A Criterion for Decoding on the BSC

Anup Rao[*]

School of Computer Science, University of Washington

anuprao@cs.washington.edu

Oscar Sprumont[†]

School of Computer Science, University of Washington

osprum@cs.washington.edu

June 15, 2022

Abstract

We present an approach to showing that a linear code is resilient to random errors. We use this approach to obtain decoding results for both transitive codes and Reed-Muller codes. We give three kinds of results about linear codes in general, and transitive linear codes in particular.

1. We give a tight bound on the weight distribution of every transitive linear code $C \subseteq \mathbb{F}_2^N$: $\Pr_{c \in C}[|c| = \alpha N] \leq 2^{-(1-h(\alpha)) \cdot \dim(C)}$.

2. We give a criterion that certifies that a linear code C can be decoded on the binary symmetric channel. Let $K_s(x)$ denote the Krawtchouk polynomial of degree s, and let $C^\perp$ denote the dual code of C. We show that bounds on $\mathbb{E}_{c \in C^\perp}[K_{\epsilon N}(|c|)^2]$ imply that C recovers from errors on the binary symmetric channel with parameter ϵ. Weaker bounds can be used to obtain list-decoding results using similar methods. One consequence of our criterion is that whenever the weight distribution of $C^\perp$ is sufficiently close to the binomial distribution in some interval around $\frac{N}{2}$, C is resilient to ϵ-errors.

3. We combine known estimates for the Krawtchouk polynomials with our weight bound for transitive codes, and with known weight bounds for Reed-Muller codes, to obtain list-decoding results for both these families of codes. In some regimes, our bounds for Reed-Muller codes achieve the information-theoretic optimal trade-off between rate and list size.

1 Introduction

In his seminal 1948 paper, Shannon laid out the bases of coding theory and introduced the concept of channel capacity, which is the maximal rate at which information can be transmitted over a communication channel [Sha48]. The two channels that have received the most attention are the Binary Symmetric Channel (BSC), where each bit is independently flipped with some probability ϵ, and the Binary Erasure Channel (BEC), where each bit is independently replaced by an erasure symbol with some probability ϵ. Shannon's work initiated a decades-long search for explicit codes that can achieve high rates over a noisy channel.

Explicit construction of codes often have a lot of symmetry. In particular, many known constructions of codes are *transitive*. The group of symmetries of a code is the subgroup G of permutations $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$ such that permuting the coordinates of each of the codewords using π does not change the code. A code is transitive if for every two coordinates i, j, there is a permutation $\pi \in G$ with $\pi(i) = j$. A code is 2-transitive if for every $i \neq k$, $j \neq \ell$ there is a permutation $\pi \in G$ with $\pi(i) = j, \pi(k) = \ell$. Many known constructions of codes are *cyclic*, and every cyclic code is transitive. Reed-Solomon codes, BCH codes and Reed-Muller codes are all transitive.

The binary code that is arguably the cleanest explicit candidate to achieving capacity over both the BSC and the BEC is the family of Reed-Muller codes. 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 d in n variables.

Reed-Muller codes enjoy strong symmetry beyond transitivity: their symmetry group is the group of invertible affine transformations over $\mathbb{F}_2^n$. Using fundamental results from Fourier analysis about the influences of symmetric boolean functions [KKL88, Tal94, BK97] has led to a very successful line of work, with [KKM+16] showing that Reed-Muller codes achieve capacity over the BEC and [HSS21] showing that they are polynomially close to achieving capacity over the BSC. In fact, [KKM+16] show that if a linear code $C \subseteq \mathbb{F}_2^N$ has a 2-transitive symmetry group G such that for every $S \subseteq \{1, \ldots, N\}$ with $|S| = (s \log N)^{0.99}$, $|\{\pi(S) : \pi \in G\}| \geq N^s$, then C can tolerate $\epsilon - O(1/s)$ fraction of random erasures. Given these results, it is natural to investigate the types of symmetry that lead to good codes. In this paper, we prove three kinds of results relevant to understanding the error resilience of general linear codes, transitive linear codes, and Reed-Muller codes.

1. We give a clean and tight weight distribution bound for every transitive linear code. We show that for any such code $C \subseteq \mathbb{F}_2^N$,

$$\Pr_{c \in C}[|c| = \alpha N] \leq 2^{-(1-h(\alpha)) \cdot \dim(C)}.$$

   This bound is proved by combining transitivity with the subadditivity of entropy. In some regimes, it improves on all previously known weight bounds for Reed-Muller codes (see Appendix A).

2. We give a new criterion to validate that a code can be decoded over the BSC. Let

$$K_t(x) = \sum_{j=0}^{t} (-1)^j \binom{x}{j} \binom{N-x}{t-j}$$

denote the Krawtchouk polynomial of degree t, and let $C^\perp$ denote the dual code of C. In spirit, our criterion says that any code C satisfying

$$\mathop{\mathbb{E}}_{c \in C^\perp} [K_{\epsilon N}(|c|)^2] < (1 + o(N^{-1})) \cdot \binom{N}{\epsilon N}$$

can be uniquely decoded on the BSC with high probability. Our actual result is a little more technically involved (see Theorem 2). This criterion implies that any code whose dual codewords are distributed sufficiently close to the binomial distribution must be resilient to ϵ-errors (see Corollary 3). Moreover, if the above expectation is bounded by $o(k\binom{N}{\epsilon N}^{-1})$, then we prove that the code can be list-decoded with a list size of about k.

3. Finally, we combine known estimates for the Krawtchouk polynomials with our weight bound for transitive codes, and with known weight bounds for Reed-Muller codes, to obtain list-decoding results for both families of codes. In some regimes, our bounds for Reed-Muller codes achieve the information-theoretic optimal trade-off between rate and list size.

Next, we discuss our results more rigorously. We note that throughout this section, for any set X we denote the uniform distribution over X by $\mathcal{D}(X)$.

I. Weight Bounds for Transitive Codes

We bound the weight distribution of any transitive linear code over any prime field. See section 6 for the proof.

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

$$\mathop{\Pr}_{c \sim \mathcal{D}(C)} \left[|c| = \alpha N \right] \leq q^{-(1 - h_q(\alpha)) \dim C},$$

where $\mathcal{D}(C)$ is the uniform distribution over all codewords in C, $|c|$ is the number of non-zero coordinates of c, and h_q is the q-ary entropy

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

Note that $h_2(\alpha)$ denotes the binary entropy function. We note that in some regimes (for e.g. when the degree satisfies $0.38n < d < 0.499n$ and α is larger than some constant depending on d/n), the bound above improves on all previously proven weight

3

distribution bounds for Reed-Muller codes, even though the only feature of the code that we use is transitivity. See Appendix A for some details.

II. A Criterion for Decoding on the BSC

We develop a new approach for proving decoding results over the BSC, i.e. the communication channel whose errors $z \in \mathbb{F}_2^N$ are sampled from the ϵ-noisy distribution

$$P_\epsilon(z) = \epsilon^{|z|}(1-\epsilon)^{N-|z|}$$

for some $\epsilon \in (0,1)$. Our approach is based on Fourier analysis, although unlike [KKM+16] and [HSS21], the ideas we use do not rely on bounds on influences. We obtain the following result (recall that $\mathcal{D}(C^\perp)$ denotes the uniform distribution over $C^\perp$):

Theorem 2. *Let $C \subseteq \mathbb{F}_2^N$ be any linear code, and denote by $C^\perp \subseteq \mathbb{F}_2^N$ its dual code. Then for any $\epsilon \in (0, \frac{1}{2})$, there exists a decoding function $d : \mathbb{F}_2^N \to \mathbb{F}_2^N$ such that for all $c \in C$ we have*

$$\Pr_{\rho \sim P_\epsilon}[d(c+\rho) \neq c] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + N \max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \mathop{\mathbb{E}}_{c \sim \mathcal{D}(C^\perp)} \left[K_S(|c|)^2 \right] - 1 \right\},$$

where $\binom{N}{S} = \sum_{j \in S} \binom{N}{S}$, and where $K_S(x) = \sum_{j \in S} K_j(x)$ for K_j the Krawtchouk polynomial of degree j.

We will now consider one interesting consequence of Theorem 2. Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and define

$$A_\epsilon = \{\alpha N : h(\alpha) > 1 - h(\epsilon) - N^{-1/5}\}.$$

Our next corollary states that whenever the dual codewords of C are distributed sufficiently close to the binomial distribution for all weights in A_ϵ, the code C must be resilient to ϵ-errors. See Appendix B for the proof.

Corollary 3. *Let $C \subseteq \mathbb{F}_2^N$ be a linear code, and let $\epsilon \in (0, \frac{1}{2})$ be arbitrary. Suppose that for every $j \in A_\epsilon$ we have*

$$\Pr_{y \sim \mathcal{D}(C^\perp)} \left[|y| = j \right] \leq \left(1 + o(N^{-1}) \right) \frac{\binom{N}{j}}{2^N},$$

and suppose that

$$\Pr_{y \sim \mathcal{D}(C^\perp)} \left[|y| \notin A_\epsilon \right] \leq 2^{N^{\frac{3}{4}}} \cdot \frac{\sum_{i \notin A_\epsilon} \binom{N}{i}}{2^N}.$$

Then C is resilient to ϵ-errors.

4

As a proof of concept, we note that a uniformly random linear code of dimension $(1 - h(\epsilon))N + \sqrt{N}$ satisfies all these conditions simultaneously with high probability.

III. List Decoding Results

Using a generalized version of Theorem 2 (namely, Theorem 21 in section 5), we obtain list decoding bounds for both transitive codes and Reed-Muller codes. We start with our bound for Reed-Muller codes.

Theorem 4. *Let $\epsilon \in (0, \frac{1}{2})$ and $\gamma \in (0, 1)$ be such that $1 - \gamma \geq 2^{-\frac{2\epsilon}{(\ln 2)^2}}$. Then the Reed-Muller code $\mathsf{RM}(n, d)$ of dimension $\binom{n}{\leq d} = (1 - \gamma)N$ can with high probability list-decode ϵ-errors using a list T of size*

$$|T| = 2^{(h(\epsilon) - \gamma)N + o(N)} + 2^{4\epsilon N + o(N)}.$$

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

$$|T| \gtrsim 2^{(h(\epsilon) - \gamma)N}. \tag{1}$$

Our bound in Theorem 4 shows that Reed-Muller codes achieve these optimal parameters, at least in some regimes (for e.g. when $\binom{n}{\leq d} \geq 1 - \frac{1.99\epsilon}{\ln 2}$ and ϵ is small enough). We now turn to our list-decoding bound for transitive codes.

Theorem 5. *Fix any $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0, 1)$, and $N > \left(\frac{5}{\epsilon}\right)^{20}$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode ϵ-errors using a list T of size*
$$|T(x)| = 2^{\epsilon N \log(\frac{2}{1-\eta}) + o(N)} + 2^{4\epsilon N}.$$

As an explicit example of the types of bounds one gets from Theorem 5, we have that any transitive linear code of dimension $\dim C = (1 - \frac{4\epsilon}{e})N$ can with high probability list-decode ϵ-errors using a list T of size

$$|T| = 2^{(h(\epsilon) - \epsilon + \frac{\epsilon^2}{\ln 2})N + o(N)} + 2^{4\epsilon N}.$$

For comparison, recall that our lower bound (1) states that any code C of dimension $(1 - \frac{4\epsilon}{e})N$ requires a list size of at least about $2^{(h(\epsilon) - \frac{4\epsilon}{e})N}$.

1.1 Techniques

Our weight distribution bound for transitive linear codes (Theorem 1) is proven by showing that the entropy of a uniformly random codeword of weight αN is small. To do

this, we analyze the entropy of the coordinates corresponding to linearly independent columns of the generator matrix. Transitivity implies that every coordinate in the code has the same entropy, and subadditivity of entropy can then be used to bound the entropy of the entire distribution.

To obtain our decoding criterion, we 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 ϵN of the coordinates in the codeword are flipped, although our results actually hold over the BSC as well. 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 $Hz^\intercal$, the codeword can be decoded. In the case where z is uniformly distributed on vectors of weight ϵN, this amounts to showing that with high probability, the coset of z does not contain any string of weight ϵN (in other words, there is no $w \in \mathbb{F}_2^N$ of weight $|w| = \epsilon N$ such that $Hz^\perp = Hw^\perp$). This can be understood by computing the norm

$$\|f\|_2^2 = \frac{1}{2^N} \sum_y f(y)^2 = \frac{1}{2^N} \sum_y \Pr[Hz^\intercal = y^\intercal]^2,$$

where $f(y) = \Pr[Hz^\intercal = y^\intercal]$. The norm above is always at least $2^{-N} \binom{N}{\epsilon N}^{-1}$, and if it is close to $2^{-N} \binom{N}{\epsilon N}^{-1}$ then the code can be decoded with high probability. If $\|f\|_2^2$ is larger than $2^{-N} \binom{N}{\epsilon N}^{-1}$, then we show that the code can be list-decoded with high probability, where the size of the list is related to $2^N \binom{N}{\epsilon N} \|f\|_2^2$.

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

$$\|f\|_2^2 = \sum_{j=0}^N \Pr[|c^\perp| = j] \cdot K_{\epsilon N}(j)^2, \tag{2}$$

where $c^\perp$ is a uniformly random codeword in the dual code and $K_{\epsilon N}$ is the Krawtchouk polynomial of degree ϵN. We note that such relations for the coset weight distribution have been used to understand the discrepancy of subsets of the sphere, as well as subsets of other homogeneous spaces. In particular, (2) was proven in a slightly different form in [Bar21] (see Theorem 2.1 and Lemma 4.1), whereas over $\mathbb{R}^N$ results of this type had previously been derived in [BDM18, Skr19].

Using estimates for the magnitude of Krawtchouk polynomials and bounds for the weight distribution of the dual code $C^\perp$, 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 ϵN. 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 $\approx \epsilon N$, i.e. z is taken uniformly at random from the set $S = \{x \in \mathbb{F}_2^N : |x| = \epsilon N \pm N^{3/4}\}$.

Our next step is then to show that the ℓ_2 norm corresponding to the ϵ-biased 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 ϵ-biased distribution and the uniform distribution must behave very similarly over strings of weight in S. It then follows that their corresponding ℓ_2 norms must be similar as well.

Our decoding criteria (Theorem 2, Corollary 3) are thus obtained by bounding the norm $\|f\|_2^2$ using estimates for Krawtchouk polynomials and for the weight distribution of the dual code $C^\perp$. Our list-decoding results (Theorems 4 and 5) then follow from our weight bound for transitive codes (Theorem 1) and from a weight bound of Samorodnitsky for Reed-Muller codes (Theorem 6).

1.2   Related Work

It has been shown that LDPC codes achieve capacity over Binary Memoryless Symmetric Channels (BMS) [LMS+97, KRU13, Gal62], which includes both the BSC and the BEC. These constructions are not deterministic, and it is only with the advent of polar codes [Ari09] that we obtained capacity-achieving codes with both a deterministic constructions and efficient encoding and decoding algorithms.

Polar codes are closely related to Reed-Muller codes, in the sense that they also consist of subspaces that correspond to polynomials over $\mathbb{F}_2$[Ari09]. In [Ari09] it was shown that Polar codes achieve capacity over the BSC, and algorithms were given to both encode and decode them.

It has long been believed that Reed-Muller codes achieve capacity, and significant progress has been made in that direction over the last few years. (See [ASY21] for a discussion on the subject, as well as a thorough exposition to Reed-Muller codes). Abbe, Shpilka and Wigderson first showed that Reed-Muller codes achieve capacity over the BSC and the BEC for sub-constant and super-constant rates [ASW15]. Kudekar, Kumar, Mondelli, Pfister, Sasoglu and Urbanke then proved that in the constant rate regime, Reed-Muller codes achieve capacity over the BEC channel [KKM+16]. Abbe and Ye showed that the Reed-Muller transform polarizes the conditional mutual information, and proved that some non-explicit variant of the Reed-Muller code achieves capacity [AY19]. (They conjecture that this variant is in fact the Reed-Muller code itself). Hazla, Samorodnitsky and Sberlo then proved that Reed-Muller codes of constant rates can decode a constant fraction of errors on the BSC [HSS21]; this had previously been shown for Almost-Reed-Muller codes by Abbe, Hazla and Nachum [AHN21]. Most recently, Reeves and Pfister showed that Reed-Muller codes achieve capacity over all BMS channels under bit-MAP decoding [RP21], i.e. that one can with high probability recover any single bit of the original codeword (but not with high enough probability that one could take a union bound). Despite these breakthroughs, the conjecture that Reed-Muller codes achieve capacity over all BMS channels under block-MAP decoding (i.e. recover the whole codeword with high probability) is ultimately still open.

Weight Bounds for Reed-Muller Codes
Several past works have proven bounds on the weight distribution of Reed-Muller codes.

Kaufman, Lovett and Porat gave asymptotically tight bounds on the weight distribution of Reed-Muller codes of constant degree [KLP12]. Abbe, Shpilka and Wigderson then built on these techniques to obtain bounds for all degrees smaller than $\frac{n}{4}$ [ASW15], before Sberlo and Shpilka again improved the approach and obtained bounds for all degrees [SS20]. Most recently, Samorodnitsky used completely different ideas to obtain weight bounds in the regime where both the rate of the code and the normalized weight of the codeword are $\Theta(1)$ [Sam20]. We will later use his following result in our list-decoding arguments:

Theorem 6 ([Sam20]). *Let $\binom{n}{\leq d} = \eta 2^n = \eta N$ for some $\eta \in (0,1)$, and denote by $\mathcal{D}(n,d)$ the uniform distribution over all codewords in $\mathsf{RM}(n,d)$. Then for any $\alpha \in (0, \frac{1}{2})$ we have*

$$\Pr_{c \sim \mathcal{D}(n,d)}[|c| \leq \alpha N] \leq 2^{o(N)} \left(\frac{1}{1-\eta} \right)^{2 \ln 2 \cdot \alpha N} 2^{-\eta N}.$$

These bounds are strong when $\alpha \ll 1/2$. For α close to $1/2$, the first results we are aware of are due to Ben-Eliezer, Hod and Lovett [BHL12]. Their bounds, which were extended to Reed-Muller codes over prime fields by Beame, Oveis Gharan and Yang [BGY20], are strongest when the degree is sublinear. Sberlo and Shpilka then obtained bounds for all degrees in [SS20], while Samorodnitsky again obtained bounds in the regime where both α and η are $\Theta(1)$ [Sam20].

We note that in some regimes (for e.g. when the degree satisfies $0.38n < d < 0.499n$ and α is larger than some constant depending on d/n), our Theorem 1 improves on all the aforementioned weight bounds. See Appendix A for some details.

List Decoding

List decoding was proposed by Elias in 1957 as an alternative to unique decoding [Eli57]. In the list decoding framework, the receiver of a corrupted codeword is asked to output a list of potential codewords, with the guarantee that with high probability one of these codewords is the original one. This of course allows for a greater fraction of errors to be tolerated.

The list decoding community has largely focused on proving results for the adversarial noise model, and many codes are now known to achieve list-decoding capacity. For example uniformly random codes achieve capacity, as do uniformly random linear codes [GHSZ02, LW18, GHK11]. Folded Reed-Solomon codes were the first explicit codes to provably achieve list-decoding capacity [GR08], followed by several others a few years later [GX12, Kop15, HRW17, MRR+20]. For the rest of this paper however, we will exclusively work in the model where the errors are stochastic. In this model, the strongest known list decoding bound for the code $\mathsf{RM}(n,d)$ with $\binom{n}{\leq d} = \eta N > N - N \log(1 + 2\sqrt{\epsilon(1-\epsilon)})$ is, to our knowledge, that one can output a list T of size

$$|T| = 2^{\epsilon N \log \frac{4\epsilon(1-\epsilon)}{(1-\eta)^4 \ln 2} + o(N)} \tag{3}$$

8

and succeed with high probability in decoding ϵ-errors. This result, although not explicitly stated in [Sam20], can be obtained from his weight bound of Theorem 6 by bounding the expected number of codewords that end up closer to the received string than the original codeword, and then applying Markov's inequality. We note that the expression in (3) stays strictly below the optimal size of $2^{h(\epsilon)N-(1-\eta)N+o(N)}$ (see Appendix D.1 for a proof of this).

Krawtchouk polynomials

The Krawtchouk polynomial of degree s is the polynomial

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

For any subset $S \subseteq \{0, ..., N\}$, we will be interested in the corresponding 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 $K_S(|v|)$. The following proposition follows from standard results (see for instance [KL99], or Theorem 16 in [MS77]).

Proposition 7. *For any N and any $S \subseteq \{1, ..., N\}$, 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 [KL95, IS98, Pol19] (see for e.g. [Pol19], Lemma 2.1). These estimates are asymptotically tight in the exponent. Note that $|K_s(x)| = |K_s(N-x)| = |K_{N-s}(x)|$, so it suffices to understand the case $x, s \leq \frac{N}{2}$.

Theorem 8 ([KL95, IS98, Pol19]). *Let $\epsilon, \delta \in (0, \frac{1}{2})$ be arbitrary. If $\delta \geq \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$, then*

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

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

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

As the second expression can be somewhat cumbersome to use, [Pol19] also gives the following weaker bound (see Lemma 2.2 and equation 2.10 in [Pol19]):

Theorem 9 ([Pol19]). *For any $\epsilon \in (0, \frac{1}{2})$ and any $\delta < \frac{1}{2}\sqrt{\epsilon(1-\epsilon)}$, we have*

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

We will need the above estimates when using our Theorem 2 to obtain list-decoding results for transitive codes and Reed-Muller codes.

9

2 Notation, Conventions and Preliminaries

For the sake of conciseness, we will use the notation

$$\{a \pm l\} = \{a - l, ..., a + l\},$$

the notation

$$\binom{n}{\leq d} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d},$$

and for $S \subseteq \{0, ..., N\}$ the notation

$$\binom{N}{S} = \sum_{s \in S} \binom{N}{s}.$$

Let $N = 2^n$. We will be working with the vector spaces $\mathbb{F}_2^n$ and $\mathbb{F}_2^N$. For convenience, we associate $\mathbb{F}_2^n$ with the set $[N] = \{1, 2, \ldots, N\}$, by ordering the elements of $\mathbb{F}_2^n$ lexicographically. For $x \in \mathbb{F}_2^N$, we write $|x| = |\{j \in [N] : x_j = 1\}|$ to denote the weight of x.

2.1 Linear Codes

An N-bit code is a subset $C \subseteq \mathbb{F}_2^N$. Whenever C is a subspace of $\mathbb{F}_2^N$, we say that C is a *linear* code. Any linear code $C \subseteq \mathbb{F}_2^N$ can be represented by its generator matrix, which is a dim $C \times N$ matrix G whose rows form a basis for C. The matrix G generates all codewords of C in the sense that

$$C = \{vG : v \in \mathbb{F}_2^{\dim C}\}.$$

Another useful way to describe a linear code $C \subseteq \mathbb{F}_2^N$ is via its parity-check matrix, which is an $(N - \dim C) \times N$ matrix H whose rows span the orthogonal complement of C. The linear code C can then be expressed as

$$C = \{c \in \mathbb{F}_2^N : Hc^\intercal = 0\}.$$

One property that will play an important role is transitivity, which we define below:

Definition 1. *A set $C \subseteq \mathbb{F}_2^N$ is transitive if for every $i, j \in [N]$ there exists a permutation $\pi : [N] \to [N]$ such that*

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

(ii) For every element $v = (v_1, ..., v_N) \in C$ we have $(v_{\pi(1)}, ... v_{\pi(N)}) \in C$

We note that the dual code of a transitive code is also transitive (see Appendix D.2 for the proof).

Claim 10. *The dual code $C^\perp$ of a transitive code $C \subseteq \mathbb{F}_2^N$ is transitive.*

2.2 Reed-Muller Codes

We will denote by $\mathsf{RM}(n, d)$ the Reed-Muller code with n variables and degree 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 correspond to sets of size at most d, ordered lexicographically, and whose columns correspond to elements of $\mathbb{F}_2^n$. For $S \subseteq [n], |S| \leq d$ and $x \in \mathbb{F}_2^n$, the corresponding entry is $M_{S,x} = \prod_{j \in S} x_j$. If S is empty, this entry is set to 1.

If $v \in \mathbb{F}_2^{\binom{n}{\leq d}}$ is a row vector, v can be thought of as describing the coefficients of a multilinear polynomial in $\mathbb{F}_2[X_1, \ldots, X_n]$ of degree at most d. The row vector vM is then the evaluations of this polynomial on all inputs from $\mathbb{F}_2^n$. It is well known that M has full rank, $\binom{n}{\leq d}$. In fact we have the following standard fact (see Appendix D.3 for the proof):

Fact 11. *The columns of M that correspond to the points $x \in \mathbb{F}_2^n$ with $|x| \leq d$ are linearly independent.*

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 $MH^\mathsf{T} = 0$. 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 useful algebraic features, notably transitivity:

Fact 12. *For all n and all $d \leq n$, the Reed-Muller code $\mathsf{RM}(n, d)$ is transitive.*

See Appendix D.3 for the proof.

2.3 Entropy

The binary entropy function $h : [0, 1] \to \mathbb{R}$ is defined to be

$$h(\epsilon) = \epsilon \cdot \log \frac{1}{\epsilon} + (1 - \epsilon) \cdot \log \frac{1}{1 - \epsilon}.$$

The following fact allows us to approximate binomial coefficients using the entropy function:

Fact 13. *For $n/2 \geq d \geq 1$, $\sqrt{\frac{8\pi}{e^4 n}} \cdot 2^{h(d/n) \cdot n} \leq \binom{n}{d} \leq \binom{n}{\leq d} \leq 2^{h(d/n) \cdot n}$.*

The leftmost inequality is a consequence of Stirling's approximation for the binomial coefficients, and the rightmost is a consequence of the sub-additivity of entropy.

The following lemma, which is essentially a 2-way version of Pinsker's inequality, gives a useful way to control the entropy function near $1/2$.

Lemma 14. *For any $\mu \in (0, 1)$, we have*

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

See Appendix D.4 for the proof.

2.4 Probability Distributions

There are two types of probability distributions that we will use frequently. The first one is the ϵ-Bernoulli distribution over $\mathbb{F}_2^N$, which we will denote by

$$P_\epsilon(z) = \epsilon^{|z|}(1 - \epsilon)^{N-|z|}.$$

The second one is the uniformly random distribution over some set T, which we will denote by

$$\mathcal{D}(T)(z) = \begin{cases} \frac{1}{|T|} & \text{if } z \in T, \\ 0 & \text{otherwise.} \end{cases}$$

There are two particular cases for the uniform distribution that will occur often enough that we attribute them their own notation. The first one is the uniform distribution over $\mathbb{F}_2^t$, which we will denote by

$$\mu_t = \mathcal{D}(\mathbb{F}_2^t).$$

The second one is the uniform distribution over all vectors $z \in \mathbb{F}_2^N$ of weight $|z| \in S$, for some $S \subseteq \{0, ..., N\}$. We will denote this probability distribution by

$$\lambda_S = \mathcal{D}(\{z \in \mathbb{F}_2^N : |z| \in S\}).$$

2.5 Probability Theory

We will need two very standard results of probability theory (see for e.g. [BLM13]): Markov's inequality and Chernoff's bound. We start with Markov's inequality.

Lemma 15. *Let X be a nonnegative random variable. Then for any $a > 0$, we have*

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

We will also need Chernoff's bound:

Lemma 16. *Let $X_1, ..., X_n$ be i.i.d. random variables taking values in $\{0, 1\}$, and define $X = X_1 + ... + X_n$. Then for any $\delta \in (0, 1)$, we have*

$$\Pr\left[\left|X - \mathbb{E}[X]\right| > \delta \cdot n \,\mathbb{E}[X_1]\right] \leq 2e^{-\frac{\delta^2 \cdot n \,\mathbb{E}[X_1]}{3}}.$$

2.6 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 for e.g. [dW08]). 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,

$$\|f\|_2^2 = \langle f, f \rangle = \sum_y \hat{f}(y)^2.$$

3 Outline of the Paper

The main question we will be looking into is whether or not a family of list-decoding codes $\{C_N\}$, with $C_N \subseteq \mathbb{F}_2^N$, is asymptotically resilient to independent errors of probability ϵ. Formally, we are given a list size $k = k(N)$ and want to know if there exists a family of decoding functions $\{d_N\}$, with $d_N : \mathbb{F}_2^N \to \left(\mathbb{F}_2^N\right)^{\otimes k}$, such that for every sequence of codewords $\{c_N\}$ we have

$$\lim_{N \to \infty} \Pr_{\rho_N \sim P_\epsilon} \left[c_N \notin d_N(c_N + \rho_N)\right] = 0.$$

We note that the unique decoding problem can be seen as setting $k = 1$ in the above set-up. Our general approach will be based on trying to identify the error string $\rho \in \mathbb{F}_2^N$ from its image $H\rho^\mathsf{T}$. In particular, we will be interested in the max-likelihood decoder

$$
\begin{aligned}
D_k(x) &= \underset{\substack{\{z_1, \ldots, z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\mathsf{T} = x \text{ for all } i}}{\operatorname{argmax}} \{P_\epsilon(z_1) + \ldots + P_\epsilon(z_k)\} \\
&= \underset{\substack{\{z_1, \ldots, z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\mathsf{T} = x \text{ for all } i}}{\operatorname{argmin}} \{|z_1| + \ldots + |z_k|\}.
\end{aligned}
\tag{4}
$$

We show in the following lemma that if the max-likelihood decoder is able to identify the error string ρ, then it is possible to recover the original codeword.

Lemma 17. *Let H be the $t \times N$ parity-check matrix of the linear code C, and let $D : \mathbb{F}_2^t \to \left(\mathbb{F}_2^N\right)^{\otimes k}$ be an arbitrary function. Then there exists a decoding function*

$$d : \mathbb{F}_2^N \to \left(\mathbb{F}_2^N\right)^{\otimes k}$$

13

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

$$\Pr_{\rho \sim P_\epsilon}[c \notin d(c + \rho)] \leq \Pr_{\rho \sim P_\epsilon}[\rho \notin D(H\rho^\mathsf{T})].$$

Proof. Given $D : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^{\otimes k}$, define $d : \mathbb{F}_2^N \to (\mathbb{F}_2^N)^{\otimes k}$ to be

$$d(z) = \{z + y : y \in D(Hz^\mathsf{T})\}.$$

We will show that whenever ρ satisfies $\rho \in D(H\rho^\mathsf{T})$, ρ also satisfies $c \in d(c + \rho)$ for every $c \in C$. Suppose $\rho \in D(H\rho^\mathsf{T})$. Note that since H is the parity-check matrix of C, every $c \in C$ satisfies $Hc^\mathsf{T} = 0$. So for every $c \in C$, any ρ that satisfies $\rho \in D(H\rho^\mathsf{T})$ must also satisfy $\rho \in D(H(c^\mathsf{T} + \rho^\mathsf{T}))$. It then follows by definition of $d(c + \rho)$ that

$$c = c + \rho + \rho \in d(c + \rho).$$

$\square$

From this point onward, our goal will thus be to prove that the max-likelihood decoder in (4) succeeds in recovering ρ with high probability. In section 4, we relate the decoding error probability of the max-likelihood decoder D_k to the collision probability

$$\sum_{x \in \mathbb{F}_2^t} \Pr[Hz^\mathsf{T} = x]^2.$$

In section 5, we build on this result to obtain a bound on the performance of D_k in terms of the weight distribution of the dual code. We then present new bounds on the weight distribution of transitive codes in section 6. These bounds are interesting in their own right, and we show that they are essentially tight. In section 7, we combine these bounds with our results from section 5 to obtain list-decoding results for transitive linear codes. We then repeat this argument with Samorodnitsky's Theorem 6 in section 8 to obtain a stronger list-decoding bound for Reed-Muller codes.

4   Collisions vs Decoding

Recall that we denote by P_ϵ the ϵ-Bernoulli distribution over $\mathbb{F}_2^N$, i.e. the distribution

$$P_\epsilon(z) = \epsilon^{|z|}(1 - \epsilon)^{N - |z|}.$$

Recall also that for any subset $S \subseteq \{0, ..., N\}$, we denote by λ_S the uniform distribution over all strings $z \in \mathbb{F}_2^N$ of weight $|z| \in S$, i.e.

$$\lambda_S(z) = \begin{cases} \frac{1}{\sum_{j \in S} \binom{N}{j}} & \text{if } |z| \in S, \\ 0 & \text{otherwise.} \end{cases}$$

14

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

Definition 2. *For any $z \in \mathbb{F}_2^N$ and any subset $S \subseteq \{0, ..., N\}$, define*

$$\Omega_z^S = \left\{ y \in \mathbb{F}_2^N : |y| \in S \text{ and } Hy^\intercal = Hz^\intercal \right\}.$$

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 $H\rho^\intercal$ and are told that with high probability the error string ρ has weight $|\rho| \in S$, then we should output the list Ω_ρ^S. For unique decoding we want to argue that $|\Omega_\rho^S| = 1$ with high probability, whereas for list decoding we want to argue that $|\Omega_\rho^S| \leq k$ with high probability, for some integer $k > 1$. The expectation of $|\Omega_\rho^S|$ will thus be a key quantity in our analysis. We will call this expectation the "collision count," because it will later be useful to interpret it as the renormalized collision probability of the map $z \mapsto Hz^\intercal$ (see for instance the proof of Proposition 20).

Definition 3. *For any subset $S \subseteq \{0, .., N\}$ and any $t \times N$ matrix H, define*

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

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 18. *For any subset $S \subseteq \{0, ..., N\}$, any matrix H with N columns, and any integer $k \geq 1$, we have*

$$\mathop{\Pr}_{\rho \sim \lambda_S} \left[|\Omega_\rho^S| > k \right] \leq \frac{\mathsf{Coll}_H(S) - 1}{k}.$$

Proof. Note that $|\Omega_z^S| \geq 1$ for any $z \in \mathbb{F}_2^N$ with weight $|z| \in S$, so the random variable $|\Omega_z^S| - 1$ is always non-negative. Applying Markov's inequality (i.e. Lemma 15), we then have

$$\mathop{\Pr}_{\rho \sim \lambda_S} \left[|\Omega_\rho^S| > k \right] = \mathop{\Pr}_{\rho \sim \lambda_S} \left[|\Omega_\rho^S| - 1 \geq k \right]$$
$$\leq \frac{\mathsf{Coll}_H(S) - 1}{k}.$$

$\square$

When the error string ρ is sampled uniformly at random from the set $\{z \in \mathbb{F}_2^N : |z| \in S\}$, the above lemma allows us to relate the decoding error probability to the collision count $\mathsf{Coll}_H(S)$. The problem we are most interested in, however, is when ρ is sampled not from some uniform distribution, but from the ϵ-noisy probability distribution P_ϵ. 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 $S = \{\epsilon N \pm l\}$, for some appropriately chosen l. But in this weight band all strings have similar weight, and so are given similar probability under the distribution P_ϵ. Intuitively, the P_ϵ-decoder must then perform very similarly to the λ_S-decoder. The following theorem makes this idea precise, and then uses Lemma 18 to bound the probability of a decoding error. Recall that $D_k : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^{\otimes k}$ is the max-likelihood decoder

$$D_k(x) = \underset{\substack{\{z_1,\ldots,z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\mathsf{T} = x \text{ for all } i}}{\mathrm{argmin}} \{|z_1| + \ldots + |z_k|\}.$$

Theorem 19. *Fix $\epsilon < \frac{1}{2}$, let H be any matrix with N columns, and let $k = (2l+1)m+1$ for some integers $m \geq 0$ and $l \leq (\frac{1}{2} - \epsilon)N$. Then*

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\mathsf{T})] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{\substack{S \subseteq \{\epsilon N \pm l\} \\ 1 \leq |S| \leq 1 + \mathbb{1}\{k=1\}}} \left\{ \mathsf{Coll}_H(S) - 1 \right\},$$

where $\mathbb{1}\{k = 1\}$ is 1 when $k = 1$ and 0 otherwise

Proof. We will consider the unique decoding case ($k = 1$, i.e. $m = 0$) and the list-decoding case ($k > 1$, i.e. $m \in \mathbb{N}$) 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 $\{z \in \mathbb{F}_2^N : Hz^\perp = x, |z| = \epsilon N \pm l\}$. If there is no such string, the decoder fails. It is clear that

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq D_1(H\rho^\mathsf{T})] \leq \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\mathsf{T})],$$

since D_1 always returns the most likely string whereas $\tilde{D}_1$ may not. We thus turn to proving the desired bound for $\tilde{D}_1$. Letting

$$B = \{z \in \mathbb{F}_2^N : \big||z| - \epsilon N\big| > l\}\},$$

we have by Chernoff's bound (i.e. Lemma 16) that

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\mathsf{T})] \leq \Pr_{\rho \sim P_\epsilon}[\rho \in B] + \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\mathsf{T})\big|\rho \notin B]$$

$$\leq 2e^{-\frac{l^2}{3\epsilon N}} + \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\mathsf{T})\big|\rho \notin B]. \tag{5}$$

16

We want to bound the second term. For any $\rho \notin B$, we define the set of "problematic weights" $S(\rho) = \{\epsilon N - l, ..., |\rho|\}$. We note that for $\rho \notin B$, our decoder $\tilde{D}_1$ can only fail if there is some string $z \neq \rho$ satisfying $Hz^\perp = H\rho^\perp$ and $|z| \in S(\rho)$. Recalling the definition $\Omega_\rho^S = \{z : Hz^\perp = H\rho^\perp, |z| \in S\}$, we can then rewrite our equation (5) as

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \Pr_{\rho \sim P_\epsilon}\left[|\Omega_\rho^{S(\rho)}| > 1 \big| \rho \notin B\right].$$

Considering the most problematic weight level w within the region $\{\epsilon N \pm l\}$ and using a union bound over all lower levels $w' \leq w$, we get

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \max_{w \in \{\epsilon N \pm l\}}\left\{\Pr_{\rho \sim P_\epsilon}\left[|\Omega_\rho^{S(\rho)}| > 1 \big| |\rho| = w\right]\right\}$$

$$\leq 2e^{-\frac{l^2}{3\epsilon N}} + (2l+1) \max_{\substack{w,w' \in \{\epsilon N \pm l\} \\ w' \leq w}}\left\{\Pr_{\rho \sim P_\epsilon}\left[|\Omega_\rho^{\{w,w'\}}| > 1 \big| |\rho| = w\right]\right\}.$$

We now note that under the condition $|\rho| = w$, the probability distributions $P_\epsilon(\rho)$ and $\lambda_{w,w'}(\rho)$ are identical (they are both uniform on strings of weight w). We can thus rewrite our last inequality as

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + (2l+1) \max_{\substack{w,w' \in \{\epsilon N \pm l\} \\ w' \leq w}}\left\{\Pr_{\rho \sim \lambda_{w,w'}}\left[|\Omega_\rho^{\{w,w'\}}| > 1 \big| |\rho| = w\right]\right\}.$$

But by basic conditional probability we know that

$$\Pr_{\rho \sim \lambda_{w,w'}}\left[|\Omega_\rho^{\{w,w'\}}| > 1\right] \geq \Pr_{\rho \sim \lambda_{w,w'}}\left[|\rho| = w\right] \cdot \Pr_{\rho \sim \lambda_{w,w'}}\left[|\Omega_\rho^{\{w,w'\}}| > 1 \big| |\rho| = w\right],$$

so we can bound our previous expression by

$$\Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + (2l+1) \max_{\substack{w,w' \in \{\epsilon N \pm l\} \\ w' \leq w}}\left\{\frac{\Pr_{\rho \sim \lambda_{w,w'}}\left[|\Omega_\rho^{\{w,w'\}}| > 1\right]}{\Pr_{\rho \sim \lambda_{w,w'}}\left[|\rho| = w\right]}\right\}.$$

Now for any $w < \frac{N}{2}$ and $w' \leq w$, we have $\Pr_{\rho \sim \lambda_{\{w,w'\}}}\left[|\rho| = w\right] = \frac{\binom{N}{w}}{\binom{N}{\{w,w'\}}} \geq \frac{\binom{N}{w}}{\binom{N}{w} + \binom{N}{w'}} \geq \frac{1}{2}$. It then follows that

$$\Pr_{\rho \sim P_\epsilon}\left[\rho \notin \tilde{D}_1(H\rho^\perp)\right] \leq 2e^{-\frac{l^2}{3\epsilon N}} + 2(2l+1) \cdot \max_{\substack{S \subseteq \{\epsilon N \pm l\} \\ |S| \in \{1,2\}}}\left\{\Pr_{\rho \sim \lambda_S}\left[\Omega_\rho^S > 1\right]\right\}.$$

The theorem statement then follows from Lemma 18.

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 $\frac{k-1}{2l+1}$ strings from $\{z \in \mathbb{F}_2^N : Hz = x, |z| = w\}$, for each $w \in \{\epsilon N \pm l\}$. If there are fewer than $\frac{k-1}{2l+1}$ strings in some level w, the decoder returns all of them. It is clear that for any l we have

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq \Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal)],$$

since D_k returns the k most likely strings while $D_{k,l}$ returns at most $k-1$ strings. We thus turn to proving the desired bound for $D_{k,l}$. We first bound the probability that the error string $|\rho|$ be far away from its mean. Letting

$$B = \left\{ z \in \mathbb{F}_2^N : \left| |z| - \epsilon N \right| > l \right\},$$

we have, by Chernoff's bound (i.e. Lemma 16), that

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal)] \leq \Pr_{\rho \sim P_\epsilon}[\rho \in B] + \Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal) \big| \rho \notin B]$$

$$\leq 2e^{-\frac{l^2}{3\epsilon N}} + \max_{w \in \{\epsilon N \pm l\}} \Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal) \big| |\rho| = w].$$

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

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \max_{w \in \{\epsilon N \pm l\}} \Pr_{\rho \sim \lambda_{\{w\}}}[\rho \notin D_{k,l}(H\rho^\intercal)]$$

$$\leq 2e^{-\frac{l^2}{3\epsilon N}} + \max_{w \in \{\epsilon N \pm l\}} \Pr_{\rho \sim \lambda_{\{w\}}}[|\Omega_\rho^{\{w\}}| > \frac{k-1}{2l+1}].$$

Applying Lemma 18, we get

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{2l+1}{k-1} \cdot \max_{w \in \{\epsilon N \pm l\}} \left\{ \mathsf{Coll}_\mathsf{H}(S) - 1 \right\}$$

$$\leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{2(l+1)}{k} \cdot \max_{w \in \{\epsilon N \pm l\}} \left\{ \mathsf{Coll}_\mathsf{H}(S) - 1 \right\},$$

where in the last line we used that $\frac{a}{b} \leq \frac{a+1}{b+1}$ whenever $a \leq b$. $\qquad\square$

5   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 ϵ. 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 as we recall is defined as

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

For vectors $v \in \mathbb{F}_2^N$, we will abuse notation and write $K_s(v)$ to mean $K_s(|v|)$. For convenience, we also define for any $S \subseteq \{0, ..., N\}$ 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 20 was previously proven in a different form in [Bar21] (see Theorem 2.1 and Lemma 4.1), and can be seen as describing the coset weight distribution of the code. Recall that we use μ_t to denote the uniform distribution over all vectors in $\mathbb{F}_2^t$, and that we use the notation $\binom{N}{S} = \sum_{s \in S} \binom{N}{s}$.

Proposition 20. *Fix $\epsilon \in (0, \frac{1}{2})$, and let H be a $t \times N$ matrix with entries in $\mathbb{F}_2$. Then for any $S \subseteq \{1, ..., N\}$, we have*

$$\mathsf{Coll}_H(S) = \frac{1}{\binom{N}{S}} \mathop{\mathbb{E}}_{v \sim \mu_t} [K_S(vH)^2].$$

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

$$\frac{1}{2^t} \sum_{x \in \mathbb{F}_2^t} f(x)^2 = \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2. \tag{6}$$

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 $|\Omega_z^S| = \{y \in \mathbb{F}_2^N : |y| \in S \text{ and } Hy^\mathsf{T} = Hz^\mathsf{T}\}$ and note that

$$\mathsf{Coll}_H(S) := \frac{1}{\binom{N}{S}} \sum_{z \in \mathbb{F}_2^N : |z| \in S} |\Omega_z^S|$$

$$= \binom{N}{S} \sum_{z \in \mathbb{F}_2^N : |z| \in S} \frac{1}{|\Omega_z^S|} \mathop{\Pr}_{a \sim \lambda_S} [Ha^\mathsf{T} = Hz^\mathsf{T}]^2$$

$$= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} \mathop{\Pr}_{z \sim \lambda_S} [Hz^\mathsf{T} = x]^2.$$

We are now ready to apply Parseval's Identity. Letting $f(x) = \Pr_{z \sim \lambda_S}[Hz^\mathsf{T} = x]$ in equation (6), we get

$$\mathsf{Coll}_H(S) = \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} f(x)^2$$

$$= 2^t \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2.$$

19

But by definition we have $\hat{f}(y) := 2^{-t} \sum_{x \in \mathbb{F}_2^t} f(x) \cdot (-1)^{y \cdot x^{\mathsf{T}}}$, so the last equation can be rewritten as

$$\mathsf{Coll}_H(S) = 2^{-t} \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \left(\sum_{x \in \mathbb{F}_2^t} f(x) \cdot (-1)^{y \cdot x^{\mathsf{T}}} \right)^2. \tag{7}$$

Define the function $L_S(z)$ to be 1 if $z \in \mathbb{F}_2^N$ satisfies $|z| \in S$, and 0 otherwise. We can then express $f(x)$ as

$$f(x) = \Pr_{z \sim \lambda_S} \left[Hz^{\mathsf{T}} = x \right] = \frac{1}{\binom{N}{S}} \sum_{\substack{z \in \mathbb{F}_2^N \\ Hz^{\mathsf{T}} = x}} L_S(z). \tag{8}$$

Combining expressions (7) and (8) and applying the definition of the Fourier transform, we get

$$\mathsf{Coll}_H(S) = 2^{-t} \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \left(\sum_{z \in \mathbb{F}_2^N} \frac{L_S(z)}{\binom{N}{S}} \cdot (-1)^{yHz^{\mathsf{T}}} \right)^2$$

$$= \frac{2^{2N-t}}{\binom{N}{S}} \sum_{y \in \mathbb{F}_2^t} \hat{L}_S(yH)^2$$

$$= \frac{2^{-t}}{\binom{N}{S}} \sum_{y \in \mathbb{F}_2^t} K_S(yH)^2.$$

$\square$

We will now combine Theorem 19 and Proposition 20 to obtain Theorem 2, i.e. to obtain a bound on the decoding error probability in terms of the Fourier coefficients of the level function L_ϵ. We prove a generalized version of Theorem 2 below. To recover Theorem 2, set $k = 1$ and $l = N^{3/4}$. (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 $e^{-\frac{\sqrt{N}}{3\epsilon}}$ is small).

Theorem 21. *Fix $\epsilon \in (0, \frac{1}{2})$, let H be any $t \times N$ Boolean matrix, and let $k = (2l+1)m+1$ for any integers $m \geq 0$ and $l \leq (\frac{1}{2} - \epsilon)N$. Then*

$$\Pr_{\rho \sim P_\epsilon} \left[\rho \notin D_k(H\rho^{\mathsf{T}}) \right] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{\substack{S \subseteq \{\epsilon N \pm l\} \\ 1 \leq |S| \leq 1 + \mathbb{1}\{k=1\}}} \left\{ \frac{1}{\binom{N}{S}} \mathbb{E}_{v \sim \mu_t} \left[K_S(vH)^2 \right] - 1 \right\},$$

where the function $\mathbb{1}\{k = 1\}$ is 1 when $k = 1$, and 0 otherwise.

Proof. Applying Theorem 19 and Proposition 20, we have

$$\Pr_{\rho \sim P_\epsilon} \left[\rho \notin D_k(H\rho^{\mathsf{T}}) \right] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{\substack{S \subseteq \{\epsilon N \pm l\} \\ 1 \leq |S| \leq 1 + \mathbb{1}\{k=1\}}} \left\{ \mathsf{Coll}_H(S) - 1 \right\}$$

$$= 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{\substack{S \subseteq \{\epsilon N \pm l\} \\ 1 \leq |S| \leq 1 + \mathbb{1}\{k=1\}}} \left\{ \frac{1}{\binom{N}{S}} \mathbb{E}_{v \sim \mu_t} \left[K_S(vH)^2 \right] - 1 \right\}.$$

20

□

One consequence of Theorem 21 is Corollary 3, which states that C is resilient to ϵ-errors if the weight distribution of $C^\perp$ is close enough to the binomial distribution (see Appendix B for the proof). As another application of Theorem 21, 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 22. *Fix any $\epsilon \in (0, \frac{1}{2})$, and define $\beta = \frac{1-2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}}{2}$ for $\tilde{\epsilon} = \epsilon + N^{-1/4}$. Let $B = \{\beta N, ..., (1-\beta)N\}$, and let $k^* = (2N^{3/4}+1)m+1$ for some integer $m \geq 0$. Then for all $N > \left(\frac{5}{\epsilon}\right)^{20}$ and all integers $k \geq k^*$, we have that any $t \times N$ matrix H with entries in $\mathbb{F}_2$ satisfies*

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\mathsf{T})] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{N}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}$$

$$+ \frac{2^{h(\epsilon)N+N^{\frac{4}{5}}}}{k^*} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot 2^{2\epsilon N \log|1-\frac{2j}{N}|} \right\}.$$

Proof. We will use Theorem 21 to bound the decoding error probability in terms of the Krawtchouk polynomials $K_S(j)$ and the probability factors $\Pr_{v \sim \mu_t}[|vH| = j]$. Some of these terms will then be bounded using Proposition 7, and some will be bounded using Theorem 9. We proceed with the proof; letting $l = N^{3/4}$ in Theorem 21, we get

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\mathsf{T})] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{N}{k^*} \max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \sum_{j=0}^{N} \Pr_{v \sim \mu_t}[|vH| = j] K_S(j)^2 - 1 \right\}.$$

$$(9)$$

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 7, which states that $\frac{2^{-N}}{\binom{N}{S}} \sum_{j=0}^{N} \binom{N}{j} \cdot K_S(j)^2 = 1$ for all $S \subseteq \{0, ..., N\}$. For any $S \subseteq \{0, ..., N\}$, we then get

$$\frac{1}{\binom{N}{S}} \sum_{j \in B} \Pr_{v \sim \mu_t}[|vH| = j] K_S(j)^2 \leq \frac{1}{\binom{N}{S}} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\} \sum_{j \in B} \binom{N}{j} \cdot K_S(j)^2$$

$$\leq 2^N \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\}. \quad (10)$$

We then want to bound the contribution of the faraway terms $j \notin B$ to the summation in (9), i.e. we want to bound

$$\max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \sum_{j \notin B} \Pr_{v \sim \mu_t}[|vH| = j] K_S(j)^2 \right\}. \quad (11)$$

21

We will want to apply Theorem 9 to every term in this sum. Note that by definition of Krawtchouk polynomials, for any w, w' we have

$$K_{\{w,w'\}}(y) = K_w(y) + K_{w'}(y)$$
$$\leq 2 \cdot \max\{K_w(y), K_{w'}(y)\}.$$

We can then bound equation 11 by

$$(11) \leq \frac{1}{\binom{N}{\epsilon N - N^{3/4}}} \cdot N \max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2 \\ j \notin B}} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] K_S(j)^2 \right\}$$

$$\leq \frac{N}{\binom{N}{\epsilon N - N^{3/4}}} \max_{\substack{w \in \{\epsilon N \pm N^{3/4}\} \\ j \notin B}} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 4 K_w(j)^2 \right\}.$$

Applying Theorem 9, we get

$$(11) \leq \frac{4N}{\binom{N}{\epsilon N - N^{\frac{3}{4}}}} \max_{\substack{w \in \{\epsilon N \pm N^{\frac{3}{4}}\} \\ j \notin B}} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 2^{2h(w)N + 2w \log |1 - \frac{2j}{N}|} \right\}$$

$$\leq \frac{2^{N^{4/5}}}{N} \cdot 2^{h(\epsilon)N} \max_{j < \beta N} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\}.$$

Combining this bound for the faraway terms with our bound (10) for the central terms of the summation, we bound the right-hand side of equation (9) by

$$\Pr_{\rho \sim P_\epsilon} \left[\rho \notin D_k(H\rho^\intercal)\right] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{N}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}$$

$$+ \frac{2^{h(\epsilon)N + N^{\frac{4}{5}}}}{k^*} \max_{j < \beta N} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\}.$$

$\square$

6 The Weight Distribution of Transitive Linear Codes

We will now prove Theorem 1. We note that the bound we get is essentially tight, since for $\eta \in (0, 1)$ the repetition code

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

is transitive, has dimension ηN, and has weight distribution

$$\Pr_{c \sim \mathcal{D}(C)}\left[|c| = \alpha N\right] = q^{-\eta N} \cdot \binom{\eta N}{(1-\alpha)\eta N}(q-1)^{\alpha \eta N}$$

$$\geq q^{-\eta N} \cdot \sqrt{\frac{8\pi}{e^4 \eta N}} \cdot 2^{h(\alpha)\eta N} \cdot q^{\alpha \eta N \log_q(q-1)}$$

$$= \sqrt{\frac{8\pi}{e^4 \eta N}} \cdot q^{-(1-h_q(\alpha))\eta N}$$

for all $\alpha \in (0, 1)$. We recall and prove our Theorem 1 below:

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

$$\Pr_{c \sim \mathcal{D}(C)}\left[|c| = \alpha N\right] \leq q^{-(1-h_q(\alpha))\dim C},$$

where $\mathcal{D}(C)$ is the uniform distribution over all codewords in C, $|c|$ is the number of non-zero coordinates of c, and 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 M be the $t \times N$ generator matrix of C, and let $r = \operatorname{rank} M = \dim C$. Without loss of generality, suppose that the first r columns of M span the column-space of M. Define

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

and let $Z = (Z_1, ..., Z_N)$ be a uniformly random codeword in $C^{(\alpha)}$. Now C is transitive, so for every $j, k \in \{1, ..., N\}$ 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 $j \in \{1, ..., N\}$,

$$\Pr_{Z \sim \mathcal{D}(C^{(\alpha)})}[Z_j = 0] = 1 - \alpha. \tag{12}$$

But under condition (12), Z_j has maximal entropy when its remaining mass is uniformly distributed over $\{1, ..., q-1\}$, i.e. when $\Pr[Z_j = m] = \frac{\alpha}{q-1}$ for all $m \in \{1, ..., q-1\}$. The entropy of Z_j is thus bounded by

$$\mathsf{H}_{Z \sim \mathcal{D}(C^{(\alpha)})}(Z_j) \leq (1-\alpha)\log \frac{1}{1-\alpha} + (q-1) \cdot \frac{\alpha}{q-1}\log \frac{q-1}{\alpha}$$

$$= h_q(\alpha)\log(q). \tag{13}$$

We will now show that $\mathsf{H}(Z_j | Z_1, ..., Z_{j-1}) = 0$ for every $j > r$. To this end, fix some $j > r$. Recall that the columns $\{M_1, ..., M_r\}$ span the column-space of M, so we can

23

write the column M_j as $M_j = \sum_{k=1}^{r} \beta_k M_k$ for some $\beta_1, ..., \beta_r \in \mathbb{F}_q$. But any codeword $c \in C$ can be expressed as $v^{(c)} M$ for some $v^{(c)} \in \mathbb{F}_q^t$, 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 $\{Z_1, ..., Z_r\}$, and so we indeed have

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

for every $j > r$. Applying (13) and the chain rule for entropy then gives

$$\mathsf{H}(Z) = \mathsf{H}(Z_1) + \sum_{i=2}^{N} \mathsf{H}(Z_i | Z_1, ..., Z_{i-1})$$

$$\leq \sum_{i=1}^{r} \mathsf{H}(Z_i)$$

$$= r \cdot h_q(\alpha) \log(q)$$

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

$$\underset{c \sim \mathcal{D}(C)}{\Pr} \left[|c| = \alpha N\right] = \frac{\left|C^{(\alpha)}\right|}{q^r}$$

$$= 2^{\mathsf{H}(Z)} \cdot q^{-r}$$

$$\leq q^{-(1 - h_q(\alpha)) \cdot r}.$$

$\square$

For Reed-Muller codes, we will abuse notation and denote by $\mathcal{D}(n, d)$ the uniform distribution over all codewords in $\mathsf{RM}(n, d)$.

Theorem 23. *For any $n, d < n$, and $\alpha \in (0, 1)$, the Reed-Muller code $\mathsf{RM}(n, d)$ over the prime field $\mathbb{F}_q$ satisfies*

$$\underset{c \sim \mathcal{D}(n,d)}{\Pr} \left[|c| = \alpha N\right] \leq q^{-(1 - h_q(\alpha)) \cdot \binom{n}{\leq d}}.$$

Proof. This follows immediately from Theorem 1, Fact 12, and Fact 11. $\square$

7 List Decoding for Transitive Codes

We now turn to proving Theorem 5. Recall that in section 5 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 mentioned in the preliminaries, the dual code of a transitive code is also transitive. For any transitive linear code C, we can thus apply our Theorem 1 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 5 below.

Theorem. *Fix any $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0, 1)$, and $N > \left(\frac{5}{\epsilon}\right)^{20}$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode ϵ-errors using a list T of size*

$$|T(x)| = 2^{N^{5/6}} \cdot (2^{4\epsilon\eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})}).$$

Proof. We will show that there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size

$$|T(x)| = 2^{N^{5/6}} \cdot (2^{4\epsilon\eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})}),$$

with the property that for every codeword $c \in C$ we have

$$\Pr_{\rho \sim P_\epsilon} \left[c \notin T(c + \rho)\right] \leq \frac{2}{N}.$$

Let H denote the parity-check matrix of C. By Lemma 17, it is sufficient to show that for any $N > \left(\frac{5}{\epsilon}\right)^{20}$ and any $k > 1$ we have

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\mathsf{T})] \leq \frac{1}{N} + \frac{2^{N^{5/6}}}{Nk} \cdot (2^{4\epsilon\eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})}). \tag{14}$$

We will thus prove (14). Recall that for $k > N$, Proposition 22 yields the following bound on the left-hand side of (14):

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\mathsf{T})] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{2N}{k} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot \frac{2^N}{\binom{N}{j}} \right\}$$

$$+ \frac{2^{h(\epsilon)N + N^{\frac{4}{5}} + 1}}{k} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\}, \tag{15}$$

where $\beta = \frac{1}{2} \left(1 - 2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}\right)$ for $\tilde{\epsilon} = \epsilon + N^{-1/4}$, 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've seen in section 2, the dual code $C^\perp$ is a transitive linear code of dimension $N - \dim C$. By Theorem 1, we thus have that for all $j \in \{0, ..., N\}$,

$$\Pr_{v \sim \mu_t} [|vH| = j] \leq 2^{-(1 - h(\frac{j}{N}))(1-\eta)N}. \tag{16}$$

For any $j \in B$, we then have by Fact 13 that

$$\Pr_{v \sim \mu_t} [|vH| = j] \cdot \frac{2^N}{\binom{N}{j}} \leq 2^{-(1-h(j/N))(1-\eta)N} \cdot \frac{2^N}{\sqrt{\frac{8\pi}{e^4 N}} \cdot 2^{h(j/N)N}}$$

$$= \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{(1-h(j/N))\eta N}.$$

But for $j \in B$ we have $\beta < \frac{j}{N} < 1 - \beta$, so the right-hand side is maximized at $j = \beta N$. Applying Lemma 14, we get

$$\max_{j \in B} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot \frac{2^N}{\binom{N}{j}} \right\} \leq \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{(1-h(\beta))\eta N}$$

$$\leq \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}. \tag{17}$$

We now turn to the faraway terms of equation (15). By equation (16), we have

$$\max_{j < \beta N} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\} \leq \max_{\delta < \beta} \left\{ 2^{-(1-h(\delta))(1-\eta)N} \cdot 2^{2\epsilon N \log(1 - 2\delta)} \right\}.$$

Note that by definition of β, any $\delta < \beta$ can be written as $\delta = \frac{1 - 2\sqrt{\alpha \tilde{\epsilon}(1-\tilde{\epsilon})}}{2}$ for some $\alpha > 1$. By Lemma 14, we can then rewrite our previous expression as

$$\max_{j < \beta N} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\} \leq \max_{\alpha > 1} \left\{ 2^{-\frac{2\alpha \tilde{\epsilon}(1-\tilde{\epsilon})}{\ln 2}(1-\eta)N} \cdot 2^{\epsilon N \log(4\alpha \tilde{\epsilon}(1-\tilde{\epsilon}))} \right\}.$$

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 $\alpha = \frac{\epsilon}{2\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)}$. We then get

$$\max_{j < \beta N} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot 2^{2\epsilon N \log(1 - \frac{2j}{N})} \right\} \leq 2^{-\frac{\epsilon N}{\ln 2}} \cdot 2^{\epsilon N \log(\frac{2\epsilon}{1-\eta})}$$

$$\leq 2^{-h(\epsilon)N} \cdot 2^{\epsilon N \log(\frac{2}{1-\eta})}, \tag{18}$$

where in the last line we used the inequality $\log(1 - x) \geq -\frac{x}{(1-x)\ln 2}$ for $x < 1$ to get $h(\epsilon) \leq -\epsilon \log(\epsilon) + \frac{\epsilon}{\ln 2}$. We now use equations (17) and (18) to bound the central and faraway terms of (15) respectively. This gives

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{2N}{k} \cdot \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N} + \frac{2^{N^{\frac{4}{5}}+1}}{k} \cdot 2^{\epsilon N \log(\frac{2}{1-\eta})}$$

$$\leq \frac{1}{N} + \frac{2^{N^{5/6}}}{Nk} \cdot \left(2^{4\epsilon \eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})} \right).$$

We have shown (14), and so we are done. $\qquad\square$

26

8 List Decoding for Reed-Muller Codes

We will now turn to proving our list-decoding bounds for Reed-Muller codes. The dual code of the Reed-Muller code $\mathsf{RM}(n, d)$ is the code $\mathsf{RM}(n, n-d-1)$, so we can apply Samorodnitsky's Theorem 6 to our Proposition 22. We restate and prove our Theorem 4 below.

Theorem. *Let $\epsilon \in (0, \frac{1}{2})$ and $\gamma \in (0, 1)$ be such that $1 - \gamma \geq 2^{-\frac{2\epsilon}{(\ln 2)^2}}$. Then the Reed-Muller code $\mathsf{RM}(n, d)$ of dimension $\binom{n}{\leq d} = (1 - \gamma)N$ can with high probability list-decode ϵ-errors using a list T of size*

$$|T| = 2^{h(\epsilon)N - \gamma N + o(N)} + 2^{4\epsilon N + o(N)}.$$

Proof. We will show that there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq \mathsf{RM}(n, d)$ of size

$$|T| = 2^{h(\epsilon)N - \gamma N + o(N)} + 2^{4\epsilon N + o(N)},$$

with the property that for every codeword $c \in \mathsf{RM}(n, d)$ we have

$$\Pr_{\rho \sim P_\epsilon} \left[c \notin T(c + \rho) \right] \leq \frac{2}{N}.$$

Let H denote the parity-check matrix of $\mathsf{RM}(n, d)$. By Lemma 17, it is sufficient to show that for any $N > \left(\frac{5}{\epsilon} \right)^{20}$ and any $k > 1$ we have

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\mathsf{T})] \leq \frac{1}{N} + \frac{2^{o(N)}}{kN} \left(2^{4\epsilon N} + 2^{h(\epsilon)N - (1-\eta)N} \right). \tag{19}$$

We will thus prove (19). Recall that for $k > N$, Proposition 22 yields the following bound on the left-hand side of (19):

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\mathsf{T})] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{2N}{k} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j \right] \cdot \frac{2^N}{\binom{N}{j}} \right\}$$

$$+ \frac{2^{h(\epsilon)N + N^{\frac{4}{5}} + 1}}{k} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j \right] \cdot 2^{2\epsilon N \log |1 - \frac{2j}{N}|} \right\}, \tag{20}$$

where $\beta = \frac{1}{2} \left(1 - 2\sqrt{\tilde{\epsilon}(1 - \tilde{\epsilon})} \right)$ for $\tilde{\epsilon} = \epsilon + N^{-1/4}$, and $B = \{\beta N, ..., (1 - \beta)N\}$. Our goal is to bound every term in these sums by using the weight distribution bounds given in Theorems 1 and 6. We bound the central terms in exactly the same way as in Theorem 5: by Theorem 23 we know that the weight distribution of the Reed-Muller code satisfies

$$\Pr_{v \sim \mu_t} \left[|vH| = j \right] \leq 2^{-(1 - h(\frac{j}{N}))(1 - \eta)N},$$

so by Fact 13 we have

$$\max_{j \in B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}} \right\} \leq \max_{j \in B} \left\{ 2^{-(1-h(j/N))(1-\eta)N} \cdot \frac{2^N}{\sqrt{\frac{8\pi}{e^4 N}} \cdot 2^{h(j/N)N}} \right\}$$

$$= \max_{j \in B} \left\{ \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{(1-h(j/N))\eta N} \right\}.$$

But $B = \{\beta N, ..., (1-\beta)N\}$, so by Lemma 14 we have

$$\max_{j \in B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}} \right\} \leq \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{(1-h(\beta))\eta N}$$

$$\leq \sqrt{\frac{e^4 N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}. \tag{21}$$

For the faraway terms, we use the weight bound from Theorem 6. By symmetry, we get that

$$\max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 2^{2\epsilon N \log\left|1 - \frac{2j}{N}\right|} \right\} \leq 2^{o(N)} \cdot \max_{j \leq \frac{N}{2}} \left\{ 2^{-(1-\eta)N} \left(\frac{1}{\eta}\right)^{2j \ln 2} \cdot 2^{2\epsilon N \log\left|1 - \frac{2j}{N}\right|} \right\}$$

$$= 2^{o(N)} \cdot 2^{-(1-\eta)N} \max_{j \leq \frac{N}{2}} \left\{ 2^{-2j \ln 2 \cdot \log(\eta) + 2\epsilon N \log\left(1 - \frac{2j}{N}\right)} \right\}. \tag{22}$$

Now the function

$$g(j) = -2j \ln 2 \cdot \log(\eta) + 2\epsilon N \log\left(1 - \frac{2j}{N}\right)$$

has first derivative

$$\frac{dg}{dj} = -2 \ln 2 \cdot \log(\eta) - \frac{4\epsilon}{\ln 2 \cdot \left(1 - \frac{2j}{N}\right)},$$

and second derivative

$$\frac{dg^2}{d^2 j} = -\frac{8\epsilon}{\ln 2 \cdot N\left(1 - \frac{2j}{N}\right)^2} < 0.$$

Thus $g(j)$ achieves its maximum at $j^* = \frac{N}{2} + \frac{\epsilon N}{(\ln 2)^2 \log(\eta)}$ and is decreasing over $[j^*, \frac{N}{2}]$. Whenever $\eta > 2^{-\frac{2\epsilon}{(\ln 2)^2}}$, we have $j^* \leq 0$; in that case the argument in equation (22) is maximized at $j = 0$, and we get

$$\max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[|vH| = j\right] \cdot 2^{2\epsilon N \log\left|1 - \frac{2j}{N}\right|} \right\} \leq 2^{-(1-\eta)N + o(N)}.$$

Combining this bound for the faraway terms with the bound (21) for the central terms, we bound the right-hand side of (20) by

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\mathsf{T})] \leq \frac{1}{N} + \frac{2^{o(N)}}{kN} \left(2^{4\epsilon \eta N} + 2^{h(\epsilon)N - (1-\eta)N} \right).$$

We have shown (19), and so we are done. □

28

Acknowledgements

A Weight Bounds Comparisons

In this section, we will compare our Theorem 23 with previously known bounds on the weight distribution of Reed-Muller codes. We recall our Theorem 23 below. Note that throughout this section, $\mathcal{D}(n,d)$ will denote the uniform distribution over all codewords in $\mathsf{RM}(n,d)$, and $|c|$ will denote the number of non-zero coordinates of c.

Theorem. *For any $n, d < n$, and $\alpha \in (0,1)$, the Reed-Muller code $\mathsf{RM}(n,d)$ over the prime field $\mathbb{F}_q$ satisfies*

$$\Pr_{c \sim \mathcal{D}(n,d)} \left[|c| = \alpha N \right] \leq q^{-(1 - h_q(\alpha)) \cdot \binom{n}{\leq d}},$$

where we have defined

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

Reed-Muller codes over odd prime fields
We start with Reed-Muller codes over odd prime fields, for which the only preexisting weight bound we are aware of is the following result of [BGY20]:

Theorem 24 ([BGY20])**.** *For any $0 < \delta < \frac{1}{2}$, there are constants $c_1, c_2 > 0$ such that for any odd prime q and for any integers d, n such that $d \leq \delta n$, we have*

$$\Pr_{c \sim \mathcal{D}(n,d)} \left[\frac{|c|}{N} \leq 1 - \frac{1}{q} - q^{-c_1 \frac{n}{d}} \right] \leq q^{-c_2 \binom{n}{\leq d}}.$$

This was a generalization of [BHL12], who proved the same result for Reed-Muller codes over $\mathbb{F}_2$. Theorem 24 is very strong for small degrees, but gets weaker as the degree increases. When d is linear in n we have $q^{-c_1 \frac{n}{d}} = \Theta(1)$, meaning that in this regime Theorem 24 can only give a nontrivial bound on normalized weights that are at least a constant away from $1 - \frac{1}{q}$. Our Theorem 23 gives nontrivial bounds for all normalized weights $< 1 - \frac{1}{q}$, for all degrees $d < n$.

Reed-Muller codes over $\mathbb{F}_2$
We now turn to Reed-Muller codes over $\mathbb{F}_2$, for which more results are known. The same bound as Theorem 24 was proven over $\mathbb{F}_2$ by [BHL12]. For comparison with our Theorem 23, see the discussion above.

In the constant-rate regime (i.e. $d = \frac{n}{2} \pm O(\sqrt{n})$), the strongest known bounds for contant weights are the following two results of [Sam20]:

Theorem 25 ([Sam20])**.** *Let* $\binom{n}{\leq d} = \eta 2^n = \eta N$ *for some* $\eta \in (0,1)$*. Then for any* $\alpha \in (0, \frac{1}{2})$ *we have*

$$\Pr_{c \sim \mathcal{D}(n,d)}[|c| \leq \alpha N] \leq 2^{o(N)} \left(\frac{1}{1-\eta}\right)^{2\ln 2 \cdot \alpha N} 2^{-\eta N}.$$

This result is strong when α is away from $1/2$. For α close to $1/2$, the following bound is stronger.

Theorem 26 ([Sam20])**.** *Let* $\binom{n}{\leq d} = \eta 2^n = \eta N$ *for some* $\eta \in (0,1)$*, and define* $A = \{\frac{1-\eta^{2\ln 2}}{2}, ..., \frac{1}{2}\}$*. Then for any* $\alpha \in (0, \frac{1}{2})$*,*

$$\Pr_{c \sim \mathcal{D}(n,d)}[|c| \leq \alpha N] \leq 2^{o(N)} \cdot \begin{cases} \frac{\binom{N}{\alpha N}}{2^N} & \textit{if } \alpha \in A, \\ \frac{1}{(1-\eta^{2\ln 2})^{\alpha N}(1+\eta^{2\ln 2})^{(1-\alpha)N}} & \textit{otherwise.} \end{cases}$$

We note that the combination of Theorems 25 and 26 is stronger than our Theorem 23 whenever both the rate of the code and the normalized weight of the codeword are constant (i.e. $\alpha = \Theta(1)$ and $d = \frac{n}{2} \pm O(\sqrt{n})$).

However, when the normalized weight is subconstant or when the degree is away from $\frac{n}{2}$ (i.e. $\alpha = o(1)$ or $d = \frac{n}{2} - \Theta(n)$), the $2^{o(N)}$ term becomes too large for Theorems 25 and 26 to give a strong bound. An approach that has been fairly successful in these two regimes (subonstant rate or subconstant weight) is the line of work of [KLP12, ASW15, SS20]. To our knowledge, the strongest results for these regimes are due to [SS20]. We start with their bound for lower weights, i.e. for weights in $[0, \frac{N}{4}]$.

Theorem 27 ([SS20])**.** *For any integers* j, n, d*, we have*

$$\Pr_{c \sim \mathcal{D}(n,d)}[|c| \leq 2^{-j} \cdot 2^n] \leq 2^{-\left(1-17\left(\frac{j}{1-\frac{d}{n}} + \frac{2-\frac{d}{n}}{(1-\frac{d}{n})^2}\right)\left(\frac{d}{n}\right)^{j-1}\right)\binom{n}{\leq d}+O(n^4)}.$$

We claim that for every $d > \frac{n}{34}$, there is some weight threshold $A_d < \frac{1}{4}$ for which our Theorem 23 is stronger than Theorem 27 for all weights larger than $A_d N$. One way to see this is to note that our Theorem 23 satisfies

$$Pr[|c| \leq 2^{-j} \cdot 2^n] \leq 2^{-\left(1-h(2^{-j})\right)\binom{n}{\leq d}}$$
$$\leq 2^{-(1-2j \cdot 2^{-j})\binom{n}{\leq d}},$$

while the expression in Theorem 27 satisfies

$$2^{-\left(1-17\left(\frac{j}{1-\frac{d}{n}} + \frac{2-\frac{d}{n}}{(1-\frac{d}{n})^2}\right)\left(\frac{d}{n}\right)^{j-1}\right)\binom{n}{\leq d}} \geq 2^{-\left(1-17j\left(\frac{d}{n}\right)^{j-1}\right)\binom{n}{\leq d}}.$$

Thus our Theorem 23 is stronger than Theorem 27 whenever $j \cdot 2^{-(j-1)} < 17j \cdot \left(\frac{d}{n}\right)^{j-1}$, i.e. whenever

$$j < \frac{\log 17}{\log \frac{n}{2d}} + 1.$$

For any $d > \frac{n}{34}$, this gives a nontrivial range.

This concludes our comparison of Theorem 23 with Theorem 27, which was the bound of [SS20] for weights in $[0, \frac{N}{4}]$. We now turn to their bounds for larger weights.

Theorem 28 ([SS20]). *Let $j, n \in \mathbb{N}$ and let $0 < \gamma(n) < \frac{1}{2} - \Omega\left(\sqrt{\frac{\log n}{n}}\right)$ be a parameter (which may be constant or depend on n) such that $\frac{j + \log \frac{1}{1-2\gamma}}{(1-2\gamma)^2} = o(n)$. Then*

$$\Pr_{c \sim \mathcal{D}(n, \gamma n)}[|c| \le \frac{1 - 2^{-j}}{2} N] \le 2^{-2^{-c(\gamma, j)}\binom{n}{\le d} + O(n^4)},$$

where $c(\gamma, j) = O\left(\frac{\gamma^2 j + \gamma \log \frac{1}{1-2\gamma}}{1-2\gamma} + \gamma\right).$

This bound holds when the degree is smaller than $\frac{n}{2}$. For arbitrary degree, [SS20] gives the following:

Theorem 29 ([SS20]). *For any integers n, d and any $\delta > 0$, we have*

$$\Pr_{c \sim \mathcal{D}(n, d)}[|c| \le \frac{1 - \delta}{2} N] \le e^{-\frac{\delta^2}{2} \cdot 2^d}.$$

We will start by comparing our Theorem 23 with Theorem 29. Applying Lemma 14, we get from Theorem 23 that

$$\Pr_{c \sim \mathcal{D}(n, d)}[|c| \le \frac{1 - \delta}{2} N] \le 2^{-(1 - h(\frac{1-\delta}{2})) \cdot \binom{n}{\le d}}$$

$$\le e^{-\frac{\delta^2}{2} \cdot \binom{n}{\le d}}.$$

Thus our Theorem 23 is strictly stronger than Theorem 29 for all $d < n$. We will now compare our Theorem 23 with Theorem 28. Applying Lemma 14, we get from Theorem 23 that

$$\Pr_{c \sim \mathcal{D}(n, d)}[|c| \le \frac{1 - 2^{-j}}{2} N] \le 2^{-(1 - h(\frac{1-2^{-j}}{2})) \cdot \binom{n}{\le d}}$$

$$\le 2^{-\frac{2^{-2j}}{2 \ln 2} \cdot \binom{n}{\le d}}.$$

It follows that our Theorem 23 is stronger than Theorem 28 whenever $2^{-(2j+1)} \ge 2^{-c(\gamma, j)}$, i.e. whenever

$$2j + 1 \le c(\gamma, j).$$

But $c(\gamma, j) := O\left(\frac{\gamma^2}{1-2\gamma} \cdot j + \frac{\gamma \log \frac{1}{1-2\gamma}}{1-2\gamma} + \gamma\right)$, and $\frac{\gamma^2}{1-2\gamma} \to \infty$ as $\gamma \to 1/2$. Thus there exists some constant $\gamma^* \in (0, \frac{1}{2})$ such that our Theorem 23 is stronger than Theorem 28 whenever $d > \gamma^* n$. In private correspondence with Amir Shpilka and Ori Sberlo, we learned that γ^* can be computed to be $\gamma^* \approx 0.38$.

B Proof of Corollary 3

Recall that for any $\epsilon \in (0,1)$ we defined

$$A_\epsilon = \{\alpha N : h(\alpha) > 1 - h(\epsilon) - N^{-1/5}\},$$

and that for any code C we denote by $\mathcal{D}(C^\perp)$ the uniform distribution over the dual code $C^\perp$. We now restate and prove our Corollary 3.

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

$$\Pr_{y \sim \mathcal{D}(C^\perp)} \big[|y| = j \big] \leq \left(1 + o(N^{-1}) \right) \frac{\binom{N}{j}}{2^N},$$

and suppose that

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

Then C is resilient to ϵ-errors.

Proof. From Theorem 2, we know that there exists some decoder $d : \mathbb{F}_2^N \to C$ such that for all $c \in C$,

$$\Pr_{\rho \sim P_\epsilon}[d(c + \rho) \neq c] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + N \max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \sum_{j=0}^{N} \Pr_{c \sim C^\perp} \big[|c| = j \big] K_S(j)^2 - 1 \right\}, \quad (23)$$

where $\binom{N}{S} = \sum_{s \in S} \binom{N}{s}$ and where $K_S = \sum_{s \in S} K_s$ for K_s the Krawtchouk polynomial of degree s. Let ν be such that $h(\nu) = 1 - h(\epsilon) - N^{-1/5}$, and define the set of weights

$$A_\epsilon = \{\nu N, ..., (1-\nu)N\}.$$

We will start by bounding the central terms $j \in A_\epsilon$ in equation (23). Applying Proposition 7 and the first condition in our theorem statement, we immediately get that for any $S \subseteq \{0, ..., N\}$,

$$\frac{1}{\binom{N}{S}} \sum_{j \in A_\epsilon} \Pr_{c \sim C^\perp} \big[|c| = j \big] K_S(j)^2 \leq 1 + o\Big(\frac{1}{N}\Big). \quad (24)$$

We now turn to the faraway terms $j \notin A_\epsilon$. For these we note that by definition of Krawtchouk polynomials, for any integer s we have

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

32

For any $S \subseteq \{0, ..., N\}$, we can then bound the faraway terms $j \notin A_\epsilon$ of equation (23) by

$$\frac{1}{\binom{N}{S}} \sum_{j \notin A_\epsilon} \Pr_{c \sim C^\perp} \left[|c| = j \right] K_S(j)^2 \leq \binom{N}{S} \Pr \left[|y| \notin A_\epsilon \right].$$

Applying the second condition in our theorem statement in combination with Fact 13 and the subadditivity of entropy, we get that

$$\max_{\substack{S \subseteq \{\epsilon N \pm N^{3/4}\} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \sum_{j \notin A_\epsilon} \Pr_{c \sim C^\perp} \left[|c| = j \right] K_S(j)^2 \right\} \leq 2 \binom{N}{\epsilon N + N^{3/4}} \cdot 2 \cdot 2^{-h(\epsilon)N - N^{4/5} + N^{3/4}}$$

$$\leq 4 \cdot 2^{h(\epsilon)N + h(N^{-1/4})N} \cdot 2^{-h(\epsilon)N - N^{4/5} + N^{3/4}}$$

$$\leq o(\frac{1}{N}).$$

Combining this bound for the faraway terms with our bound (24) for the central terms, we bound equation (23) by

$$\Pr_{\rho \sim P_\epsilon} [d(c + \rho) \neq c)] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + N \cdot o(\frac{1}{N})$$

$$\leq o(1).$$

$\square$

C   Lower Bounds on List Decoding

Claim 30. *Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and consider any $N > \frac{10}{\epsilon^2}$. Suppose a code $C \subseteq \mathbb{F}_2^N$ and a decoder $d_k : \mathbb{F}_2^N \to C^{\otimes k}$ satisfy*

$$\Pr_{\substack{\rho \sim P_\epsilon \\ c \sim \mathcal{D}(C)}} [c \in d_k(c + \rho)] \geq \frac{3}{4},$$

for P_ϵ the ϵ-noisy distribution and $\mathcal{D}(C)$ the uniform distribution on C. Then we must have

$$k \geq |C| \cdot 2^{-(1-h(\epsilon))N} \cdot \frac{2^{-h(\epsilon)N^{3/4}}}{8}.$$

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

$$|\{x \in \mathbb{F}_2^N : c \in d_k(x)\}| \gtrsim 2^{h(\epsilon)N}.$$

Intuitively this is because the sphere of radius ϵN around any codeword c contains $\approx 2^{h(\epsilon)N}$ points (and for any sent codeword c, with high probability the received message

m will satisfy $|m + c| \approx \epsilon N$). 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 $|C| \cdot 2^{h(\epsilon)N}$ pairs, since as we've just argued most codewords in C need to be matched to at least $\approx 2^{h(\epsilon)N}$ points. It follows that we must have

$$k \gtrsim |C| \cdot \frac{2^{h(\epsilon)N}}{2^N}.$$

Formally, we first note that the theorem condition implies that at least $\frac{|C|}{2}$ codewords $c \in C$ must satisfy

$$\Pr_{\rho \sim P_\epsilon} [c \in d_k(c + \rho)] \geq \frac{1}{2}. \tag{25}$$

Fix any such c. Now from Chernoff's bound (i.e Lemma 16), we have for N large enough that

$$\Pr_{\rho \sim P_\epsilon} \left[|\rho| \leq \epsilon N - \epsilon N^{3/4} \right] \leq \frac{1}{4}.$$

In order for c to satisfy $c \in d_k(c + \rho)$ with probability at least $\frac{1}{2}$, there must then be a subset $S_c \subseteq \{x \in \mathbb{F}_2^N : |c + x| \geq \epsilon N - \epsilon N^{3/4}\}$ satisfying both

$$x \in S_c \implies c \in d_k(x) \tag{26}$$

and

$$\Pr_{\rho \sim P_\epsilon} [\rho \in S_c] \geq \frac{1}{4}. \tag{27}$$

But every element $x \in S_c$ satisfies $|c + x| \geq \epsilon N - \epsilon N^{3/4}$, so every $x \in S_c$ satisfies

$$\Pr_{\rho \sim P_\epsilon} [\rho = c + x] \leq \epsilon^{\epsilon N - \epsilon N^{3/4}} (1 - \epsilon)^{(1-\epsilon)N + \epsilon N^{3/4}}$$

$$\leq 2^{-(1 - N^{-1/4})h(\epsilon)N} \tag{28}$$

Equations (27) and (28) imply that any $c \in C$ that can be list-decoded by d_k with probability $\geq \frac{1}{2}$ must satisfy $|S_c| \geq \frac{2^{(1 - N^{-1/4})h(\epsilon)N}}{4}$. It then follows from (26) that any such c must satisfy

$$\left| \{x \in \mathbb{F}_2^N : c \in d_k(x)\} \right| \geq \frac{2^{(1 - N^{-1/4})h(\epsilon)N}}{4}.$$

By double counting, we get

$$2^N \cdot k = \sum_{c \in C} \left| \{x \in \mathbb{F}_2^N : c \in d_k(x)\} \right|$$

$$\geq \frac{|C|}{2} \cdot \frac{2^{(1 - N^{-1/4})h(\epsilon)N}}{4}$$

$$= \frac{|C|}{8} \cdot 2^{h(\epsilon)N - h(\epsilon)N^{3/4}}.$$

The result then follows from rearranging terms. $\qquad\square$

D   Other Proofs for Sections 1 and 2

D.1   On Known List-Decoding Bounds for Reed-Muller Codes

We recall the known list-decoding bound for Reed-Muller codes (see equation (3) in section 1):

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

We claim this bound never achieves the information-theoretic $2^{h(\epsilon)N - (N - \dim C) + o(N)}$.

Claim 31. *For any $\epsilon \in (0, \frac{1}{2})$ and any $\gamma = \gamma(\epsilon) \in (0, 1)$, we have*

$$\epsilon \log \frac{4\epsilon(1 - \epsilon)}{\gamma^4 \ln 2} > h(\epsilon) - \gamma.$$

Proof. We will show that for any $\epsilon \in (0, \frac{1}{2})$ and $c = \frac{\gamma}{\epsilon} < \frac{1}{\epsilon}$ we have

$$\epsilon \log \frac{4\epsilon(1 - \epsilon)}{(c\epsilon)^2} > h(\epsilon) - c\epsilon,$$

i.e. that

$$f(\epsilon, c) := \log(1 - \epsilon) + 2\epsilon - 2\epsilon \log c + c\epsilon > 0. \tag{29}$$

We first fix some $\epsilon \in (0, \frac{1}{2})$ and compute the c maximizing $f(\epsilon, c)$. Note that

$$\frac{\partial}{\partial c} f(\epsilon, c) = -\frac{2\epsilon}{c \ln 2} + \epsilon$$

and

$$\frac{\partial^2}{\partial c^2} f(\epsilon, c) = \frac{2\epsilon}{c^2 \ln 2} > 0,$$

so $f(\epsilon, c)$ is minimized at $c = \frac{2}{\ln 2}$ and increasing over $c \in [0, \frac{2}{\ln 2}]$. We thus have

$$\min_{c < \frac{1}{\epsilon}} f(\epsilon, c) = \begin{cases} f(\epsilon, \frac{2}{\ln 2}) & \text{if } \epsilon < \frac{\ln 2}{2}, \\ f(\epsilon, \frac{1}{\epsilon}) & \text{otherwise.} \end{cases} \tag{30}$$

We deal with each case separately. For the case $\epsilon < \frac{\ln 2}{2}$, we want to show that

$$f(\epsilon, \frac{2}{\ln 2}) = \log(1 - \epsilon) + 2\epsilon \log(\ln 2) + \frac{2\epsilon}{\ln 2} \geq 0.$$

The first derivative is

$$\frac{\partial}{\partial \epsilon} f(\epsilon, \frac{2}{\ln 2}) = -\frac{1}{(1 - \epsilon) \ln 2} + 2 \log(\ln 2) + \frac{2}{\ln 2},$$

35

and the second derivative is

$$\frac{\partial^2}{\partial \epsilon^2} f(\epsilon, \frac{2}{\ln 2}) = -\frac{1}{(1-\epsilon)^2 \ln 2} < 0.$$

Thus the function $f(\epsilon, \frac{2}{\ln 2})$ is maximized at $\epsilon^* = 1 - \frac{1}{(2\log(\ln 2) + \frac{2}{\ln 2})\ln 2} \approx 0.21$, and monotone on each side of ϵ^*. In particular, since $\epsilon^* \in [0, \frac{\ln 2}{2}]$ we know that over the interval $[0, \frac{\ln 2}{2}]$ the function $f(\epsilon, \frac{2}{\ln 2})$ achieves its minimum at either $\epsilon = 0$ or $\epsilon = \frac{\ln 2}{2}$. But $f(0, \frac{2}{\ln 2}) = 0 < f(\frac{2}{\ln 2}, \frac{\ln 2}{2})$, so we indeed have that

$$f(\epsilon, \frac{2}{\ln 2}) \geq 0$$

for all $\epsilon < \frac{\ln 2}{2}$. This deals with the first case of (30). For the second case of (30), we want to show that for all $\epsilon \in (0, \frac{1}{2})$ we have

$$f(\epsilon, \frac{1}{\epsilon}) \geq 0.$$

But

$$f(\epsilon, \frac{1}{\epsilon}) = \log(1-\epsilon) + 2\epsilon + 2\epsilon \log \epsilon + 1$$

is decreasing in ϵ and $f(\frac{1}{2}, 2) = 0$, so we indeed have $f(\epsilon, \frac{1}{\epsilon}) \geq 0$ for all ϵ. $\square$

D.2   Duals of Transitive Codes - Proof of Fact 10

Claim. *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 that for any $c = (c_1, ..., c_N) \in C$, we have $c_\pi := (c_{\pi(1)}, ..., \pi(N)) \in C$. 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

$$v \in C^\perp \implies \sum_k v_k c_{\pi(k)} = 0 \text{ for all } c \in C$$

$$\implies \sum_k v_{\pi^{-1}(k)} c_k = 0 \text{ for all } c \in C$$

$$\implies v_{\pi^{-1}} \in C^\perp.$$

$\square$

D.3 Basic Properties of Reed-Muller Codes - Proof of Facts 11 and 12

Fact. *Let M be the $\binom{n}{\leq d} \times N$ generator matrix of the Reed-Muller code. The columns of M that correspond to the points $x \in \mathbb{F}_2^n$ with $|x| \leq d$ are linearly independent.*

Proof. Let M' be the submatrix of M whose columns correspond to the points $v \in \mathbb{F}_2^n$ with $|v| \leq d$. It suffices to show that when you order the columns M'_v of M' in increasing order of $|v|$, every column is linearly independent from the preceding ones. But this is clearly the case, as for the monomial $m = \prod_{i:v_i=1} x_i$ we have $M_{m,v} = 1$ and $M_{m,v'} = 0$ for all v' preceding v. $\square$

Fact. *For all n and all $d < n$, the Reed-Muller code $\mathsf{RM}(n,d) \subseteq \mathbb{F}_2^N$ is transitive.*

Proof. Recall that we view each coordinate $i \in [N]$ as a point $v_i \in \mathbb{F}_2^n$, and that every codeword in $\mathsf{RM}(n,d)$ is the evaluation vector $\big(f(v_1),...,f(v_N)\big)$ of a polynomial f of degree $\leq d$ in n variables.

Now fix two points $v_i, v_j \in \mathbb{F}_2^n$. We want to show that there is a permutation $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that

(i) $\pi(v_i) = v_j$

(ii) If $\big(z_{v_1},...,z_{v_N}\big) \in \mathsf{RM}(n,d)$ then $\big(z_{\pi(v_1)},...,z_{\pi(v_N)}\big) \in \mathsf{RM}(n,d)$

To this end, we choose the permutation $\pi(x) = x + v_i + v_j$. Then:

(i) $\pi(v_i) = v_i + v_i + v_j = v_j$.

(ii) If $\big(z_{v_1},...,z_{v_N}\big)$ is a codeword, it can be written as $\big(z_{v_1},...,z_{v_N}\big) = \big(f(v_1),...,f(v_N)\big)$ for some polynomial f of degree $\leq d$. But then the polynomial

$g(x) = f(x + v_i + v_j)$ satisfies $\deg(g) = \deg(f) \leq d$, so $\big(g(v_1),...,g(v_N)\big)$ must be a codeword. Then since $g(x) = f \circ \pi(x)$ by definition, we have that $\big(z_{\pi(v_1)},...,z_{\pi(v_N)}\big) = \big(f \circ \pi(v_1),...,f \circ \pi(v_N)\big) = \big(g(v_1),...,g(v_N)\big) \in \mathsf{RM}(n,d)$.

$\square$

D.4 A version of Pinsker's inequality - Proof of Lemma 14

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

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

Proof.

$$1 - h(\frac{1-\mu}{2}) = 1 + \frac{1-\mu}{2}\log\left(\frac{1-\mu}{2}\right) + \frac{1+\mu}{2}\log\left(\frac{1+\mu}{2}\right)$$

$$= \frac{1-\mu}{2}\log(1-\mu) + \frac{1+\mu}{2}\log(1+\mu)$$

$$= \frac{1}{2\ln 2}\left[-(1-\mu)\sum_{i=1}^{\infty}\frac{\mu^i}{i} - (1+\mu)\sum_{i=1}^{\infty}(-1)^i\frac{\mu^i}{i}\right]$$

$$= \frac{1}{2\ln 2}\left[2\mu\sum_{i=1}^{\infty}\frac{\mu^{2i-1}}{2i-1} - 2\sum_{i=1}^{\infty}\frac{\mu^{2i}}{2i}\right]$$

$$= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\mu^{2i}\left(\frac{1}{2i-1} - \frac{1}{2i}\right)$$

$$= \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^{2i}}{i(2i-1)}$$

Thus $1 - h(\frac{1-\mu}{2}) \geq \frac{\mu^2}{2\ln 2}$ and $1 - h(\frac{1-\mu}{2}) \leq \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^2}{i(2i-1)} = \frac{1}{2\ln 2}\cdot 2\ln 2\cdot\mu^2 = \mu^2$. $\square$

References

[AHN21]   Emmanuel Abbe, Jan Hazla, and Ido Nachum. Almost-reed-muller codes achieve constant rates for random errors. *IEEE Trans. Inf. Theory*, 67(12):8034–8050, 2021.

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

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

[ASY21]   Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed-muller codes: Theory and algorithms. *IEEE Trans. Inf. Theory*, 67(6):3251–3277, 2021.

[AY19]   Emmanuel Abbe and Min Ye. Reed-muller codes polarize. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 273–286. IEEE Computer Society, 2019.

[Bar21]   Alexander Barg. Stolarsky's invariance principle for finite metric spaces. *Mathematika*, 67(1):158–186, 2021.

[BDM18]   Dmitriy Bilyk, Feng Dai, and Ryan Matzke. Stolarsky principle and energy optimization on the sphere. *Constructive Approximation*, 48, 08 2018.

[BGY20]   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.

[BHL12]   Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complex.*, 21(1):63–81, 2012.

[BK97]   Jean Bourgain and Gil Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric & Functional Analysis GAFA*, 7(3):438–461, 1997.

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

[dW08]   Ronald de Wolf. A brief introduction to fourier analysis on the boolean cube. *Theory Comput.*, 1:1–20, 2008.

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

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

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

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

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

[GX12]   Venkatesan Guruswami and Chaoping Xing. Folded codes from function field towers and improved optimal rate list decoding. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 339–350. ACM, 2012.

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

[HSS21]   Jan Hazla, Alex Samorodnitsky, and Ori Sberlo. On codes decoding a constant fraction of errors on the BSC. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1479–1488. ACM, 2021.

[IS98]   M. E. H. Ismail and P. Simeonov. Strong asymptotics for krawtchouk polynomials. *SIAM J. Comput. Appl. Math.*, 100(2):121–144, 1998.

[KKL88]   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.

[KKM+16]   Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Sasoglu, and Rüdiger L. Urbanke. Reed-muller codes achieve capacity on erasure channels. In Daniel Wichs and Yishay Mansour, editors,

Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016, pages 658–669. ACM, 2016.

[KL95]   Gil Kalai and Nathan Linial. On the distance distribution of codes. *IEEE Trans. Inf. Theory*, 41(5):1467–1472, 1995.

[KL99]   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.

[KLP12]   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.

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

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

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

[LW18]   Ray Li and Mary Wootters. Improved list-decodability of random linear binary codes. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, volume 116 of *LIPIcs*, pages 50:1–50:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

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

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

[Pol19]   Yury Polyanskiy. Hypercontractivity of spherical averages in hamming space. *SIAM J. Discret. Math.*, 33(2):731–754, 2019.

[RP21]   Galen Reeves and Henry D. Pfister. Reed-muller codes achieve capacity on BMS channels. *CoRR*, abs/2110.14631, 2021.

[Sam20]   Alex Samorodnitsky. An upper bound on l_q norms of noisy functions. *IEEE Trans. Inf. Theory*, 66(2):742–748, 2020.

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

[Skr19]   M. Skriganov. Point distributions in two-point homogeneous spaces. *Mathematika*, 65:557–587, 03 2019.

[SS20]   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.

[Tal94]   Michel Talagrand. On russo's approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.