On List Decoding Transitive Codes From Random Errors

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

January 31, 2022

Abstract

We study the error resilience of transitive linear codes over $\mathbb{F}_2$. We give tight bounds on the weight distribution of every such code C, and we show how these bounds can be used to infer bounds on the error rates that C can tolerate on the binary symmetric channel. Using this connection, we show that every transitive code can be list-decoded from random errors. As an application, our results imply list-decoding bounds for Reed-Muller codes even when the rate exceeds the channel capacity.

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]. This initiated a decades-long search for capacity-achieving codes, i.e. codes that achieve these optimal rates. 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.

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. There has recently been significant progress in understanding the performance of Reed-Muller codes over various channels, and one key realization has been that some of these results hold for *all* doubly transitive linear codes. A code is *transitive* if for every two coordinates i, j, there is a permutation π with $\pi(i) = \pi(j)$, and permuting the coordinates of each of the codewords using π does not change the code. Similarly, a code is *doubly transitive* if for $i \neq k, j \neq w$, there is a permutation as above with $\pi(i) = j, \pi(k) = w$. Reed-Muller codes are doubly transitive because applying an affine transformation to the input preserves the degree of polynomials over $\mathbb{F}_2$.

Combining double transitivity with fundamental results about the influences of 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 this paper, we show that even the weaker property of transitivity is fairly useful for binary codes. We first prove the following bound on the weight distribution of any transitive code (see section 6):

Theorem 1. *Let $C \subseteq \mathbb{F}_2^N$ be a transitive linear code. Then for any $\alpha \in (0,1)$ we have*

$$\Pr_{c \sim \mathcal{D}(C)} \left[|c| = \alpha N \right] \leq 2^{-(1-h(\alpha)) \dim C},$$

where $\mathcal{D}(C)$ is the uniform distribution over all codewords in C.

Here $h(\alpha)$ denotes the binary entropy function. We note that the bound above is stronger than many previously proved weight distribution bounds for Reed-Muller codes, even though the only feature of the code that we use is transitivity.

We also 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. A strong enough bound on the weight distribution would lead to results about unique decoding using our approach, however our weight distribution bounds are not strong enough to imply unique decoding for Reed-Muller codes. Still, our methods, combined with the above weight distribution bound for transitive codes, and the weight distribution bounds proved by [HSS21] for Reed-Muller codes, are enough to give list decoding results for these codes, which we present in Theorems 2 and 3 (see sections 7 and 8).

Theorem 2. *Fix some $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0,1)$, and $N > \left(\frac{5}{\epsilon}\right)^{20}$. Then for any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$, there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size*

$$|T(x)| = N^7 \cdot 2^{(\epsilon N + N^{3/4}) \log \frac{e^4}{1-\eta}},$$

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin T(c+\rho)\right] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}},$$

where P_ϵ is the ϵ-noisy distribution.

Theorem 3. *Fix some $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0,1)$ and $N > \left(\frac{5}{\epsilon}\right)^{20}$, and consider the Reed-Muller code $\mathsf{RM}(n,d)$ of dimension $\binom{n}{\le d} = \eta 2^n = \eta N$. 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*

$$|T(x)| = N^7 \cdot 2^{8N^{7/8}\log\frac{1}{1-\eta}} \cdot \left(2^{4\epsilon N} + 2^{(\epsilon \log\frac{\epsilon}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2)N}\right),$$

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin L(c+\rho)\right] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}}.$$

Although our lists have exponential size, for codes of dimension smaller than $(1-1000\epsilon)N$ the list size is non-trivial, in the sense that it is much smaller than the number of noise vectors or the number of codewords in the code. We present the following two corollaries as explicit examples of the bounds one gets from Theorems 2 and 3.

Corollary 4. *Let $\epsilon \in (0, \frac{1}{2})$ and $N > \left(\frac{5}{\epsilon}\right)^{20}$ be arbitrary, and let $C \subseteq \mathbb{F}_2^N$ be a transitive linear code of dimension $\dim C < (1-\epsilon^{0.99})N$. Then there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size*

$$|T(x)| = 2^{(0.99h(\epsilon)+5\epsilon)N},$$

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin T(c+\rho)\right] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}}.$$

Corollary 5. *Let $\epsilon \in (0, \frac{1}{2})$ and $N > \left(\frac{5}{\epsilon}\right)^{20}$ be arbitrary, and consider any Reed-Muller code $\mathsf{RM}(n,d)$ of dimension $\binom{n}{\le d} < (1-10\epsilon)2^n = (1-10\epsilon)N$. 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*

$$|T(x)| = 2^{(h(\epsilon)-3\epsilon+100\epsilon^2)N},$$

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin L(c+\rho)\right] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}}.$$

It can be shown that random codes $C \subseteq \mathbb{F}_2^N$ can successfully return a list of size L under errors of probability ϵ with

$$\log|C| \approx (1 - h(\epsilon))N + \log L.$$

Our bound in Corollary 5 shows that Reed-Muller codes achieve similar parameters for exponentially large L.

An important part of our analysis, which is of interest in its own right, is to understand the Fourier coefficients of the level function

$$L_S(x) = \begin{cases} 1 & |x| \in S \\ 0 & \text{otherwise,} \end{cases}$$

where S is some subset of $\{0, ..., N\}$. One can also view the Fourier coefficient $\hat{L}_{\{\epsilon N\}}(1_{\delta N})$ as the renormalized coefficient of a Krawtchouk polynomial, or as the renormalized expectation of the parity of $|X \cap Y|$, where $X \subseteq \{0, ..., N\}$ is a uniformly random subset of size ϵN and $Y \subseteq \{0, ..., N\}$ is a uniformly random subset of size δN. Using techniques from complex analysis (see for e.g. [FS09], chapter 8), we obtain the following bounds on $\hat{L}_{\{\epsilon N\}}(1_{\delta N})$ (see section 9):

Theorem 6. *For any $\epsilon, \delta \in (0, 1)$ and any integer N, we have*

$$|\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq \begin{cases} 2^{-N/2} \cdot \left(\frac{(1/2-\delta)\cdot e^2}{\epsilon}\right)^{\epsilon N} & \text{if } (1-2\delta)^2 - 4\epsilon(1-\epsilon) \geq 0, \\ 2^{(h(\epsilon)-h(\delta))N/2} & \text{otherwise.} \end{cases}$$

1.1 Techniques

Our weight distribution bound for transitive linear codes (Theorem 1) is based on a simple calculation. We show 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 rows 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 list decoding results, we make use of a connection between the decoding of a codeword and the ℓ_2 norm of a certain distribution. To explain the intuition, we 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 $Hz^\intercal$ with high probability, the code can be decoded. In the case that z is uniformly distributed on vectors of weight ϵN, this amounts to showing that for most pairs z, w of weight ϵN, $Hz^\intercal$ and $Hw^\intercal$ are distinct. This can be understood by computing the norm

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

where $f(y) = \Pr[Hz^\intercal = y^\intercal]$. The norm above is always at least $\binom{N}{\epsilon N}^{-1}$, and if $\binom{N}{\epsilon N}\|f\|_2^2$ is close to 1 then the code can be decoded with high probability. If $\binom{N}{\epsilon N}\|f\|_2^2$ is larger than 1, then we show that the code can be list-decoded with high probability, where the size of the list is related to $\binom{N}{\epsilon N}\|f\|_2^2$ (see Theorem 16 for the exact statement).

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

$$\|f\|_2^2 = 2^N \cdot \sum_{j=0}^{N} \Pr[|c^\perp| = j] \cdot \hat{L}_{\epsilon N}(1_j)^2,$$

where $c^\perp$ is a random codeword in the dual code, and $L_{\epsilon N}$ is the indicator function for strings of weight ϵN. This explains the connection between the Fourier coefficients of the level functions $L_{\epsilon N}$, the weight distribution of the code, and the probability of a decoding failure.

Our bounds on the Fourier coefficients of $L_{\epsilon N}$ (Theorem 6) are proven using ideas from complex analysis. For any $y \in \mathbb{F}_2^N$, we express $\hat{L}_{\epsilon N}(y)$ as the coefficient of $z^{\epsilon N}$ in the polynomial $(1-z)^{|y|}(1+z)^{N-|y|}$. Cauchy's residue theorem then allows us to rewrite $\hat{L}_{\epsilon N}(y)$ in terms of a contour integral around the origin of the complex plane. By choosing a well-behaved curve, we evaluate and bound this integral.

Our list-decoding results (Theorems 2 and 3) are then obtained by using the weight distribution bounds for transitive or Reed-Muller codes in conjunction with our bounds on the Fourier coefficients of the level function.

1.2 Related work

It has been shown that LDPC codes achieve capacity over Binary Memoryless Symmetric Channels (BMS) [LMS+97, KRU13], 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 $\mathbb{F}_2$. 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 (they conjecture that this variant is in fact the Reed-Muller code itself) [AY19]. 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 in [AHN20]). 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.

Several past works have proved 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 the following bound, which is stronger for weights that are linear in N [Sam20]:

Theorem 7. *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 $1/2$, results have been obtained by [BHL12, SS20], but perhaps the strongest one for Reed-Muller codes of constant rates is again due to Samorodnitsky [Sam20]:

Theorem 8. *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 $\mathsf{RM}(n,d)$. Defining $A = \{\frac{1-\eta^{2\ln 2}}{2}, ..., \frac{1}{2}\}$, we have that for any $\alpha \in (0, \frac{1}{2})$,*

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

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

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

and succeed with high probability in decoding ϵ-errors. This result, although not explicitly stated in [Sam20], can be obtained from their weight bound of Theorem 7 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 for $\eta >> 1 - \epsilon^{\frac{1}{4\ln 2}}$ and $\epsilon \in (0,1)$ small enough, (1) behaves like $2^{\epsilon N \log \frac{\epsilon}{(1-\eta)^{4\ln 2}}}$. But for $\eta >> 1-\sqrt{\epsilon}$ the list-decoding bound given by our Theorem 3 behaves like $2^{\epsilon N \log \frac{\epsilon}{(1-\eta)^2}}$, so in that regime our Theorem 3 improves (1) by an exponential factor.

2 Notation, Conventions and Preliminaries

For the sake of conciseness, we will use the notation

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

and will use the notation

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

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. For $x \in \mathbb{F}_2^N$ and $S \subseteq \{0, ...N\}$, we define the level function

$$L_S(x) = \begin{cases} 1 & |x| \in S, \\ 0 & \text{otherwise.} \end{cases}$$

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 $(N - \dim\ \mathrm{C}) \times N$ 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 A.1 for the proof).

Claim 9. *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 A.2 for the proof):

Fact 10. *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 $MH^\intercal = 0$. So, the rows of H are a basis for the orthogonal complement of the span of the rows of M. Reed-Muller codes also have useful algebraic features, notably transitivity:

Fact 11. *For all n and all $d \leq n$, the Reed-Muller code $\mathsf{RM}(n, d)$ is transitive.*

See appendix A.2 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 12. *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 13. *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 A.3 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 Fourier Analysis

The Fourier basis is a useful basis for the space of functions mapping $\mathbb{F}_2^N$ to the real numbers. For $f, g \in \mathbb{F}_2^N \to \mathbb{R}$, define the inner product

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

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

$$\chi_y(x) = \frac{(-1)^{\langle x,y \rangle}}{2^{N/2}} = \frac{(-1)^{\sum_{j=1}^N x_j y_j}}{2^{N/2}}.$$

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 resistant 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 $H\rho^\intercal$. In particular, we will be interested in the max-likelihood decoder

$$\begin{aligned} D_k(x) &= \underset{\substack{\{z_1,\dots,z_k\} \subseteq \mathbb{F}_2^N \\ H{z_i}^\intercal = x \text{ for all } i}}{\operatorname{argmax}} \{P_\epsilon(z_1) + \dots + P_\epsilon(z_k)\} \\ &= \underset{\substack{\{z_1,\dots,z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\intercal = x \text{ for all } i}}{\operatorname{argmin}} \{|z_1| + \dots + |z_k|\}. \end{aligned} \tag{2}$$

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 14. *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*

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

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

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

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

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

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

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

□

From this point onward, our goal will thus be to prove that the max-likelihood decoder in (2) 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

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

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 7 in section 8 to obtain a stronger list-decoding bound for Reed-Muller codes. Our arguments make use of some upper bounds on the Fourier coefficients of the level function, which we derive in section 9.

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.

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

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 $S = \{\epsilon N \pm l\}$, 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 collide with 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 quantity we will use to analyze the probability of $|\Omega_\rho^S|$ being large is the "collision count" $\mathsf{Coll}_H(S)$:

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

$$\mathsf{Coll}_H(S) = \binom{N}{S} \sum_{x\in\mathbb{F}_2^t} \Pr_{z\sim\lambda_S}[Hz^\intercal = x]^2,$$

where we recall the definition $\binom{N}{S} = \sum_{j\in S}\binom{N}{j}$.

The collision count of S can be seen as a measure of injectivity for the map $z \mapsto Hz^\intercal$ over the domain $\{z \in \mathbb{F}_2^N : |z| \in S\}$. When this map is injective, we have $\mathsf{Coll}_H(S) = 1$. When the map is not injective, we have $\mathsf{Coll}_H(S) > 1$, and $\mathsf{Coll}_H(S)$ increases as the number of collisions increase (i.e. it is larger when the map $z \mapsto Hz^\intercal$ is "further away" from being injective). For a uniformly random error string ρ of weight $|\rho| \in S$, we get the following relationship between the collision count $\mathsf{Coll}_\mathsf{H}(S)$ and the list size $|\Omega_\rho^S|$:

Lemma 15. *For any subset $S \subseteq \{0, ..., N\}$, any matrix H with N columns, and any integer $k > 1$, we have*

$$\Pr_{\rho\sim\lambda_S}[|\Omega_\rho^S| > k] \leq \sqrt{\frac{\mathsf{Coll}_\mathsf{H}(S)}{k}}.$$

Proof. Fix any $N \times t$ matrix H, and for any $x \in \mathbb{F}_2^t$ define

$$A_x = |\{z \in \mathbb{F}_2^N : |z| \in S, Hz^\intercal = x\}|.$$

Note that this definition is closely linked to our definition of S-colliders, since $A_{Hz^\intercal} = |\Omega_z^S|$ for every $z \in \mathbb{F}_2^N$. We can thus rewrite our quantity of interest as

$$\begin{aligned}\Pr_{\rho\sim\lambda_S}\left[|\Omega_\rho^S| > k\right] &= \Pr_{\rho\sim\lambda_S}\left[A_{H\rho^\intercal} > k\right] \\ &= \sum_{x\in\mathbb{F}_2^t : A_x > k} \Pr_{\rho\sim\lambda_S}[H\rho^\intercal = x] \\ &= \sum_{x\in\mathbb{F}_2^t : A_x > k} \frac{A_x}{\binom{N}{S}}.\end{aligned}$$

By Cauchy-Schwartz's inequality, we get

$$\begin{aligned}\Pr_{\rho\sim\lambda_S}\left[|\Omega_\rho^S| > k\right] &= \binom{N}{S}^{-1} \sum_{x\in\mathbb{F}_2^t : A_x > k} A_x \\ &\leq \sqrt{\sum_{x\in\mathbb{F}_2^t : A_x > k} \binom{N}{S}^{-2} A_x^2} \cdot \sqrt{\sum_{x\in\mathbb{F}_2^t : A_x > k} 1}.\end{aligned}$$

For the first term of the product, we note that by definition $\Pr_\rho[H\rho^\intercal = x] = \binom{N}{S}^{-1} A_x$. For the second term, we note that there can be at most $\frac{\binom{N}{S}}{k}$ vectors $x \in \mathbb{F}_2^t$ with $A_x > k$.

We can then bound the right-hand side of the previous expression by

$$\begin{aligned}
\Pr_{\rho\sim\lambda_S}[X_\rho > k] &\le \sqrt{\sum_{x\in\mathbb{F}_2^t : A_x > k} \Pr_{\rho\sim\lambda_S}[H\rho^\intercal = x]^2} \cdot \sqrt{\frac{\binom{N}{S}}{k}} \\
&\le \sqrt{\frac{\binom{N}{S}}{k} \sum_{x\in\mathbb{F}_2^t} \Pr_{\rho\sim\lambda_S}[H\rho^\intercal = x]^2} \\
&= \sqrt{\frac{\mathsf{Coll}_\mathsf{H}(S)}{k}}.
\end{aligned}$$

□

In order to obtain a list decoding result from Lemma 15, we will simply consider all weight levels near ϵN; this is done in Theorem 16. We recall that upon receiving a corrupted message $x \in \mathbb{F}_2^N$, the decoder $D_k : \mathbb{F}_2^N \to \left(\mathbb{F}_2^N\right)^{\otimes k}$ is the function that outputs the k closest codewords to x (see definition 2).

Theorem 16. *Fix $\epsilon < \frac{1}{2}$ and $l \le (\frac{1}{2} - \epsilon)N$, and let H be a matrix with N columns. For any $k > 1$, we have*

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

Proof. We will show that a slightly less performant decoding function $D_{k,l} : \mathbb{F}_2^t \to \mathbb{F}_2^N$ satisfies the desired probability bound. We define $D_{k,l}$ as follows: upon receiving input $x \in \mathbb{F}_2^t$, $D_{k,l}$ outputs $\frac{k}{2l+1}$ strings from $\{z \in \mathbb{F}_2^N : Hz = x, |z| = w\}$, for each $w \in \{\epsilon N \pm l\}$. If there are fewer than $\frac{k}{2l+1}$ strings in some level w, the decoder returns all of them. It is clear that for any l we have

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

since D_k returns the k most likely strings while $D_{k,l}$ returns at most k 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

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

we have, by Chernoff's bound, that

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

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

$$\begin{aligned}\Pr_{\rho\sim P_\epsilon}[\rho \notin D_{k,l}(H\rho^\intercal)] &\leq e^{-\frac{l^2}{3\epsilon N}} + \max_{w\in\{\epsilon N\pm l\}} \Pr_{\rho\sim\lambda_{\{w\}}}[\rho \notin D_{k,l}(H\rho^\intercal)] \\ &\leq e^{-\frac{l^2}{3\epsilon N}} + \max_{w\in\{\epsilon N\pm l\}} \Pr_{\rho\sim\lambda_{\{w\}}}[|\Omega_\rho^{\{w\}}| > \frac{k}{2l+1}].\end{aligned}$$

The theorem statement then follows from Lemma 15. □

5 Weight Bounds and Decoding

In this section, we show that strong enough bounds on the weight distribution of a dual code $C^\perp$ imply that the original linear code C is resistant to bitwise-independent errors. Our goal will be to relate the collision count $\mathsf{Coll}_H(S)$ of C's parity-check matrix H to the weight distribution of $C^\perp$, for any subset $S \subseteq \{0, ..., N\}$. Once we have a bound on $\mathsf{Coll}_H(S)$, we can then use Theorem 16 to bound the decoding error probability of the max-likelihood decoder.

The function we will need to make the link between $\mathsf{Coll}_H(S)$ and the weight distribution of $C^\perp$ is the level function of S, which as we recall is defined as

$$L_S(z) = \begin{cases} 1 & \text{if } |z| \in S, \\ 0 & \text{otherwise} \end{cases}.$$

In the following theorem, we use basic Fourier analysis tools to rewrite the collision count $\mathsf{Coll}_H(S)$ in terms of the Fourier coefficients $\{\hat{L}_S(1_j)\}$ and of the weight distribution of $C^\perp$. Recall that we use μ_t to denote the uniform distribution over all vectors in $\mathbb{F}_2^t$.

Theorem 17. *Fix $\epsilon \in (0, \frac{1}{2})$, and let H be a $N \times t$ matrix with entries in $\mathbb{F}_2$. Then for any $S \subseteq \{1, ..., N\}$, we have*

$$\mathsf{Coll}_H(S) = \frac{2^N}{\binom{N}{S}} \cdot \sum_{j=0}^{N} \Pr_{v\sim\mu_t}[|vH| = j] \cdot \hat{L}_S(1_j)^2.$$

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

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

Letting $f(x) = \Pr_{z \sim \lambda_S}[Hz^\intercal = x]$, we get

$$\begin{aligned}
\mathsf{Coll}_H(S) &= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} \Pr_{z \sim \lambda_S}[Hz^\intercal = x]^2 \\
&= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} f(x)^2 \\
&= \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2.
\end{aligned}$$

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

$$\mathsf{Coll}_H(S) = \binom{N}{S} \cdot 2^{-t} \sum_{y \in \mathbb{F}_2^t} \Big(\sum_{x \in \mathbb{F}_2^t} f(x) \cdot (-1)^{y \cdot x^\intercal} \Big)^2. \tag{3}$$

Now recall that by definition, a string $z \in \mathbb{F}_2^N$ satisfies $|z| \in S$ if and only if $L_S(z) = 1$. We can thus express $f(x)$ as

$$f(x) = \Pr_{z \sim \lambda_S}[Hz^\intercal = x] = \frac{1}{\binom{N}{S}} \sum_{\substack{z \in \mathbb{F}_2^N \\ Hz^\intercal = x}} L_S(z). \tag{4}$$

Combining expressions (3) and (4) and applying the definition of the Fourier transform, we get

$$\begin{aligned}
\mathsf{Coll}_H(S) &= \binom{N}{S} \cdot 2^{-t} \sum_{y \in \mathbb{F}_2^t} \Big(\sum_{z \in \mathbb{F}_2^N} \frac{L_S(z)}{\binom{N}{S}} \cdot (-1)^{yHz^\intercal} \Big)^2 \\
&= \frac{2^{N-t}}{\binom{N}{S}} \sum_{y \in \mathbb{F}_2^t} \hat{L}_S(yH)^2.
\end{aligned}$$

Now since $L_S(z)$ only depends on $|z|$, by definition of the Fourier transform we must have that $\hat{L}_S(u)$ is also a function of $|u|$ only. We can then rewrite the previous expression as

$$\begin{aligned}
\mathsf{Coll}_H(S) &= \frac{2^{N-t}}{\binom{N}{S}} \sum_{j=0}^{N} 2^t \cdot \Pr_{v \sim \mu_t}[|vH| = j] \cdot \hat{L}_S(1_j)^2 \\
&= \frac{2^N}{\binom{N}{S}} \cdot \sum_{j=0}^{N} \Pr_{v \sim \mu_t}[|vH| = j] \cdot \hat{L}_S(1_j)^2.
\end{aligned}$$

□

The following corollary will be very useful, as it gives an implicit bound on the Fourier coefficients $\hat{L}_S(1_j)$ of the level function:

Corollary 18. *For any N and any $S \subseteq \{1, ..., N\}$, we have*

$$\frac{1}{\binom{N}{S}} \sum_{j=0}^{N} \binom{N}{j} \cdot \hat{L}_S(1_j)^2 = 1.$$

Proof. Letting the matrix H in Theorem 17 be the $N \times N$ identity matrix I, we get

$$\begin{aligned}
\binom{N}{S}^{-1} &= \sum_{x \in \mathbb{F}_2^N} \Pr_{z \sim \lambda_S} [Iz = x]^2 \\
&= \frac{1}{\binom{N}{S}} \mathsf{Coll}_I(S) \\
&= \frac{2^N}{\binom{N}{S}^2} \cdot \sum_{j=0}^{N} \Pr[|v^\intercal I| = j] \cdot \hat{L}_S(1_j)^2 \\
&= \frac{1}{\binom{N}{S}^2} \sum_{j=0}^{N} \binom{N}{j} \cdot \hat{L}_S(1_j)^2.
\end{aligned}$$

$\square$

Given any weight distribution for a dual code $C^\perp$, we will now combine Theorems 16 and 17 to obtain an upper bound on the list size needed to decode codewords from C. For this we will need the bounds of Theorem 6 on the Fourier coefficients of the level function (see section 9 for the proof of Theorem 6).

Theorem 19. *Fix any $\epsilon \in (0, \frac{1}{2})$, define $\tilde{\epsilon} = \epsilon + N^{-\frac{1}{4}}$ and let $B = \{\beta N, ..., (1-\beta)N\}$ for $\beta = \frac{1}{2}\left(1 - 2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}\right)$. Then for all $N > \left(\frac{5}{\epsilon}\right)^{20}$ and any integer $k > 1$, we have that any $N \times t$ matrix H with entries in $\mathbb{F}_2$ satisfies*

$$\begin{aligned}
\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \max_{j \in B} \left\{ \sqrt{\Pr_{v \sim \mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}} \right\} \\
+ \sqrt{\frac{3}{k}} \cdot 2^{4N^{4/5}} \cdot \sqrt{\frac{N^{7/4}}{\binom{N}{\tilde{\epsilon}N}}} \left(\frac{e^2}{\tilde{\epsilon}}\right)^{\tilde{\epsilon}N} \max_{j \notin B} \left\{ \sqrt{\Pr_{v \sim \mu_t}[|vH| = j] \left(\frac{1}{2} - \frac{j}{N}\right)^{2\tilde{\epsilon}N}} \right\}.
\end{aligned}$$

Proof. Our general approach will be to use Theorem 16 to bound the decoding error probability in terms of the collision count $\mathsf{Coll}_H(S)$, and then use Theorem 17 to express $\mathsf{Coll}_H(S)$ in terms of the Fourier coefficients $\hat{L}_S(1_j)$ and the probability factors $\Pr_{v \sim \mu_t}\left[|vH| = j\right]$. Some of these factors will then be bounded by applying Corollary

18, and some will be bounded by applying Theorem 6. We proceed with the proof; letting $l = N^{3/4}$ in Theorem 16 and applying Theorem 17, we get

$$\begin{aligned}
&\Pr_{\rho \sim P_\epsilon}[D_k(H\rho^\intercal) \neq \rho] \\
&\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \max_{w \in \{\epsilon N \pm N^{3/4}\}} \left\{ \sqrt{\frac{(2N^{3/4}+1) \cdot \mathsf{Coll}_{\mathsf{H}}(\{w\})}{k}} \right\} \\
&\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \sqrt{\max_{w \in \{\epsilon N \pm N^{3/4}\}} \left\{ \frac{2^N}{\binom{N}{w}} \cdot \sum_{j=0}^{N} \Pr_{v \sim \mu_t}[|vH| = j] \cdot \hat{L}_{\{w\}}(1_j)^2 \right\}}. \qquad (5)
\end{aligned}$$

We will start by bounding the central terms $j \in B$ in the summation above. Applying Corollary 18, we get that for any $w \in \{\epsilon N \pm N^{3/4}\}\}$ the Fourier coefficients $\hat{L}_{\{w\}}(1_j)$ satisfy

$$\begin{aligned}
\frac{2^N}{\binom{N}{w}} \sum_{j \in B} \Pr_{v \sim \mu_t}\big[|vH| = j\big] \hat{L}_{\{w\}}(1_j)^2 &\leq \frac{2^N}{\binom{N}{w}} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\} \sum_{j \in B} \binom{N}{j} \cdot \hat{L}_{\{w\}}(1_j)^2 \\
&\leq 2^N \cdot \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\}. \qquad (6)
\end{aligned}$$

It remains to bound the contribution of the faraway terms $j \notin B$ to the summation in equation (5). For this, we use our bound on the Fourier coefficients proven in Theorem 6 to get

$$\begin{aligned}
&\max_{w \in \{\epsilon N \pm N^{\frac{3}{4}}\}} \left\{ \frac{2^N}{\binom{N}{w}} \sum_{j \notin B} \Pr_{v \sim \mu_t}\big[|vH| = j\big] \hat{L}_{\{w\}}(1_j)^2 \right\} \\
&\leq \frac{N}{\binom{N}{\epsilon N - N^{\frac{3}{4}}}} \max_{\substack{w \in \{\epsilon N \pm N^{\frac{3}{4}}\} \\ j \notin B}} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \Big(\frac{(\frac{1}{2} - \frac{j}{N})e^2}{\frac{w}{N}} \Big)^{2w} \right\}. \qquad (7)
\end{aligned}$$

Now note that by definition, if $j \notin B$ then $|\frac{1}{2} - \frac{j}{N}| \geq \sqrt{\epsilon(1-\epsilon)} > \epsilon \geq \epsilon - N^{-\frac{1}{4}}$. Thus the quotient in (7) can be bounded by $\left| \frac{\frac{1}{2} - \frac{j}{N}}{\epsilon - N^{-\frac{1}{4}}} \right| \cdot e^2 > e^2 > 1$, which then implies that $\left(\frac{(\frac{1}{2} - \frac{j}{N})e^2}{\epsilon - N^{-\frac{1}{4}}} \right)^{2w}$ is maximized at the largest possible w. We thus get

$$(7) \leq \frac{N}{\binom{N}{\epsilon N - N^{\frac{3}{4}}}} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \Big(\frac{(\frac{1}{2} - \frac{j}{N})e^2}{\epsilon - N^{-\frac{1}{4}}} \Big)^{2(\epsilon N + N^{\frac{3}{4}})} \right\}.$$

Combining this bound for the faraway terms with our bound (6) for the central terms of the summation, we bound the right-hand side of equation (5) by

$$\begin{aligned}
&\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \\
&\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \max_{j\in B}\left\{\sqrt{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}}\right\} \\
&+ \sqrt{\frac{3N^{3/4}}{k}} \cdot \sqrt{\frac{N}{\binom{N}{\epsilon N - N^{\frac{3}{4}}}} \max_{j\notin B}\left\{\Pr_{v\sim\mu_t}[|vH| = j]\Big(\frac{(\frac{1}{2}-\frac{j}{N})e^2}{\epsilon - N^{-\frac{1}{4}}}\Big)^{2\tilde{\epsilon}N}\right\}}, \qquad (8)
\end{aligned}$$

where we have defined $\tilde{\epsilon} = \epsilon + N^{-\frac{1}{4}}$. Now we note that

$$\begin{aligned}
\binom{N}{\epsilon N - N^{3/4}}^{-1} &= \binom{N}{\epsilon N + N^{3/4}}^{-1} \cdot \frac{(N - \epsilon N + N^{3/4}) \cdot \ldots \cdot (N - \epsilon N - N^{3/4} + 1)}{(\epsilon N + N^{3/4}) \cdot \ldots \cdot (\epsilon N - N^{3/4} + 1)} \\
&\leq \left(\frac{2}{\epsilon}\right)^{2N^{3/4}} \binom{N}{\tilde{\epsilon}N}^{-1},
\end{aligned}$$

and that

$$\begin{aligned}
\left(\epsilon - N^{-\frac{1}{4}}\right)^{-2\tilde{\epsilon}N} &= \left(\frac{1}{\tilde{\epsilon}}\right)^{2\tilde{\epsilon}N}\left(\frac{\epsilon + N^{-\frac{1}{4}}}{\epsilon - N^{-\frac{1}{4}}}\right)^{2\tilde{\epsilon}N} \\
&\leq \left(1 + \frac{4N^{-\frac{1}{4}}}{\epsilon}\right)^{2\tilde{\epsilon}N}\left(\frac{1}{\tilde{\epsilon}}\right)^{2\tilde{\epsilon}N} \\
&\leq 2^{\frac{8}{\ln 2}(1+\frac{N^{-\frac{1}{4}}}{\epsilon})N^{3/4}}\left(\frac{1}{\tilde{\epsilon}}\right)^{2\tilde{\epsilon}N}.
\end{aligned}$$

For $N > \left(\frac{5}{\epsilon}\right)^{20}$, equation (8) can then be bounded by

$$\begin{aligned}
\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] &\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \max_{j\in B}\left\{\sqrt{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}}\right\} \\
&+ \sqrt{\frac{3}{k}} \cdot 2^{4N^{4/5}} \cdot \sqrt{\frac{N^{7/4}}{\binom{N}{\tilde{\epsilon}N}}}\left(\frac{e^2}{\tilde{\epsilon}}\right)^{\tilde{\epsilon}N} \max_{j\notin B}\left\{\sqrt{\Pr_{v\sim\mu_t}[|vH| = j]\Big(\frac{1}{2} - \frac{j}{N}\Big)^{2\tilde{\epsilon}N}}\right\}.
\end{aligned}$$

□

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 any $\eta \in (0, 1)$ the repetition code

$$C = \left\{(z, ..., z) \in \mathbb{F}_2^N : z \in \mathbb{F}_2^{\eta N}\right\}$$

is transitive, has dimension ηN, and has weight distribution

$$\Pr_{c \sim \mathcal{D}(C)} \Big[|c| = \alpha N \Big] = 2^{-\eta N} \binom{\eta N}{\alpha \eta N}$$
$$\geq \sqrt{\frac{8\pi}{e^4 \eta N}} \cdot 2^{-(1-h(\alpha)) \dim C}$$

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

Theorem. *Let $C \subseteq \mathbb{F}_2^N$ be a transitive linear code. Then for any $\alpha \in (0,1)$ we have*

$$\Pr_{c \sim \mathcal{D}(C)} \Big[|c| = \alpha N \Big] \leq 2^{-(1-h(\alpha)) \dim C}.$$

Proof. Let M be the $t \times N$ generator matrix of C, and let $r = \text{rank } M = \dim C$. 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 $j, k \in [N]$ 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 $j \in [N]$

$$\Pr_{Z \sim \mathcal{D}(C^{(\alpha)})} [Z_j = 1] = \alpha,$$

or equivalently that the entropy of Z_j is

$$\mathsf{H}_{Z \sim \mathcal{D}(C^{(\alpha)})} (Z_j) = h(\alpha). \tag{9}$$

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 write the column M_j as $M_j = \sum_{k=1}^r \beta_k M_k$ for some $\beta_1, ..., \beta_r \in \{0,1\}$. But any codeword $c \in C$ can be expressed as $v^{(c)} M$ for some $v^{(c)} \in \mathbb{F}_2^t$, 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

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

for every $j > r$. Applying (9) and the chain rule for entropy then gives

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

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

$$\begin{aligned}\Pr_{c\sim\mathcal{D}(C)}\Big[|c| = \alpha N\Big] &= \frac{\left|C^{(\alpha)}\right|}{2^r} \\ &= 2^{\mathsf{H}(Z)-r} \\ &\leq 2^{-(1-h(\alpha))\cdot r}.\end{aligned}$$

□

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

Corollary 20. *For any $n, d < n$, and $\alpha \in (0,1)$, the Reed-Muller code $\mathsf{RM}(n,d)$ satisfies*

$$\Pr_{c\sim\mathcal{D}(n,d)}\Big[|c| = \alpha N\Big] \leq 2^{-(1-h(\alpha))\cdot\binom{n}{\leq d}}.$$

Proof. This follows immediately from Theorem 1, Fact 11, and Fact 10. □

7 List Decoding for Transitive Codes

We now turn to proving Theorem 2. 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 an exponential bound on the size of the decoding list for C. We restate and prove Theorem 2 below:

Theorem. *Fix some $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0,1)$, and $N > \left(\frac{5}{\epsilon}\right)^{20}$. Then for any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$, there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size*

$$|T(x)| = N^7 \cdot 2^{(\epsilon N + N^{3/4})\log\frac{e^4}{1-\eta}},$$

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

$$\Pr_{\rho\sim P_\epsilon}\left[c \notin T(c+\rho)\right] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}}.$$

Proof. Let H denote the parity-check matrix of C. By Lemma 14, it is sufficient to show that for any $N > \left(\frac{5}{\epsilon}\right)^{20}$ and any $k > 1$ we have

$$\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} + N^3 \cdot \sqrt{\frac{2^{(\epsilon N + N^{3/4})\log\frac{e^4}{1-\eta}}}{k}}. \tag{10}$$

We will thus prove (10). Recall that Theorem 19 yields the following bound on the left-hand side of (10):

$$\begin{aligned}
&\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \\
&\le e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \max_{j\in B}\left\{\sqrt{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}}\right\} \\
&+ \sqrt{\frac{3}{k}} \cdot 2^{4N^{4/5}} \cdot \sqrt{\frac{N^{7/4}}{\binom{N}{\tilde{\epsilon}N}}\Big(\frac{e^2}{\tilde{\epsilon}}\Big)^{\tilde{\epsilon}N}} \max_{j\notin B}\left\{\sqrt{\Pr_{v\sim\mu_t}[|vH| = j]\Big(\frac{1}{2} - \frac{j}{N}\Big)^{2\tilde{\epsilon}N}}\right\},
\end{aligned} \tag{11}$$

where $\tilde{\epsilon} = \epsilon + N^{3/4}$ and $B = \{\beta N, ..., (1-\beta)N\}$ for $\beta = \frac{1}{2} - \sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}$. 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 $N - \dim C$. We thus have by Theorem 1 that for all $j \in \{0, ..., N\}$,

$$\Pr_{v\sim\mu_t}\left[|vH| = j\right] \le 2^{-(1-h(\frac{j}{N}))(1-\eta)N}. \tag{12}$$

We recall also that for any $j \in \mathbb{N}$ and α such that $\left|\frac{1}{2} - \frac{j}{N}\right| = \sqrt{\alpha\tilde{\epsilon}(1-\tilde{\epsilon})}$, we have by Lemma 13 that

$$\frac{2\alpha\tilde{\epsilon}(1-\tilde{\epsilon})}{\ln 2} < 1 - h(j/N) < 4\alpha\tilde{\epsilon}(1-\tilde{\epsilon}). \tag{13}$$

We will use equations (12) and (13) to bound every term in (11). We start with the central terms. Fixing any $j \in B$, we have by Fact 12 and equation (12) that

$$\begin{aligned}
\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}} &\le 2^{-(1-h(j/N))(1-\eta)N} \cdot \frac{2^N}{\sqrt{\frac{8\pi}{e^4N}} \cdot 2^{h(j/N)N}} \\
&= \sqrt{\frac{e^4N}{8\pi}} \cdot 2^{(1-h(j/N))\eta N}.
\end{aligned}$$

But for $j \in B$ we have $\beta < \frac{j}{N} < 1-\beta$, so the right-hand side is maximized at $j = \beta N$. Combining this with (13), we get that

$$\begin{aligned}
\max_{j\in B}\left\{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}\right\} &\le \sqrt{\frac{e^4N}{8\pi}} \cdot 2^{(1-h(\beta))\eta N} \\
&\le \sqrt{\frac{e^4N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}.
\end{aligned} \tag{14}$$

We now turn to the faraway terms. Fix $j \notin B$ and define $\alpha > 1$ such that $\left|\frac{1}{2} - \frac{j}{N}\right| = \sqrt{\alpha\tilde{\epsilon}(1-\tilde{\epsilon})}$. By equations (12) and (13), we then have

$$\begin{aligned}
\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \left(\frac{1}{2} - \frac{j}{N}\right)^{2\tilde{\epsilon}N} &\leq 2^{-(1-h(j/N))(1-\eta)N} \cdot \left(\alpha\tilde{\epsilon}(1-\tilde{\epsilon})\right)^{\tilde{\epsilon}N} \\
&\leq 2^{-\frac{2\alpha}{\ln 2}\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)N} \cdot 2^{\tilde{\epsilon}N\log\alpha} \cdot \left(\tilde{\epsilon}(1-\tilde{\epsilon})\right)^{\tilde{\epsilon}N} \\
&= 2^{\tilde{\epsilon}N\left(\log\alpha - \frac{2\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)N}{\tilde{\epsilon}N\ln 2}\cdot\alpha\right)} \cdot \left(\tilde{\epsilon}(1-\tilde{\epsilon})\right)^{\tilde{\epsilon}N}.
\end{aligned}$$

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

$$\begin{aligned}
\max_{j\notin B}\left\{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \left(\frac{1}{2} - \frac{j}{N}\right)^{2\tilde{\epsilon}N}\right\} &\leq 2^{\tilde{\epsilon}N\left(\log\frac{\tilde{\epsilon}N}{2\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)N} - \frac{1}{\ln 2}\right)} \cdot \left(\tilde{\epsilon}(1-\tilde{\epsilon})\right)^{\tilde{\epsilon}N} \\
&= \left(\frac{\tilde{\epsilon}N}{2\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)N}\right)^{\tilde{\epsilon}N}\left(\frac{1}{e}\right)^{\tilde{\epsilon}N}\left(\tilde{\epsilon}(1-\tilde{\epsilon})\right)^{\tilde{\epsilon}N} \\
&= \left(\frac{\tilde{\epsilon}}{2e(1-\eta)}\right)^{\tilde{\epsilon}N}. \qquad (15)
\end{aligned}$$

We now use equations (14) and (15) to bound the central and faraway terms of (11) respectively. This gives

$$\begin{aligned}
\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} &+ \sqrt{\frac{3N^{3/4}}{k} \cdot \left(\frac{e^4 N}{8\pi}\right)^{1/2} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}} \\
&+ \sqrt{\frac{1}{k}} \cdot \sqrt{\frac{3N^{7/4}\cdot 2^{8N^{4/5}}}{\binom{N}{\tilde{\epsilon}N}} \cdot \left(\frac{e^2}{\tilde{\epsilon}}\right)^{2\tilde{\epsilon}N}\left(\frac{\tilde{\epsilon}}{2e(1-\eta)}\right)^{\tilde{\epsilon}N}}.
\end{aligned}$$

Using Fact 12 to bound $\binom{N}{\tilde{\epsilon}N} \geq \sqrt{\frac{8\pi}{e^4 N}} \cdot 2^{h(\tilde{\epsilon})N} \geq \sqrt{\frac{8\pi}{e^4 N}} \cdot \left(\frac{1}{\tilde{\epsilon}}\right)^{\tilde{\epsilon}N}$, we get

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

Now for $N > \left(\frac{5}{\tilde{\epsilon}}\right)^{20}$ we have $8N^{4/5} < 0.1 \cdot \tilde{\epsilon}N\log\frac{e^3}{2}$. Using this and the fact that $\sqrt{a} + \sqrt{b} \leq 2\sqrt{a+b}$ for any a, b, we get

$$\begin{aligned}
\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] &\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + 2N^2 \cdot \sqrt{\frac{2^{0.9\tilde{\epsilon}N\log\frac{e^3}{2} + \tilde{\epsilon}N\log\frac{1}{1-\eta}} + 2^{4\tilde{\epsilon}\eta N}}{k}} \\
&\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + N^3 \cdot \sqrt{\frac{2^{\tilde{\epsilon}N\log\frac{e^4}{1-\eta}}}{k}}.
\end{aligned}$$

We have shown (10), and so we are done. $\square$

8 List Decoding for Reed-Muller Codes

We will now turn to proving our Theorem 3. 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 7 to our list-decoding Theorem 19. We note without proof that for $\eta < 1 - 2\sqrt{\epsilon(1-\epsilon)}$ this would yield a $2^{o(N)}$ bound on the list size, as long as one uses the exact form for the Fourier coefficients of the level function. This result is however strictly weaker than that of [HSS21], who proved that Reed-Muller codes of rate $\eta < 1 - \log(1 + 2\sqrt{\epsilon(1-\epsilon)})$ can uniquely decode error messages of rate ϵ. For $\eta > 1 - \log(1 + 2\sqrt{\epsilon(1-\epsilon)})$ however, we get the following new exponential bounds on the list size of Reed-Muller codes:

Theorem. *Fix some $\epsilon \in (0, \frac{1}{2})$, $\eta \in (0,1)$ and $N > \left(\frac{5}{\epsilon}\right)^{20}$, and consider the Reed-Muller code $\mathsf{RM}(n,d)$ with $\binom{n}{\leq d} = \eta 2^n = \eta N$. 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*

$$|T(x)| = N^7 \cdot 2^{8N^{7/8}\log\frac{1}{1-\eta}} \cdot \left(2^{4\epsilon N} + 2^{(\epsilon\log\frac{\epsilon}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2)N}\right),$$

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin T(c+\rho)\right] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \frac{1}{\sqrt{N}}.$$

Proof. Let H denote the parity-check matrix of $\mathsf{RM}(n,d)$. By Lemma 14, it is sufficient to show that for any $N > \left(\frac{5}{\epsilon}\right)^{20}$ and any $k > 1$ we have

$$\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq e^{-\frac{\sqrt{N}}{3\epsilon}} + N^3\sqrt{2^{8N^{7/8}\log\frac{1}{1-\eta}} \cdot \frac{2^{4\epsilon N} + 2^{(\epsilon\log\frac{\epsilon}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2)N}}{k}}. \tag{16}$$

We will thus prove (16). We define $\tilde{\epsilon} = \epsilon + N^{-\frac{1}{4}}$, and we recall that Theorem 19 yields the following bound on the left-hand side of (16):

$$\begin{aligned}
&\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \\
&\leq e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}} \cdot \max_{j\in B}\left\{\sqrt{\Pr_{v\sim\mu_t}\left[|vH| = j\right] \cdot \frac{2^N}{\binom{N}{j}}}\right\} \\
&+ \sqrt{\frac{3}{k}} \cdot 2^{4N^{4/5}} \cdot \sqrt{\frac{N^{7/4}}{\binom{N}{\tilde{\epsilon}N}}\left(\frac{e^2}{\tilde{\epsilon}}\right)^{\tilde{\epsilon}N}} \max_{j\notin B}\left\{\sqrt{\Pr_{v\sim\mu_t}[|vH| = j]\left(\frac{1}{2} - \frac{j}{N}\right)^{2\tilde{\epsilon}N}}\right\},
\end{aligned} \tag{17}$$

where $B = \{\beta N, ..., (1-\beta)N\}$ for $\beta = \frac{1}{2} - \sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}$. Our goal is to bound every term in these sums by using the weight distribution bounds given in Theorems 1 and 7. We bound the central terms in exactly the same way as in Theorem 2: by Corollary 20 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 12 we have

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

But $B=\{\beta N,...,(1-\beta)N\}$, so by Lemma 13 we have

$$\max_{j\in B}\left\{\Pr_{v\sim\mu_t}\left[|vH|=j\right]\cdot\frac{2^N}{\binom{N}{j}}\right\}\leq\sqrt{\frac{e^4N}{8\pi}}\cdot 2^{(1-h(\beta))\eta N}$$
$$\leq\sqrt{\frac{e^4N}{8\pi}}\cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}. \tag{18}$$

For the faraway terms, we use the weight bound from Theorem 7. By symmetry, we have that

$$\max_{j\notin B}\left\{\Pr_{v\sim\mu_t}[|vH|=j]\cdot\left(\frac{1}{2}-\frac{j}{N}\right)^{2\tilde{\epsilon}N}\right\}\leq 2^{o(N)}\cdot\max_{j\leq\frac{N}{2}}\left\{2^{-(1-\eta)N}\left(\frac{1}{\eta}\right)^{2j\ln 2}\left(\frac{1}{2}-\frac{j}{N}\right)^{2\tilde{\epsilon}N}\right\}$$
$$=2^{o(N)}2^{-(1-\eta)N}\max_{j\leq\frac{N}{2}}\left\{2^{-2j\ln 2\cdot\log(\eta)+2\tilde{\epsilon}N\log(\frac{1}{2}-\frac{j}{N})}\right\}. \tag{19}$$

Now the function

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

has first derivative

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

and second derivative

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

Thus $g(j)$ achieves its maximum at $j=\frac{N}{2}+\frac{\tilde{\epsilon}N}{(\ln 2)^2\log(\eta)}$, and we can bound the right side of equation (19) by

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

Letting $\gamma=1-\eta$ and using the fact that $\log(1-x)\in[-\frac{x+x^2}{\ln 2},-\frac{x}{\ln 2}]$ for all $x\in[0,\frac{1}{2}]$, we get

$$\max_{j\notin B}\left\{\Pr_{v\sim\mu_t}[|vH|=j]\cdot\left(\frac{1}{2}-\frac{j}{N}\right)^{2\tilde{\epsilon}N}\right\}\leq 2^{-\gamma N}\cdot 2^{\frac{\gamma+\gamma^2}{\ln 2}\cdot N\ln 2}\cdot 2^{-\frac{2\tilde{\epsilon}N}{\ln 2}}\cdot 2^{2\tilde{\epsilon}N\log\left(\frac{\tilde{\epsilon}}{\gamma\ln 2}\right)}$$
$$\leq 2^{\left(\gamma^2-\frac{2\tilde{\epsilon}}{\ln 2}+2\tilde{\epsilon}\log\tilde{\epsilon}+2\tilde{\epsilon}\log\frac{1}{\ln 2}+2\tilde{\epsilon}\log\frac{1}{\gamma}\right)N}. \tag{20}$$

Using inequalities (18) and (20) and Lemma 13, we bound the right-hand side of (17) by

$$\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}\cdot\left(\frac{e^4N}{8\pi}\right)^{1/2}\cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}}$$
$$+\sqrt{\frac{3}{k}}\cdot 2^{4N^{4/5}}\cdot\sqrt{\frac{N^{7/4}}{\binom{N}{\tilde{\epsilon}N}}\left(\frac{e^2}{\tilde{\epsilon}}\right)^{2\tilde{\epsilon}N}2^{\left((1-\eta)^2-\frac{2\tilde{\epsilon}}{\ln 2}+2\tilde{\epsilon}\log\tilde{\epsilon}+2\tilde{\epsilon}\log\frac{1}{\ln 2}+2\tilde{\epsilon}\log\frac{1}{1-\eta}\right)N}}.$$

Now by Fact 12 we know that $\binom{N}{\tilde{\epsilon}N} \ge \sqrt{\frac{8\pi}{e^4N}}\cdot 2^{h(\tilde{\epsilon})N}$, so we get

$$\begin{aligned}
&\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)]\\
&\le e^{-\frac{\sqrt{N}}{3\epsilon}} + \sqrt{\frac{3N^{3/4}}{k}\cdot\left(\frac{e^4N}{8\pi}\right)^{1/2}\cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}}\\
&+\sqrt{\frac{3}{k}}\cdot 2^{4N^{4/5}}\cdot\sqrt{N^{7/4}\cdot\left(\frac{e^4N}{8\pi}\right)^{1/2}\cdot 2^{-h(\tilde{\epsilon})N}\cdot 2^{\frac{2\tilde{\epsilon}}{\ln 2}}\cdot 2^{\left((1-\eta)^2+2\tilde{\epsilon}\log\frac{1}{\ln 2}+2\tilde{\epsilon}\log\frac{1}{1-\eta}\right)N}}\\
&\le e^{-\frac{\sqrt{N}}{3\epsilon}} + N^2\left(\sqrt{\frac{2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}}{k}}+\sqrt{\frac{2^{(-h(\tilde{\epsilon})+2\tilde{\epsilon}\log\frac{1}{1-\eta}+4\tilde{\epsilon}+(1-\eta)^2)N}}{k}}\right)
\end{aligned}$$

But $\tilde{\epsilon} \le (1+N^{-\frac{1}{8}})\epsilon$ for all $N > \left(\frac{5}{\epsilon}\right)^{20}$ and $\sqrt{a}+\sqrt{b} \le 2\sqrt{a+b}$ for all $a, b > 0$, so we get

$$\Pr_{\rho\sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \le e^{-\frac{\sqrt{N}}{3\epsilon}} + N^3\sqrt{2^{8N^{7/8}\log\frac{1}{1-\eta}}\cdot\frac{2^{4\epsilon N}+2^{(\epsilon\log\frac{\epsilon}{(1-\eta)^2}+4\epsilon+(1-\eta)^2)N}}{k}}.$$

We have shown (16), and so we are done. □

9 Fourier Coefficients of the Level Function

In this section, we compute bounds on the Fourier coefficients of the level function. For $x \in \mathbb{F}_2^N$ and $S \subseteq \{0, ...N\}$, we recall the definition of the level function $L_S(x)$,

$$L_S(x) = \begin{cases} 1 & |x| \in S, \\ 0 & \text{otherwise.}\end{cases}$$

By definition, for any $y \in \mathbb{F}_2^N$ the Fourier coefficient $\hat{L}_S(y)$ is then

$$\begin{aligned}
\hat{L}_S(y) &= 2^{-N/2}\sum_{x\in\mathbb{F}_2^N} L_S(x)(-1)^{\langle x,y\rangle}\\
&= \frac{\binom{N}{S}}{2^{N/2}}\mathop{\mathbb{E}}_{x\sim\lambda_S}(-1)^{\langle x,y\rangle}, \qquad (21)
\end{aligned}$$

where we recall the definition $\binom{N}{S} = \sum_{j\in S}\binom{N}{j}$. We observe from (21) that the Fourier coefficients of the level function are symmetric in multiple ways. First, $|\hat{L}_S(y)| = |\hat{L}_S(y')|$ for any $S \subseteq \{0,...,N\}$ and any $y, y' \in \mathbb{F}_2^N$ such that $|y| = |y'|$. Second, since $\langle a,b\rangle + \langle a, b+1\rangle = |a|$ for any $a, b \in \mathbb{F}_2^N$, we have $|\hat{L}_{\{j\}}(y)| = |\hat{L}_{\{j\}}(y+1)|$ and $|\hat{L}_{\{j\}}(y)| = |\hat{L}_{\{N-j\}}(y)|$ for any $y \in \mathbb{F}_2^N$ and any $j \in \{0,...,N\}$. When computing bounds on the Fourier coefficients $\hat{L}_{\{j\}}(y)$, it will thus suffice to restrict our attention to the case where $j \le \frac{N}{2}$ and y is of the form $y = 1_w$ for some $w \le \frac{N}{2}$ (we recall that $1_w \in \mathbb{F}_2^N$ is the vector with ones in the first w indices and zeroes in the last $N-w$ indices). Our first bound on the Fourier coefficients of the level function is an immediate consequence of Corollary 18:

Theorem 21. *For any $S \subseteq \{0,...,N\}$ and $w \in \{0,...,N\}$, we have*

$$|\hat{L}_S(1_w)| \le \sqrt{\frac{\binom{N}{S}}{\binom{N}{w}}}.$$

Proof. Recall from Corollary 18 that we have

$$\frac{1}{\binom{N}{S}}\sum_{j=0}^{N}\binom{N}{j}\cdot \hat{L}_S(1_j)^2 = 1.$$

In particular, for every $w \in \{0,...,N\}$ we must have

$$\hat{L}_S(1_w)^2 \le \frac{\binom{N}{S}}{\binom{N}{w}}.$$

□

For our purposes, we will be most interested in the case $|S| = 1$, i.e. we will want to estimate Fourier coefficients of the form $\hat{L}_{\{\epsilon N\}}(1_{\delta N})$, with $\epsilon, \delta \in (0,1)$. In the rest of this section, we show that the bounds given on $\hat{L}_{\{\epsilon N\}}(1_{\delta N})$ by Theorem 21 can be significantly improved whenever $|\frac{1}{2} - \delta| \ge \sqrt{\epsilon(1-\epsilon)}$. The main tool we will need is a simple case of the residue theorem, which states that for any Laurent series $f(z) = \sum_{j=-\infty}^{\infty} a_j z^j$ and any integer m, we have

$$a_m = \frac{1}{2\pi i}\cdot \oint_\gamma \frac{f(z)}{z^{m+1}}\,dz, \tag{22}$$

where γ is any closed curve around the origin of the complex plane. To evaluate the explicit integral we obtain in the complex plane, we will use the so-called saddlepoint method (see for e.g. [FS09], chapter 8 for some exposition). We note that the only bounds needed in our paper concern the case $|\frac{1}{2} - \delta| \ge \sqrt{\epsilon(1-\epsilon)}$, but that for completeness we prove bounds for all regimes. We now state and prove the following strengthened version of Theorem 6:

Theorem. *For any $\epsilon, \delta \in (0, \frac{1}{2})$, we have*

$$|\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq 2^{-N/2} \cdot \left| \frac{(1-s)^{\delta}(1+s)^{(1-\delta)}}{s^{\epsilon}} \right|^{N},$$

where

$$s = \begin{cases} \frac{(1-2\delta)-\sqrt{(1-2\delta)^2-4\epsilon(1-\epsilon)}}{2(1-\epsilon)} & \text{if } \delta < \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}, \\ \frac{(1-2\delta)+i\sqrt{4\epsilon(1-\epsilon)-(1-2\delta)^2}}{2(1-\epsilon)} & \text{otherwise.} \end{cases}$$

Moreover, we have

$$|\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq \begin{cases} 2^{-N/2} \cdot \left(\frac{(1/2-\delta)\cdot e^2}{\epsilon} \right)^{\epsilon N} & \text{if } \delta < \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}, \\ 2^{(h(\epsilon)-h(\delta))N/2} & \text{otherwise.} \end{cases}$$

Proof. By definition of the Fourier transform, we have

$$\begin{aligned} \hat{L}_{\{\epsilon N\}}(1_{\delta N}) &= 2^{-N/2} \sum_{\substack{x \in \mathbb{F}_2^N \\ |x| = \epsilon N}} \prod_{j=1}^{\delta N} (-1)^{x_j} \\ &= 2^{-N/2} \cdot \text{coefficient of } z^{\epsilon N} \text{ in } (1-z)^{\delta N}(1+z)^{(1-\delta)N}. \end{aligned}$$

Applying the residue theorem (22), we then get that

$$\hat{L}_{\{\epsilon N\}}(1_{\delta N}) = \frac{2^{-N/2}}{2\pi i} \cdot \oint_{\gamma} \frac{(1-z)^{\delta N}(1+z)^{(1-\delta)N}}{z^{\epsilon N+1}} \, dz \tag{23}$$

for any curve γ around the origin of the complex plane. We now define the polar coordinates r, t to be such that $s = re^{it}$, where s is the complex number defined in the theorem statement. Letting the contour γ in equation (23) be the circle of radius r around the origin (i.e. $\gamma(\theta) = re^{i\theta}$), we get

$$\hat{L}_{\{\epsilon N\}}(1_{\delta N}) = \frac{2^{-N/2}}{2\pi} \cdot \int_{-\pi}^{\pi} \frac{(1-re^{i\theta})^{\delta N}(1+re^{i\theta})^{(1-\delta)N}}{(re^{i\theta})^{\epsilon N}} \, d\theta. \tag{24}$$

Our approach will be to bound the integrand in (24) by its maximal value over the interval $\theta \in [-\pi, \pi]$. For this, we define the magnitude

$$\begin{aligned} \tau(\theta) &= \left| (1-re^{i\theta})^{\delta N}(1+re^{i\theta})^{(1-\delta)N} \right|^2 \\ &= (1 - r\cos\theta - ir\sin\theta)^{\delta N}(1 - r\cos\theta + ir\sin\theta)^{\delta N} \\ &\cdot (1 + r\cos\theta + ir\sin\theta)^{(1-\delta)N}(1 + r\cos\theta - ir\sin\theta)^{(1-\delta)N} \\ &= (1 - 2r\cos\theta + r^2)^{\delta N}(1 + 2r\cos\theta + r^2)^{(1-\delta)N}. \end{aligned}$$

The derivative of $\tau(\theta)$ is then

$$\begin{aligned}\tau'(\theta) &= 2r\delta N \sin\theta \cdot (1-2r\cos\theta+r^2)^{\delta N-1}(1+2r\cos\theta+r^2)^{(1-\delta)N} \\ &\quad - 2r(1-\delta)N\sin\theta\cdot(1-2r\cos\theta+r^2)^{\delta N}(1+2r\cos\theta+r^2)^{(1-\delta)N-1} \\ &= 2Nr\sin\theta\cdot(1-2r\cos\theta+r^2)^{\delta N-1}(1+2r\cos\theta+r^2)^{(1-\delta)N-1} \\ &\quad\cdot\left(2r\cos\theta-(1-2\delta)(1+r^2)\right).\end{aligned}$$

We note that $1-2r\cos(\theta)+r^2 \geq (1-r)^2 > 0$ for all $\theta\in[-\pi,\pi]$. For the same reason, we have $1+2r\cos(\theta)+r^2>0$. Thus for all $\theta\in[-\pi,\pi]$, we have

$$\operatorname{sgn}(\tau'(\theta)) = \operatorname{sgn}(\sin\theta)\cdot\operatorname{sgn}(2r\cos\theta-(1-2\delta)(1+r^2)). \tag{25}$$

Case 1: $\delta < \frac{1}{2}-\sqrt{\epsilon(1-\epsilon)}$
We will rely on the following two facts, which are proven in claim 22:

$$(1-2\delta)r^2-2r+1-2\delta>0, \tag{26}$$

and

$$r=\omega\cdot\frac{2\epsilon}{1-2\delta} \tag{27}$$

for some $\omega\in[\frac{1}{2},1]$. It follows from (25) and (26) that for every θ we have

$$\operatorname{sgn}(\tau'(\theta)) = -\operatorname{sgn}(\sin\theta),$$

which implies that $\tau(\theta)$ is increasing over $[-\pi,0]$ and decreasing over $[0,\pi]$. By equation (24) and since $s=r$ when $\delta<\sqrt{\epsilon(1-\epsilon)}$, we then have

$$\begin{aligned}|\hat{L}_{\{\epsilon N\}}(1_{\delta N})| &\leq 2^{-N/2}\cdot\max_{\theta\in[-\pi,\pi]}\left|\frac{(1-re^{i\theta})^{\delta N}(1+re^{i\theta})^{(1-\delta)N}}{(re^{i\theta})^{\epsilon N}}\right| \\ &= 2^{-N/2}\cdot\left|\frac{(1-s)^{\delta N}(1+s)^{(1-\delta)N}}{s^{\epsilon N}}\right|.\end{aligned}$$

This proves our theorem's first inequality. To obtain the more explicit second inequality, we use (27) and the inequality $1+x\leq e^x$ to bound

$$\begin{aligned}\hat{L}_{\{\epsilon N\}}(1_{\delta N}) &\leq 2^{-N/2}\cdot\frac{e^{(1-2\delta)rN}}{r^{\epsilon N}} \\ &= 2^{-N/2}\cdot\left(\frac{(1-2\delta)e^{2\omega}}{2\omega\epsilon}\right)^{\epsilon N},\end{aligned}$$

for some $\omega\in[\frac{1}{2},1]$. Now the function $\nu(\omega):=\frac{e^{2\omega}}{\omega}$ has first derivative $\frac{d\nu}{d\omega}=\frac{e^{2\omega}(2\omega-1)}{\omega^2}$ and second derivative $\frac{d^2\nu}{d\omega^2}=\frac{2e^{2\omega}(2\omega^2-2\omega+1)}{\omega^3}>0$, so $\nu(\omega)$ achieves its global minimum at

$\omega = \frac{1}{2}$ and is increasing over the interval $\omega \in [\frac{1}{2}, 1]$. We can then bound our previous equation by

$$\hat{L}_{\{\epsilon N\}}(1_{\delta N}) \leq 2^{-N/2} \cdot \left(\frac{(1-2\delta)e^2}{2\epsilon} \right)^{\epsilon N}.$$

Case 2: $\delta \geq \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$

In this case, by definition we have

$$r = \sqrt{\frac{\epsilon}{1-\epsilon}} \text{ and } s = re^{it} = \sqrt{\frac{\epsilon}{1-\epsilon}} \left(\frac{1-2\delta}{2\sqrt{\epsilon(1-\epsilon)}} + \frac{\sqrt{4\epsilon(1-\epsilon) - (1-2\delta)^2}}{2\sqrt{\epsilon(1-\epsilon)}} \cdot i \right).$$

It then follows that

$$\cos t = \frac{1-2\delta}{2\sqrt{\epsilon(1-\epsilon)}}. \tag{28}$$

But from equation (25) we know that

$$\operatorname{sgn}(\tau'(\theta)) = \begin{cases} \operatorname{sgn}(\sin\theta) & \text{if } \cos\theta > \frac{(1-2\delta)(1+r^2)}{2r} \\ -\operatorname{sgn}(\sin\theta) & \cos\theta < \frac{(1-2\delta)(1+r^2)}{2r} \end{cases},$$

and so since $r = \sqrt{\frac{\epsilon}{1-\epsilon}}$ we have

$$\operatorname{sgn}(\tau'(\theta)) = \begin{cases} \operatorname{sgn}(\sin\theta) & \text{if } \cos\theta > \frac{1-2\delta}{2\sqrt{\epsilon(1-\epsilon)}}, \\ -\operatorname{sgn}(\sin\theta) & \text{if } \cos\theta < \frac{1-2\delta}{2\sqrt{\epsilon(1-\epsilon)}}. \end{cases} \tag{29}$$

It follows from (28) and (29) that $\tau(\theta)$ is increasing over $[-\pi, -t]$, decreasing over $[-t, 0]$, increasing over $[0, t]$, and decreasing over $[t, \pi]$. But $\tau(\theta)$ is clearly symmetric, so we know that $\tau(-t) = \tau(t)$. Thus $\tau(\theta)$ is maximized at $\theta = t$, and so by equation (24) we have

$$\begin{aligned} |\hat{L}_{\{\epsilon N\}}(1_{\delta N})| &\leq 2^{-N/2} \cdot \max_{\theta \in [-\pi,\pi]} \left| \frac{(1-re^{i\theta})^{\delta N}(1+re^{i\theta})^{(1-\delta)N}}{(re^{i\theta})^{\epsilon N}} \right| \\ &= 2^{-N/2} \cdot \left| \frac{(1-s)^{\delta N}(1+s)^{(1-\delta)N}}{s^{\epsilon N}} \right|. \end{aligned} \tag{30}$$

This proves our theorem's first inequality. To obtain the more explicit second inequality, we define $\alpha = \frac{(1-2\delta)^2}{4\epsilon(1-\epsilon)} < 1$ and note that we can rewrite s as $s = \sqrt{\frac{\epsilon}{1-\epsilon}} \cdot (\sqrt{\alpha} + i\sqrt{1-\alpha})$. We then compute

$$|s|^2 = \frac{\epsilon}{1-\epsilon}.$$

We also compute

$$\begin{aligned}
|1-s|^2 &= 1 + \frac{\alpha\epsilon}{1-\epsilon} - 2\sqrt{\frac{\alpha\epsilon}{1-\epsilon}} + \frac{(1-\alpha)\epsilon}{1-\epsilon} \\
&= \frac{1}{1-\epsilon} \cdot \left(1-\epsilon+\epsilon-\sqrt{4\alpha\epsilon(1-\epsilon)}\right) \\
&= \frac{2\delta}{1-\epsilon},
\end{aligned}$$

and we compute

$$\begin{aligned}
|1+s|^2 &= 1 + \frac{\alpha\epsilon}{1-\epsilon} + 2\sqrt{\frac{\alpha\epsilon}{1-\epsilon}} + \frac{(1-\alpha)\epsilon}{1-\epsilon} \\
&= \frac{1}{1-\epsilon} \cdot \left(1-\epsilon+\epsilon+\sqrt{4\alpha\epsilon(1-\epsilon)}\right) \\
&= \frac{2(1-\delta)}{1-\epsilon}.
\end{aligned}$$

From equation (30), we then have

$$\begin{aligned}
|\hat{L}_S^w| &\leq 2^{-N/2} \cdot \frac{\left(\frac{2\delta}{1-\epsilon}\right)^{\frac{\delta N}{2}} \left(\frac{2(1-\delta)}{1-\epsilon}\right)^{\frac{(1-\delta)N}{2}}}{\left(\frac{\epsilon}{1-\epsilon}\right)^{\frac{\epsilon N}{2}}} \\
&= 2^{-N/2} \cdot \left(\frac{2\delta^\delta(1-\delta)^{1-\delta}}{\epsilon^\epsilon(1-\epsilon)^{1-\epsilon}}\right)^{N/2} \\
&= 2^{(-h(\delta)+h(\epsilon))N/2}.
\end{aligned}$$

□

Claim 22. *For any $\epsilon \in (0, \frac{1}{2})$ and $\delta < \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$, define $r = \frac{(1-2\delta)-\sqrt{(1-2\delta)^2-4\epsilon(1-\epsilon)}}{2(1-\epsilon)}$. Then the following two claims hold:*

$$(1-2\delta)r^2 - 2r + 1 - 2\delta > 0, \tag{31}$$

$$r = \omega \cdot \frac{2\epsilon}{1-2\delta} \tag{32}$$

for some $\omega \in [\frac{1}{2}, 1]$.

Proof. We note that since $1 - x \leq \sqrt{1-x} \leq 1 - \frac{x}{2}$ for all $x \in [0, 1]$, we can write

$\sqrt{1-\frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}} = 1-\frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}\cdot\omega$ for some $\omega\in[\frac{1}{2},1]$. We then have

$$
\begin{aligned}
r &= \frac{(1-2\delta)-(1-2\delta)\sqrt{1-\frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}}}{2(1-\epsilon)} \\
&= \frac{1-2\delta}{2(1-\epsilon)}\left(1-\sqrt{1-\frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}}\right) \\
&= \omega\cdot\frac{2\epsilon}{1-2\delta} \qquad (33)
\end{aligned}
$$

for some $\omega\in[\frac{1}{2},1]$, which proves the first claim. Now that we have (33), in order to prove the second claim it will suffice to show that for all $\omega\in[\frac{1}{2},1]$, we have

$$\eta(\omega) := \frac{4\omega^2\epsilon^2}{1-2\delta} - \frac{4\omega\epsilon}{1-2\delta} + 1 - 2\delta > 0.$$

But $\frac{d\eta}{d\omega} = \frac{1}{1-2\delta}(8\omega\epsilon^2 - 4\epsilon)$ and $\frac{d^2\eta}{d\omega^2}$ is always positive, so $\eta(\omega)$ achieves its global minimum at $\omega=\frac{1}{2\epsilon}>1$ and is decreasing over the interval $[\frac{1}{2},1]$. Thus for any $\omega\in[\frac{1}{2},1]$ we have

$$\eta(\omega)\geq\eta(1) = \frac{1}{1-2\delta}((1-2\delta)^2 - 4\epsilon(1-\epsilon)) > 0.$$

□

Acknowledgements

We thank Noam Elkies, Madhu Sudan and Amir Yehudayoff for useful discussions.

A Proofs for section 2

A.1 Duals of Transitive Codes - Proof of Fact 9

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 $c_\pi := (c_{\pi(1),\dots,\pi(N)})\in C$. Clearly π^{-1} satisfies $\pi^{-1}(i)=j$, and we claim that it also satisfies that $v_{\pi^{-1}}\in C^\perp$ for all $v\in C^\perp$. For this we note that since $c_\pi\in C$ for every $c\in C$, we have by definition that every $v\in C^\perp$ satisfies

$$\sum_k v_k c_{\pi(k)} = 0 \text{ for all } c\in C.$$

We thus have

$$\begin{aligned} v \in C^{\perp} &\implies \sum_k v_k c_{\pi(k)} = 0 \text{ for all } c \in C \\ &\implies \sum_k v_{\pi^{-1}(k)} c_k = 0 \text{ for all } c \in C \\ &\implies v_{\pi^{-1}} \in C^{\perp}. \end{aligned}$$

□

A.2 Basic Properties of Reed-Muller Codes - Proof of Facts 10 and 11

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. □

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 $\mathsf{deg}(g) = \mathsf{deg}(f) \leq d$, 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 $\big(z_{\pi(v_1)}, ..., z_{\pi(v_N)}\big) = \big(f \circ \pi(v_1), ..., f \circ \pi(v_N)\big) = \big(g(v_1), ..., g(v_N)\big) \in \mathsf{RM}(n,d)$.

□

A.3 A version of Pinsker's inequality - Proof of Lemma 13

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

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

Proof.

$$\begin{aligned}
1 - h(\frac{1-\mu}{2}) &= 1 + \frac{1-\mu}{2}\log\left(\frac{1-\mu}{2}\right) + \frac{1+\mu}{2}\log\left(\frac{1+\mu}{2}\right) \\
&= \frac{1-\mu}{2}\log(1-\mu) + \frac{1+\mu}{2}\log(1+\mu) \\
&= \frac{1}{2\ln 2}\left[-(1-\mu)\sum_{i=1}^{\infty}\frac{\mu^i}{i} - (1+\mu)\sum_{i=1}^{\infty}(-1)^i\frac{\mu^i}{i}\right] \\
&= \frac{1}{2\ln 2}\left[2\mu\sum_{i=1}^{\infty}\frac{\mu^{2i-1}}{2i-1} - 2\sum_{i=1}^{\infty}\frac{\mu^{2i}}{2i}\right] \\
&= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\mu^{2i}\left(\frac{1}{2i-1} - \frac{1}{2i}\right) \\
&= \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{\mu^{2i}}{i(2i-1)}
\end{aligned}$$

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

References

[AHN20] Emmanuel Abbe, Jan Hazla, and Ido Nachum. Almost-reed-muller codes achieve constant rates for random errors. *CoRR*, abs/2004.09590, 2020.

[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.

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

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

[FS09] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[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.

[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.

[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.

[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.

[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.