# FROM BIT TO BLOCK: DECODING ON ERASURE CHANNELS

HENRY D. PFISTER*, OSCAR SPRUMONT†, AND GILLES ZÉMOR‡

ABSTRACT. We provide a general framework for bounding the block error threshold of a linear code $C \subseteq \mathbb{F}_2^N$ over the erasure channel in terms of its bit error threshold. Our approach relies on understanding the minimum support weight of any $r$-dimensional subcode of $C$, for all small values of $r$. As a proof of concept, we use our machinery to obtain a new proof of the celebrated result that Reed–Muller codes achieve capacity on the erasure channel with respect to block error probability.

## 1. INTRODUCTION

We will be interested in the performance of linear codes over noisy communication channels. Ever since Shannon's seminal paper [Sha48], information theorists have spent considerable time and effort designing codes that *achieve capacity*, i.e. codes $C \subseteq \mathbb{F}_2^N$ that approach the optimal tradeoff between their rate $R := \frac{\log_2 |C|}{N}$ and the amount of noise they can tolerate.

Shannon's probabilistic argument [Sha48] established that uniform random codes achieve capacity on both the binary symmetric channel and the erasure channel. The first explicit families of codes to provably achieve capacity were only obtained decades later, when Forney introduced concatenated codes in [For66]. More recently, Arikan showed that polar codes - which have both a deterministic construction and efficient encoding and decoding algorithms - achieve capacity on all memoryless channels [Ari09]. This brought renewed attention to the closely related Reed–Muller codes, which were shown to achieve capacity on the binary symmetric channel and the erasure channel in [ASW15, KKM$^+$17, KKM$^+$16, RP24, AS23b, AS23a].

The proof that Reed–Muller codes of constant rate achieve capacity on the erasure channel [KKM$^+$17, KKM$^+$16] relies heavily on the double transitivity of their automorphism group. In fact, Kudekar, Kumar, Mondelli, Pfister, Şaşoğlu and Urbanke showed in [KKM$^+$17] that *any* doubly transitive linear code achieves capacity under bit-MAP decoding. A very natural question is then to understand whether or not the bit and block error thresholds of doubly transitive codes are asymptotically equal. (By the result of [KKM$^+$17], this is equivalent to asking whether or not every doubly transitive code achieves capacity on the erasure channel under block-MAP decoding.)

### 1.1. **Main Results.**

Our main contribution is to develop a framework for bounding the gap between a code's bit and block error thresholds on the erasure channel. Let $C \subseteq \mathbb{F}_2^N$

* DUKE UNIVERSITY, HENRY.PFISTER@DUKE.EDU
† UNIVERSITY OF WASHINGTON, OSPRUM@CS.WASHINGTON.EDU
‡ UNIVERSITÉ DE BORDEAUX AND INSTITUT UNIVERSITAIRE DE FRANCE, ZEMOR@MATH.U-BORDEAUX.FR

be any linear code. We show that if, for all integers $r < r_0$ (where $r_0$ depends on the sharpness of the bit threshold of $C$), any $r$-dimensional subcode of $C$ has support[1] growing faster than $r \log N$, then the block error threshold of $C$ is close to its bit error threshold. An informal version of our main result is stated below. For the formal version, see Theorem 7.

**Theorem 1** (Informal)**.** *Consider any linear code $C \subseteq \mathbb{F}_2^N$ and suppose that a uniform random codeword $c \in C$ is sent over the erasure channel with erasure probability $p \in [0,1]$. Let $\delta \in [0,1]$ be such that, for every $i \in [N]$, the probability we fail to decode the $i^{\text{th}}$ coordinate of the codeword $c$ is bounded by*

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

*Suppose additionally that for all $r = 1, 2, \ldots, \sqrt{\delta}N$, the support of any $r$-dimensional subcode of $C$ has size $\omega(r \log N)$. Then there exists $p' = p - o(1)$ such that the probability we fail to decode the codeword $c$ from erasures with probability $p'$ is bounded by*

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

As a proof of concept, we show in Section 5 that Reed–Muller codes of constant rate satisfy the conditions of Theorem 1. Our analysis relies on the work of Wei, who obtained in [Wei91] exact expressions for the minimum support weight of any $r$-dimensional subcode of a Reed–Muller code. Bounding these expressions appropriately and combining them with Theorem 1 gives an alternative proof that Reed–Muller codes achieve capacity on the erasure channel under block-MAP decoding, a fact that was shown by Kudekar, Kumar, Mondelli, Pfister, Şaşoğlu and Urbanke in [KKM+17, KKM+16]. See Theorem 12.

1.2. **Techniques.** Our proof of Theorem 1 is based on a well-known work of Tillich and Zémor [TZ00], who established sharp threshold results for the symmetric and erasure channels. For a linear code $C \subseteq \mathbb{F}_2^N$ and a string $x \in \{0,1\}^N$, we define

$$S_C(x) := \Big\{c \in C : c_i \leq x_i \text{ for all } i \in [N]\Big\} \tag{1}$$

to be the set of codewords that are indistinguishable from the 0-vector once you erase all coordinates $i \in [N]$ where $x_i = 1$. Note that $S_C(x)$ is in fact a subcode of $C$. Tillich and Zémor showed in [TZ00] that for any fixed $r \in \{1, 2, \ldots, N\}$, the function

$$f_r(p) := \Pr_{x \sim p}\Big[\dim S_C(x) \geq r\Big] \tag{2}$$

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

[1]The support of a subcode $S \subseteq C$ is the set of indices $j \in [N]$ where at least one codeword $c \in S$ has $c_j = 1$.

In this work, we first strengthen their result to show that the curves $f_r(p)$ and $f_{r+1}(p)$ in fact stay within a distance of about $\frac{1}{d_r(C)}$ from each other (see the proof of Theorem 6). We then leverage the fact that every doubly transitive code achieves capacity under bit-MAP decoding [KKM+17] to prove that for some $r_0 = N^{1-o(1)}$, the function $f_{r_0}$ corresponding to any doubly transitive code $C \subseteq \mathbb{F}_2^N$ satisfies

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

Since the distance between the curves $f_r(p)$ and $f_{r+1}(p)$ can be bounded in terms of the minimum support weight $d_r(C)$, we are then able to prove (see Theorem 9) that the function $f_1(p)$ corresponding to any doubly transitive code $C \subseteq \mathbb{F}_2^N$ with large enough minimum support weights satisfies

$$f_1\Big(1 - \text{rate}(C) - o(1)\Big) = o(1). \tag{3}$$

But $f_1\Big(1 - \text{rate}(C) - o(1)\Big)$ is exactly the probability that any sent codeword $c \in C$ can be uniquely recovered from random erasures of probability $1 - \text{rate}(C) - o(1)$, so equation (3) shows that any doubly transitive linear code $C \subseteq \mathbb{F}_2^N$ with large enough minimum support weights $\{d_r(C)\}$ achieves capacity on the erasure channel under block-MAP decoding. See Theorem 7 for the formal statement.

1.3. **Known Results.** Our work makes use of several well-known results, which we state below.

The first result we need is the following bound on the bit error decay of any doubly transitive linear code. It is due to Kudekar, Kumar, Mondelli, Pfister, Şaşoğlu and Urbanke. We state a somewhat informal version of their result below. The more formal statement can be found in Theorem 8.

**Theorem 2** (Informal. Follows from [KKM+17], Lemma 34 and Theorem 19)**.** *Let $C \subseteq \mathbb{F}_2^N$ be any doubly transitive linear code and fix any index $i \in [N]$. Define $p^* \in [0,1]$ to be the erasure probability at which we fail to decode the $i^{\text{th}}$ coordinate of a uniform random codeword $c \in C$ with probability*

$$\Pr_{p^*\text{-erasures}}[\text{can't recover } c_i] = \frac{1}{2}.$$

*Then, for all $p \leq p^*$, we have*

$$\Pr_{p\text{-erasures}}\Big[\text{can't recover } c_i\Big] \leq e^{-(p^*-p)\log(N-1)}.$$

We will also need the following result of Wei, who proved that the subcodes of Reed–Muller codes with smallest supports correspond to Reed–Muller codes of smaller degrees on fewer variables.

**Theorem 3** (follows from [Wei91], Theorem 7)**.** *For every $t \leq d \leq n$, a $\binom{n-t}{\leq d-t}$-dimensional subcode of $\mathsf{RM}(n,d)$ with smallest support is the code spanned by the monomials*

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

*where* $m(x_{t+1}, x_{t+2}, \ldots, x_n)$ *is any monomial of degree* $\leq d-t$ *in the variables* $x_{t+1}, x_{t+2}, \ldots, x_n$.

Finally, we need some tools to understand the function $g_r \colon \{0,1\}^N \to \{0,1\}$ defined by

$$g_r(x) := \begin{cases} 1 & \text{if } \dim S_C(x) \geq r, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $S_C(x)$ is the subcode defined in (1). Note that taking the expectation of $g_r$ gives us the function $f_r$ defined in (2).

For any Boolean function $g \colon \mathbb{F}_2^N \to \{0,1\}$, we define the function

$$h_g(x) := \begin{cases} \left|\left\{i \in [N] : g(x+e_i) = 0\right\}\right| & \text{if } g(x) = 1, \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

and its minimal nonzero value

$$\nu_g := \min\left\{h_g(x) : x \in \mathbb{F}_2^n \text{ such that } h_g(x) \neq 0\right\}. \tag{6}$$

The quantities (5) and (6) were introduced by Margulis [Mar74] and Russo [Rus82] to prove sharp transition results for the expectation of monotone[2] Boolean functions. The following lemma, which relates the derivative of the expectation of any monotone function $g$ to the expectation of its corresponding function $h_g$, is often called the *Margulis-Russo Lemma.*

**Lemma 4** (Margulis-Russo Lemma, [Mar74, Rus82])**.** *For any monotone Boolean function* $g \colon \{0,1\}^N \to \{0,1\}$ *and any* $p \in [0,1]$*, we have*

$$\frac{d}{dp} \mathop{\mathbb{E}}_{x\sim p}[g(x)] = \frac{1}{p} \cdot \mathop{\mathbb{E}}_{x\sim p}[h_g(x)],$$

*where* $x \sim p$ *denotes a* $p$*-noisy random string* $x \in \{0,1\}^N$.

For any $r = 1, 2, \ldots, N$, let $d_r(C)$ denote the minimum size of the support of any $r$-dimensional subcode of $C$. Tillich and Zémor showed that the quantity $\nu_{g_r}$ defined in (4) and (6) can be bounded by $d_r(C)$.

**Lemma 5** (follows from [TZ00], page 476)**.** *Consider any linear code* $C \subseteq \mathbb{F}_2^N$. *For every* $1 \leq r \leq N$*, the corresponding function* $g_r$ *satisfies*

$$\nu_{g_r} \geq d_r(C).$$

*Additionally, for* $1 \leq r \leq N-1$ *we have*

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

See [TZ00] for the proof. Intuitively, the two statements in Lemma 5 follow from the following two facts:

[2]A Boolean function $g : \{0,1\}^N \to \{0,1\}$ is said to be *monotone* if for all $x, y \in \{0,1\}^N$ with $x_i \geq y_i$ for all $i \in [N]$, we have $g(x) \geq g(y)$.

(1) If an index $i \in [N]$ is in the support of the subcode covered by the erasure pattern $x \in \{0,1\}^N$, then half of the codewords $c \in S_C(x)$ have $c_i = 1$. Thus, if you remove the erasure symbol on the $i^{\text{th}}$ coordinate of the erasure pattern $x$, then you only cover half as many codewords as you did before (and thus, the dimension of the covered subcode has gone down by 1).

(2) Removing the erasure symbol on any coordinate $i \in [N]$ of an erasure pattern $x$ can reduce the number of covered codewords by at most half (for every $i \in [N]$, at least half of the codewords $c \in S_C(x)$ have $c_i = 0$). Thus, removing the erasure symbol on any coordinate $i \in [N]$ of an erasure pattern $x$ can decrease the dimension of the covered subcode by at most 1.

1.4. **Comparison with the work of Kudekar et al.** Kudekar, Kumar, Mondelli, Pfister and Urbanke showed in [KKM+16] that on any memoryless channel, if the bit error probability has a sharp enough threshold, then the bit and block error thresholds of Reed–Muller codes are asymptotically equal. One key argument in their proof is the claim that the low-weight codewords do not contribute much to the decoding error probability, and thus one only needs to worry about codewords of high weight. In order to prove this claim, [KKM+16] rely on a very careful analysis of Reed–Muller codes' weight enumerator, which may be difficult to generalize to other codes. In contrast, our approach only requires somewhat loose bounds on the minimum support weight of any subcode. For $r < N^{1-o(1)}$, we need any $r$-dimensional subcode to have support weight larger than $r \log N$; for constant-rate Reed–Muller codes, it turns out that any such subcode has support weight larger than $r2^{(\log N)^{0.99}}$ (see Corollary 11).

## 2. Notation

We will be interested in the behavior of linear codes $C \subseteq \mathbb{F}_2^N$ over the erasure channel $\mathrm{BEC}_p$, for some $p \in [0,1]$. Throughout this paper, we use $\ln X$ to denote the natural logarithm of a real number $X > 0$ and use $\log X$ to denote the logarithm of $X$ in base 2. We denote by $[N]$ the set of integers $\{1, 2, \ldots, N\}$ and use the shorthand

$$x \sim p$$

to signify that the random variable $x \in \mathbb{F}_2^N$ has independent Bernoulli coordinates

$$x_i = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{otherwise.} \end{cases}$$

We will think of the coordinates $j \in [N]$ where $x_j = 1$ as the coordinates that are erased by the channel and will call $x$ the *erasure pattern*. Formally, when a sender sends some codeword $c \in C$ through the channel $\mathrm{BEC}_p$, an erasure pattern $x \sim p$ is sampled and the receiver receives a vector $y \in \{0, 1, *\}^N$ with coordinates

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

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

$$x \succ c.$$

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

$$S_C(x) := \Big\{c \in C : x \succ c\Big\}.$$

A sent codeword $c \in C$ can be recovered uniquely from an erasure pattern $x \in \mathbb{F}_2^N$ if and only if $S_C(x) = \{0\}$. Similarly, for any coordinate $i \in [N]$, the bit $c_i$ can be recovered if and only if $i \notin \operatorname{supp}\big(S_C(x)\big)$, where for any subcode $S \subseteq C$ we define

$$\operatorname{supp}\big(S\big) := \Big\{j \in [N] : \exists c \in S \text{ with } c_j = 1\Big\} \tag{7}$$

to be the *support* of $S$. For every $r = 1, 2, \ldots, N$, we define the function

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

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

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

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

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

The main quantity that will allow us to control the behavior of these two functions is the minimum support size of any $r$-dimensional subcode of $C$,

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

One explicit family of codes we will be interested in is the family of Reed–Muller codes. We will denote by $\mathsf{RM}(n,d)$ the Reed–Muller code with $n$ variables and degree $d$. The codewords of $\mathsf{RM}(n,d)$ are the evaluation vectors (over all points in $\mathbb{F}_2^n$) of all multivariate polynomials of degree $\leq d$ in $n$ variables. The Reed–Muller code $\mathsf{RM}(n,d)$ has dimension $\binom{n}{\leq d}$ (see for instance page 5 of [ASY21]).

## 3. General Linear Codes

In this section, we will provide general conditions under which the bit and block error thresholds of an arbitrary linear code $C \subseteq \mathbb{F}_2^N$ are close to one another. We start by bounding the distance between the curves $\theta_1(\alpha)$ and $\theta_{N_0}(\alpha)$, for any $N_0 > 1$.

**Theorem 6.** *Fix any linear code $C \subseteq \mathbb{F}_2^N$ and any integer $N_0 \leq \dim C$. Then, letting $\gamma_{N_0} := \sqrt{\sum_{r=1}^{N_0-1} \frac{1}{d_r(C)}}$, we have that for any*

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

*the functions $\theta_1$ and $\theta_{N_0}$ associated with $C$ satisfy*

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

*Proof.* Consider any $r \leq \dim C$. By Lemmas 4 and 5, the function $f_r$ associated with our code $C$ satisfies

$$\begin{aligned} \frac{d}{dp} f_r(p) &\geq \Pr_{x\sim p}\left[h_{f_r}(x) \neq 0\right] \cdot d_r(C) \\ &= \Big(f_r(p) - f_{r+1}(p)\Big) d_r(C). \end{aligned} \tag{8}$$

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

$$\begin{aligned} \int_0^1 \Big(\theta_{r+1}(\alpha) - \theta_r(\alpha)\Big) d\alpha &= \int_0^1 \Big(f_r(p) - f_{r+1}(p)\Big) dp \\ &\leq \frac{1}{d_r(C)}, \end{aligned} \tag{9}$$

where the last line follows from applying the Fundamental Theorem of Calculus to (8). Taking $\alpha$ uniformly at random from the interval $[0, 1]$ and applying (9) with $r = 1, 2, \ldots, N_0 - 1$ then gives

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

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

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

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

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

Since the functions $\theta_r$ are increasing, it follows that

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

$\square$

We are now ready to prove our main result. The following theorem is a formal version of Theorem 1.

**Theorem 7.** *Consider any linear code $C \subseteq \mathbb{F}_2^N$ with $N \geq 10$. Let $p, \delta \in [0,1]$ be such that* $\dim C \geq \sqrt{\delta} N$ *and*

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

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

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

*Then we have*

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

*Proof.* Note that we may assume that

$$\sqrt{\delta} + \sqrt{\frac{\log N}{\Delta}} \leq 1, \tag{10}$$

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

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

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

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

or equivalently,

$$\theta_{\sqrt{\delta}N}(\sqrt{\delta}) \geq p. \tag{11}$$

On the other hand, by definition of $\Delta$ we have

$$\begin{aligned}\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{d_r(C)} &\leq \frac{1}{\Delta}\sum_{r=1}^{\sqrt{\delta}N} \frac{1}{r} \\ &\leq \frac{\log N}{\Delta},\end{aligned} \tag{12}$$

where in the last line we used the assumption that $N \geq 10$ and the fact that $\sum_{r=1}^{N'} \frac{1}{r} \leq \ln(N')+1$ for all $N' \geq 1$. Combining this inequality with (10), we get that the conditions of Theorem 6 are satisfied for $N_0 = \sqrt{\delta}N$ and $\alpha = \sqrt{\delta}$. Applying Theorem 6, we then have

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

where in the first and third lines we used the inequality (12). By equation (11), we get

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

The inequality above is equivalent to

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

This completes the proof. $\square$

## 4. Doubly Transitive Codes

In this section, we will apply our Theorem 7 to doubly transitive codes. We will need the following celebrated result, which is a formal version of Theorem 2.

**Theorem 8** (follows from [KKM+17], Lemma 34 and Theorem 19)**.** *Let $C \subseteq \mathbb{F}_2^N$ be a doubly transitive code, and fix any index $i \in [N]$. Define $p^* \in [0,1]$ to be the noise parameter at which*

$$\Pr_{x \sim p^*}[i \in \operatorname{supp}(S(x))] = \frac{1}{2}.$$

*Then for all $p \leq p^*$, we have*

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

Combining Theorem 8 with our Theorem 7 yields the following bound on the gap between the bit and block error thresholds of any doubly transitive linear code.

**Theorem 9.** *Let $C \subseteq \mathbb{F}_2^N$ be any doubly transitive linear code with $N \geq 10$. Define $p^* \in [0,1]$ to be the noise parameter at which $\Pr_{x\sim p^*}[1 \in S(x)] = \frac{1}{2}$, and fix any $p \leq p^*$. Suppose* $\dim C \geq Ne^{-\frac{p^*-p}{2}\log(N-1)}$*. Then defining*

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

*we have*

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

*Proof.* By Theorem 8, we have that for all $i \in [N]$,

$$\Pr_{x\sim p}\left[i \in \operatorname{supp}\bigl(S(x)\bigr)\right] \leq e^{-(p^*-p)\log(N-1)}.$$

Applying Theorem 7 with $\delta = e^{-(p^*-p)\log(N-1)}$, we then get

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

□

## 5. Reed–Muller Codes

In this section, we will use our Theorem 9 for doubly transitive codes to show that Reed–Muller codes achieve capacity on the erasure channel. For any $t \leq d$, define the subcode $S_t \subseteq \mathsf{RM}(n,d)$ to be the set of evaluation vectors for all polynomials of the form

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

where $p(x_{t+1}, x_{t+2}, \ldots, x_n)$ is any polynomial of degree $\leq d-t$ in the variables $x_{t+1}, x_{t+2}, \ldots, x_n$. Recall that Wei showed in [Wei91] that $S_t$ is a subcode of minimal support for its dimension (see Theorem 3). We now bound its dimension and support explicitly.

**Lemma 10.** *For every $n \in \mathbb{N}$, every $d \leq \frac{n}{2} + \sqrt{n\log n}$ and every $t \in [5\sqrt{n\log n}, d]$, the corresponding subcode $S_t \subseteq \mathsf{RM}(n,d)$ satisfies*

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

*and*

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

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

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

where in the second line we used the fact that $d \leq \frac{n}{2} + \sqrt{n \log n} \leq \frac{n}{2} + \frac{t}{5}$. Since $\binom{m}{\leq s} \leq 2^{h(\frac{s}{m})m}$ for all $s \leq \frac{m}{2}$ (see for example [Gal14], Theorem 3.1) we then get

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

where in the second line we used the fact that $\frac{\frac{1}{2} - \frac{4x}{5}}{1-x} \leq \frac{1}{2} - \frac{3x}{10}$ for all $x \in [0,1)$ (see Appendix A for the proof). Applying the inequality $h(\frac{1-x}{2}) \leq 1 - \frac{x^2}{2 \ln 2}$, we then get

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

as desired. □

In particular, we get the following bound on the minimum size of the support of any $r$-dimensional subcode of a Reed–Muller code.

**Corollary 11.** *For every $n$ large enough, every $d \leq \frac{n}{2} + \sqrt{n \log n}$ and every $\epsilon \in \left[6\sqrt{\frac{\log n}{n}}, \frac{1}{2}\right]$, we have*

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

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

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

$$2^{n-t-1-\frac{\epsilon^2 n}{10}} \leq r \leq 2^{n-t-\frac{\epsilon^2 n}{10}}. \tag{13}$$

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

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

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

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

where in the second line we used the fact that $\epsilon \leq \frac{1}{2}$. Combining this with the leftmost inequality of (13), we get $r \geq 2^{n-t-1-\frac{\epsilon^2 n}{10}} \geq \dim(S_t)$. By Theorem 3 and Lemma 10, we must then have

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

Combining this inequality with the right-hand side of (13), we get

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

□

We are now ready to prove that the bit and block error thresholds of Reed–Muller codes are asymptotically equal. Since every doubly transitive code achieves capacity under bit-MAP decoding [KKM+17], this implies that Reed–Muller codes achieve capacity under block-MAP decoding.

**Theorem 12.** *For every $n$ large enough, every $d \in [\frac{n}{2} - \sqrt{n \log n}, \frac{n}{2} + \sqrt{n \log n}]$ and every $\epsilon \geq 20\sqrt{\frac{\log n}{n}}$, the Reed–Muller code $\mathsf{RM}(n,d)$ satisfies*

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

*where $p^* \in [0,1]$ denotes the noise value at which $\Pr_{x\sim p^*}\big[1 \in \mathrm{supp}(S(x))\big] = \frac{1}{2}$.*

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

$$\begin{aligned} \dim\big(\mathsf{RM}(n,d)\big) &= \binom{n}{\leq d} \\ &\geq \binom{n}{\leq \frac{n}{2} - \sqrt{n \log n}} \\ &\geq \frac{1}{\sqrt{2n}} \cdot 2^{h\left(\frac{1}{2} - \sqrt{\frac{\log n}{n}}\right)n} \\ &\geq \frac{1}{\sqrt{2n}} 2^{(1 - 4\frac{\log n}{n})n} \\ &\geq 2^n e^{-\frac{\epsilon(n-1)}{6}}, \end{aligned}$$

where in the third line we used the inequality $\binom{n}{k} \geq \frac{1}{\sqrt{2n}} \cdot 2^{h(k/n)n}$ (see for example [MS77], page 309, Lemma 7) and in the fourth line we used the inequality $h(\frac{1-x}{2}) \geq 1 - x^2$. By Theorem 9, for every $p \leq p^* - \frac{\epsilon}{3}$, we have

$$\begin{aligned} \Pr_{x \sim p - \sqrt{\frac{n}{\Delta}}}\Big[\exists c \in C : x \succ c\Big] &\leq \sqrt{\frac{n}{\Delta}} + e^{-\frac{p^*-p}{2}\log(2^n - 1)} \\ &\leq \sqrt{\frac{n}{\Delta}} + e^{-\frac{p^*-p}{2}(n-1)}, \end{aligned} \tag{14}$$

with

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

Letting $p = p^* - \frac{2\epsilon}{3} \cdot \frac{n}{n-1}$ and applying Corollary 11 with parameter $\frac{\epsilon}{3\ln 2}$ (the conditions of Corollary 11 are satisfied by our theorem's condition on $\epsilon$), we get

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

and thus equation (14) becomes

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

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

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

□

**Remark 1.** *The work of Tillich and Zémor gives very sharp block error decays in terms of the minimum distance of a linear code* [TZ00]*. Thus, the primary purpose here is to bound the distance in p between the block-error and bit-error thresholds. Once we have a bound such as the one given by our Theorem 12, we immediately get a strong error decay by* [TZ00]*.*

**Remark 2.** *For Reed–Muller codes specifically, the ratios* $\{\frac{d_r}{r}\}$ *are large enough that one does not need the full power of Theorem 8's bit-error decay,* $\Pr_{x\sim p}[\text{bit error}] \leq e^{-(p^*-p)n}$*. To show that the bit-error and block-error thresholds of Reed–Muller codes are asymptotically equal, it would have been sufficient to use a bit error decay of* $e^{-(p^*-p)a\sqrt{n\log n}}$ *for any* $a = \omega(1)$*. This is because by Corollary 11, we have* $\frac{d_r}{r} \geq n^{10}$ *for all* $r \leq 2^{n-10\sqrt{n\log n}}$*. Setting* $p = p^* - \frac{20\ln 2}{a}$ *and applying Theorem 7 would then have given that the block-error probability under noise* $p^* - \frac{20\ln 2}{a} - \frac{1}{n^4}$ *is bounded by* $\frac{2}{n^4}$*.*


## Acknowledgments

This work was initiated while the authors were visiting the Simons Institute for the Theory of Computing. The work of Henry D. Pfister was supported in part by NSF CCF-2106213. The work of Oscar Sprumont was supported in part by NSF CCF-2131899, NSF CCF-1813135 and Anna Karlin's Bill and Melinda Gates Endowed Chair. The work of Gilles Zémor was supported in part by Plan France 2030 through the project NISQ2LSQ, ANR-22-PETQ-0006.

## Appendix A. An Inequality for Lemma 10

We prove the following inequality, which was used in the proof of Lemma 10.

**Claim 13.** *For all $x \in [0, 1)$, we have*

$$\frac{\frac{1}{2} - \frac{4x}{5}}{1 - x} \leq \frac{1}{2} - \frac{3x}{10}.$$

*Proof.* Note that we have equality at $x = 0$. It will thus suffice to show that the function

$$f(x) := \frac{1}{2} - \frac{3x}{10} - \frac{\frac{1}{2} - \frac{4x}{5}}{1 - x}$$

is increasing over $[0, 1)$. We compute the derivative of $f$,

$$\begin{aligned} \frac{df}{dx} &= -\frac{3}{10} - \frac{-\frac{4}{5}(1 - x) + \frac{1}{2} - \frac{4x}{5}}{(1 - x)^2} \\ &= -\frac{3}{10} + \frac{\frac{4}{5} - \frac{1}{2}}{(1 - x)^2} \\ &= -\frac{3}{10} + \frac{3}{10(1 - x)^2} \\ &\geq 0 \end{aligned}$$

for every $x \in [0, 1)$.

□

## References


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

[AS23a] Emmanuel Abbe and Colin Sandon. A proof that Reed-Muller codes achieve Shannon capacity on symmetric channels. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 177–193. IEEE, 2023.

[AS23b] Emmanuel Abbe and Colin Sandon. Reed-Muller codes have vanishing bit-error probability below capacity: a simple tighter proof via camellia boosting. *CoRR*, abs/2312.04329, 2023.

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

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

[For66] George D. Forney. Concatenated codes. *MIT Press*, 1966.

[Gal14] David Galvin. Three tutorial lectures on entropy and counting, 2014.

[KKM+16] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, and Rüdiger L. Urbanke. Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels. In *IEEE International Symposium on Information Theory, ISIT 2016, Barcelona, Spain, July 10-15, 2016*, pages 1755–1759. IEEE, 2016.

[KKM+17] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Sasoglu, and Rüdiger L. Urbanke. Reed-Muller codes achieve capacity on erasure channels. *IEEE Trans. Inf. Theory*, 63(7):4298–4316, 2017.

[Mar74] Grigory A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problems of Information Transmission*, 10(2):101–108, 1974.

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

[RP24] Galen Reeves and Henry D. Pfister. Reed-Muller codes on BMS channels achieve vanishing bit-error probability for all rates below capacity. *IEEE Trans. Inf. Theory*, 70(2):920–949, 2024.

[Rus82] Lucio Russo. An approximate zero-one law. *Probability Theory and Related Fields*, 61(1):129–139, 1982.

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

[TZ00] Jean-Pierre Tillich and Gilles Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Comb. Probab. Comput.*, 9(5):465–479, 2000.

[Wei91] Victor K.-W. Wei. Generalized Hamming weights for linear codes. *IEEE Trans. Inf. Theory*, 37(5):1412–1418, 1991.