pN)\right|\right] &\geq q^{RN-o(N)} \cdot q^{-(1-h_q(p))N-o(N)} \\
&\geq q^{\varepsilon N-o(N)}.
\end{aligned}$$

By the probabilistic method, there must then exist a point $x \in \mathbb{F}_q^N$ such that $\left|B(x, pN) \cap C\right| \geq q^{\varepsilon N-o(N)}$. □

Theorem 8 motivates the following definition of list decoding capacity.

Definition 7. *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* $p(N) = o(1)$ *such that each* C_N *is* $(p - p(N), L(N))$*-list decodable.*

We note that by our proof of Theorem 8, uniformly random codes achieve list decoding capacity with $p(N) = O(\frac{1}{L(N)})$. But in general, we allow for worse dependence of p on L: we only require that p go to 0 as N goes to infinity.

Many families of codes are now known to achieve list decoding capacity. 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 [84, 29, 80, 53, 30].

The first explicit constructions were obtained by Guruswami and Rudra, who showed in [31] that folded Reed-Solomon codes achieve list decoding capacity (see also the follow-up work of [47, 75, 72, 18]). Guruswami, Kopparty and Wang then extended these results to other families of codes, including multiplicity codes [34, 46].

All of these families have growing alphabet sizes. Turning to constructions of capacity-achieving list-decodable codes with constant alphabet sizes, we have Gallager's ensemble of LDPC codes [57], Guruswami and Rudra's concatenated codes [32], and constructions based on algebraic geometry (AG) codes, for example [35, 38, 27, 36, 16].

All of these codes use an alphabet of size larger than 2. One of the most important open problems in coding theory is to find an explicit family of *binary* codes $C \subseteq \mathbb{F}_2^N$ that achieves list decoding capacity.

2.4 Decoding Algorithms

In order for a code to be useful in practice, one should be able to decode it efficiently. An important subfield of coding theory is thus the design of decoders $D : \mathbb{F}_2^N \to C$ with efficient algorithms. The following standard lemma states that for any linear code C with parity check matrix H, if one can recover the string z from its image Hz, then one can also recover

any codeword $c \in C$ that is corrupted by the error pattern z. This will simplify the analysis in many of our calculations.

Lemma 9. *Let H be an $N \times (N - \dim C)$ parity-check matrix of a linear code $C \subseteq \mathbb{F}_q^N$ and let $p \in (0, \frac{1}{2})$, $k \in \mathbb{N}$ be arbitrary. Then for every decoder $D' : \mathbb{F}_q^{N-\dim C} \to \left(\mathbb{F}_q^N\right)^{\times k}$, there exists a decoder*

$$D : \mathbb{F}_q^N \to C^{\times k}$$

such that for every $c \in C$, we have

$$z \in D'(Hz) \implies c \in D(c + z).$$

We note that the lemma above applies to both unique decoding ($k = 1$) and list decoding ($k > 1$).

Proof. Consider the following decoder D: on any input $y \in \mathbb{F}_q^N$, compute $y' = Hy$ and return the set $y - D'(y')$. We claim that whenever z, D' satisfy

$$z \in D'(Hz), \tag{2.4}$$

we also have $c \in D(c+z)$. To see this, consider any codeword $c \in C$. Since H is a parity-check matrix of C, we have

$$\begin{aligned} H(c + z) &= Hc + Hz \\ &= Hz. \end{aligned}$$

Thus on input $y = c + z$, by definition, D will output the set $c + z - D'(Hz)$. By our assumption (2.4), we then have $c \in D(c + z)$.

□

Perhaps the most basic decoder is the *maximum-likelihood* decoder, which can be used on any code.

Definition 8. *For any code $C \subseteq \mathbb{F}_q^N$, the maximum-likelihood decoder $D_{\mathrm{ML}} : \mathbb{F}_q^N \to C$ does the following: given input $x \in \mathbb{F}_q^N$, compute $d(x, c)$ for every $c \in C$. Return a codeword c that is closest to x.*

Note that the maximum-likelihood decoder has runtime polynomial in $|C|$, which renders it impractical for most applications. One major goal of coding theory is to obtain practical (i.e., polynomal-time in N) decoding algorithms for codes that are known to behave well under maximum-likelihood decoding. For instance, [66] showed that there exists a polynomial-time decoder D for the binary Reed-Muller code $\mathsf{RM}(m - 2t, m)$ with the following property: if the maximum-likelihood decoder for $\mathsf{RM}(m - t, m)$ correctly decodes an *erasure* pattern $z \in \{0, 1\}^N$, then D correctly decodes the corresponding *error* pattern z. See Section 2.8 for a description of Reed-Muller codes.

Theorem 10 ([66])**.** *Fix any integers n and $t \leq n$. There exists an $O\left(2^n \cdot \mathrm{poly}(\binom{n}{\leq t})\right)$ - time decoder D for the code $\mathsf{RM}(n, n - 2t)$ with the following property. For every $z \in \{0, 1\}^{2^n}$, if*

$$\left\{x \in \mathsf{RM}(n, n - t) : z_i \geq x_i \text{ for all } i\right\} = \{0\},$$

then every $c \in \mathsf{RM}(n, n - 2t)$ satisfies

$$D(c + z) = c.$$

One tool that will come in handy when analyzing the runtime of certain algorithms is the Master theorem.

Lemma 11 (The Master theorem)**.** *Suppose $T(n)$ denotes the running time of an algorithm on an input of size n, and suppose $T(n)$ can be expressed recursively as*

$$T(n) \leq aT\left(\frac{n}{a}\right) + O(n)$$

for some constant $a > 0$. Then if $T(1) = O(1)$, we have $T(n) \leq O(n \log n)$.

2.5 Probability Theory

A real-valued random variable is a function $X : A \to \mathbb{R}$, where A is a probability space. We will make use of three very standard results in probability theory (see e.g. [14]): Markov's inequality, Chernoff's bound, and Hoeffding's inequality.

Lemma 12 (Markov's inequality)**.** *Let X be a non-negative random variable. Then for any $a > 0$, we have*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Lemma 13 (Chernoff bound)**.** *Let $X_1, X_2, \ldots, X_m$ be i.i.d. random variables taking values in $\{0, 1\}$ and define $p := \mathbb{E}[X_1]$. Then for any $\varepsilon \in (0, 1)$, we have*

$$\Pr\left[\left|\sum_{i=1}^{m} X_i - pm\right| > \varepsilon \cdot pm\right] \leq 2e^{-\frac{\varepsilon^2 pm}{3}}.$$

Furthermore, for any $\delta \in [0, 1-p]$ we have

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} X_i \geq p + \delta\right] \leq 2^{-D(p+\delta\|p)m},$$

where $D(x \parallel y) := x \log \frac{x}{y} + (1-x) \log\left(\frac{1-x}{1-y}\right)$.

Lemma 14 (Hoeffding's inequality)**.** *Let $X_1, X_2, \ldots, X_m$ be independent random variables taking values in $[0, 1]$. Then for any $t > 0$, we have*

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

and

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

For any probability distribution $\mathcal{D}$ over a set X, we will use the notation $z \sim \mathcal{D}$ to denote a random element $z \in X$ sampled according to $\mathcal{D}$. The main probability distribution that

we will consider is the *p-noisy* distribution over $\mathbb{F}_q^N$, and for this specific distribution we use the following shorthand notation. For $p \in [0,1]$, we write

$$z \sim p$$

to denote a p-noisy string $z \in \mathbb{F}_q^N$, meaning that for each $i \in [N]$, the i^{th} entry is independently

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

Given any subset of points $S \subseteq \mathbb{F}_q^N$, we will abuse notation and write

$$\Pr_{x \in S}[\cdot]$$

to identify the uniform distribution over the set S.

2.6 Entropy and Volume Bounds

In this section, we will go over one of the most useful tools we have for analyzing the behavior of codes: entropy. Consider any probability space A with probability distribution p. For a random variable $X : A \to B$, we abuse notation and write $p(b) := \sum_{a:X(a)=b} p(a)$ for every $b \in B$. The entropy of X is then the quantity

$$H(X) := \sum_{b \in B} p(b) \log \frac{1}{p(b)}.$$

For any random variable $Y : A \to C$, the conditional entropy of X on Y is the quantity

$$H(X|Y) := \sum_{\substack{b \in B \\ c \in C}} p(b,c) \log \frac{1}{p(b|c)}.$$

The following theorem states a few well-known properties of entropy.

Theorem 15. *For any random variables $X : A \to B$ and $Y : A \to C$, we have:*

- *Extremal values:* $0 \leq H(X) \leq \log |B|$.
- *Subbaditivity:* $H(X, Y) \leq H(X) + H(Y)$.
- *Conditioning doesn't increase entropy:* $H(X|Y) \leq H(X)$.

One random variable that will be of great importance to us is the identity random variable $X : \mathbb{F}_q \to \mathbb{F}_q$ (i.e. the map $X(a) = a$) with probability distribution

$$p(a) = \begin{cases} 1 - \varepsilon & \text{if } a = 0, \\ \frac{\varepsilon}{q-1} & \text{otherwise.} \end{cases}$$

This random variable corresponds to the errors introduced by the communication channel qSC_p. The entropy of this random variable will be of such use to us that we attribute it its own name: multiplying by a constant factor that simply keeps the function in $[0, 1]$, we define the q-ary entropy function $h_q : [0, 1] \to [0, 1]$ to be

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

For the case $q = 2$, we will often omit the subscript and simply write $h(x) := h_2(x)$. One useful property of the entropy function is that it is subadditive.

Lemma 16. *For any positive integer q, any $x \in [0, 1]$ and any $y \in [0, 1 - x]$, we have*

$$h_q(x + y) \leq h_q(x) + h_q(y).$$

This can be seen as a consequence of the second bullet point in Theorem 15. Alternatively, it can be proven by noting that the q-ary entropy function is concave and that any concave, positive function is subadditive (see e.g. [26], page 83, statement 103).

One major use of the q-ary entropy function is that it gives good bounds on the number of points in the ball of radius r around any point $x \in \mathbb{F}_q^N$. Define

$$B(x, r) := \{y \in \mathbb{F}_q^N : d(x, y) \leq r\}$$

and

$$S(x,r) := \{y \in \mathbb{F}_q^N : d(x,y) = r\}.$$

Then the quantity $q^{h_q(r/N)N}$ is a very good approximation for the size of both these objects.

Lemma 17. *For any integer $q \geq 2$, any point $x \in \mathbb{F}_q^N$ and any radius $0 \leq r \leq (1-\frac{1}{q})N$, we have*

$$\frac{1}{\sqrt{3N}} \cdot q^{h_q(r/N)N} \leq |S(x,r)| \leq |B(x,r)| \leq q^{h_q(r/N)N}.$$

Proof. We start with the rightmost inequality. Suppose without loss of generality that $x = \vec{0}$ and let $X = (X_1, X_2, \ldots, X_N)$ denote a uniformly random element in $B(\vec{0}, r)$. Note that by symmetry, since every point in $B(\vec{0}, r)$ has weight $\leq r$, we must have

$$p := \Pr[X_1 \neq 0] = \cdots = \Pr[X_N \neq 0] \leq \frac{r}{N}. \tag{2.5}$$

By subadditivity of entropy (the second point of Lemma 15), we have

$$\begin{aligned} H(X) &\leq \sum_{i=1}^{N} H(X_i) \\ &= \sum_{i=1}^{N} \Big((1-p) \log \frac{1}{1-p} + H(X_i | X_i \neq 0) \Big). \end{aligned}$$

But for any $s, t \in \{1, 2, \ldots, q-1\}$ and any $v \in \mathbb{F}_q^N$, the vectors $v^{i \mapsto s}$ and $v^{i \mapsto t}$ have the same Hamming weight (thus either both of them are in $B(\vec{0}, r)$ or neither of them are). It follows that under the condition $X_i \neq 0$, the entry X_i is uniformly distributed on $\{1, 2, \ldots, q-1\}$, which means we get

$$\begin{aligned} H(X) &\leq \sum_{i=1}^{N} \Big((1-p) \log \frac{1}{1-p} + (q-1) \cdot \frac{p}{q-1} \log \frac{q-1}{p} \Big) \\ &= \sum_{i=1}^{N} h_q(p) \cdot \log(q). \end{aligned}$$

Since X was taken uniformly at random from $B(\vec{0}, r)$, we have $H(X) = \log |B(\vec{0}, r)|$, and thus

$$\begin{aligned} |B(\vec{0}, r)| &= 2^{H(X)} \\ &\leq 2^{N h_q(p) \cdot \log(q)} \\ &= q^{h_q(p) N} \\ &\leq q^{h_q(r/N) N}, \end{aligned}$$

where the last line follows from (2.5) and from the fact that $h_q(p)$ is increasing on $p \in [0, 1 - \frac{1}{q}]$. We now turn to proving the leftmost inequality. We may assume that $1 \leq r \leq N - 1$, as otherwise the claim is trivial. We will need Stirling's inequality, which states that for any $m \in \mathbb{N}$,

$$\sqrt{2\pi m} \cdot m^m e^{-m} \leq m! \leq \sqrt{2\pi m} \cdot m^m e^{-m + \frac{1}{12m}}.$$

From Stirling's inequality, we compute

$$\begin{aligned} |S(x, r)| &= \binom{N}{r} \cdot (q-1)^r \\ &= \frac{N!}{r!(N-r)!} \cdot (q-1)^r \\ &\geq \sqrt{\frac{N}{2\pi r(N-r)}} \cdot \frac{N^N}{r^r (N-r)^{N-r}} \cdot e^{-\frac{1}{12r} - \frac{1}{12(N-r)}} \cdot (q-1)^r. \end{aligned}$$

But $r(N-r) = \frac{r}{N}(1 - \frac{r}{N}) N^2 \leq \frac{N^2}{4}$, and $e^{-\frac{1}{12r} - \frac{1}{12(N-r)}} \geq e^{-\frac{1}{6}}$ since we assumed that $1 \leq r \leq N - r$. Thus we get

$$\begin{aligned} |S(x, r)| &\geq \sqrt{\frac{2}{\pi N}} \cdot e^{-\frac{1}{6}} \cdot \frac{N^N}{r^r (N-r)^{N-r}} \cdot (q-1)^r \\ &\geq \frac{1}{\sqrt{3N}} \cdot q^{h_q(r/N) N}. \end{aligned}$$

□

Some useful facts to know about the q-ary entropy function $h_q(p)$ are that it is always larger than p and that it tends to p as q goes to infinity.

Lemma 18. *For any $q \geq 2$ and any $p \in [0, 1 - \frac{1}{q}]$, we have*

$$h_q(p) \geq p.$$

Furthermore, for any constant $p \in (0, 1)$ we have

$$\lim_{q \to \infty} h_q(p) = p.$$

Proof. For the first inequality, we note that for any $p \leq 1 - \frac{1}{q} = \frac{q-1}{q}$ we have $\log_q \frac{q-1}{p} \geq \log_q q = 1$. It then follows that

$$\begin{aligned} h_q(p) &= (1-p) \log_q \frac{1}{1-p} + p \log_q \frac{q-1}{p} \\ &\geq p \log_q \frac{q-1}{p} \\ &\geq p. \end{aligned}$$

For the second inequality, we note that by definition, we have

$$\begin{aligned} h_q(p) &= (1-p) \log_q \frac{1}{1-p} + p \log_q \frac{1}{p} + p \log_q (q-1) \\ &= \frac{h_2(p)}{\log_2 q} + p \log_q (q-1). \end{aligned}$$

Since $\frac{1}{\log_2 q} \to 0$ and $\log_q (q-1) \to 1$ as $q \to \infty$, we indeed get

$$\lim_{q \to \infty} h_q(p) = p.$$

□

Finally, we will need a way to bound the binary entropy near 1/2. The following lemma is essentially a 2-way version of Pinsker's inequality - see Appendix C.1 for the proof.

Lemma 19. *For any $\mu \in (0, 1)$, we have*

$$\frac{\mu^2}{2 \ln 2} \leq 1 - h\left(\frac{1-\mu}{2}\right) \leq \mu^2.$$

2.7 Distance and Symmetry

Two properties that will play a key role throughout this thesis are distance and symmetry.

Definition 9. *The minimum distance of a code $C \subseteq \mathbb{F}_q^N$ is the quantity*

$$d_{\min}(C) := \min_{c,c' \in C} \big\{ \mathrm{wt}(c - c') \big\}.$$

We note that if C is linear, then its minimum distance is simply the minimum Hamming weight of any codeword $c \in C$. For linear codes, we generalize the concept of minimum distance to larger subspaces as follows.

Definition 10. *For any linear code $C \subseteq \mathbb{F}_q^N$ and any $r \leq \dim C$, the r^{th} generalized distance of C is*

$$d_r(C) := \min_{\substack{S \subseteq C \\ \dim S = r}} \big\{ |\mathrm{supp}(S)| \big\},$$

where we define the support of the subcode S to be

$$\mathrm{supp}(S) := \big\{ i \in [N] : \exists c \in S \text{ with } c_i = 1 \big\}.$$

We note that for any linear code $C \subseteq \mathbb{F}_q^N$, the quantity $d_1(C)$ is simply the minimum distance of C. As we will see in Chapters 4 and 5, linear codes with large generalized distances possess useful properties. We have already seen that a code C can decode pN adversarial errors if and only if its minimum distance is at least $2pN$ (see Claim 3). But even for random errors, the distance of a code plays an important role. For instance, the probability of a decoding success for any linear code C with superconstant minimum distance transitions rapidly from $1 - o(1)$ to $o(1)$ as a function of the error rate, and the bit and block decoding thresholds of any linear code with large generalized distances lie very close to one another. See Section 4.3 and Chapter 5 for more details.

The concepts of symmetry we will rely on speak of symmetry between the different coordinates of our code C. The first level of symmetry we will consider is *transitivity*, which intuitively requires that any two coordinates $i, j \in [N]$ of the code be equivalent up to some rearrangement of the coordinates.

Definition 11. *A code $C \subseteq \mathbb{F}_q^N$ is transitive if for every coordinates $i, j \in [N]$, there exists a permutation $\pi : [N] \to [N]$ such that*

(i) $\pi(i) = j$

(ii) For every codeword $c = (c_1, c_2, \ldots, c_N) \in C$, we have $(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$.

Many well-known and widely used codes are transitive, e.g. Reed-Muller codes, Reed-Solomon codes, general BCH codes, and all cyclic codes. In addition, Reed-Muller codes and extended primitive narrow-sense BCH codes are doubly transitive. Intuitively, *double transitivity* is the requirement that any two *pairs* of coordinates (i, k), (j, l) be equivalent, again up to some rearrangement of the coordinates.

Definition 12. *A code $C \subseteq \mathbb{F}_q^N$ is doubly transitive if for every coordinates $i, j, k, \ell \in [N]$ with $i \neq k$ and $j \neq \ell$, there exists a permutation $\pi : [N] \to [N]$ such that*

(i) $\pi(i) = j$ *and* $\pi(k) = \ell$

(ii) For every codeword $c = (c_1, c_2, \ldots, c_N) \in C$, we have $(c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$.

We note that the dual code of a transitive code is transitive, and that the dual code of a doubly transitive code is doubly transitive (see Appendix B.2 for the proofs).

Claim 20. *The dual code $C^\perp$ of a transitive code $C \subseteq \mathbb{F}_2^N$ is transitive.*

Claim 21. *The dual code $C^\perp$ of a doubly transitive code $C \subseteq \mathbb{F}_2^N$ is doubly transitive.*

Kudekar, Kumar, Mondelli, Pfister, Şaşoğlu and Urbanke showed in [49] that any doubly transitive code $C \subseteq \mathbb{F}_2^N$ of rate $R \leq 1 - p$ can with high probability recover any one erased bit under $(p - o(1))$-noisy erasures[4].

[4]In the theorem below, the quantity $S_C(z)$ is the *subcode covered by* z - see (2.2).

Theorem 22 ([49]). *Let $C \subseteq \mathbb{F}_2^N$ be a doubly transitive linear code of rate R. Fix any coordinate $i \in [N]$, and define $p^* \in [0,1]$ to be the noise parameter at which*

$$\Pr_{z \sim p^*}\Big[i \in \mathrm{supp}\big(S_C(z)\big)\Big] = \frac{1}{2}.$$

Then $p^* \geq 1 - R - o(1)$. Moreover, for all $p \leq p^*$, we have

$$\Pr_{z \sim p}\Big[i \in \mathrm{supp}\big(S_C(z)\big)\Big] \leq e^{-(p^*-p)\log(N-1)}.$$

2.8 Reed-Muller Codes

Reed-Muller codes will be featured heavily in this thesis, as they have both large distances and good symmetry. We will denote by $\mathsf{RM}(n,d)$ the Reed-Muller code with n variables and degree d. The codewords of the Reed-Muller code $\mathsf{RM}(n,d)$ are the evaluation vectors (over all points in $\mathbb{F}_2^n$) of all multivariate polynomials of degree $\leq d$ in n variables. The dimension of the code is known to be $\binom{n}{\leq d}$. (See e.g. page 5 of [5]).

Fact 23. *The dimension of the Reed-Muller code $\mathsf{RM}(n,d)$ is*

$$\dim\Big(\mathsf{RM}(n,d)\Big) = \binom{n}{\leq d}.$$

Throughout this section, we let M be the generator matrix of $\mathsf{RM}(n,d)$; this is an $\binom{n}{\leq d} \times N$ matrix whose rows are indexed by subsets of $[n]$ of size at most d, and whose columns are indexed by elements of $\mathbb{F}_2^n$. For $S \subseteq [n], |S| \leq d$ and $x \in \mathbb{F}_2^n$, the entry of M whose row is indexed by S and whose column is indexed by x is

$$M_{S,x} := \prod_{j \in S} x_j.$$

If S is empty, this entry is set to 1. The parity-check matrix of the Reed-Muller code is known to be the same as the generator matrix of a different Reed-Muller code. Namely, let H be the $\binom{n}{\leq n-d-1} \times N$ generator matrix for the code $\mathsf{RM}(n, n-d-1)$. Then H has full rank, and $MH^\intercal = 0$. So, the rows of H are a basis for the orthogonal complement of the span of the rows of M. Reed-Muller codes also have well-known algebraic features, notably double transitivity.

Fact 24. *For all non-negative integers n and $d \leq n$, the Reed-Muller code $\mathsf{RM}(n,d)$ is doubly transitive.*

Proof. Recall that the codewords of the Reed-Muller code $\mathsf{RM}(n,d)$ are the evaluations (over $\mathbb{F}_2^N$) of all multilinear polynomials of degree $\leq d$ in the variables $x_1, \ldots, x_n$. But for any points $u, v, u', v' \in \mathbb{F}_2^N$ with $u \neq v$ and $u' \neq v'$, there exists an affine map $A : \mathbb{F}_2^N \to \mathbb{F}_2^N$ such that $A(u) = v$ and $A(u') = v'$. Moreover, for any affine map $A : \mathbb{F}_2^N \to \mathbb{F}_2^N$ and any polynomial $p(x_1, \ldots, x_n)$ of degree $\leq d$, the function $q(x_1, \ldots, x_n) := p(A(x_1, \ldots, x_n))$ is a polynomial of degree $\leq d$ in the variables $x_1, \ldots, x_n$. Thus the permutation of the coordinates given by $A : \mathbb{F}_2^N \to \mathbb{F}_2^N$ maps u to u', v to v', and maps any codeword of $\mathsf{RM}(n,d)$ to another codeword of $\mathsf{RM}(n,d)$. □

In addition to good symmetry, Reed-Muller codes also have large generalized distances. The following result of Wei, which states that the subcodes of Reed–Muller codes with smallest supports are Reed–Muller codes of smaller degrees on fewer variables, gives us a useful way to bound the generalized distances of $\mathsf{RM}(n,d)$.

Theorem 25 ([79])**.** *For every $t \leq d \leq n$, a $\binom{n-t}{\leq d-t}$-dimensional subcode of $\mathsf{RM}(n,d)$ with smallest support is the subcode $S_t \subseteq \mathsf{RM}(n,d)$, which contains the evaluation vectors of all polynomials of the form*

$$p(x_{t+1}, x_{t+2}, \ldots, x_n) \prod_{i=1}^{t} x_i,$$

for $p(x_{t+1}, x_{t+2}, \ldots, x_n)$ any polynomial of degree $\leq d - t$ in the variables $x_{t+1}, x_{t+2}, \ldots, x_n$.

Kudekar, Kumar, Mondelli, Pfister, Şaşoğlu and Urbanke showed in [49] that Reed-Muller codes achieve capacity on the erasure channel.

Theorem 26 ([49])**.** *For every constant $R \in (0,1)$, any sequence of Reed-Muller codes $\{\mathsf{RM}(n_i, d_i)\}$ with rate R can with high probability decode from $(1 - R - o(1))$-noisy erasures.*

Abbe and Sandon proved in [1] that Reed-Muller codes of constant rate achieve capacity on the symmetric channel.

Theorem 27 ([1]). *Consider any error parameter $p \in (0, \frac{1}{2})$, and let $\{\mathsf{RM}(n_i, d_i)\}$ be a sequence of Reed-Muller codes with rate $R < 1 - h(p)$. Then the maximum-likelihood decoder D_{ML} satisfies that for every $c \in \mathsf{RM}(n_i, d_i)$,*

$$\Pr_{z \sim p}\Big[D_{\mathrm{ML}}(c + z) = c\Big] \geq 1 - 2^{-\Omega\left(\sqrt{2^{d_i}}\right)}.$$

One major open problem in coding theory is to find polynomial-time decoding algorithms for Reed-Muller codes of constant rates.

2.9 Tensor Reed-Muller Codes

For any choice of Reed-Muller codes $\mathsf{RM}(n_1, d_1)$, $\mathsf{RM}(n_2, d_2), \ldots, \mathsf{RM}(n_t, d_t)$, we define the Tensor Reed-Muller code $\mathsf{TRM}(n_1, d_1; n_2, d_2; \ldots; n_t, d_t)$ as follows: consider $n := \sum_i n_i$ variables $\{x_{ij}\}_{i=1,\ldots,t}^{j=1,\ldots n_i}$ and define the set

$$\mathcal{S} := \Big\{S_1 \cup S_2 \cup \cdots \cup S_t : \text{ for all } i,\, S_i \subseteq \{x_{i1}, \ldots, x_{in_i}\} \text{ and } |S_i| \leq d_i\Big\}.$$

Abusing notation, we say that a monomial is in $\mathcal{S}$ if the set of its constituent variables is in $\mathcal{S}$. Then the evaluation vector of a polynomial $f(x_{11}, \ldots, x_{tn_t})$ over all points in $\{0,1\}^m$ is in $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ if and only if all the monomials of f are in $\mathcal{S}$. We note that the generator matrix of the code $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ is the tensor product of the generator matrices of the Reed-Muller codes $\mathsf{RM}(n_1, d_1), \ldots, \mathsf{RM}(n_t, d_t)$. We also note that the codewords of $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ can be seen as the t-dimensional tensors $A \in \{0,1\}^{2^{n_1} \times \cdots \times 2^{n_t}}$ satisfying the condition that for every $i \in [t]$, every i-axis vector of A is a codeword of $\mathsf{RM}(n_i, d_i)$.[5] Finally, we note that the rate of $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ is equal to the product of the rates of the codes $\big\{\mathsf{RM}(n_i, d_i)\big\}_{i=1}^{t}$.

Although we do not yet have polynomial algorithms for decoding Reed-Muller codes, Emmanuel Abbe, Colin Sandon and I showed that certain families of Tensor Reed-Muller codes can be decoded to capacity in quasilinear time. See Chapter 8.

---

[5] We say that a vector $v \in \{0,1\}^{n_i}$ is an i-axis vector of a tensor $A \in \{0,1\}^{n_1 \times \cdots \times n_t}$ if it is a "row" of A along the i^{th} axis - formally, if there exist indices $\{j_k \in [n_k]\}_{k \in [t] \setminus i}$ such that for all $s \in [n_i]$, we have $v_s = A_{j_1, \ldots, j_{i-1}, s, j_{i+1}, \ldots, j_t}$.

2.10 Fourier Analysis

The Fourier basis is a useful basis for the space of functions mapping $\mathbb{F}_2^N$ to the real numbers. We recall some of its properties below (see e.g. [19]). For $f, g \in \mathbb{F}_2^N \to \mathbb{R}$, define the inner product

$$\langle f, g \rangle := \frac{1}{2^N} \sum_{x \in \mathbb{F}_2^N} f(x)g(x).$$

For every $x, y \in \mathbb{F}_2^N$, define the character

$$\chi_y(x) := (-1)^{\sum_{j=1}^N x_j y_j}.$$

These functions form an orthonormal basis, namely for $y, y' \in \mathbb{F}_2^N$,

$$\langle \chi_y, \chi_{y'} \rangle = \begin{cases} 1 & \text{if } y = y', \\ 0 & \text{otherwise.} \end{cases}$$

We define the Fourier coefficients $\hat{f}(y) := \langle f, \chi_y \rangle$. Then for $f, g : \mathbb{F}_2^N \to \mathbb{R}$, we have

$$\langle f, g \rangle = \sum_{y \in \mathbb{F}_2^N} \hat{f}(y) \cdot \hat{g}(y).$$

In particular,

$$\frac{1}{2^N} \sum_{x \in \mathbb{F}_2^N} f(x)^2 = \sum_{y \in \mathbb{F}_2^N} \hat{f}(y)^2.$$

2.11 Krawtchouk Polynomials

For any non-negative integers N and $s \leq N$, the Krawtchouk polynomial of degree s is the real polynomial

$$K_s(x) := \sum_{j=0}^{s} (-1)^j \binom{x}{j} \binom{N-x}{s-j},$$

where for any polynomial $p(x)$ we abused notation to write $\binom{p(x)}{j} := \frac{p(x)(p(x)-1)\ldots(p(x)-j+1)}{j!}$. For any subset $S \subseteq \{0, 1, \ldots, N\}$, we will be interested in the polynomial $K_S(x) := \sum_{s \in S} K_s(x)$. For $v \in \mathbb{F}_2^N$, we will sometimes abuse notation and use $K_S(v)$ to mean $K_S(\mathrm{wt}(v))$. The

following proposition follows from standard results (see for instance [48], or Theorem 16 in [55]).

Proposition 28. *For any N and any $S \subseteq \{0, 1, \ldots, N\}$, we have*

$$\frac{2^{-N}}{\sum_{s \in S} \binom{N}{s}} \sum_{j=0}^{N} \binom{N}{j} K_S(j)^2 = 1.$$

Good estimates for Krawtchouk polynomials of any degree were obtained in [43, 40, 60] (see for e.g. [60], Lemma 2.1). These estimates are asymptotically tight in the exponent. Note that $|K_s(x)| = |K_s(N-x)| = |K_{N-s}(x)|$ by symmetry (see for e.g. equations (2.8) and (2.9) in [60]), so it suffices to understand the case $x, s \leq \frac{N}{2}$.

Theorem 29 ([43, 40, 60])**.** *Let $p, \delta \in (0, \frac{1}{2})$ be arbitrary. If $\delta \geq \frac{1}{2} - \sqrt{p(1-p)}$, then*

$$|K_{pN}(\delta N)| \leq 2^{(1+h(p)-h(\delta))\frac{N}{2}}.$$

If $\delta < \frac{1}{2} - \sqrt{p(1-p)}$, define $\omega := \frac{1-2\delta-\operatorname{sgn}(1-2\delta)\sqrt{(1-2\delta)^2-4p(1-p)}}{2(1-2\delta)}$. Then

$$|K_{pN}(\delta N)| \leq \frac{(1-\omega)^{\delta N}(1+\omega)^{(1-\delta)N}}{\omega^{pN}}.$$

As the second expression can be somewhat cumbersome to use, [60] also gives the following weaker bound.

Theorem 30 (Lemma 2.2 and equation 2.10 in [60])**.** *For any $p \in (0, \frac{1}{2})$ and any $\delta < \frac{1}{2} - \sqrt{p(1-p)}$, we have*

$$|K_{pN}(\delta N)| \leq 2^{h(p)N + pN\log(1-2\delta)}.$$

Chapter 3

FORMAL STATEMENT OF MAIN RESULTS

In this chapter, we formalize the results mentioned in Section 1.1

3.1 Leveraging Distance

Francisco Pernice, Mary Wootters and I showed that any linear code $C \subseteq \mathbb{F}_q^N$ that achieves list decoding capacity and has superconstant minimum distance also achieves unique decoding capacity over the qSC.

Theorem 31. *Fix any finite field $\mathbb{F}_q$ and any $p \in (0, \frac{1}{2})$. 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 $d_{\min}(C_n) = \omega(1)$, then $\{C_n\}$ achieves capacity on the qSC_p.*

The main tool we used to prove this fact is the following sharp transition result. The case $q = 2$ had previously been proven by Tillich and Zémor in [76].

Theorem 32 ([76, 58])**.** *Let $C \subseteq \mathbb{F}_q^N$ be any linear code with superconstant minimum distance. Then as a function of the error rate p, the interval between a decoding error probability of $o(1)$ and a decoding error probability of $1 - o(1)$ is of subconstant width.*

See Chapter 4 for the proof of Theorems 31 and 32. In another paper with Henry Pfister and Gilles Zémor, we showed that on the erasure channel, the bit-error threshold and the block-error threshold of any linear code C with large generalized distances must be asymptotically equal (see Chapter 5).

Theorem 33. *Consider any linear code $C \subseteq \mathbb{F}_2^N$ and let $p, \delta \in (0, 1)$ be such that upon transmitting any codeword $c \in C$ along the channel BEC_p, we have for any individual $i \in [N]$*

that

$$\Pr_{p\text{-erasures}}\Big[\text{cannot recover } c_i\Big] \leq \delta.$$

Suppose additionally that for all $r = 1, 2, \ldots, \sqrt{\delta}N$*, the* r^{th} *generalized distance of* C *satisfies* $d_r(C) = \omega(r \log N)$*. Then there exists* $p' = p - o(1)$ *such that*

$$\Pr_{p'\text{-erasures}}\Big[\text{cannot recover } c\Big] \leq \sqrt{\delta} + o(1).$$

In the same paper, we also showed that Reed-Muller codes satisfy the above criterion, yielding an alternative proof of the celebrated result [49, 50] that Reed-Muller codes achieve capacity on the erasure channel.

3.2 Leveraging Symmetry

Anup Rao and I derived the following bound on the weight distribution of any transitive linear code. See Chapter 6 for the proof.

Theorem 34. *Let* $C \subseteq \mathbb{F}_q^N$ *be any transitive linear code. Then for any* $\alpha \in (0, 1)$*, we have*

$$\Big|\big\{c \in C : \mathrm{wt}(c) = \alpha N\big\}\Big| \leq q^{h_q(\alpha)\cdot \dim C}.$$

As one might expect, requiring further symmetry from a linear code induces stronger bounds on its weight distribution. Samorodnitsky proved in [64] that for every doubly transitive linear code $C \subseteq \mathbb{F}_2^N$, there exists an interval $I \subseteq [0, N]$ of linear size over which the weight distribution of C is very close to the binomial distribution (see Theorem 58). We provide here a simpler argument that yields a weaker result. See Chapter 6 for the proof.

Theorem 35. *There exists a constant* $\beta > 0$ *such that the following holds: for any doubly transitive linear code* $C \subseteq \mathbb{F}_2^N$ *of rate* R *and any relative Hamming weight* $\alpha \in \left[\frac{1-R+\sqrt{\frac{1-R}{\beta}}+o(1)}{2}, \frac{1}{2}\right]$*, we have*

$$\Pr_{c\in C}\Big[\mathrm{wt}(c) = \alpha N\Big] \leq 2^{-(1-h(\alpha))N+o(N)}.$$

In joint work with Anup Rao, we obtained the following list decoding bounds for transitive and doubly transitive linear codes. See Chapter 7 for the proof.

Theorem 36. *Fix any $p \in (0, \frac{1}{2})$ and $\eta \in (0, 1)$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode p-noisy errors using a list T of size*

$$|T| = 2^{pN \log(\frac{2}{1-\eta}) + o(N)} + 2^{4pN + o(N)}.$$

Theorem 37. *Fix any $p \in (0, \frac{1}{2})$ and any $\gamma \leq 1 - \log(1 + 2^{-4p})$. Then any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1 - \gamma)N$ can with high probability list-decode p-noisy errors using a list T of size*

$$|T| = 2^{h(p)N - \gamma N + o(N)}.$$

Although our lists have exponential size, the list size is non-trivial in the sense that it is much smaller than the number of noise vectors (which is about $\binom{N}{pN} \approx 2^{h(p)N}$) and the number of codewords in the code (which is $2^{\dim C} = 2^{(1-\gamma)N}$). In fact, a standard calculation (see Section A.2) shows that any code $C \subseteq \mathbb{F}_2^N$ of dimension $(1 - \gamma)N$ that can successfully list-decode errors of probability p with list size $|T|$ must satisfy

$$|T| \gtrsim 2^{(h(p) - \gamma)N}. \tag{3.1}$$

Our bound in Theorem 37 shows that doubly transitive codes achieve these optimal parameters, at least in some regimes. (Since the requirement $\gamma \leq 1 - \log(1 + 2^{-4p})$ can be a bit hard to digest, we note e.g. that $1.3p < 1 - \log(1 + 2^{-4p})$ for all $p \in (0, \frac{1}{2})$, so Theorem 37 implies that any doubly transitive code of rate $\geq 1 - 1.3p$ achieves the optimal list size for decoding p-noisy errors).

3.3 Reed-Muller Codes: Distance and Symmetry

Reed-Muller codes are the prototypical example of a linear code with large distance, large generalized distances, and doubly transitive permutation group. They were shown to achieve

capacity on the erasure channel by Kudekar, Kumar, Mondelli, Pfister, Sasoglu and Urbanke [49], and on the symmetric channel by Abbe and Sandon [1]. However, apart from some extreme regimes - where either the rate of the code is non-constant or the capacity of the channel is much higher than the code rate - there is currently no known algorithm for decoding Reed-Muller codes in (provably) polynomial time[1]. In joint work with Emmanuel Abbe and Colin Sandon, we used existing results to prove that tensor products of Reed-Muller codes achieve capacity on the symmetric channel and are decodable in quasilinear time.

Theorem 38. *Consider any noise probability p and any rate $R < 1 - h(p)$. Then we can construct a Tensor Reed-Muller code $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ of rate R that achieves capacity and has quasilinear decoding time. For any blocklength N, we provide two constructions of such codes:*

1. *Our first construction (with $t = 3$) has error probability $N^{-\omega(\log N)}$ and decoding time $O(N \log \log N)$.*

2. *Our second construction, for any $t \geq 4$, has error probability $2^{-N^{\frac{1}{2} - \frac{1}{2(t-2)} - o(1)}}$ and decoding time $O(N \log N)$.*

The remainder of this thesis will be organized as follows:

- In Chapter 4, we prove our results relating list decoding capacity and capacity on the symmetric channel (Theorems 31 and 32).

- In Chapter 5, we prove our bound on the gap between the bit and block thresholds on the erasure channel (Theorem 33).

- In Chapter 6, we prove our weight bounds for transitive and doubly transitive codes (Theorems 34 and 35).

[1] Although they lack theoretical guarantees, several algorithms have been developed which in practice seem to perform quite well. See e.g. [65, 82, 24].

- In Chapter 7, we prove our list decoding results for transitive and doubly transitive codes (Theorems 36 and 37).

- In Chapter 8, we prove that Tensor Reed-Muller codes achieve capacity with quasilinear decoding time (Theorem 38).

Chapter 4

LARGE DISTANCE I - LIST DECODING CAPACITY IMPLIES CAPACITY ON THE SYMMETRIC CHANNEL

In this chapter, we will prove that any linear code $C \subseteq \mathbb{F}_q^N$ that has superconstant minimum distance and achieves list decoding capacity also achieves capacity on the symmetric channel. The following theorem will be our main goal.

Theorem 31. *Fix any finite field $\mathbb{F}_q$ and any $p \in (0, \frac{1}{2})$. 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 $d_{\min}(C_N) = \omega(1)$, then $\{C_N\}$ achieves capacity on the q-ary Symmetric Channel.*

We note that in the binary case $q = 2$, Sasoglu showed in [67] that if a linear code $C \subseteq \mathbb{F}_2^N$ has an $o\left(\frac{1}{N}\right)$ error probability over the BSC channel with capacity γ, then it has an $o(1)$ error probability over all binary memoryless symmetric channels (BMS) of capacity $\gamma' \geq \gamma$. Thus one corollary of our Theorem 31 is that any binary linear code C that achieves list decoding capacity and has large enough minimum distance also achieves capacity on all BMS channels.

Theorem 39. *Suppose a family of linear codes $\{C_N \subseteq \mathbb{F}_2^N\}$ achieves list-decoding capacity on the adversarial channel. If $d_{\min}(C_N) = \omega(\log^4 N)$, then $\{C_N\}$ also achieves capacity on all binary memoryless symmetric channels.*

Theorems 31 and 39 follow from a more general statement about (p, L)-list-decodability, which we state below.

Theorem 40. *Let $C \subseteq \mathbb{F}_q^N$ be any linear, (p, L)-list decodable code with minimum distance $d_{\min} \geq 4q$. Then there exists a decoder $D : \mathbb{F}_q^N \to C$ such that for any $\delta > 0$ and any $c \in C$,*

we have

$$\Pr_{z\sim p-N^{-\frac{1}{4}}-\delta}\left[D(c+z)=c\right]\geq 1-2Le^{-\frac{1-p}{4}\frac{\sqrt{d_{\min}}}{q^{3/2}}\delta},$$

where z is a $(p-N^{-\frac{1}{4}}-\delta)$-noisy random string.

We first present a high level overview of our argument in Section 4.1, before formally proving our main results in Sections 4.2 - 4.4. We also prove the following result for the erasure channel in Section 4.5. Note that in this case, we are able to drop the linearity requirement.

Theorem 41. *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 $qEC_{p'}$ for $p'=p-\frac{\log n}{\sqrt{n}}$.*

We show in Section B.3 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.

4.1 Overview of the Proof

Consider any code $C\subseteq\mathbb{F}_q^N$ that achieves list decoding capacity and let $D^*:\mathbb{F}_q^N\to C$ be a maximum-likelihood decoder on the qSC_p. 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

$$f(z):=\begin{cases}1 & \text{if } D^*(z)=\vec{0},\\ 0 & \text{otherwise.}\end{cases}\tag{4.1}$$

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

$$\mathop{\mathbb{E}}_{z\sim p'}[f(z)]=1-o(1).\tag{4.2}$$

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

$$\mathop{\mathbb{E}}_{z\sim p'}[f(z)] \geq \frac{1}{L} - o(1) \tag{4.3}$$

for, say, $p' = p - N^{-1/4}$. Indeed, consider the following decoder D: upon receiving some corrupted codeword $z \in \mathbb{F}_q^N$, find all the codewords $c \in C$ such that $d(x, c) \leq pN$, 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 $d(x, c) \leq pN$. 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. So when $z \sim p'$, 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 (4.3).

In order to deduce (4.2) from (4.3), 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 $q = 2$ by Tillich and Zémor in [83, 76] and generalized to larger q in our work [58]. Formally, our goal will be to bound the derivative of $\mathbb{E}[f]$ by

$$\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}[f(z)] \leq -\omega(1) \cdot \mathop{\mathbb{E}}_{z\sim p}[f(z)]\big(1 - \mathop{\mathbb{E}}_{z\sim p}[f(z)]\big). \tag{4.4}$$

Margulis [56] and Russo [63] pioneered the use of such inequalities for proving that the expectations of certain Boolean functions transition 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 (4.4) ensures that in this regime the derivative of $\mathbb{E}[f]$ is large. Now for any monotone[1] decreasing function $f : \mathbb{F}_q^N \to \{0, 1\}$, we can bound the derivative of the expectation as

$$\frac{d}{dp}\mathop{\mathbb{E}}_{z\sim p}[f(z)] \leq -\frac{1}{q-1}\mathop{\mathbb{E}}_{z\sim p}[h_f(z)], \tag{4.5}$$

[1]See (2.1) for a definition of monotonicity.

where

$$h_f(z) := \begin{cases} \Big|\big\{i \in [N] : z_i = 0 \text{ and } \exists a \in \mathbb{F}_q \text{ s.t. } f(z^{i \mapsto a}) = 0\big\}\Big| & \text{if } f(z) = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

For $q = 2$, the inequality (4.5) first appeared in [56, 63] and is called Russo's Lemma; our generalization to larger q is stated as Lemma 44. To make use of (4.5), we prove in Section 4.2 the following isoperimetric inequality, which is a generalization of a bound proven by Talagrand for $q = 2$:

$$\mathop{\mathbb{E}}_{z \sim p}\big[h_f(z)\big] \geq \frac{1-p}{2}\sqrt{\Delta_f} \cdot \mathop{\mathbb{E}}_{z \sim p}[f(z)]\big(1 - \mathop{\mathbb{E}}_{z \sim p}[f(z)]\big), \tag{4.7}$$

where we denoted the minimum positive value of h_f by

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

To obtain our desired inequality (4.4) from the bounds (4.5) and (4.7), it will then suffice to show that our specific function f (the indicator function of a successful decoding defined in (4.1)) satisfies $\Delta_f = \omega(1)$. That is, we want to show that if $z \in \mathbb{F}_q^N$ is some error string that leads to correct decoding, and if one of the neighbors of z leads to incorrect decoding, then there must be many such "bad" neighbors of z. Intuitively, this is because in order for z and one of its neighbors to be mapped to different codewords, it must be the case that z is about halfway between the transmitted codeword and some other codeword c. Formally, it must be the case that

$$d(z, \vec{0}) \leq d(z, c) \leq d(z, \vec{0}) + 2,$$

assuming again that the sent codeword was $\vec{0}$. For simplicity, in this section we will assume that $d(z, c) = d(z, \vec{0}) + 1$. Then for any coordinate $i \in [N]$ where $z_i = 0$ and $c_i \neq 0$, the neighbor z' of z obtained by setting the i^{th} coordinate to c_i is closer to c than to $\vec{0}$. Each such neighbor z' thus satisfies $f(z') = 0$, and we have

$$h_f(z) \geq \big|\big\{i \in [N] : z_i = 0, c_i \neq 0\big\}\big|.$$

The exact number of such coordinates $i \in [N]$ can be bounded in terms of the distance between $\vec{0}$ and c, which itself is bounded by the minimum distance of the code. See Section 4.3 for more details. Once we obtain a lower bound on Δ_f, our desired inequality (4.4) follows from equations (4.5) and (4.7).

4.2 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 monotone function $f : \mathbb{F}_q^N \to \{0,1\}$ and any point $z \in \mathbb{F}_q^N$, our goal will be to relate the quantities $\mathbb{E}[f]$ and $\mathbb{E}[h_f]$. The theorem below was proven for the case of $q = 2$ by Talagrand in [73]; we extended his result to arbitrary field sizes in [58].

Theorem 42 ([73, 58])**.** *For any monotone decreasing function $f : \mathbb{F}_q^N \to \{0,1\}$ and any noise parameter $p \in [0,1]$, we have*

$$\mathop{\mathbb{E}}_{z\sim p}\left[\sqrt{h_f(z)}\right] \geq \frac{1-p}{2} \mathop{\mathbb{E}}_{z\sim p}[f(z)]\Big(1 - \mathop{\mathbb{E}}_{z\sim p}[f(z)]\Big).$$

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 $a \in \{1, 2, \ldots, q-1\}$, 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

$$\mathbb{E}\left[\sqrt{h_f}\right] \leq \frac{1-p}{2}, \tag{4.8}$$

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

$$f_a(x) := f(xa),$$

where we define the vector $xa \in \mathbb{F}_q^N$ to have the same entries as x in coordinates $1, 2, \ldots, N-1$ and to have value a in the last coordinate. For convenience, we will denote the expectation of each of these functions by $E_a := \mathbb{E}_{z\sim p}[f_a(z)]$. By definition, we have

$$\mathop{\mathbb{E}}_{z\sim p}[f(z)] = (1-p)E_0 + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a, \tag{4.9}$$

and thus

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

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

$$\begin{aligned}
\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &= (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a(1-E_a) \\
&\quad + (1-p)pE_0\Big(E_0 - \frac{1}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a\Big) \\
&\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a\Big((1-\frac{p}{q-1})E_a - (1-p)E_0\Big) \\
&\quad - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q^*\\ b\in\mathbb{F}_q^*\setminus\{a\}}} E_aE_b.
\end{aligned} \tag{4.10}$$

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 (4.10) as

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

and

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

Combining the two equations above with (4.10), we get

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big)=(1-p)E_0(1-E_0)+\frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*}E_a(1-E_a)&\\ +(1-\frac{p}{q-1})\frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*}\big(E_0-E_a\big)^2&\\ -\frac{p^2(q-2)}{(q-1)^2}\sum_{a\in\mathbb{F}_q^*}E_0(E_0-E_a)&\\ +\sum_{b\in\mathbb{F}_q^*}\frac{p^2(q-2)}{(q-1)^2}E_bE_0&\\ -\frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q^*\\ b\in\mathbb{F}_q^*\setminus\{a\}}}E_aE_b.&\end{aligned}\tag{4.11}$$

Replacing the factor of $q-2$ in the fourth and fifth sums of (4.11) by a summation over $q-2$ elements of $\mathbb{F}_q^*$, we get

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

Defining the quantity

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

we can then bound the variance of f by

$$\begin{aligned}\mathbb{E}[f]\big(1-\mathbb{E}[f]\big) &\le (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a(1-E_a)\\ &\quad + (1-\frac{p}{q-1})p\big(E_0-E_{\min}\big)^2\\ &\quad - \frac{p^2}{(q-1)^2}\sum_{\substack{a\in\mathbb{F}_q^*\\ b\in\mathbb{F}_q^*\setminus\{a\}}}(E_0-E_b)(E_0-E_a)\\ &\le (1-p)E_0(1-E_0) + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*} E_a(1-E_a)\\ &\quad + (1-\frac{p}{q-1})p\big(E_0-E_{\min}\big)^2. \qquad (4.12)\end{aligned}$$

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. Recall that for any $z\in\mathbb{F}_q^{n-1}$ and any $a\in\mathbb{F}_q$, we defined $za\in\mathbb{F}_q^n$ to be the vector with coordinates

$$[za]_i = \begin{cases} z_i & \text{for } i\neq n,\\ a & \text{for } i=n.\end{cases}$$

By the definition of h_f, for any function $f:\mathbb{F}_q^N\to\{0,1\}$, any $z\in\mathbb{F}_q^{N-1}$ and any $a\in\mathbb{F}_q^*$, we must have

$$h_f(za) = h_{f_a}(z). \qquad (4.13)$$

In the case where we append a 0 instead of an nonzero element a, we have

$$h_f(z0) = h_{f_0}(z) + \mathbb{1}\big\{\exists b \in \mathbb{F}_q \text{ s.t. } f(zb) < f(z0)\big\}. \tag{4.14}$$

Defining the function $f^-(z) := f(z0) - \min_{a\in\mathbb{F}_q^*}\big\{f(za)\big\}$, we then get

$$\begin{aligned}
\mathop{\mathbb{E}}_{z\sim p^N}\Big[\sqrt{h_f(z)}\Big] &= (1-p)\mathop{\mathbb{E}}_{z\sim p^{N-1}}\Big[\sqrt{h_f(z0)}\Big] \\
&\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*}\mathop{\mathbb{E}}_{z\sim p^{N-1}}\Big[\sqrt{h_f(za)}\Big] \\
&= (1-p)\mathop{\mathbb{E}}_{z\sim p^{N-1}}\Big[\sqrt{h_{f_0}(z) + f^-(z)}\Big] \\
&\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*}\mathop{\mathbb{E}}_{z\sim p^{N-1}}\Big[\sqrt{h_{f_a}(z)}\Big],
\end{aligned} \tag{4.15}$$

where the second equality follows from (4.13) and (4.14). 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

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

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

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

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 (4.15), we get

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

where in the last line we used assumption (4.8) and equation (4.15) to get $\mathbb{E}\left[\sqrt{h_{f_0}}\right] \leq \frac{1}{2}$. Applying our induction hypothesis to the functions $\{f_a\}_{a\in\mathbb{F}_q}$, we then have

$$\begin{aligned}\mathbb{E}\left[\sqrt{h_f}\right] &\geq \frac{(1-p)^2}{2}E_0(1-E_0) \\ &\quad + \frac{p}{q-1}\sum_{a\in\mathbb{F}_q^*}\frac{1-p}{2}E_a(1-E_a) + \frac{1-p}{2}(E_0 - E_{\min})^2.\end{aligned}$$

Combining this with equation (4.12), we indeed get

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

□

We now denote the minimum non-zero value of $h_f(z)$ 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 43. *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 42 and the Cauchy-Schwarz inequality
$\mathbb{E}[\sqrt{g_1 g_2}]^2 \leq \mathbb{E}[g_1]\,\mathbb{E}[g_2]$, we have

$$\frac{1-p}{2} \cdot \mathbb{E}[f]\big(1-\mathbb{E}[f]\big) \leq \mathbb{E}\left[\sqrt{h_f}\right] \leq \sqrt{\mathbb{E}\left[h_f\right] \Pr_z\left[h_f(z) \neq 0\right]}.$$

By definition of Δ_f, we then get

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

□

4.3 Sharp Transition of the Probability of a Decoding Error

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 [83, 76]. We generalize their results to arbitrary field size q in [58]. The first building block of our argument is Russo's Lemma, which for $q = 2$ first appeared in [56, 63]. Over arbitrary field size q and for our particular definition of monotonicity, it generalizes as follows.

Lemma 44. *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

$$\begin{aligned}\frac{d}{dp} \mathop{\mathbb{E}}_{z\sim p}\left[f(z)\right] &= \sum_{i=1}^{N} \frac{d}{dp_i} \mathop{\mathbb{E}}_{z}\left[(1-p_i) f(z^{i\mapsto 0}) + p_i \mathop{\mathbb{E}}_{a\neq 0}\left[f(z^{i\mapsto a})\right]\right] \\ &= \sum_{i=1}^{N} \mathop{\mathbb{E}}_{\substack{z\sim p \\ a\in\mathbb{F}_q^*}}\big[-f(z^{i\mapsto 0}) + f(z^{i\mapsto a})\big], \end{aligned} \tag{4.16}$$

where a is taken uniformly at random from $\mathbb{F}_q^*$. Since f is monotone decreasing, we have $f(z^{i\mapsto 0}) \geq f(z^{i\mapsto a})$ for all $a \in \mathbb{F}_q^*$, and thus every $z \in \mathbb{F}_q^n$ has a non-positive contribution to

the expectation in (4.16). We can then bound the expectation by

$$\begin{aligned}
&\frac{d}{dp} \mathop{\mathbb{E}}_{z\sim p}\big[f(z)\big] \\
&\leq -\sum_{i\in[N]} \mathop{\mathbb{E}}_{z\sim p}\Big[\mathbb{1}\{z_i = 0\} \cdot \mathop{\mathbb{E}}_{a\in\mathbb{F}_q^*}\big[|f(z) - f(z^{i\mapsto a})|\big]\Big] \\
&\leq -\sum_{i\in[N]} \mathop{\mathbb{E}}_{z\sim p}\Big[\mathbb{1}\{z_i = 0\} \cdot \frac{1}{q-1}\mathbb{1}\big\{\exists a \text{ s.t. } |f(z) - f(z^{i\mapsto a})| = 1\big\}\Big] \\
&= -\frac{1}{q-1} \mathop{\mathbb{E}}_{z\sim p}[h_f(z)].
\end{aligned}$$

□

From Lemma 44, it is clear that for any monotone decreasing 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 45. *Let $C \subseteq \mathbb{F}_q^N$ be any linear code. Suppose $z \in \mathbb{F}_q^N$ and $c \in C \setminus \{0\}$ satisfy*

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

Then we must have

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

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

$$|S| \geq \big|\{i \in \mathrm{supp}(c) \cap \mathrm{supp}(z) : c_i = z_i\}\big|. \tag{4.17}$$

We also note that

$$\big|\{i \in \mathrm{supp}(z) \cap \mathrm{supp}(c) : c_i = z_i\}\big| \geq \frac{1}{q-1}\big|\mathrm{supp}(c) \cap \mathrm{supp}(z)\big| - \nu. \tag{4.18}$$

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

$$\begin{aligned} d(z, \alpha c) = &\big|\mathrm{supp}(z) \setminus \mathrm{supp}(c)\big| + \big|\mathrm{supp}(c) \setminus \mathrm{supp}(z)\big| \\ &+ \big|\{i \in \mathrm{supp}(z) \cap \mathrm{supp}(c) : \alpha c_i \neq z_i\}\big|, \end{aligned}$$

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

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

Since $d(z,c) \leq d(z, \alpha c) + \nu$ for every codeword αc, we then get equation (4.18). With respect to the minimum distance of our code C, this gives us

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

where in the third line we used equations (4.17) and (4.18). □

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 $\mathrm{wt}(a) < \mathrm{wt}(b)$, then $a \prec b$. If $\mathrm{wt}(a) = \mathrm{wt}(b)$ 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 $0 < 1 < 2 < ... < q-1$ over $\mathbb{F}_q$), then we say $a \prec b$. Consider the max-likelihood decoder $D^* : \mathbb{F}_q^N \to C$ defined by

$$D^*(z) := \min_{c \in C} \{z - c\}, \tag{4.19}$$

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 46. *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 Ω_0 and Ω_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 Δ_Ω and h_Ω to mean $\Delta_{\mathbb{1}_\Omega}$ and $h_{\mathbb{1}_\Omega}$ respectively.

Lemma 47. *Consider any linear code $C \subseteq \mathbb{F}_q^N$. Its corresponding decoding region Ω 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 $D^*(z^{i \mapsto c_i}) = c$.

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

$$d(z, 0) \leq d(z, c) \leq d(z, 0) + 2. \tag{4.20}$$

By Lemma 45, we must then have

$$\big|\mathrm{supp}(c) \setminus \mathrm{supp}(z)\big| \geq \frac{d_{\min}}{q} - 2. \tag{4.21}$$

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

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

$$\begin{aligned} d(z^{j \mapsto c_j}, c) &= d(z, c) - 1 \\ &\leq d(z, 0) \\ &= d(z^{j \mapsto c_j}, 0) - 1, \end{aligned}$$

and thus $z^{j \mapsto c_j} \notin \Omega$. By equation 4.21, 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

$$\begin{aligned} d(z^{j \mapsto c_j}, c) &= d(z, c) - 1 \\ &= d(z, 0) + 1 \\ &= d(z^{j \mapsto c_j}, 0). \end{aligned} \tag{4.22}$$

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

$$z^{j \mapsto c_j} - c \prec z^{j \mapsto c_j},$$

or equivalently that the support of $z^{j\mapsto c_j} - c$ comes after the support of $z^{j\mapsto c_j}$ in the lexicographic order. We note that by point (ii) above, there exists a coordinate $i \in [N]$ such that $\mathrm{supp}(z^{i\mapsto c_i} - c)$ comes after $\mathrm{supp}(z^{i\mapsto c_i})$ 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

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

where the minimum is taken over the standard order $1 < 2 < 3 \ldots < n$. Then for any coordinate $j > j^*$, $\mathrm{supp}(z^{j\mapsto c_j} - c)$ must come after $\mathrm{supp}(z^{j\mapsto c_j})$ in the lexicographic order. Combining this with equation (4.22), we get that for every coordinate $j \in \mathrm{supp}(c) \setminus \mathrm{supp}(z)$, $j \neq j^*$, we have

$$z^{j\mapsto c_j} - c \prec z^{j\mapsto c_j}.$$

By equation (4.21), there are at least $\frac{d_{\min}}{q} - 3$ such coordinates. Thus

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

□

Combining our results from Sections 4.2 and 4.3, we get the following bound on the derivative of the decoding success probability.

Lemma 48. *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 Ω 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 Ω is monotone decreasing. By Lemma 44 and Theorem 43, 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 47, 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. 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 states 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 49. *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 the function g associated with C satisfies*

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

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 48, 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

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

□

4.4 Proof of Main Results

In this section, we use our results from Section 4.3 to prove Theorems 31, 39 and 40. We start with Theorem 40.

Theorem 40. *Let $C \subseteq \mathbb{F}_q^N$ be any 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

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

where in the second inequality we used Hoeffding's inequality (Lemma 14) for the first term, and the fact that C is (p, L)-decodable for the second term. Now the maximum-likelihood decoder D^* can only have a better decoding probability than D, so we have

$$\Pr_{z \sim p - N^{-\frac{1}{4}}}[D^*(c+z) = c] \geq \frac{1}{2L}. \tag{4.23}$$

By Theorem 49 and Claim 46, 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 now turn to proving Theorems 31 and 39 from Theorem 40.

Theorem 31. *Fix any finite field $\mathbb{F}_q$ and any $p \in (0, \frac{1}{2})$. 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 $d_{\min}(C_N) = \omega(1)$, then $\{C_N\}$ achieves capacity on the q-ary Symmetric Channel.*

Proof. By Definition 7, there exists a function $p(N) = o(1)$ such that each C_N is $\big(p - p_N, d_{\min}\big)$-list decodable. Applying Theorem 40 with $\delta = d_{\min}^{-1/4} = o(1)$, we then get

$$\begin{aligned}\Pr_{z \sim p - p_N - N^{-\frac{1}{4}} - \delta}\Big[D^*(c+z) \neq c\Big] &\leq 2d_{\min} e^{-\Omega\left(d_{\min}^{1/4}\right)} \\ &\leq o(1).\end{aligned}$$

□

In order to prove Theorem 39, we will need the following result of Sasoglu [67].

Proposition 50. *[follows from [67], Proposition 7.1] Let V be the binary symmetric Channel with capacity c and V' be any binary memoryless symmetric channel with capacity $c' \geq c$. Fix any linear code $C \subseteq \{0,1\}^N$. Denote by P_V and $P_{V'}$ the probabilities that the maximum-likelihood decoder fails to recover a codeword of C sent through the channels V and V' respectively. Then we have*

$$P_{V'} \leq N \cdot P_V + h(P_V).$$

Combining Sasoglu's result with our Theorem 40 immediately gives us the following.

Theorem 39. *Suppose a family of linear codes $\{C_N \subseteq \mathbb{F}_2^N\}$ achieves list-decoding capacity on the adversarial channel. If $d_{\min}(C_n) = \omega(\log^4 N)$, then $\{C_N\}$ also achieves capacity on all binary memoryless symmetric channels.*

Proof. By Definition 7, there exists a function $p(N) = o(1)$ such that each C_N is $\big(p - p_N, N\big)$-list decodable. Applying Theorem 40 with $\delta = \frac{1}{\log N}$, we then get

$$\begin{aligned}\Pr_{z \sim p - p_N - N^{-\frac{1}{4}} - \delta}\Big[D^*(c+z) \neq c\Big] &\leq 2Ne^{-\omega(\log N)} \\ &\leq o\left(\frac{1}{N}\right).\end{aligned}$$

The claim then follows immediately from Proposition 50. □

4.5 Proof of the Erasure Case

In this Section, we give a short proof of Theorem 41, which states that list-decodability also gives good decoding guarantees on the erasure channel.

Theorem 41. *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 $qEC_{p'}$ for $p' = p - \frac{\log n}{\sqrt{n}}$.*

Proof. Fix any sent codeword $c \in C$. For every 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 14), we have

$$\Pr_{z \sim p - \frac{\log n}{\sqrt{n}}}[\mathrm{wt}(z) > pn] < e^{-2\log^2 n}. \tag{4.24}$$

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

$$\begin{aligned}\Pr_{z \sim p - \frac{\log n}{\sqrt{n}}}[c' \in S(z)] &= \left(p - \frac{\log n}{\sqrt{n}}\right)^{\mathrm{wt}(c+c')} \\ &\leq p^{\omega(\log L)} \\ &\leq o(\frac{1}{L}). \end{aligned} \tag{4.25}$$

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 (4.24) and 4.25) and applying the union bound, we thus get

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

□

Chapter 5

LARGE DISTANCE II - BIT AND BLOCK THRESHOLDS ON THE ERASURE CHANNEL

In this chapter, we will prove that the bit error threshold and the block error threshold of any linear code with large generalized Hamming weights are asymptotically equal. The following theorem is a formal version of Theorem 33.

Theorem 51. *Consider any linear code $C \subseteq \mathbb{F}_2^N$ with $N \geq 10$. Let $p, \delta \in [0,1]$ be such that* $\dim C \geq \sqrt{\delta}N$ *and*

$$\Pr_{z \sim p}\Big[i \in \operatorname{supp}\big(S_C(z)\big)\Big] \leq \delta$$

for every $i \in [N]$. Define

$$\Delta := \min_{r=1,2,\ldots,\sqrt{\delta}N} \Big\{\frac{d_r(C)}{r}\Big\}.$$

Then we have

$$\Pr_{z \sim p-\sqrt{\frac{\log N}{\Delta}}}\Big[\exists c \in C : z \succ c\Big] \leq \sqrt{\delta} + \sqrt{\frac{\log N}{\Delta}}.$$

5.1 Overview of the Proof

Our proof of Theorem 51 is based on a well-known work of Tillich and Zémor [76], who established sharp threshold results for the symmetric and erasure channels. For a linear code $C \subseteq \mathbb{F}_2^N$ and a string $x \in \{0,1\}^N$, we define

$$S_C(z) := \Big\{c \in C : c_i \leq z_i \text{ for all } i \in [N]\Big\} \tag{5.1}$$

to be the set of codewords that are indistinguishable from the 0-vector once you erase all coordinates $i \in [N]$ where $z_i = 1$. Note that $S_C(z)$ is a subcode of C. Tillich and Zémor showed in [76] that for any fixed $r \in \{1, 2, \ldots, \dim C\}$, the function

$$f_r(p) := \Pr_{z \sim p}\Big[\dim S_C(z) \geq r\Big] \tag{5.2}$$

transitions rapidly from ≈ 0 to ≈ 1 as a function of p. They also showed that for any r, the curves $f_r(p)$ and $f_{r+1}(p)$ stay within a distance of about $\frac{1}{\sqrt{d_r(C)}}$ from each other, where $d_r(C)$ denotes the minimum support weight of any r-dimensional subcode of C.

In Section 5.2, we first strengthen their result to show that the curves $f_r(p)$ and $f_{r+1}(p)$ in fact stay within a distance of about $\frac{1}{d_r(C)}$ from each other (see the proof of Theorem 53). We then leverage the fact that every doubly transitive code achieves capacity under bit-MAP decoding to prove that for some $r_0 = N^{1-o(1)}$, the function f_{r_0} corresponding to any doubly transitive code $C \subseteq \mathbb{F}_2^N$ satisfies

$$f_{r_0}\Big(1 - \mathrm{rate}(C) - o(1)\Big) = o(1).$$

Since the distance between the curves $f_r(p)$ and $f_{r+1}(p)$ can be bounded in terms of the minimum support weight $d_r(C)$, we are then able to prove (see Theorem 54) that the function $f_1(p)$ corresponding to any doubly transitive code $C \subseteq \mathbb{F}_2^N$ with large enough minimum support weights satisfies

$$f_1\Big(1 - \mathrm{rate}(C) - o(1)\Big) = o(1). \tag{5.3}$$

But the left hand side above is exactly the probability that a sent codeword $c \in C$ can be uniquely recovered from random erasures of probability $1 - \mathrm{rate}(C) - o(1)$, so equation (5.3) shows that any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ with large enough minimum support weights $\{d_r(C)\}$ achieves capacity on the erasure channel under block-MAP decoding. See Theorem 51 for the formal statement.

5.2 General Linear Codes

In this section, we will provide general conditions under which the bit and block error thresholds of an arbitrary linear code $C \subseteq \mathbb{F}_2^N$ are close to one another. For every $r \leq \dim C$, we define the function $g_r \colon \{0,1\}^N \to \{0,1\}$ to be

$$g_r(z) := \begin{cases} 1 & \text{if } \dim S_C(z) \geq r, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

For any fixed linear code $C \subseteq \mathbb{F}_2^N$, we will be interested in the expected value of its corresponding function g_r,

$$f_r(p) := \mathop{\mathbb{E}}_{z \sim p} \big[g_r(z) \big],$$

as well as the inverse map $\theta_r \colon [0,1] \to [0,1]$,

$$\theta_r(\alpha) := f_r^{-1}(\alpha). \tag{5.5}$$

We will make use of a well-known work of Tillich and Zémor [76], who proved the following result.

Lemma 52 (follows from [76], page 476)**.** *Consider any linear code $C \subseteq \mathbb{F}_2^N$. Then for every $1 \leq r \leq \dim C$ and every $z \in \mathbb{F}_2^N$ such that $g_r(z) > 0$, we have*

$$g_r(z) \geq d_r(C).$$

Additionally, for $1 \leq r \leq \dim C - 1$ we have

$$\Pr_{z \sim p} \Big[h_{g_r}(z) \neq 0 \Big] = \Pr_{z \sim p} \Big[g_r(z) = 1 \Big] - \Pr_{z \sim p} \Big[g_{r+1}(z) = 1 \Big].$$

Our first step will be to bound the distance between the curves $\theta_1(\alpha)$ and $\theta_{N_0}(\alpha)$, for any $N_0 > 1$. Recall the definition of the θ_r curves in (5.5).

Theorem 53. *Fix any linear code $C \subseteq \mathbb{F}_2^N$ and any integer $N_0 \leq \dim C$. Then, letting $\gamma_{N_0} := \sqrt{\sum_{r=1}^{N_0-1} \frac{1}{d_r(C)}}$, we have that for any*

$$0 \leq \alpha \leq 1 - \gamma_{N_0},$$

the functions θ_1 and θ_{N_0} associated with C satisfy

$$\theta_1\big(\alpha + \gamma_{N_0}\big) \geq \theta_{N_0}(\alpha) - \gamma_{N_0}.$$

Proof. Consider any $r \leq \dim C$. By Lemmas 44 and 52, the function f_r associated with our code C satisfies

$$\begin{aligned} \frac{d}{dp} f_r(p) &\geq \Pr_{z \sim p}\Big[h_{g_r}(z) \neq 0\Big] \cdot d_r(C) \\ &= \Big(f_r(p) - f_{r+1}(p)\Big) d_r(C). \end{aligned} \tag{5.6}$$

Since each θ_r is the inverse function of f_r, the area between the curves $\theta_r(\alpha)$ and $\theta_{r+1}(\alpha)$ is the same as the area between the curves $f_r(p)$ and $f_{r+1}(p)$. Thus, we find that

$$\begin{aligned} \int_0^1 \Big(\theta_{r+1}(\alpha) - \theta_r(\alpha)\Big) d\alpha &= \int_0^1 \Big(f_r(p) - f_{r+1}(p)\Big) dp \\ &\leq \frac{1}{d_r(C)}, \end{aligned} \tag{5.7}$$

where the last line follows from applying the Fundamental Theorem of Calculus to (5.6). Taking α uniformly at random from the interval $[0,1]$ and applying (5.7) with $r = 1, 2, \ldots, N_0 - 1$ then gives

$$\begin{aligned} \mathop{\mathbb{E}}_{\alpha \in [0,1]}\Big[\theta_{N_0}(\alpha) - \theta_1(\alpha)\Big] &= \int_0^1 \Big(\theta_{N_0}(\alpha) - \theta_1(\alpha)\Big) d\alpha \\ &= \sum_{r=1}^{N_0-1} \int_0^1 \Big(\theta_{r+1}(\alpha) - \theta_r(\alpha)\Big) d\alpha \\ &\leq \sum_{r=1}^{N_0-1} \frac{1}{d_r(C)} \\ &= \gamma_{N_0}^2. \end{aligned}$$

By Markov's inequality, we must then have that for α uniformly random in $[0,1]$,

$$\Pr_{\alpha\in[0,1]}\Big[\theta_{N_0}(\alpha)-\theta_1(\alpha)>\gamma_{N_0}\Big]\leq\gamma_{N_0}.$$

In particular, we see that, for any $\alpha\in\big[0,1-\gamma_{N_0}\big]$, there must be some $\alpha'\in\big[\alpha,\alpha+\gamma_{N_0}\big]$ such that

$$\theta_{N_0}(\alpha')-\theta_1(\alpha')\leq\gamma_{N_0}.$$

Since the functions θ_r are increasing, it follows that

$$\theta_{N_0}(\alpha)-\theta_1\big(\alpha+\gamma_{N_0}\big)\leq\gamma_{N_0}.$$

□

We now recall and prove our main result.

Theorem 51. *Consider any linear code $C\subseteq\mathbb{F}_2^N$ with $N\geq 10$. Let $p,\delta\in[0,1]$ be such that* $\dim C\geq\sqrt{\delta}N$ *and*

$$\Pr_{z\sim p}\Big[i\in\operatorname{supp}\big(S_C(z)\big)\Big]\leq\delta$$

for every $i\in[N]$. Define

$$\Delta:=\min_{r=1,2,\ldots,\sqrt{\delta}N}\Big\{\frac{d_r(C)}{r}\Big\}.$$

Then we have

$$\Pr_{z\sim p-\sqrt{\frac{\log N}{\Delta}}}\Big[\exists c\in C:z\succ c\Big]\leq\sqrt{\delta}+\sqrt{\frac{\log N}{\Delta}}.$$

Proof. Note that we may assume that

$$\sqrt{\delta}+\sqrt{\frac{\log N}{\Delta}}\leq 1, \tag{5.8}$$

as otherwise the claim is trivial. By linearity of expectation, we have

$$\mathop{\mathbb{E}}_{z\sim p}\Big[\operatorname{supp}\big(S_C(z)\big)\Big]\leq\delta N.$$

Since the dimension of a subspace can at most be as large as its support, Markov's inequality then gives us

$$\begin{aligned}\Pr_{z\sim p}\Big[\dim\big(S_C(z)\big) \geq \sqrt{\delta}N\Big] &\leq \frac{\mathbb{E}_{z\sim p}\Big[\mathrm{supp}(S_C(z))\Big]}{\sqrt{\delta}N} \\ &\leq \sqrt{\delta},\end{aligned}$$

or equivalently,

$$\theta_{\sqrt{\delta}N}(\sqrt{\delta}) \geq p. \tag{5.9}$$

On the other hand, by definition of Δ we have

$$\begin{aligned}\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{d_r(C)} &\leq \frac{1}{\Delta}\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{r} \\ &\leq \frac{\log N}{\Delta},\end{aligned} \tag{5.10}$$

where in the last line we used the facts that $N \geq 10$ and that $\sum_{r=1}^{N'} \frac{1}{r} \leq \ln(N') + 1$ for all $N' \geq 1$. By inequalities (5.8) and (5.10), the conditions of Theorem 53 are satisfied for $N_0 = \sqrt{\delta}N$ and $\alpha = \sqrt{\delta}$. Applying Theorem 53, we then have

$$\begin{aligned}\theta_1\left(\sqrt{\delta} + \sqrt{\frac{\log N}{\Delta}}\right) &\geq \theta_1\left(\sqrt{\delta} + \sqrt{\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{d_r(C)}}\right) \\ &\geq \theta_{\sqrt{\delta}N}(\sqrt{\delta}) - \sqrt{\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{d_r(C)}} \\ &\geq \theta_{\sqrt{\delta}N}(\sqrt{\delta}) - \sqrt{\frac{\log N}{\Delta}},\end{aligned}$$

where in the first and third lines we used the inequality (5.10). By equation (5.9), we get

$$\theta_1\left(\sqrt{\delta} + \sqrt{\frac{\log N}{\Delta}}\right) \geq p - \sqrt{\frac{\log N}{\Delta}}.$$

Since θ_1 is the inverse function of f_1, this inequality is equivalent to our desired claim. $\square$

5.3 Doubly Transitive Codes

In this section, we apply our Theorem 51 to doubly transitive codes to bound the gap between the bit and block error thresholds of any doubly transitive linear code.

Theorem 54. *Let $C \subseteq \mathbb{F}_2^N$ be any doubly transitive linear code with $N \geq 10$. Define $p^* \in [0,1]$ to be the noise parameter at which $\Pr_{z\sim p^*}\left[1 \in S_C(z)\right] = \frac{1}{2}$, and fix any $p \leq p^*$. Suppose $\dim C \geq Ne^{-\frac{p^*-p}{2}\log(N-1)}$. Then defining*

$$\Delta := \min_{r=1,2,\ldots,Ne^{-\frac{p^*-p}{2}\log(N-1)}} \left\{\frac{d_r(C)}{r}\right\},$$

we have

$$\Pr_{z\sim p-\sqrt{\frac{\log N}{\Delta}}}\left[\exists c \in C : z \succ c\right] \leq \sqrt{\frac{\log N}{\Delta}} + e^{-\frac{p^*-p}{2}\log(N-1)}.$$

Proof. By Theorem 22, we have that for all $i \in [N]$,

$$\Pr_{z\sim p}\left[i \in \operatorname{supp}\big(S_C(z)\big)\right] \leq e^{-(p^*-p)\log(N-1)}.$$

Applying Theorem 51 with $\delta = e^{-(p^*-p)\log(N-1)}$, we then get

$$\Pr_{z\sim p-\sqrt{\frac{\log N}{\Delta}}}\left[\exists c \in C : z \succ c\right] \leq \sqrt{\frac{\log N}{\Delta}} + e^{-\frac{p^*-p}{2}\log(N-1)}.$$

□

5.4 Reed–Muller Codes

We will now use Theorem 54 to show that Reed–Muller codes achieve capacity on the erasure channel. For any $t \leq d \leq n$, consider the subcode $S_t \subseteq \mathsf{RM}(n,d)$ defined in Theorem 25. Recall that Wei showed in [79] that S_t is a subcode of minimal support for its dimension (see Theorem 25). We bound its dimension and support explicitly below.

Lemma 55. *For every $n \in \mathbb{N}$, every $d \leq \frac{n}{2} + \sqrt{n\log n}$ and every $t \in [5\sqrt{n\log n}, d]$, the corresponding subcode $S_t \subseteq \mathsf{RM}(n,d)$ satisfies*

$$\left|\operatorname{supp}(S_t)\right| \geq 2^{n-t},$$

$$\dim(S_t) \leq 2^{n-t} \cdot 2^{-\frac{1}{4}\left(\frac{t^2}{n} - \frac{t^3}{n^2}\right)}.$$

Proof. The first statement follows from the fact that the evaluation vector of the monomial $\prod_{i=1}^{t} x_i$ is in S_t, and the fact that $\prod_{i=1}^{t} x_i$ evaluates to 1 on all points $x \in \mathbb{F}_2^n$ with $x_1 = x_2 = \cdots = x_t = 1$. For the second statement, we compute

$$\begin{aligned}\dim(S_t) &= \binom{n-t}{\leq d-t} \\ &\leq \binom{n-t}{\leq \frac{n}{2} - \frac{4t}{5}},\end{aligned}$$

where in the second line we used the fact that $d \leq \frac{n}{2} + \sqrt{n \log n} \leq \frac{n}{2} + \frac{t}{5}$. Since $\binom{m}{\leq s} \leq 2^{h(\frac{s}{m})m}$ for all $s \leq \frac{m}{2}$ (see for example [23], Theorem 3.1) we then get

$$\begin{aligned}\dim(S_t) &\leq 2^{h\left(\frac{\frac{1}{2} - \frac{4t}{5n}}{1 - \frac{t}{n}}\right)(n-t)} \\ &\leq 2^{h\left(\frac{1 - \frac{3t}{5n}}{2}\right)(n-t)},\end{aligned}$$

where in the second line we used the Taylor expansion $\frac{1}{1-x} = 1 + \sum_{j=1}^{\infty} x^j$ to bound $\frac{\frac{1}{2} - \frac{4x}{5}}{1-x} = \frac{1}{2} - \frac{3x}{10(1-x)} \leq \frac{1}{2} - \frac{3x}{10}$ for all $x \in [0, 1)$. Applying the inequality $h(\frac{1-x}{2}) \leq 1 - \frac{x^2}{2 \ln 2}$, we then get

$$\dim(S_t) \leq 2^{(1 - \frac{t^2}{4n^2})(n-t)},$$

as desired. □

As a corollary of Lemma 55, we get the following bound on the minimum size of the support of any r-dimensional subcode of a Reed–Muller code.

Corollary 56. *For every n large enough, every $d \leq \frac{n}{2} + \sqrt{n \log n}$ and every $p \in \left[6\sqrt{\frac{\log n}{n}}, \frac{1}{2}\right]$, we have*

$$\frac{d_r(\mathsf{RM}(n,d))}{r} \geq 2^{\frac{p^2 n}{10}}$$

for all $r \leq 2^{n - pn}$.

Proof. We first note that since the minimum distance of $\mathsf{RM}(n,d)$ is 2^{n-d}, it will suffice to prove our claim for every $r \in [2^{n-d-\frac{p^2 n}{10}}, 2^{n-pn}]$. Consider any such r, and let the integer $t \in [pn - \frac{p^2 n}{10}, d-1]$ be such that

$$2^{n-t-1-\frac{p^2 n}{10}} \leq r \leq 2^{n-t-\frac{p^2 n}{10}}. \tag{5.11}$$

Note that by Lemma 55 and our theorem's condition on p, for any $t \in [pn - \frac{p^2 n}{10}, d-1]$ we have

$$\dim(S_t) \leq 2^{n-t} \cdot 2^{-\frac{n}{4} \cdot \frac{t^2}{n^2}\left(1-\frac{t}{n}\right)}.$$

But the function $x^2(1-x)$ is increasing over $[0, \frac{2}{3}]$. Since $t \leq \frac{n}{2} + \sqrt{n \log n} \leq \frac{2n}{3}$ for all n large enough, we then get (because $\frac{t}{n} \geq p - \frac{p^2}{10}$)

$$\begin{aligned}
\dim(S_t) &\leq 2^{n-t} \cdot 2^{-\frac{n}{4}\left(p-\frac{p^2}{10}\right)^2\left(1-p+\frac{p^2}{10}\right)} \\
&\leq 2^{n-t} \cdot 2^{-\frac{n}{4}\left(\frac{19p}{20}\right)^2 \cdot \frac{1}{2}} \\
&\leq 2^{n-t} \cdot 2^{-\frac{p^2 n}{10}-1},
\end{aligned}$$

where in the second line we used the fact that $p \leq \frac{1}{2}$. Combining this with the leftmost inequality of (5.11), we get $r \geq 2^{n-t-1-\frac{p^2 n}{10}} \geq \dim(S_t)$. By Theorem 25 and Lemma 55, we must then have

$$\begin{aligned}
d_r\big(\mathsf{RM}(n,d)\big) &\geq \big|\mathrm{supp}(S_t)\big| \\
&\geq 2^{n-t}.
\end{aligned}$$

Combining this inequality with the right-hand side of (5.11), we get

$$\frac{d_r\big(\mathsf{RM}(n,d)\big)}{r} \geq 2^{\frac{p^2 n}{10}}.$$

□

We are now ready to prove that the bit and block error thresholds of Reed–Muller codes are asymptotically equal. Since every doubly transitive code achieves capacity under bit-MAP decoding [49], this implies that Reed–Muller codes achieve capacity under block-MAP decoding.

Theorem 57. *For every n large enough, every $d \in [\frac{n}{2} - \sqrt{n \log n}, \frac{n}{2} + \sqrt{n \log n}]$ and every $p \geq 20\sqrt{\frac{\log n}{n}}$, the Reed–Muller code $\mathsf{RM}(n,d)$ satisfies*

$$\Pr_{z \sim p^* - p}\Big[\exists c \in C : z \succ c\Big] \leq 2^{-\frac{p^2 n}{100}},$$

where $p^* \in [0,1]$ is such that $\Pr_{z \sim p^*}\big[1 \in \mathrm{supp}(S_C(z))\big] = \frac{1}{2}$.

Proof. Note that we may assume that $p \leq 1$. Note also that

$$\begin{aligned}\dim\big(\mathsf{RM}(n,d)\big) &= \binom{n}{\leq d} \\ &\geq \binom{n}{\leq \frac{n}{2} - \sqrt{n \log n}} \\ &\geq \frac{1}{\sqrt{2n}} \cdot 2^{h\left(\frac{1}{2} - \sqrt{\frac{\log n}{n}}\right) n},\end{aligned}$$

where in the third line we used the inequality $\binom{n}{k} \geq \frac{1}{\sqrt{2n}} \cdot 2^{h(k/n)n}$ (see for example [55], page 309, Lemma 7). Applying the inequality $h(\frac{1-x}{2}) \geq 1 - x^2$, we then get

$$\begin{aligned}\dim\big(\mathsf{RM}(n,d)\big) &\geq \frac{1}{\sqrt{2n}} 2^{(1 - 4\frac{\log n}{n})n} \\ &\geq 2^n e^{-\frac{p(n-1)}{6}}.\end{aligned}$$

By Theorem 54, for every $p \leq p^* - \frac{p}{3}$, we then have

$$\begin{aligned}\Pr_{z \sim p - \sqrt{\frac{n}{\Delta}}}\Big[\exists c \in C : z \succ c\Big] &\leq \sqrt{\frac{n}{\Delta}} + e^{-\frac{p^* - p}{2} \log(2^n - 1)} \\ &\leq \sqrt{\frac{n}{\Delta}} + e^{-\frac{p^* - p}{2}(n-1)} \qquad (5.12)\end{aligned}$$

for

$$\Delta := \min_{r = 1,2,\ldots, N e^{-\frac{p^* - p}{2}(n-1)}} \Big\{\frac{d_r\big(\mathsf{RM}(n,d)\big)}{r}\Big\}.$$

Letting $p = p^* - \frac{2p}{3} \cdot \frac{n}{n-1}$ and applying Corollary 56 with parameter $\frac{p}{3 \ln 2}$ (the conditions of Corollary 56 are satisfied by our theorem's condition on p), we get

$$\Delta \geq 2^{\frac{p^2 n}{90(\ln 2)^2}} \geq n 2^{\frac{p^2 n}{50} + 2},$$

and thus equation (5.12) becomes

$$\begin{aligned}\Pr_{z\sim p-2^{-\frac{p^2n}{100}}}\left[\exists c\in C : z\succ c\right] &\leq 2^{-\frac{p^2n}{100}-1}+e^{-\frac{pn}{3}}\\ &\leq 2^{-\frac{p^2n}{100}}.\end{aligned}$$

But by definition $p = p^* - \frac{2p}{3}\cdot\frac{n}{n-1}$ and by our theorem's conditions $2^{-\frac{p^2n}{100}} < \frac{p}{4}$, so we get

$$\Pr_{z\sim p^*-p}\left[\exists c\in C : z\succ c\right] \leq 2^{-\frac{p^2n}{100}}.$$

□

Remark. *The work of Tillich and Zémor gives very sharp block error decays in terms of the minimum distance of a linear code [76]. Thus, the primary purpose here is to bound the distance in p between the block-error and bit-error thresholds. Once we have a bound such as the one given by our Theorem 57, we immediately get a strong error decay by [76].*

Remark. *For Reed–Muller codes specifically, the ratios* $\{\frac{d_r}{r}\}$ *are large enough that one does not need the full power of Theorem 22's bit-error decay,* $\Pr_{z\sim p}[\text{bit error}] \leq e^{-(p^*-p)n}$. *To show that the bit-error and block-error thresholds of Reed–Muller codes are asymptotically equal, it would have been sufficient to use a bit error decay of* $e^{-(p^*-p)a\sqrt{n\log n}}$ *for any* $a = \omega(1)$. *This is because by Corollary 56, we have* $\frac{d_r}{r} \geq n^{10}$ *for all* $r \leq 2^{n-10\sqrt{n\log n}}$. *Setting* $p = p^* - \frac{20\ln 2}{a}$ *and applying Theorem 51 would then have given that the block-error probability under noise* $p^* - \frac{20\ln 2}{a} - \frac{1}{n^4}$ *is bounded by* $\frac{2}{n^4}$.

Chapter 6

SYMMETRY I - WEIGHT BOUNDS FOR TRANSITIVE AND DOUBLY TRANSITIVE CODES

In this chapter, we will obtain bounds on the weight distribution of transitive and doubly transitive linear codes.

6.1 Transitive Codes

We first recall and prove Theorem 34. We note that the bound we get is essentially tight, since for any finite field $\mathbb{F}_q$ and any integer divider j of N, the repetition code

$$C = \left\{(z, z, \ldots, z) \in \mathbb{F}_q^N : z \in \mathbb{F}_q^j\right\}$$

is transitive, has dimension j, and has weight distribution

$$\begin{aligned}\Pr_{c\sim\mathcal{D}(C)}\left[\mathrm{wt}(c) = \alpha N\right] &= q^{-j} \cdot \binom{j}{(1-\alpha)j}(q-1)^{\alpha j} \\ &\geq q^{-j} \cdot \sqrt{\frac{1}{2j}} \cdot 2^{h(\alpha)j} \cdot q^{\alpha j \log_q(q-1)} \\ &= \sqrt{\frac{1}{2j}} \cdot q^{-(1-h_q(\alpha))j}\end{aligned}$$

for all $\alpha \in (0,1)$ such that $\alpha j \in \mathbb{N}$.

Theorem 34. *Consider any finite field $\mathbb{F}_q$ and let $C \subseteq \mathbb{F}_q^N$ be any transitive linear code. Then for any $\alpha \in (0,1)$, we have*

$$\left|\left\{c \in C : \mathrm{wt}(c) = \alpha N\right\}\right| \leq q^{h_q(\alpha)\dim C},$$

where h_q is the q-ary entropy

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

Proof. Let $r = \dim C$, and let M the $r \times N$ generator matrix of C. Without loss of generality, suppose that the first r columns of M span the column-space of M. Define

$$C^{(\alpha)} := \{c \in C : \mathrm{wt}(c) = \alpha N\},$$

and let $Z = (Z_1, Z_2, \ldots, Z_N)$ be a uniformly random codeword in $C^{(\alpha)}$. Now C is transitive, so for every $j, k \in \{1, 2, \ldots, N\}$ the random variables Z_j and Z_k are identically distributed. By linearity of expectation and by definition of $C^{(\alpha)}$, we thus have that for every $j \in \{1, 2, \ldots, N\}$,

$$\Pr_{Z \sim \mathcal{D}(C^{(\alpha)})}[Z_j = 0] = 1 - \alpha. \tag{6.1}$$

Now for any nonzero $a, b \in \mathbb{F}_q$, there must be as many codewords $c \in C_\alpha$ with $c_j = a$ as there are codewords $c' \in C_\alpha$ with $c'_j = b$ (because C is a linear subspace, so the mapping $c \mapsto ba^{-1} \cdot c$ maps codewords to codewords). The entropy of Z_j can thus be expressed as

$$\begin{aligned} \mathop{\mathsf{H}}_{Z \sim \mathcal{D}(C^{(\alpha)})}(Z_j) &= (1-\alpha)\log\frac{1}{1-\alpha} + (q-1)\cdot\frac{\alpha}{q-1}\log\frac{q-1}{\alpha} \\ &= h_q(\alpha)\log(q). \end{aligned} \tag{6.2}$$

We will now show that $\mathsf{H}(Z_j | Z_1, Z_2, \ldots, Z_{j-1}) = 0$ for every $j > r$. To this end, fix some $j > r$. Recall that the columns $\{M_1, M_2, \ldots, M_r\}$ span the column-space of M, so we can write the column M_j as $M_j = \sum_{k=1}^{r} \beta_k M_k$ for some $\beta_1, \beta_2 \ldots, \beta_r \in \mathbb{F}$. But any codeword $c \in C$ can be expressed as $v^{(c)}M$ for some $v^{(c)} \in \mathbb{F}^r$, so any codeword $c \in C$ satisfies

$$c_j = v^{(c)} M_j = \sum_{k=1}^{r} \beta_k v^{(c)} M_k = \sum_{k=1}^{r} \beta_k c_k.$$

The random variable Z_j is thus determined by $\{Z_1, Z_2, \ldots, Z_r\}$, and so we indeed have

$$\mathop{\mathsf{H}}_{Z \sim \mathcal{D}(C^{(\alpha)})}(Z_j | Z_1, Z_2, \ldots, Z_{j-1}) = 0$$

for every $j > r$. Applying (6.2) and the chain rule for entropy then gives

$$\begin{aligned}\mathsf{H}(Z) &= \mathsf{H}(Z_1) + \sum_{i=2}^{N} \mathsf{H}(Z_i | Z_1, Z_2, \ldots, Z_{i-1}) \\ &\leq \sum_{i=1}^{r} \mathsf{H}(Z_i) \\ &\leq r \cdot h_q(\alpha) \log(q)\end{aligned}$$

But Z is sampled uniformly from $C^{(\alpha)}$, so $\mathsf{H}(Z) = \log\left(\left|C^{(\alpha)}\right|\right)$. We thus have

$$\begin{aligned}\left|C^{(\alpha)}\right| &= 2^{\mathsf{H}(Z)} \\ &\leq q^{h_q(\alpha) \cdot r}.\end{aligned}$$

□

6.2 Doubly Transitive codes

For doubly transitive codes, Samorodnitsky obtained the following weight bound.

Theorem 58 ([64])**.** *Let $C \subseteq \mathbb{F}_2^N$ be a doubly transitive linear code of rate $R := \frac{\dim C}{N}$. For any $j \in \{1, 2, \ldots, N\}$, define $j^* := \min\{j, N - j\}$. Then for any $j \in \{1, 2, \ldots, N\}$,*

$$\left|\left\{c \in C : \mathrm{wt}(c) = j\right\}\right| \leq 2^{o(N)} \cdot \left(\frac{1}{2^{1-R} - 1}\right)^{j^*}.$$

Moreover, if $j^* \geq (1 - 2^{R-1})N$,

$$\left|\left\{c \in C : \mathrm{wt}(c) = j\right\}\right| \leq 2^{o(N)} \cdot \frac{\binom{N}{j^*}|C|}{2^N}.$$

We note that in particular, there is an interval of linear length centered around $\frac{N}{2}$ where the weight distribution of any doubly transitive code of rate R behaves essentially like the weight distribution of a uniformly random code. In unpublished work, I found a much simpler proof of a weaker version of Theorem 58. It also gives that the weight distribution of a doubly transitive linear code is essentially identical to the uniform distribution over some interval of linear length, but the length of this interval is smaller than in Samorodnitsky's result.

Theorem 35. *There exists a constant $\beta > 0$ such that the following holds. For any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of rate R and any relative Hamming weight $\alpha \in \left[\frac{1-R+\sqrt{\frac{1-R}{\beta}}+o(1)}{2}, \frac{1}{2}\right]$, we have*

$$\Pr_{c \in C}\Big[\mathrm{wt}(c) = \alpha N\Big] \leq 2^{-(1-h(\alpha))N+o(N)}.$$

I will first give an overview of the proof in Section 6.2.1, before proving it formally in Sections 6.2.2 - 6.2.4.

6.2.1 Overview of the Proof

The proof of Theorem 35 is based on the result of [49] that every doubly transitive linear code achieves capacity on the erasure channel under bit-MAP decoding, as well as the idea that if a random string $x \in \{0,1\}^N$ has very small probability of having all-0s on any small set of indices, then it must be very unlikely to have too small of a Hamming weight overall (for a formal statement see our Lemma 59, which was heavily inspired by the Chernoff bound proof of [39]).

Consider any linear code $C \subseteq \mathbb{F}_2^N$ of rate R whose dual code achieves bit-capacity on the erasure channel. Then a random subset $T \subseteq [N]$ of about Rn coordinates is unlikely to fully cover the support of any dual codeword. But for any subset T that doesn't cover any dual codeword, it must be the case that a uniformly random codeword of C has uniformly random distribution on the indices of T (see Claim 60). Thus for most choices of T of size $|T| = Rn$, a random codeword in C has probability 2^{-Rn} of being all-0 on those indices.

If we could show that $\Pr_{c \in C, |T|=Rn}[c_T = \vec{0}] = 2^{-Rn+o(n)}$, then our Lemma 59 would kick in and we would get our desired weight bound. But even for codes whose dual achieves capacity on the erasure channel, we need to worry about the possibility that with probability $o(1)$, an R-noisy random string could cover $2^{\Omega(n)}$ codewords.

One way to ensure that this doesn't happen is to rely on generalized distances: as we show in Lemma 62, for any transitive code C, all the minimum subcode supports $d_r(C)$

satisfy $d_r(C) \geq \frac{r}{\text{rate}(C)}$. Combining this with the work of Tillich and Zémor [76], one can bound the probability that the covered subcode of $C^\perp$ has dimension larger than r.

6.2.2 A Criterion for Binomial Distribution

In this section, we will give a sufficient criterion for ensuring that the weight distribution of a linear code is close to the binomial distribution. We start with the following useful lemma, which is a small spin on an argument of [39].

Lemma 59. *Consider any code $C \subseteq \mathbb{F}_2^N$ and any parameter $\alpha \in (0,1)$. Suppose that for $X \in C$ a uniformly random codeword and $T \subseteq [N]$ a uniformly random subset of size $|T| = (1-2\alpha)N$, we have*

$$\Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T\Big] \leq 2^{-(1-2\alpha)N+o(N)}.$$

Then we have

$$\Pr_{X}\Big[\text{wt}(X) = \alpha N\Big] \leq 2^{-(1-h(\alpha))N+o(N)}.$$

Proof. Note that by our lemma's assumption,

$$\Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T\Big] \leq 2^{-(1-2\alpha)N+o(N)}. \tag{6.3}$$

On the other hand, we have

$$\begin{aligned}\Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T\Big] &\geq \Pr_{X}\Big[\text{wt}(X) = \alpha N\Big] \cdot \Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T \Big| \text{wt}(X) = \alpha N\Big] \\ &\geq \Pr_{X}\Big[\text{wt}(X) = \alpha N\Big] \cdot \frac{\binom{(1-\alpha)N}{(1-2\alpha)N}}{\binom{N}{(1-2\alpha)N}}. \end{aligned} \tag{6.4}$$

Combining inequalities (6.3) and (6.4), we get

$$\begin{aligned}\Pr_{X}\Big[\text{wt}(X) = \alpha N\Big] &\leq 2^{-(1-2\alpha)N+o(N)} \cdot 2^{h(1-2\alpha)N} \cdot 2^{-(1-\alpha)h(\frac{1-2\alpha}{1-\alpha})N} \\ &= 2^{-(1-h(\alpha))n+o(N)},\end{aligned}$$

where the last line follows from simple algebraic manipulations (see Appendix C.2).

□

Now that we have Lemma 59, it will suffice to establish that for a uniformly random subset T of size exactly $(1-2\alpha)N$, the probability that a uniformly random codeword c of a doubly transitive linear code C satisfies $c_T = \vec{0}$ is close to to $2^{-(1-2\alpha)n}$. As we will now show, this is equivalent to showing that the subset T of indices is unlikely to cover any codeword in the dual code $C^\perp$.

Claim 60. *Consider any linear code $C \subseteq \mathbb{F}_2^N$ and any fixed subset $T \subseteq [N]$. Then for a uniformly random codeword $c \in C$, we have*

$$\Pr_{c \in C}\Big[c_i = 0 \text{ for all } i \in T\Big] = 2^{-|T| + \dim S_{C^\perp}(T)}$$

Proof. We will first need to show that for any linear code $C \subseteq \mathbb{F}_2^N$ and any subset of coordinates $T \subseteq [N]$, the restriction $C_T \subseteq \mathbb{F}_2^{|T|}$ of C to T satisfies

$$(C_T)^\perp = \Big\{u \in \mathbb{F}_2^{|T|} : u^{\bar{T} \mapsto 0} \in C^\perp\Big\}. \tag{6.5}$$

To prove (6.5), we note that for any $u \in \mathbb{F}_2^{|T|}$, we have

$$\begin{aligned} u \in (C_T)^\perp &\iff \forall c \in C,\ \langle c_T, u\rangle = 0 \\ &\iff \forall c \in C,\ \langle c, u^{\bar{T} \mapsto 0}\rangle = 0 \\ &\iff u^{\bar{T} \mapsto 0} \in C^\perp. \end{aligned}$$

Now that we have proven (6.5), we are ready to prove our claim. Note that by linearity of C, for every vector $u \in C_T$ there are exactly $\frac{|C|}{|C_T|}$ codewords $c \in C$ with $c_T = u$. Thus when we sample a uniformly random codeword, we have

$$\begin{aligned} \Pr_{X \in C}\Big[X_i = 0 \text{ for all } i \in T\Big] &= \frac{1}{|C_T|} \\ &= \frac{|(C_T)^\perp|}{2^{|T|}}. \end{aligned}$$

Combining this with (6.5), we then get

$$\Pr_{X \in C}\Big[X_i = 0 \text{ for all } i \in T\Big] = 2^{-|T| + \dim S_{C^\perp}(T)}.$$

□

We are now ready to obtain a concrete criterion that ensures that the weight distribution of a linear code $C \subseteq \mathbb{F}_2^N$ is close to the binomial distribution.

Lemma 61. *Suppose a linear code $C \subseteq \mathbb{F}_2^N$ satisfies*

$$\mathop{\mathbb{E}}_{|x|=\tau N}\left[|S_{C^\perp}(x)|\right] \leq 2^{o(N)}$$

for some $\tau \in (0,1)$. Then for every $\alpha \in \left[\frac{1-\tau}{2}, \frac{1}{2}\right]$, we have

$$\Pr_{c \in C}\Big[\mathrm{wt}(c) = \alpha N\Big] \leq 2^{-(1-h(\alpha))N+o(N)}.$$

Proof. By Lemma 59, it suffices to prove that for $X \in C$ a uniformly random codeword and $T \subseteq [N]$ a uniformly random subset of size $|T| = (1-2\alpha)N$, we have

$$\Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T\Big] \leq 2^{-(1-2\alpha)N+o(N)}. \tag{6.6}$$

But by Claim 60, we have

$$\begin{aligned}\Pr_{X,T}\Big[X_i = 0 \text{ for all } i \in T\Big] &= \mathop{\mathbb{E}}_{T}\left[2^{-|T|+\dim S_{C^\perp}(T)}\right] \\ &= 2^{-(1-2\alpha)N} \cdot \mathop{\mathbb{E}}_{T}\left[|S_{C^\perp}(T)|\right] \\ &\leq 2^{-(1-2\alpha)N+o(N)},\end{aligned}$$

where the last line follows from our lemma's assumption and the fact that the function $S_{C^\perp}(x)$ is monotone. □

6.2.3 Bounding the Generalized Distances of Any Transitive Linear Code

In this section, we show that the r^{th} generalized distance of any transitive linear code C can be lower bounded by $\frac{r}{\mathrm{rate}(C)}$.

Lemma 62. *For any transitive linear code $C \subseteq \mathbb{F}_2^N$ and any $1 \leq r \leq \dim C$, we have*

$$\frac{d_r(C)}{r} \geq \frac{N}{\dim C}.$$

Proof. Suppose for contradiction that our claim is false, and let

$$\Delta := \min_{r=1,\ldots,\dim C} \left\{ \frac{d_r(C)}{r} \right\}.$$

Consider a subspace U of maximal dimension satisfying $\frac{|\mathrm{supp}(U)|}{\dim U} = \Delta$. Since C is transitive, there exists a permutation $\pi : [N] \to [N]$ such that

1. $\pi(U) \neq U$,

2. $\pi(U)$ is a subspace of C.

Now $U \cap \pi(U)$ is a subspace of C, so either it has dimension 0 or the ratio between its support size and its dimension is at least Δ. In either case, we must have

$$|\mathrm{supp}(U \cap \pi(U))| \geq \frac{|\mathrm{supp}(U)| \cdot \dim U \cap \pi(U)}{\dim U}. \tag{6.7}$$

But if this is true, we claim that the subspace $V := U + \pi(U)$ satisfies

$$\frac{|\mathrm{supp}(V)|}{\dim V} \leq \frac{|\mathrm{supp}(U)|}{\dim U}, \tag{6.8}$$

which would contradict the dimension-maximality of U. To prove (6.8), we note that by equation (6.7) we have

$$\begin{aligned} |\mathrm{supp}(V)| &= |\mathrm{supp}(U)| + |\mathrm{supp}(\pi(U))| - |\mathrm{supp}(U) \cap \mathrm{supp}(\pi(U))| \\ &\leq |\mathrm{supp}(U)| + |\mathrm{supp}(\pi(U))| - |\mathrm{supp}(U \cap \pi(U))| \\ &\leq 2 \cdot |\mathrm{supp}(U)| - \frac{|\mathrm{supp}(U)| \cdot \dim(U \cap \pi(U))}{\dim(U)}, \end{aligned}$$

while the dimension of V is simply

$$\begin{aligned} \dim V &= \dim(U) + \dim(\pi(U)) - \dim(U \cap \pi(U)) \\ &= 2 \cdot \dim(U) - \dim(U \cap \pi(U)). \end{aligned}$$

Combining these two expressions gives us

$$\begin{aligned}\frac{|\mathrm{supp}(V)|}{\dim V} &\leq \frac{|\mathrm{supp}(U)|}{\dim U} \cdot \frac{2 - \frac{\dim(U \cap \pi(U))}{\dim(U)}}{2 - \frac{\dim(U \cap \pi(U))}{\dim(U)}} \\ &= \frac{|\mathrm{supp}(U)|}{\dim U},\end{aligned}$$

as desired. We have proven (6.8), so we are done. □

6.2.4 Proof of our Weight Bound for Doubly Transitive Codes

We are now ready to prove our section's main result. We recall and prove Theorem 35 below.

Theorem 35. *There exists a constant $\beta > 0$ such that the following holds. For any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of rate R and any relative Hamming weight $\alpha \in \left[\frac{1-R+\sqrt{\frac{1-R}{\beta}}+o(1)}{2}, \frac{1}{2}\right]$, we have*

$$\Pr_{c \in C}\left[\mathrm{wt}(c) = \alpha N\right] \leq 2^{-(1-h(\alpha))N+o(N)}.$$

Proof. For every $r = 1, \ldots, \dim C$, define the function

$$f_r(p) := \Pr_{z \sim p}\left[\dim S_{C^\perp}(z) \geq r\right].$$

Note that since doubly transitive codes achieve capacity on the erasure channel with respect to bit-MAP decoding [49], there exists some $R_0 = R - o(1)$ and some $r_0 = o(N)$ such that

$$f_r(R_0) = o(1)$$

for all $r \geq r_0$. We pick an absolute constant $\beta > 0$ such that

$$f_r(p) \leq 2^{-\beta \cdot d_r(C^\perp)(R_0 - p)^2}$$

for every $r \geq r_0$, which is possible by [76]. For $p_0 = R_0 - \sqrt{\frac{1-R}{\beta}}$, we then get

$$\begin{aligned}f_r(p_0) &\leq 2^{-(1-R)d_r(C^\perp)} \\ &\leq 2^{-r},\end{aligned} \tag{6.9}$$

where the last line follows from Lemma 62. Since

$$\Pr_{z\sim p_0}\left[|z| = \lceil p_0 N\rceil\right] \geq \Omega\left(\frac{1}{\sqrt{N}}\right),$$

we have

$$\begin{aligned}\mathop{\mathbb{E}}_{|T|=\lceil p_0 N\rceil}\left[\left|S_{C^\perp}(T)\right|\right] &\leq 2^{r_0} + \sum_{r=r_0}^{\dim C} \Pr_{|T|=\lceil p_0 N\rceil}[\dim\, S_{C^\perp}(T) = r]\cdot 2^r \\ &\leq 2^{o(N)} + O\left(\sqrt{N}\right)\sum_{r=1}^{\dim C} f_r(p_0)\cdot 2^r \\ &\leq 2^{o(N)},\end{aligned}$$

where the third line follows from (6.9). By Lemma 61, we then get

$$\Pr_{c\in C}\left[\mathrm{wt}(c) = \alpha N\right] \leq 2^{-(1-h(\alpha))N+o(N)}$$

for every $\alpha \in \left[\frac{1-p_0}{2}, \frac{1}{2}\right]$. □

Chapter 7

SYMMETRY II - LIST DECODING BOUNDS FOR TRANSITIVE AND DOUBLY TRANSITIVE CODES

In this chapter, we will prove the list decoding results that Anup Rao and I obtained in [61] for transitive and doubly transitive codes.

Theorem 36. *Fix any $p \in (0, \frac{1}{2})$ and $\eta \in (0, 1)$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode p-noisy errors using a list T of size*

$$|T| = 2^{pN \log(\frac{2}{1-\eta}) + o(N)} + 2^{4pN + o(N)}.$$

Theorem 37. *Fix any $p \in (0, \frac{1}{2})$ and any $\gamma \leq 1 - \log(1 + 2^{-4p})$. Then any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1 - \gamma)N$ can with high probability list-decode p-noisy errors using a list T of size*

$$|T| = 2^{h(p)N - \gamma N + o(N)}.$$

7.1 Overview of the Proof

We will make use of a connection between the probability of a decoding error and the ℓ_2 norm of the coset distribution of the code. To explain the intuition, let us start by assuming that exactly pN coordinates of the codeword are flipped. Let z be the vector in $\mathbb{F}_2^N$ that represents the errors introduced by the channel and let H be the parity check matrix of the code. Then by standard arguments, if z can be recovered from Hz, the codeword can be decoded. In the case where z is uniformly distributed on vectors of weight pN, this amounts to showing that with high probability, there is no $w \in \mathbb{F}_2^N$, $w \neq z$ of weight $\mathrm{wt}(w) = pN$

such that $Hz = Hw$. This can be understood by computing the norm

$$\|f\|_2^2 := \sum_y f(y)^2 = \sum_y \Pr[Hz = y]^2,$$

where $f(y) = \Pr[Hz = y]$. This norm computes the probability that two independent, uniformly random strings z, z' of weight pN collide under the mapping $z \mapsto Hz$. Note that $\|f\|_2^2$ is always at least $\binom{N}{pN}^{-1}$, because with probability $\binom{N}{pN}^{-1}$ we have $z = z'$. If $\|f\|_2^2$ is close to $\binom{N}{pN}^{-1}$, then the code can be decoded with high probability. If $\|f\|_2^2$ is larger than $\binom{N}{pN}^{-1}$, then we show that the code can be list-decoded with high probability, where the size of the list is proportional to $\binom{N}{pN}\|f\|_2^2$.

Thus, to understand decoding, it suffices to understand $\|f\|_2^2$. Using Fourier analysis, we express this quantity as

$$\|f\|_2^2 = \frac{1}{\binom{N}{pN}^2} \sum_{j=0}^{N} \Pr[\mathrm{wt}(c^\perp) = j] \cdot K_{pN}(j)^2, \tag{7.1}$$

where $c^\perp$ is a uniformly random codeword in the dual code and K_{pN} is the Krawtchouk polynomial of degree pN. We note that (7.1) was proven in a slightly different form in [8] (see Theorem 2.1 and Lemma 4.1), whereas over $\mathbb{R}^N$ results of this type had previously been derived in [13, 71].

Using estimates for the magnitude of Krawtchouk polynomials and bounds for the weight distribution of the dual code $C^\perp$ (our Theorem 34 for transitive codes and Samorodnitsky's Theorem 58 for doubly transitive codes), one can thus bound the norm $\|f\|_2^2$ in the set-up where the error string z is a random vector of weight exactly pN. Using essentially the same techniques, one can also bound the norm $\|f\|_2^2$ when the error string z is a random vector of weight $\approx pN$, i.e. z is taken uniformly at random from the set $S = \{z \in \mathbb{F}_2^N : \mathrm{wt}(z) = pN \pm N^{3/4}\}$.

Our next step is then to show that the ℓ_2 norm corresponding to the p-noisy distribution is very similar to the ℓ_2 norm corresponding to the uniform distribution over S. Intuitively, this is because S only contains a very small range of weights, so the p-noisy distribution and the uniform distribution must behave very similarly over strings of weight in S.

7.2 Collisions vs Decoding

The goal of this section will be to analyze the relationship between the decoding of an error string and the collision probability of strings $z \in \mathbb{F}_2^N$ within the map $z \mapsto Hz$. Intuitively, the more collisions there are within this mapping, the harder it is for our decoder to correctly identify the error string z upon seeing only its image Hz. However, certain error strings might be unlikely enough to occur that our decoder can safely ignore them. For example, if we are interested in a p-noisy error string z, then z is unlikely to have weight $\mathrm{wt}(z)$ far away from pN. We could thus choose to ignore all strings whose weights do not lie in the set $S = [pN \pm l]$, for some integer l. In order to analyze the collisions that occur when strings are required to have weight $\mathrm{wt}(z) \in S$, we define for every $z \in \mathbb{F}_2^N$ and every $S \subseteq \{0, 1, \ldots, N\}$ the set of S-colliders of z, i.e. the set of strings $y \in \mathbb{F}_2^N$ that lie in the coset of z and have weight $\mathrm{wt}(y) \in S$:

Definition 13. *For any $z \in \mathbb{F}_2^N$, any matrix H with N columns and entries in $\mathbb{F}_2$, and any subset $S \subseteq \{0, 1, \ldots, N\}$, define*

$$\Omega_z^{S,H} := \left\{ y \in \mathbb{F}_2^N : \mathrm{wt}(y) \in S \text{ and } Hy = Hz \right\}.$$

When H is clear from context, we will drop the superscript and write Ω_z^S.

This definition captures a natural parameter for how large of a list we need before we can confidently claim that it contains the error string: if we are given Hz and are told that with high probability the error string z has weight $\mathrm{wt}(z) \in S$, then we should output the list Ω_z^S. For unique decoding we want to argue that $|\Omega_z^S| = 1$ with high probability, whereas for list decoding we want to argue that $|\Omega_z^S| \leq k$ with high probability, for some integer $k > 1$. The expectation of $|\Omega_z^S|$ will thus be a key quantity in our analysis. We will call this expectation the "collision count."

Definition 14. *For any subset $S \subseteq \{0, 1, \ldots, N\}$ and any matrix H with N columns and*

entries in $\mathbb{F}_2$*, define*

$$\mathsf{Coll}_H(S) := \mathop{\mathbb{E}}_{\mathrm{wt}(z)\in S}\left[|\Omega_z^S|\right].$$

When the set S only contains one or two weights (i.e. $S = \{w\}$ or $S = \{w, w'\}$), we will abuse notation and write $\mathsf{Coll}_H(w)$ and $\mathsf{Coll}_H(w, w')$ to mean $\mathsf{Coll}_H(\{w\})$ and $\mathsf{Coll}_H(\{w, w'\})$ respectively. In the following lemma, we use Markov's inequality to bound the probability of a list decoding error in terms of $\mathsf{Coll}_H(S)$.

Lemma 63. *For any subset* $S \subseteq \{0, 1, \ldots, N\}$*, any matrix* H *with* N *columns and entries in* $\mathbb{F}_2$*, and any integer* $k \geq 1$*, we have*

$$\Pr_{\mathrm{wt}(z)\in S}\left[|\Omega_z^S| > k\right] \leq \frac{\mathsf{Coll}_\mathsf{H}(S) - 1}{k}.$$

Proof. Note that $|\Omega_z^S| \geq 1$ for any $z \in \mathbb{F}_2^N$ with weight $\mathrm{wt}(z) \in S$, so the random variable $|\Omega_z^S| - 1$ is always non-negative. Applying Markov's inequality (i.e. Lemma 12), we then have

$$\begin{aligned}\Pr_{\mathrm{wt}(z)\in S}\left[|\Omega_z^S| > k\right] &= \Pr_{\mathrm{wt}(z)\in S}\left[|\Omega_z^S| - 1 \geq k\right] \\ &\leq \frac{\mathsf{Coll}_\mathsf{H}(S) - 1}{k}.\end{aligned}$$

□

When the error string z is sampled uniformly at random from the set $\{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S\}$, the above lemma allows us to relate the decoding error probability to the collision count $\mathsf{Coll}_\mathsf{H}(S)$. The problem we are most interested in, however, is when z is sampled not from some uniform distribution, but from the p-noisy probability distribution. We will now show how to connect these two decoding problems. The intuition is that by the Chernoff bound, we only need to concern ourselves with strings whose weights lie in $S = [pN \pm l]$, for some appropriately chosen l. But in this weight band all strings have similar weight, and so are given similar probability under the p-noisy distribution. Intuitively, the p-noisy decoder must then perform very similarly to the decoder that considers the uniform distribution on strings

whose weight lies in S. The following proposition makes this idea precise, and then uses Lemma 63 to bound the probability of a decoding error. We define $D_k : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^{\otimes k}$ to be the maximum-likelihood decoder

$$D_k(z) := \operatorname*{argmin}_{\substack{\{x^1,x^2,\ldots,x^k\}\subseteq\mathbb{F}_2^N \\ Hx^i=z \text{ for all } i}} \{\mathrm{wt}(x^1) + \mathrm{wt}(x^2) + \ldots + \mathrm{wt}(x^k)\},$$

where ties are broken according to the lexicographic order.

Proposition 64. *Let H be any matrix with N columns and entries in $\mathbb{F}_2$. Consider any noise parameter $p \in (0, \frac{1}{2})$ and any $l \in [1, \min\{pN, (\frac{1}{2} - p)N\}]$. Then*

(i) We have the following unique-decoding bound.

$$\Pr_{z\sim p}[z \notin D_1(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + 4(l+1) \max_{\substack{S\subseteq[pN\pm l] \\ 1\leq|S|\leq 2}} \Big\{\mathsf{Coll}_{\mathsf{H}}(S) - 1\Big\}.$$

(ii) Consider some integer $k > 1$ satisfying $\frac{k}{2l+1} \in \mathbb{N}$. Then we have the following list-decoding bound for list size k.

$$\Pr_{z\sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + \frac{4(l+1)}{k} \max_{w\in[pN\pm l]} \Big\{\mathsf{Coll}_{\mathsf{H}}(w) - 1\Big\}.$$

Proof. We will consider the unique decoding case ($k = 1$) and the list-decoding case ($k > 1$) separately.

Case 1: Unique decoding, i.e. $k = 1$

Let t be the number of rows in the matrix H. We will show that a slightly less performant decoder $\tilde{D}_1 : \mathbb{F}_2^t \to \mathbb{F}_2^N$ satisfies the desired probability bound. We define $\tilde{D}_1$ as follows: upon receiving input $x \in \mathbb{F}_2^t$, $\tilde{D}_1$ outputs the minimum-weight string from the set $\big\{z \in \mathbb{F}_2^N : Hz = x, \mathrm{wt}(z) \in [pN \pm l]\big\}$. If this set is empty, the decoder fails. If there are multiple minimal-weight strings in the set, the decoder outputs the first one in the lexicographic order. It is clear that

$$\Pr_{z\sim p}[z \neq D_1(Hz)] \leq \Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)],$$

since D_1 always returns the most likely string whereas $\tilde{D}_1$ may not. We thus turn to proving the desired bound for $\tilde{D}_1$. We first bound the probability that the error string $\mathrm{wt}(z)$ be far away from its mean. Letting

$$B = \{z \in \mathbb{F}_2^N : \big|\mathrm{wt}(z) - pN\big| \leq l\},$$

we have by Chernoff's bound (i.e. Lemma 13) that

$$\begin{aligned}\Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)] &\leq \Pr_{z\sim p}[z \notin B] + \Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)\big|z \in B] \\ &\leq 2e^{-\frac{l^2}{3pN}} + \Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)\big|z \in B]. \end{aligned} \tag{7.2}$$

We want to bound the second term. For any $z \in B$, we define the set of "problematic weights" $S(z) := \{\lceil pN - l\rceil, \lceil pN - l\rceil + 1, \ldots, \mathrm{wt}(z)\}$. We note that for $z \in B$, our decoder $\tilde{D}_1$ can only fail if there is some string $z' \neq z$ satisfying $Hz' = Hz$ and $\mathrm{wt}(z') \in S(z)$. Recalling the definition $\Omega_z^S := \{z : Hz = Hz, \mathrm{wt}(z) \in S\}$, we can then rewrite our equation (7.2) as

$$\Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + \Pr_{z\sim p}\left[|\Omega_z^{S(z)}| > 1\big|z \in B\right].$$

Considering the most problematic weight level w within the region $[pN \pm l]$ and using a union bound over all lower levels $w' \leq w$, we get

$$\begin{aligned}\Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)] &\leq 2e^{-\frac{l^2}{3pN}} + \max_{w\in[pN\pm l]}\left\{\Pr_{z\sim p}\left[|\Omega_z^{S(z)}| > 1\big|\mathrm{wt}(z) = w\right]\right\} \\ &\leq 2e^{-\frac{l^2}{3pN}} + (2l+1)\max_{\substack{w,w'\in[pN\pm l]\\ w'\leq w}}\left\{\Pr_{z\sim p}\left[|\Omega_z^{\{w,w'\}}| > 1\big|\mathrm{wt}(z) = w\right]\right\}.\end{aligned}$$

We now note that under the condition $\mathrm{wt}(z) = w$, the p-noisy probability distribution and the uniform distribution on strings of weight $\{w, w'\}$ are identical (they are both uniform on strings of weight w). We can thus rewrite our last inequality as

$$\Pr_{z\sim p}[z \neq \tilde{D}_1(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + (2l+1)\max_{\substack{w,w'\in[pN\pm l]\\ w'\leq w}}\left\{\Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[|\Omega_z^{\{w,w'\}}| > 1\big|\mathrm{wt}(z) = w\right]\right\}.$$

But by basic conditional probability we know that

$$\Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[|\Omega_z^{\{w,w'\}}|>1\right] \geq \Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[\mathrm{wt}(z)=w\right]\cdot \Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[|\Omega_z^{\{w,w'\}}|>1\middle|\mathrm{wt}(z)=w\right],$$

so we can bound our previous expression by

$$\Pr_{z\sim p}[z\neq \tilde{D}_1(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + (2l+1)\max_{\substack{w,w'\in[pN\pm l]\\ w'\leq w}}\left\{\frac{\Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[|\Omega_z^{\{w,w'\}}|>1\right]}{\Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[\mathrm{wt}(z)=w\right]}\right\}. \tag{7.3}$$

Now, from our theorem's assumption on l, we know that any $w, w' \in [pN \pm l]$ must lie in the interval $[0, \frac{N}{2}]$. Combining this with the fact that $w' \leq w$, we have

$$\Pr_{\mathrm{wt}(z)\in\{w,w'\}}\left[\mathrm{wt}(z)=w\right] = \frac{\binom{N}{w}}{\binom{N}{\{w,w'\}}} \geq \frac{\binom{N}{w}}{\binom{N}{w}+\binom{N}{w'}} \geq \frac{1}{2}. \tag{7.4}$$

It then follows from (7.3) and (7.4) that

$$\Pr_{z\sim p}\left[z\notin \tilde{D}_1(Hz)\right] \leq 2e^{-\frac{l^2}{3pN}} + 2(2l+1)\cdot\max_{\substack{S\subseteq[pN\pm l]\\ |S|\in\{1,2\}}}\left\{\Pr_{\mathrm{wt}(z)\in S}\left[|\Omega_z^S|>1\right]\right\}.$$

The theorem statement then follows from Lemma 63.

Case 2: List decoding, i.e. $k > 1$

Let t be the number of rows in the matrix H. We will show that a slightly less performant decoding function $D_{k,l} : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^{\otimes k}$ satisfies the desired probability bound. We define $D_{k,l}$ as follows: upon receiving input $x \in \mathbb{F}_2^t$, $D_{k,l}$ outputs $\frac{k}{2l+1}$ strings from $\{z \in \mathbb{F}_2^N : Hz = x, \mathrm{wt}(z) = w\}$, for each $w \in [pN \pm l]$. If there are fewer than $\frac{k}{2l+1}$ strings in some level w, the decoder returns all of them. If there are more than $\frac{k}{2l+1}$ strings in some level w, the decoder returns the first $\frac{k}{2l+1}$ ones in lexicographic order. It is clear that for any l we have

$$\Pr_{z\sim p}[z\notin D_k(Hz)] \leq \Pr_{z\sim p}[z\notin D_{k,l}(Hz)],$$

since D_k returns the k most likely strings while $D_{k,l}$ returns at most k strings. We thus turn to proving the desired bound for $D_{k,l}$. Letting

$$B = \left\{z\in\mathbb{F}_2^N : \left|\mathrm{wt}(z)-pN\right|\leq l\right\},$$

we have by Chernoff's bound (i.e. Lemma 13) that

$$\begin{aligned}\Pr_{z\sim p}[z \notin D_{k,l}(Hz)] &\leq \Pr_{z\sim p}[z \notin B] + \Pr_{z\sim p}[z \notin D_{k,l}(Hz)\big|z \in B] \\ &\leq 2e^{-\frac{l^2}{3pN}} + \max_{w\in[pN\pm l]}\left\{\Pr_{z\sim p}[z \notin D_{k,l}(Hz)\big|\mathrm{wt}(z)=w]\right\}.\end{aligned}$$

Since the distribution p gives the same probability to any two strings of equal weights, we get

$$\begin{aligned}\Pr_{z\sim p}[z \notin D_{k,l}(Hz)] &\leq 2e^{-\frac{l^2}{3pN}} + \max_{w\in[pN\pm l]}\left\{\Pr_{\mathrm{wt}(z)=w}[z \notin D_{k,l}(Hz)]\right\} \\ &\leq 2e^{-\frac{l^2}{3pN}} + \max_{w\in[pN\pm l]}\left\{\Pr_{\mathrm{wt}(z)=w}\left[|\Omega_z^{\{w\}}| > \frac{k}{2l+1}\right]\right\}.\end{aligned}$$

Applying Lemma 63, we get

$$\Pr_{z\sim p}[z \notin D_{k,l}(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + \frac{2l+1}{k}\cdot\max_{w\in[pN\pm l]}\Big\{\mathsf{Coll}_{\mathsf{H}}(w)-1\Big\}.$$

□

7.3 A Criterion for Decoding

In this section, we give a criterion that certifies that a linear code $C \subseteq \mathbb{F}_2^N$ is resilient to errors of probability p. We give such a criterion for both unique decoding and list decoding. The function we will need to make this connection is the Krawtchouk polynomial of degree s, which is defined as

$$K_s(x) := \sum_{j=0}^{s}(-1)^j\binom{x}{j}\binom{N-x}{s-j},$$

where for any polynomial $p(x)$ we abused notation to write $\binom{p(x)}{j} := \frac{p(x)(p(x)-1)\ldots(p(x)-j+1)}{j!}$. For vectors $v \in \mathbb{F}_2^N$, we will abuse notation and write $K_s(v)$ to mean $K_s(\mathrm{wt}(v))$. For convenience, we also define for any $S \subseteq \{0,1,\ldots,N\}$ the function

$$K_S(x) := \sum_{s\in S} K_s(x).$$

In the following proposition, we use basic Fourier analysis tools to rewrite the collision count $\mathsf{Coll}_H(S)$ in terms of the Krawtchouk polynomial K_S. We note that Proposition 65 was

previously proven in a different form in [8] (see Theorem 2.1 and Lemma 4.1), and can be seen as describing the coset weight distribution of the code.

Proposition 65. *Fix $p \in (0, \frac{1}{2})$, and let H be a $t \times N$ matrix with entries in $\mathbb{F}_2$. Then for any $S \subseteq \{0, 1, \ldots, N\}$, we have*

$$\mathsf{Coll}_H(S) = \frac{1}{\binom{N}{S}} \mathop{\mathbb{E}}_{v \in \mathbb{F}_2^t} [K_S(v^\intercal H)^2].$$

Proof. The main tool we will use is Parseval's Identity, which relates the evaluations $f(x)$ of a function $f : \mathbb{F}_2^t \to \mathbb{R}$ to its Fourier coefficients $\hat{f}(y)$ by

$$\frac{1}{2^t} \sum_{x \in \mathbb{F}_2^t} f(x)^2 = \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2. \tag{7.5}$$

We will first need to rewrite $\mathsf{Coll}_H(S)$ as the ℓ_2 norm of some function f. For this, we recall the definition $\Omega_z^S := \left\{ y \in \mathbb{F}_2^N : \mathrm{wt}(y) \in S \text{ and } Hy = Hz \right\}$ and note that

$$\begin{aligned}
\mathsf{Coll}_H(S) &:= \frac{1}{\binom{N}{S}} \sum_{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S} |\Omega_z^S| \\
&= \binom{N}{S} \sum_{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S} \frac{1}{|\Omega_z^S|} \Pr_{\mathrm{wt}(z') \in S} [Hz' = Hz]^2 \\
&= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} \Pr_{\mathrm{wt}(z) \in S} [Hz = x]^2.
\end{aligned}$$

We are now ready to apply Parseval's Identity. Letting $f(x) = \Pr_{\mathrm{wt}(z) \in S}[Hz = x]$ in equation (7.5), we get

$$\begin{aligned}
\mathsf{Coll}_H(S) &= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} f(x)^2 \\
&= 2^t \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2.
\end{aligned}$$

But by definition of the Fourier transform, we have

$$\hat{f}(y) := 2^{-t} \sum_{x \in \mathbb{F}_2^t} \frac{1}{\binom{N}{S}} \left| \left\{ z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S \text{ and } Hz = x \right\} \right| \cdot (-1)^{y \cdot x},$$

so our previous equation can be rewritten as

$$\begin{aligned}\mathsf{Coll}_H(S) &= 2^t\binom{N}{S}\sum_{y\in\mathbb{F}_2^t}\Big(2^{-t}\sum_{x\in\mathbb{F}_2^t}\frac{1}{\binom{N}{S}}(-1)^{y\cdot x}\cdot\big|\{z\in\mathbb{F}_2^N:\mathrm{wt}(z)\in S \text{ and } Hz=x\}\big|\Big)^2\\ &= 2^{-t}\frac{1}{\binom{N}{S}}\sum_{y\in\mathbb{F}_2^t}\Big(\sum_{\substack{z\in\mathbb{F}_2^N\\ \mathrm{wt}(z)\in S}}(-1)^{y\cdot Hz}\Big)^2. \qquad (7.6)\end{aligned}$$

We now note that by definition, for any non-negative integer $s\le N$ we have

$$\begin{aligned}K_s(y^\intercal H) &:= \sum_{j=0}^{s}(-1)^j\binom{\mathrm{wt}(y^\intercal H)}{j}\binom{N-\mathrm{wt}(y^\intercal H)}{s-j}\\ &= \sum_{\substack{z\in\mathbb{F}_2^N\\ \mathrm{wt}(z)=s}}(-1)^{y^\intercal H\cdot z},\end{aligned}$$

where we used the convention that $\binom{a}{b}=0$ when $a<b$. Combining this with equation (7.6), we get

$$\mathsf{Coll}_H(S)=\frac{2^{-t}}{\binom{N}{S}}\sum_{y\in\mathbb{F}_2^t}K_S(y^\intercal H)^2.$$

□

We will now combine Propositions 64 and 65 to obtain a bound on the decoding error probability in terms of Krawtchouk polynomials. (You want to think of the parameter l as being $l >> \sqrt{N}$ in both the case $k=1$ and the case $k>1$, so that the error term $e^{-\frac{l^2}{3pN}}$ is small).

Theorem 66. *Let H be any $t\times N$ matrix with entries in $\mathbb{F}_2$. Consider any noise parameter $p\in(0,\frac{1}{2})$ and any $l\in[1,\min\{pN,(\frac{1}{2}-p)N\}]$. Then*

(i) We have the following unique-decoding bound.

$$\Pr_{z\sim p}[z\notin D_1(Hz)]\le 2e^{-\frac{l^2}{3pN}}+4(l+1)\max_{\substack{S\subseteq[pN\pm l]\\ 1\le|S|\le 2}}\Big\{\frac{1}{\binom{N}{S}}\mathop{\mathbb{E}}_{v\in\mathbb{F}_2^t}\big[K_S(v^\intercal H)^2\big]-1\Big\}.$$

(ii) Consider some integer $k > 1$ satisfying $\frac{k}{2l+1} \in \mathbb{N}$. Then we have the following list-decoding bound for list size k.

$$\Pr_{z\sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{l^2}{3pN}} + \frac{4(l+1)}{k} \max_{w\in[pN\pm l]} \Big\{ \frac{1}{\binom{N}{w}} \mathop{\mathbb{E}}_{v\in\mathbb{F}_2^t} \big[K_w(v^\intercal H)^2\big] - 1 \Big\}.$$

Proof. The theorem statement follows directly from Propositions 64 and 65. □

One interesting consequence of Theorem 66 that if the weight distribution of the dual code $C^\perp$ is close enough to the binomial distribution, then C is resilient to p-errors. In order to state our result, we first define the set

$$A_p := \{\alpha N \in \mathbb{N} : h(\alpha) > 1 - h(p) - N^{-1/5}\}.$$

Corollary 67. *Let $p \in (0, \frac{1}{2})$ be arbitrary and let $C \subseteq \mathbb{F}_2^N$ be any linear code. Suppose that for every $j \in A_p$, we have*

$$\Pr_{y\in C^\perp} \big[\mathrm{wt}(y) = j\big] \leq \big(1 + o(N^{-1})\big) \frac{\binom{N}{j}}{2^N}.$$

Suppose also that

$$\Pr_{y\sim\mathcal{D}(C^\perp)} \big[\mathrm{wt}(y) \notin A_p\big] \leq 2^{N^{\frac{3}{4}}} \cdot \frac{\sum_{i\notin A_p} \binom{N}{i}}{2^N}.$$

Then C is resilient to p-noisy errors.

See Appendix B.1 for the proof. As a proof of concept, we note that a uniformly random linear code of dimension $(1 - h(p))N - \sqrt{N}$ satisfies all the conditions of Corollary B.1 simultaneously with high probability.

As another application of Theorem 66, we present the following bound on the probability of making a list-decoding error for a code C. We note that once again, our bound depends only on the weight distribution of the dual code $C^\perp$.

Proposition 68. *Fix any $p \in (0, \frac{1}{2})$, and define $\beta := \frac{1-2\sqrt{\tilde{p}(1-\tilde{p})}}{2}$ for $\tilde{p} = p + \frac{1}{\sqrt{\log N}}$. Let $B = [\beta N, (1-\beta)N]$, and let $k^* = (2\lfloor \frac{N}{\sqrt{\log N}} \rfloor + 1)m$ for some integer $m > 0$. Then for any*

integer $N > 2^{\frac{1}{p^2(1-p)^2}+1}$ *and all list sizes* $k \geq k^*$*, we have that any* $t \times N$ *matrix* H *with entries in* $\mathbb{F}_2$ *satisfies*

$$\Pr_{z\sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p\log N}} + \frac{N}{k^*}\max_{j\in B}\left\{\Pr_{v\in\mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H)=j]\cdot\frac{2^N}{\binom{N}{j}}-1\right\} + \frac{2^{h(p)N+5h(\frac{1}{\sqrt{\log N}})N}}{k^*}\max_{j\notin B}\left\{\Pr_{v\in\mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H)=j\right]\cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\}.$$

Proof. We will use Theorem 66 to bound the decoding error probability in terms of the Krawtchouk polynomials $K_S(j)$ and the probability factors $\Pr_{v\sim\mu_t}\left[\mathrm{wt}(v^\intercal H)=j\right]$. Some of these terms will then be bounded using Proposition 28, and some will be bounded using Theorem 30. We proceed with the proof; applying Theorem 66 to the list size k^* with parameter $l = \lfloor\frac{N}{\sqrt{\log N}}\rfloor$, we get

$$\begin{aligned}\Pr_{z\sim p}[z \notin D_k(Hz)] &\leq \Pr_{z\sim p}[z \notin D_{k^*}(Hz)] \\ &\leq 2e^{-\frac{N}{4p\log N}} + \frac{N}{k^*}\max_{w\in[pN\pm\frac{N}{\sqrt{\log N}}]}\left\{\frac{1}{\binom{N}{w}}\sum_{j=0}^{N}\Pr_{v\sim\mu_t}\left[\mathrm{wt}(v^\intercal H)=j\right]K_w(j)^2-1\right\}.\end{aligned} \tag{7.7}$$

We want to bound the summation in the second term. We will start with the central terms $j \in B$. For these we rely on Proposition 28, which states that $\frac{2^{-N}}{\binom{N}{w}}\sum_{j=0}^{N}\binom{N}{j}\cdot K_w(j)^2 = 1$ for all $w \in \{0,1,\ldots,N\}$. For any $w \in \{0,1,\ldots,N\}$, we thus get

$$\begin{aligned}\frac{1}{\binom{N}{w}}\sum_{j\in B}\Pr_{v\sim\mu_t}\left[\mathrm{wt}(v^\intercal H)=j\right]K_w(j)^2 &\leq \frac{1}{\binom{N}{w}}\max_{j\in B}\left\{\Pr_{v\in\mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H)=j]\cdot\frac{1}{\binom{N}{j}}\right\}\sum_{j\in B}\binom{N}{j}\cdot K_w(j)^2 \\ &\leq 2^N\max_{j\in B}\left\{\Pr_{v\in\mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H)=j]\cdot\frac{1}{\binom{N}{j}}\right\}.\end{aligned} \tag{7.8}$$

We then want to bound the contribution of the faraway terms $j \notin B$ to the summation in (7.7), i.e. we want to bound

$$\max_{w\in[pN\pm\frac{N}{\sqrt{\log N}}]}\left\{\frac{1}{\binom{N}{w}}\sum_{j\notin B}\Pr_{v\sim\mu_t}\left[\mathrm{wt}(v^\intercal H)=j\right]K_w(j)^2\right\}. \tag{$*$}$$

Bounding this quantity by N times its maximum value over j and applying Theorem 30, we get

$$\begin{aligned}(*) &\leq \frac{N}{\binom{N}{\lceil pN - \frac{N}{\sqrt{\log N}} \rceil}} \max_{\substack{w \in [pN \pm \frac{N}{\sqrt{\log N}}] \\ j \notin B}} \left\{ \Pr_{v \sim \mu_t} \left[\mathrm{wt}(v^\intercal H) = j \right] K_w(j)^2 \right\} \\ &\leq \frac{N}{\binom{N}{\lceil pN - \frac{N}{\sqrt{\log N}} \rceil}} \max_{\substack{w \in [pN \pm \frac{N}{\sqrt{\log N}}] \\ j \notin B}} \left\{ \Pr_{v \in \mathbb{F}_2^t} \left[\mathrm{wt}(v^\intercal H) = j \right] \cdot 2^{2h(\frac{w}{N})N + 2w \log |1 - \frac{2j}{N}|} \right\}.\end{aligned}$$

But by Lemma 17 and subadditivity of entropy (i.e. Lemma 16), we know that

$$\binom{N}{\lceil pN - \frac{N}{\sqrt{\log N}} \rceil} \geq \frac{1}{\sqrt{3N}} 2^{h(p - \frac{1}{\sqrt{\log N}})N} \geq \frac{1}{\sqrt{3N}} 2^{h(p)N - h(\frac{1}{\sqrt{\log N}})N}.$$

Additionally, for any $w \in \{pN \pm \frac{N}{\sqrt{\log N}}\}$ we have (again by subadditivity of entropy, i.e. Lemma 16)

$$2h(\frac{w}{N})N \leq 2h(p + \frac{1}{\sqrt{\log N}})N \leq 2h(p)N + 2h(\frac{1}{\sqrt{\log N}})N.$$

Finally, for any $w \in \{pN \pm \frac{N}{\sqrt{\log N}}\}$ and any $j \notin B$, we have $2w \log |1 - \frac{2j}{N}| \leq 2pN \log |1 - \frac{2j}{N}| - 2\frac{N}{\sqrt{\log N}} \log |1 - 2\beta| \leq 2pN \log |1 - \frac{2j}{N}| + h(\frac{1}{\sqrt{\log N}})N$, where the last inequality follows from our assumption that $N > 2^{\frac{1}{p^2(1-p)^2} + 1}$. Overall, we then get

$$\begin{aligned}(*) &\leq \sqrt{3} N^{\frac{3}{2}} \cdot 2^{4h(\frac{1}{\sqrt{\log N}})N} \cdot 2^{h(p)N} \max_{j \notin B} \left\{ \Pr_{v \in \mathbb{F}_2^t} \left[\mathrm{wt}(v^\intercal H) = j \right] \cdot 2^{2pN \log |1 - \frac{2j}{N}|} \right\} \\ &\leq \frac{1}{N} \cdot 2^{5h(\frac{1}{\sqrt{\log N}})N} \cdot 2^{h(p)N} \max_{j \notin B} \left\{ \Pr_{v \in \mathbb{F}_2^t} \left[\mathrm{wt}(v^\intercal H) = j \right] \cdot 2^{2pN \log |1 - \frac{2j}{N}|} \right\},\end{aligned}$$

where the last line follows from our assumption that $N > 2^{17} > 50$ and the fact that for all $N > 50$, we have $\log(\sqrt{2} N^{\frac{5}{2}}) \leq 3 \log N \leq \frac{N}{\sqrt{\log N}} \leq h(\frac{1}{\sqrt{\log N}})N$. Combining this bound for the faraway terms with our bound (7.8) for the central terms of the summation, we bound the right-hand side of equation (7.7) by

$$\begin{aligned}\Pr_{z \sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p \log N}} &+ \frac{N}{k^*} \max_{j \in B} \left\{ \Pr_{v \in \mathbb{F}_2^t} [\mathrm{wt}(v^\intercal H) = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\} \\ &+ \frac{2^{h(p)N + 5h(\frac{1}{\sqrt{\log N}})N}}{k^*} \max_{j \notin B} \left\{ \Pr_{v \in \mathbb{F}_2^t} \left[\mathrm{wt}(v^\intercal H) = j \right] \cdot 2^{2pN \log |1 - \frac{2j}{N}|} \right\}.\end{aligned}$$

□

7.4 List Decoding for Transitive Codes

We now turn to proving Theorem 36. In section 7.3, we bounded the minimum size for the decoding list of a linear code in terms of the weight distribution of its dual code. But as we stated in Claim 20, the dual code of a transitive code is also transitive. For any transitive linear code C, we can thus apply our Theorem 34 for the weight distribution of $C^\perp$ to get a bound on the size of the decoding list for C. We restate and prove our Theorem 36 below.

Theorem 36. *Fix any $p \in (0, \frac{1}{2})$ and $\eta \in (0, 1)$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode p-errors using a list T of size*

$$|T| = 2^{pN \log(\frac{2}{1-\eta}) + o(N)} + 2^{4pN + o(N)}.$$

Proof. We will show that for all $N > 2^{\frac{1}{p^2(1-p)^2}+1}$, there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size

$$|T(x)| = e^{\frac{N}{4p \log N}} \cdot 2^{5h(\frac{1}{\sqrt{\log N}})N} \cdot (2^{4p\eta N} + 2^{pN \log(\frac{2}{1-\eta})}),$$

with the property that for every codeword $c \in C$ we have

$$\Pr_{z \sim p} \left[c \notin T(c + z) \right] \leq 4e^{-\frac{N}{4p \log N}}.$$

Let H denote the parity-check matrix of C. By Lemma 9, it is sufficient to show that for any list size $k > N$, we have

$$\Pr_{z \sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p \log N}} + \frac{2^{5h(\frac{1}{\sqrt{\log N}})N+1}}{k} \cdot (2^{4p\eta N} + 2^{pN \log(\frac{2}{1-\eta})}). \tag{7.9}$$

Setting the list size $k = e^{\frac{N}{4p \log N}} \cdot 2^{5h(\frac{1}{\sqrt{\log N}})N} \cdot (2^{4p\eta N} + 2^{pN \log(\frac{2}{1-\eta})})$ in equation (7.9) will then recover our theorem statement. We thus turn to proving (7.9). We note that $2\lfloor \frac{N}{\sqrt{\log N}} \rfloor + 1 < \frac{k}{2}$, so there exists some $k^* \in [\frac{k}{2}, k]$ satisfying the conditions of Proposition 68. Proposition 68 then yields the following bound on the left-hand side of (7.9):

$$\begin{aligned} \Pr_{z \sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p \log N}} &+ \frac{2N}{k} \max_{j \in B} \left\{ \Pr_{v \in \mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H) = j] \cdot \frac{2^N}{\binom{N}{j}} \right\} \\ &+ \frac{2^{h(p)N + 5h(\frac{1}{\sqrt{\log N}})N+1}}{k} \max_{j \notin B} \left\{ \Pr_{v \in \mathbb{F}_2^t} \left[\mathrm{wt}(v^\intercal H) = j \right] \cdot 2^{2pN \log |1 - \frac{2j}{N}|} \right\}, \end{aligned} \tag{7.10}$$

where $\beta := \frac{1}{2}\left(1 - 2\sqrt{\tilde{p}(1-\tilde{p})}\right)$ for $\tilde{p} := p + \frac{1}{\sqrt{\log N}}$, and $B := [\beta N, (1-\beta)N]$. Our goal will be to bound both the central terms $j \in B$ and the faraway terms $j \notin B$ by using our bounds on the weight distribution of transitive codes. As we saw in Chapter 2, the dual code $C^\perp$ is a transitive linear code of dimension $N - \dim C$. By Theorem 34, we thus have that for all $j \in \{0, 1, \ldots, N\}$,

$$\Pr_{v \in \mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H) = j\right] \leq 2^{-(1-h(\frac{j}{N}))(1-\eta)N}. \tag{7.11}$$

For any $j \in B$, we then have by Lemma 17 that

$$\begin{aligned}\Pr_{v \in \mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H) = j\right] \cdot \frac{2^N}{\binom{N}{j}} &\leq 2^{-(1-h(j/N))(1-\eta)N} \cdot \frac{2^N}{\sqrt{\frac{1}{2N}} \cdot 2^{h(j/N)N}} \\ &= \sqrt{3N} \cdot 2^{(1-h(j/N))\eta N}.\end{aligned}$$

But for $j \in B$ we have $\beta \leq \frac{j}{N} \leq 1 - \beta$, so the right-hand side is maximized at $j = \lceil \beta N \rceil$. Applying Lemma 19, we get

$$\begin{aligned}\max_{j \in B}\left\{\Pr_{v \in \mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H) = j\right] \cdot \frac{2^N}{\binom{N}{j}}\right\} &\leq \sqrt{3N} \cdot 2^{(1-h(\beta))\eta N} \\ &\leq \sqrt{3N} \cdot 2^{4\tilde{p}(1-\tilde{p})\eta N}.\end{aligned} \tag{7.12}$$

We now turn to the faraway terms of equation (7.10). By equation (7.11), we have

$$\max_{j \notin B}\left\{\Pr_{v \in \mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H) = j\right] \cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\} \leq \max_{\delta < \beta}\left\{2^{-(1-h(\delta))(1-\eta)N} \cdot 2^{2pN\log(1-2\delta)}\right\}.$$

Note that by definition of β, any $\delta \in (0, \beta)$ can be written as $\delta = \frac{1-2\sqrt{\alpha\tilde{p}(1-\tilde{p})}}{2}$ for some $\alpha > 1$. By Lemma 19, we can then rewrite our previous expression as

$$\max_{j \notin B}\left\{\Pr_{v \in \mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H) = j\right] \cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\} \leq \max_{\alpha > 1}\left\{2^{-\frac{2\alpha\tilde{p}(1-\tilde{p})}{\ln 2}(1-\eta)N} \cdot 2^{pN\log(4\alpha\tilde{p}(1-\tilde{p}))}\right\}.$$

But for any positive constant c, the derivative of $\log(\alpha) - c\alpha$ is $\frac{1}{\alpha \cdot \ln 2} - c$, and the second derivative is always negative. Thus, the above expression achieves its maximum when $\alpha =$

$\frac{p}{2\tilde{p}(1-\tilde{p})(1-\eta)}$. We then get

$$\max_{j\notin B}\left\{\Pr_{v\in\mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H)=j]\cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\} \le 2^{-\frac{pN}{\ln 2}}\cdot 2^{pN\log(\frac{2p}{1-\eta})}$$
$$\le 2^{-h(p)N}\cdot 2^{pN\log(\frac{2}{1-\eta})}, \tag{7.13}$$

where in the last line we used the inequality $\log(1-x)\ge -\frac{x}{(1-x)\ln 2}$ for $x<1$ to get $h(p)\le -p\log(p)+\frac{p}{\ln 2}$. We now use equations (7.12) and (7.13) to bound the central and faraway terms of (7.10) respectively. This gives

$$\Pr_{z\sim p}[z\notin D_k(Hz)] \le 2e^{-\frac{N}{4p\log N}}+\frac{2N}{k}\cdot\sqrt{3N}\cdot 2^{4\tilde{p}(1-\tilde{p})\eta N}+\frac{2^{5h(\frac{1}{\sqrt{\log N}})N+1}}{k}\cdot 2^{pN\log(\frac{2}{1-\eta})}$$
$$\le 2e^{-\frac{N}{4p\log N}}+\frac{2^{5h(\frac{1}{\sqrt{\log N}})N+1}}{k}\cdot(2^{4p\eta N}+2^{pN\log(\frac{2}{1-\eta})}).$$

We have shown (7.9), and so we are done. □

7.5 List Decoding for Doubly Transitive Codes

We will now turn to proving our list-decoding bounds for doubly transitive codes. We restate and prove our Theorem 37 below.

Theorem 37. *Fix any $p\in(0,\frac{1}{2})$ and any $\gamma\le 1-\log(1+2^{-4p})$. Then any doubly transitive linear code $C\subseteq\mathbb{F}_2^N$ of dimension $\dim C=(1-\gamma)N$ can with high probability list-decode p-errors using a list T of size*

$$|T|=2^{h(p)N-\gamma N+o(N)}.$$

Proof. We will show that for all $N>2^{\frac{1}{p^2(1-p)^2}+1}$, there exists a function T mapping every $x\in\mathbb{F}_2^N$ to a subset $T(x)\subseteq C$ of size

$$|T(x)|=2^{h(p)N-\gamma N+o(N)},$$

with the property that for every codeword $c\in C$ we have

$$\Pr_{z\sim p}\left[c\notin T(c+z)\right]\le 3e^{-\frac{N}{4p\log N}}.$$

Let H denote the parity-check matrix of C. By Lemma 9, it is sufficient to show that for any $N > 2^{\frac{1}{p^2(1-p)^2}+1}$ and any list size $k > N$, we have

$$\Pr_{z\sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p\log N}} + \frac{a}{k}\cdot 2^{h(p)N-\gamma N} \tag{7.14}$$

for some $a = 2^{o(N)}$. Setting the list size $k = a\cdot e^{\frac{N}{4p\log N}}\cdot 2^{h(p)N-\gamma N}$ in equation (7.14) will then recover our theorem statement. We thus turn to proving (7.14). We note that $2\lfloor\frac{N}{\sqrt{\log N}}\rfloor+1 < \frac{k}{2}$, so there exists some $k^* \in [\frac{k}{2}, k]$ satisfying the conditions of Proposition 68. Proposition 68 then yields the following bound on the left-hand side of (7.14).

$$\begin{aligned}\Pr_{z\sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p\log N}} &+ \frac{2N}{k}\max_{j\in B}\left\{\Pr_{v\in\mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H)=j]\cdot\frac{2^N}{\binom{N}{j}}\right\}\\ &+ \frac{2^{h(p)N+5h(\frac{1}{\sqrt{\log N}})N+1}}{k}\max_{j\notin B}\left\{\Pr_{v\in\mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H)=j\right]\cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\},\end{aligned} \tag{7.15}$$

where $\beta := \frac{1}{2}\left(1-2\sqrt{\tilde{p}(1-\tilde{p})}\right)$ for $\tilde{p} := p+\frac{1}{\sqrt{\log N}}$, and $B := [\beta N, (1-\beta)N]$. Our goal will be to bound both the central terms $j \in B$ and the faraway terms $j \notin B$ by using Samorodnitsky's weight distribution bound for doubly transitive codes. Now by Claim 21, the dual code of a doubly transitive code is itself doubly transitive. Applying Theorem 58, we thus get that for all $j \in \{0, 1, \ldots, N\}$,

$$\Pr_{v\in\mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H)=j\right] \leq 2^{-\gamma N+o(N)}\cdot\left(\frac{1}{2^{1-\gamma}-1}\right)^{\min\{j,N-j\}}. \tag{7.16}$$

It then follows that

$$\max_{j\in B}\left\{\Pr_{v\in\mathbb{F}_2^t}\left[\mathrm{wt}(v^\intercal H)=j\right]\cdot\frac{2^N}{\binom{N}{j}}\right\} \leq \max_{\alpha\in[\beta,\frac{1}{2}]}\left\{2^{-\gamma N-\alpha N\log(2^{1-\gamma}-1)+N-h(\alpha)N+o(N)}\right\}. \tag{7.17}$$

We want to bound the expression on the right-hand side by $2^{h(p)N-\gamma N+o(N)}$. For this we define the function

$$f(\alpha) := -\gamma N - \alpha N\log(2^{1-\gamma}-1) + N - h(\alpha)N$$

and compute its derivative

$$\frac{df}{d\alpha} = -N\log(2^{1-\gamma}-1) - N\log\frac{1-\alpha}{\alpha}.$$

We note that over the interval $[0,1]$, the second derivative $\frac{d^2 f}{d\alpha^2} = \frac{N}{\alpha(1-\alpha)\ln 2}$ is positive. Thus over $[0,1]$, the function f is minimized at the point α^* satisfying $\frac{1-\alpha^*}{\alpha^*} = 2^{1-\gamma} - 1$ (i.e. $\alpha^* = 1 - 2^{\gamma-1}$), and f is monotone on either side of α^*. In particular, over the interval $[\beta, \frac{1}{2}]$ the function f must be maximized at either $\alpha = \beta$ or $\alpha = \frac{1}{2}$. But since $\gamma \leq 1 - \log(1 + 2^{-4p})$ by our theorem assumption, we have

$$\begin{aligned} f\big(\frac{1}{2}\big) &\leq -\gamma N + 2pN \\ &\leq -\gamma N + h(p)N. \end{aligned} \tag{7.18}$$

On the other hand we have $\beta = \frac{1-\sqrt{4p(1-p)}}{2} - o(1)$, so in order to show that

$$f(\beta) \leq h(p)N - \gamma N + o(N), \tag{7.19}$$

it suffices to show that

$$-\frac{1-\sqrt{4p(1-p)}}{2}\log(2^{1-\gamma}-1) + 1 - h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) - h(p) \leq 0.$$

But the left-hand is an increasing function of γ, so by our theorem assumption that $\gamma \leq 1 - \log(1 + 2^{-4p})$, it suffices to show that

$$2p(1-\sqrt{4p(1-p)}) + 1 - h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) - h(p) \leq 0. \tag{7.20}$$

We postpone the proof of this fact to Appendix C.3. Assuming this fact we get equation (7.19), which when combined with (7.18) and (7.17) gives us

$$\max_{j \in B}\left\{\Pr_{v \in \mathbb{F}_2^t}\big[\mathrm{wt}(v^\intercal H) = j\big] \cdot \frac{2^N}{\binom{N}{j}}\right\} \leq 2^{h(p)N - \gamma N + o(N)}. \tag{7.21}$$

This finishes our analysis of the central terms of equation (7.15). For the faraway terms, by (7.16) we have

$$\begin{aligned} \max_{j \notin B}\left\{\Pr_{v \in \mathbb{F}_2^t}[\mathrm{wt}(v^\intercal H) = j] \cdot 2^{2pN\log|1-\frac{2j}{N}|}\right\} &\leq \max_{j \leq \frac{N}{2}}\left\{2^{-\gamma N + o(N)}\left(\frac{1}{2^{1-\gamma}-1}\right)^j \cdot 2^{2pN\log(1-\frac{2j}{N})}\right\} \\ &= 2^{-\gamma N + o(N)} \max_{j \leq \frac{N}{2}}\left\{2^{-j\log(2^{1-\gamma}-1) + 2pN\log(1-\frac{2j}{N})}\right\}. \end{aligned} \tag{7.22}$$

Now the function

$$g(j) := -j \log(2^{1-\gamma} - 1) + 2pN \log(1 - \frac{2j}{N})$$

has first derivative

$$\frac{dg}{dj} = -\log(2^{1-\gamma} - 1) - \frac{4p}{\ln 2 \cdot (1 - \frac{2j}{N})},$$

and second derivative

$$\frac{dg^2}{d^2 j} = -\frac{8p}{\ln 2 \cdot N(1 - \frac{2j}{N})^2} < 0.$$

Thus $g(j)$ achieves its maximum at $j^* = \frac{N}{2} + \frac{2pN}{\ln 2 \log(2^{1-\gamma}-1)}$ and is decreasing over $[j^*, \frac{N}{2}]$. Whenever $1 - \gamma \geq \log(1 + 2^{-\frac{4p}{\ln 2}})$, we have $j^* \leq 0$; in that case the argument in equation (7.22) is maximized at $j = 0$ and we get

$$\max_{j \notin B} \left\{ \Pr_{v \in \mathbb{F}_2^t} [\mathrm{wt}(v^\intercal H) = j] \cdot 2^{2pN \log |1 - \frac{2j}{N}|} \right\} \leq 2^{-\gamma N + o(N)}.$$

We now combine this bound for the faraway terms with the bound (7.21) for the central terms to bound the right-hand side of (7.15). We get that for all $N > 2^{\frac{1}{p^2(1-p)^2}+1}$, we have

$$\Pr_{z \sim p}[z \notin D_k(Hz)] \leq 2e^{-\frac{N}{4p \log N}} + \frac{2^{o(N)}}{k} \cdot 2^{h(p)N - \gamma N}.$$

We have shown (7.14), so we are done. □

Chapter 8

SYMMETRY III - TENSOR REED-MULLER CODES ACHIEVE CAPACITY WITH QUASILINEAR DECODING TIME

This chapter will be dedicated to proving a result of Emmanuel Abbe, Colin Sandon and I, where we show that Tensor Reed-Muller codes achieve capacity on the symmetric channel and are decodable in quasilinear time.

Theorem 38. *Consider any noise probability p and any rate $R < 1 - h(p)$. Then we can construct a Tensor Reed-Muller code $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ of rate R that achieves capacity and has quasilinear decoding time. For any blocklength N, we provide two constructions of such codes:*

- *Our first construction (with $t = 3$) has error probability $N^{-\omega(\log N)}$ and decoding time $O(N \log\log N)$.*
- *Our second construction, for any $t \geq 4$, has error probability $2^{-N^{\frac{1}{2} - \frac{1}{2(t-2)} - o(1)}}$ and decoding time $O(N \log N)$.*

We will give an overview of the proof in Section 8.1, before proving it formally in Sections 8.2 to 8.4

8.1 Overview of the Proof

The starting point of our approach is the fact that there are two regimes in which we currently have efficient decoding algorithms for Reed-Muller codes:

1. When the noise is smaller than $1 - \frac{\binom{n}{\leq n-t}}{2^n}$, one can use Theorem 10 to decode a code $\mathsf{RM}(n - 2t, n)$ of length $N = 2^n$.

2. When the blocklength of the code is very small, brute-force decoding, which runs in time $O(2^N)$, may have reasonable runtime.

We will combine the two regimes above to obtain an efficient decoder for Tensor Reed-Muller codes. We take a short code $\mathsf{RM}(n_1, d_1)$ of rate $R - o(1)$ and tensor it with a longer code $\mathsf{RM}(n_2, d_2)$ of rate $1 - o(1)$. The codewords of the resulting code $\mathsf{TRM}(n_1, d_1; n_2, d_2)$ are all the matrices $A \in \{0,1\}^{2^{n_2} \times 2^{n_1}}$ such that each row of A is a codeword of $\mathsf{RM}(n_1, d_1)$ and each column of A is a codeword of $\mathsf{RM}(n_2, d_2)$. To decode our tensor code, we first use the brute-force algorithm to decode each row of A independently. After this first step, which takes polynomial time as long as $2^{n_1} \approx \log N$, only a $o(1)$ fraction of A's entries will have been incorrectly decoded. We then use the high-rate algorithm of [66] (see Theorem 10) to decode each column of A independently. At the end of this second decoding step, the fraction of incorrectly decoded entries will have dropped below $o\left(\frac{1}{N}\right)$, allowing us to take a union bound over all coordinates. To further reduce the decoding error probability, we can take the tensor product of $\mathsf{TRM}(n_1, d_1; n_2, d_2)$ with an even longer RM code $\mathsf{RM}(n_3, d_3)$ of rate $1 - o(1)$ and repeat the same argument.

The ideas outlined above, further iterated, would allow us to decode t-Tensor Reed-Muller codes (for some t) with a decoding failure probability of about $2^{-N^{1/4}}$. To bring this error rate closer to the distance-optimal $2^{-\Omega(\sqrt{N})}$, we introduce a new algorithm for decoding tensor codes from adversarial errors. This algorithm works for any tensor code $C_1 \otimes \ldots \otimes C_t$ and relies on the fact that erasures are generally easier to decode than errors.[1] First, replace each row that is not a codeword of C_1 by an all-erasures row. Then, go through each column and determine whether or not there is a unique codeword $c \in C_2$ compatible with the (now partially-erased) column. If so, replace the column by c; otherwise, replace every entry in the column by an erasure symbol. For $t > 2$, repeat this process with every additional axis. By

[1]For any linear code $C \subseteq \{0,1\}^N$, given a partially-erased codeword of C, one can use the parity-check matrix to obtain a system of $N - \dim C$ linear equations in $e \leq N$ unknowns, where e is the number of erased coordinates. This can be solved by Gaussian elimination in time $O(N^3)$. For Reed-Muller codes, we improve this decoding time to $O(N \log N)$ whenever the number of erasures is below the minimum distance - see Lemma 69.

adding additional checks that ensure we never return subtensors that are too far away from the corresponding input, we obtain in Theorem 71 an algorithm for decoding any arbitrary tensor code $C = C_1 \otimes \ldots \otimes C_t$ from $\frac{d_{\min}(C)}{2 \max_i \{d_{\min}(C_i)\}} - 1$ adversarial errors. For Reed-Muller codes, this algorithm runs in time $O(N \log N)$.

Our final construction combines the ideas of the above two paragraphs: the first two Reed-Muller codes $\mathsf{RM}(n_1, d_1)$ and $\mathsf{RM}(n_2, d_2)$ are of subpolynomial lengths and taken as in the first paragraph. As mentioned above, this allows us to bring the error probability down to about $N^{-\omega(\log N)}$. The remaining Reed-Muller codes $\left\{\mathsf{RM}(n_i, d_i)\right\}_{i=3,\ldots,t}$ all have $n_i = \frac{\log N - n_1 - n_2}{t-2}$ and $d_i = \frac{n_i + n_i^{3/4}}{2}$. By the arguments we outlined in the second paragraph, we can recover any sent codeword of $\mathsf{TRM}(n_3, d_3; \ldots; n_t, d_t)$ with fewer than about $N^{\frac{1}{2} - \frac{1}{2(t-2)}}$ errors. But since the first pass on $\mathsf{RM}(n_1, d_1)$ and $\mathsf{RM}(n_2, d_2)$ brought the error rate down to $N^{-\omega(\log N)}$, by the Chernoff bound, the probability that there are more than $N^{\frac{1}{2} - \frac{1}{2(t-2)}}$ errors is bounded by $2^{-N^{\frac{1}{2} - \frac{1}{2(t-2)} - o(1)}}$.

8.2 Helpful Lemmas

In this section, we will prove performance guarantees for two decoding algorithms that will be used as subroutines throughout this paper. We start with an algorithm for efficiently testing and correcting Reed-Muller codes from adversarial erasures.

Lemma 69. *For any nonnegative integers $d \leq n$, there is an $O(n2^n)$-time algorithm which, given an input string $y \in \{0, 1, *\}^{2^n}$ with fewer than 2^{n-d} erasure symbols, determines whether or not there exists a codeword $c \in \mathsf{RM}(n, d)$ such that $y_i \in \{c_i, *\}$ for all $i \in [2^n]$. The algorithm returns such a codeword c if it exists and an error message otherwise.*

Proof. Our decoder is given in Algorithm 1. We first prove by induction on n that it always finds the desired codeword $c \in \mathsf{RM}(n, d)$ if such a codeword exists. Note that the base case $n = 1$ holds trivially, since Algorithm 1 always succeeds when $d = 0$ or $d = n$.

For the inductive case, suppose there exists a codeword $c \in \mathsf{RM}(n, d)$ such that y agrees with c on all non-erased entries. Let $f(x_1, \ldots, x_n)$ be the unique polynomial whose evaluation

Algorithm 1: Codeword testing and erasure correction for Reed-Muller codes

Input: Two integers $0 \leq d \leq n$ and a vector $y \in \{0,1,*\}^{2^n}$ with fewer than 2^{n-d} erasure entries.

Output: A codeword $c \in \mathsf{RM}(n,d)$ with $y_i \in \{c_i, *\}$ for all $i \in [2^n]$, if such a c exists; an error message otherwise.

if $d = 0$ **then**
  If there exists $b \in \{0,1\}$ such that $y_i \in \{b,*\}$ for all i, output $(b,b,\ldots,b)$. Otherwise, output an error message.
end
else if $d = n$ **then**
  Output y.
end
else
  Define $y^0 := (y_1,\ldots,y_{2^{n-1}})$ and $y^1 := (y_{2^{n-1}+1},\ldots,y_{2^n})$ and let $y^{\mathrm{sum}} := y^0 + y^1$ (defining $y_i^{\mathrm{sum}} = *$ whenever either $y_i^0 = *$ or $y_i^1 = *$). Run Algorithm 1 on input $(n-1, d-1, y^{\mathrm{sum}})$ and denote the output you receive by c^{sum}. If c^{sum} is an error message, abort and output an error.
  if y^0 *contains fewer erasure symbols than* y^1 **then**
    Run Algorithm 1 on input $(n-1, d, y^0)$, denoting the output you receive by c^0. If c^0 is an error message, abort and output an error. Otherwise, define c to be the concatenation $c := (c^0, c^0 + c^{\mathrm{sum}})$. If $y_i \in \{c_i, *\}$ for all $i \in [2^n]$, output c; otherwise, output an error.
  end
  else
    Run Algorithm 1 on input $(n-1, d, y^1)$, denoting the output you receive by c^1. If c^1 is an error message, abort and output an error. Otherwise, define c to be the concatenation $c := (c^1 + c^{\mathrm{sum}}, c^1)$. If $y_i \in \{c_i, *\}$ for all $i \in [2^n]$, output c; otherwise, output an error.
  end
end

vector is c and express f as

$$f(x_1, \ldots, x_n) = f_0(x_2, \ldots, x_n) + x_1 \cdot f_1(x_2, \ldots, x_n). \tag{8.1}$$

We make the following two observations:

1. Define $c^0 \in \{0,1\}^{2^{n-1}}$ to be the vector containing the first half of c's entries. Then c^0 is the evaluation vector of the polynomial $f_0(x_2, \ldots, x_n)$.

2. Define $c^1 \in \{0,1\}^{2^{n-1}}$ to be the vector containing the second half of c's entries. Then c^1 is the evaluation vector of $f_0(x_2, \ldots, x_n) + f_1(x_2, \ldots, x_n)$.

Both (1) and (2) follow immediately from the fact that the indices of c are ordered lexicographically (and thus an index $v \in \mathbb{F}_2^n$ is in the first half if and only if $v_1 = 0$).

Now, since $y^0 := (y_1, \ldots, y_{2^n})$ is a corrupted evaluation vector for the polynomial f_0 and $y^1 := (y_{2^n+1}, \ldots, y_{2^n})$ is a corrupted evaluation vector for the polynomial $f_0 + f_1$, the vector $y^{\text{sum}} := y^0 + y^1$ must be an evaluation vector for f_1 corrupted with fewer than 2^{n-r} erasures. By induction, since f_1 has degree $\leq f - 1$, running Algorithm 1 on input $(n-1, d-1, y^{\text{sum}})$ will then return the correct evaluation vector c^{sum} for the polynomial $f_1(x_2, \ldots, x_n)$. (See line 8.)

Furthermore, since y contains fewer than 2^{n-d} erasures, one of y^0 or y^1 must contain fewer than 2^{n-d-1} erasures. Without loss of generality, we assume that y^0 contains fewer erasures[2]. By induction, running Algorithm 1 on input $(n, d-1, y^0)$ will then return the correct evaluation vector c^0 for the polynomial f_0. (See line 10.) Since c^1 is the evaluation vector for $f_0 + f_1$ and we have obtained the evaluation vectors c^0, c^{sum} for f_0 and f_1, setting $c^1 = c^0 + c^{\text{sum}}$ will successfully recover c^1. Thus the algorithm indeed outputs the correct codeword $c = (c_0, c_1)$. (See line 10.)

We have proven that whenever there exists a (by our theorem's requirement, unique) codeword $c \in \mathsf{RM}(n, d)$ that is consistent with y, our algorithm returns it. Note also that

[2]The proof is identical in the other case, with the roles of y^0 and y^1 reversed.

Algorithm 1 never outputs a codeword $c \in \mathsf{RM}(\mathsf{n}, \mathsf{d})$ that is not consistent with y; this is because before returning c, the algorithm verifies that $y_i \in \{c_i, *\}$ for all $i \in [2^n]$ (see lines 10 and 13). Thus Algorithm 1 always succeeds. As for the runtime analysis, we note that at each step, the algorithm spends $O(2^n)$ time performing basic computations and then makes 2 recursive calls on instances of length 2^{n-1}. By the Master theorem (Lemma 11), the total runtime will thus be $O(n2^n)$.

□

Our second lemma is essentially a special case of the work of [66] (Theorem 10), which we will use in the following form to bound the running time and error probability of a decoder for high-rate Reed-Muller codes.

Lemma 70. *Consider any integers $n > t > 0$ and any $p \leq \frac{2^{-n+\frac{t}{2}}}{5}$. Then there exists a decoder $\tilde{D}$ for the Reed-Muller code $\mathsf{RM}(n, n-t)$ with the following two properties.*

1. *$\tilde{D}$ runs in time $O\Big(2^n \cdot \mathrm{poly}(\binom{n}{\leq \frac{t}{2}})\Big)$.*

2. *Under random errors of probability p, $\tilde{D}$ succeeds in decoding any sent codeword $c \in \mathsf{RM}(n-t, n)$ with probability*

$$\Pr_{z \sim p}\left[\tilde{D}(c+z) = c\right] \geq 1 - 2^{-2^{t/2}}.$$

Proof. By Theorem 10, it will suffice to show that the Reed-Muller code $\mathsf{RM}(n, n-\frac{t}{2})$ can recover from p-noisy erasures with success probability $1-2^{-2^{t/2}}$. Since the code $\mathsf{RM}(n, n-\frac{t}{2})$ has minimum distance $2^{t/2}$, it is enough to prove that for independent Bernoulli variables $X_1, X_2, \ldots, X_{2^n}$ with Bernoulli parameter $p \leq \frac{2^{-n+\frac{t}{2}}}{5}$, we have

$$\Pr\left[X_1 + X_2 + \ldots X_{2^n} \geq 2^{t/2}\right] \leq 2^{-2^{t/2}}.$$

But this follows immediately from the Chernoff bound (Lemma 13). □

8.3 Decoding Arbitrary Tensor Codes from Adversarial Errors

In this section, we will give an algorithm for decoding any tensor product code from adversarial errors. For any code $C \subseteq \{0,1\}^N$, define the function $f_C : \{0,1,*\}^N \to \{0,1,*\}^N$ to be

$$f_C(x) := \begin{cases} c & \text{if } c \in C,\ x_i \in \{c_i, *\} \text{ for all } i, \text{ and } x \text{ has } < d_{\min}(C) \text{ erasures} \\ (*, \ldots, *) & \text{otherwise.} \end{cases} \tag{8.2}$$

The function f_C essentially tells us whether or not a partially-erased binary string with fewer than $d_{\min}(C)$ erasures is consistent with any codeword of C. It can always be computed in time $O(N^3)$ (see Note 1), and its runtime dictates the runtime of our following decoder for adversarial errors.

Theorem 71. *Consider any linear codes $C_1 \subseteq \{0,1\}^{N_1}, \ldots, C_t \subseteq \{0,1\}^{N_t}$ and define $N := \prod_{i=1}^t N_i$. Suppose there exists a function $T : \mathbb{N} \to \mathbb{N}$ such that $T(n) \geq n$ for all $n \in \mathbb{N}$ and such that for all $i \in [t]$, there is a $T(N_i)$-time algorithm for computing the function f_{C_i} defined in (8.2). Then there is an $O\left(\sum_{i=1}^t \frac{N}{N_i} \cdot T(N_i)\right)$-time algorithm for decoding the tensor code $C := C_1 \otimes \ldots \otimes C_t$ from*

$$\left\lceil \frac{d_{\min}(C)}{2 \max\left\{d_{\min}(C_1), \ldots, d_{\min}(C_t)\right\}} \right\rceil - 1$$

adversarial errors.

Remark. *By Note 1, for any linear codes $C_1, \ldots, C_t$ we can take $T(N) = O(N^3)$, which gives us a runtime of $O\left(N \sum_i N_i^2\right)$. But for Reed-Muller codes, we can do better: by Lemma 69, we can take $T(N) = O(N \log N)$, which gives a runtime of $O\big(\sum_i N \log N_i\big) = O(N \log N)$.*

Remark. *Note that in Algorithm 2, at each layer $i = 2, \ldots, t$ of the decoding process, the i-axis erasure pattern is the same for all i-axis vectors within the same i-subtensor of A. One could thus use Gaussian elimination to express the erased entries in this erasure pattern as a linear combination of the non-erased entries, then go through each i-axis vector of A*

and correct the erasures accordingly. If C_1 is taken to be the code of maximum length among $C_1, \ldots, C_t$, this will give a running time of $O(N \sum_i N_i)$ for decoding the tensor product of any linear codes $C_1, \ldots, C_t$ of lengths $N_1, \ldots, N_t$.

Proof. Note that we may assume that each N_i is greater than 1, as otherwise C trivially reduces to a tensor product of $t-1$ codes. Our algorithm for the case where each N_i is greater than 1 is given in Algorithm 2. We first show by induction that it always outputs either a codeword of $C_1 \otimes \ldots \otimes C_t$ or the all-erasures tensor. The base case $t = 1$ is trivial. For the case $t > 1$, we note that by induction, after the line-5 for loop completes, each $(t-1)$-subtensor of A is either a valid codeword of $C_1 \otimes \ldots \otimes C_t$ or a tensor filled with erasure symbols. Thus the erasure pattern of each t-axis vector in the line 8-for loop is identical. In particular, if there is a unique consistent codeword for each of these t-axis vectors, then the erased entries of each t-axis vector can be expressed as the same linear combination of non-erased coordinates. This means that the erased $(t-1)$-subtensors can be expressed as linear combinations of the non-erased $(t-1)$-subtensors; since the code is linear, the newly recovered subtensors must then be codewords of $C_1 \otimes \ldots \otimes C_{t-1}$. Combining this with the fact that by line 9, each t-axis vector is a codeword of C_t (otherwise by line 12, we would output an all-erasures tensor), we get that any Boolean output must indeed be a codeword of $C_1 \otimes \ldots \otimes C_t$.

Now that we have proven that our algorithm always outputs either a valid codeword or a tensor filled with erasures, we turn to showing that the algorithm correctly ouputs the sent codeword as long as there are fewer than $\frac{d_{\min}(C)}{2 \max\left\{d_{\min}(C_1), \ldots, d_{\min}(C_t)\right\}}$ errors. We again proceed by induction. The base case $t = 1$ is trivial. For the general case, we note that in order for our Algorithm 2 to fail in decoding a noisy codeword containing fewer than $\frac{d_{\min}(C_1) \cdot \ldots \cdot d_{\min}(C_t)}{2}$ errors, one of the following two statements must hold:

1. After we decode all the $(t-1)$-dimensional subtensors $\{A^i\}_{i \in [N_t]}$ (see line 6), there is a non-erasure erroneous entry in at least one of the updated subtensors A^i.

2. After we decode all the $(t-1)$-dimensional subtensors $\{A^i\}_{i \in [N_t]}$, there are at least

Algorithm 2: A polynomial-time decoder for arbitrary tensor codes

Input: An $N_1 \times \ldots \times N_t$ Boolean tensor A.

Output: Either a codeword of the tensor code $C_1 \otimes \ldots \otimes C_t$ or an $N_1 \times \ldots \times N_t$ tensor filled with erasure symbols.

if $t = 1$ **then**

  If A has fewer than $d_{\min}(C_1)$ erasures and there is a codeword $c \in C_1$ with $A_i \in \{c_i, *\}$ for all $i \in [N_1]$, replace A by c. Otherwise replace every entry in A by an erasure symbol.

end

else

  for $i = 1, 2, \ldots, N_t$ **do**

    Run Algorithm 2 on the $N_1 \times \ldots \times N_{t-1}$ tensor A^i whose entries are given by $A^i_{j_1 \ldots j_{t-1}} = A_{j_1 \ldots j_{t-1} i}$. Replace the entries $\{A_{j_1 \ldots j_{t-1} i}\}$ of A by the output entries.

  end

  for *every* $t-$*axis vector* $v \in \{0, 1, *\}^{N_t}$ *of* A *(see definition 5)* **do**

    If v has fewer than $d_{\min}(C_t)$ erasures and there is a codeword $c \in C_t$ with $v_i \in \{c_i, *\}$ for all i, replace v by c. Otherwise, replace v by $(*, *, \ldots, *)$.

  end

  if *the updated tensor* A *contains erasure symbols or its Hamming distance from the original input is at least* $\frac{d_{\min}(C_1) \cdot \ldots \cdot d_{\min}(C_t)}{2}$ **then**

    Replace every entry of A by an erasure symbol.

  end

end

Output the updated tensor A.

$d_{\min}(C_t)$ values of $i \in [N_t]$ for which the updated subtensor A^i contains one or more erasures.

Indeed, if neither of these occur, then our line 9 will allow us to recover every entry of A correctly. We now show that neither point (1) nor point (2) can occur. Suppose for contradiction that there exists a subtensor A^i as described in point (1). Note that by line 12, the Hamming distance between A^i and its corresponding input must be less than $\frac{d_{\min}(C_1)\cdot\ldots\cdot d_{\min}(C_{t-1})}{2}$. Since A^i is a codeword of $C_1 \otimes \ldots \otimes C_{t-1}$ (we proved in the first paragraph that any output of our algorithm is a codeword) and since $C_1 \otimes \ldots \otimes C_{t-1}$ has minimum distance $d_{\min}(C_1) \cdot \ldots \cdot d_{\min}(C_{t-1})$, there must have been at least $\frac{d_{\min}(C_1)\cdot\ldots\cdot d_{\min}(C_{t-1})}{2}$ corrupted entries in the i^{th} subtensor to begin with. This contradicts our theorem's requirement on the total number of errors.

Suppose instead that there are $d_{\min}(C_t)$ subtensors $\{A^{i_1}, \ldots, A^{i_{d_{\min}(C_t)}}\}$ satisfying point (2) above. Since each of these subtensors is decoded independently, by our inductive hypothesis there must have been at least $d_{\min}(C_t) \cdot \frac{d_{\min}(C_1)\cdot\ldots\cdot d_{\min}(C_{t-1})}{2\max\left\{d_{\min}(C_1),\ldots,d_{\min}(C_{t-1})\right\}} \geq \frac{d_{\min}(C)}{2\max\left\{d_{\min}(C_1),\ldots,d_{\min}(C_t)\right\}}$ errors. But this again contradicts our theorem's requirement on the total number of errors.

This concludes the proof of correctness. We now turn to the runtime analysis. Define $R(N_1, \ldots, N_t)$ to be the maximal runtime of Algorithm 2 on any code $C' = C'_1 \otimes \ldots \otimes C'_t$ with $C'_i \subseteq \{0,1\}^{N_i}$ for all i. Note that for $t > 1$, we have

$$R(N_1, \ldots, N_t) \leq N_t \cdot R(N_1, \ldots, N_{t-1}) + \frac{N}{N_t} T(N_t) + \alpha N,$$

where the first and second terms correspond to the bulk of the runtime of Algorithm 2 for the for-loops 5 and 8 respectively, whereas the constant α is chosen to be big enough that the αN-term covers all the other operations needed throughout the algorithm. We claim that

$$R(N_1, \ldots, N_t) \leq (\alpha + 1) \sum_{i=1}^{t} \frac{N}{N_i} \cdot T(N_i).$$

For $t = 1$, the statement is obvious. For $t > 1$, by induction we have

$$\begin{aligned} R(N_1, \ldots, N_t) &\leq (\alpha + 1) N_t \sum_{i=1}^{t-1} \frac{N/N_t}{N_i} T(N_i) + \frac{N}{N_t} T(N_t) + \alpha N \\ &\leq (\alpha + 1) \sum_{i=1}^{t} \frac{N}{N_i} T(N_i), \end{aligned}$$

where in the last line we used the fact that by our theorem's requirement on T, we have $\alpha N = \frac{\alpha N}{N_t} N_t \leq \frac{\alpha N}{N_t} T(N_t)$.

□

8.4 Decoding Tensor Reed-Muller Codes from Random Errors

In this section, we leverage our Algorithm 2 to prove the following formal version of Theorem 38. Note that it is sufficient to prove Theorem 38 for $t \leq \sqrt{\log N}$, since after that the $O\left(\frac{1}{t}\right)$ improvement in the error probability is subsumed by the $o(1)$ term (one can always artificially increase t by taking tensor products with the trivial Reed-Muller code $\{0, 1\}$).

Theorem 72. *Consider any constants $p \in (0, \frac{1}{2})$ and $R < 1 - h(p)$. Let $N \in \mathbb{N}$ be some growing parameter and consider any corresponding integer $3 \leq t \leq \sqrt{\log N}$. Define*

- $n_1 := \lceil \log \log N - 3 \rceil$ *and d_1 to be any integer such that* $\frac{\binom{n_1}{\leq d_1}}{2^{n_1}} = R \pm o(1)$
- $n_2 := \lceil 10 \log \log N \rceil$ *and* $d_2 := \lceil \frac{n_2}{2} + \sqrt{n_2} \log n_2 \rceil$
- $n_3 = \ldots = n_t := \lceil \frac{\log N - n_1 - n_2}{t-2} \rceil$ *and* $d_3 = \ldots = d_t := \lceil \frac{n_3 + n_3^{3/4}}{2} \rceil$

Then the Tensor Reed-Muller code $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ *has length $N^{1+o(1)}$ and rate $R \pm o(1)$. Moreover, there exists a decoder D with the following properties:*

1. *D has runtime $O(N \log \log N)$ in the case $t = 3$ and runtime $O(N \log N)$ in the case $t > 3$.*

2. For every codeword $c \in \mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$, D has decoding failure probability

$$\Pr_{z\sim p}\Big[D(c+z) \neq c\Big] \leq \begin{cases} N^{-\omega(\log N)} & \text{if } t = 3, \\ 2^{-N^{\frac{1}{2}-\frac{1}{2(t-2)}-o(1)}} & \text{otherwise.} \end{cases}$$

Proof. Note that $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ has rate

$$\begin{aligned} \prod_{i=1}^{t} \frac{\binom{n_i}{\leq d_i}}{2^{n_i}} &\geq \big(R - o(1)\big) \prod_{i=3}^{t} \left(1 - \frac{2^{h(\frac{n_i-d_i}{n_i})n_i}}{2^{n_i}}\right) \\ &\geq (R - o(1)) \prod_{i=3}^{t} \left(1 - 2^{-\frac{\sqrt{n_3}}{2\ln 2}}\right) \\ &\geq (R - o(1))(1 - t2^{-\frac{\sqrt{n_3}}{2\ln 2}}), \end{aligned}$$

where the first inequality follows from the fact that $\binom{n}{\leq d} \leq 2^{h(\frac{d}{n})n}$ for all integers n and $d \leq \frac{n}{2}$, the second inequality follows from the fact that $h\left(\frac{1-\mu}{2}\right) \leq 1 - \frac{\mu^2}{2\ln 2}$ for any $\mu \in (0, 1)$, and the third inequality follows from the fact that $(1+x)^r \geq 1 + rx$ for all $x \geq -1$ and $r \geq 1$. Since $t \leq \sqrt{\log N}$ and $n_3 = \Omega(\frac{\log N}{t})$, our code $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ indeed has rate $R \pm o(1)$. It also has length $2^{n_1+n_2} \cdot 2^{(t-2)\lceil \frac{\log N - n_1 - n_2}{t-2} \rceil} = N^{1+o(1)}$.

The decoder D we use for decoding our code is given in Algorithm 3. We represent each codeword of $\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)$ as a $2^{n_1} \times \ldots \times 2^{n_t}$ Boolean tensor. For any tensor $A \in \{0,1\}^{2^{n_1} \times \ldots \times 2^{n_t}}$, we call any vector along the first axis of A a "row" of A and call any vector along the second axis of A a "column" of A. We first bound the probability that Algorithm 3 makes a decoding mistake. Note that the only way a decoding mistake can occur is if either:

1. Our algorithm would fail even if the counter condition (last sentence of line 10) was disregarded.

2. The counter eventually exceeds $N2^{-2^{(\log\log N)^{1/4}}}$.

Algorithm 3: An efficient decoder for t-tensor Reed-Muller codes

Input: A $2^{n_1} \times \ldots \times 2^{n_t}$ Boolean tensor A.

Output: A $2^{n_1} \times \ldots \times 2^{n_t}$ Boolean tensor.

1 Create a look-up table $D_1 : \{0,1\}^{2^{n_1}} \to \mathsf{RM}(n_1, d_1)$.

2 **for** *each vector* $s \in \{0,1\}^{2^{n_1}}$ **do**

3 Use brute-force search to find the $c \in \mathsf{RM}(n_1, d_1)$ closest to s. Set $D_1[s] = c$.

4 **end**

5 **for** *each row* u *of* A **do**

6 Replace u by $D_1[u]$.

7 **end**

8 $\mathsf{counter} \leftarrow 0$.

9 **for** *each column* v *of our updated tensor* A **do**

10 Use Lemma 69 to check if $v \in \mathsf{RM}(n_2, d_2)$. If not, increase $\mathsf{counter}$ by 1 and replace the column v by the codeword $D_2(v) \in \mathsf{RM}(n_2, d_2)$, where D_2 is the decoder from Lemma 70 for the code $\mathsf{RM}(n_2, d_2)$. If $\mathsf{counter} > N2^{-2^{(\log\log N)^{1/4}}}$, abort the entire algorithm and return the $\vec{0}$ codeword.

11 **end**

12 If $t = 3$, output the updated tensor A. If $t > 3$, run Algorithm 2 on A and return the output (if Algorithm 2 outputs a tensor filled with erasure symbols, return the $\vec{0}$ codeword).

We first show that point (1) is very unlikely to occur. Note that by Theorem 27, the lookup table D_1 used in the row-for loop of our Algorithm 3 (line 6) satisfies that for any $c_1 \in \mathsf{RM}(n_1, d_1)$,

$$\begin{aligned} \Pr_{z \sim p}\Big[D_1[c_1 + z] \neq c_1\Big] &\leq 2^{-2^{\Omega(\sqrt{n_1})}} \\ &\leq 2^{-2^{\Omega(\sqrt{\log\log N})}}. \end{aligned} \tag{8.3}$$

Thus at the end of the row-for loop, every entry of A has probability $\varepsilon \leq 2^{-2^{\Omega(\sqrt{\log\log N})}}$ of

being different from the corresponding entry of the sent codeword B. Furthermore, since D_1 was applied independently to every row of A, any coordinates of A that are not in the same row have uncorrelated probabilities of being correct.

In particular, at the end of the row-for loop, the entries of any given column of A have uncorrelated probabilities of being incorrect. Now, since $\varepsilon \leq \frac{2^{-n_2}}{5}$ for all n large enough, we get from Lemma 70 that for any $c_2 \in \mathsf{RM}(n_2, d_2)$, the decoder D_2 used in the column-for loop (line 10) of our Algorithm 3 satisfies

$$\begin{aligned}\Pr_{z\sim\varepsilon}\Big[D_2(c_2+z) \neq c_2\Big] &\leq O\left(2^{-2^{\frac{n_2-d_2}{2}}}\right)\\ &= 2^{-2^{2.5\log\log N \pm o(\log\log N)}}\\ &\leq 2^{-\omega\left(\log^2 N\right)} \end{aligned} \tag{8.4}$$

Thus at the end of our column-for loop, if we disregard the $\mathsf{counter}$ condition when running the algorithm, then every entry of A has probability

$$\varepsilon' \leq N^{-\omega(\log N)} \tag{8.5}$$

of being incorrect. Taking a union bound over all coordinates then concludes the analysis of point (1) for the case $t = 3$. For the case $t > 3$, we define for every $k \in [2^{n_3}] \times \ldots \times [2^{n_t}]$ the Boolean random variable X_k that is 1 if and only if upon running Algorithm 3 without the $\mathsf{counter}$ condition, at the end of the column-for loop (line 11), there exists $(i, j) \in [2^{n_1}] \times [2^{n_2}]$ such that the entry A_{ijk} is incorrect. By (8.5), since D_1 was applied independently to every row and D_2 was applied independently to every column, the random variables $\{X_k\}$ are independent Bernoulli random variables with probability parameter at most $2^{n_1+n_2}\varepsilon' = N^{-\omega(\log N)}$. By the Chernoff bound (Lemma 13), we then have

$$\begin{aligned}\Pr\left[\sum_k X_k \geq \frac{d_{\min}\big(\mathsf{TRM}(n_1, d_1; \ldots; n_t, d_t)\big)}{2^{n_1+n_2+1} d_{\min}\big(\mathsf{RM}(n_3, d_3)\big)}\right] &\leq 2^{-\frac{d_{\min}\left(\mathsf{TRM}(n_1,d_1;\ldots;n_t,d_t)\right)}{d_{\min}\left(\mathsf{RM}(n_3,d_3)\right)}\cdot n^{-o(1)}}\\ &\leq 2^{-N^{\frac{t-3}{2(t-2)}-o(1)}}.\end{aligned}$$

Thus if we disregard the counter condition when running the algorithm, we get that with probability $1 - 2^{-N^{\frac{t-3}{2(t-2)}-o(1)}}$, there are $\leq \frac{d_{\min}\left(\mathsf{TRM}(n_1,d_1;\ldots;n_t,d_t)\right)}{2d_{\min}\left(\mathsf{RM}(n_3,d_3)\right)}$ errors remaining after the line-9 for loop. By Theorem 71, line 12 will then succeed with probability at least $1 - 2^{-N^{\frac{1}{2}-\frac{1}{2(t-2)}-o(1)}}$. This concludes our analysis for point (1). We then turn to showing that point (2) is very unlikely to occur. Note that by (8.3), at the end of the row-for loop (line 7) of our algorithm, for any $k \in [2^{n_3}] \times \ldots \times [2^{n_t}]$ we have

$$\begin{aligned}\Pr\left[\exists (i,j) \in [2^{n_1}] \times [2^{n_2}] \text{ such that } A_{i,j,k} \text{ is incorrect}\right] &\leq 2^{n_2} \cdot 2^{-2^{\Omega(\sqrt{\log\log N})}} \\ &\leq 2^{-2^{\Omega(\sqrt{\log\log N})}}.\end{aligned}$$

Since Algorithm 3 processes each $k \in [2^{n_3}] \times \ldots \times [2^{n_t}]$ independently up to the end of the column-for loop (line 11), we get

$$\Pr\left[\mathsf{counter} \text{ exceeds } N2^{-2^{(\log\log N)^{1/4}}}\right] \leq \Pr\left[\sum_{k=1}^{N2^{-n_1-n_2}} 2^{n_1} Y_k \geq N2^{-2^{(\log\log N)^{1/4}}}\right]$$

for $\{Y_k\}$ independent Bernoulli variables of probability $2^{-2^{\Omega(\sqrt{\log\log N})}}$. By the Chernoff bound (Lemma 13), we then have

$$\Pr\left[\mathsf{counter} \text{ exceeds } N2^{-2^{(\log\log N)^{1/4}}}\right] \leq 2^{-\Omega\left(N2^{-2^{(\log\log N)^{1/4}}}/2^{n_1}\right)} \leq O\left(2^{-\sqrt{N}}\right).$$

Combining this bound for point (2) with our previously established bound for point (1), we get

$$\Pr_{z\sim p}\left[D(c+z) \neq c\right] \leq N^{-\omega(\log N)} + O\left(2^{-\sqrt{N}}\right)$$

for the case $t = 3$ and

$$\Pr_{z\sim p}\left[D(c+z) \neq c\right] \leq 2^{-N^{\frac{1}{2}-\frac{1}{2(t-2)}-o(1)}} + O\left(2^{-\sqrt{N}}\right)$$

for the case $t > 3$, as desired. We now turn to bounding our algorithm's runtime. Since there are $2^{2^{n_1}}$ vectors in $\{0,1\}^{2^{n_1}}$, creating the look-up table D_1 takes time

$$2^{2^{n_1}} \cdot O(2^{2^{n_1}} 2^{n_1}) = o(N). \tag{8.6}$$

Since there are $\frac{N}{2^{n_1}}$ rows in the tensor A and since looking up a value in the table D_1 takes time $O(2^{n_1})$, the row-for loop (line 5) then takes time

$$\frac{N}{2^{n_1}} \cdot O\left(2^{n_1}\right) = O\left(N\right). \tag{8.7}$$

For the column-for loop (line 9), since there are $\frac{N}{2^{n_2}}$ columns in the tensor A, by Lemmas 69 and 70 the algorithm takes time

$$\frac{N}{2^{n_2}} \cdot O(n_2 2^{n_2}) + N2^{-2^{(\log\log N)^{1/4}}} \cdot 2^{O(n_2)} = O\left(N \log\log N\right). \tag{8.8}$$

Combining equations (8.6), (8.7) and (8.8), we get that our decoder D has total runtime $O\left(N \log\log N\right)$ when $t = 3$. When $t > 3$, the algorithm additionally has to process line 12, which takes time

$$O\left(N \log N\right) \tag{8.9}$$

by Theorem 71 and Lemma 69. This brings the total runtime for the case $t > 3$ to $O\left(N \log N\right)$.

□

Appendix A

CHANNEL LOWER BOUNDS

In this appendix, we will prove some fundamental lower bounds on the capacity of communication channels.

A.1 Proof of Erasure Channel Capacity

We restate and prove Theorem 2.

Theorem 2. *The capacity of the q-ary erasure channel qEC_p is $1-p$.*

Proof. We will first show that a uniformly random codeword c of a uniformly random code $C \subseteq \mathbb{F}_q^N$ of size $|C| = q^{(1-p)N-\sqrt{N}}$ is resilient to p-noisy erasures with high probability. This would prove our claim, since we can then take C' to be the code C minus all the codewords of C that are not resilient to p-noisy erasures. Thus take a uniformly random code $C \subseteq \mathbb{F}_q^N$ of size $q^{(1-p)N-\sqrt{N}}$ and consider a uniformly random codeword $c \in C$. Without loss of generality, we may assume that $c = \vec{0}$. Now note that with probability $1 - \frac{1}{N}$, for each $i = 1, 2, \ldots, N$ there are at most $N^2 \cdot \frac{\binom{N}{i}(q-1)^i \cdot q^{(1-p)N-\sqrt{N}}}{q^N}$ codewords $c \in C$ of weight $\mathrm{wt}(c) = i$. Also, with high probability the erasure pattern has weight smaller than pN. Thus by taking a union bound over every codeword in C, we get that the probability of a decoding mistake when sending the codeword $\vec{0}$ can be bounded by

$$\begin{aligned}
\Pr[\text{decoding mistake}] &\leq o(1) + \sum_{i=1}^{pN} N^2 \cdot \frac{\binom{N}{i}(q-1)^i \cdot q^{(1-p)N-\sqrt{N}}}{q^N} \cdot \frac{\binom{N-i}{pN-i}}{\binom{N}{pN}} \\
&= o(1) + N^2 q^{-pN-\sqrt{N}} \sum_{i=1}^{pN} \binom{pN}{i} (q-1)^i \\
&\leq o(1) + N^2 q^{-\sqrt{N}},
\end{aligned}$$

where in the last line we used the fact that $\sum_{i=0}^{m} \binom{m}{i} a^i = (a+1)^m$. For the second part of the statement, we want to show that for any constant $\varepsilon > 0$, no code $C \subseteq \mathbb{F}_q^N$ of size $q^{(1-p+\varepsilon)N}$ is resilient to p-noisy erasures. But note that with high probability, there will be at least $(p-\varepsilon/2)N$ erasures. Whenever this happens, there are $\leq q^{(1-p+\varepsilon/2)N}$ strings consistent with the erasure pattern z. Thus there can at most be $q^{(1-p+\varepsilon/2)N} - 1$ codewords that are uniquely-decodable from erasure pattern z, which means that a uniformly random codeword $c \in C$ has probability $\leq \frac{q^{(1-p+\varepsilon/2)N}}{q^{(1-p+\varepsilon)N}} \leq q^{-\frac{\varepsilon N}{2}}$ of being uniquely decodable from z. □

A.2 Lower Bounds for List Decoding

In this section, we prove the result mentioned in equation (3.1), Section 3.2.

Claim 73. *Let $p \in (0, \frac{1}{2})$ be arbitrary, and consider any $N > \frac{100}{p^2}$. Suppose a code $C \subseteq \mathbb{F}_2^N$ and a decoder $d_k : \mathbb{F}_2^N \to C^{\otimes k}$ satisfy*

$$\Pr_{\substack{z \sim p \\ c \in C}} [c \in d_k(c+z)] \geq \frac{3}{4}.$$

Then we must have

$$k \geq |C| \cdot 2^{-(1-h(p))N} \cdot \frac{2^{-h(p)N^{3/4}}}{8}.$$

Proof. We will first show that in order for the decoder d_k to succeed with high probability, there must be many codewords $c \in C$ for which

$$|\{x \in \mathbb{F}_2^N : c \in d_k(x)\}| \gtrsim 2^{h(p)N}.$$

Intuitively, this is because the sphere of radius pN around any codeword c contains $\approx 2^{h(p)N}$ points (and for any transmitted codeword c, with high probability the received message m will satisfy $\mathrm{wt}(m+c) \approx pN$). We will then simply double-count the number of pairs (x, c) for which $c \in d_k(x)$. On the one hand, there are $2^N \cdot k$ such pairs, since every received message is mapped to k codewords; on the other hand, there must be at least about $|C| \cdot 2^{h(p)N}$ pairs, since most codewords in C need to be matched to at least $\approx 2^{h(p)N}$ points. It follows that we must have

$$k \gtrsim |C| \cdot \frac{2^{h(p)N}}{2^N}.$$

Formally, we first note that the theorem condition implies that at least $\frac{|C|}{2}$ codewords $c \in C$ must satisfy

$$\Pr_{z\sim p}[c \in d_k(c+z)] \geq \frac{1}{2}. \tag{A.1}$$

Fix any such c. Now from Chernoff's bound (i.e Lemma 13), we have for $N > \frac{100}{p^2}$ that

$$\begin{aligned}\Pr_{z\sim p}\left[\mathrm{wt}(z) \leq pN - pN^{3/4}\right] &\leq 2e^{-\frac{10}{3}} \\ &\leq \frac{1}{4}.\end{aligned}$$

In order for c to satisfy $c \in d_k(c+z)$ with probability at least $\frac{1}{2}$, there must then be a subset $S_c \subseteq \{x \in \mathbb{F}_2^N : \mathrm{wt}(c+x) \geq pN - pN^{3/4}\}$ satisfying both

$$x \in S_c \implies c \in d_k(x) \tag{A.2}$$

and

$$\Pr_{z\sim p}\left[z \in S_c\right] \geq \frac{1}{4}. \tag{A.3}$$

But every element $x \in S_c$ satisfies $\mathrm{wt}(c+x) \geq pN - pN^{3/4}$, so every $x \in S_c$ satisfies

$$\begin{aligned}\Pr_{z\sim p}\left[z = c+x\right] &\leq p^{pN-pN^{3/4}}(1-p)^{(1-p)N+pN^{3/4}} \\ &\leq 2^{-(1-N^{-1/4})h(p)N}\end{aligned} \tag{A.4}$$

Equations (A.3) and (A.4) imply that any $c \in C$ that can be list-decoded by d_k with probability $\geq \frac{1}{2}$ must satisfy $|S_c| \geq \frac{2^{(1-N^{-1/4})h(p)N}}{4}$. It then follows from (A.2) that any such c must satisfy

$$\left|\{x \in \mathbb{F}_2^N : c \in d_k(x)\}\right| \geq \frac{2^{(1-N^{-1/4})h(p)N}}{4}.$$

By double counting, we get

$$\begin{aligned}2^N \cdot k &= \sum_{c\in C}\left|\{x \in \mathbb{F}_2^N : c \in d_k(x)\}\right| \\ &\geq \frac{|C|}{2} \cdot \frac{2^{(1-N^{-1/4})h(p)N}}{4} \\ &= \frac{|C|}{8} \cdot 2^{h(p)N - h(p)N^{3/4}}.\end{aligned}$$

The result then follows from rearranging terms. □

Appendix B

MINOR RESULTS

In this appendix, we will prove some minor results we stated but did not prove in the main text.

B.1 Proof of Corollary 67

Recall that for any $p \in (0,1)$, we defined

$$A_p := \{\alpha N \in \mathbb{N} : h(\alpha) > 1 - h(p) - N^{-1/5}\}.$$

We now restate and prove our Corollary 67.

Corollary 67. *Let $p \in (0, \frac{1}{2})$ be arbitrary and let $C \subseteq \mathbb{F}_2^N$ be any linear code. Suppose that for every $j \in A_p$, we have*

$$\Pr_{y \sim \mathcal{D}(C^\perp)} \big[\mathrm{wt}(y) = j\big] \leq \big(1 + o(N^{-1})\big) \frac{\binom{N}{j}}{2^N}.$$

Suppose also that

$$\Pr_{y \sim \mathcal{D}(C^\perp)} \big[\mathrm{wt}(y) \notin A_p\big] \leq 2^{N^{\frac{3}{4}}} \cdot \frac{\sum_{i \notin A_p} \binom{N}{i}}{2^N}.$$

Then C is resilient to p-noisy errors.

Proof. Applying Lemma 9 and Theorem 66 with $k = 1$ and $l = N^{3/4}$, we get that whenever $N > \frac{1}{p^4(\frac{1}{2}-p)^4}$, there exists some decoder $d : \mathbb{F}_2^N \to C$ such that for all $c \in C$,

$$\Pr_{z \sim p}[d(c+z) \neq c] \leq 2e^{-\frac{\sqrt{N}}{3p}} + N \max_{\substack{S \subseteq [pN \pm N^{3/4}] \\ 1 \leq |S| \leq 2}} \Big\{ \frac{1}{\binom{N}{S}} \sum_{j=0}^{N} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 - 1 \Big\}. \quad \text{(B.1)}$$

Let $\nu \in (0, \frac{1}{2})$ be such that $h(\nu) = 1 - h(p) - N^{-1/5}$, and note that we have

$$A_p = \{\lceil \nu N \rceil, \lceil \nu N \rceil + 1, \ldots, \lfloor (1-\nu)N \rfloor\}.$$

We will start by bounding the central terms $j \in A_p$ in equation (B.1). Applying Proposition 28 and the first condition in our theorem statement, we immediately get that for any $S \subseteq \{0, 1, \ldots, N\}$,

$$\frac{1}{\binom{N}{S}} \sum_{j \in A_p} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \leq 1 + o\big(\frac{1}{N}\big). \tag{B.2}$$

We now turn to the faraway terms $j \notin A_p$. For these, we note that for any non-negative integers $j, s \leq N$ we have

$$\begin{aligned} |K_s(j)| &= \left| \sum_{t=0}^{s} (-1)^t \binom{j}{t} \binom{N-j}{s-t} \right| \\ &\leq \sum_{t=0}^{s} \binom{j}{t} \binom{N-j}{s-t} \\ &= \binom{N}{s}, \end{aligned}$$

where we used the convention that $\binom{a}{b} = 0$ when $a < b$. For any $S \subseteq \{0, 1, \ldots, N\}$, we can then bound the faraway terms $j \notin A_p$ of equation (B.1) by

$$\frac{1}{\binom{N}{S}} \sum_{j \notin A_p} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \leq \binom{N}{S} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) \notin A_p\big].$$

Applying the second condition in our theorem statement in combination with Lemma 17 and the subadditivity of entropy (Lemma 16), we get

$$\begin{aligned} \max_{\substack{S \subseteq [pN \pm N^{3/4}] \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \sum_{j \notin A_p} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \right\} &\leq 2 \binom{N}{\lfloor pN + N^{3/4} \rfloor} \cdot 2 \cdot 2^{-h(p)N - N^{4/5} + N^{3/4}} \\ &\leq 4 \cdot 2^{h(p)N + h(N^{-1/4})N} \cdot 2^{-h(p)N - N^{4/5} + N^{3/4}} \\ &\leq o(\frac{1}{N}). \end{aligned}$$

Combining this bound for the faraway terms with our bound (B.2) for the central terms, we bound equation (B.1) by

$$\begin{aligned}\Pr_{z\sim p}[d(c+z)\neq c)] &\leq 2e^{-\frac{\sqrt{N}}{3p}} + N\cdot o\big(\frac{1}{N}\big)\\ &\leq o(1).\end{aligned}$$

□

B.2 Duals of Transitive Codes

In this section, we prove Claims 20 and 21; that is, we show that the dual of a (doubly) transitive code is itself (doubly) transitive.

Claim 20. *The dual code $C^\perp$ of a transitive code $C \subseteq \mathbb{F}_2^N$ is transitive.*

Proof. Let $i, j \in [N]$ be arbitrary. Since C is transitive, we know there exists a permutation $\pi : [N] \to [N]$ such that $\pi(j) = i$ and for any $c = (c_1, c_2, \ldots, c_N) \in C$, we have $c_\pi := (c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$. Clearly π^{-1} satisfies $\pi^{-1}(i) = j$, and we claim that it also satisfies that $v_{\pi^{-1}} \in C^\perp$ for all $v \in C^\perp$. For this we note that since $c_\pi \in C$ for every $c \in C$, we have by definition that every $v \in C^\perp$ satisfies

$$\sum_k v_k c_{\pi(k)} = 0 \text{ for all } c \in C.$$

We thus have

$$\begin{aligned}v \in C^\perp &\implies \sum_k v_k c_{\pi(k)} = 0 \text{ for all } c \in C\\ &\implies \sum_k v_{\pi^{-1}(k)} c_k = 0 \text{ for all } c \in C\\ &\implies v_{\pi^{-1}} \in C^\perp.\end{aligned}$$

□

Claim 21. *The dual code $C^\perp$ of a doubly transitive code $C \subseteq \mathbb{F}_2^N$ is doubly transitive.*

Proof. Let $i, j, k, l \in [N]$ be such that $i \neq k$ and $j \neq l$. Since C is doubly transitive, we know there exists a permutation $\pi : [N] \to [N]$ such that $\pi(j) = i$, $\pi(l) = k$, and for any $c = (c_1, c_2, \ldots, c_N) \in C$, we have $c_\pi := (c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$. Clearly π^{-1} satisfies $\pi^{-1}(i) = j$ and $\pi^{-1}(k) = l$, and we claim that it also satisfies that $v_{\pi^{-1}} \in C^\perp$ for all $v \in C^\perp$. For this we note that since $c_\pi \in C$ for every $c \in C$, we have by definition that every $v \in C^\perp$ satisfies

$$\sum_{t=1}^{N} v_t c_{\pi(t)} = 0 \text{ for all } c \in C.$$

We thus have

$$\begin{aligned} v \in C^\perp &\implies \sum_t v_t c_{\pi(t)} = 0 \text{ for all } c \in C \\ &\implies \sum_t v_{\pi^{-1}(t)} c_t = 0 \text{ for all } c \in C \\ &\implies v_{\pi^{-1}} \in C^\perp. \end{aligned}$$

□

B.3 Necessity of the Distance Condition in Theorem 31

We will show that there exists a linear code $C \subseteq \mathbb{F}_2^N$ of constant distance that achieves list-decoding capacity but not BSC_p capacity (we mentioned this fact in the introduction to Chapter 4). First, start with a linear code $C' \subseteq \mathbb{F}_2^N$ of rate $1 - h(p)$ that achieves list-decoding capacity. Our goal will be to show that the code $C = \mathrm{span}\{e_1, C'\}$ also achieves list decoding capacity. Indeed, consider any sequence $L_N = \omega(1)$. Since C' achieves list decoding capacity, there exists a sequence $\epsilon_N = o(1)$ such that C' is $(p - \epsilon, \frac{L}{2})$-list decodable. We claim that C must be $(p - 2\epsilon, L)$-list decodable. Indeed, consider any $z \in \mathbb{F}_q^N$ and its associated ball $C' \cap B_{(p-2p')n+1}(z) = \{c_1, \ldots, c_t\}$. Since $(p - 2\epsilon)N < (p - \epsilon)N - 1$, we must have $C \cap B_{(p-2p')N}(z) \subseteq \{c_1, \ldots, c_t, c_1 + e_1, \ldots, c_t + e_1\}$. As C' was $(p - \epsilon, \frac{L}{2})$-list decodable, C must be $(p - 2\epsilon, L)$ list decodable.

However, it is clear that C cannot achieve BSC capacity. Indeed, if we send some $c \in C$ and the first bit gets corrupted, we cannot distinguish between c and $c + e_1$.

Appendix C

PROOF OF VARIOUS INEQUALITIES

In this appendix, we will go over a few technical inequalities whose proof we deferred to the appendix.

C.1 A version of Pinsker's Inequality

In this section, we prove Lemma 19.

Lemma 19. *For any $\mu \in (0,1)$, we have*

$$1 - h\left(\frac{1-\mu}{2}\right) = \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^{2i}}{i(2i-1)},$$

and thus

$$\frac{\mu^2}{2\ln 2} \le 1 - h\left(\frac{1-\mu}{2}\right) \le \mu^2.$$

Proof.

$$\begin{aligned}
1 - h(\frac{1-\mu}{2}) &= 1 + \frac{1-\mu}{2}\log\left(\frac{1-\mu}{2}\right) + \frac{1+\mu}{2}\log\left(\frac{1+\mu}{2}\right) \\
&= \frac{1-\mu}{2}\log(1-\mu) + \frac{1+\mu}{2}\log(1+\mu) \\
&= \frac{1}{2\ln 2}\left[-(1-\mu)\sum_{i=1}^{\infty}\frac{\mu^i}{i} - (1+\mu)\sum_{i=1}^{\infty}(-1)^i\frac{\mu^i}{i}\right] \\
&= \frac{1}{2\ln 2}\left[2\mu\sum_{i=1}^{\infty}\frac{\mu^{2i-1}}{2i-1} - 2\sum_{i=1}^{\infty}\frac{\mu^{2i}}{2i}\right] \\
&= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\mu^{2i}\left(\frac{1}{2i-1} - \frac{1}{2i}\right) \\
&= \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^{2i}}{i(2i-1)}
\end{aligned}$$

Thus $1 - h(\frac{1-\mu}{2}) \ge \frac{\mu^2}{2\ln 2}$ and $1 - h(\frac{1-\mu}{2}) \le \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^2}{i(2i-1)} = \frac{1}{2\ln 2}\cdot 2\ln 2\cdot\mu^2 = \mu^2$. □

C.2 An Inequality for Lemma 59

In this section, we prove an equality that was used in the proof of Lemma 59.

Claim 74. *For any $\alpha \in (0,1)$, we have*

$$1 - 2\alpha - h(1-2\alpha) + (1-\alpha)h(\frac{1-2\alpha}{1-\alpha}) = 1 - h(\alpha).$$

Proof. We compute

$$\begin{aligned}
1 - 2\alpha - h(1-2\alpha) + (1-\alpha)h(\frac{1-2\alpha}{1-\alpha}) &= 1 - 2\alpha + (1-2\alpha)\log(1-2\alpha) + 2\alpha\log(2\alpha) \\
&\quad - (1-\alpha)\left(\frac{1-2\alpha}{1-\alpha}\log\frac{1-2\alpha}{1-\alpha} + \frac{\alpha}{1-\alpha}\log\frac{\alpha}{1-\alpha}\right) \\
&= 1 - 2\alpha + 2\alpha + 2\alpha\log\alpha + (1-2\alpha)\log(1-\alpha) - \alpha\log\frac{\alpha}{1-\alpha} \\
&= 1 + \alpha\log\alpha + (1-\alpha)\log(1-\alpha) \\
&= 1 - h(\alpha).
\end{aligned}$$

□

C.3 An Inequality for Theorem 37

In this section, we prove the inequality (7.20), which was used in the proof of Theorem 37.

Claim 75. *For any $p \in [0, \frac{1}{2}]$, we have*

$$h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) + h(p) \geq 1 + 2p(1-\sqrt{4p(1-p)}).$$

Proof. Writing the Taylor expansion of h as in the proof of Lemma 19, we have

$$h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) + h(p) = 2 - \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{(4p(1-p))^i + (1-2p)^{2i}}{i(2i-1)}.$$

But $\sum_{i=1}^{\infty} \frac{1}{i(2i-1)} = 2\ln 2$, so our previous expression can be rewritten as

$$h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) + h(p) = 1 + \frac{1}{2\ln 2}\sum_{i=1}^{\infty} \frac{1-(4p(1-p))^i - (1-4p(1-p))^i}{i(2i-1)}$$
$$\geq 1 + \frac{1}{2\ln 2}\sum_{i=2}^{\infty} \frac{1-(4p(1-p))^2 - (1-4p(1-p))^2}{i(2i-1)},$$

where in the second line we used the fact that the term $i = 1$ in the summation is 0. We will now need the following inequality:

$$1-(4p(1-p))^2 - (1-4p(1-p))^2 \geq \frac{4\ln 2\cdot(1-\sqrt{4p(1-p)})}{2\ln 2 - 1}. \tag{C.1}$$

Once we establish (C.1), our claim follows from bounding our previous inequality by

$$h\Big(\frac{1-\sqrt{4p(1-p)}}{2}\Big) + h(p) \geq 1 + \frac{1}{2\ln 2}\cdot\frac{4\ln 2\cdot(1-\sqrt{4p(1-p)})}{2\ln 2-1}\sum_{i=2}^{\infty}\frac{1}{i(2i-1)}$$
$$= 1 + \frac{2p(1-\sqrt{4p(1-p)})}{2\ln 2-1}\Big(\sum_{i=1}^{\infty}\frac{1}{i(2i-1)} - 1\Big)$$
$$= 1 + 2p(1-\sqrt{4p(1-p)})$$

It thus only remains to prove (C.1). For this, we note that the right-hand side of (C.1) can be bounded by

$$\frac{4\ln 2\cdot(1-\sqrt{4p(1-p)})}{2\ln 2-1} \leq 8p(1-\sqrt{4p(1-p)}),$$

while the left-hand side of (C.1) expands to

$$1-(4p(1-p))^2-(1-4p(1-p))^2 = 8p - 40p^2 + 64p^3 - 32p^4.$$

Thus it is sufficient to show that

$$5p - 8p^2 + 4p^3 \leq \sqrt{4p(1-p)},$$

or equivalently (squaring both sides and dividing by p) that the function

$$g(p) := 16p^5 - 64p^4 + 104p^3 - 80p^2 + 29p - 4$$

satisfies

$$g(p) \leq 0 \tag{C.2}$$

for all $p \in [0, \frac{1}{2}]$. But the derivative of g is

$$\begin{aligned} \frac{dg}{dp} &= 80p^4 - 256p^3 + 312p^2 - 160p + 29 \\ &= (1-2p)^2(20p^2 - 44p + 29), \end{aligned}$$

and the polynomial $20p^2 - 44p + 29$ has the two complex roots $\frac{11 \pm 2\sqrt{6} \cdot i}{10}$. Thus over the interval $[0, \frac{1}{2}]$, the function $g(p)$ must be maximized at either $p = 0$ or $p = \frac{1}{2}$. Since $g(0) = -4$ and $g(\frac{1}{2}) = 0$, we have

$$g(p) \leq 0$$

for all $p \in [0, \frac{1}{2}]$. We have thus shown (C.2), and we are done. □

BIBLIOGRAPHY

[1] 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.

[2] Emmanuel Abbe, Colin Sandon, and Oscar Sprumont. Tensor Reed-Muller codes: Achieving capacity with quasilinear decoding time. *CoRR*, abs/2601.16164, 2026.

[3] 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.

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

[5] Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed-muller codes: Theory and algorithms. *IEEE Trans. Inf. Theory*, 67(6):3251–3277, 2021.

[6] 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.

[7] 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.

[8] Alexander Barg. Stolarsky's invariance principle for finite metric spaces. *Mathematika*, 67(1):158–186, 2021.

[9] Paul Beame, Shayan Oveis Gharan, and Xin Yang. On the bias of reed-muller codes over odd prime fields. *SIAM J. Discret. Math.*, 34(2):1232–1247, 2020.

[10] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complex.*, 21(1):63–81, 2012.

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

[12] Claude Berrou, Alain Glavieux, and Punya Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo-codes. 1. In *Proceedings of ICC '93 - IEEE International Conference on Communications*, volume 2, pages 1064–1070 vol.2, 1993.

[13] Dmitriy Bilyk, Feng Dai, and Ryan Matzke. Stolarsky principle and energy optimization on the sphere. *Constructive Approximation*, 48, 08 2018.

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

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

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

[17] 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.

[18] Yeyuan Chen and Zihan Zhang. Explicit folded reed-solomon and multiplicity codes achieve relaxed generalized singleton bounds. In Michal Koucký and Nikhil Bansal, editors, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*, pages 1–12. ACM, 2025.

[19] Ronald de Wolf. A brief introduction to fourier analysis on the boolean cube. *Theory Comput.*, 1:1–20, 2008.

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

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

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

[23] David Galvin. Three tutorial lectures on entropy and counting, 2014.

[24] Marvin Geiselhart, Ahmed Elkelesh, Moustafa Ebada, Sebastian Cammerer, and Stephan ten Brink. Automorphism ensemble decoding of reed-muller codes. *IEEE Trans. Commun.*, 69(10):6424–6438, 2021.

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

[26] John Edensor Littlewood Godfrey Harold Hardy and George Polya. *Inequalities*. Cambridge University Press, 1934.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

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

[35] 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.

[36] 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.

[37] R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950.

[38] 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.

[39] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *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*, volume 6302 of *Lecture Notes in Computer Science*, pages 617–631. Springer, 2010.

[40] Mourad E.H Ismail and Plamen Simeonov. Strong asymptotics for krawtchouk polynomials. *Journal of Computational and Applied Mathematics*, 100(2):121–144, 1998.

[41] Kirill Ivanov and Rüdiger L. Urbanke. Capacity-achieving codes: a review on double transitivity. *CoRR*, abs/2010.15453, 2020.

[42] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 68–80. IEEE Computer Society, 1988.

[43] Gil Kalai and Nathan Linial. On the distance distribution of codes. *IEEE Trans. Inf. Theory*, 41(5):1467–1472, 1995.

[44] Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of reed-muller codes. *IEEE Trans. Inf. Theory*, 58(5):2689–2696, 2012.

[45] 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.

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

[47] 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.

[48] Ilia Krasikov and Simon Litsyn. Survey of binary krawtchouk polynomials. In Alexander Barg and Simon Litsyn, editors, *Codes and Association Schemes, Proceedings of a DIMACS Workshop, Piscataway, New Jersey, USA, November 9-12, 1999*, volume 56 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 199–211. DIMACS/AMS, 1999.

[49] 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.

[50] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, and Rüdiger L. Urbanke. Comparing the bit-map and block-map decoding thresholds of reed-muller codes on BMS channels. In *IEEE International Symposium on Information Theory, ISIT 2016, Barcelona, Spain, July 10-15, 2016*, pages 1755–1759. IEEE, 2016.

[51] 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.

[52] 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.

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

[54] 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.

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

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

[57] 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.

[58] Francisco Pernice, Oscar Sprumont, and Mary Wootters. List-decoding capacity implies capacity on the q-ary symmetric channel. In Michal Koucký and Nikhil Bansal, editors, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*, pages 855–866. ACM, 2025.

[59] Henry D. Pfister, Oscar Sprumont, and Gilles Zémor. From bit to block: Decoding on erasure channels. In *IEEE International Symposium on Information Theory, ISIT 2025, Ann Arbor, MI, USA, June 22-27, 2025*, pages 1–6. IEEE, 2025.

[60] Yury Polyanskiy. Hypercontractivity of spherical averages in hamming space. *SIAM J. Discret. Math.*, 33(2):731–754, 2019.

[61] Anup Rao and Oscar Sprumont. A criterion for decoding on the binary symmetric channel. *Adv. Math. Commun.*, 19(2):437–477, 2025.

[62] 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.

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

[64] Alex Samorodnitsky. An improved bound on l_q norms of noisy functions. *CoRR*, abs/2010.02721, 2020.

[65] Elia Santi, Christian Häger, and Henry D. Pfister. Decoding reed-muller codes using minimum- weight parity checks. In *2018 IEEE International Symposium on Information Theory, ISIT 2018, Vail, CO, USA, June 17-22, 2018*, pages 1296–1300. IEEE, 2018.

[66] Ramprasad Saptharishi, Amir Shpilka, and Ben Lee Volk. Efficiently decoding Reed-Muller codes from random errors. *IEEE Trans. Inf. Theory*, 63(4):1954–1960, 2017.

[67] Eren Sasoglu. Polar coding theorems for discrete systems. *PhD thesis*, 2011.

[68] Ori Sberlo and Amir Shpilka. On the performance of reed-muller codes with respect to random errors and erasures. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1357–1376. SIAM, 2020.

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

[70] R. Singleton. Maximum distanceq-nary codes. *IEEE Transactions on Information Theory*, 10(2):116–118, 1964.

[71] M. Skriganov. Point distributions in two-point homogeneous spaces. *Mathematika*, 65:557–587, 03 2019.

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

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

[74] Michel Talagrand. On russo's approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.

[75] 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.

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

[77] 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.

[78] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Dokl. Akad. Nauk SSSR Soviet Math*, 117:739–741, 1957.

[79] Victor K.-W. Wei. Generalized Hamming weights for linear codes. *IEEE Trans. Inf. Theory*, 37(5):1412–1418, 1991.

[80] 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.

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

[82] Min Ye and Emmanuel Abbe. Recursive projection-aggregation decoding of reed-muller codes. *IEEE Trans. Inf. Theory*, 66(8):4948–4965, 2020.

[83] 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.

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