A Fourier Analytic 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

May 10, 2022

Abstract

We present an approach to showing that a linear code is resilient to random errors. We then 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 $C \subseteq F_2^N$: $\Pr_{\mathbf{c} \in C}[|\mathbf{c}| = \alpha N] \leq 2^{-h(\alpha) \cdot \dim(C)}$.

2. We give a Fourier analytic criterion that certifies that a linear code $C$ can be decoded on the binary symmetric channel. Let $L_w(x)$ denote the level function that is 1 if $|x| = w$ and 0 otherwise, and let $C^\perp$ denote the dual code of $C$. We show that bounds on $E_{\mathbf{c} \in C^\perp} [\hat{L}_\epsilon N(\mathbf{c})]^2$ imply that $C$ recovers from errors on the binary symmetric channel with parameter $\epsilon$. Weaker bounds can be used to obtain list-decoding results using similar methods.

3. Motivated by the above results, we use complex analysis to give tight estimates for the Fourier coefficients of the level function. We then combine these estimates with our weight distribution bound to give list-decoding results for transitive linear codes and Reed-Muller codes.

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 $\epsilon$, and the Binary Erasure Channel (BEC), where each bit is independently replaced by an erasure symbol with some probability $\epsilon$. 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. A code is transitive if for every two coordinates $i, j$, there is a permutation $\pi$ with $\pi(i) = j$, and permuting the coordinates of each of the codewords using $\pi$ does not change the code. Reed-Solomon codes, BCH codes and Reed-Muller codes are all transitive. Indeed, many known constructions of codes are cyclic, and every cyclic code is 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 $\text{RM}(n, d)$ 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. Reed-Muller codes enjoy strong symmetry: the symmetry group is the group of invertible affine transformations over $\mathbb{F}_2^n$. Using fundamental results from Fourier analysis about the influences of such highly 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 this paper, we prove three kinds of results relevant to understanding the error resilience of linear codes in general, and transitive linear codes in particular.

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.

2. We give a new Fourier analysis based criterion to validate that a code can be decoded over the BSC. We use Fourier analysis in a different way from [KKM+16, HSS21]; we do not use bounds on influences. Let

$$L_w(x) = \begin{cases} 1 & |x| = w, \\ 0 & \text{otherwise} \end{cases}$$

denote the symmetric level function. Let $C^\perp$ denote the dual code, and $\hat{L}_\epsilon N(c)$ denote the Fourier coefficient of $L_\epsilon N$ corresponding to the set $c$. In spirit, our criterion says that any code satisfying

$$\mathbb{E}_{c \in C^\perp}[\hat{L}_\epsilon N(c)^2] < (1 + o(N^{-1/2})) \cdot 2^N \left(\frac{N}{\epsilon N}\right)^{-1}$$


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 $\epsilon$-errors (see Corollary 3). Moreover, if the above expectation is bounded by $o(k2^N(N\epsilon)^{-1})$, then we prove that the code can be list-decoded with a list size of about $k$.

3. Finally, motivated by our Fourier analytic criterion, we use methods from complex analysis to give tight bounds on the Fourier coefficients of the level function, which allows us to estimate the above expectation in several cases. Using known weight distribution bounds for Reed-Muller codes as well as the weight distribution bounds for transitive codes that we proved, we then obtain list-decoding results for both kinds of codes.

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 3 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

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

where we have defined 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, the bound above improves on all previously proven weight distribution bounds for Reed-Muller codes, even though the only feature of the code that we use is transitivity. See appendix A.1 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 $\epsilon$-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

$$\Pr_{\rho \sim P_\epsilon} [d(c + \rho) \neq c] \leq \Delta_\epsilon + N \max_{S \subseteq \{cN \pm N^{3/4}\}} \left\{ \frac{2N}{|S|} \mathbb{E}_{c \sim D(C^\perp)} \left[ \hat{L}_S(c)^2 \right] - 1 \right\},$$

where we defined $L_S(x)$ to be the function that is 1 if $|x| \in S$ and 0 otherwise, $\binom{N}{S}$ to be the quantity $\sum_{j \in S} \binom{N}{j}$, and $\Delta_\epsilon$ to be the quantity $\Delta_\epsilon = \left( \frac{\epsilon}{1 - \epsilon} \right)^{N^{3/4}} + e^{-\sqrt{\frac{N}{2}}}$.

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 $A_\epsilon$, the code $C$ must be resilient to $\epsilon$-errors. See appendix A.2 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

$$\Pr_{c \sim D(C^\perp)} [ |c| = j ] \leq (1 + o(N^{-1/4})) \frac{\binom{N}{j}}{2N},$$

and suppose that

$$\Pr_{c \sim D(C^\perp)} [ |c| \not\in A_\epsilon ] \leq (1 + o(N^{-1/4})) \frac{\sum_{i \not\in A_\epsilon} \binom{N}{i}}{2N}.$$

Then $C$ is resilient to $\epsilon$-errors.

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

III. List Decoding Results

Using a generalized version of Theorem 2 (namely, Theorem 19 in section 5), we obtain list decoding bounds for both transitive codes and Reed-Muller codes. We present below two explicit examples of the bounds one gets using our techniques. More general bounds, that hold for codes of any rate, are presented in Theorems 22 and 23. See sections 7 and 8 for all proofs of this subsection.

Theorem 4. Fix some $\epsilon \in (0, \frac{1}{2})$ and $N > \left( \frac{5}{\epsilon} \right)^{20}$. Then any transitive linear code of dimension $\dim C < (1 - \epsilon^{0.99})N$ can with high probability list-decode $\epsilon$-errors using a list $T$ of size

$$|T| = 2^{(0.99h(\epsilon)+5\epsilon)N}.$$
Theorem 5. Fix some $\epsilon \in (0, \frac{1}{2})$ and $N > \left(\frac{5}{\epsilon}\right)^{20}$. Then any Reed-Muller code $RM(n, d)$ of dimension $\left(\frac{n}{d}\right) \leq (1 - 10\epsilon)N$ can with high probability list-decode $\epsilon$-errors using a list $T$ of size 
\[ |T| = 2^{(h(\epsilon) - 3\epsilon + 100\epsilon^2)N}. \]

Although our lists have exponential size, for small $\epsilon$ 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 A.3) shows that any code $C \subseteq \mathbb{F}_2^N$ that can successfully list-decode errors of probability $\epsilon$ with list size $|T|$ must satisfy 
\[ |T| \gtrsim 2^{\dim C - (1 - h(\epsilon))N}. \]

Our bound in Theorem 5 shows that Reed-Muller codes achieve similar parameters. (To compare with Theorem 5, note that the lower bound above states that any code $C \subseteq \mathbb{F}_2^N$ of dimension $\left(\frac{n}{d}\right) \geq (1 - 10\epsilon)N$ needs list size $\gtrsim 2^{(h(\epsilon) - 10\epsilon)N}$).

IV. Fourier Coefficients of the Level Function

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 $\{0, ..., N\}$. For $1_{\delta N} \in \mathbb{F}_2^N$ the vector with 1s in the first $\delta N$ coordinates and 0s in the last $(1 - \delta)N$ coordinates, one can 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 $\epsilon N$ and $Y \subseteq \{0, ..., N\}$ is a uniformly random subset of size $\delta N$. 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 
\[ |\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq \sqrt{\frac{\binom{N}{\epsilon N}}{\binom{N}{\delta N}}} = 2^{(h(\epsilon) - h(\delta))\frac{N}{\epsilon}} + o(N). \]

Additionally, we have 
\[ |\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq \begin{cases} 2^{-N/2} \cdot \left(\frac{(1/2 - \delta)^2}{\epsilon}\right)^{\epsilon N} & \text{if } (1 - 2\delta)^2 - 4\epsilon(1 - \epsilon) \geq 0, \\ 2^{(h(\epsilon) - h(\delta))\frac{N}{\epsilon}} & \text{otherwise}. \end{cases} \]
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 $\alpha N$ 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 Fourier analytic criterion, we make use of a connection between the decoding of a codeword and the $\ell_2$ norm of a certain distribution. To explain the intuition, we start by assuming that exactly $\epsilon N$ 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 $Hz^\top$ with high probability, the codeword can be decoded. In the case that $z$ is uniformly distributed on vectors of weight $\epsilon N$, this amounts to showing that for most pairs $z, w$ of weight $\epsilon N$, $Hz^\top$ and $Hw^\top$ are distinct. This can be understood by computing the norm

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

where $f(y) = \Pr[Hz^\top = y^\top]$. 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

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

where $c^\perp$ is a random codeword in the dual code, and $L_{\epsilon N}$ is the indicator function for strings of weight $\epsilon N$. This explains the connection between the Fourier coefficients of the level functions $L_{\epsilon N}$, the weight distribution of the dual 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 $\check{L}_{\epsilon N}(y)$ as the coefficient of $z^{|y|} \check{L}_{\epsilon N}(1_j)$ in the polynomial $(1 - z)^{|y|} (1 + z)^{N-|y|}$. Cauchy’s residue theorem then allows us to rewrite $\check{L}_{\epsilon N}(y)$ 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 4 and 5) are then obtained by using 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) \[LMS^{+}97, KRU13, Gal62\], 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 Shberlo 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 \[AHN20\]. 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.

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 Shberlo 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 7** \([Sam20]\). 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 \(\text{RM}(n, d)\). 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 \(\alpha\) 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 the rate of the code and the normalized weight of the codeword are \(\Theta(1)\) \[Sam20\].

We note that in some regimes (for e.g. when the degree satisfies \(0.38n < d < 0.499n\) and the normalized weight \(\alpha\) is larger than some constant), our Theorem 1 improves on all the aforementioned weight bounds. See appendix \(\ref{app:weird} \) 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 \(\text{RM}(n,d)\) 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

\[
|T| = 2^{\left(\epsilon \log \frac{\epsilon}{(1-\eta)\log(2-2\epsilon)}\right)N}
\]

and succeed with high probability in decoding \(\epsilon\)-errors. This result, although not explicitly stated in \[Sam20\], can be obtained from his 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 rates beyond capacity (i.e. \(\eta > 1 - h(\epsilon)\)), the list size in (1) is trivial (it is larger than \(\binom{N}{\epsilon N} \approx 2^{h(\epsilon)N}\), which is the total number of noise vectors). To our knowledge, our Theorem 5 is the first nontrivial list-decoding bound for Reed-Muller codes in the regime where the rate is larger than the channel capacity.
2 Notation, Conventions and Preliminaries

For the sake of conciseness, we will use the notation

\[
\binom{n}{d} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d},
\]

and will use the notation

\[
\{a \pm l\} = \{a - l, \ldots, a + l\}.
\]

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 \( S \subseteq \{0, \ldots, N\} \), 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 \((N - \dim C) \times N\) matrix \( H \) whose rows span the orthogonal complement of \( C \). The linear code \( C \) can then be expressed as

\[
C = \{c \in \mathbb{F}_2^N : Hc^\top = 0\}.
\]

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 \( v = (v_1, \ldots, v_N) \in C \) we have \( (v_{\pi(1)}, \ldots, v_{\pi(N)}) \in C \)

We note that the dual code of a transitive code is also transitive (see appendix B.1 for the proof).

Claim 8. 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 $\text{RM}(n, d)$ the Reed-Muller code with $n$ variables and degree $d$. Throughout this section, we let $M$ be the generator matrix of $\text{RM}(n, d)$; 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, $(n \leq d)$. In fact we have the following standard fact (see appendix B.2 for the proof):

**Fact 9.** 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 $\text{RM}(n, n-d-1)$. Then $H$ has full rank, and $MH^\top = 0$. 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 10.** For all $n$ and all $d \leq n$, the Reed-Muller code $\text{RM}(n, d)$ is transitive.

See appendix B.2 for the proof.

2.3 Entropy

The binary entropy function $h : [0, 1] \rightarrow \mathbb{R}$ 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 11.** For $n/2 \geq d \geq 1$, $\sqrt{\frac{8\pi}{e^{2n}}} \cdot 2^{h(d/n) \cdot n} \leq \binom{n}{d} \leq 2^{h(d/n) \cdot n}$.

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 12.** 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 B.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 $\epsilon$-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 $f, g : \mathbb{F}_2^N \rightarrow \mathbb{R}$, 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 $f, g : \mathbb{F}_2^N \rightarrow \mathbb{R}$, 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 \( \epsilon \). 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 \( d_N : \mathbb{F}_2^N \to (\mathbb{F}_2^N)^\otimes k \), 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 \( H\rho^\top \). In particular, we will be interested in the max-likelihood decoder

\[
D_k(x) = \arg\max_{\{ z_1, \ldots, z_k \} \subseteq \mathbb{F}_2^N} \{ P_i(z_1) + \ldots + P_i(z_k) \}
\]

\[
\text{subject to } H_{zi}^\top = x \text{ for all } i.
\]

\[
= \arg\min_{\{ z_1, \ldots, z_k \} \subseteq \mathbb{F}_2^N} \{ |z_1| + \ldots + |z_k| \}. \quad (2)
\]

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

**Lemma 13.** Let \( H \) be the \( t \times N \) parity-check matrix of the linear code \( C \), and let \( D : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^\otimes k \) be an arbitrary function. Then there exists a decoding function

\[
d : \mathbb{F}_2^N \to (\mathbb{F}_2^N)^\otimes k
\]

such that for every \( c \in C \) we have

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

**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

\[
d(z) = \{ z + y : y \in D(Hz^\top) \}.
\]

We will show that whenever \( \rho \) satisfies \( \rho \in D(H\rho^\top) \), \( \rho \) also satisfies \( c \in d(c + \rho) \) for every \( c \in C \). Suppose \( \rho \in D(H\rho^\top) \). Note that since \( H \) is the parity-check matrix of \( C \), every \( c \in C \) satisfies \( Hc^\top = 0 \). So for every \( c \in C \), any \( \rho \) that satisfies \( \rho \in D(H\rho^\top) \) must also satisfy \( \rho \in D(H(c^\top + \rho^\top)) \). 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 (2) succeeds in recovering $\rho$ with high probability. In section 4, we relate the decoding error probability of the max-likelihood decoder $D_k$ to the collision probability

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

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 $P_\epsilon$ the $\epsilon$-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 $\lambda_S$ 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 $z \mapsto Hz^\top$. 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 $H\rho^\top \in \mathbb{F}_2^n$. 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 $\epsilon$-noisy error string $\rho$, then $\rho$ is unlikely to have weight $|\rho|$ far away from $\epsilon N$. 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 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

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

13
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 $H\rho^t$ and are told that with high probability the error string $\rho$ has weight $|\rho| \in S$, then we should output the list $\Omega^S_\rho$. For unique decoding we want to argue that $|\Omega^S_\rho| = 1$ with high probability, whereas for list decoding we want to argue that $|\Omega^S_\rho| \leq k$ with high probability, for some integer $k > 1$. The function we will use to analyze the probability of $|\Omega^S_\rho|$ being large is the ”collision count” $\text{Coll}_H(S)$:

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

$$\text{Coll}_H(S) = \binom{N}{S} \sum_{x \in \mathbb{F}_t^2} \Pr \rho \sim \lambda_S [Hz^T = x]^2,$$

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 $z \mapsto Hz^T$ over the domain $\{z \in \mathbb{F}_t^N : |z| \in S\}$. When this map is injective, we have $\text{Coll}_H(S) = 1$. When the map is not injective, we have $\text{Coll}_H(S) > 1$, and $\text{Coll}_H(S)$ increases as the number of collisions increase (i.e. it is larger when the map $z \mapsto Hz^T$ is ”further away” from being injective). For a uniformly random error string $\rho$ of weight $|\rho| \in S$, we have the following relationship between the collision count $\text{Coll}_H(S)$ and the list size $|\Omega^S_\rho|$:

**Lemma 14.** For any subset $S \subseteq \{0, \ldots, N\}$, any matrix $H$ with $N$ columns, and any integer $k > 1$, we have

$$\Pr_{\rho \sim \lambda_S} [\Omega^S_\rho > k] \leq \frac{\text{Coll}_H(S) - 1}{k}.$$

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

$$H^{-1}(x) = \{z \in \mathbb{F}_2^N : |z| \in S, Hz^T = x\}.$$

Now by definition, the collision count can be expressed as

$$\text{Coll}_H(S) = \binom{N}{S} \sum_{x \in \mathbb{F}_t^2} \Pr_{\rho \sim \lambda_S} [H\rho^T = x]^2$$

$$= \binom{N}{S} \sum_{x \in \mathbb{F}_t^2} \Pr_{\rho \sim \lambda_S} [H\rho^T = x] \cdot \frac{|H^{-1}(x)|}{\binom{N}{S}}$$

$$= \sum_{x \in \mathbb{F}_t^2} |H^{-1}(x)| \cdot \Pr_{\rho \sim \lambda_S} [H\rho^T = x].$$

We now break the sum up into two: the strings $x \in \mathbb{F}_2^t$ whose preimages have size smaller than $k$, and the strings $x \in \mathbb{F}_2^t$ whose preimages have size greater than $k$. We
get

\[ \text{Coll}_{H}(S) \geq \sum_{x \in \mathbb{F}_2^N : |H^{-1}(x)| > k} (k + 1) \Pr_{\rho \sim \lambda_S} [H \rho^\top = x] + \sum_{x \in \mathbb{F}_2^N : |H^{-1}(x)| \leq k} \Pr_{\rho \sim \lambda_S} [H \rho^\top = x] \]

\[ = (k + 1) \Pr_{\rho \sim \lambda_S} [\Omega^S_\rho > k] + \left( 1 - \Pr_{\rho \sim \lambda_S} [\Omega^S_\rho > k] \right) \]

\[ = k \cdot \Pr_{\rho \sim \lambda_S} [\Omega^S_\rho > k] + 1. \]

The theorem statement then follows from rearranging terms. \( \square \)

When the error string \( \rho \) is sampled uniformly at random from the set \( \{ z \in \mathbb{F}_2^N : |z| \in S \} \), the above theorem allows us to relate the decoding error probability to the Collision Count \( \text{Coll}_{H}(S) \). The problem we are most interested in, however, is when \( \rho \) is sampled not from some uniform distribution, but from the \( \epsilon \)-noisy probability distribution \( P_\epsilon \). In the rest of this section, we will show how to connect these two decoding problems. More precisely, we will show that the max-likelihood decoder that operates under the distribution \( P_\epsilon \) is essentially identical to the max-likelihood decoder that operates under the uniform distribution \( \lambda_S \) over strings of weight \( |z| \in S \), where \( S \) is some interval centered around \( \epsilon N \) (i.e. \( S = \{ \epsilon N - l, ..., \epsilon N + l \} \) for some integer \( l \)). We first introduce the following definition, which gives us a success/failure criterion for the decoding of an error string \( \rho \) under the max-likelihood \( P_\epsilon \)-decoder:

**Definition 4.** For any \( z \in \mathbb{F}_2^N \), define its set \( \Omega_z \) of lower-weight colliders,

\[ \Omega_z = \Omega_{\{0, ..., |z|\}}, \]

where for \( A \subseteq \{0, ..., N\} \) we recall the definition \( \Omega^A = \{ y \in \mathbb{F}_2^N : |y| \in A \text{ and } Hy^\top = Hz^\top \} \).

An error string \( \rho \in \mathbb{F}_2^N \) will be successfully decoded by the \( P_\epsilon \)-decoder if and only if \( \rho \) has no lower-weight colliders, i.e. \( |\Omega_\rho| = 1 \). This is because lower-weight strings have higher probability under the distribution \( P_\epsilon \) than higher-weight strings, so the max-likelihood \( P_\epsilon \)-decoder will output the lowest-weight valid candidate. For comparison, recall that the \( \lambda_S \)-decoder succeeds in decoding the string \( \rho \) if and only if \( |\Omega^S_\rho| = 1 \). We thus want to show that with high probability, if a string \( \rho \) has no \( S \)-colliders (for some \( S = \{ \epsilon N - l, ..., \epsilon N + l \} \) then the string \( \rho \) has no lower-weight colliders. First note that we need not worry about error strings \( \rho \notin S \), as by Chernoff bound these are very unlikely to occur. Our real problematic event thus occurs when some error string \( \rho \) with weight \( |\rho| \approx \epsilon N \) has the same image under the mapping \( z \mapsto Hz^\top \) as some string \( \varphi \) of weight \( |\varphi| << \epsilon N \). In order to look into this issue, we introduce the following definition:

**Definition 5.** For any \( z \in \mathbb{F}_2^N \) and any matrix \( H \) with \( N \) columns, define \( z \)’s neighbor

\[ v_z^H = \arg\max_{y \in \Omega_z \setminus \{z\}} \{|y|\}. \]

When the matrix \( H \) is clear from context, we will drop the superscript and write \( v_z \).
Intuitively, we will see that the only error strings that are problematic for the \( P_e \)-decoder but not for the \( \lambda_S \)-decoder are the strings \( \rho \) whose neighbor \( v_\rho \) satisfies \(|v_\rho| \notin S\). Indeed, assuming that \(|\rho| \in S\) there are three cases: If the error string \( \rho \) has no neighbor, then there is no string \( \varphi \neq \rho \) of weight \(|\varphi| \leq |\rho|\) that satisfies \( H\varphi^T = H\rho^T\); in this case the \( P_e \)-decoder succeeds, since the string \( \rho \) has maximal probability within the set \( \{ z \in \mathbb{F}_2^N : Hz^T = H\rho^T \} \). If \( \rho \) has a neighbor \( v_\rho \) and \( v_\rho \) has weight \(|v_\rho| \in S\), then both decoders fail: in this case we have both \( P_e(v_\rho) \geq P_e(\rho) \) and \( \lambda_S(v_\rho) = \lambda_S(\rho) \), so neither decoder can uniquely decode \( \rho \). It is only when \( \rho \) has a neighbor and that neighbor satisfies \(|v_\rho| \notin S\) that the \( P_e \)-decoder can become worse than the \( \lambda_S \)-decoder: in this case the \( P_e \) max-likelihood decoder fails, since \( P_e(v_\rho) > P_e(\rho) \). On the other hand the \( \lambda_S \) max-likelihood decoder might succeed, since there are no strings \( \varphi \in \Omega^S_p \) satisfying \(|\varphi| < |\rho|\) (there might still be a string \( \varphi \in \Omega^S_p \) satisfying \(|\varphi| > |\rho|\), but we have no such guarantee). We thus want to make sure that the error string \( \rho \) is very unlikely to have a neighbor \( v_\rho \) with weight \(|v_\rho| << |\rho|\). This is proven in the following Lemma:

**Lemma 15.** Fix \( \epsilon > 0 \) and \( l \geq 1 \), and let \( H \) be a matrix with \( N \) columns. For the set

\[
A = \{ z \in \mathbb{F}_2^N : |v_z| < |z| - l \},
\]

we have

\[
\Pr_{\rho \sim P_e} [\rho \in A] \leq \left( \frac{\epsilon}{1 - \epsilon} \right)^l.
\]

**Proof.** Note that since every \( z \in A \) satisfies \(|v_z| < |z| - l\), every \( z \in A \) must satisfy

\[
\Pr_{\rho \sim P_e} [\rho = z] < \left( \frac{\epsilon}{1 - \epsilon} \right)^l \Pr_{\rho \sim P_e} [\rho = v_z].
\] (3)

We now claim that the function \( z \mapsto v_z \) is injective over \( A \). Suppose \( z, y \in A \) satisfy \( v_z = v_y \), and WLOG suppose that \(|y| \leq |z|\). Since \( v_z = v_y \) we must have \( Hz^T = Hy^T \), and so since \(|y| \leq |z|\) we have \( y \in \Omega_z \). By definition of \( v_z \) we must then have \(|y| \leq |v_z| = |v_y|\). But this is impossible, since by definition of \( A \) we know \(|v_y| < |y|\). The function \( z \mapsto v_z \) is thus injective over \( A \), and so we can bound

\[
\sum_{z \in A} \Pr_{\rho \sim P_e} [\rho = z] < \left( \frac{\epsilon}{1 - \epsilon} \right)^l \sum_{z \in A} \Pr_{\rho \sim P_e} [\rho = v_z]
\]

\[
\leq \left( \frac{\epsilon}{1 - \epsilon} \right)^l \sum_{v' \in \mathbb{F}_2^N} \Pr_{\rho \sim P_e} [\rho = v']
\]

\[
= \left( \frac{\epsilon}{1 - \epsilon} \right)^l.
\]
We have shown that two “bad” events (events in which the $P_\epsilon$-decoder performs worse than the $\lambda_S$-decoder) are unlikely to occur: the bad event where the error string $\rho$ has weight far away from $\epsilon N$, and the bad event where the neighbor $v_\rho$ of the error string $\rho$ has small weight $|v_\rho| << \epsilon N$. The only potential issues that could differentiate the two decoders must then occur when both the error string $\rho$ and its closest collider $v_\rho$ are in the set $\{z \in F_2^N : |z| \in S\}$. But in this regime all strings have similar weight, and so are given similar probability under the distribution $P_\epsilon$. Intuitively, the $P_\epsilon$-decoder must then be very similar to the $\lambda_S$-decoder over this regime. The following theorem makes this idea precise, and then use Lemma 14 to bound the probability of a decoding error.

**Theorem 16.** 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

$$\Pr_\rho [\rho \notin D_k(H\rho^\top)] \leq \left(\frac{\epsilon}{1 - \epsilon}\right)^l + e^{-\frac{l^2}{2N}} + \frac{2(l + 1)}{k} \max_{S \subseteq \{\epsilon N - 2l, ..., \epsilon N + l\}} \{|\text{Coll}_H(S)| - 1\}. $$

*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. $k = 1$**

We start by noting that in order for the decoder to make a mistake on the string $\rho$, the set $\Omega_\rho$ of lower-weight colliders must have size at least 2, i.e.

$$\Pr_\rho [\rho \notin D_1(H\rho^\top)] \leq \Pr_\rho [|\Omega_\rho| > 1]. \quad (4)$$

We will show that this is unlikely to happen. We first bound the probability that one of two ”bad” events occurs: the error string $\rho$ being far away from its neighbor $v_\rho$, or the error string $\rho$ having unusual weight. For this, we define the sets

$$A = \{z \in F_2^N : |v_z| < |z| - l\},$$

$$B = \{z \in F_2^N : |z| - \epsilon N > l\}.$$  

We note that $A$ is a subset of $\{z : |\Omega_z| > 1\}$, as $v_z$ is only defined for strings $z \in F_2^N$ that have lower-weight colliders. Using basic conditional probability, we get

$$\Pr_\rho [|\Omega_\rho| > 1] \leq \Pr_\rho [\rho \notin A] + \Pr_\rho [|\Omega_\rho| > 1 \text{ and } \rho \notin A]$$

$$\leq \Pr_\rho [\rho \notin A] + \Pr_\rho [\rho \in B] + \Pr_\rho [|\Omega_\rho| > 1 \text{ and } \rho \notin A | \rho \notin B].$$

To bound the first and second terms, we apply Lemma 15 and Chernoff’s bound respectively. For the third term, we consider the most problematic weight level within
We proceed to bound the third term of the right-hand side. Note that the fact that $|\Omega_\rho| > 1$ ensures that $v_\rho$ is well-defined. But under the condition $|\rho| = w$, we can only have $|v_\rho| = w'$ if $|\Omega_\rho^{(w,w')}| > 1$, where we recall the definition

$$\Omega_\rho^{(w,w')} = \{|y| \in F_2^N : |y| \in \{w, w'\} \text{ and } Hy^T = Hz^T\}.$$ 

For any $w$ and $w'$, we thus bound the corresponding term in (5) by

$$\Pr_{\rho \sim P_\rho} \left[ |\Omega_\rho| > 1 \text{ and } |v_\rho| = w' \mid |\rho| = w \right] \leq \Pr_{\rho \sim P_\rho} \left[ |\Omega_\rho^{(w,w')}| > 1 \mid |\rho| = w \right] \leq \Pr_{\rho \sim \lambda_{(w)}} \left[ |\Omega_\rho^{(w,w')}| > 1 \right] \leq \Pr_{\rho \sim \lambda_{(w,w')}} \left[ |\Omega_\rho^{(w,w')}| > 1 \mid |\rho| = w \right].$$

Using basic conditional probability, we get

$$\Pr_{\rho \sim P_\rho} \left[ |\Omega_\rho| > 1 \text{ and } |v_\rho| = w' \mid |\rho| = w \right] \leq \frac{\Pr_{\rho \sim \lambda_{(w,w')}} \left[ |\Omega_\rho^{(w,w')}| > 1 \right]}{\Pr_{\rho \sim \lambda_{(w,w')}} \left[ |\rho| = w \right]}. \quad (7)$$

Now for any $w < \frac{N}{2}$ and $w' \leq w$, we have $\Pr_{\rho \sim \lambda_{(w,w')}} \left[ |\rho| = w \right] = \frac{\binom{N}{w}}{2^N} \geq \frac{\binom{N}{w}}{\binom{N}{w'}} \geq \frac{1}{2}$. It then follows from (7) that

$$\Pr_{\rho \sim P_\rho} \left[ |\Omega_\rho| > 1 \text{ and } |v_\rho| = w' \mid |\rho| = w \right] \leq 2 \Pr_{\rho \sim \lambda_{(w,w')}} \left[ |\Omega_\rho^{(w,w')}| > 1 \right].$$

Combining this with equations (4) and (5), we get

$$\Pr_{\rho \sim P_\rho} \left[ \rho \notin D_1(H\rho^+) \right] \leq \left( \frac{\epsilon}{1 - \epsilon} \right)^l + e^{-\frac{2}{5N^2}} + 2(l + 1) \cdot \max_{S \subseteq \{\epsilon N - 2l, \ldots, \epsilon N + l\}} \left\{ \Pr_{\rho \sim \lambda_S} \left[ |\Omega_\rho^S| > 1 \right] \right\}. \quad (8)$$

The theorem statement then follows from Lemma 14.

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

We will show that a slightly less performant decoding function $D_{k,l} : F_2^k \to F_2^N$ satisfies
the desired probability bound. We define $D_{k,l}$ as follows: upon receiving input $x \in \mathbb{F}_2$, $D_{k,l}$ outputs $\frac{k-1}{2l+1}$ strings from $\{z \in \mathbb{F}_2^N : H z = x, |z| = w\}$, 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

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

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

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

we have, by Chernoff’s bound, that

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

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

Since the distribution $P_\epsilon$ gives the same probability to any two strings of equal weights, we get

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

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

Applying Lemma 14, we get

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

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

5 A Fourier Analytic Criterion

In this section, we give a Fourier analytic criterion that certifies that a linear code $C \subseteq \mathbb{F}_2^N$ is resilient to errors of probability $\epsilon$. We give such a criterion for both unique decoding and list decoding. The function we will need to make this connection is the level function of any set $S \subseteq \{0, ..., N\}$, 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 \( \text{Coll}_H(S) \) in terms of the Fourier coefficients of the level function \( L_S \). Recall that we use \( \mu_t \) to denote the uniform distribution over all vectors in \( \mathbb{F}_2^t \).

**Proposition 17.** 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 \( S \subseteq \{1, \ldots, N\} \), we have

\[
\text{Coll}_H(S) = \frac{2^N}{\binom{N}{S}} \cdot \mathbb{E}_{v \sim \mu_t} [\hat{L}_S(vH)^2].
\]

**Proof.** The main tool we will use is Parseval’s Identity, which relates the evaluations \( f(x) \) of a function \( f : \mathbb{F}_2^t \rightarrow \mathbb{R} \) 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 \( f(x) = \Pr_{z \sim \lambda_S} [Hz^\top = x] \), we get

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

\[
= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} f(x)^2
\]

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

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

\[
\text{Coll}_H(S) = \binom{N}{S} \cdot 2^{-t} \sum_{y \in \mathbb{F}_2^t} \left( \sum_{x \in \mathbb{F}_2^t} f(x) \cdot (-1)^{y \cdot x} \right)^2. \tag{8}
\]

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

\[
f(x) = \Pr_{z \sim \lambda_S} [Hz^\top = x] = \frac{1}{\binom{N}{S}} \sum_{z \in \mathbb{F}_2^N \atop Hz^\top = x} L_S(z). \tag{9}
\]

Combining expressions (8) and (9) and applying the definition of the Fourier transform, we get

\[
\text{Coll}_H(S) = \binom{N}{S} \cdot 2^{-t} \sum_{y \in \mathbb{F}_2^t} \left( \sum_{z \in \mathbb{F}_2^N \atop Hz^\top = x} \frac{L_S(z)}{\binom{N}{S}} \cdot (-1)^{yHz^\top} \right)^2
\]

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

\(\square\)
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

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

**Proof.** Applying Proposition 17 with \( H \) the \( N \times N \) identity matrix \( I \), we have

\[
\left( \frac{N}{S} \right) \sum_{x \in \mathbb{F}_N^2} \Pr_{z \sim \lambda_S} [z = x]^2 = \frac{2^N}{(N_S)^j} \mathbb{E}_{v \sim \mu_N} [\hat{L}_S(v)^2].
\]

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

\[
1 = \frac{2^N}{(N_S)^j} \sum_{j=0}^{N} \Pr_{v \sim \mu_N} [|v| = j] \cdot \hat{L}_S(1_j)^2
\]

\[
= \frac{1}{(N_S)^j} \sum_{j=0}^{N} \binom{N}{j} \hat{L}_S(1_j)^2,
\]

where \( 1_j \in \mathbb{F}_2^N \) denotes the vector with 1s in the first \( j \) coordinates and 0s in the last \( N - j \) coordinates. \[\square\]

We will now combine Theorem 16 and Proposition 17 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 \( L_\epsilon \). We prove a generalized version of Theorem 2 below. To recover Theorem 2, set \( k = 1 \) and \( l = \frac{N^3}{4} \). 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 quantity \( C_{\epsilon,l} \) is small.

**Theorem 19.** 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 \( l \leq \left( \frac{1}{2} - \epsilon \right)N \). Then

\[
\Pr_{\rho \sim P_t}[\rho \notin D_k(H\rho^T)] \leq C_{\epsilon,l} + \frac{2(l+1)}{k} \max_{S \subseteq \{\epsilon N - 2l, ..., \epsilon N + l\}} \left\{ \frac{2^N}{(N_S)^j} \mathbb{E}_{v \sim \mu} [\hat{L}_S(vH)^2] - 1 \right\},
\]

where we have defined \( C_{\epsilon,l} = \left( \frac{\epsilon}{1 - \epsilon} \right)^l + e^{-\frac{l^2}{2N} \epsilon} \), and where the function \( 1\{k = 1\} \) is 1 when \( k = 1 \), and 0 otherwise.

21
Proof. Applying Theorem 16 and Proposition 17 we have

\[
\Pr_{\rho \sim P_c}[\rho \notin D_k(H\rho^T)] \leq C_{\epsilon,l} + \frac{2(l + 1)}{k} \max_{S \subseteq \{\epsilon N - 2l, \ldots, \epsilon N + l\}} \frac{1}{1 \leq |S| \leq 1 + 1(k = 1)} \left\{ \text{Col}_H(S) - 1 \right\}
\]

\[
= C_{\epsilon,l} + \frac{2(l + 1)}{k} \max_{S \subseteq \{\epsilon N - 2l, \ldots, \epsilon N + l\}} \frac{1}{1 \leq |S| \leq 1 + 1(k = 1)} \left\{ \frac{2^N}{\binom{N}{S}} \mathbb{E}_{v \sim \mu_t} \left[ \hat{L}_S(vH)^2 \right] - 1 \right\}.
\]

As an application of Theorem 19 we present the following bound on the probability of making a decoding error on any code \( C \subseteq \mathbb{F}_2^N \). We note that our bound depends only on the weight distribution of the dual code \( C^\perp \).

**Proposition 20.** Fix any \( \epsilon \in (0, \frac{1}{2}) \), define \( \bar{\epsilon} = \epsilon + N^{-\frac{1}{2}} \) and let \( B = \{\beta N, \ldots, (1 - \beta)N\} \) for \( \beta = \frac{1}{2} \left( 1 - 2 \sqrt{\bar{\epsilon}(1 - \bar{\epsilon})} \right) \). Let \( k^* = (2N^{3/4} + 1)m + 1 \) for some integer \( m \geq 0 \). Then for all \( N > \left( \frac{5}{16} \right)^{20} \) and all \( k \geq k^* \), we have that any \( t \times N \) matrix \( H \) with entries in \( \mathbb{F}_2 \) satisfies

\[
\Pr_{\rho \sim P_c}[\rho \notin D_k(H\rho^T)] \leq C_\epsilon + \frac{3N^{3/4}}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[|vH| = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}
\]

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

where \( C_\epsilon = \left( \frac{\epsilon}{1 - \epsilon} \right)^{N^{3/4}} + \frac{2^N}{\epsilon} \).

Proof. We will use Theorem 19 to bound the decoding error probability in terms of the Fourier coefficients \( \hat{L}_S(1j) \) and the probability factors \( \Pr_{v \sim \mu_t}[|vH| = j] \). Some of these factors will then be bounded using Corollary 18 and some will be bounded using Theorem 6. We proceed with the proof; letting \( l = N^{3/4} \) in Theorem 19 we get

\[
\Pr_{\rho \sim P_c}[\rho \notin D_k(H\rho^T)] \leq C_\epsilon + \frac{3N^{3/4}}{k^*} \max_{S \subseteq \{\epsilon N - 2l, \ldots, \epsilon N + l\}} \frac{1}{1 \leq |S| \leq 1 + 1(k^* = 1)} \left\{ \frac{2^N}{\binom{N}{S}} \sum_{j = 0}^N \Pr_{v \sim \mu_t}[|vH| = j] \hat{L}_S(1j)^2 - 1 \right\},
\]

(10)

where \( C_\epsilon = \left( \frac{\epsilon}{1 - \epsilon} \right)^{N^{3/4}} + \frac{2^N}{\epsilon N} \). We want to bound the summation in the second term. We will start with the central terms \( j \in B := \{\beta N, \ldots, (1 - \beta)N\} \). For these we rely on Corollary 18 which states that \( \frac{1}{\binom{N}{S}} \sum_{j = 0}^N \binom{N}{j} \cdot \hat{L}_S(1j)^2 = 1 \) for all \( S \subseteq \{0, \ldots, N\} \). For
any such \( S \), we then get

\[
\frac{2^N}{\binom{N}{S}} \sum_{j \in B} \Pr_{v \sim \mu_B} [|vH| = j] \hat{L}_S(1_j)^2 \leq \frac{2^N}{\binom{N}{S}} \max_{j \in B} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\} \sum_{j \in B} \binom{N}{j} \cdot \hat{L}_S(1_j)^2 \leq 2^N \max_{j \in B} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \cdot \frac{1}{\binom{N}{j}} \right\}.
\]

We then want to bound the contribution of the faraway terms \( j \notin B \) to the summation in (10), i.e. we want to bound

\[
\max_{S \subseteq \{ \epsilon N - 2l, \ldots, \epsilon N + l \}} \left\{ \frac{2^N}{\binom{N}{S}} \sum_{j \notin B} \Pr_{v \sim \mu_B} [|vH| = j] \hat{L}_S(1_j)^2 \right\}.
\]

We will want to apply Theorem 6 to every Fourier coefficient in this sum. To do this, we will first need to bound the Fourier coefficients \( \hat{L}_{S,v} \). Now note that by definition, if \( j \notin B \) then \( \hat{L}_{S,v}(y) = 2^{-N/2} \sum_{x \in \mathbb{F}_2^N} (-1)^{x \cdot y} \hat{L}_{S,v}(y) \leq 2 \cdot \max \{ \hat{L}_{S}(y), \hat{L}_{S,v}(y) \} \).

Combining this fact with Theorem 6, we bound the expression in (12) by

\[
\leq \frac{2^N}{\binom{N}{\epsilon N - N^{3/4}}} \cdot N \max_{S \subseteq \{ \epsilon N - 2N^{3/4} \}} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \hat{L}_S(1_j)^2 \right\} \leq \frac{2^N}{\binom{N}{\epsilon N - N^{3/4}}} \cdot N \max_{j \notin B} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \cdot 4 \hat{L}_S(1_j)^2 \right\} \leq \frac{4N}{\epsilon N^{2/4}} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \left( \frac{1}{\epsilon - N^{-1/4}} \right)^2 \right\}.
\]

Now note that by definition, if \( j \notin B \) then \( \frac{1}{\epsilon - \frac{j}{N}} \geq \epsilon(1 - \epsilon) > \epsilon \geq \epsilon - N^{-\frac{1}{4}}. \) Thus the quotient in (13) can be bounded by \( \left| \frac{1}{\epsilon - N^{-\frac{1}{4}}} \right| \cdot e^2 > e^2 > 1 \), which then implies that \( \left( \frac{1}{\epsilon - N^{-1/4}} \right)^2 \) is maximized at the largest possible \( w \). We thus get

\[
\leq \frac{4N}{\epsilon N^{2/4}} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_B} [|vH| = j] \left( \frac{1}{\epsilon - N^{-1/4}} \right)^2 \right\}.
\]
Combining this bound for the faraway terms with our bound (13) for the central terms of the summation, we bound the right-hand side of equation (10) by

$$\Pr [\rho \notin D_k(H\rho^T)] \leq C_\epsilon + \frac{3N^{3/4}}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} [\|vH\| = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}$$

$$+ \frac{3N^{3/4}}{k^*} \max_{j \notin B} \left\{ \frac{4N}{N} \Pr_{v \sim \mu_t} [\|vH\| = j] \left( \frac{1}{\epsilon - N^{-1/4}} e^{2(\epsilon N + N^{3/4})} - 1 \right) \right\},$$

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

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

and that

$$\left( \epsilon - N^{-1/4} \right)^{-2\epsilon N} = \left( \frac{1}{\epsilon} \right)^{2\epsilon N} \left( \frac{\epsilon + N^{-1/4}}{\epsilon - N^{-1/4}} \right)^{2\epsilon N} \leq \left( 1 + \frac{4N^{-1/4}}{\epsilon} \right)^{2\epsilon N} \left( \frac{1}{\epsilon} \right)^{2\epsilon N} \leq 2^{8\frac{1}{\epsilon} (1 + \frac{N^{-1/4}}{\epsilon}) N^{3/4}} \left( \frac{1}{\epsilon} \right)^{2\epsilon N}.$$

For $$N > \left( \frac{2}{\epsilon} \right)^{20}$$, equation (14) can then be bounded by

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^T)] \leq C_\epsilon + \frac{3N^{3/4}}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} [\|vH\| = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}$$

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

\(\square\)

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

$$C = \{ (z, \ldots, z) \in \mathbb{F}_q^N : z \in \mathbb{F}_q^{\eta N} \}$$
is transitive, has dimension $\eta N$, and has weight distribution
\[
\Pr_{c \sim D(C)}[|c| = \alpha N] = q^{-\eta N} \cdot \left(\frac{\eta N}{(1 - \alpha)\eta N}\right) (q - 1)^{\alpha N}
\geq q^{-\eta N} \cdot \sqrt{\frac{8\pi}{e^4} \cdot 2^{h_\alpha N} \cdot q^{\alpha N \log_q(q - 1)}}
\]
\[
= \sqrt{\frac{8\pi}{e^4} \cdot q^{-(1 - h_\alpha(q))\eta N}}
\]
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
\[
\Pr_{c \sim D(C)}[|c| = \alpha N] \leq q^{-(1 - h_\alpha(q))\dim C},
\]
where $D(C)$ is the uniform distribution over all codewords in $C$, and $h_\alpha$ is the $q$-ary entropy
\[
h_\alpha = (1 - \alpha) \log_q \frac{1}{1 - \alpha} + \alpha \log_q \frac{q - 1}{\alpha}.
\]

**Proof.** Let $M$ be the $t \times N$ generator matrix of $C$, and let $r = \rank M = \dim C$. 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 $j \in \{1, ..., N\}$,
\[
\Pr_{Z \sim D(C^{(\alpha)})}[Z_j = 0] = 1 - \alpha. \quad (15)
\]
But under condition (15), $Z_j$ 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
\[
\mathbb{H}_{Z \sim D(C^{(\alpha)})}(Z_j) \leq (1 - \alpha) \log \frac{1}{1 - \alpha} + (q - 1) \cdot \frac{\alpha}{q - 1} \log \frac{q - 1}{\alpha}
\]
\[
= h_\alpha(\alpha) \log(q). \quad (16)
\]
We will now show that $\mathbb{H}(Z_j | Z_1, ..., Z_{j-1}) = 0$ 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 $\beta_1, ..., \beta_r \in \mathbb{F}_q$. 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 $\{Z_1, \ldots, Z_r\}$, and so we indeed have

$$H_{Z \sim D(C(\alpha))}(Z_j|Z_1, \ldots, Z_{j-1}) = 0$$

for every $j > r$. Applying (16) and the chain rule for entropy then gives

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

Now $Z$ is sampled uniformly from $C(\alpha)$, so $H(Z) = \log(|C(\alpha)|)$. We thus have

$$\Pr_{c \sim D(C)}(|c| = \alpha N) = \frac{|C(\alpha)|}{q^r} = 2^{H(Z)} \cdot q^{-r} \leq q^{-(1-h_q(\alpha)) \cdot r}.$$  

\[\square\]

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

**Theorem 21.** For any $n, d < n$, and $\alpha \in (0, 1)$, the Reed-Muller code $\text{RM}(n, d)$ over the prime field $\mathbb{F}_q$ satisfies

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

**Proof.** This follows immediately from Theorem 1, Fact 10, and Fact 9.  

\[\square\]

## 7 List Decoding for Transitive Codes

We now turn to proving Theorem 4. 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 state and prove a generalized version of Theorem 4 below:
Theorem 22. 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 $\epsilon$-errors using a list $T$ of size

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

Proof. Let $\tilde{\epsilon} = \epsilon + N^{3/4}$. 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

$$|T(x)| = 2^{\tilde{\epsilon} N \log \frac{4}{1-\eta}},$$

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

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

Let $H$ denote the parity-check matrix of $C$. By Lemma 13, it is sufficient to show that for any $N > \left(\frac{5}{\epsilon}\right)^{20}$ and any $k > 1$ we have

$$\Pr_{\rho \sim P_\epsilon} \left[ \rho \notin D_k(H\rho^\perp) \right] \leq \frac{1}{N} + \frac{2^{\epsilon N \log \frac{4}{1-\eta}}}{Nk}. \quad (17)$$

We will thus prove (17). Recall that Proposition 20 yields the following bound on the left-hand side of (17):

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

$$+ \frac{2N^{3/5}}{k} \cdot \frac{N}{\binom{N}{\frac{1}{2} - \frac{\epsilon}{N}}} \left( \frac{\epsilon^2}{\tilde{\epsilon}} \right) 2^{\epsilon N} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_\epsilon} [|vH| = j] \cdot \left( \frac{1}{2} - \frac{j}{N} \right)^{2\epsilon N} \right\}, \quad (18)$$

where $C_\epsilon = \left(\frac{\epsilon}{1-\epsilon}\right)^{N^{3/4}} + e^{-\frac{\sqrt{\alpha}}{\ln 2}}$, $\tilde{\epsilon} = \epsilon + N^{3/4}$, and $B = \{\beta N, ..., (1 - \beta)N\}$ for $\beta = \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$. 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 11 that for all $j \in \{0, ..., N\}$,

$$\Pr_{v \sim \mu_\epsilon} [|vH| = j] \leq 2^{-(1-h(\frac{j}{N}))(1-\eta)N}. \quad (19)$$

We recall also that for any $j \in \mathbb{N}$ and $\alpha$ such that $\left| \frac{1}{2} - \frac{j}{N} \right| = \sqrt{\alpha \tilde{\epsilon}(1-\tilde{\epsilon})}$, we have by Lemma 12 that

$$\frac{2\alpha \tilde{\epsilon}(1-\tilde{\epsilon})}{\ln 2} < 1 - h(j/N) < 4\alpha \tilde{\epsilon}(1-\tilde{\epsilon}). \quad (20)$$
We will use equations (19) and (20) to bound every term in (18). We start with the central terms. Fixing any \( j \in B \), we have by Fact 11 and equation (19) that

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

But for \( j \in B \) we have \( \beta < \frac{j}{N} < 1 - \beta \), so the right-hand side is maximized at \( j = \beta N \). Combining this with (20), we get that

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

(21)

We now turn to the faraway terms. Fix \( j \notin B \) and define \( \alpha > 1 \) such that \( \left| \frac{1}{2} - \frac{j}{N} \right| = \sqrt{\alpha \bar{\epsilon}(1 - \bar{\epsilon})} \). By equations (19) and (20), we then have

\[
\Pr_{v \sim \mu_t} [|vH| = j] \cdot \left( \frac{1}{2} - \frac{j}{N} \right)^{2\varepsilon N} \leq 2^{-(1-h(j/N))(1-\eta)N} \cdot \left( \alpha \bar{\epsilon}(1 - \bar{\epsilon}) \right)^{\varepsilon N} \cdot \left( \alpha \bar{\epsilon}(1 - \bar{\epsilon}) \right)^{\varepsilon N} \cdot \frac{\beta N}{\sqrt{e^N}} \cdot 2^{(1-h(\beta))\eta N} \cdot \left( \frac{\beta N}{\sqrt{e^N}} \right)^{\varepsilon N}.
\]

For any positive constant \( c \), the derivative of \( \log(\alpha) - c \alpha \) is \( \frac{1}{\alpha \ln 2} - c \), and the second derivative is always negative. Thus, the above expression achieves its maximum when

\[
\alpha = \frac{2\varepsilon N}{\log \frac{\varepsilon N}{\frac{1}{(1-c)}(1-\eta)N}}.
\]

We then get

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

(22)

We now use equations (21) and (22) to bound the central and faraway terms of (18) respectively. This gives

\[
\Pr_{\rho \sim P_t} [\rho \notin D_k(H \rho^*)] \leq \frac{1}{N} + \frac{3N}{k} \cdot \sqrt{\frac{\varepsilon N}{8\pi}} \cdot 2^{4(1-\bar{\epsilon})\eta N} \cdot \left( \frac{\varepsilon N}{\frac{1}{(1-c)}(1-\eta)N} \right)^{\varepsilon N} \cdot \left( \frac{\bar{\epsilon}}{2e(1-\eta)} \right)^{\varepsilon N}.
\]

(28)
Using Fact 11 to bound \( \frac{N}{\tilde{\epsilon}N} \geq \sqrt{\frac{8\pi}{eN}} \cdot 2^{h(\tilde{\epsilon})N} \geq \sqrt{\frac{8\pi}{eN}} \cdot \left( \frac{1}{2} \right)^{\tilde{\epsilon}N} \), we get

\[
\Pr_{\rho \sim \mathcal{P}_{\tilde{\epsilon}}} [\rho \notin D_k(H\rho^T)] \leq \frac{1}{N} + \frac{N^2}{k} \left( 2^{4\tilde{\epsilon}N} + 2^{\tilde{\epsilon}N} \log \frac{2^{e^{\frac{4}{2(1-\eta)}}} + 2N^{4/5}}{Nk} \right)
\]

\[
\leq \frac{1}{N} + \frac{2^{\tilde{\epsilon}N} \log \frac{e^{\frac{4}{2(1-\eta)}}}{Nk}}{Nk}.
\]

We have shown (17), and so we are done. \(\square\)

### 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 \(\text{RM}(n, d)\) is the code \(\text{RM}(n, n - d - 1)\), so we can apply Samorodnitsky’s Theorem 7 to our Proposition 20. We state and prove a generalized version of our Theorem 5 below.

**Theorem 23.** Fix any \(\epsilon \in (0, \frac{1}{2})\), \(\eta \in (0, 1)\), and \(N > \left( \frac{5}{\epsilon} \right)^{20}\). Then any Reed-Muller code \(\text{RM}(n, d)\) of dimension \(\left( \frac{n}{d} \right) \leq \eta N\) can with high probability list-decode \(\epsilon\)-errors using a list \(T\) of size

\[
|T| = 2^{(\epsilon \log \frac{e}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2)N + o(N)}.
\]

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

\[
|T(x)| = 2^{(\epsilon \log \frac{e}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2)N + N^{9/10}}
\]

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

\[
\Pr_{\rho \sim \mathcal{P}_{\tilde{\epsilon}}} [c \notin T(c + \rho)] \leq \frac{2}{N}.
\]

Let \(H\) denote the parity-check matrix of \(\text{RM}(n, d)\). By Lemma 13, it is sufficient to show that for any \(N > \left( \frac{5}{\epsilon} \right)^{20}\) and any \(k > 1\) we have

\[
\Pr_{\rho \sim \mathcal{P}_{\tilde{\epsilon}}} [\rho \notin D_k(H\rho^T)] \leq \frac{1}{N} + \frac{2^{\frac{\epsilon}{(1-\eta)^2} + 4\epsilon + (1-\eta)^2}N}{kN}.
\]

(23)

We will thus prove (23). Recall that for \(k > 2\sqrt{N}\) Proposition 20 yields the following bound on the left-hand side of (23):

\[
\Pr_{\rho \sim \mathcal{P}_{\tilde{\epsilon}}} [\rho \notin D_k(H\rho^T)] \leq C_\epsilon + \frac{3N}{k} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot \frac{2N}{\binom{N}{j}} - 1 \right\}
\]

\[
+ \frac{2^{N^{4/5}}}{k} \cdot \frac{N}{\binom{N}{\frac{4}{5}\tilde{\epsilon}N}} (\frac{e^2}{\tilde{\epsilon}}) \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} [|vH| = j] \cdot \left( \frac{1}{2} - \frac{j}{N} \right)^{2\tilde{\epsilon}N} \right\},
\]

(24)
where \( C_\epsilon = (\epsilon/1-\epsilon)^{N^{3/4}} + e^{-\sqrt{N} \epsilon}, \epsilon = \epsilon + N^{3/4}, \) and \( B = \{ \beta N, ... , (1-\beta)N \} \) for \( \beta = \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}. \) 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 22: by Theorem 21 we know that the weight distribution of the Reed-Muller code satisfies

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

so by Fact 11 we have

\[
\max_{j \in B} \left\{ \Pr_{v \sim \mu} [|vH| = j] \cdot \left(\frac{2^N}{N}ight)^j \right\} \leq \max_{j \in B} \left\{ 2^{-(1-h(j/N))\gamma N} \cdot \frac{2^N}{\sqrt{8\pi \epsilon^3 N}} \cdot 2^{h(j/N)N} \right\}
\]

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

But \( B = \{ \beta N, ... , (1-\beta)N \}, \) so by Lemma 12 we have

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

\[
\leq \sqrt{e^4 N/8\pi} \cdot 2^{4\epsilon(1-\epsilon)\eta N}.
\] (25)

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

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

\[
= 2^{o(N)} 2^{-(1-\eta)N} \max_{j \leq \frac{N}{2}} \left\{ 2^{-2j \ln \log(\eta) + 2\epsilon N \log(\frac{1}{2} - \frac{j}{N})} \right\}.
\] (26)

Now the function

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

has first derivative

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

and second derivative

\[
\frac{dg^2}{d^2j} = -\frac{2\epsilon}{\ln 2 \cdot N(\frac{1}{2} - \frac{j}{N})^2} < 0.
\]
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 (24) by

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

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

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

$$\leq 2\left(\gamma^2 - \frac{\tilde{\epsilon}^2}{2N} + 2\tilde{\epsilon} \log \tilde{\epsilon} + 2\tilde{\epsilon} \log \frac{1}{\ln 2} + 2\tilde{\epsilon} \log \frac{1}{\gamma}\right) N. \quad (27)$$

Using inequalities (25) and (27) and recalling that $\gamma = 1 - \eta$, we bound the right-hand side of (24) by

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

$$+ \frac{2^{N^{4/5}}}{k} \cdot \frac{N}{\tilde{\epsilon} N} \cdot \left(\frac{\tilde{\epsilon}}{\epsilon}\right)^{2\tilde{\epsilon}N} \cdot 2^{(1 - \eta)^2 - \frac{2\tilde{\epsilon}N}{\ln 2} + 2\tilde{\epsilon} \log \tilde{\epsilon} + 2\tilde{\epsilon} \log \frac{1}{\ln 2} + 2\tilde{\epsilon} \log \frac{1}{\gamma} - \frac{1}{1 - \eta}} N.$$  

Now by Fact [11] we know that $\left(\frac{N}{\tilde{\epsilon} N}\right) \geq \frac{8}{\eta} \cdot 2^{h(\tilde{\epsilon}) N}$, so we get

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

$$+ \frac{2^{N^{4/5}}}{k} \cdot \frac{2^{-h(\epsilon)N} \cdot 2^{2\tilde{\epsilon}N \log \frac{1}{\epsilon}} \cdot 2^{\left(1 - \eta\right)^2 - \frac{2\tilde{\epsilon}}{\ln 2} + 2\tilde{\epsilon} \log \tilde{\epsilon} + 2\tilde{\epsilon} \log \frac{1}{\ln 2} + 2\tilde{\epsilon} \log \frac{1}{\gamma} - \frac{1}{1 - \eta}} N}$$  

$$= C_\epsilon + \frac{3N}{k} \sqrt{\frac{e^4N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1 - \tilde{\epsilon})\eta N} + \frac{2^{N^{4/5}}}{k} \cdot 2^{\left(h(\epsilon) + \frac{2\tilde{\epsilon}}{\ln 2} + 2\tilde{\epsilon} \log \frac{1}{\ln 2} + 2\tilde{\epsilon} \log \frac{1}{\gamma} - (1 - \eta)^2\right) N}$$  

$$\leq C_\epsilon + \frac{3N}{k} \sqrt{\frac{e^4N}{8\pi}} \cdot 2^{4\tilde{\epsilon}(1 - \tilde{\epsilon})\eta N} + \frac{2^{N^{4/5}}}{k} \cdot 2^{\left(h(\epsilon) + 2\tilde{\epsilon} \log \frac{1}{1 - \eta} + (1 - \eta)^2\right) N}.$$  

Now $\tilde{\epsilon} \leq (1 + N^{-\frac{1}{10}})\epsilon$ for all $N > \left(\frac{5}{2}\right)^{20}$, so we get

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\top)] \leq \frac{1}{N} + 2^{N^{9/10}} \cdot \frac{2^{4\epsilon N} + 2^{(\epsilon \log \frac{\epsilon}{1 - \eta} + 4\epsilon) + (1 - \eta)^2} N}{2kN}.$$  

Since $\epsilon \log \frac{\epsilon}{a} + a \geq 0$ for all $\epsilon, a \in (0, 1)$, we have

$$\Pr_{\rho \sim P_\epsilon} [\rho \notin D_k(H\rho^\top)] \leq \frac{1}{N} + 2^{N^{9/10}} \cdot \frac{2^{(\epsilon \log \frac{\epsilon}{1 - \eta} + 4\epsilon) + (1 - \eta)^2} N}{kN}.$$  

We have shown (23), 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 \( S \subseteq \{0, \ldots, N\} \), 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

\[
\hat{L}_S(y) = 2^{-N/2} \sum_{x \in \mathbb{F}_2^N} L_S(x)(-1)^{\langle x, y \rangle} = \frac{(N_S)}{2^{N/2}} \mathbb{E}_{x \sim \lambda_S} (-1)^{\langle x, y \rangle},
\]

where we recall the definition \( (N_S) = \sum_{j \in S} (N_j) \). We observe from (28) that the Fourier coefficients of the level function are symmetric in multiple ways. First, \( |\hat{L}_S(y)| = |\hat{L}_S(y')| \) for any \( S \subseteq \{0, \ldots, N\} \) 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 \( j \in \{0, \ldots, N\} \). When computing bounds on the Fourier coefficients \( \hat{L}_{(j)}(y) \), it will thus suffice to restrict our attention to the case where \( j \leq N/2 \) and \( y \) is of the form \( y = 1_w \) for some \( w \leq N/2 \) (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 24.** For any \( S \subseteq \{0, \ldots, N\} \) and \( w \in \{0, \ldots, N\} \), we have

\[
|\hat{L}_S(1_w)| \leq \sqrt{(N_S) \left( \frac{N}{N_w} \right)}.
\]

**Proof.** Recall from Corollary 18 that we have

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

In particular, for every \( w \in \{0, \ldots, N\} \) we must have

\[
\hat{L}_S(1_w)^2 \leq \frac{(N)}{(N_w)}.
\]

\[\square\]
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 24 can be significantly improved whenever $|1/2 - \delta| \geq \sqrt{\epsilon(1 - \epsilon)}$. 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

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

where $\gamma$ 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 $|1/2 - \delta| \geq \sqrt{\epsilon(1 - \epsilon)}$, but that for completeness we prove bounds for all regimes. The first half of Theorem 6 follows directly from Theorem 24. For the second half, we state and prove the following:

**Theorem.** For any $\epsilon, \delta \in (0, 1/2)$, 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

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

Moreover, we have

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

**Proof.** By definition of the Fourier transform, we have

$$\hat{L}_{(\epsilon N)}(1_{\delta N}) = 2^{-N/2} \sum_{x \in \mathbb{F}_2^N} \prod_{j=1}^{\delta N} (-1)^{x_j}$$

$$= 2^{-N/2} \cdot \text{coefficient of } z^{\epsilon N} \text{ in } (1 - z)^{\delta N}(1 + z)^{(1 - \delta)N}.$$ 

Applying the residue theorem \([29]\), we then get that

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

(30)
for any curve $\gamma$ around the origin of the complex plane. We now define the polar coordinates $r, \theta$ to be such that $s = re^{i\theta}$, where $s$ is the complex number defined in the theorem statement. Letting the contour $\gamma$ in equation (30) be the circle of radius $r$ around the origin (i.e. $\gamma(\theta) = re^{i\theta}$), we get

$$\hat{L}_{\{\epsilon N\}}(1_{\delta N}) = \frac{2^{-N/2}}{2\pi} \cdot \int_{-\pi}^{\pi} \frac{(1 - re^{i\theta})\delta N (1 + re^{i\theta})^{(1-\delta)N}}{(re^{i\theta})^{\epsilon N}} \, d\theta. \quad (31)$$

Our approach will be to bound the integrand in (31) by its maximal value over the interval $\theta \in [-\pi, \pi]$. For this, we define the magnitude

$$\tau(\theta) = \left|(1 - re^{i\theta})\delta N (1 + re^{i\theta})^{(1-\delta)N}\right|^2$$

$$= (1 - r \cos \theta - ir \sin \theta)^\delta N (1 - r \cos \theta + ir \sin \theta)^\delta N \cdot (1 + r \cos \theta + ir \sin \theta)^{(1-\delta)N} (1 + r \cos \theta - ir \sin \theta)^{(1-\delta)N}$$

$$= (1 - 2r \cos \theta + r^2)^\delta N (1 + 2r \cos \theta + r^2)^{(1-\delta)N}.$$ 

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

$$\tau'(\theta) = 2r\delta N \sin \theta \cdot (1 - 2r \cos \theta + r^2)^{\delta N - 1} (1 + 2r \cos \theta + r^2)^{(1-\delta)N}$$

$$- 2r(1 - \delta)N \sin \theta \cdot (1 - 2r \cos \theta + r^2)^{\delta N} (1 + 2r \cos \theta + r^2)^{(1-\delta)N-1}$$

$$= 2N r \sin \theta \cdot (1 - 2r \cos \theta + r^2)^{\delta N - 1} (1 + 2r \cos \theta + r^2)^{(1-\delta)N-1}$$

$$\cdot (2r \cos \theta - (1 - 2\delta)(1 + r^2)).$$

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

$$\text{sgn}(\tau'(\theta)) = \text{sgn}(\sin \theta) \cdot \text{sgn}(2r \cos \theta - (1 - 2\delta)(1 + r^2)). \quad (32)$$

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

We will rely on the following two facts, which are proven in claim 25

$$(1 - 2\delta) r^2 - 2r + 1 - 2\delta > 0, \quad (33)$$

and

$$r = \omega \cdot \frac{2\epsilon}{1 - 2\delta} \quad (34)$$

for some $\omega \in [\frac{1}{2}, 1]$. It follows from (32) and (33) that for every $\theta$ we have

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

which implies that $\tau(\theta)$ is increasing over $[-\pi, 0]$ and decreasing over $[0, \pi]$. By equation (31) and since $s = r$ when $\delta < \sqrt{\epsilon(1 - \epsilon)}$, we then have

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

$$= 2^{-N/2} \cdot \left| \frac{(1 - s)\delta N (1 + s)^{(1-\delta)N}}{s^{\epsilon N}} \right|.$$

34
This proves our theorem’s first inequality. To obtain the more explicit second inequality, we use (34) and the inequality \(1 + x \leq e^x\) to bound

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

for some \(\omega \in \left[\frac{1}{2}, 1\right]\). Now the function \(\nu(\omega) := \frac{e^{\epsilon N}}{\omega}\) has first derivative \(\frac{d\nu}{d\omega} = \frac{e^{\epsilon N}(2\omega - 1)}{\omega^2}\) and second derivative \(\frac{d^2\nu}{d\omega^2} = \frac{2e^{\epsilon N}(2\omega^2 - 2\omega + 1)}{\omega^3} > 0\), so \(\nu(\omega)\) achieves its global minimum at \(\omega = \frac{1}{2}\) and is increasing over the interval \(\omega \in \left[\frac{1}{2}, 1\right]\). We can then bound our previous equation by

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

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

In this case, by definition we have

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

It then follows that

\[
\cos t = \frac{1 - 2\delta}{2\sqrt{\epsilon(1 - \epsilon)}}.
\]

But from equation (32) we know that

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

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

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

It follows from (35) and (36) 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 (31) we have

\[
|\hat{L}_{\{\epsilon N\}}(1_{\delta N})| \leq 2^{-N/2} \cdot \max_{\theta \in [-\pi, \pi]} \left|\frac{(1 - re^{i\theta})\delta N(1 + re^{i\theta})(1 - \delta)N}{(re^{i\theta})\epsilon N}\right|
\leq 2^{-N/2} \cdot \frac{(1 - s)\delta N(1 + s)(1 - \delta)N}{s\epsilon N}.
\]
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 \( s = \sqrt{\frac{1}{1-\epsilon}} \cdot (\sqrt{\alpha} + i\sqrt{1-\alpha}) \). We then compute

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

We also compute

\[
|1-s|^2 = 1 + \frac{\alpha\epsilon}{1-\epsilon} - 2\sqrt{\frac{\alpha\epsilon}{1-\epsilon}} + \frac{(1-\alpha)\epsilon}{1-\epsilon} = \frac{1}{1-\epsilon} \cdot (1 - \epsilon + \epsilon - \sqrt{4\alpha\epsilon(1-\epsilon)}) = \frac{2\delta}{1-\epsilon},
\]

and we compute

\[
|1+s|^2 = 1 + \frac{\alpha\epsilon}{1-\epsilon} + 2\sqrt{\frac{\alpha\epsilon}{1-\epsilon}} + \frac{(1-\alpha)\epsilon}{1-\epsilon} = \frac{1}{1-\epsilon} \cdot (1 - \epsilon + \epsilon + \sqrt{4\alpha\epsilon(1-\epsilon)}) = \frac{2(1-\delta)}{1-\epsilon}.
\]

From equation (37), we then have

\[
|\hat{L}_{SS}^w| \leq 2^{-N/2} \cdot \left( \frac{2\delta}{1-\epsilon} \right)^{\frac{4N}{2}} \left( \frac{2(1-\delta)}{1-\epsilon} \right)^{\frac{(1-\delta)N}{2}} \left( \frac{1}{1-\epsilon} \right)^{\epsilon N^2} = 2^{-N/2} \cdot \left( \frac{2\delta(1-\delta)^{1-\delta}}{\epsilon^{\epsilon(1-\epsilon)^{1-\epsilon}}} \right)^{N/2} = 2^{-h(2\delta+h(\epsilon))N/2}.
\]

\[\square\]

Claim 25. 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:

\[
(1-2\delta)r^2 - 2r + 1 - 2\delta > 0, \tag{38}
\]

\[
r = \omega \cdot \frac{2\epsilon}{1-2\delta}, \tag{39}
\]

for some \( \omega \in \left[\frac{1}{2}, 1\right] \).
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
\[
\sqrt{1 - \frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}} = 1 - \frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2} \cdot \omega \quad \text{for some } \omega \in [\frac{1}{2}, 1].
\]
We then have
\[
r = \frac{(1 - 2\delta) - (1 - 2\delta)\sqrt{1 - \frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}}}{2(1 - \epsilon)}
\]
\[
= \frac{1 - 2\delta}{2(1 - \epsilon)} \left( 1 - \sqrt{1 - \frac{4\epsilon(1-\epsilon)}{(1-2\delta)^2}} \right)
\]
\[
= \omega \cdot \frac{2\epsilon}{1 - 2\delta}
\]
for some \(\omega \in [\frac{1}{2}, 1]\), which proves the first claim. Now that we have (40), 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
\[
\eta(\omega) \geq \eta(1) = \frac{1}{1 - 2\delta}((1 - 2\delta)^2 - 4\epsilon(1 - \epsilon)) > 0.
\]
\[\square\]

Acknowledgements

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

A Proofs for Section [1]

A.1 Weight Bounds Comparisons

In this section, we will compare our Theorem [21] with previously known bounds on the weight distribution of Reed-Muller codes. We recall our Theorem [21] below. Note that throughout this section, \(\mathcal{D}(n, d)\) will denote the uniform distribution over all codewords in \(\text{RM}(n, d)\), 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 \(\text{RM}(n, d)\) over the prime field \(\mathbb{F}_q\) satisfies
\[
\Pr_{c \sim \mathcal{D}(n, d)} \left[ |c| = \alpha N \right] \leq q^{-(1-h_q(\alpha))(\frac{n}{d})},
\]

37
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 26 ([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
\[
\Pr_{c \sim D(n, d)} \left[ |c| N \leq 1 - \frac{1}{q} - q^{-c_1 \frac{n}{2}} \right] \leq q^{-c_2 \left( \frac{n}{\delta d} \right)}. \]

This was a generalization of [BHL12], who proved the same result for Reed-Muller codes over \(\mathbb{F}_2\). Theorem 26 is very strong for small degrees, but gets weaker as the degree increases. When \(d\) is linear in \(n\) we have \(q^{-c_1 \frac{n}{2}} = \Theta(1)\), meaning that in this regime Theorem 26 can only give a nontrivial bound on normalized weights that are at least a constant away from \(1 - \frac{1}{q}\). Our Theorem 21 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 26 was proven over \(\mathbb{F}_2\) by [BHL12]. For comparison with our Theorem 21 see the discussion above.

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

**Theorem 27 ([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
\[
\Pr_{c \sim D(n, d)} \left[ |c| \leq \alpha N \right] \leq 2^{o(N)} \left( \frac{1}{1 - \eta} \right)^{2 \ln 2 \cdot \alpha^N} 2^{-\eta N}. \]

This result is strong when \(\alpha\) is away from 1/2. For \(\alpha\) close to 1/2, the following bound is stronger.

**Theorem 28 ([Sam20]).** Let \(\binom{n}{\leq d} = \eta 2^n = \eta N\) for some \(\eta \in (0, 1)\), and define \(A = \{ \frac{1 - \eta^2 \ln 2}{2}, \ldots, \frac{1}{2} \}\). Then for any \(\alpha \in (0, \frac{1}{2})\),
\[
\Pr_{c \sim D(n, d)} \left[ |c| \leq \alpha N \right] \leq 2^{o(N)} \begin{cases} \frac{(\eta N)^{2N}}{2^N} & \text{if } \alpha \in A, \\ \frac{1}{(1 - \eta^2 \ln 2 \alpha N (1 + \eta^2 \ln 2)(1 - \alpha) N)} & \text{otherwise.} \end{cases} \]
We note that the combination of Theorems 27 and 28 is stronger than our Theorem 21 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 27 and 28 to give a strong bound. An approach that has been fairly successful in these two regimes (subconstant 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 29** ([SS20]). For any integers \( j, n, d \), we have

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

We claim that for every \( d > \frac{n}{3} \), there is some weight threshold \( A_d < \frac{1}{4} \) for which our Theorem 21 is stronger than Theorem 29 for all weights larger than \( A_d N \). One way to see this is to note that our Theorem 21 satisfies

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

while the expression in Theorem 29 satisfies

\[
2^{-\left(1-17\left(\frac{4}{n} + \frac{2}{(1-\frac{4}{n})^2}\right)j^{-1}\right)}(\frac{n}{d}) \geq 2^{-\left(1-17j\left(\frac{4}{n}\right)^{-1}\right)}(\frac{n}{d}).
\]

Thus our Theorem 21 is stronger than Theorem 29 whenever \( j \cdot 2^{-j} < 17j \cdot \left(\frac{4}{n}\right)^{-1} \), i.e. whenever

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

For any \( d > \frac{n}{3} \), this gives a nontrivial range.

This concludes our comparison of Theorem 21 with Theorem 29, which was the bound of [SS20] for weights in \([0, \frac{N}{4}]\). We now turn to their bounds for larger weights.

**Theorem 30** ([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 \( \frac{j + \log \frac{1}{2\gamma} \cdot \left(1 - 2\gamma\right)^2}{(1-2\gamma)^2} = o(n) \). Then

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

where \( c(\gamma, j) = O\left(\gamma^2 j + \gamma \log \frac{1}{1-2\gamma} + \gamma\right) \).
This bound holds when the degree is smaller than $\frac{n}{2}$. For arbitrary degree, [SS20] gives the following:

**Theorem 31 ([SS20]).** For any integers $n, d$ and any $\delta > 0$, we have

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

We will start by comparing our Theorem 21 with Theorem 31. Applying Lemma 12, we get from Theorem 21 that

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

$$\leq e^{-\frac{\delta^2}{2} \cdot (\frac{n}{\delta_d})}.$$  

Thus our Theorem 21 is strictly stronger than Theorem 31 for all $d < n$. We will now compare our Theorem 21 with Theorem 30. Applying Lemma 12, we get from Theorem 21 that

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

$$\leq 2^{-\frac{2^{-2j}}{2n^2}}(\frac{n}{\delta_d}).$$  

It follows that our Theorem 21 is stronger than Theorem 30 whenever $2^{-(2j+2)} \geq 2^{-c(\gamma, j)}$, i.e. whenever

$$2j + 2 \leq c(\gamma, j).$$  

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 21 is stronger than Theorem 30 whenever $d > \gamma^* n$. In private correspondence with Amir Shpilka and Ori Sberlo, we learned that $\gamma^*$ can be computed to be $\gamma^* \approx 0.38$.

### A.2 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 $D(C^\perp)$ 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

$$\Pr_{y \sim D(C^\perp)} [ |y| = j ] \leq (1 + o(N^{-1/4})) \frac{(N)_j}{2^N},$$

and suppose that

$$\Pr_{y \sim D(C^\perp)} [ |y| \notin A_\epsilon ] \leq (1 + o(N^{-1/4})) \frac{\sum_{j \notin A_\epsilon} (N)_j}{2^N}.$$ 

Then $C$ is resilient to $\epsilon$-errors.

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

$$\Pr_{\rho \sim P_c} [ d(c + \rho) \neq c ] \leq \Delta_\epsilon + N \max_{S \subseteq \{ \epsilon N \pm 2N^{3/4} \}} \left\{ \frac{2^N}{(N/S)} \mathbb{E}_{y \sim C^\perp} [ \hat{L}_S(y)^2 ] - 1 \right\},$$

where we have defined $\Delta_\epsilon = \left( \frac{1}{1-\epsilon} \right)^{N^{3/4}} + e^{-\frac{N}{8}}$. Now by symmetry of the level function and of the Fourier transform, we have $\hat{L}_S(y) = \hat{L}_S(y')$ for all $|y| = |y'|$. Denoting by $1_j \in \mathbb{F}_2^N$ the vector with 1s in the first $j$ coordinates and 0s in the last $N-j$ coordinates, we can then rewrite the inequality above as

$$\Pr[d(c + \rho) \neq c] \leq \Delta_\epsilon + N \max_{S \subseteq \{ \epsilon N \pm 2N^{3/4} \}} \left\{ \frac{2^N}{(N/S)} \sum_{j=0}^{N} \Pr_{y \sim C^\perp} [ |y| = j ] \hat{L}_S(1_j)^2 - 1 \right\}. \quad (41)$$

Let $\nu$ be such that $h(\nu) = 1 - h(\epsilon) - N^{-1/5}$, and define the set of weights $A = \{ \nu N, \ldots, (1-\nu)N \}$.

We will start by bounding the faraway terms $j \notin A$ in equation (41). We note that $\epsilon N + 2N^{3/4} < \frac{N}{2}$, and so for any $S \subseteq \{ \epsilon N \pm 2N^{3/4} \}$ with $|S| \in \{ 1, 2 \}$ we have $(N/S) \leq 2 (\epsilon N + 2N^{3/4})$. The corresponding Fourier coefficients $\hat{L}_S(1_j)$ can then be crudely bounded as follows:

$$\hat{L}_S(1_j)^2 = \left( 2^{-N/2} \sum_{z \in \mathbb{F}_2^N \setminus \{z\} \in S} (-1)^{\langle z, 1_j \rangle} \right)^2 \leq \left( \frac{(N/S)^2}{2^N} \right) \leq \frac{(N/S) \cdot 2^{(\epsilon N + 2N^{3/4})}}{2^N}.$$ 

We thus get that for any $S \subseteq \{ \epsilon N \pm 2N^{3/4} \}$ with $|S| \in \{ 1, 2 \}$, the faraway terms in summation (41) can be bounded by

$$\frac{2^N}{(N/S)} \sum_{j \notin A} \Pr_{y \sim C^\perp} [ |y| = j ] \hat{L}_S(1_j)^2 \leq \frac{2^N}{(N/S)} \sum_{j \notin A} \Pr_{y \sim C^\perp} [ |y| = j ] \cdot \frac{(N/S) \cdot 2^{(\epsilon N + 2N^{3/4})}}{2^N} \cdot \Pr_{y \sim C^\perp} [ |y| \notin A ]. \quad (42)$$
It remains to bound the central terms in the summation, i.e. the terms in (41) corresponding to \( j \in A = \{ \nu N, \ldots, (1 - \nu) N \} \). For these we will rely on Corollary 18 which states that 
\[
\frac{1}{S} \sum_{j=0}^{N} \binom{N}{j} \cdot \hat{L}_S(1_j)^2 = 1 \text{ for all } S \subseteq \{ 0, \ldots, N \}.
\]
For any such \( S \), we then get
\[
2N \left( \frac{N}{S} \right) \sum_{j \in A} \Pr_{y \sim C^\perp} [ |y| = j ] \hat{L}_S(1_j)^2 \leq 2N \max_{j \in A} \left\{ \Pr_{y \sim C^\perp} [ |y| = j ] \cdot \frac{1}{\binom{N}{j}} \right\} \sum_{j \in A} \binom{N}{j} \cdot \hat{L}_S(1_j)^2 
\]
\[
\leq 2N \cdot \max_{j \in A} \left\{ \Pr_{y \sim C^\perp} [ |y| = j ] \cdot \frac{1}{\binom{N}{j}} \right\}.
\]
(43)

Now from the theorem's assumptions, we know that
\[
\Pr_{y \sim D} [ |y| = j ] \leq (1 + o(N^{-1/4})) \frac{\binom{N}{j}}{2N}
\]
for all \( j \in A \), and that
\[
\Pr_{y \sim D(C^\perp)} [ |y| \not\in A ] \leq (1 + o(N^{-1/4})) \frac{2(\frac{N}{\nu N})}{2N}.
\]
Combining this with Lemma 11 and with our bounds (42) and (43) for the faraway and central terms of the summation respectively, we bound the right-hand side of equation (41) by
\[
\Pr_{\rho \sim P} [ d(c + \rho) \neq c ] \leq \Delta_\epsilon + 3N \left( \frac{N}{\epsilon N + 2N^{3/4}} \right) \frac{2(\frac{N}{\nu N})}{2N} + o(N^{-1/4})
\]
\[
\leq \Delta_\epsilon + 3N \cdot 2^{h(\epsilon + 2N^{-1/4})N} \cdot 2^{h(\nu)N} + o(N^{-1/4}).
\]
By definition of \( \nu \) and by the subadditivity of entropy, we then get
\[
\Pr_{\rho \sim P_c} [ d(c + \rho) \neq c ] \leq \Delta_\epsilon + 3N \cdot 2^{h(\epsilon)N + 2h(N^{-1/4})N} \cdot 2^{-h(\nu)N - N^{3/5}} + o(N^{-1/4})
\]
\[
\leq \left( \frac{\epsilon}{1 - \epsilon} \right)^{N^{3/4}} + e^{-\frac{\sqrt{N}}{10}} + 2^{-N^{4/5} + o(N^{4/5})} + o(N^{-1/4})
\]
\[
\leq o(1).
\]

\[\Box\]

A.3 List-decoding capacity

Claim 32. 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
\[
\Pr_{\rho \sim P_c, c \sim D(C)} [ c \in d_k(c + \rho) ] \geq \frac{3}{4},
\]

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

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

**Proof.** We first note that the theorem condition implies that at least \( \frac{|C|}{2} \) codewords $c \in C$ must satisfy

$$\Pr_{\rho \sim P_\epsilon} [c \in d_k(c + \rho)] \geq \frac{1}{2}. \quad (44)$$

Fix any such $c$. Now from Chernoff’s bound, we have that

$$\Pr_{\rho \sim P_\epsilon} [|\rho| \leq \epsilon N - \epsilon N^{3/4}] \leq \frac{1}{4}. \quad (45)$$

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

$$x \in S_c \implies c \in d_k(x) \quad (45)$$

and

$$\Pr_{\rho \sim P_\epsilon} [\rho \in S_c] \geq \frac{1}{4}. \quad (46)$$

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

$$\Pr_{\rho \sim P_\epsilon} [\rho = c + x] \leq e^{\epsilon N - \epsilon N^{3/4} (1-\epsilon)(1-\epsilon)N + \epsilon N^{3/4}} \leq 2^{-2(1-N^{-1/4})h(\epsilon)N} \quad (47)$$

Equations (46) and (47) 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 equation (45) that any such $c$ must satisfy

$$\left| \{ x \in \mathbb{F}_2^N : c \in d_k(x) \} \right| \geq \frac{2^{(1-N^{-1/4})h(\epsilon)N}}{4}. \quad (48)$$

By double counting, we get

$$2^N \cdot k = \sum_{c \in C} \left| \{ x \in \mathbb{F}_2^N : c \in d_k(x) \} \right| \geq \frac{|C|}{2} \cdot \frac{2^{2(1-N^{-1/4})h(\epsilon)N}}{4} = \frac{1}{8} \cdot 2^{\dim C + h(\epsilon)N - h(\epsilon)N^{3/4}}. \quad (49)$$

The result then follows from rearranging terms.
B Proofs for section 2

B.1 Duals of Transitive Codes - Proof of Fact 8

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 $c_\pi := (c_{\pi(1)}, ..., c_{\pi(N)}) \in C$. Clearly $\pi^{-1}$ 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

$$\sum_k v_k c_{\pi(k)} = 0$$

for all $c \in C$.

We thus have

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

for all $c \in C$

$$\implies \sum_k v_{\pi^{-1}(k)} c_k = 0$$

for all $c \in C$

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

\qed

B.2 Basic Properties of Reed-Muller Codes - Proof of Facts 9 and 10

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 $v$.  

\qed

Fact. For all $n$ and all $d < n$, the Reed-Muller code $\text{RM}(n,d) \subseteq \mathbb{F}_2^N$ 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 $\text{RM}(n,d)$ is the evaluation vector $(f(v_1), ..., f(v_N))$ 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 \((z_{v_1}, ..., z_{v_N}) \in \text{RM}(n, d)\) then \((z_{\pi(v_1)}, ..., z_{\pi(v_N)}) \in \text{RM}(n, d)\)

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 \((z_{v_1}, ..., z_{v_N})\) is a codeword, it can be written as \((z_{v_1}, ..., z_{v_N}) = (f(v_1), ..., f(v_N))\) 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 \((g(v_1), ..., g(v_N))\) must be a codeword. Then since \(g(x) = f \circ \pi(x)\) by definition, we have that \((z_{\pi(v_1)}, ..., z_{\pi(v_N)}) = (f \circ \pi(v_1), ..., f \circ \pi(v_N)) = (g(v_1), ..., g(v_N)) \in \text{RM}(n, d)\).

\(\square\)

### B.3 A version of Pinsker’s inequality - Proof of Lemma [12]

**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.**

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

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

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

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

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

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

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

[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] 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.

[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.

[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.