Reed-Muller Codes Achieve Capacity on Erasure Channels*

Shrinivas Kudekar
Qualcomm Research
New Jersey, USA
skudekar@qti.qualcomm.com

Santhosh Kumar
Texas A&M University
College Station, Texas, USA
santhosh.kumar@tamu.edu

Marco Mondelli
EPFL
Lausanne, Switzerland
marco.mondelli@epfl.ch

Henry D. Pfister
Duke University
Durham, North Carolina, USA
henry.pfister@duke.edu

Eren Şaşoğlu
Intel Corporation
Santa Clara, California, USA
eren.sasoglu@gmail.com

Rüdiger Urbanke
EPFL
Lausanne, Switzerland
ruediger.urbanke@epfl.ch

ABSTRACT

We introduce a new approach to proving that a sequence of deterministic linear codes achieves capacity on an erasure channel under maximum a posteriori decoding. Rather than relying on the precise structure of the codes, our method exploits code symmetry. In particular, the technique applies to any sequence of linear codes where the block lengths are strictly increasing, the code rates converge, and the permutation group of each code is doubly transitive. In a nutshell, we show that symmetry alone implies near-optimal performance.

An important consequence of this result is that a sequence of Reed-Muller codes with increasing block length and converging rate achieves capacity. This possibility has been suggested previously in the literature, but it has only been proven for cases where the limiting code rate is 0 or 1. Moreover, these results extend naturally to affine-invariant codes and, thus, to all extended primitive narrow-sense BCH codes. This is used to resolve, in the affirmative, the existence question for capacity-achieving sequences of binary cyclic codes. The primary tools used in the proofs are the sharp threshold property for symmetric monotone boolean functions and the area theorem for extrinsic information transfer (EXIT) functions.

Categories and Subject Descriptors

E.4 [**Coding and Information Theory**]: Error control codes

General Terms

Theory

*An extended version of this paper is available as [43].

Keywords

Affine-invariant codes, BCH codes, capacity-achieving codes, erasure channels, EXIT functions, linear codes, MAP decoding, monotone boolean functions, Reed-Muller codes.

1. INTRODUCTION

1.1 Overview

Since the introduction of channel capacity by Shannon in his seminal paper [65], theorists have been fascinated by the idea of constructing codes that achieve *capacity* (e.g., under optimal decoding). Ideally, one would also like these codes to have low-complexity encoding/decoding *algorithms*, algebraic or geometric *structure*, and *deterministic* constructions.

The advent of Turbo codes [9] and low-density parity-check (LDPC) codes [31, 52, 68] has made it possible to construct practical codes that achieve good performance near the Shannon limit. It was even proven that sequences of irregular LDPC codes can achieve *capacity* on the binary erasure channel (BEC) using low-complexity message-passing *algorithms* [51].

Recently, spatially-coupled LDPC codes were shown to achieve *capacity* universally over the class of binary memoryless symmetric (BMS) channels using low-complexity *algorithms* [45–48]. In regard to the other desirable properties, these codes also have some *structure* (e.g., low-density graph structure) but their construction is not *deterministic*.

For an arbitrary BMS channel, however, polar codes [6] were the first codes proven to achieve *capacity* with low-complexity encoding and decoding *algorithms*. In addition, polar codes inherit some *structure* from the Hadamard matrix and also have a *deterministic* construction.

This article considers the performance of *structured* and *deterministic* binary linear codes transmitted over the BEC under bitwise maximum-a-posteriori (MAP) decoding. In particular, our primary technical result is the following.

Theorem: A sequence of linear codes achieves capacity on a memoryless erasure channel under bit-MAP decoding if its block lengths are strictly increasing, its code rates converge

to some $r \in (0,1)$, and the permutation group[1] of each code is doubly transitive.

The analysis focuses primarily on the bit erasure rate under bit-MAP decoding, but it can be extended to the block erasure rate in some cases. One important consequence is a proof of the fact that binary Reed-Muller codes achieve capacity on the BEC under block-MAP decoding, which settles a rather old conjecture in coding theory.

The main theorem for bit-MAP decoding applies also to affine invariant codes and it extends naturally from binary linear codes to $\mathbb{F}_q$-linear codes transmitted over a q-ary erasure channel. Additionally, we show that extended primitive narrow-sense BCH codes achieve capacity under block-MAP decoding. Finally, this allows us to resolve, in the affirmative, the existence question for capacity-achieving sequences of binary cyclic codes [59].

These results are perhaps surprising. Until the discovery of polar codes, it was unclear whether or not codes with a simple deterministic structure could achieve capacity [4,17,19]. But even though polar codes (as well as Reed-Muller codes) derive from the Hadamard matrix, the ability of polar codes to achieve capacity appears unrelated to the inherent symmetry of this matrix. In contrast, the performance guarantees obtained here are a consequence only of linearity and the structure induced by the doubly-transitive permutation group.

1.2 Reed-Muller Codes

Reed-Muller codes were introduced by Muller in [58] and, soon after, Reed proposed a majority logic decoder in [60]. For integers v, n satisfying $0 \le v \le n$, a binary Reed-Muller code $\mathrm{RM}(v,n)$ is a linear code of length 2^n and dimension $\binom{n}{0} + \cdots + \binom{n}{v}$. It is well known that the minimum distance of this code is 2^{n-v} [20, 50, 53]. Thus, it is impossible to simultaneously have a non-vanishing rate and a minimum distance that scales linearly with block length.

The idea that Reed-Muller codes might achieve capacity appears to be rather old. In a personal communication with Shu Lin, we learned that this possibility was discussed privately by Kasami, Lin, and Peterson in the late 1960s. Later the idea was mentioned explicitly in a 1993 talk by Shu Lin, entitled "RM Codes are Not So Bad" [49]. To the best of the authors' knowledge, a 1994 paper by Dumer and Farrell contains the earliest printed discussion of this question [27]. In that paper, they show that some sequences of BCH codes with rates approaching 1 have a vanishing gap to capacity on the BEC. They also suggest, as an open problem, the evaluation of a quantity which equals 1 if and only if Reed-Muller codes achieve capacity on the BEC. Since then, similar ideas have been discussed by a variety of authors [1,2,6,7,16,19,21,57]. In particular, short Reed-Muller codes with erasures were investigated in [16, 21] and it was observed numerically that the block erasure rate is quite close to that of random codes. In [57], a modified construction of polar codes is analyzed and the results again suggest that Reed-Muller codes achieve capacity on the BEC. For rates approaching either 0 or 1 with sufficient speed, it has recently been shown by Abbe et al. that Reed-Muller codes can correct almost all erasure patterns up to the capacity limit[2] [1, 2]. Beyond erasure channels, it is conjectured in [19] that the sequence of rate-1/2 self-dual Reed-Muller codes achieves capacity on the binary-input AWGN channel.

More than 50 years after their discovery, Reed-Muller codes remain an active area of research in theoretical computer science and coding theory. The early works in [39, 40, 67] culminated in obtaining asymptotically tight bounds for their weight distribution for the case of fixed order v and asymptotic n [41]. Also, there is considerable interest in constructing low-complexity decoding algorithms, see [63, 66] and a series of papers by Dumer et al. [25, 26, 28]. Undoubtedly, the interest in the coding theory community for these codes was rekindled by the tremendous success of polar codes and their close connection to Reed-Muller codes [5, 6, 57].

Due to their desirable structure, constructions based on these codes are used extensively in cryptography [13–16, 22, 34, 64, 69]. Reed-Muller codes are also known for their locality [75] and some of the earliest known constructions for the locally correctable codes are based on them [32, 33]. Interestingly, the local correctability of Reed-Muller codes is also a consequence of their permutation group being doubly transitive [42], a crucial requirement in our approach. However, a doubly transitive permutation group is not sufficient for local testability [35].

1.3 Outline of the Proof

The central object in our analysis is the extrinsic information transfer (EXIT) function. EXIT charts were introduced by ten Brink [72] in the context of turbo decoding as a visual tool to understand iterative decoding. For a given input bit, the EXIT function is defined to be the conditional entropy of the input bit given the outputs associated with *all other* input bits. The *average* EXIT function is formed by averaging all of the bit EXIT functions. We note that these functions are also instrumental in the design and analysis of LDPC codes [61].

The crucial property we exploit is the so called area theorem, originally proved in [8] and further generalized in [56], which says that the area under the average EXIT function equals the rate of the code. The average EXIT function is also directly related to the bit erasure probability under MAP decoding. Indeed, for a sequence of binary linear codes with rate r to be capacity achieving, the average EXIT function must converge to 0 for any erasure rate below $1-r$. Since the areas under the average EXIT curves are fixed to r, the EXIT functions for these codes must also converge to 1 for any erasure rate above $1-r$. Thus, the EXIT curves must exhibit a *sharp transition* from 0 to 1, and, as a consequence of area theorem, this transition must occur at the erasure value of $1-r$.

We investigate the threshold behavior of EXIT functions for binary linear codes via sharp thresholds for monotone boolean functions [10, 37]. The general method was pioneered by Margulis [54] and Russo [62]. Later, it was significantly generalized by Talagrand in [70] and [71]. This approach has been applied to many problems in theoretical computer science with remarkable success [24, 29, 30]. In the context of coding theory, this technique was first introduced by Zémor in [76], refined further in [73], and also extended to

[1] The permutation group of a linear code is the set of permutations on code bits under which the code is invariant.

[2] It requires some effort to define precisely what capacity limit is for rates approaching 0 or 1. See [2, Definition 16] for details.

AWGN channels in [74]. For the BEC, it is shown in [73,76] that the block erasure rate jumps from 0 to 1 as the minimum distance of the code grows. However, focusing on the block erasure rate does not allow one to establish the location of the threshold. In order to show the threshold behavior for EXIT functions, we instead focus on symmetry [30] which follows if the codes have doubly-transitive permutation groups.

2. PRELIMINARIES

2.1 Basic Setup and Notation

A linear code is proper if no codeword position is 0 in all codewords. In the following, all codes are understood to be proper binary linear codes with minimum distance at least 2. Let $\mathcal{C}$ denote an (N, K) binary linear code with length N and dimension K. The rate of this code is given by $r \triangleq K/N$.

We say that a sequence $\underline{a}$ *covers* a sequence $\underline{b}$, namely $\underline{a} \geq \underline{b}$, if $a_i \geq b_i$ for all i. Denote $[N] \triangleq \{1, \cdots, N\}$. A set $A \subseteq [N]$ is said to *cover* a sequence $\underline{a} \in \{0,1\}^N$ if the set of non-zero indices of $\underline{a}$ is a subset of A. Let $H(\cdot)$ and $H(\cdot|\cdot)$ be the entropy and conditional entropy of a discrete random variable in bits, respectively. Below, all logarithms are natural unless the base is explicitly mentioned.

Let π be a permutation on N elements, i.e., $\pi : [N] \to [N]$, a bijection. Let $\underline{x}$ be a vector with components indexed by $[N]$. Abusing notation, we will also let $\pi(\underline{x})$ denote a length-N vector, say $\underline{z}$, with components satisfying $z_{\pi(i)} = x_i$.

2.2 Bit and Block Erasure Probability

The input and output alphabets of the BEC are denoted by $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,*\}$, respectively, where $*$ denotes the erasure symbol. Let $\underline{X} = (X_1, \ldots, X_N) \in \mathcal{X}^N$ be a uniformly random codeword and $\underline{Y} = (Y_1, \ldots, Y_N) \in \mathcal{Y}^N$ be the received sequence obtained by transmitting $\underline{X}$ through a memoryless BEC(p), for $p \in (0,1)$. The erasure pattern of a received sequence in $\mathcal{Y}^N$ is the binary sequence in $\{0,1\}^N$ that indicates the positions of erasures.

For linear codes and erasure channels, it is possible to recover the transmitted codeword if and only if the erasure pattern does not cover any non-zero codeword [61, Section 3.2.1]. Similarly, for $i \in [N]$, it is possible to recover bit i from $\underline{Y} = \underline{y}$ if and only if the erasure pattern does not cover any codeword where bit i is non-zero. In this case, $H(X_i|\underline{Y} = \underline{y}) = 0$. Whenever bit i cannot be recovered uniquely, the linearity of the code implies that the set of codewords matching the unerased observations has an equal number of 0's and 1's in bit position i [61, Section 3.2.2]. In this case, the uniform codeword assumption implies that $\Pr(X_i = x|\underline{Y} = \underline{y}) = \frac{1}{2}$ and, therefore, $H(X_i|\underline{Y} = \underline{y}) = 1$.

Let $D_i \colon \mathcal{Y}^N \to \mathcal{X} \cup \{*\}$ denote the bit-MAP decoder for bit i. For a received sequence $\underline{Y}$, if X_i can be recovered uniquely, then $D_i(\underline{Y}) = X_i$. Otherwise, D_i declares an erasure and returns $*$. Thus, the erasure probability for bit i and the average erasure probability are given by

$$P_{b,i} \triangleq \Pr(D_i(\underline{Y}) = *) = H(X_i|\underline{Y}),$$
$$P_b \triangleq \frac{1}{N}\sum_{i=1}^{N} P_{b,i} = \frac{1}{N}\sum_{i=1}^{N} H(X_i|\underline{Y}).$$

Let $D \colon \mathcal{Y}^N \to \mathcal{X}^N \cup \{*\}$ denote the block-MAP decoder. Given a received sequence $\underline{Y}$, $D(\underline{Y})$ is equal to $\underline{X}$ whenever it is possible to uniquely recover $\underline{X}$ from $\underline{Y}$. Otherwise, D declares an erasure and returns $*$. Therefore, the block erasure probability is given by

$$P_B \triangleq \Pr(D(\underline{Y}) \neq \underline{X}).$$

Using the set equivalence

$$\{D(\underline{Y}) \neq \underline{X}\} = \bigcup_{i \in [N]} \{D_i(\underline{Y}) \neq X_i\}$$

and the union bound, it is easy to see that

$$P_B \leq N P_b.$$

Also, if D declares an erasure, there will be at least $d_{\min}$ bits in erasure, where $d_{\min}$ is the minimum distance of the code $\mathcal{C}$. Therefore,

$$d_{\min} \mathbf{1}_{\{D(\underline{Y}) \neq \underline{X}\}} \leq \sum_{i \in [N]} \mathbf{1}_{\{D_i(\underline{Y}) \neq X_i\}}.$$

Taking expectations on both sides gives a tighter bound on P_B in terms of P_b,

$$P_B \leq \frac{N}{d_{\min}} P_b. \tag{1}$$

2.3 MAP EXIT Functions

The extrinsic information transfer functions are closely related to the bit erasure probabilities. The EXIT function associated with bit i and the average EXIT function are defined by,

$$h_i(p) \triangleq H(X_i|\underline{Y}_{\sim i}), \qquad h(p) \triangleq \frac{1}{N}\sum_{i=1}^{N} H(X_i|\underline{Y}_{\sim i}),$$

respectively, where $\underline{Y}_{\sim i} \triangleq (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_N)$ and p represents the erasure probability of the channel. The erasure probability of bit i can be written, in terms of the EXIT function, as

$$\begin{aligned} P_{b,i} = H(X_i|\underline{Y}) &= \Pr(Y_i = X_i) \underbrace{H(X_i|\underline{Y}_{\sim i}, Y_i = X_i)}_{=0} \\ &\quad + \underbrace{\Pr(Y_i = *)}_{=p} \underbrace{H(X_i|\underline{Y}_{\sim i}, Y_i = *)}_{=H(X_i|\underline{Y}_{\sim i}) = h_i(p)} \\ &= p h_i(p). \end{aligned}$$

It follows immediately that

$$P_b(p) = p h(p). \tag{2}$$

The following two propositions restate some known results in the notation of this paper. The area theorem, stated in Proposition 1, first appeared in [8, Theorem 1]. The proof we give below for completeness is from [56]. The explicit evaluation of $h_i(p)$, stated in Proposition 2, is a restatement of [61, Lemma 3.74(iv)].

Proposition 1 (Area Theorem)**:** For a code $\mathcal{C}$ with rate r and transmission over a BEC, the average EXIT function satisfies

$$\int_0^1 h(p) \mathrm{d}p = r.$$

Proof. For mathematical convenience, we assume that X_i is transmitted through a BEC with parameter p_i, and denote the received vector by $\underline{Y}(\underline{p})$ to emphasize this. Then,

$$H(\underline{X}|\underline{Y}(\underline{p})) \stackrel{(a)}{=} H(X_i|\underline{Y}(\underline{p})) + H(\underline{X}_{\sim i}|X_i, \underline{Y}(\underline{p})) \stackrel{(b)}{=} p_i H(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})) + H(\underline{X}_{\sim i}|X_i, \underline{Y}_{\sim i}),$$

where (a) follows from the chain rule of entropy, and (b) uses the same procedure as (2) to obtain the first term and the fact that $\underline{X}_{\sim i}$ and Y_i are conditionally independent given X_i to obtain the second term. Since $H(\underline{X}_{\sim i}|X_i, \underline{Y}_{\sim i})$ is independent of p_i,

$$\frac{\partial H(\underline{X}|\underline{Y}(\underline{p}))}{\partial p_i}\Big|_{\underline{p}=(p,\ldots,p)} = H(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i}))\Big|_{\underline{p}=(p,\ldots,p)} = h_i(p).$$

By using the total derivative rule, we gather that

$$\frac{\mathrm{d}}{\mathrm{d}p}\left[H(\underline{X}|\underline{Y}(\underline{p}))\Big|_{\underline{p}=(p,\ldots,p)}\right] = \sum_{i=1}^{N} \frac{\partial H(\underline{X}|\underline{Y}(\underline{p}))}{\partial p_i}\Big|_{\underline{p}=(p,\ldots,p)} = \sum_{i=1}^{N} h_i(p) = N h(p).$$

By integrating the above equation from 0 to 1 and by observing that $H(\underline{X}|\underline{Y}(\underline{0})) = 0$ and $H(\underline{X}|\underline{Y}(\underline{1})) = H(\underline{X}) = K$, the result follows. □

Proposition 2: Consider transmission over the BEC(p) of a code $\mathcal{C}$ of block length N and define

$$\Omega_i = \{\underline{z} \in \{0,1\}^{N-1} \mid \underline{z} \geq \underline{x}_{\sim i}, x_i = 1, \underline{x} \in \mathcal{C}\}. \quad (3)$$

Then, the EXIT function associated with bit i explicitly evaluates to

$$h_i(p) = \sum_{\underline{z} \in \Omega_i} p^{|\underline{z}|}(1-p)^{N-1-|\underline{z}|}.$$

Proof. The definition of h_i implies

$$h_i(p) = H(X_i|\underline{Y}_{\sim i}) = \sum_{\underline{y}_{\sim i} \in \mathcal{Y}^{N-1}} \Pr(\underline{Y}_{\sim i} = \underline{y}_{\sim i}) H\Big(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}\Big).$$

The fact that the decoding process is successful depends only on the erasure pattern in $\underline{Y} = \underline{y}$. Hence, we can assume that the all-zero codeword has been transmitted. In such a case, for $\ell \in [N]$, either $y_\ell = 0$ or $y_\ell = *$. Let $A \subseteq [N]\backslash\{i\}$ be the set of indices where $y_\ell = *$. Then,

$$\Pr(\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = p^{|A|}(1-p)^{N-1-|A|}.$$

If $A \cup \{i\}$ covers a codeword in $\mathcal{C}$ whose i-th bit is non-zero, then the bit-MAP decoder fails to decode bit i and $H(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1$. If $A \cup \{i\}$ does not cover any codeword in $\mathcal{C}$ with non-zero bit i, then bit i can be correctly recovered and $H(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 0$.

Thus, the EXIT function $h_i(p)$ is given by summing over the set of erasure patterns where the entropy is 1. This set is precisely Ω_i, the set of all erasure patterns that cover a codeword whose i-th bit is non-zero. □

By assumption, the code $\mathcal{C}$ is proper and has a minimum distance of at least 2. As such, Ω_i is non-empty (it contains the all-one sequence) and also does not contain the all-zero sequence. Thus, $h(0) = h_i(0) = 0$ and $h(1) = h_i(1) = 1$. Further, from the above proposition, the EXIT functions $h_i(p)$, and therefore $h(p)$, are continuous, polynomial functions on $[0,1]$. Under the assumptions on the code, it is also possible to show that $h_i(p)$ and $h(p)$ are strictly increasing.

Definition 3 (Monotone Sets)**:** A non-empty proper subset $\Omega \subset \{0,1\}^M$ is called *monotone* if $\underline{x} \in \Omega$ and $\underline{x} \leq \underline{y}$ imply that $\underline{y} \in \Omega$.

Monotone sets appear frequently in the theory of random graphs and satisfiability problems. The following observation relates this concept to our setting.

Remark 4 (Monotonicity of Ω_i)**:** It is easy to verify that the set Ω_i defined in (3) is monotone. This is because adding erasures to the received vector can only prevent the recovery of bit i.

2.4 Permutations of Linear Codes

Let S_N be the symmetric group on N elements. The permutation group of a code is defined as the subgroup of S_N whose group action on the bit ordering preserves the set of codewords [36, Section 1.6].

Definition 5 (Permutation Group)**:** The *permutation group* $\mathcal{G}$ of a code $\mathcal{C}$ is defined to be

$$\mathcal{G} \triangleq \{\pi \in S_N \mid \pi(\underline{x}) \in \mathcal{C} \text{ for all } \underline{x} \in \mathcal{C}\}.$$

Proposition 6: The permutation group allows the following properties on the bit-MAP decoders.

a) If $\pi \in \mathcal{G}$ is such that $\pi(i) = j$, then $D_i(\underline{y}) = D_j(\pi(\underline{y}))$.

b) For $\pi \in \mathcal{G}$, $\Pr(D_i(\underline{Y}) = *) = \Pr(D_i(\pi(\underline{Y})) = *)$.

Proof. From the discussion in Section 2.2, $D_i(\underline{y}) = *$ if and only if the erasure pattern in $\underline{y}$ covers some codeword with a value of 1 at bit position i. Since $\pi(i) = j$, the erasure pattern in $\underline{y}$ covers such a codeword if and only if the erasure pattern in $\pi(\underline{y})$ covers a codeword with a value of 1 at bit position j. This implies that $D_i(\underline{y}) = *$ if and only if $D_j(\pi(\underline{y})) = *$.

Now, suppose $D_i(\underline{y}) = 0$. This happens only if all the codewords $\underline{x}$ that can result in $\underline{y}$ have a 0 at position i. For such an $\underline{x}$, $\pi(\underline{x})$ can result in $\pi(\underline{y})$, and the codeword $\pi(\underline{x})$ has a 0 at position j. From the discussion above, $D_j(\pi(\underline{y})) \neq *$. This implies that $D_j(\pi(\underline{y})) = 0$.

Similarly, $D_j(\pi(\underline{y})) = 0$ implies that $D_i(\underline{y}) = 0$. Thus, $D_i(\underline{y}) = 0$ if and only if $D_j(\pi(\underline{y})) = 0$. This immediately shows that $D_i(\underline{y}) = D_j(\pi(\underline{y}))$.

The second part of the proposition follows from the fact that $\underline{Y}$ and $\pi(\underline{Y})$ have the same distribution for any $\pi \in \mathcal{G}$. □

Definition 7 (Transitivity)**:** Suppose $\mathcal{G}$ is a permutation group. Then,

a) $\mathcal{G}$ is *transitive* if for any $i, j \in [N]$, there exists $\pi \in \mathcal{G}$ such that $\pi(i) = j$;

b) $\mathcal{G}$ is *doubly transitive* if for distinct $i, j, k \in [N]$, there exists $\pi \in \mathcal{G}$ such that $\pi(i) = i$, $\pi(j) = k$.

Below, we study the consequences of the transitive/doubly transitive permutation groups on the decoding process.

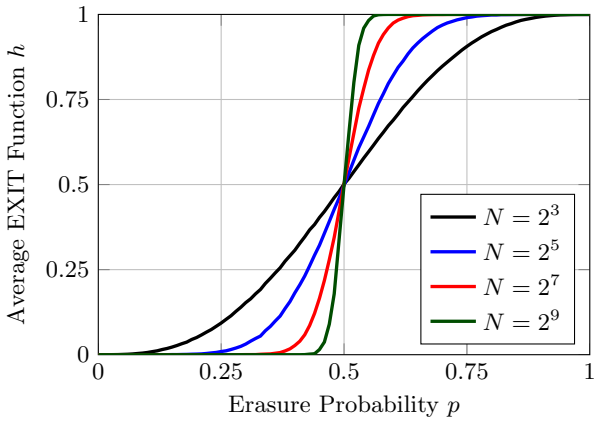

Figure 1: Average EXIT functions, $h^{(n)}(p)$, of rate-1/2 Reed-Muller codes of length-N.

Proposition 8: If the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is transitive, then, for any $0 \leq p \leq 1$ and any $i, j \in [N]$,

$$P_{b,i}(p) = P_{b,j}(p) = P_b(p), \qquad h_i(p) = h_j(p) = h(p).$$

Proof. Since $\mathcal{G}$ is transitive, there exists $\pi \in \mathcal{G}$ such that $\pi(i) = j$. Therefore, by Proposition 6,

$$\begin{aligned} P_{b,i} &= \Pr(D_i(\underline{Y}) = *) \\ &= \Pr(D_j(\pi(\underline{Y})) = *) \\ &= \Pr(D_j(\underline{Y}) = *) \\ &= P_{b,j}. \end{aligned}$$

Thus, for any $i, j \in [N]$, $P_{b,i}(p) = P_{b,j}(p)$ and, consequently, $P_{b,i}(p) = P_b(p)$. Together with (2), we get $h_i(p) = h_j(p) = h(p)$. □

Proposition 9 (Symmetry of Ω_i)**:** If the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is doubly transitive, then the set Ω_i is *symmetric*, i.e., it is invariant under a transitive permutation group.

Proof. Since $\mathcal{G}$ is doubly transitive, for distinct $i, j, k \in [N]$, there exists a permutation $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$. Since $\pi(i) = i$, the restriction of π to $[N]\backslash\{i\}$ acts as a permutation and transposes j to k. Such restrictions induce transitive permutations on the set of codewords whose value is 1 at bit position i. From the definition of Ω_i in (3), it is easy to see that these restrictions leave Ω_i invariant. □

2.5 Capacity-Achieving Codes

Definition 10 (Capacity-Achieving Codes)**:** Suppose $\{\mathcal{C}_n\}$ is a sequence of codes with rates $\{r_n\}$, where $r_n \to r$ for $r \in (0, 1)$.

a) $\{\mathcal{C}_n\}$ is said to be capacity achieving on the BEC under bit-MAP decoding, if for any $p \in [0, 1 - r)$, the average bit erasure probabilities satisfy $\lim_{n\to\infty} P_b^{(n)}(p) = 0$.

b) $\{\mathcal{C}_n\}$ is said to be capacity achieving on the BEC under block-MAP decoding, if for any $p \in [0, 1-r)$, the average block erasure probabilities satisfy $\lim_{n\to\infty} P_B^{(n)}(p) = 0$.

The following proposition encapsulates the approach we use to show that a sequence of codes achieves capacity. It connects capacity-achieving codes, average EXIT functions, and the sharp transitions associated with boolean functions. For any $\varepsilon \in (0, 1/2]$, the *transition width* of the EXIT function $h^{(n)}(p)$ is the interval over which it transitions from ε to $1-\varepsilon$. Mathematically, this is given by $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}$, where $p_t^{(n)}$ is the functional inverse of $h^{(n)}$ given by

$$p_t^{(n)} \triangleq \inf\{p \in [0, 1] \,|\, h^{(n)}(p) \geq t\}. \tag{4}$$

The average EXIT functions of some rate-1/2 Reed-Muller codes are shown in Figure 1. Observe that as the block length increases, the transition width of the average EXIT function decreases. According to the following proposition, if this width converges to 0, the Reed-Muller code sequence achieves capacity on the BEC under bit-MAP decoding.

Proposition 11: Let $\{\mathcal{C}_n\}$ be a sequence of codes with rates $\{r_n\}$, where $r_n \to r$ for some $r \in (0, 1)$. Then, the following statements are equivalent.

S1: $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.

S2: The sequence of average EXIT functions satisfies

$$\lim_{n\to\infty} h^{(n)}(p) = \begin{cases} 0 & \text{if } 0 \leq p < 1 - r, \\ 1 & \text{if } 1 - r < p \leq 1. \end{cases}$$

S3: For any $\varepsilon \in (0, 1/2]$, the sequence of transition widths satisfies $\lim_{n\to\infty} \left(p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}\right) = 0$.

Proof. See Appendix A.1. □

The equivalence between the first two statements is due to the close relationship between the bit erasure probability and the EXIT function in (2), while the equivalence between the last two statements is a consequence of the area theorem in Proposition 1.

While the above result appears deceptively simple, our approach is successful largely because the transition point of the limiting EXIT function is known a priori due to the area theorem. Even though the sharp transition framework presented in the next section is widely applicable in theoretical computer science and allows one to deduce that the transition width of certain functions goes to 0, establishing the existence of a threshold and determining its precise location can be notoriously difficult[3] [3, 18, 23].

3. REED-MULLER CODES ACHIEVE CAPACITY

3.1 General Results

Consider a measure μ_p on $\{0, 1\}^M$ such that

$$\mu_p(\Omega) = \sum_{\underline{a} \in \Omega} p^{|\underline{a}|}(1-p)^{M-|\underline{a}|}, \quad \text{for } \Omega \subseteq \{0, 1\}^M.$$

One important result in the theory of boolean functions is that symmetric monotone sets always exhibit a sharp transition [30], i.e., $\mu_p(\Omega)$ transitions quickly from 0 to 1 as a

[3]Existence of a threshold means for some $0 < a < 1$, $p_\varepsilon^{(n)} \to a$ for all $\varepsilon > 0$. Note that this implies that the transition width $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)} \to 0$ and not vice versa.

function of p. Specifically, for a symmetric monotone set Ω, the results in [11, 30, 71] imply that, for $0 < p < 1$,

$$\frac{\mathrm{d}\mu_p(\Omega)}{\mathrm{d}p} \geq C(\log M)\mu_p(\Omega)(1-\mu_p(\Omega)),$$

where the constant $C > 0$ is independent of p, Ω, M.

For a monotone set $\Omega \subseteq \{0,1\}^M$, let $g(p) = \log \frac{\mu_p(\Omega)}{1-\mu_p(\Omega)}$ and observe that

$$\frac{\mathrm{d}g(p)}{\mathrm{d}p} = \frac{1}{\mu_p(\Omega)(1-\mu_p(\Omega))}\frac{\mathrm{d}\mu_p(\Omega)}{\mathrm{d}p} \geq C\log M. \quad (5)$$

Define

$$p_t \triangleq \inf\{p \in [0,1] \,|\, \mu_p(\Omega) \geq t\},$$

and note that $\mu_{p_t}(\Omega) = t$ and $g(p_t) = \log \frac{t}{1-t}$. For $\varepsilon \in (0, 1/2]$, by integrating (5) from p_ε to $p_{1-\varepsilon}$, we obtain

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{2}{C}\frac{\log\frac{1-\varepsilon}{\varepsilon}}{\log M}. \quad (6)$$

At this point, we have all the ingredients to prove the main theorem of the paper.

Theorem 12 (Capacity-Achieving Codes under Bit-MAP Decoding)**:** Let $\{\mathcal{C}_n\}$ be a sequence of codes where the block lengths satisfy $N_n \to \infty$, the rates satisfy $r_n \to r$, for $r \in (0,1)$, and the permutation group $\mathcal{G}^{(n)}$ of the code $\mathcal{C}_n$ is doubly transitive for each n. Then, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.

Proof. Let $h^{(n)}$ be the average EXIT function of $\mathcal{C}_n$ and fix $i \in [N]$. The quantities N, $\mathcal{G}$, h, h_i, Ω_i, and p_t that appear in this proof should be indexed by n, but we drop the index to avoid cluttering.

The set Ω_i is monotone by Remark 4 and, since $\mathcal{G}$ is doubly transitive, it is symmetric from Proposition 9.

Consider a measure μ_p on $\{0,1\}^{N-1}$ such that

$$\mu_p(\Omega) = \sum_{\underline{a}\in\Omega} p^{|\underline{a}|}(1-p)^{N-1-|\underline{a}|}, \quad \text{for } \Omega \subseteq \{0,1\}^{N-1}.$$

Then, $h_i(p) = \mu_p(\Omega_i)$. Also, since $\mathcal{G}$ is transitive, from Proposition 8, $h(p) = h_i(p)$, for all $p \in [0,1]$. Therefore, from (6), we have

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{2}{C}\frac{\log\frac{1-\varepsilon}{\varepsilon}}{\log(N-1)}, \quad (7)$$

where p_t is the inverse of h as in (4). Since $N \to \infty$ from the hypothesis, $p_{1-\varepsilon} - p_\varepsilon \to 0$. As a result, from Proposition 11, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding. □

Let us now consider the block erasure probability P_B. Recall that $P_B \leq NP_b$ and, therefore, if $P_b \to 0$ with sufficient speed, then $P_B \to 0$ as well. The following theorem provides sufficient conditions for a sequence of codes to achieve capacity under block-MAP decoding.

Theorem 13 (Capacity-Achieving Codes under Block-MAP Decoding)**:** Let $\{\mathcal{C}_n\}$ be a sequence of codes where the block lengths satisfy $N_n \to \infty$ and the rates satisfy $r_n \to r$ for $r \in (0,1)$. Let $h^{(n)}(p)$ be the average EXIT function of $\mathcal{C}_n$ and suppose that it satisfies, for $a_n < p < b_n$,

$$\frac{\mathrm{d}h^{(n)}(p)}{\mathrm{d}p} \geq w_n \log(N_n) h^{(n)}(p)(1-h^{(n)}(p)), \quad (8)$$

where $w_n \to \infty$, $a_n \to 0$, $b_n \to 1$, and $0 \leq a_n < b_n \leq 1$. Then, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

Proof. See Appendix A.2. □

3.2 Reed-Muller Codes

Recall that, for integers v, n satisfying $0 \leq v \leq n$, the Reed-Muller code $\mathrm{RM}(v,n)$ is a binary linear code with block length $N = 2^n$ and rate $r = 2^{-n}\left(\binom{n}{0} + \ldots + \binom{n}{v}\right)$.

Consider the set of n variables, $x_1, \ldots, x_n$. For a monomial $x_1^{i_1}\cdots x_n^{i_n}$ in these variables, define its degree to be $i_1 + \cdots + i_n$. A polynomial in n variables is a linear combination of such monomials using coefficients from a field, and the degree of such a polynomial is defined to be the maximum degree of any monomial it contains. It is well-known that the set of all n-variable polynomials of degree at most v is a vector space over its field of coefficients. In this section, the coefficient field is the Galois field $\mathbb{F}_2$ and the vector space of interest is given by

$$P(n,v) = \mathrm{span}\Big\{x_1^{t_1}\ldots x_n^{t_n} \mid t_1 + \cdots + t_n \leq v, t_i \in \{0,1\}, \text{for } i \in [n]\Big\}.$$

For a polynomial $f \in P(n,v)$, $f(\underline{x}) \in \{0,1\}$ denotes the evaluation of f at $\underline{x} \in \{0,1\}^n$. Let the elements of the vector space $\{0,1\}^n$ over $\mathbb{F}_2$ be enumerated by $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_N$ with $\underline{e}_N = \underline{0}$ as the all-zero vector. Then, the code $\mathrm{RM}(v,n)$ is defined as

$$\mathrm{RM}(v,n) \triangleq \{(f(\underline{e}_1), \ldots, f(\underline{e}_N)) \mid f \in P(n,v)\}.$$

The following is a classical result for Reed-Muller codes [38, Corollary 4].

Lemma 14: The permutation group $\mathcal{G}$ of $\mathrm{RM}(v,n)$ is doubly transitive.

Proof. See Appendix A.3. □

For $r \in (0,1)$, consider the code $\mathrm{RM}(v_n(r), n)$, with

$$v_n(r) \triangleq \max\left\{\left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}Q^{-1}(1-r)\right\rfloor, 0\right\},$$

where $Q(t) \triangleq \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\tau^2/2}\mathrm{d}\tau$. Then, the sequence of codes $\{\mathrm{RM}(v_n(r),n)\}$ has rates $r_n \to r$ and increasing block lengths. The sequence of average EXIT functions of Reed-Muller codes exhibits a sharp transition from 0 to 1 as n grows (see Figure 1), and the theorem below follows from Lemma 14 and Theorem 12.

Theorem 15: For any $r \in (0,1)$, the sequence of codes $\{\mathrm{RM}(v_n(r),n)\}$ has rates $r_n \to r$ and is capacity achieving on the BEC under bit-MAP decoding.

Now, let's consider the block erasure probability of Reed-Muller codes. It turns out that the factor $\log(N-1)$ in Theorem 15 is not sufficient for $NP_b \to 0$. Fortunately, it is possible to exploit symmetries beyond the double transitivity of the permutation group, to obtain factors that grow asymptotically faster than $\log(N-1)$ [12]. First, we require the following result.

Lemma 16: For a Reed-Muller code of block length N, consider the set Ω_N defined in (3) for bit N and let its permutation group be

$$\mathcal{G}_N \triangleq \{\pi \in S_{N-1} \mid \pi(\underline{a}) \in \Omega_N \text{ for all } \underline{a} \in \Omega_N\}.$$

Then, $\mathcal{G}_N$ has a transitive subgroup isomorphic to $\mathrm{GL}(n, \mathbb{F}_2)$, the general linear group of degree n over the Galois field $\mathbb{F}_2$.

Proof. See Appendix A.4. □

Theorem 17: For any $r \in (0,1)$, the sequence of codes $\{\mathrm{RM}(v_n(r), n)\}$ has rates $r_n \to r$ and is capacity achieving on the BEC under block-MAP decoding.

Proof. Let the EXIT function associated with the last bit and the average EXIT function of the code $\mathrm{RM}(v_n(r), n)$ be h_N and h, respectively. The permutation group of this code is transitive by Lemma 14. Hence, from Proposition 8, we have $h = h_N$. Moreover, by Lemma 16, $\mathcal{G}_N$ contains a transitive subgroup isomorphic to $\mathrm{GL}(n, \mathbb{F}_2)$.

Now, we can exploit the $\mathrm{GL}(n, \mathbb{F}_2)$ symmetry of Ω_N within the framework of [12]. In particular, [12, Theorem 1, Corollary 4.1] implies that there exists a universal constant $C > 0$, independent of n and p, such that, for $a_n < p < b_n$,

$$\frac{\mathrm{d}h_N(p)}{\mathrm{d}p} \geq C \log(\log(N_n)) \log(N_n) h_N(p)(1 - h_N(p)),$$

where $N_n = 2^n$, $0 < a_n < b_n < 1$, and $a_n \to 0$, $b_n \to 1$ as $n \to \infty$. Since $h = h_N$, Theorem 13 implies that $\{\mathrm{RM}(v_n(r), n)\}$ is capacity achieving on the BEC under block-MAP decoding. □

Remark 18 (From Bit-MAP to Block-MAP via Weight Distribution)**:** The proof presented above of Theorem 17 is based on the framework of [12], which yields an extra factor of $\log(\log(N))$ in the expression of the derivative of the average EXIT function. However, it is also possible to prove that the block erasure probability goes to 0, for all $0 \leq p < 1 - r$, by combining the analysis in Theorem 12 with a careful upper bound on the weight distribution of Reed-Muller codes (see [44] for details).

4. DISCUSSION

In this paper, we show that a sequence of binary linear codes achieves capacity if its block lengths are strictly increasing, its code rates converge to some $r \in (0,1)$, and the permutation group of each code is doubly transitive. As a consequence, we prove that Reed-Muller codes achieve capacity on the BEC both under bit-MAP and block-MAP decoding, thus settling a long standing conjecture. To achieve this, we exploit the symmetry of the codes to prove the existence of 0-1 transitions for EXIT functions and apply the area theorem to locate the thresholds. One remarkable aspect of this method is its simplicity. In particular, it does not rely on the precise structure of the codes.

Our results lead to a number of open questions. The most natural question is whether or not one can extend this approach to BMS channels via generalized EXIT (GEXIT) functions [55]. For this, some new ideas are required because the straightforward approach leads to the analysis of functions that are neither boolean nor monotonic. It would also be interesting to find boolean functions outside of coding theory where area theorems can be used to pinpoint thresholds.

We can also analyze the case where the rates $r_n \to 0$. In Section 4.4, we show that a sequence of Reed-Muller codes (and, in general, any sequence of codes whose permutation groups are doubly transitive) with rates satisfying $r_n \log(N_n) \to \infty$ is capacity achieving under bit-MAP decoding according to [2, Definition 16]. However, our method does not apply to cases where the rates satisfy $r_n = N_n^{-t}$ for some $t \in (0,1)$. Hence, it would be interesting to extend the proof of Theorem 15 to show that Reed-Muller codes are capacity achieving in this regime. We note that part of this range is already covered by the results in [2].

Finally, the conditions required by Theorem 12 can certainly be relaxed. For example, puncturing a finite number of code bits destroys the double transitivity of the code but does not affect the conclusion of the theorem. Thus, an interesting open problem is "To what extent can the conditions of Theorem 12 be weakened?".

The main results have several natural extensions. Below, we discuss affine invariant codes, BCH codes, $\mathbb{F}_q$-linear codes over q-ary erasure channels, the analysis of the case where the rates of the codes asymptotically go to 0, and finally some algorithmic consequences.

4.1 Affine-Invariant Codes

Consider a code $\mathcal{C}$ of length $N = 2^n$ and the Galois field $\mathbb{F}_N$. Let $\Theta\colon [N] \to \mathbb{F}_N$ denote a bijection between the codeword positions and the elements of the field. Take a pair $\beta, \gamma \in \mathbb{F}_N$ with $\beta \neq 0$ and define $\pi_{\beta,\gamma} \in S_N$ such that

$$\pi_{\beta,\gamma}(\ell) \triangleq \Theta^{-1}(\beta\Theta(\ell) + \gamma).$$

Note that $\pi_{\beta,\gamma}$ is well-defined since Θ is bijective and $\beta \neq 0$, and observe that $\pi_{\beta_1,\gamma_1} \circ \pi_{\beta_2,\gamma_2} = \pi_{\beta_1\beta_2,\beta_1\gamma_2+\gamma_1}$. As such, the collection of permutations $\pi_{\beta,\gamma}$ forms a group. Now, the code $\mathcal{C}$ is called *affine-invariant* if its permutation group contains the subgroup

$$\{\pi_{\beta,\gamma} \in S_N \mid \beta, \gamma \in \mathbb{F}_N, \beta \neq 0\},$$

for some bijection Θ [36, Section 4.7].

Affine-invariant codes are of interest to us because their permutation groups are doubly transitive. To see this, consider distinct $i, j, k \in [N]$ and choose $\beta, \gamma \in \mathbb{F}_N$ as

$$\beta = \frac{\Theta(i) - \Theta(k)}{\Theta(i) - \Theta(j)}, \qquad \gamma = \Theta(i)\left(\frac{\Theta(k) - \Theta(j)}{\Theta(i) - \Theta(j)}\right).$$

Then, $\pi_{\beta,\gamma}(i) = i$ and $\pi_{\beta,\gamma}(j) = k$. Thus, by Theorem 12, a sequence of affine-invariant codes of increasing block lengths and rates converging to $r \in (0,1)$, achieves capacity on the BEC under bit-MAP decoding. Another class of affine-invariant codes, besides Reed-Muller codes, is the family of extended primitive narrow-sense BCH codes [36, Theorem 5.1.9]. The next subsection contains a detailed discussion of BCH codes along with a proof that they are capacity achieving on the BEC both under bit-MAP and block-MAP decoding.

4.2 BCH Codes

Let α be a primitive element of $\mathbb{F}_{2^n}$. Recall that a binary BCH code is *primitive* if its block length is of the form $2^n - 1$, and *narrow-sense* if the roots of its generator polynomial include consecutive powers of a primitive element starting from α. In this work, we consider only primitive narrow-sense BCH codes and we follow closely the treatment of BCH codes in [36].

For integers v, n with $1 \leq v \leq 2^n - 1$, let $f(n, v)$ be the polynomial of lowest-degree over $\mathbb{F}_2$ that has the roots $\alpha, \alpha^2, \ldots, \alpha^v$. Then, BCH$(v, n)$ is defined to be the binary cyclic code with the generator polynomial $f(n, v)$ and block length $N = 2^n - 1$. This is precisely the primitive narrow-sense BCH code with block length N and designed distance $v + 1$. The dimension K of the code is determined by the degree of the generator polynomial as $K = N - \text{degree}(f(n, v))$ [36, Theorem 4.2.1], and the minimum distance $d_{\min}$ is at least $v + 1$ [36, Theorem 5.1.1]. Furthermore, for any $r \in (0, 1)$, one can choose $v_n \in [N]$ such that

$$N(1 - r) \leq \text{degree}(f(n, v_n)) \leq N(1 - r) + n.$$

Thus, the rates of the sequence of codes $\{\text{BCH}(v_n, n)\}$ converge to r.

Consider the extended BCH code of length 2^n, namely eBCH(v, n), which is formed by adding a single parity bit to the code BCH(v, n) so that the overall codeword parity is always even [36, Section 5.1]. The code eBCH(v, n) has the same dimension as BCH(v, n) and a minimum distance of at least $v + 1$. Thus, for any $r \in (0, 1)$, there exists a sequence of codes $\{\text{eBCH}(v_n, n)\}$ with block lengths $N_n = 2^n$, rates $r_n \to r$ and minimum distances

$$d_{\min}^{(n)} \geq 1 + v_n \geq 1 + \frac{N_n(1 - r)}{n}. \tag{9}$$

The crucial property of the extended BCH codes is that they are affine-invariant [36, Theorem 5.1.9]. Thus, by the argument of Section 4.1, the code sequence $\{\text{eBCH}(v_n, n)\}$ is capacity achieving on the BEC under bit-MAP decoding.

Recall from (1) that

$$P_B \leq \frac{N}{d_{\min}} P_b.$$

Choosing $\varepsilon_n = d_{\min}^{(n)}/(N_n \log(N_n))$ in (6) shows that $p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} \to 0$. Since the code sequence $\{\text{eBCH}(v_n, n)\}$ achieves capacity on the BEC under bit-MAP decoding, from statement S3 of Proposition 11, for any $\varepsilon \in (0, 1/2]$, $p_\varepsilon^{(n)} \to 1 - r$. Combined with the fact that $p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} \to 0$ and $p_{1-\varepsilon_n}^{(n)} \geq p_\varepsilon^{(n)} \geq p_{\varepsilon_n}^{(n)}$ shows that $p_{\varepsilon_n}^{(n)} \to 1 - r$. Using the improved upper bound on P_B implied by minimum distance in (1), we also find that

$$P_B^{(n)}\left(p_{\varepsilon_n}^{(n)}\right) \leq \frac{N_n}{d_{\min}^{(n)}} P_b^{(n)}\left(p_{\varepsilon_n}^{(n)}\right) \leq \frac{N_n}{d_{\min}^{(n)}} \varepsilon_n = \frac{1}{\log(N_n)} \to 0.$$

Thus, the code sequence $\{\text{eBCH}(v_n, n)\}$ is capacity achieving on the BEC under block-MAP decoding.

By definition, the code BCH(v, n) can be constructed from the code eBCH(v, n) by erasing the overall parity bit. Consequently, the sequence $\{\text{BCH}(v_n, n)\}$ of binary cyclic codes satisfies $r_n \to r$ and can be shown to achieve capacity on the BEC under both bit-MAP and block-MAP decoding. As far as we know, this provides the first proof that sequences of binary cyclic codes can achieve capacity under MAP decoding [59].

4.3 $\mathbb{F}_q$-Linear Codes

While our exposition focuses on binary linear codes over the BEC, all results extend easily to $\mathbb{F}_q$-linear codes over the q-ary erasure channel.

In this more general framework, the set Ω_i is redefined to be the set of erasure patterns that prevent the recovery of the symbol X_i given the output $\underline{Y}_{\sim i}$. Note that Ω_i is still a set of binary sequences and *not* a set of sequences over the alphabet $\{0, 1, \ldots, q - 1\}$. When the recovery of X_i is not possible, the linearity of the code implies that the posterior marginal of symbol i given the observations $\underline{Y}_{\sim i} = \underline{y}_{\sim i}$ is $\Pr(X_i = x | \underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1/q$. In this case, in order to have $H(X_i | \underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1$, we rescale the logarithm in the entropy function to base q.

As a result, the sharp threshold framework for symmetric monotone boolean functions can be applied without change. With these straightforward modifications, the results in Section 3 hold true verbatim.

The concept of affine-invariance also extends naturally to $\mathbb{F}_q$-linear codes of length q^n over the Galois field $\mathbb{F}_q$. Similarly, affine-invariance implies that the permutation group is doubly transitive. Thus, sequences of affine-invariant $\mathbb{F}_q$-linear codes of increasing length, whose rates converge to $r \in (0, 1)$, achieve capacity on the q-ary erasure channel under symbol-MAP decoding. The results for the block-MAP decoder also extend without change. Thus, one finds that generalized Reed-Muller codes [20] and extended primitive narrow-sense BCH codes over $\mathbb{F}_q$ achieve capacity on the q-ary erasure channel under block-MAP decoding.

4.4 Rates Converging to Zero

Consider a sequence of Reed-Muller codes $\{\text{RM}(v_n, n)\}$ where the rates $r_n \to 0$ sufficiently fast. A key result of [2] is that Reed-Muller codes are capacity-achieving in this scenario. That is, for any $\delta > 0$,

$$P_B^{(n)}(p_n) \to 0, \quad \text{for any } 0 \leq p_n < 1 - (1 + \delta) r_n.$$

Looking closely at [2, Corollary 44], we see that the rates must decay as $r_n = O(N_n^{-\kappa})$ for some $\kappa > 0$.

Let's analyze the bit erasure probability using our method. From the proof of Theorem 12, it is possible to deduce that $P_b^{(n)}(p_{\varepsilon_n}) \to 0$ if we choose $\varepsilon_n = o(1)$ such that $\log(1/\varepsilon_n) = o(\log(N_n))$. In addition, the following lower bound on p_{ε_n} holds

$$\begin{aligned} p_{\varepsilon_n} &\overset{(a)}{\geq} 1 - \frac{r_n}{1 - \varepsilon_n} - (p_{1-\varepsilon_n} - p_{\varepsilon_n}) \\ &\overset{(b)}{\geq} 1 - \frac{r_n}{1 - \varepsilon_n} - \frac{2}{C} \frac{\log \frac{1}{\varepsilon_n}}{\log(N_n - 1)} \overset{(c)}{=} 1 - (1 + \delta_n) r_n, \end{aligned}$$

where (a) comes from (10) in Appendix A.1, (b) comes from (7) in the proof of Theorem 12, and in (c) we set

$$\delta_n = \frac{\varepsilon_n}{1 - \varepsilon_n} + \frac{2}{C} \frac{\log \frac{1}{\varepsilon_n}}{r_n \log(N_n - 1)}.$$

Therefore, we have

$$P_b^{(n)}(p_n) \to 0, \quad \text{for any } 0 \leq p_n < 1 - (1 + \delta_n) r_n.$$

In order to obtain a capacity-achieving result under bit-MAP decoding, we require that $\delta_n \to 0$. This can be guaranteed by taking $r_n \log(N_n) \to \infty$. Indeed, under this condition, we can choose $\varepsilon_n = 1/\log(r_n \log(N_n))$ so that

$$\varepsilon_n \to 0, \qquad \frac{\log \frac{1}{\varepsilon_n}}{r_n \log(N_n - 1)} \to 0, \qquad \delta_n \to 0.$$

In conclusion, in order to have a capacity-achieving sequence of Reed-Muller codes with rates $r_n \to 0$, our method requires $r_n \log(N_n) \to \infty$ while the results in [2, Corollary 44] require $r_n = O(N_n^{-\kappa})$ for some $\kappa > 0$. Thus, the results in the two papers apply to two asymptotic rate regimes that are non-overlapping.

4.5 Algorithmic Consequences

One attraction of Reed-Muller codes is the availability of efficient decoding algorithms [25, 26, 28, 63, 66], at least in the low-rate regime.

Our results for the erasure channel also have implications for the decoding of Reed-Muller codes over the binary symmetric channel. In particular, [2, Theorem 8] shows that an error pattern can be corrected by $\mathrm{RM}(n-(2t+2),n)$ under block-MAP decoding whenever an erasure pattern with the same support can be corrected by $\mathrm{RM}(n-(t+1),n)$ under block-MAP decoding. Using the algorithm in [63], these error patterns can even be corrected efficiently. Combined with our results for the BEC, [63, Corollary 14] shows that there exists a deterministic algorithm that runs in time at most n^4 and is able to correct $(1/2-o(1))2^n$ random errors in $\mathrm{RM}(n, o(\sqrt{n}))$ with probability $1-o(1)$.

ACKNOWLEDGMENTS

The authors' interest in this problem was piqued by its listing as an open problem during the 2015 Simons Institute program on Information Theory. We gratefully acknowledge discussions with Hamed Hassani and Tom Richardson.

The work of S. Kumar and H. D. Pfister was supported in part by the National Science Foundation (NSF) under Grant No. 1218398. The work of M. Mondelli and R. Urbanke was supported by grant No. 200020_146832/1 of the Swiss National Science Foundation. Any opinions, findings, recommendations, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of these sponsors.

5. REFERENCES

[1] E. Abbe, A. Shpilka, and A. Wigderson. Reed-Muller codes for random erasures and errors. In *Proc. of the Annual ACM Symp. on Theory of Comp.*, STOC '15, pages 297–306, New York, NY, USA, 2015. ACM.

[2] E. Abbe, A. Shpilka, and A. Wigderson. Reed-Muller codes for random erasures and errors. *IEEE Trans. Inform. Theory*, 61(10):5229–5252, Oct 2015.

[3] D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, 2005.

[4] R. Ahlswede and G. Dueck. Good codes can be produced by a few permutations. *IEEE Trans. Inform. Theory*, 28(3):430–443, May 1982.

[5] E. Arıkan. A performance comparison of polar codes and Reed-Muller codes. *IEEE Commun. Letters*, 12(6):447–449, June 2008.

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

[7] E. Arıkan. A survey of Reed-Muller codes from polar coding perspective. In *Proc. IEEE Inform. Theory Workshop*, pages 1–5, Jan 2010.

[8] A. Ashikhmin, G. Kramer, and S. ten Brink. Extrinsic information transfer functions: model and erasure channel properties. *IEEE Trans. Inform. Theory*, 50(11):2657–2674, Nov. 2004.

[9] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *Proc. IEEE Int. Conf. Commun.*, volume 2, pages 1064–1070, Geneva, Switzerland, May 1993. IEEE.

[10] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[11] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1-2):55–64, 1992.

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

[13] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In *Advances in Cryptology–CRYPTO'91*, pages 86–100. Springer, 1992.

[14] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine. On cryptographic properties of the cosets of R(1, m). *IEEE Trans. Inform. Theory*, 47(4):1494–1513, 2001.

[15] C. Carlet, D. K. Dalai, K. C. Gupta, and S. Maitra. Algebraic immunity for cryptographically significant boolean functions: analysis and construction. *IEEE Trans. Inform. Theory*, 52(7):3105–3121, 2006.

[16] C. Carlet and P. Gaborit. On the construction of balanced boolean functions with a good algebraic immunity. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1101–1105, Sept 2005.

[17] J. Coffey and R. Goodman. Any code of which we cannot think is good. *IEEE Trans. Inform. Theory*, 36(6):1453–1461, Nov 1990.

[18] A. Coja-Oghlan. The asymptotic k-SAT threshold. In *Proc. of the Annual ACM Symp. on Theory of Comp.*, STOC '14, pages 804–813, New York, NY, USA, 2014. ACM.

[19] D. J. Costello, Jr. and G. D. Forney, Jr. Channel coding: The road to channel capacity. *Proc. of the IEEE*, 95(6):1150–1177, June 2007.

[20] P. Delsarte, J. Goethals, and F. M. Williams. On generalized Reed-Muller codes and their relatives. *Inform. and Control*, 16(5):403–442, 1970.

[21] F. Didier. A new upper bound on the block error probability after decoding over the erasure channel. *IEEE Trans. Inform. Theory*, 52(10):4496–4503, Oct 2006.

[22] F. Didier and J.-P. Tillich. Computing the algebraic immunity efficiently. In *Fast Software Encryption*, pages 359–374. Springer, 2006.

[23] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. In *Proc. of the Annual ACM Symp. on Theory of Comp.*, STOC '15, pages 59–68, New York, NY, USA, 2015. ACM.

[24] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Ann. of Math.*, pages 439–485, 2005.

[25] I. Dumer. Recursive decoding and its performance for

low-rate Reed-Muller codes. *IEEE Trans. Inform. Theory*, 50(5):811–823, May 2004.

[26] I. Dumer. Soft-decision decoding of Reed-Muller codes: a simplified algorithm. *IEEE Trans. Inform. Theory*, 52(3):954–963, March 2006.

[27] I. Dumer and P. G. Farrell. Erasure correction performance of linear block codes. In *Algebraic Coding*, pages 316–326. Springer, 1994.

[28] I. Dumer and K. Shabunov. Soft-decision decoding of Reed-Muller codes: recursive lists. *IEEE Trans. Inform. Theory*, 52(3):1260–1266, March 2006.

[29] E. Friedgut and J. Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999.

[30] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996.

[31] R. G. Gallager. *Low-Density Parity-Check Codes*. The M.I.T. Press, Cambridge, MA, USA, 1963.

[32] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proc. of the Annual ACM Symp. on Theory of Comp.*, STOC '91, pages 33–42, New York, NY, USA, 1991. ACM.

[33] P. Gemmell and M. Sudan. Highly resilient correctors for polynomials. *Information processing letters*, 43(4):169–174, 1992.

[34] B. Gérard and J.-P. Tillich. Using tools from error correcting theory in linear cryptanalysis. *Adv. Linear Cryptanalysis of Block and Stream Ciphers*, 7:87, 2011.

[35] E. Grigorescu, T. Kaufman, and M. Sudan. 2-transitivity is insufficient for local testability. In *Annual IEEE Conf. on Comp. Complex.*, pages 259–267, June 2008.

[36] W. C. Huffman and V. Pless. *Fundamentals of error-correcting codes*. Cambridge University Press, 2003.

[37] G. Kalai and S. Safra. Threshold phenomena and influence with some perspectives from mathematics, computer science, and economics. *Comp. Complexity and Stat. Phy., Santa Fe Institute Studies in Sci. of Complexity*, 19517738, 2005.

[38] T. Kasami, S. Lin, and W. W. Peterson. New generalizations of the Reed-Muller codes–I: Primitive codes. *IEEE Trans. Inform. Theory*, 14(2):189–199, Mar 1968.

[39] T. Kasami and N. Tokura. On the weight structure of Reed-Muller codes. *IEEE Trans. Inform. Theory*, 16(6):752–759, Nov 1970.

[40] T. Kasami, N. Tokura, and S. Azumi. On the weight enumeration of weights less than 2.5d of Reed-Muller codes. *Inform. and Control*, 30(4):380 – 395, 1976.

[41] T. Kaufman, S. Lovett, and E. Porat. Weight distribution and list-decoding size of Reed-Muller codes. *IEEE Trans. Inform. Theory*, 58(5):2689–2696, May 2012.

[42] T. Kaufman and M. Viderman. Locally testable vs. locally decodable codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 670–682. Springer, 2010.

[43] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. L. Urbanke. Reed-Muller codes achieve capacity on erasure channels. Submitted to *IEEE Trans. Inform. Theory*, 2016. [Online]. Available: http://arxiv.org/pdf/1601.04689.pdf.

[44] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. L. Urbanke. Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 1755–1759, Barcelona, Spain, 2016.

[45] S. Kudekar, T. Richardson, and R. L. Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Trans. Inform. Theory*, 59(12):7761–7813, Dec. 2013.

[46] S. Kudekar, T. J. Richardson, and R. L. Urbanke. Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC. *IEEE Trans. Inform. Theory*, 57(2):803–834, Feb. 2011.

[47] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister. Threshold saturation for spatially-coupled LDPC and LDGM codes on BMS channels. *IEEE Trans. Inform. Theory*, 60(12):7389–7415, Dec. 2014.

[48] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov. Iterative decoding threshold analysis for LDPC convolutional codes. *IEEE Trans. Inform. Theory*, 56(10):5274–5289, Oct. 2010.

[49] S. Lin. RM codes are not so bad. In *Proc. IEEE Inform. Theory Workshop*, June 1993. Invited talk.

[50] S. Lin and D. J. Costello, Jr. *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 2004. ISBN-13: 978-0130426727.

[51] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. Efficient erasure correcting codes. *IEEE Trans. Inform. Theory*, 47(2):569–584, Feb. 2001.

[52] D. J. C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Trans. Inform. Theory*, 45(2):399–431, March 1999.

[53] F. J. MacWilliams and N. J. A. Sloane. *The theory of error correcting codes*, volume 16. Elsevier, 1977.

[54] G. A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problems of Inform. Transm.*, 10(2):101–108, 1974.

[55] C. Méasson, A. Montanari, T. J. Richardson, and R. Urbanke. The generalized area theorem and some of its consequences. *IEEE Trans. Inform. Theory*, 55(11):4793–4821, Nov. 2009.

[56] C. Méasson, A. Montanari, and R. L. Urbanke. Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding. *IEEE Trans. Inform. Theory*, 54(12):5277–5307, Dec. 2008.

[57] M. Mondelli, S. H. Hassani, and R. L. Urbanke. From polar to Reed-Muller codes: A technique to improve the finite-length performance. *IEEE Trans. Commun.*, 62(9):3084–3091, Sept 2014.

[58] D. Muller. Application of Boolean algebra to switching circuit design and to error detection. *IRE Tran. on Electronic Computers*, EC-3(3):6–12, Sept 1954.

[59] O. Ordentlich and U. Erez. Cyclic-coded

integer-forcing equalization. *IEEE Trans. Inform. Theory*, 58(9):5804–5815, 2012.

[60] I. Reed. A class of multiple-error-correcting codes and the decoding scheme. *IRE Tran. on Information Theory*, 4(4):38–49, September 1954.

[61] T. J. Richardson and R. L. Urbanke. *Modern Coding Theory*. Cambridge University Press, New York, NY, 2008.

[62] L. Russo. An approximate zero-one law. *Prob. Th. and Related Fields*, 61(1):129–139, 1982.

[63] R. Saptharishi, A. Shpilka, and B. L. Volk. Efficiently decoding Reed-Muller codes from random errors. [Online]. Available: http://arxiv.org/abs/1503.09092v2, 2015.

[64] R. Shaltiel and C. Umans. Simple extractors for all min-entropies and a new pseudo-random generator. In *Proc. IEEE Symp. on the Found. of Comp. Sci.*, pages 648–657, 2001.

[65] C. E. Shannon. A mathematical theory of communication. *The Bell Syst. Techn. J.*, 27:379–423, 623–656, July / Oct. 1948.

[66] V. M. Sidel'nikov and A. Pershakov. Decoding of Reed-Muller codes with a large number of errors. *Problems of Inform. Transm.*, 28(3):80–94, 1992.

[67] N. Sloane and E. Berlekamp. Weight enumerator for second-order Reed-Muller codes. *IEEE Trans. Inform. Theory*, 16(6):745–751, Nov 1970.

[68] D. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inform. Theory*, 42(6):1723–1731, Nov 1996.

[69] A. Ta-Shma, D. Zuckerman, and S. Safra. Extractors from Reed-Muller codes. In *Proc. IEEE Symp. on the Found. of Comp. Sci.*, pages 638–647, 2001.

[70] M. Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric & Functional Analysis*, 3(3):295–314, 1993.

[71] M. Talagrand. On Russo's approximate zero-one law. *The Ann. of Prob.*, pages 1576–1587, 1994.

[72] S. ten Brink. Convergence of iterative decoding. *Electronic Letters*, 35(10):806–808, May 1999.

[73] J.-P. Tillich and G. Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Combinatorics, Probability and Computing*, 9(05):465–479, 2000.

[74] J.-P. Tillich and G. Zemor. The Gaussian isoperimetric inequality and decoding error probabilities for the Gaussian channel. *IEEE Trans. Inform. Theory*, 50(2):328–331, Feb 2004.

[75] S. Yekhanin. Locally decodable codes. *Found. Trends Theor. Comput. Sci.*, 7(4):169–174, 1992.

[76] G. Zémor. Threshold effects in codes. In *Algebraic Coding*, pages 278–286. Springer, 1994.

APPENDIX

A. PROOFS

A.1 Proof of Proposition 11

S1 $\implies$ S2: By hypothesis, $P_b^{(n)}(p) \to 0$ for $0 \le p < 1-r$. As $P_b(p) = ph(p)$ from (2), and $h^{(n)}(0) = 0$, we have for $0 \le p < 1-r$,

$$\lim_{n\to\infty} h^{(n)}(p) = 0.$$

Now, let us focus on the limit of $h^{(n)}(p)$ for $1-r < p \le 1$. Fix $q \in (1-r, 1]$ and $\varepsilon > 0$. Choose n_0 large enough so that, for all $n > n_0$, we have $r_n > r - \varepsilon$ and $h^{(n)}(1-r-\varepsilon) \le \varepsilon$. Such an n_0 exists because $r_n \to r$ and $h^{(n)}(p) \to 0$ for $0 \le p < 1-r$. Since the function $h^{(n)}$ is increasing for all n, Proposition 1 implies that for all $n > n_0$,

$$\begin{aligned} r - \varepsilon < r_n &= \int_0^1 h^{(n)}(p)\mathrm{d}p \\ &= \int_0^{1-r-\varepsilon} h^{(n)}(p)\mathrm{d}p + \int_{1-r-\varepsilon}^{q} h^{(n)}(p)\mathrm{d}p \\ &\qquad + \int_q^1 h^{(n)}(p)\mathrm{d}p \\ &\le (1-r-\varepsilon)\varepsilon + (q-(1-r)+\varepsilon)h^{(n)}(q) \\ &\qquad + (1-q). \end{aligned}$$

This implies that

$$h^{(n)}(q) \ge \frac{q-(1-r)-\varepsilon(2-r-\varepsilon)}{q-(1-r)+\varepsilon} \ge 1 - \frac{3\varepsilon}{q-(1-r)}.$$

As such, $\lim_{n\to\infty} h^{(n)}(q) = 1$, for any $1-r < q \le 1$.

S2 $\implies$ S3: Since $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}$ is the width of the interval over which $h^{(n)}$ transitions from ε to $1-\varepsilon$, this follows immediately from S2.

S3 $\implies$ S1: It suffices to show that for any $q < 1-r$ and $\varepsilon > 0$, $p_\varepsilon^{(n)} \ge q$ for large enough n. This shows that $P_b^{(n)}(q) = qh^{(n)}(q) \le h^{(n)}(q) \le h^{(n)}(p_\varepsilon^{(n)}) = \varepsilon$ for large enough n, as desired.

Fix $q < 1-r$ and choose $\varepsilon > 0$ such that

$$\frac{1-r-2\varepsilon}{1-\varepsilon} - \varepsilon \ge q.$$

From the hypothesis, let n_0 be such that, for all $n > n_0$, $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)} \le \varepsilon$, and $r_n \le r + \varepsilon$. Then, from Proposition 1, we have

$$r_n = \int_0^1 h^{(n)}(\alpha)\mathrm{d}\alpha \ge \int_{p_{1-\varepsilon}^{(n)}}^1 h^{(n)}(\alpha)\mathrm{d}\alpha \ge (1-p_{1-\varepsilon}^{(n)})(1-\varepsilon),$$

which implies

$$p_{1-\varepsilon}^{(n)} \ge \frac{1-r_n-\varepsilon}{1-\varepsilon}. \tag{10}$$

Thus, for $n > n_0$,

$$\begin{aligned} p_\varepsilon^{(n)} &= p_{1-\varepsilon}^{(n)} - (p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}) \\ &\ge \frac{1-r_n-\varepsilon}{1-\varepsilon} - \varepsilon \\ &\ge \frac{1-r-2\varepsilon}{1-\varepsilon} - \varepsilon \ge q, \end{aligned}$$

by the choice of ε, which gives the desired result.

A.2 Proof of Theorem 13

Define $g(p) = \log \frac{h^{(n)}(p)}{1-h^{(n)}(p)}$ and observe that for $a_n < p < b_n$, we have

$$\frac{\mathrm{d}g(p)}{\mathrm{d}p} = \frac{1}{h^{(n)}(p)(1-h^{(n)}(p))}\frac{\mathrm{d}h^{(n)}(p)}{\mathrm{d}p} \ge w_n \log(N_n), \tag{11}$$

where the lower bound comes from the hypothesis.

Let $p_t^{(n)}$ be the inverse of $h^{(n)}$ as in (4) and fix $\varepsilon \in (0, 1/2]$. Assume that $a_n < p_\varepsilon^{(n)} \leq p_{1-\varepsilon}^{(n)} < b_n$. Then, by definition of $g(p)$,

$$\int_{p_\varepsilon^{(n)}}^{p_{1-\varepsilon}^{(n)}} \frac{\mathrm{d}g(p)}{\mathrm{d}p} dp = \log \frac{1-\varepsilon}{\varepsilon} - \log \frac{\varepsilon}{1-\varepsilon} = 2 \log \frac{1-\varepsilon}{\varepsilon}.$$

Also, from (11), we have

$$\begin{aligned} \int_{p_\varepsilon^{(n)}}^{p_{1-\varepsilon}^{(n)}} \frac{\mathrm{d}g(p)}{\mathrm{d}p} dp &\geq \int_{p_\varepsilon^{(n)}}^{p_{1-\varepsilon}^{(n)}} w_n \log(N_n) dp \\ &= w_n \log(N_n)(p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}), \end{aligned}$$

which implies

$$p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)} \leq \frac{2 \log \frac{1-\varepsilon}{\varepsilon}}{w_n \log(N_n)}.$$

When $p_\varepsilon^{(n)}$ and $p_{1-\varepsilon}^{(n)}$ do not lie between a_n and b_n, by considering different cases, it is straightforward to show that

$$p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)} \leq a_n + (1 - b_n) + \frac{2 \log \frac{1-\varepsilon}{\varepsilon}}{w_n \log(N_n)}.$$

Since $a_n \to 0$, $1 - b_n \to 0$, and $w_n \log(N_n) \to \infty$ from the hypothesis, we have $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)} \to 0$. Using this fact, from statement S2 of Proposition 11, we see that

$$p_{1/2}^{(n)} \to 1 - r. \tag{12}$$

Now, choose $\varepsilon_n = 1/N_n^2$ and observe that

$$\begin{aligned} p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} &\leq a_n + (1 - b_n) + \frac{1}{w_n \log(N_n)} \cdot 2 \log \frac{1-\varepsilon_n}{\varepsilon_n} \\ &\leq a_n + (1 - b_n) + \frac{1}{w_n \log(N_n)} \cdot 4 \log(N_n) \\ &= a_n + (1 - b_n) + \frac{4}{w_n}, \end{aligned}$$

which implies

$$p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} \to 0. \tag{13}$$

By combining (12), (13), and the fact that $p_{\varepsilon_n}^{(n)} \leq p_{1/2}^{(n)} \leq p_{1-\varepsilon_n}^{(n)}$, one obtains $p_{\varepsilon_n}^{(n)} \to 1 - r$. Hence, for any $p \in [0, 1 - r)$ and for n large enough, we conclude

$$\begin{aligned} P_B^{(n)}(p) &\leq N_n \cdot P_b^{(n)}(p) \\ &\leq N_n \cdot P_b^{(n)}(p_{\varepsilon_n}^{(n)}) \\ &= N_n \cdot p_{\varepsilon_n}^{(n)} \cdot h^{(n)}(p_{\varepsilon_n}^{(n)}) \\ &\leq N_n \cdot h^{(n)}(p_{\varepsilon_n}^{(n)}) = N_n \cdot \varepsilon_n = \frac{1}{N_n} \to 0, \end{aligned}$$

which implies that $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

A.3 Proof of Lemma 14

Take any distinct $i, j, k \in [N]$. In order to prove that $\mathcal{G}$ is doubly transitive, we will produce a $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$.

It is well known that for any vector space with two ordered bases $(\underline{u}_1, \ldots, \underline{u}_n)$ and $(\underline{u}'_1, \ldots, \underline{u}'_n)$, there exists an invertible $n \times n$ matrix T such that $\underline{u}_i = T\underline{u}'_i$ for all $i \in [n]$.

Recall that the elements of the vector space $\{0,1\}^n$ are enumerated by $\underline{e}_1, \ldots, \underline{e}_N$, where the bit i of a codeword is given by $f(\underline{e}_i)$ for some $f \in P(n, v)$. Note that, since i, j, k are distinct, $\underline{e}_j - \underline{e}_i \neq 0^n$ and $\underline{e}_k - \underline{e}_i \neq 0^n$. Therefore, there exists an invertible $n \times n$ binary matrix T such that $T(\underline{e}_j - \underline{e}_i) = \underline{e}_k - \underline{e}_i$. Now, we construct $\pi \colon [N] \to [N]$ by defining $\pi(\ell)$ as the unique ℓ' such that

$$\underline{e}_{\ell'} = T(\underline{e}_\ell - \underline{e}_i) + \underline{e}_i.$$

Since T is invertible, it follows that $\pi \in S_N$. Also, by construction, $\pi(i) = i$ and $\pi(j) = k$.

It remains to show that $\pi \in \mathcal{G}$. Consider a codeword in $\mathrm{RM}(v, n)$ given by $f \in P(n, v)$. It suffices to produce a $g \in P(n, v)$ such that $g(\underline{e}_{\pi(\ell)}) = f(\underline{e}_\ell)$ for all $\ell \in [N]$. Let

$$g(x_1, \ldots, x_n) = f(T^{-1}[x_1, \ldots, x_n]^{\mathrm{T}} - T^{-1}\underline{e}_i + \underline{e}_i).$$

Then, $\mathrm{degree}(f) = \mathrm{degree}(g)$, and $g(\underline{e}_{\pi(\ell)}) = f(\underline{e}_\ell)$ for all $\ell \in [N]$. Therefore, $g \in P(n, v)$ and $\pi \in \mathcal{G}$.

A.4 Proof of Lemma 16

Note that the elements of the vector space $\{0,1\}^n$ are enumerated by $\underline{e}_1, \ldots, \underline{e}_N$, where $\underline{e}_N = \underline{0}$ and the bit i of a codeword is given by $f(\underline{e}_i)$ for some $f \in P(n, v)$.

Given $T \in \mathrm{GL}(n, \mathbb{F}_2)$, we construct $\pi_T \in S_{N-1}$ by defining $\pi_T(\ell)$ as the unique ℓ' such that

$$\underline{e}_{\ell'} = T\underline{e}_\ell, \quad \text{for } \ell \in [N-1].$$

Since T is invertible, above, $\ell' \neq N$ and π_T is well-defined. Moreover, it is easy to check that $\pi_{T_1} \circ \pi_{T_2} = \pi_{T_1 T_2}$ for $T_1, T_2 \in \mathrm{GL}(n, \mathbb{F}_2)$. As such, the collection of permutations

$$\mathcal{H} = \{\pi_T \in S_{N-1} \mid T \in \mathrm{GL}(n, \mathbb{F}_2)\}$$

forms a subgroup of S_{N-1} isomorphic to $\mathrm{GL}(n, \mathbb{F}_2)$. Also, for $i, j \in [N-1]$, there exists $T \in \mathrm{GL}(n, \mathbb{F}_2)$ such that $\underline{e}_j = T\underline{e}_i$. Consequently, there exists $\pi_T \in \mathcal{H}$ such that $\pi_T(i) = j$ and, therefore, $\mathcal{H}$ is transitive. Below, we show that $\mathcal{H} \subseteq \mathcal{G}_N$.

For this, associate $\pi_T \in \mathcal{H}$ with $\pi'_T \in S_N$ where

$$\pi'_T(\ell) = \pi_T(\ell) \quad \text{for } \ell \in [N-1], \qquad \pi'_T(N) = N.$$

Also, it is easy to show that $\pi_T \in \mathcal{G}_N$ if $\pi'_T \in \mathcal{G}$, the permutation group of $\mathrm{RM}(v, n)$. To see that $\pi'_T \in \mathcal{G}$, consider a codeword given by $f \in P(n, v)$. It suffices to produce a $g \in P(n, v)$, where $g(\underline{e}_{\pi'_T(\ell)}) = f(\underline{e}_\ell)$ for $\ell \in [N]$.

The desired g is given by

$$g(x_1, \ldots, x_n) = f(T^{-1}[x_1, \ldots, x_n]^{\mathrm{T}}),$$

by observing that $\mathrm{degree}(g) = \mathrm{degree}(f)$ and

$$g(\underline{e}_N) = f(T^{-1}0^n) = f(\underline{e}_N).$$