Advances in Mathematics of Communications

doi:10.3934/amc.2024007

A CRITERION FOR DECODING ON THE BINARY SYMMETRIC CHANNEL

Anup Rao[✉] and Oscar Sprumont[✉*]

School of Computer Science, University of Washington, USA

(Communicated by Sihem Mesnager)

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

1. We give a tight bound on the weight distribution of every transitive linear code $C \subseteq \mathbb{F}_2^N$:
$$\Pr_{c \in C}[\mathrm{wt}(c) = \alpha N] \leq 2^{-(1-h(\alpha))\dim(C)}.$$
2. We give a criterion that certifies that a linear code C can be decoded on the binary symmetric channel. Let $K_s(x)$ denote the Krawtchouk polynomial of degree s, and let $C^\perp$ denote the dual code of C. We show that bounds on $\mathbb{E}_{c \in C^\perp}[K_{\epsilon N}(\mathrm{wt}(c))^2]$ imply that C recovers from errors on the binary symmetric channel with parameter ϵ. Weaker bounds can be used to obtain list-decoding results using similar methods. One consequence of our criterion is that whenever the weight distribution of $C^\perp$ is sufficiently close to the binomial distribution in some interval around $\frac{N}{2}$, C is resilient to ϵ-errors.
3. We combine known estimates for the Krawtchouk polynomials with our weight bound for transitive codes, and with known weight bounds for doubly transitive codes, to obtain list-decoding results for both these families of codes. In some regimes, our bounds for doubly transitive codes achieve the information-theoretic optimal trade-off between rate and list size.

1. **Introduction.** In his seminal 1948 paper, Shannon laid out the bases of coding theory and introduced the concept of channel capacity, which is the maximal rate at which information can be transmitted over a communication channel [52]. The two channels that have received the most attention are the Binary Symmetric Channel (BSC), where each bit is independently flipped with some probability ϵ, and the Binary Erasure Channel (BEC), where each bit is independently replaced by an erasure symbol with some probability ϵ. Shannon's work initiated a decades-long search for explicit codes that can achieve high rates over a noisy channel.

2020 *Mathematics Subject Classification.* 94B05, 94B65, 94B70, 94B15.

Key words and phrases. Linear codes, transitive codes, doubly transitive codes, weight enumerator, error channel, decoding criterion, list decoding.

The first author was supported by NSF CCF-2131899. The second author was supported in part by NSERC PGSD3-545945-2020, NSF CCF-2131899, NSF CCF-1813135 and Anna Karlin's Bill and Melinda Gates Endowed Chair.

*Corresponding author: Oscar Sprumont.

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

Using fundamental results from Fourier analysis about the influences of symmetric boolean functions [29, 55, 14] has led to a very successful line of work, with [38] showing that Reed-Muller codes achieve capacity over the BEC and [47, 2] showing that they achieve capacity over the BSC. In fact, [38] show that for any $s > 1$, if a linear code $C \subseteq \mathbb{F}_2^N$ of size $|C| \leq 2^{(1-\epsilon)N}$ has a doubly-transitive symmetry group G such that for every $S \subseteq \{1, 2, \ldots, N\}$ with $|S| = (s \log N)^{0.99}$, $|\{\pi(S) : \pi \in G\}| \geq N^{s+1}$, then C can tolerate $\epsilon - O(1/s)$ fraction of random erasures [1]. Given these results, it is natural to investigate the types of symmetry that lead to good codes. In this paper, we prove three kinds of results relevant to understanding the error resilience of general linear codes, transitive linear codes, and doubly transitive linear codes.

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

$$\Pr_{c \in C}[\mathrm{wt}(c) = \alpha N] \leq 2^{-(1-h(\alpha))\mathsf{dim}(C)},$$

   where $\mathrm{wt}(c)$ denotes the Hamming weight of the binary vector c. This bound is proven by combining transitivity with the subadditivity of entropy. In some regimes, it improves on all previously known weight bounds for Reed-Muller codes (See Table 1 and Appendix A for a comparison of our weight bound with previous results).
2. We give a new criterion to validate that a code can be decoded over the BSC. For any fixed integers $0 \leq s \leq N$, define the Krawtchouk polynomial of degree s to be the real polynomial

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

   where for any polynomial $p(x)$ we abused notation to write $\binom{p(x)}{j} := \frac{p(x)(p(x)-1)\ldots(p(x)-j+1)}{j!}$. Let $C^\perp$ denote the dual code of C. In spirit, our criterion says that any linear code C satisfying

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

   can be uniquely decoded on the BSC with high probability. Our actual result is a little more technically involved (see Theorem 1.2). This criterion implies that any linear code whose dual codewords are distributed sufficiently close to the binomial distribution must be resilient to ϵ-errors (see Corollary 1.3).

[1] Their result is not stated in this form, but we believe this follows from their analysis.

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

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

Next, we discuss our results more rigorously. 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 finite field. See section 6 for the proof.

Theorem 1.1. *Consider any finite field $\mathbb{F}_q$, and let $C \subseteq \mathbb{F}_q^N$ be any transitive linear code. Then for any $\alpha \in (0,1)$, we have*

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

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

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

Note that $h_2(\alpha)$ denotes the binary entropy function. We note that in some regimes (see Table 1), 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 for a comparison of our Theorem 1.1 with previous weight bounds.

II. a criterion for decoding on the BSC

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

$$P_\epsilon(z) := \epsilon^{\mathrm{wt}(z)}(1-\epsilon)^{N-\mathrm{wt}(z)}$$

for some $\epsilon \in (0, \frac{1}{2})$. Our approach is based on Fourier analysis, although unlike [38] and [25], the ideas we use do not rely on bounds on influences. We obtain the following result.

Theorem 1.2. *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 $\epsilon \in (0, \frac{1}{2})$ satisfying $N > \frac{1}{\epsilon^4(\frac{1}{2}-\epsilon)^4}$, there exists a decoding function $d : \mathbb{F}_2^N \to C$ such that for all $c \in C$ we have*

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

where $\binom{N}{S} := \sum_{j\in S}\binom{N}{j}$, and $K_S(x) := \sum_{j\in S} K_j(x)$ for K_j the Krawtchouk polynomial of degree j, and where $[\epsilon N \pm N^{3/4}]$ denotes the interval $[\epsilon N - N^{3/4}, \epsilon N + N^{3/4}]$.

See section 5 (Theorem 5.2) for the proof. We will now consider one interesting consequence of Theorem 1.2. Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and define

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

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

Corollary 1.3. *Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and let $C \subseteq \mathbb{F}_2^N$ be a linear code. Suppose that for every $j \in A_\epsilon$ we have*

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

and suppose that

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

Then C is resilient to ϵ-errors.

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

III. list decoding results

Using a generalized version of Theorem 1.2 (namely, Theorem 5.2 in section 5), we obtain list decoding bounds for both transitive and doubly transitive codes. We start with our bound for doubly transitive codes (see Section 8 for the proof).

Theorem 1.4. *Fix any $\epsilon \in (0, \frac{1}{2})$ and any $\gamma \leq 1 - \log(1 + 2^{-4\epsilon})$. Then any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1-\gamma)N$ can with high probability list-decode ϵ-errors using a list T of size*

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

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

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

Our bound in Theorem 1.4 shows that doubly transitive codes achieve these optimal parameters, at least in some regimes. (Since the requirement $\gamma \leq 1 - \log(1 + 2^{-4\epsilon})$ can be a bit hard to digest, we note for e.g. that $1.3\epsilon < 1 - \log(1 + 2^{-4\epsilon})$ for all $\epsilon \in (0, \frac{1}{2})$, so Theorem 1.4 implies that any doubly transitive code of rate $\geq 1 - 1.3\epsilon$ achieves the optimal list size for decoding ϵ-errors). We now turn to our list-decoding bound for transitive codes (see section 7 for the proof).

Theorem 1.5. *Fix any $\epsilon \in (0, \frac{1}{2})$ and $\eta \in (0, 1)$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode ϵ-errors using a list T of size*

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

As an explicit example of the types of bounds one gets from Theorem 1.5, we have that any transitive linear code of dimension $\dim C = (1 - \frac{4\epsilon}{e})N$ can with high probability list-decode ϵ-errors using a list T of size

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

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

1.1. **Techniques.** Our weight distribution bound for transitive linear codes (Theorem 1.1) is proven by showing that the entropy of a uniformly random codeword of weight αN is small. To do this, we analyze the entropy of the coordinates corresponding to linearly independent columns of the generator matrix. Transitivity implies that every coordinate in the code has the same entropy, and subadditivity of entropy can then be used to bound the entropy of the entire distribution.

To obtain our decoding criterion, we make use of a connection between the probability of a decoding error and the ℓ_2 norm of the coset distribution of the code. To explain the intuition, let us start by assuming that exactly ϵN of the coordinates in the codeword are flipped, although our results actually hold over the BSC as well. Let z be the vector in $\mathbb{F}_2^N$ that represents the errors introduced by the channel, and let H be the parity check matrix of the code. Then by standard arguments, if z can be recovered from $Hz^\intercal$, the codeword can be decoded. In the case where z is uniformly distributed on vectors of weight ϵN, this amounts to showing that with high probability, the coset of z does not contain any other string of weight ϵN (in other words, there is no $w \in \mathbb{F}_2^N$, $w \neq z$ of weight $\mathrm{wt}(w) = \epsilon N$ such that $Hz^\intercal = Hw^\intercal$). This can be understood by computing the norm

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

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

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

$$\|f\|_2^2 = \frac{1}{\binom{N}{\epsilon N}^2} \sum_{j=0}^{N} \Pr[\mathrm{wt}(c^\perp) = j] \cdot K_{\epsilon N}(j)^2, \tag{3}$$

where $c^\perp$ is a uniformly random codeword in the dual code and $K_{\epsilon N}$ is the Krawtchouk polynomial of degree ϵN. We note that such relations for the coset weight distribution have been used to understand the discrepancy of subsets of the sphere, as well as subsets of other homogeneous spaces. In particular, (3) was proven in a slightly different form in [9] (see Theorem 2.1 and Lemma 4.1), whereas over $\mathbb{R}^N$ results of this type had previously been derived in [12, 53].

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

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

Our decoding criteria (Theorem 1.2, Corollary 1.3) are thus obtained by bounding the norm $\|f\|_2^2$ using estimates for Krawtchouk polynomials and for the weight distribution of the dual code $C^\perp$. Our list-decoding results (Theorems 1.4 and 1.5) then follow from our weight bound for transitive codes (Theorem 1.1) and from a weight bound of Samorodnitsky for doubly transitive codes (Theorem 1.6).

1.2. **Related work.** It has been shown that LDPC codes achieve capacity over Binary Memoryless Symmetric Channels (BMS) [42, 39, 18], which includes both the BSC and the BEC. These constructions are not deterministic, and it is only with the advent of polar codes [7] 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 $\mathbb{F}_2$ [7]. For this reason, when Arikan showed that polar codes achieve capacity over the BSC, Reed-Muller codes received renewed attention from the coding theory community. A long and fruitful line of work [4, 38, 6, 25, 1, 47, 50] has recently culminated in Abbe and Sandon showing that Reed-Muller codes achieve capacity over all BMS channels [2].

One of the key properties of Reed-Muller codes, which is strongly leveraged in all the papers above, is that they are doubly transitive. In fact, Kudekar, Kumar, Mondelli, Pfister, Sasoglu and Urbanke showed that any doubly transitive linear code achieves bit-decoding capacity over the BEC [38], 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). An important open question is thus whether general doubly transitive codes achieve capacity over all BMS channels under block-MAP decoding, or whether one really needs the additional symmetry that Reed-Muller codes possess. Some of the key techniques used in [47] and [2] are very much tailored to Reed-Muller codes, or at least to codes consisting of evaluations of polynomials over $\mathbb{F}_2^N$; in order to prove the same results for arbitrary doubly transitive codes, it may be necessary to develop a more general framework.

Weight bounds for doubly transitive codes

As far as we know, there were no previously known weight bound for general transitive linear codes. There are however two known weight bounds for doubly transitive codes (which we'll give here), as well as many known weight bounds for Reed-Muller codes (which we'll discuss in the next section). We compare all these results in Appendix A. We state below the weight bounds of Samorodnitsky, which to the best of our knowledge are the only previously known weight bounds for doubly transitive codes.

TABLE 1. Best known upper bounds on the number of codewords of weight w in $\mathsf{RM}(n,d)$

	$o(n) \le d \lesssim 0.38n$	$0.38n \lesssim d \le \frac{n}{2} - \tilde{\Omega}(\sqrt{n})$	$d = \frac{n}{2} \pm \tilde{O}(\sqrt{n})$
$o(N) \le w < \tau N^*$	$2^{O\left(\binom{n}{\le d}\left(\frac{d}{n}\right)^{\lceil \log \frac{N}{w} \rceil} \log \frac{N}{w}\right)}$ [51]	← as previous	$\left(\frac{1}{2^{1-\binom{n}{\le d}/N}-1}\right)^{w+o(N)}$ [48]**
$\tau N \le w < \frac{N}{4}$	$2^{h\left(\frac{w}{N}\right)\binom{n}{\le d}}$ (Theorem 1.1)	← as previous	as above ↑
$\frac{N}{4} \le w \le \frac{N}{2} - o(N)$	$2^{(1-2^{-O(\log \frac{N}{w})})\binom{n}{\le d}}$ [51]	$2^{h\left(\frac{w}{N}\right)\binom{n}{\le d}}$ (Theorem 1.1)	as above ↑
$(1-2^{\binom{n}{\le d}/N-1})N \le w \le \frac{N}{2}$			$\binom{N}{w} \cdot 2^{\binom{n}{\le d}-N+o(N)}$ [48]

*τ is a threshold that depends on $\frac{d}{n}$. See (4) and the surrounding discussion.

**Unless $w \ge (1-2^{\frac{\binom{n}{\le d}}{N}-1})N$, in which case see row below dashed line.

Theorem 1.6 (Proposition 1.4 in [48]). *Let $C \subseteq \mathbb{F}_2^N$ be a doubly transitive linear code of rate $\eta := \frac{\dim C}{N}$. For any $j \in \{1,2,\dots,N\}$, define $j^* := \min\{j, N-j\}$. Then for any $j \in \{1,2,\dots,N\}$,*

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

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

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

Weight bounds for reed-muller codes

As we mentioned earlier, one specific family of doubly transitive codes that has received a lot of attention is the family of Reed-Muller codes. Several past works have proven bounds on their weight distribution. We give here a brief history of these results, although for space reasons (there are over 10 different weight bounds), we will not state them here. We delve deeper into some prior results in Appendix A, where we compare them to our weight bound of Theorem 1.1. We also refer the reader to [5, 3] for a discussion on the subject, as well as a thorough exposition to Reed-Muller codes.

The earliest work we are aware of is that of Sloane and Berlekamp, who characterized all codewords in Reed-Muller codes of degree 2 [54]. For arbitrary degree, Kasami and Tokura then characterized all codewords of weight smaller than twice the minimum distance [31], before Kasami, Tokura and Azumi improved this characterization to include all codewords of weight up to 2.5 times the minimum distance [32].

A few decades later, Kaufman, Lovett and Porat gave asymptotically tight bounds on the weight distribution of Reed-Muller codes of constant degree [33]. Abbe, Shpilka and Wigderson then built on these techniques to obtain bounds for all degrees smaller than $\frac{n}{4}$ [4], before Sberlo and Shpilka again improved the approach to obtain bounds for all degrees [51]. Most recently, Samorodnitsky used completely different ideas to obtain weight bounds for codes of constant rate [49, 48] (see previous section).

The bounds mentioned above are strong when $j/N \ll 1/2$. For j/N close to $1/2$, the first results we are aware of are due to Ben-Eliezer, Hod and Lovett [11]. Their bounds were extended to Reed-Muller codes over prime fields by Beame, Oveis Gharan and Yang [10]. Sberlo and Shpilka then obtained the first results to hold for all degrees in [51], while Samorodnitsky again obtained bounds for codes of constant rate in [48].

We summarize in Table 1 the best known upper bounds on the weight distribution of Reed-Muller codes. We note that in some regimes, our Theorem 1.1 improves on all the aforementioned weight bounds. See Appendix A for some details; see also [3], section 4.

In Table 1, τ is a threshold that depends on $\frac{d}{n}$. We show for e.g. that

$$\tau \leq \frac{1}{2} \cdot 2^{-\frac{\log 17}{\log \frac{n}{2d}}} \tag{4}$$

in Appendix A, which is below the trivial $\frac{1}{4}$ for any $\frac{d}{n} > \frac{1}{34}$. We note that when $\frac{d}{n}$ is small enough (smaller than some constant), then $\tau = \frac{1}{4}$.

List decoding

List decoding was proposed by Elias in 1957 as an alternative to unique decoding [17]. 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 [21, 40, 20]. Folded Reed-Solomon codes were the first explicit codes to provably achieve list-decoding capacity [22], followed by several others a few years later [23, 34, 26, 45, 15]. For the rest of this paper however, we will exclusively work in the model where the errors are stochastic. In this model, as far as we know, there was no known list-decoding bound for transitive codes prior to our Theorem 1.5. For doubly transitive codes, the strongest previously known list decoding bound was, to the best of our knowledge, that any doubly transitive code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1-\gamma)N$ can list-decode ϵ-errors with a list T of size

$$|T| = 2^{\epsilon N \log \frac{4\epsilon(1-\epsilon)}{(2\gamma-1)^2} + o(N)}. \tag{5}$$

This result, although not explicitly stated in [48], can be obtained from his weight bound of Theorem 1.6 by bounding the expected number of codewords that end up closer to the received string than the original codeword, and then applying Markov's inequality. We summarize in Table 2 our list-decoding results and compare them to previous work. We note that the previously known bound for doubly transitive codes stays strictly above the optimal size of $2^{h(\epsilon)N - \gamma N \pm o(N)}$ (see Appendix D.3).

Following the publication of the present paper on arxiv, Hazla showed in [24] that any code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1-\gamma)N \geq (1-4\epsilon(1-\epsilon))N$ that achieves capacity over the BEC can list-decode ϵ-errors with a list T of size

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

TABLE 2. Upper bounds on the list size needed for a code of rate $1-\gamma$ to recover from ϵ-errors

	Previous work	**Our results**	**Information-theoretic lower bound**
Transitive codes (any γ)	-	$2^{\epsilon N \log(\frac{2}{\gamma})+o(N)} + 2^{4\epsilon N+o(N)}$	$2^{h(\epsilon)N-\gamma N-o(N)}$
Doubly transitive codes (any $\gamma \leq 1-\log(1+2^{-4\epsilon})$)	$2^{\epsilon N \log \frac{4\epsilon(1-\epsilon)}{(2^\gamma-1)^2}+o(N)}$	$2^{h(\epsilon)N-\gamma N+o(N)}$	$2^{h(\epsilon)N-\gamma N-o(N)}$

Krawtchouk polynomials

Fix any non-negative integers N and $s \leq N$. The Krawtchouk polynomial of degree s is the real polynomial

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

where for any polynomial $p(x)$ we defined $\binom{p(x)}{j} := \frac{p(x)(p(x)-1)\ldots(p(x)-j+1)}{j!}$. For any subset $S \subseteq \{0,1,\ldots,N\}$, we will be interested in the polynomial $K_S(x) := \sum_{s\in S} K_s(x)$. For $v \in \mathbb{F}_2^N$, we will sometimes abuse notation and use $K_S(v)$ to mean $K_S(\mathrm{wt}(v))$, where $\mathrm{wt}(v)$ denotes the Hamming weight of v. The following proposition follows from standard results (see for instance [37], or Theorem 16 in [44]).

Proposition 1.7. *For any N and any $S \subseteq \{0,1,\ldots,N\}$, we have*

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

Good estimates for Krawtchouk polynomials of any degree were obtained in [30, 27, 46] (see for e.g. [46], Lemma 2.1). These estimates are asymptotically tight in the exponent. Note that $|K_s(x)| = |K_s(N-x)| = |K_{N-s}(x)|$ by symmetry (see for e.g. equations (2.8) and (2.9) in [46]), so it suffices to understand the case $x, s \leq \frac{N}{2}$.

Theorem 1.8 ([30, 27, 46])**.** *Let $\epsilon, \delta \in (0, \frac{1}{2})$ be arbitrary. If $\delta \geq \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$, then*

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

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

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

As the second expression can be somewhat cumbersome to use, [46] also gives the following weaker bound.

Theorem 1.9 (Lemma 2.2 and equation 2.10 in [46])**.** *For any $\epsilon \in (0, \frac{1}{2})$ and any $\delta < \frac{1}{2} - \sqrt{\epsilon(1-\epsilon)}$, we have*

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

We will need the above estimate when using our Theorem 1.2 to obtain list-decoding results for transitive and doubly transitive codes.

Relations between a code and its dual

Several connections have been established between the properties of a code $C \subseteq \mathbb{F}_2^N$ and those of its dual $C^\perp$. MacWilliams proved in [43] the MacWilliams identities, relating the weight distributions of C and $C^\perp$ by

$$\sum_{c \in C} (1+z)^{N-\mathrm{wt}(c)}(1-z)^{\mathrm{wt}(c)} = |C| \sum_{c \in C^\perp} z^{\mathrm{wt}(c)},$$

where z is an indeterminate. Krasikov and Litsyn then bounded the weight distribution of any linear code with large dual distance [35, 36], while Ashikhmin, Honkala, Laihonen and Litsyn derived bounds for the covering radius of any such code [8]. To the best of our knowledge however, the present paper is the first work to relate the decoding performance of a code $C \subseteq \mathbb{F}_2^N$ to the weight distribution of its dual. As far as we know, there is no known way to apply the results mentioned above to obtain a unique-decoding criterion like our Theorem 1.2 or our Corollary 1.3.

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

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

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

$$\begin{aligned} D_k(x) &:= \underset{\substack{\{z_1, z_2 \ldots, z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i{}^\intercal = x^\intercal \text{ for all } i}}{\operatorname{argmax}} \{P_\epsilon(z_1) + P_\epsilon(z_2) + \ldots + P_\epsilon(z_k)\} \\ &= \underset{\substack{\{z_1, z_2 \ldots, z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\intercal = x^\intercal \text{ for all } i}}{\operatorname{argmin}} \{\mathrm{wt}(z_1) + \mathrm{wt}(z_2) + \ldots + \mathrm{wt}(z_k)\}, \end{aligned} \tag{6}$$

where ties are broken according to the lexicographic order. The following standard lemma (see for e.g. page 17, Theorem 5 in [44]) states that if the max-likelihood decoder is able to identify the error string ρ, then it is possible to recover the original codeword.

Lemma 2.1. *Let H be the $t \times N$ parity-check matrix of the linear code C, and let $D : \mathbb{F}_2^t \to \left(\mathbb{F}_2^N\right)^{\otimes k}$ be arbitrary. Then there exists a decoder*

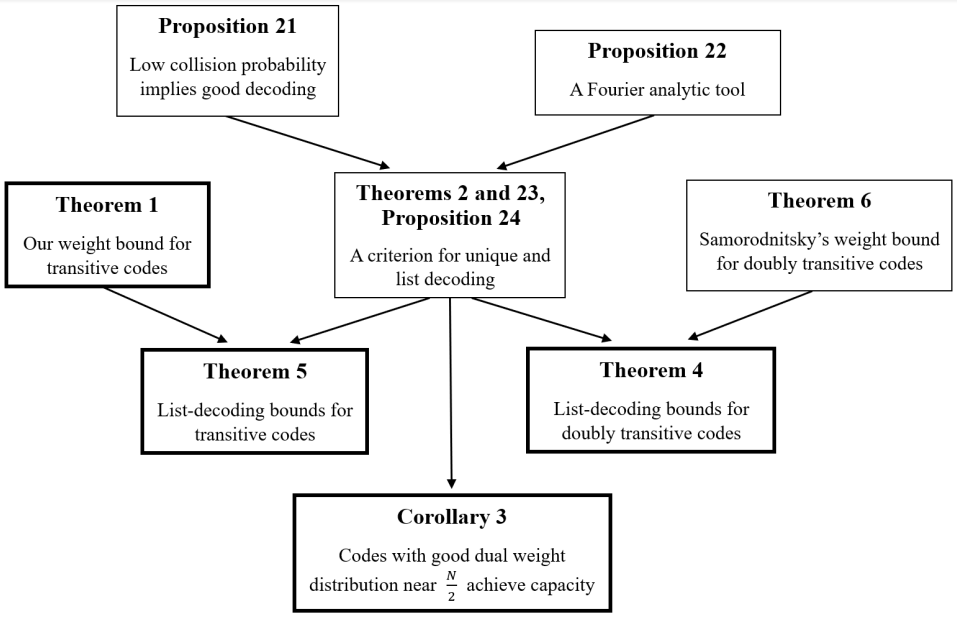
$$d : \mathbb{F}_2^N \to C^{\otimes k}$$

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

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

From this point onward, our goal will thus be to prove that the max-likelihood decoder in (6) succeeds in recovering ρ with high probability. In section 4, we relate

FIGURE 1. Organization of our paper and connections between our results.



the decoding error probability of the max-likelihood decoder D_k to the collision probability

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

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 1.6 in section 8 to obtain stronger list-decoding bounds for doubly transitive codes.

See Figure 1 for a description of the connections between our various propositions and theorems.

3. **Notation, conventions and preliminaries.** For the sake of conciseness, we will use the notation

$$[a \pm l] := [a - l, a + l]$$

for intervals, the notation

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

for binomial coefficients, and for $S \subseteq \{0, 1, \ldots, N\}$ the notation

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

We denote the set of all non-negative integers by

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

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 $[N] := \{1, 2, \dots, N\}$, by ordering the elements of $\mathbb{F}_2^n$ lexicographically. For $x \in \mathbb{F}_2^N$, we write

$$\mathrm{wt}(x) := |\{j \in [N] : x_j = 1\}|$$

to denote the Hamming weight of x.

3.1. **Coding theory definitions and terminology.** An N-bit code is a subset $C \subseteq \mathbb{F}_2^N$, and we call any element $c \in C$ a *codeword* of C. Throughout the paper, we will use N to denote the length of the code, i.e. the number of bits in any given codeword.

We will be interested in the performance of various codes over the so-called Binary Symmetric Channel (BSC for short). When a codeword $c \in C$ is sent through the Binary Symmetric Channel, each one of its bits is flipped independently at random with probability ϵ, for some $\epsilon \in (0, \frac{1}{2})$. Throughout the paper, we will use ϵ to denote this error probability, and we will use ρ to denote the vector $(\rho_1, \rho_2, \dots, \rho_N)$ whose i^{th} coordinate is 1 with probability ϵ and 0 with probability $1 - \epsilon$, for all $i \in \{1, 2, \dots N\}$. We will call the original codeword $c \in C$ the *transmitted codeword*, we will call the noisy vector ρ the *error string*, and we will call $c + \rho$ the *received message*.

We say that the code C is *resilient to ϵ-errors* if there exists a decoding function $d : \mathbb{F}_2^N \to C$ such that for every $c \in C$, with high probability over the choice of an ϵ-noisy error string ρ we have

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

We will also be interested in the performance of a code with respect to list decoding. In this set-up, the decoder is now a function $d : \mathbb{F}_2^N \to C^{\otimes k}$. We say that a code C can list-decode ϵ-errors with a list size of k if with high probability (again, over the choice of an ϵ-noisy error string ρ), we have

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

Throughout the paper, we will denote by k the size of the list. We note that the unique decoding problem can be seen as setting $k = 1$ in the list decoding set-up.

3.2. **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 $\dim\ \mathrm{C} \times N$ matrix G whose rows form a basis of C. The matrix G generates all codewords of C in the sense that

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

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

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

One property that will play an important role in our analysis is transitivity, which we define below.

Definition 3.1. A code $C \subseteq \mathbb{F}_2^N$ is transitive if for every $i, j \in [N]$ there exists a permutation $\pi : [N] \to [N]$ such that

(i) $\pi(i) = j$
(ii) For every element $v = (v_1, v_2, \ldots, v_N) \in C$ we have $(v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(N)}) \in C$.

Many well-known and widely used codes are transitive, for e.g. Reed-Muller codes, Reed-Solomon codes, general BCH codes, and all cyclic codes. In addition, Reed-Muller codes and extended primitive narrow-sense BCH codes are doubly transitive.

Definition 3.2. A code $C \subseteq \mathbb{F}_2^N$ is doubly transitive if for every $i, j, k, \ell \in [N]$ with $i \neq k$ and $j \neq \ell$, there exists a permutation $\pi : [N] \to [N]$ such that

(i) $\pi(i) = j$ and $\pi(k) = \ell$
(ii) For every element $v = (v_1, v_2, \ldots, v_N) \in C$ we have $(v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(N)}) \in C$.

For a review on doubly transitive codes, see [28] We note that the dual code of a transitive code is transitive, and that the dual code of a doubly transitive code is doubly transitive (see Appendix D.2 for the proof).

Claim 3.3. *The dual code $C^\perp$ of a transitive code $C \subseteq \mathbb{F}_2^N$ is transitive.*

Claim 3.4. *The dual code $C^\perp$ of a doubly transitive code $C \subseteq \mathbb{F}_2^N$ is doubly transitive.*

3.3. **Reed-muller codes.** We will denote by $\mathsf{RM}(n, d)$ the Reed-Muller code with n variables and degree d. The codewords of the Reed-Muller code $\mathsf{RM}(n, d)$ are the evaluation vectors (over all points in $\mathbb{F}_2^n$) of all multivariate polynomials of degree $\leq d$ in n variables. The dimension of the code is known to be $\binom{n}{\leq d}$. (See for e.g. page 5 of [5]).

Fact 3.5. *The dimension of the Reed-Muller code $\mathsf{RM}(n, d)$ is*

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

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

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

If S is empty, this entry is set to 1. The parity-check matrix of the Reed-Muller code is known to be the same as the generator matrix of a different Reed-Muller code. Namely, let H be the $\binom{n}{\leq n-d-1} \times N$ generator matrix for the code $\mathsf{RM}(n, n-d-1)$. Then H has full rank, and $MH^\intercal = 0$. So, the rows of H are a basis for the orthogonal complement of the span of the rows of M. Reed-Muller codes also have well-known algebraic features, notably transitivity (see for e.g. Lemma 23 in [38]).

Fact 3.6. *For all non-negative integers n and $d \leq n$, the Reed-Muller code $\mathsf{RM}(n, d)$ is transitive.*

3.4. **Entropy.** The binary entropy function $h : [0,1] \to [0,1]$ is defined to be

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

One useful property of the binary entropy function is that it is subadditive.

Lemma 3.7. *For any $x \in [0,1]$ and any $y \in [0, 1-x]$, we have*

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

This is because the binary entropy function is concave, and any concave, positive function is subadditive (see for e.g. [41], page 83, statement 103). The entropy function can be used to approximate binomial coefficients.

Lemma 3.8. *For any integer $d \in \{1, 2, \dots, \frac{n}{2}\}$, we have*

$$\frac{1}{\sqrt{2n}} \cdot 2^{h(d/n) \cdot n} \leq \binom{n}{d} \leq \binom{n}{\leq d} \leq 2^{h(d/n) \cdot n}.$$

See for e.g. page 309, Lemma 7 in [44] for the proof of the leftmost inequality, and Theorem 3.1 in [19] for the proof of the rightmost inequality.

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

Lemma 3.9. *For any $\mu \in (0,1)$, we have*

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

See Appendix D.4 for the proof.

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

$$P_\epsilon(z) := \epsilon^{\mathrm{wt}(z)} (1-\epsilon)^{N - \mathrm{wt}(z)}.$$

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 $\mathrm{wt}(z) \in S$, for some $S \subseteq \{0, 1, \dots, N\}$. We will denote this probability distribution by

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

3.6. **Probability theory.** We will need two very standard results of probability theory (see for e.g. [13]): Markov's inequality and Chernoff's bound. We start with Markov's inequality.

Lemma 3.10. *Let X be a non-negative random variable. Then for any $a > 0$, we have*

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

We will also need Chernoff's bound:

Lemma 3.11. *Let $X_1, X_2, \ldots, X_m$ be i.i.d. random variables taking values in $\{0,1\}$, and define $X := X_1 + X_2 + \ldots + X_m$. Then for any $\delta \in (0,1)$, we have*

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

3.7. **Fourier analysis.** The Fourier basis is a useful basis for the space of functions mapping $\mathbb{F}_2^N$ to the real numbers. We recall some of its properties below (see for e.g. [16]). For $f, g \in \mathbb{F}_2^N \to \mathbb{R}$, define the inner product

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

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

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

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

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

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

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

In particular,

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

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

$$P_\epsilon(z) := \epsilon^{\mathrm{wt}(z)}(1-\epsilon)^{N - \mathrm{wt}(z)}.$$

Recall also that for any subset $S \subseteq \{0, 1, \ldots, N\}$, we denote by λ_S the uniform distribution over all strings $z \in \mathbb{F}_2^N$ of weight $\mathrm{wt}(z) \in S$, i.e.

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

The goal of this section will be to analyze the relationship between the decoding of an error string $\rho \in \mathbb{F}_2^N$ and the collision probability of strings $z \in \mathbb{F}_2^N$ within the map $z \mapsto Hz^\intercal$. Intuitively, the more collisions there are within this mapping, the harder it is for our decoder to correctly identify the error string ρ upon seeing only its image $H\rho^\intercal$. However, certain error strings might be unlikely enough to

occur that our decoder can safely ignore them. For example, if we are interested in an ϵ-noisy error string ρ, then ρ is unlikely to have weight $\mathrm{wt}(\rho)$ far away from ϵN. We could thus choose to ignore all strings whose weights do not lie in the set $S = [\epsilon N \pm l] \cap \mathbb{N}$, for some integer l. In order to analyze the collisions that occur when strings are required to have weight $\mathrm{wt}(z) \in S$, we define for every $z \in \mathbb{F}_2^N$ and every $S \subseteq \{0, 1, \dots, N\}$ the set of S-colliders of z, i.e. the set of strings y that lie in the coset of z and have weight $\mathrm{wt}(y) \in S$:

Definition 4.1. For any $z \in \mathbb{F}_2^N$, any matrix H with N columns and entries in $\mathbb{F}_2$, and any subset $S \subseteq \{0, 1, \dots, N\}$, define

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

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

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

Definition 4.2. For any subset $S \subseteq \{0, 1, \dots, N\}$ and any matrix H with N columns and entries in $\mathbb{F}_2$, define

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

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

Lemma 4.3. *For any subset $S \subseteq \{0, 1, \dots, N\}$, any matrix H with N columns and entries in $\mathbb{F}_2$, and any integer $k \geq 1$, we have*

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

Proof. Note that $|\Omega_z^S| \geq 1$ for any $z \in \mathbb{F}_2^N$ with weight $\mathrm{wt}(z) \in S$, so the random variable $|\Omega_\rho^S| - 1$ is always non-negative. Applying Markov's inequality (i.e. Lemma 3.10), we then have

$$\begin{aligned} \Pr_{\rho \sim \lambda_S} \left[|\Omega_\rho^S| > k \right] &= \Pr_{\rho \sim \lambda_S} \left[|\Omega_\rho^S| - 1 \geq k \right] \\ &\leq \frac{\mathsf{Coll}_H(S) - 1}{k}. \end{aligned}$$

□

When the error string ρ is sampled uniformly at random from the set $\{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S\}$, the above lemma allows us to relate the decoding error probability to the collision count $\mathsf{Coll}_H(S)$. The problem we are most interested in, however, is when ρ is sampled not from some uniform distribution, but from the ϵ-noisy probability distribution P_ϵ. We will now show how to connect these two decoding problems. The intuition is that by the Chernoff bound, we only need to concern

ourselves with strings whose weights lie in $S = [\epsilon N \pm l] \cap \mathbb{N}$, for some appropriately chosen l. But in this weight band all strings have similar weight, and so are given similar probability under the distribution P_ϵ. Intuitively, the P_ϵ-decoder must then perform very similarly to the λ_S-decoder. The following proposition makes this idea precise, and then uses Lemma 4.3 to bound the probability of a decoding error. Recall that $D_k : \mathbb{F}_2^t \to (\mathbb{F}_2^N)^{\otimes k}$ is the max-likelihood decoder

$$D_k(x) := \operatorname*{argmin}_{\substack{\{z_1, z_2 \ldots, z_k\} \subseteq \mathbb{F}_2^N \\ Hz_i^\intercal = x^\intercal \text{ for all } i}} \{\mathrm{wt}(z_1) + \mathrm{wt}(z_2) + \ldots + \mathrm{wt}(z_k)\},$$

where ties are broken according to the lexicographic order.

Proposition 4.4. *Let H be any matrix with N columns and entries in $\mathbb{F}_2$. Consider any noise parameter $\epsilon \in (0, \frac{1}{2})$ and any $l \in [1, \min\{\epsilon N, (\frac{1}{2} - \epsilon)N\}]$. Then*

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

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + 4(l+1) \max_{\substack{S \subseteq [\epsilon N \pm l] \cap \mathbb{N} \\ 1 \leq |S| \leq 2}} \Big\{ \mathsf{Coll}_{\mathsf{H}}(S) - 1 \Big\}.$$

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

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{w \in [\epsilon N \pm l] \cap \mathbb{N}} \Big\{ \mathsf{Coll}_{\mathsf{H}}(w) - 1 \Big\}.$$

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

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

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

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

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

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

we have by Chernoff's bound (i.e. Lemma 3.11) that

$$\begin{aligned} \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] &\leq \Pr_{\rho \sim P_\epsilon}[\rho \notin B] + \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal) \big| \rho \in B] \\ &\leq 2e^{-\frac{l^2}{3\epsilon N}} + \Pr_{\rho \sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal) \big| \rho \in B]. \end{aligned} \tag{7}$$

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

$\mathrm{wt}(z) \in S(\rho)$. Recalling the definition $\Omega_\rho^S := \{z : Hz^\intercal = H\rho^\intercal, \mathrm{wt}(z) \in S\}$, we can then rewrite our equation (7) as

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

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

$$\begin{aligned}\Pr_{\rho\sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] &\leq 2e^{-\frac{l^2}{3\epsilon N}} + \max_{w\in[\epsilon N\pm l]\cap\mathbb{N}}\left\{\Pr_{\rho\sim P_\epsilon}\left[|\Omega_\rho^{S(\rho)}| > 1 \middle| \mathrm{wt}(\rho) = w\right]\right\} \\ &\leq 2e^{-\frac{l^2}{3\epsilon N}} \\ &\quad + (2l+1)\max_{\substack{w,w'\in[\epsilon N\pm l]\cap\mathbb{N}\\ w'\leq w}}\left\{\Pr_{\rho\sim P_\epsilon}\left[|\Omega_\rho^{\{w,w'\}}| > 1 \middle| \mathrm{wt}(\rho) = w\right]\right\}.\end{aligned}$$

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

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

But by basic conditional probability we know that

$$\Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[|\Omega_\rho^{\{w,w'\}}| > 1\right] \geq \Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[\mathrm{wt}(\rho) = w\right]\Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[|\Omega_\rho^{\{w,w'\}}| > 1 \middle| \mathrm{wt}(\rho) = w\right],$$

so we can bound our previous expression by

$$\begin{aligned}\Pr_{\rho\sim P_\epsilon}[\rho \neq \tilde{D}_1(H\rho^\intercal)] &\leq 2e^{-\frac{l^2}{3\epsilon N}} \\ &\quad + (2l+1)\max_{\substack{w,w'\in[\epsilon N\pm l]\cap\mathbb{N}\\ w'\leq w}}\left\{\frac{\Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[|\Omega_\rho^{\{w,w'\}}| > 1\right]}{\Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[\mathrm{wt}(\rho) = w\right]}\right\}.\end{aligned} \tag{8}$$

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

$$\Pr_{\rho\sim\lambda_{\{w,w'\}}}\left[\mathrm{wt}(\rho) = w\right] = \frac{\binom{N}{w}}{\binom{N}{\{w,w'\}}} \geq \frac{\binom{N}{w}}{\binom{N}{w}+\binom{N}{w'}} \geq \frac{1}{2}. \tag{9}$$

We note that the inequality above holds for both the case $w \neq w'$ and the case $w = w'$. (When $w = w'$, we have $\binom{N}{\{w,w'\}} = \binom{N}{w} \leq \binom{N}{w} + \binom{N}{w'}$). It then follows from (8) and (9) that

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

The theorem statement then follows from Lemma 4.3.

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

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

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

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

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

we have by Chernoff's bound (i.e. Lemma 3.11) that

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

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

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

Applying Lemma 4.3, we get

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

□

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

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

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

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

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

that Proposition 5.1 was previously proven in a different form in [9] (see Theorem 2.1 and Lemma 4.1), and can be seen as describing the coset weight distribution of the code. Recall that we use μ_t to denote the uniform distribution over all vectors in $\mathbb{F}_2^t$, and that we use the notation $\binom{N}{S} := \sum_{s \in S} \binom{N}{s}$.

Proposition 5.1. *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 $S \subseteq \{0, 1, \dots, N\}$, we have*

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

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

$$\frac{1}{2^t} \sum_{x \in \mathbb{F}_2^t} f(x)^2 = \sum_{y \in \mathbb{F}_2^t} \hat{f}(y)^2. \tag{10}$$

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

$$\begin{aligned}
\mathsf{Coll}_H(S) &:= \frac{1}{\binom{N}{S}} \sum_{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S} |\Omega_z^S| \\
&= \binom{N}{S} \sum_{z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S} \frac{1}{|\Omega_z^S|} \Pr_{a \sim \lambda_S} [Ha^\intercal = Hz^\intercal]^2 \\
&= \binom{N}{S} \sum_{x \in \mathbb{F}_2^t} \Pr_{z \sim \lambda_S} [Hz^\intercal = x^\intercal]^2.
\end{aligned}$$

We are now ready to apply Parseval's Identity. Letting $f(x) = \Pr_{z \sim \lambda_S}[Hz^\intercal = x^\intercal]$ in equation (10), we get

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

But by definition of the Fourier transform, we have

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

so our previous equation can be rewritten as

$$\begin{aligned}
&\mathsf{Coll}_H(S) \\
&= 2^t \binom{N}{S} \sum_{y \in \mathbb{F}_2^t} \left(2^{-t} \sum_{x \in \mathbb{F}_2^t} \frac{1}{\binom{N}{S}} (-1)^{y \cdot x^\intercal} \cdot \left| \left\{ z \in \mathbb{F}_2^N : \mathrm{wt}(z) \in S \text{ and } Hz^\intercal = x^\intercal \right\} \right| \right)^2 \\
&= 2^{-t} \frac{1}{\binom{N}{S}} \sum_{y \in \mathbb{F}_2^t} \Big(\sum_{\substack{z \in \mathbb{F}_2^N \\ \mathrm{wt}(z) \in S}} (-1)^{y \cdot Hz^\intercal} \Big)^2.
\end{aligned} \tag{11}$$

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

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

where we used the convention that $\binom{a}{b} = 0$ when $a < b$. Combining this with equation (11), we get

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

□

We will now combine Propositions 4.4 and 5.1 to obtain Theorem 1.2, i.e. to obtain a bound on the decoding error probability in terms of Krawtchouk polynomials. We prove a generalized version of Theorem 1.2 below. To recover Theorem 1.2, set the list size $k = 1$ and set $l = N^{3/4}$, and apply Lemma 2.1. (You want to think of the parameter l as being $l >> \sqrt{N}$ in both the case $k = 1$ and the case $k > 1$, so that the error term $e^{-\frac{l^2}{3\epsilon N}}$ is small).

Theorem 5.2. *Let H be any $t \times N$ matrix with entries in $\mathbb{F}_2$. Consider any noise parameter $\epsilon \in (0, \frac{1}{2})$ and any $l \in [1, \min\{\epsilon N, (\frac{1}{2} - \epsilon)N\}]$. Then*

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

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_1(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + 4(l+1) \max_{\substack{S \subseteq [\epsilon N \pm l] \cap \mathbb{N} \\ 1 \leq |S| \leq 2}} \left\{ \frac{1}{\binom{N}{S}} \mathop{\mathbb{E}}_{v \sim \mu_t} \left[K_S(vH)^2 \right] - 1 \right\}.$$

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

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{l^2}{3\epsilon N}} + \frac{4(l+1)}{k} \max_{w \in [\epsilon N \pm l] \cap \mathbb{N}} \left\{ \frac{1}{\binom{N}{w}} \mathop{\mathbb{E}}_{v \sim \mu_t} \left[K_w(vH)^2 \right] - 1 \right\}.$$

Proof. The theorem statement follows directly from Propositions 4.4 and 5.1. □

One consequence of Theorem 5.2 is Corollary 1.3, which states that C is resilient to ϵ-errors if the weight distribution of $C^\perp$ is close enough to the binomial distribution (see Appendix B for the proof). As another application of Theorem 5.2, we present the following bound on the probability of making a list-decoding error for a code C. We note that once again, our bound depends only on the weight distribution of the dual code $C^\perp$.

Proposition 5.3. *Fix any $\epsilon \in (0, \frac{1}{2})$, and define $\beta := \frac{1 - 2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}}{2}$ for $\tilde{\epsilon} = \epsilon + \frac{1}{\sqrt{\log N}}$. Let $B = [\beta N, (1-\beta)N] \cap \mathbb{N}$, and let $k^* = (2\lfloor \frac{N}{\sqrt{\log N}} \rfloor + 1)m$ for some integer $m > 0$. Then for any integer $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$ and all list sizes $k \geq k^*$, we have that any $t \times N$ matrix H with entries in $\mathbb{F}_2$ satisfies*

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{N}{4\epsilon \log N}} + \frac{N}{k^*} \max_{j \in B} \left\{ \Pr_{v \sim \mu_t}[\mathrm{wt}(vH) = j] \cdot \frac{2^N}{\binom{N}{j}} - 1 \right\}$$

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

Proof. We will use Theorem 5.2 to bound the decoding error probability in terms of the Krawtchouk polynomials $K_S(j)$ and the probability factors $\Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH) = j\right]$. Some of these terms will then be bounded using Proposition 1.7, and some will be bounded using Theorem 1.9. We proceed with the proof; applying Theorem 5.2 to the list size k^* with parameter $l = \lfloor \frac{N}{\sqrt{\log N}} \rfloor$, we get

$$\begin{aligned}
\Pr_{\rho\sim P_\epsilon}\left[\rho \notin D_k(H\rho^\intercal)\right] &\leq \Pr_{\rho\sim P_\epsilon}\left[\rho \notin D_{k^*}(H\rho^\intercal)\right] \\
&\leq 2e^{-\frac{N}{4\epsilon \log N}} \\
&+ \frac{N}{k^*} \max_{w\in[\epsilon N \pm \frac{N}{\sqrt{\log N}}]\cap\mathbb{N}} \left\{ \frac{1}{\binom{N}{w}} \sum_{j=0}^{N} \Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH) = j\right] K_w(j)^2 - 1 \right\}.
\end{aligned} \tag{12}$$

We want to bound the summation in the second term. We will start with the central terms $j \in B$. For these we rely on Proposition 1.7, which states that $\frac{2^{-N}}{\binom{N}{w}} \sum_{j=0}^{N} \binom{N}{j} \cdot K_w(j)^2 = 1$ for all $w \in \{0, 1, \ldots, N\}$. For any $w \in \{0, 1, \ldots, N\}$, we thus get

$$\begin{aligned}
&\frac{1}{\binom{N}{w}} \sum_{j\in B} \Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH) = j\right] K_w(j)^2 \\
&\qquad \leq \frac{1}{\binom{N}{w}} \max_{j\in B} \left\{ \Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH) = j\right] \cdot \frac{1}{\binom{N}{j}} \right\} \sum_{j\in B} \binom{N}{j} \cdot K_w(j)^2 \\
&\qquad \leq 2^N \max_{j\in B} \left\{ \Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH) = j\right] \cdot \frac{1}{\binom{N}{j}} \right\}.
\end{aligned} \tag{13}$$

We then want to bound the contribution of the faraway terms $j \notin B$ to the summation in (12), i.e. we want to bound

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

Bounding this quantity by N times its maximum value over j and applying Theorem 1.9, we get

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

But by Lemma 3.8 and subadditivity of entropy (i.e. Lemma 3.7), we know that

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

Additionally, for any $w \in \{\epsilon N \pm \frac{N}{\sqrt{\log N}}\}$ we have (again by subadditivity of entropy, i.e. Lemma 3.7)

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

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

$$\begin{aligned}(*) &\leq \sqrt{2}N^{\frac{3}{2}} \cdot 2^{4h(\frac{1}{\sqrt{\log N}})N} \cdot 2^{h(\epsilon)N} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[\mathrm{wt}(vH) = j\right] \cdot 2^{2\epsilon N \log|1 - \frac{2j}{N}|} \right\} \\ &\leq \frac{1}{N} \cdot 2^{5h(\frac{1}{\sqrt{\log N}})N} \cdot 2^{h(\epsilon)N} \max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[\mathrm{wt}(vH) = j\right] \cdot 2^{2\epsilon N \log|1 - \frac{2j}{N}|} \right\},\end{aligned}$$

where the last line follows from our assumption that $N > 2^{17} > 50$ and the fact that for all $N > 50$, we have $\log(\sqrt{2}N^{\frac{5}{2}}) \leq 3 \log N \leq \frac{N}{\sqrt{\log N}} \leq h(\frac{1}{\sqrt{\log N}})N$. Combining this bound for the faraway terms with our bound (13) for the central terms of the summation, we bound the right-hand side of equation (12) by

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

□

6. **The weight distribution of transitive linear codes.** We will now prove Theorem 1.1. We note that the bound we get is essentially tight, since for any finite field $\mathbb{F}_q$ and any integer divider j of N, the repetition code

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

is transitive, has dimension j, and has weight distribution

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

for all $\alpha \in (0,1)$ such that $\alpha j \in \mathbb{N}$. We recall and prove our Theorem 1.1 below:

Theorem 1.1. Consider any finite field $\mathbb{F}_q$, and let $C \subseteq \mathbb{F}_q^N$ be any transitive linear code. Then for any $\alpha \in (0,1)$, we have

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

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

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

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

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

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

$$\Pr_{Z \sim \mathcal{D}(C^{(\alpha)})}[Z_j = 0] = 1 - \alpha. \tag{14}$$

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

$$\begin{aligned} \mathsf{H}_{Z \sim \mathcal{D}(C^{(\alpha)})}(Z_j) &= (1-\alpha)\log\frac{1}{1-\alpha} + (q-1)\cdot\frac{\alpha}{q-1}\log\frac{q-1}{\alpha} \\ &= h_q(\alpha)\log(q). \end{aligned} \tag{15}$$

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

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

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

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

for every $j > r$. Applying (15) and the chain rule for entropy then gives

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

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

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

□

7. **List decoding for transitive codes.** We now turn to proving Theorem 1.5. 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 stated in Claim 3.3, the dual code of a transitive code is also transitive. For any transitive linear code C, we can thus apply our Theorem 1.1 for the weight distribution of $C^\perp$ to get a bound on the size of the decoding list for C. We restate and prove our Theorem 1.5 below.

Theorem 1.5. Fix any $\epsilon \in (0, \frac{1}{2})$ and $\eta \in (0,1)$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = \eta N$ can with high probability list-decode ϵ-errors using a list T of size

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

Proof. We will show that for all $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$, there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size

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

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin T(c+\rho)\right] \leq 4e^{-\frac{N}{4\epsilon \log N}}.$$

Let H denote the parity-check matrix of C. By Lemma 2.1, it is sufficient to show that for any list size $k > N$, we have

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{N}{4\epsilon \log N}} + \frac{2^{5h(\frac{1}{\sqrt{\log N}})N+1}}{k} \cdot (2^{4\epsilon\eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})}). \quad (16)$$

Setting the list size $k = e^{\frac{N}{4\epsilon \log N}} \cdot 2^{5h(\frac{1}{\sqrt{\log N}})N} \cdot (2^{4\epsilon\eta N} + 2^{\epsilon N \log(\frac{2}{1-\eta})})$ in equation (16) will then recover our theorem statement. We thus turn to proving (16). We note that $2\lfloor \frac{N}{\sqrt{\log N}} \rfloor + 1 < \frac{k}{2}$, so there exists some $k^* \in [\frac{k}{2}, k]$ satisfying the conditions of Proposition 5.3. Proposition 5.3 then yields the following bound on the left-hand side of (16):

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

where $\beta := \frac{1}{2}\left(1 - 2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}\right)$ for $\tilde{\epsilon} := \epsilon + \frac{1}{\sqrt{\log N}}$, and $B := [\beta N, (1-\beta)N] \cap \mathbb{N}$. Our goal will be to bound both the central terms $j \in B$ and the faraway terms $j \notin B$ by using our bounds on the weight distribution of transitive codes. As we've seen in section 3, the dual code $C^\perp$ is a transitive linear code of dimension $N - \dim C$. By Theorem 1.1, we thus have that for all $j \in \{0, 1, \ldots, N\}$,

$$\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] \leq 2^{-(1-h(\frac{j}{N}))(1-\eta)N}. \quad (18)$$

For any $j \in B$, we then have by Lemma 3.8 that

$$\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] \cdot \frac{2^N}{\binom{N}{j}} \leq 2^{-(1-h(j/N))(1-\eta)N} \cdot \frac{2^N}{\sqrt{\frac{1}{2N}} \cdot 2^{h(j/N)N}}$$

$$= \sqrt{2N} \cdot 2^{(1-h(j/N))\eta N}.$$

But for $j \in B$ we have $\beta \leq \frac{j}{N} \leq 1-\beta$, so the right-hand side is maximized at $j = \lceil \beta N \rceil$. Applying Lemma 3.9, we get

$$\max_{j \in B} \left\{ \Pr_{v \sim \mu_t} \left[\mathrm{wt}(vH) = j \right] \cdot \frac{2^N}{\binom{N}{j}} \right\} \leq \sqrt{2N} \cdot 2^{(1-h(\beta))\eta N}$$
$$\leq \sqrt{2N} \cdot 2^{4\tilde{\epsilon}(1-\tilde{\epsilon})\eta N}. \tag{19}$$

We now turn to the faraway terms of equation (17). By equation (18), we have

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

Note that by definition of β, any $\delta \in (0, \beta)$ can be written as $\delta = \frac{1 - 2\sqrt{\alpha\tilde{\epsilon}(1-\tilde{\epsilon})}}{2}$ for some $\alpha > 1$. By Lemma 3.9, we can then rewrite our previous expression as

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

But for any positive constant c, the derivative of $\log(\alpha) - c\alpha$ is $\frac{1}{\alpha \cdot \ln 2} - c$, and the second derivative is always negative. Thus, the above expression achieves its maximum when $\alpha = \frac{\epsilon}{2\tilde{\epsilon}(1-\tilde{\epsilon})(1-\eta)}$. We then get

$$\max_{j \notin B} \left\{ \Pr_{v \sim \mu_t} \left[\mathrm{wt}(vH) = j \right] \cdot 2^{2\epsilon N \log |1 - \frac{2j}{N}|} \right\} \leq 2^{-\frac{\epsilon N}{\ln 2}} \cdot 2^{\epsilon N \log(\frac{2\epsilon}{1-\eta})}$$
$$\leq 2^{-h(\epsilon)N} \cdot 2^{\epsilon N \log(\frac{2}{1-\eta})}, \tag{20}$$

where in the last line we used the inequality $\log(1-x) \geq -\frac{x}{(1-x)\ln 2}$ for $x < 1$ to get $h(\epsilon) \leq -\epsilon \log(\epsilon) + \frac{\epsilon}{\ln 2}$. We now use equations (19) and (20) to bound the central and faraway terms of (17) respectively. This gives

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

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

8. **List decoding for doubly transitive codes.** We will now turn to proving our list-decoding bounds for doubly transitive codes. We restate and prove our Theorem 1.4 below.

Theorem 1.4. Fix any $\epsilon \in (0, \frac{1}{2})$ and any $\gamma \leq 1 - \log(1 + 2^{-4\epsilon})$. Then any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1-\gamma)N$ can with high probability list-decode ϵ-errors using a list T of size

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

Proof. We will show that for all $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$, there exists a function T mapping every $x \in \mathbb{F}_2^N$ to a subset $T(x) \subseteq C$ of size

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

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

$$\Pr_{\rho \sim P_\epsilon}\left[c \notin T(c+\rho)\right] \leq 3e^{-\frac{N}{4\epsilon \log N}}.$$

Let H denote the parity-check matrix of C. By Lemma 2.1, it is sufficient to show that for any $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$ and any list size $k > N$, we have

$$\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{N}{4\epsilon \log N}} + \frac{a}{k} \cdot 2^{h(\epsilon)N - \gamma N} \tag{21}$$

for some $a = 2^{o(N)}$. Setting the list size $k = a \cdot e^{\frac{N}{4\epsilon \log N}} \cdot 2^{h(\epsilon)N-\gamma N}$ in equation (21) will then recover our theorem statement. We thus turn to proving (21). We note that $2\lfloor \frac{N}{\sqrt{\log N}} \rfloor + 1 < \frac{k}{2}$, so there exists some $k^* \in [\frac{k}{2}, k]$ satisfying the conditions of Proposition 5.3. Proposition 5.3 then yields the following bound on the left-hand side of (21).

$$\begin{aligned}\Pr_{\rho \sim P_\epsilon}[\rho \notin D_k(H\rho^\intercal)] \leq 2e^{-\frac{N}{4\epsilon \log N}} &+ \frac{2N}{k} \max_{j \in B}\left\{\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] \cdot \frac{2^N}{\binom{N}{j}}\right\} \\ &+ \frac{2^{h(\epsilon)N + 5h(\frac{1}{\sqrt{\log N}})N + 1}}{k} \max_{j \notin B}\left\{\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] 2^{2\epsilon N \log|1-\frac{2j}{N}|}\right\},\end{aligned} \tag{22}$$

where $\beta := \frac{1}{2}\left(1 - 2\sqrt{\tilde{\epsilon}(1-\tilde{\epsilon})}\right)$ for $\tilde{\epsilon} := \epsilon + \frac{1}{\sqrt{\log N}}$, and $B := [\beta N, (1-\beta)N] \cap \mathbb{N}$. Our goal will be to bound both the central terms $j \in B$ and the faraway terms $j \notin B$ by using Samorodnitsky's weight distribution bound for doubly transitive codes. Now by Claim 3.4, the dual code of a doubly transitive code is itself doubly transitive. Applying Theorem 1.6, we thus get that for all $j \in \{0, 1, \ldots, N\}$,

$$\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] \leq 2^{-\gamma N + o(N)} \cdot \left(\frac{1}{2^{1-\gamma} - 1}\right)^{\min\{j, N-j\}}. \tag{23}$$

It then follows that

$$\max_{j \in B}\left\{\Pr_{v \sim \mu_t}\left[\mathrm{wt}(vH) = j\right] \cdot \frac{2^N}{\binom{N}{j}}\right\} \leq \max_{\alpha \in [\beta, \frac{1}{2}]}\left\{2^{-\gamma N - \alpha N \log(2^{1-\gamma}-1) + N - h(\alpha)N + o(N)}\right\}. \tag{24}$$

We want to bound the expression on the right-hand side by $2^{h(\epsilon)N - \gamma N + o(N)}$. For this we define the function

$$f(\alpha) := -\gamma N - \alpha N \log(2^{1-\gamma} - 1) + N - h(\alpha)N$$

and compute its derivative

$$\frac{df}{d\alpha} = -N \log(2^{1-\gamma} - 1) - N \log \frac{1-\alpha}{\alpha}.$$

We note that over the interval $[0, 1]$, the second derivative $\frac{d^2 f}{d\alpha^2} = \frac{N}{\alpha(1-\alpha)\ln 2}$ is positive. Thus over $[0, 1]$, the function f is minimized at the point α^* satisfying $\frac{1-\alpha^*}{\alpha^*} = 2^{1-\gamma} - 1$ (i.e. $\alpha^* = 1 - 2^{\gamma-1}$), and f is monotone on either side of α^*. In particular, over the interval $[\beta, \frac{1}{2}]$ the function f must be maximized at either $\alpha = \beta$ or $\alpha = \frac{1}{2}$. But since $\gamma \leq 1 - \log(1 + 2^{-4\epsilon})$ by our theorem assumption, we have

$$f\left(\frac{1}{2}\right) \leq -\gamma N + 2\epsilon N$$

$$\leq -\gamma N + h(\epsilon)N. \tag{25}$$

On the other hand we have $\beta = \frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2} - o(1)$, so in order to show that

$$f(\beta) \leq h(\epsilon)N - \gamma N + o(N), \tag{26}$$

it suffices to show that

$$-\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\log(2^{1-\gamma}-1) + 1 - h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) - h(\epsilon) \leq 0.$$

But the left-hand is an increasing function of γ, so by our theorem assumption that $\gamma \leq 1 - \log(1+2^{-4\epsilon})$, it suffices to show that

$$2\epsilon(1-\sqrt{4\epsilon(1-\epsilon)}) + 1 - h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) - h(\epsilon) \leq 0. \tag{27}$$

We postpone the proof of this fact to Appendix D.5. Assuming this fact we get equation (26), which when combined with (25) and (24) gives us

$$\max_{j\in B}\left\{\Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH)=j\right]\cdot\frac{2^N}{\binom{N}{j}}\right\} \leq 2^{h(\epsilon)N-\gamma N+o(N)}. \tag{28}$$

This finishes our analysis of the central terms of equation (22). For the faraway terms, by (23) we have

$$\begin{aligned}
&\max_{j\notin B}\left\{\Pr_{v\sim\mu_t}\left[\mathrm{wt}(vH)=j\right]\cdot 2^{2\epsilon N\log|1-\frac{2j}{N}|}\right\} \\
&\qquad\leq \max_{j\leq\frac{N}{2}}\left\{2^{-\gamma N+o(N)}\left(\frac{1}{2^{1-\gamma}-1}\right)^j\cdot 2^{2\epsilon N\log(1-\frac{2j}{N})}\right\} \\
&\qquad= 2^{-\gamma N+o(N)}\max_{j\leq\frac{N}{2}}\left\{2^{-j\log(2^{1-\gamma}-1)+2\epsilon N\log(1-\frac{2j}{N})}\right\}.
\end{aligned} \tag{29}$$

Now the function

$$g(j) := -j\log(2^{1-\gamma}-1) + 2\epsilon N\log(1-\frac{2j}{N})$$

has first derivative

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

and second derivative

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

Thus $g(j)$ achieves its maximum at $j^* = \frac{N}{2} + \frac{2\epsilon N}{\ln 2\log(2^{1-\gamma}-1)}$ and is decreasing over $[j^*, \frac{N}{2}]$. Whenever $1-\gamma \geq \log(1+2^{-\frac{4\epsilon}{\ln 2}})$, we have $j^* \leq 0$; in that case the argument in equation (29) is maximized at $j=0$ and we get

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

We now combine this bound for the faraway terms with the bound (28) for the central terms to bound the right-hand side of (22). We get that for all $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$, we have

$$\Pr_{\rho\sim P_\epsilon}\left[\rho\notin D_k(H\rho^\intercal)\right] \leq 2e^{-\frac{N}{4\epsilon\log N}} + \frac{2^{o(N)}}{k}\cdot 2^{h(\epsilon)N-\gamma N}.$$

We have shown (21), so we are done. □

Acknowledgments. We thank Alexander Barg, Paul Beame, Noam Elkies, Jan Hazla, Amir Shpilka, Madhu Sudan and Amir Yehudayoff for useful discussions.

Appendix A. **Weight bounds comparisons.** In this section, we compare our Theorem 1.1 with previously known bounds on the weight distribution of Reed-Muller codes. We will denote by $\mathsf{RM}_q(n,d)$ the Reed-Muller code over $\mathbb{F}_q$ with n variables and degree d. The codewords of $\mathsf{RM}_q(n,d)$ are the evaluation vectors (over all points in $\mathbb{F}_q^n$) of all multivariate polynomials of degree $\leq d$ in n variables. Let $M_q(n,d)$ denote the set of monomials $m = \prod_{i=1}^n x_i^{d_i}$ satisfying

1. $d_i < q$ for every $i \in \{1,2,\ldots,n\}$
2. $\sum_{i=1}^n d_i \leq d$.

Then the dimension of the corresponding Reed-Muller code is

$$\dim \mathsf{RM}_q(n,d) = |M_q(n,d)|. \tag{30}$$

We note that when $q = 2$, we have

$$|M_2(n,d)| = \binom{n}{\leq d}.$$

Throughout this section, we will denote by $\mathcal{D}_q(n,d)$ the uniform distribution over all codewords in $\mathsf{RM}_q(n,d)$, and by $\mathrm{wt}(c)$ the number of non-zero coordinates of c. When $q = 2$, we will simply write $\mathsf{RM}(n,d)$ and $\mathcal{D}(n,d)$ to mean $\mathsf{RM}_2(n,d)$ and $\mathcal{D}_2(n,d)$. The following result is an immediate consequence of our Theorem 1.1.

Theorem A.1. *Consider any finite field $\mathbb{F}_q$. For any non-negative integers $n, d \leq n$, and any $\alpha \in (0,1)$, the Reed-Muller code $\mathsf{RM}_q(n,d)$ satisfies*

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

Proof. This follows immediately from Theorem 1.1, Fact 3.6, and equation (30). □

Reed-Muller codes over non-prime fields

To the best of our knowledge, our Theorem A.1 is the first weight bound for Reed-Muller codes over non-prime fields.

Reed-Muller codes over odd prime fields

For Reed-Muller codes over odd prime fields, the only preexisting weight bound we are aware of is the following result of [10]:

Theorem A.2 (Corollary 1.2 in [10])**.** *For any $0 < \delta < \frac{1}{2}$, there are constants $c_1, c_2 > 0$ such that for any odd prime q and for any integers d, n such that $d \leq \delta n$, we have*

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

This was a generalization of [11], who proved the same result for Reed-Muller codes over $\mathbb{F}_2$. Theorem A.2 is very strong for small degrees, but gets weaker as the degree increases. When d is linear in n we have $q^{-c_1\frac{n}{d}} = \Theta(1)$, meaning that in this regime Theorem A.2 can only give a nontrivial bound on relative weights $\frac{\mathrm{wt}(c)}{N}$ that are at least a constant away from $1 - \frac{1}{q}$. Our Theorem A.1 gives nontrivial bounds for all relative weights and all degrees.

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 A.2 was proven over $\mathbb{F}_2$ by [11]. For comparison with our Theorem A.1, see the discussion above.

In the constant-rate regime (i.e. $d = \frac{n}{2} \pm O(\sqrt{n})$), the strongest known weight bound (for all weights) is due to Samorodnitsky. It follows immediately from Theorem 1.6, i.e. from Proposition 1.4 in [48].

Theorem A.3 (follows from Proposition 1.4 in [48])**.** *For any $\alpha \in (0,1)$, define $\alpha^* := \min\{\alpha, 1-\alpha\}$. Then for any non-negative integers $n, d \le n$ and any $\alpha \in (0,1)$, the Reed-Muller code $\mathsf{RM}(n,d)$ satisfies*

$$\Pr_{c\sim\mathcal{D}(n,d)}\left[\mathrm{wt}(c) = \alpha N\right] \le 2^{-\binom{n}{\le d}+o(N)}\left(2^{1-\frac{\binom{n}{\le d}}{N}} - 1\right)^{-\alpha^* N}.$$

Moreover, if $\alpha^* \ge 1 - 2^{\frac{\binom{n}{\le d}}{N}-1}$,

$$\Pr_{c\sim\mathcal{D}(n,d)}\left[\mathrm{wt}(c) = \alpha N\right] \le 2^{o(N)} \cdot \frac{\binom{N}{\alpha N}}{2^N}.$$

When the rate of the code is subconstant (i.e. when the degree is away from $\frac{n}{2}$), Theorem A.3 does not give strong bounds. An approach that has been fairly successful in this regime is the line of work of [33, 4, 51]. To our knowledge, the strongest results for these regimes are due to [51]. We start with their bound for lower weights, i.e. for weights in $[0, \frac{N}{4}]$.

Theorem A.4 (Theorem 1.1 in [51])**.** *For any $j, n, d \in \mathbb{N}$ with $d \le n$, we have*

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

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

$$\begin{aligned}\Pr[\mathrm{wt}(c) \le 2^{-j}\cdot 2^n] &\le 2^{-\left(1-h(2^{-j})\right)\binom{n}{\le d}} \\ &\le 2^{-(1-2j\cdot 2^{-j})\binom{n}{\le d}},\end{aligned}$$

while the expression in Theorem A.4 satisfies

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

Thus our Theorem A.1 is stronger than Theorem A.4 whenever

$$j\cdot 2^{-(j-1)} < 17j\cdot\left(\frac{d}{n}\right)^{j-1}. \tag{31}$$

This condition is always satisfied when $d \ge \frac{n}{2}$, so in this range our Theorem A.1 is stronger than Theorem A.4 for all weights. When $d < \frac{n}{2}$, condition (31) is satisfied whenever

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

For any $\frac{n}{34} < d < \frac{n}{2}$, this gives a nontrivial range.

This concludes our comparison of Theorem A.1 with Theorem A.4, which was the bound of [51] for weights in $[0, \frac{N}{4}]$. We now turn to their bounds for larger weights.

Theorem A.5 (Theorem 1.3 in [51]). *Let $j, n \in \mathbb{N}$ and let $0 < \gamma(n) < \frac{1}{2} - \Omega\left(\sqrt{\frac{\log n}{n}}\right)$ be a parameter (which may be constant or depend on n) such that $\frac{j+\log \frac{1}{1-2\gamma}}{(1-2\gamma)^2} = o(n)$. Then*

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

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

This bound holds when the degree is smaller than $\frac{n}{2}$. For arbitrary degree, [51] gives the following:

Theorem A.6 (Theorem 1.5 in [51]). *For any $n, d \in \mathbb{N}$ with $d \leq n$ and any $\delta > 0$, we have*

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

We will start by comparing our Theorem A.1 with Theorem A.6. Applying Lemma 3.9, we get from Theorem A.1 that

$$\begin{aligned}\Pr_{c\sim\mathcal{D}(n,d)}[\mathrm{wt}(c) \leq \frac{1-\delta}{2}N] &\leq 2^{-(1-h(\frac{1-\delta}{2}))\cdot\binom{n}{\leq d}} \\ &\leq e^{-\frac{\delta^2}{2}\cdot\binom{n}{\leq d}}.\end{aligned}$$

Thus our Theorem A.1 is strictly stronger than Theorem A.6 for all $d < n$. We will now compare our Theorem A.1 with Theorem A.5. Applying Lemma 3.9, we get from Theorem A.1 that

$$\begin{aligned}\Pr_{c\sim\mathcal{D}(n,d)}[\mathrm{wt}(c) \leq \frac{1-2^{-j}}{2}N] &\leq 2^{-(1-h(\frac{1-2^{-j}}{2}))\cdot\binom{n}{\leq d}} \\ &\leq 2^{-\frac{2^{-2j}}{2\ln 2}\cdot\binom{n}{\leq d}}.\end{aligned}$$

It follows that our Theorem A.1 is stronger than Theorem A.5 whenever $2^{-(2j+1)} \geq 2^{-c(\gamma,j)}$, i.e. whenever

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

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 A.1 is stronger than Theorem A.5 whenever $d > \gamma^* n$. In private correspondence with Amir Shpilka and Ori Sberlo, we learned that γ^* can be computed to be $\gamma^* \approx 0.38$.

Appendix B. Proof of corollary 1.3.

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

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

and that for any code C we denote by $\mathcal{D}(C^\perp)$ the uniform distribution over the dual code $C^\perp$. We now restate and prove our Corollary 1.3.

Corollary 1.3. Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and let $C \subseteq \mathbb{F}_2^N$ be a linear code. Suppose that for every $j \in A_\epsilon$ we have

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

and suppose that

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

Then C is resilient to ϵ-errors.

Proof. From Theorem 1.2, we know that whenever $N > \frac{1}{\epsilon^4(\frac{1}{2}-\epsilon)^4}$, there exists some decoder $d : \mathbb{F}_2^N \to C$ such that for all $c \in C$,

$$\Pr_{\rho \sim P_\epsilon} [d(c+\rho) \neq c] \leq 2e^{-\frac{\sqrt{N}}{3\epsilon}} + N \max_{\substack{S \subseteq [\epsilon N \pm N^{3/4}] \cap \mathbb{N} \\ 1 \leq |S| \leq 2}} \Big\{ \frac{1}{\binom{N}{S}} \sum_{j=0}^{N} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 - 1 \Big\}. \tag{32}$$

Let $\nu \in (0, \frac{1}{2})$ be such that $h(\nu) = 1 - h(\epsilon) - N^{-1/5}$, and note that we have

$$A_\epsilon = \{\lceil \nu N \rceil, \lceil \nu N \rceil + 1, \ldots, \lfloor (1-\nu) N \rfloor\}.$$

We will start by bounding the central terms $j \in A_\epsilon$ in equation (32). Applying Proposition 1.7 and the first condition in our theorem statement, we immediately get that for any $S \subseteq \{0, 1, \ldots, N\}$,

$$\frac{1}{\binom{N}{S}} \sum_{j \in A_\epsilon} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \leq 1 + o\big(\frac{1}{N}\big). \tag{33}$$

We now turn to the faraway terms $j \notin A_\epsilon$. For these, we note that for any non-negative integers $j, s \leq N$ we have

$$\begin{aligned} |K_s(j)| &= \left| \sum_{t=0}^{s} (-1)^t \binom{j}{t} \binom{N-j}{s-t} \right| \\ &\leq \sum_{t=0}^{s} \binom{j}{t} \binom{N-j}{s-t} \\ &= \binom{N}{s}, \end{aligned}$$

where we used the convention that $\binom{a}{b} = 0$ when $a < b$. For any $S \subseteq \{0, 1, \ldots, N\}$, we can then bound the faraway terms $j \notin A_\epsilon$ of equation (32) by

$$\frac{1}{\binom{N}{S}} \sum_{j \notin A_\epsilon} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \leq \binom{N}{S} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) \notin A_\epsilon\big].$$

Applying the second condition in our theorem statement in combination with Lemma 3.8 and the subadditivity of entropy (Lemma 3.7), we get

$$\max_{\substack{S \subseteq [\epsilon N \pm N^{3/4}] \cap \mathbb{N} \\ 1 \leq |S| \leq 2}} \Big\{ \frac{1}{\binom{N}{S}} \sum_{j \notin A_\epsilon} \Pr_{y \sim C^\perp} \big[\mathrm{wt}(y) = j\big] K_S(j)^2 \Big\}$$

$$\leq 2\binom{N}{\lfloor \epsilon N + N^{3/4} \rfloor} \cdot 2 \cdot 2^{-h(\epsilon)N - N^{4/5} + N^{3/4}}$$
$$\leq 4 \cdot 2^{h(\epsilon)N + h(N^{-1/4})N} \cdot 2^{-h(\epsilon)N - N^{4/5} + N^{3/4}}$$
$$\leq o(\frac{1}{N}).$$

Combining this bound for the faraway terms with our bound (33) for the central terms, we bound equation (32) by

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

□

Appendix C. **Lower bounds on list decoding.** In this section, we prove the result mentioned in equation (1), section 1.

Claim C.1. *Let $\epsilon \in (0, \frac{1}{2})$ be arbitrary, and consider any $N > \frac{100}{\epsilon^2}$. Suppose a code $C \subseteq \mathbb{F}_2^N$ and a decoder $d_k : \mathbb{F}_2^N \to C^{\otimes k}$ satisfy*

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

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

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

Proof. We will first show that in order for the decoder d_k to succeed with high probability, there must be many codewords $c \in C$ for which

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

Intuitively, this is because the sphere of radius ϵN around any codeword c contains $\approx 2^{h(\epsilon)N}$ points (and for any transmitted codeword c, with high probability the received message m will satisfy $\mathrm{wt}(m+c) \approx \epsilon N$). We will then simply double-count the number of pairs (x, c) for which $c \in d_k(x)$. On the one hand, there are $2^N \cdot k$ such pairs, since every received message is mapped to k codewords; on the other hand, there must be at least about $|C| \cdot 2^{h(\epsilon)N}$ pairs, since as we've just argued most codewords in C need to be matched to at least $\approx 2^{h(\epsilon)N}$ points. It follows that we must have

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

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

$$\Pr_{\rho \sim P_\epsilon}[c \in d_k(c+\rho)] \geq \frac{1}{2}. \tag{34}$$

Fix any such c. Now from Chernoff's bound (i.e Lemma 3.11), we have for $N > \frac{100}{\epsilon^2}$ that

$$\Pr_{\rho \sim P_\epsilon}\left[\mathrm{wt}(\rho) \leq \epsilon N - \epsilon N^{3/4}\right] \leq 2e^{-\frac{10}{3}}$$
$$\leq \frac{1}{4}.$$

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

$$x \in S_c \implies c \in d_k(x) \tag{35}$$

and

$$\Pr_{\rho \sim P_\epsilon}\left[\rho \in S_c\right] \geq \frac{1}{4}. \tag{36}$$

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

$$\begin{aligned}\Pr_{\rho \sim P_\epsilon}\left[\rho = c+x\right] &\leq \epsilon^{\epsilon N - \epsilon N^{3/4}}(1-\epsilon)^{(1-\epsilon)N+\epsilon N^{3/4}} \\ &\leq 2^{-(1-N^{-1/4})h(\epsilon)N}\end{aligned} \tag{37}$$

Equations (36) and (37) 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 (35) that any such c must satisfy

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

By double counting, we get

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

The result then follows from rearranging terms. □

Appendix D. Other proofs for sections 1, 3 and 2.

D.1. Explicit bounds from Theorem 1.5.

In this section, we prove the result we mentioned in equation (2).

Claim D.1. *Fix any $\epsilon \in (0, \frac{1}{2})$ and $N > 2^{\frac{1}{\epsilon^2(1-\epsilon)^2}+1}$. Then any transitive linear code $C \subseteq \mathbb{F}_2^N$ of dimension $\dim C = (1-\frac{4\epsilon}{e})N$ can with high probability list-decode ϵ-errors using a list T of size*

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

Proof. From Theorem 1.5, we know that C can with high probability list-decode ϵ-errors using a list T of size

$$\begin{aligned}|T| &= 2^{\epsilon N \log(\frac{2e}{4\epsilon})+o(N)} + 2^{4\epsilon N+o(N)} \\ &= 2^{\epsilon N \log(\frac{1}{\epsilon})+\epsilon N \log e - \epsilon N+o(N)} + 2^{4\epsilon N+o(N)} \\ &= 2^{\epsilon N \log(\frac{1}{\epsilon})+(1-\epsilon)N\frac{\epsilon}{\ln 2}-\epsilon N+\frac{\epsilon^2}{\ln 2}N+o(N)} + 2^{4\epsilon N+o(N)} \\ &\leq 2^{(h(\epsilon)-\epsilon+\frac{\epsilon^2}{\ln 2})N+o(N)} + 2^{4\epsilon N+o(N)},\end{aligned}$$

where in the last line we used the inequality $\log(1-x) \leq -\frac{x}{\ln 2}$ for all x to get $h(\epsilon) \geq \epsilon \log \frac{1}{\epsilon} + (1-\epsilon)\frac{\epsilon}{\ln 2}$. □

D.2. **Duals of transitive codes - proof of claims 3.3 and 3.4.** We show that the dual of a transitive code is itself transitive.

Claim 3.3. The dual code $C^\perp$ of a transitive code $C \subseteq \mathbb{F}_2^N$ is transitive.

Proof. Let $i, j \in [N]$ be arbitrary. Since C is transitive, we know there exists a permutation $\pi : [N] \to [N]$ such that $\pi(j) = i$ and for any $c = (c_1, c_2, \ldots, c_N) \in C$, we have $c_\pi := (c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$. Clearly π^{-1} satisfies $\pi^{-1}(i) = j$, and we claim that it also satisfies that $v_{\pi^{-1}} \in C^\perp$ for all $v \in C^\perp$. For this we note that since $c_\pi \in C$ for every $c \in C$, we have by definition that every $v \in C^\perp$ satisfies

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

We thus have

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

□

Claim 3.4. The dual code $C^\perp$ of a doubly transitive code $C \subseteq \mathbb{F}_2^N$ is doubly transitive.

Proof. Let $i, j, k, l \in [N]$ be such that $i \neq k$ and $j \neq l$. Since C is doubly transitive, we know there exists a permutation $\pi : [N] \to [N]$ such that $\pi(j) = i$, $\pi(l) = k$, and for any $c = (c_1, c_2, \ldots, c_N) \in C$, we have $c_\pi := (c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(N)}) \in C$. Clearly π^{-1} satisfies $\pi^{-1}(i) = j$ and $\pi^{-1}(k) = l$, and we claim that it also satisfies that $v_{\pi^{-1}} \in C^\perp$ for all $v \in C^\perp$. For this we note that since $c_\pi \in C$ for every $c \in C$, we have by definition that every $v \in C^\perp$ satisfies

$$\sum_{t=1}^{N} v_t c_{\pi(t)} = 0 \text{ for all } c \in C.$$

We thus have

$$\begin{aligned} v \in C^\perp &\implies \sum_t v_t c_{\pi(t)} = 0 \text{ for all } c \in C \\ &\implies \sum_t v_{\pi^{-1}(t)} c_t = 0 \text{ for all } c \in C \\ &\implies v_{\pi^{-1}} \in C^\perp. \end{aligned}$$

□

D.3. **On known list-decoding bounds for doubly transitive codes.** We recall the known list-decoding bound for doubly transitive codes (see equation (5) in section 1):

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

where $1 - \gamma \in (0, 1)$ is the rate of the code. We claim that for constant γ, this bound never achieves the information-theoretic $2^{h(\epsilon)N - \gamma N + o(N)}$.

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

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

Proof. Since $2^x < 1 + x$ for all $x \in (0,1)$, it will be sufficient to show that

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

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

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

i.e. that

$$f(\epsilon, c) := \log(1-\epsilon) + 2\epsilon - 2\epsilon \log c + c\epsilon \geq 0. \tag{38}$$

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

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

and

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

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

$$\min_{c \leq \frac{1}{\epsilon}} f(\epsilon, c) = \begin{cases} f(\epsilon, \frac{2}{\ln 2}) & \text{if } \epsilon \leq \frac{\ln 2}{2}, \\ f(\epsilon, \frac{1}{\epsilon}) & \text{otherwise.} \end{cases} \tag{39}$$

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

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

The first derivative is

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

and the second derivative is

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

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

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

for all $0 \leq \epsilon \leq \frac{\ln 2}{2}$. This deals with the first case of (39). For the second case of (39), we want to show that for all $\epsilon \in (0, \frac{1}{2})$ we have

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

But

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

is maximized at $\epsilon = \frac{1}{2}$, since $\frac{\partial^2}{\partial\epsilon^2}f(\epsilon,\frac{1}{\epsilon}) = \frac{1}{\ln 2}(\frac{2}{\epsilon} - \frac{1}{(1-\epsilon)^2})$ and $2(1-\epsilon)^2 \geq \frac{1}{2} \geq \epsilon$ for $\epsilon \in (0,\frac{1}{2})$. It then follows that for $\epsilon \in (0,\frac{1}{2})$, we have

$$\begin{aligned}\frac{\partial}{\partial\epsilon}f(\epsilon,\frac{1}{\epsilon}) &\leq -\frac{1}{(1-\frac{1}{2})\ln 2} + 2 - 2\log(2) + \frac{2}{\ln 2}\\ &= 0,\end{aligned}$$

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

D.4. **A version of Pinsker's inequality - Proof of Lemma 3.9.**

Lemma 3.9. For any $\mu \in (0,1)$, we have

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

and thus

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

Proof.

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

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

D.5. **Proof of (27).**

Claim D.3. *For any $\epsilon \in [0,\frac{1}{2}]$, we have*

$$h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) + h(\epsilon) \geq 1 + 2\epsilon(1-\sqrt{4\epsilon(1-\epsilon)}).$$

Proof. Writing the Taylor expansion of h as in the proof of Lemma 3.9, we have

$$h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) + h(\epsilon) = 2 - \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{(4\epsilon(1-\epsilon))^i + (1-2\epsilon)^{2i}}{i(2i-1)}.$$

But $\sum_{i=1}^{\infty}\frac{1}{i(2i-1)} = 2\ln 2$, so our previous expression can be rewritten as

$$\begin{aligned} h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) + h(\epsilon) &= 1 + \frac{1}{2\ln 2}\sum_{i=1}^{\infty}\frac{1-(4\epsilon(1-\epsilon))^i - (1-4\epsilon(1-\epsilon))^i}{i(2i-1)} \\ &\geq 1 + \frac{1}{2\ln 2}\sum_{i=2}^{\infty}\frac{1-(4\epsilon(1-\epsilon))^2 - (1-4\epsilon(1-\epsilon))^2}{i(2i-1)}, \end{aligned}$$

where in the second line we used the fact that the term $i=1$ in the summation is 0. We will now need the following inequality:

$$1-(4\epsilon(1-\epsilon))^2 - (1-4\epsilon(1-\epsilon))^2 \geq \frac{4\ln 2\cdot\epsilon(1-\sqrt{4\epsilon(1-\epsilon)})}{2\ln 2 - 1}. \tag{40}$$

Once we establish (40), our claim follows from bounding our previous inequality by

$$\begin{aligned} h\Big(\frac{1-\sqrt{4\epsilon(1-\epsilon)}}{2}\Big) + h(\epsilon) &\geq 1 + \frac{1}{2\ln 2}\cdot\frac{4\ln 2\cdot\epsilon(1-\sqrt{4\epsilon(1-\epsilon)})}{2\ln 2-1}\sum_{i=2}^{\infty}\frac{1}{i(2i-1)} \\ &= 1 + \frac{2\epsilon(1-\sqrt{4\epsilon(1-\epsilon)})}{2\ln 2-1}\Big(\sum_{i=1}^{\infty}\frac{1}{i(2i-1)} - 1\Big) \\ &= 1 + 2\epsilon(1-\sqrt{4\epsilon(1-\epsilon)}) \end{aligned}$$

It thus only remains to prove (40). For this, we note that the right-hand side of (40) can be bounded by

$$\frac{4\ln 2\cdot\epsilon(1-\sqrt{4\epsilon(1-\epsilon)})}{2\ln 2-1} \leq 8\epsilon(1-\sqrt{4\epsilon(1-\epsilon)}),$$

while the left-hand side of (40) expands to

$$1-(4\epsilon(1-\epsilon))^2 - (1-4\epsilon(1-\epsilon))^2 = 8\epsilon - 40\epsilon^2 + 64\epsilon^3 - 32\epsilon^4.$$

Thus it is sufficient to show that

$$5\epsilon - 8\epsilon^2 + 4\epsilon^3 \leq \sqrt{4\epsilon(1-\epsilon)},$$

or equivalently (squaring both sides and dividing by ϵ) that the function

$$g(\epsilon) := 16\epsilon^5 - 64\epsilon^4 + 104\epsilon^3 - 80\epsilon^2 + 29\epsilon - 4$$

satisfies

$$g(\epsilon) \leq 0 \tag{41}$$

for all $\epsilon \in [0, \frac{1}{2}]$. But the derivative of g is

$$\begin{aligned} \frac{dg}{d\epsilon} &= 80\epsilon^4 - 256\epsilon^3 + 312\epsilon^2 - 160\epsilon + 29 \\ &= (1-2\epsilon)^2(20\epsilon^2 - 44\epsilon + 29), \end{aligned}$$

and the polynomial $20\epsilon^2 - 44\epsilon + 29$ has the two complex roots $\frac{11 \pm 2\sqrt{6} \cdot i}{10}$. Thus over the interval $[0, \frac{1}{2}]$, the function $g(\epsilon)$ must be maximized at either $\epsilon = 0$ or $\epsilon = \frac{1}{2}$. Since $g(0) = -4$ and $g(\frac{1}{2}) = 0$, we have

$$g(\epsilon) \leq 0$$

for all $\epsilon \in [0, \frac{1}{2}]$. We have thus shown (41), and we are done. □

REFERENCES

[1] E. Abbe, J. Hazla and I. Nachum, Almost-reed-muller codes achieve constant rates for random errors, *IEEE Trans. Inf. Theory*, **67** (2021), 8034-8050.

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

[3] E. Abbe, O. Sberlo, A. Shpilka and M. Ye, Reed-muller codes, *Foundations and Trends in Communications and Information Theory*, **20** (2023), 1-156.

[4] E. Abbe, A. Shpilka and A. Wigderson, Reed-muller codes for random erasures and errors, *IEEE Trans. Inf. Theory*, **61** (2015), 5229-5252.

[5] E. Abbe, A. Shpilka and M. Ye, Reed-muller codes: Theory and algorithms, *IEEE Trans. Inf. Theory*, **67** (2021), 3251-3277.

[6] E. Abbe and M. Ye, Reed-muller codes polarize, *2019 IEEE 60th Annual Symposium on Foundations of Computer Science,*, IEEE Computer Society Press, Los Alamitos, CA, (2019), 273-286.

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

[8] A. E. Ashikhmin, I. S. Honkala, T. Laihonen and S. Litsyn, On relations between covering radius and dual distance, *IEEE Trans. Inf. Theory*, **45** (1999), 1808-1816.

[9] A. Barg, Stolarsky's invariance principle for finite metric spaces, *Mathematika*, **67** (2021), 158-186.

[10] P. Beame, S. O. Gharan and X. Yang, On the bias of reed-muller codes over odd prime fields, *SIAM J. Discret. Math.*, **34** (2020), 1232-1247.

[11] I. Ben-Eliezer, R. Hod and S. Lovett, Random low-degree polynomials are hard to approximate, *Comput. Complex.*, **21** (2012), 63-81.

[12] D. Bilyk, F. Dai and R. Matzke, Stolarsky principle and energy optimization on the sphere, *Constructive Approximation*, **48** (2018), 31-60.

[13] S. Boucheron, G. Lugosi and P. Massart, *Concentration Inequalities - A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013.

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

[15] J. Brakensiek, S. Gopi and V. Makam, Generic reed-solomon codes achieve list-decoding capacity, *STOC'23—Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, Association for Computing Machinery (ACM), New York, (2023), 1488-1501.

[16] R. de Wolf, A brief introduction to fourier analysis on the boolean cube, *Theory Comput.*, **1** (2008), 1-20.

[17] P. Elias, *List Decoding for Noisy Channels*, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, MA, 1957.

[18] R. G. Gallager, Low-density parity-check codes, *IRE Trans. Inf. Theory*, **8** (1962), 21-28.

[19] D. Galvin, Three tutorial lectures on entropy and counting, 2014.

[20] V. Guruswami, J. Håstad and S. Kopparty, On the list-decodability of random linear codes, *IEEE Trans. Inf. Theory*, **57** (2011), 718-725.

[21] V. Guruswami, J. Håstad, M. Sudan and D. Zuckerman, Combinatorial bounds for list decoding, *IEEE Trans. Inf. Theory*, **48** (2002), 1021-1034.

[22] V. Guruswami and A. Rudra, Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy, *IEEE Trans. Inf. Theory*, **54** (2008), 135-150.

[23] V. Guruswami and C. Xing, Folded codes from function field towers and improved optimal rate list decoding, *STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing*, Association for Computing Machinery (ACM), New York, (2012), 339-350.

[24] J. Hazla, Optimal list decoding from noisy entropy inequality, In *IEEE International Symposium on Information Theory, ISIT 2023, Taipei, Taiwan, June 25-30, 2023*, IEEE, (2023), 15-18.

[25] J. Hazla, A. Samorodnitsky and O. 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*, ACM, (2021), 1479-1488.

[26] B. Hemenway, N.Ron-Zewi and M. Wootters, Local list recovery of high-rate tensor codes & applications, *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, IEEE Computer Society, Los Alamitos, CA, (2017), 204-215.

[27] M. E. H. Ismail and P. Simeonov, Strong asymptotics for krawtchouk polynomials, *Journal of Computational and Applied Mathematics*, **100** (1998), 121-144.

[28] K. Ivanov and R. L. Urbanke, Capacity-achieving codes: A review on double transitivity, *CoRR*, abs/2010.15453, 2020.

[29] J. Kahn, G. Kalai and N. 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 IEEE Computer Society, (1988), 68-80.

[30] G. Kalai and N. Linial, On the distance distribution of codes, *IEEE Trans. Inf. Theory*, **41** (1995), 1467-1472.

[31] T. Kasami and N. Tokura, On the weight structure of reed-muller codes, *IEEE Trans. Inf. Theory*, **16** (1970), 752-759.

[32] T. Kasami, N. Tokura and S. Azumi, On the weight enumeration of weights less than 2.5d of reed-muller codes, *Information and Control*, **30** (1976), 380-395.

[33] T. Kaufman, S. Lovett and E. Porat, Weight distribution and list-decoding size of reed-muller codes, *IEEE Trans. Inf. Theory*, **58** (2012), 2689-2696.

[34] S. Kopparty, List-decoding multiplicity codes, *Theory Comput.*, **11** (2015), 149-182.

[35] I. Krasikov and S. Litsyn, On the accuracy of the binomial approximation to the distance distribution of codes, *IEEE Trans. Inf. Theory*, **41** (1995), 1472-1474.

[36] I. Krasikov and S. Litsyn, Bounds on spectra of codes with known dual distance, *Des. Codes Cryptogr.*, **13** (1998), 285-297.

[37] I. Krasikov and S. Litsyn, Survey of binary krawtchouk polynomials, In Alexander Barg and Simon Litsyn, editors, *Codes and Association Schemes, Proceedings of a DIMACS Workshop, Piscataway, New Jersey, USA, November 9-12, 1999*, volume 56 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, DIMACS/AMS, (1999), 199-211.

[38] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Sasoglu and R. 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*, ACM, (2016), 658-669.

[39] S. Kudekar, T. Richardson and R. L. Urbanke, Spatially coupled ensembles universally achieve capacity under belief propagation, *IEEE Trans. Inf. Theory*, **59** (2013), 7761-7813.

[40] R. Li and M. 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*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, (2018), Art. No. 50, 19 pp.

[41] J. E. Littlewood, G, H, Hardy and G, Polya, *Inequalities*, Cambridge University Press, 1934.

[42] M. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman and V. 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*, ACM, (1997), 150-159.

[43] J. Macwilliams, A theorem on the distribution of weights in a systematic code, *Bell System Technical Journal*, **42** (1963), 79-94.

[44] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*, North-Holland Publishing Company, 1977.

[45] J. Mosheiff, N. Resch, N. Ron-Zewi, S. Silas and M. 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*, IEEE, (2020), 458-469.

[46] Y. Polyanskiy, Hypercontractivity of spherical averages in hamming space, *SIAM J. Discret. Math.*, **33** (2019), 731-754.

[47] G. Reeves and H. D. Pfister, Reed-muller codes on BMS channels achieve vanishing bit-error probability for all rates below capacity, *IEEE Trans. Inf. Theory*, **70** (2024), 920-949.
[48] A. Samorodnitsky, An improved bound on l_q norms of noisy functions, preprint, 2020, arXiv:2010.02721.
[49] A. Samorodnitsky, An upper bound on l_q norms of noisy functions, *IEEE Trans. Inf. Theory*, **66** (2020), 742-748.
[50] A. Samorodnitsky, On some properties of random and pseudorandom codes, *CoRR*, abs/2206.05135, 2022.
[51] O. Sberlo and A. 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*, SIAM, (2020), 1357-1376.
[52] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), 379-423.
[53] M. Skriganov, Point distributions in two-point homogeneous spaces, *Mathematika*, **65** (2019), 557-587.
[54] N. J. A. Sloane and E. R. Berlekamp, Weight enumerator for second-order reed-muller codes, *IEEE Trans. Inf. Theory*, **16** (1970), 745-751.
[55] M. Talagrand, On russo's approximate zero-one law, *The Annals of Probability*, **22** (1994), 1576-1587.

Received April 2023; 1st revision October 2023; 2nd revision February 2024; early access February 2024.