REED-MULLER CODES ACHIEVE CAPACITY ON ERASURE CHANNELS

E2-207: Concentration Inequalities

COURSE PROJECT REPORT [*]

A report by

RAMAKRISHNAN ,SOUMYA SUBHRA BANERJEE
DEPARTMENT OF ECE,
INDIAN INSTITUTE OF SCIENCE.

November 28, 2017

CONTENTS

ABSTRACT

Efficiently decodable deterministic coding schemes which achieve channel capacity provably have been elusive until the advent of polar codes[**?**] in the last decade.Further,the recent results by Urbanke et al.[**?**] show that doubly transitive codes achieve capacity on erasure channel under MAP decoding.Urbanke and his group use threshold phenomenon observed in EXIT function , which capture the error probability ,to prove the same.These results were applied to Reed-Muller codes [**?**].Alternative proof of the fact Polar codes achieve capacity was suggested in [**?**].This report is a comprehensive study of threshold phenomenon in EXIT function and its applications as indicated above.

[*] Original work by *S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşŏglu and R. Urbanke.*

INTRODUCTION

The possibility of construction of low-complexity structured codes which provably achieve channel capacity ,having geometric structure and deterministic construction, have been explored since the conceptualization of channel capacity in [?] . Turbo codes,LDPC ,spatially coupled LDPC were practical constructions towards this end.Polar codes [?] were the first provably capacity achieving codes for arbitrary Binary Input Symmetric Discrete Memoryless channels (BMS) with desirable complexity and deterministic structure.

Interestingly [?],exploits the symmetry of deterministic linear codes, and threshold phenomenon in boolean function analysis to arrive at results summarized by the following theorem for transmission over a BEC channel.It is useful to point out that capacity of BEC(p),$r = 1 - p$.

Theorem - A sequence of linear codes achieves capacity on a memoryless erasure channel under MAP decoding if its blocklengths are strictly increasing, its code rates converge to some $r \in (0, 1)$, and the permutation group of each code is doubly transitive. The result, coupled with the fact that Reed-Muller are doubly transitive,extends to prove the capacity achieving nature of RM codes.It further provides an alternative prove of the same for Polar codes,without employing the inherent symmetry of the Hadamard matrix (consequently Arikan transform), from which polar codes derive their structure.

Brief introduction to Reed Muller codes:

Reed-Muller codes are one of the oldest families of error correcting codes and use concepts from algebra for the encoding and decoding process. The idea is to look at the message as the coefficients of a multivariate polynomial of a suitable degree and pass its evaluations.They were introduced by Muller [?] ,later Reed [?] proposed a majority logic decoder for these codes. A binary Reed-Muller (RM) code with parameters m and ν (the order) is a linear code of length 2^m, dimension $\binom{m}{0} + \ldots + \binom{m}{\nu}$ and minimum distance $2^{m-\nu}$. Although the possibility of RM codes achieving capacity under MAP decoding has been discussed in several papers, they being able to correct *almost all* erasure patterns up to the capacity limit was not clear until this work.

Overall proof idea from the perspective of Threshold Phenomenon:

Here we assume the reader has prior exposure to threshold phenomenon of boolean functions.

For a sequence of binary linear codes with rate r to be capacity achieving, the bit error probability,must converge to 0 for any erasure rate below $1 - r$.Towards this, (1)if the bit error probability under MAP-decoding can be captured by a function of erasure probability (p),(2)which in turn is the measure of a symmetric monotone set ,then we shall observe threshold phenomenon,(3) if the threshold is sharp occuring at $p = 1 - r$ under the settings of stated theorem ,it provides a proof for the theorem. Extrinsic information transfer (EXIT) function [?] denoted by $h(p)$ in fig:1, and the area theorem for EXIT functions [?] occupy a central role in this work. For a given input bit, the EXIT function is defined to be the conditional entropy of the input bit i given the outputs associated with all other input bits.The value of the EXIT function at a particular erasure value is also directly related to the bit error probability under bit-MAP decoding,hence EXIT functions serve as (1).Furthermore, EXIT function can be associated with measure of set of erasure patterns Ω_i which are symmetric monotone sets for doubly transitive codes, solving(2).Finally application of area theorem solves (3). The rest of the report elaborates on (1),(2) and (3) with focus on bit error probability under bit-MAP decoding.It extends the results to block error probability.

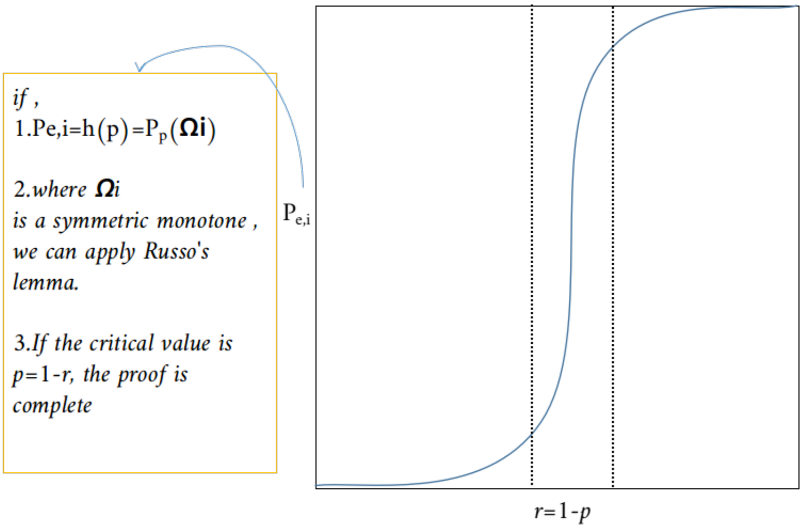

Figure 1: "Achieving Capacity" from the perspective of threshold phenomenon.

PRELIMINARIES

Let $\mathcal{C}$ denote an (N, K) proper binary linear code with length N ,dimension K ,minimum distance (d_{min}) atleast 2,and rate defined by $r \triangleq K/N$. We assume that a random codeword is chosen uniformly from this code and transmitted over a memoryless Binary Erasure Channel (BEC).A BEC with erasure probability p is denoted by BEC(p),or BEC($\underline{p}$) in case the erasure probability is different for each bit where $\underline{p} = (p_1, \ldots, p_n)$ and p_i indicates the erasure probability for bit i. The input and output alphabets of the BEC are denoted by $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, e\}$, respectively. Let $\underline{X} = (X_1, \ldots, X_N) \in \mathcal{X}_N$ be a uniform 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 BEC(p).For a vector $a = (a_1, a_2, \ldots, a_N)$, the shorthand $\underline{a}_{\sim i}$ denotes $(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N)$. Let $\underline{a}$,$\underline{b}$ denote the indicator vectors of the sets $A \subseteq [N]$,$B \subseteq [N]$.We say that A *covers* B if $B \subseteq A$, equivalently $\underline{a} \preccurlyeq \underline{b}$.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 codeword.Similarly, it is possible to recover bit i if and only if the erasure pattern does not cover any codeword where bit i is non-zero.

Bit and Block Erasure Probability

BIT ERROR PROBABILITY: Let $D_i : \mathcal{Y}^N \rightarrow \mathcal{X} \cup \{e\}$ denote the bit-MAP decoder for bit i of $\mathcal{C}$. 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 e. Let the erasure probability for bit $i \in [N]$ be

$$P_{b,i} \triangleq \mathbb{P}[D_i(\underline{Y}) \neq X_i].$$

and the average bit erasure probability be

$$P_b \triangleq \frac{1}{N} \sum_{i=1}^{N} P_{b,i}.$$

Whenever bit i can be recovered from a received sequence $\underline{Y} = \underline{y}$, $H(X_i|\underline{Y} = \underline{y}) = 0$. Otherwise, the uniform codeword assumption implies that the posterior marginal of bit i given the observations is $\mathbb{P}(X_i = x|\underline{Y} = \underline{y}) = 1/2$ and $H(X_i|\underline{Y} = \underline{y}) = 1$. This immediately implies that

$$P_{b,i} = H(X_i|\underline{Y})$$

and,

$$P_b = \frac{1}{N}\sum_{i=1}^{N} H(X_i|\underline{Y}).$$

BLOCK ERROR PROBABILITY: TO BE EDITED BY RAMKI

MAP EXIT Functions

The vector EXIT function associated with bit i of the (uniformly randomly chosen) codeword is

$$h_i(\underline{p}) \triangleq H(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})).$$

The average vector EXIT function is defined by

$$h(\underline{p}) \triangleq \frac{1}{N}\sum_{i=1}^{N} h_i(\underline{p}).$$

Scalar EXIT functions are defined by choosing $\underline{p} = (p, p, \dots, p)$.

$$\begin{aligned} H(X_i|\underline{Y}) &= \mathbb{P}(Y_i = e)H(X_i|\underline{Y}_{\sim i}, Y_i = e) + \mathbb{P}(X_i = Y_i)H(X_i|\underline{Y}_{\sim i}, Y_i = X_i) \\ &= \mathbb{P}(Y_i = e)H(X_i|\underline{Y}_{\sim i}). \end{aligned}$$

Therefore,

$$P_{b,i}(p) = ph_i(p)$$

and

$$P_b(p) = ph(p).$$

Proposition 1. The MAP EXIT function for the ith bit satisfies $h_i(p) = \frac{\partial H(\underline{X}/\underline{Y}(\underline{p}))}{\partial p_i}$

Proof.

$$\begin{aligned} H(\underline{X}/\underline{Y}(\underline{p})) &= H(X_i/\underline{Y}(\underline{p})) + H(X_{\sim i}/X_i, \underline{Y}(\underline{p}) \\ &= H(X_i/\underline{Y}(\underline{p})) + H(X_{\sim i}/X_i, Y_{\sim i}) \text{ ,by memorylessness} \\ &= p_i h_i(p) + H(X_{\sim i}/X_i, Y_{\sim i}) \end{aligned}$$

We note the second term is independent of p_i,the proposition follows on differentiation. □

INDIRECT RECOVERY Consider a code $\mathcal{C}$ and the *indirect recovery* of X_i from the subvector $\underline{Y}_{\sim i}$ (i.e., the bit-MAP decoding of Y_i from $\underline{Y}$ when $Y_i = e$). For $i \in [N]$, the set of erasure patterns that prevent indirect recovery of X_i under bit-MAP decoding is given by

Definition 2. $\Omega_i \triangleq \{A \subseteq [N]\backslash\{i\} : \exists B \subseteq [N]\backslash\{i\}, B \cup \{i\} \in \mathcal{C}, B \subseteq A\}$.

For distinct $i, j \in [N]$, the set of erasure patterns where the j-th bit is *pivotal* for the indirect recovery of X_i is given by

Definition 3. $\partial_j\Omega_i \triangleq \{A \subseteq [N]\backslash\{i\} : A\backslash\{j\} \notin \Omega_i, A \cup \{j\} \in \Omega_i\}$

These are the erasure patterns where X_i can be recovered from $\underline{Y}_{\sim i}$ if and only if $Y_j \neq e$.Note $\partial_j\Omega_i$ includes patterns from both Ω_i and Ω_i^c.

Proposition 4. For a code $\mathcal{C}$ and transmission over a BEC, we have the following properties for the EXIT functions.

(a) The EXIT function associated with bit i satisfies

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

(b) For $j \in [N] \backslash \{i\}$, the partial derivative satisfies

$$\left. \frac{\partial h_i(\underline{p})}{\partial p_j} \right|_{\underline{p}=(p,p,\ldots,p)} = \mu_p(\partial_j \Omega_i) = \sum_{A \in \partial_j \Omega_i} p^{|A|}(1-p)^{N-1-|A|}.$$

(c) The average EXIT function satisfies the *area theorem*

$$\int_0^1 h(p)dp = \frac{K}{N}.$$

$H(\underline{X}/UnderlineY(1)) -$

Where $\mu_p(\Omega)$ is the measure of the set of erasure patterns Ω. Here (a) and (b) follow from definition of conditional entropy and the fact that $H(X_i/\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1$ when $A \cup \{i\}$ covers some codeword and decoding fails, and 0 otherwise.(c) is a direct consequence of proposition 1 From the above discussion it is clear that the measure of the set Ω_i is equal to the probability of error for bit i , which in turn is equal to the ith EXIT function due to the uniform input assumption.

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.

Definition 5. The permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is defined to be

$$\mathcal{G} = \{\pi \in S_N : \pi(A) \in \mathcal{C} \text{ for all } A \in \mathcal{C}\}.$$

Definition 6. - Suppose $\mathcal{G}$ is a permutation group. Then,

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

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

Proposition 7. *All EXIT functions are equal.* Suppose the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is transitive. Then, for any $i \in [N]$,

$$h(p) = h_i(p) \text{ for } 0 \leqslant p \leqslant 1.$$

Proof. claim: *if $\mathcal{G}$ is transitive, then so is Ω_i.*
As $A \in \Omega_i$,by definition $\exists B$, s.t., $B \cup \{i\} \in \mathcal{C}$, *but by transitivity of $\mathcal{G}$,* $\pi(B \cup \{i\}) \in \mathcal{C}$.
Observe,

$$\pi(B \cup \{i\}) = \pi(B) \cup \pi(\{i\}) = \pi(B) \cup j$$

.

*Since $\pi(B) \subseteq \pi(A)$,it follows $\pi(A) \in \Omega_j$ .This indicates a bijection between Ω_i and Ω_j,i.e,$|\Omega_i| = |\Omega_j|$.Moreover since, $|A| = |\pi(A)|$, propostion follows from proposition*4 (a) □

Proposition 8. Suppose the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is doubly transitive. Then, for distinct $i, j, k \in [N]$, and any $0 \leqslant p \leqslant 1$,

$$\left.\frac{\partial h_i(\underline{p})}{\partial p'_j}\right|_{\underline{p}=(p,p,\ldots,p)} = \left.\frac{\partial h_i(\underline{p})}{\partial p'_k}\right|_{\underline{p}=(p,p,\ldots,p)}.$$

Proof. Similar to the proof of proposition 2.Intuitively we expect that once we permute the locations , the bits which were pivotal must continue to remain so.Otherwise we could have decoded the concerned bit using simple permutations. □

SUMMARY OF PROPERTIES OF EXIT FUNCTION.

1 $h_i(p)$ captures the bit error probability of MAP decoder.

2 $h_i(p)$ is measure of Ω_i.

3 All EXIT functions are equal to average EXIT function h(p).

4 $h_i(p)$ is strictly increasing and invertible.(follows from proposition 4 (b))

5 The area under the h vs p curve is the rate(by *Area Theorem*).

We may notice here, that is $A \in \Omega_i$, it is a bit erasure pattern that causes error at position i, then $B \supset A$ will surely cause errors, and $B \in \Omega_i$.Thus Ω_i is monotone.Proving Ω_i is symmetric and has a sharp threshold at $p = 1 - r$, will establish that 2-transitive codes achieve Capacity.We will formalize this in the following subsections.

Capacity achieving codes and EXIT function

Definition 9. 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) BLOCK MAP DECODING TO BE EDITED BY RAMKI

The following theorem bridges capacity achieving codes, average EXIT functions, and the sharp transition framework that allows us to show that the transition width of certain functions goes to 0. The average EXIT functions of some rate-1/2 Reed-Muller codes are shown in Figure 2. Observe that as the blocklength increases, the transition width of the average EXIT function decreases. According to the following proposition, if this width converges to 0, then Reed-Muller codes achieve capacity on the BEC under bit-MAP decoding.

Proposition 10. Let $\{\mathcal{C}_n\}$ be a seq. of codes with rates $\{r_n\}$, $r_n \to r$ for $r \in (0, 1)$. The following are equivalent -

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

S2: The sequence of average EXIT functions satisfies

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

S3: For any $0 < \epsilon \leqslant 1/2$,

$$\lim_{n\to\infty} p_{1-\epsilon}^{(n)} - p_{\epsilon}^{(n)} = 0.$$

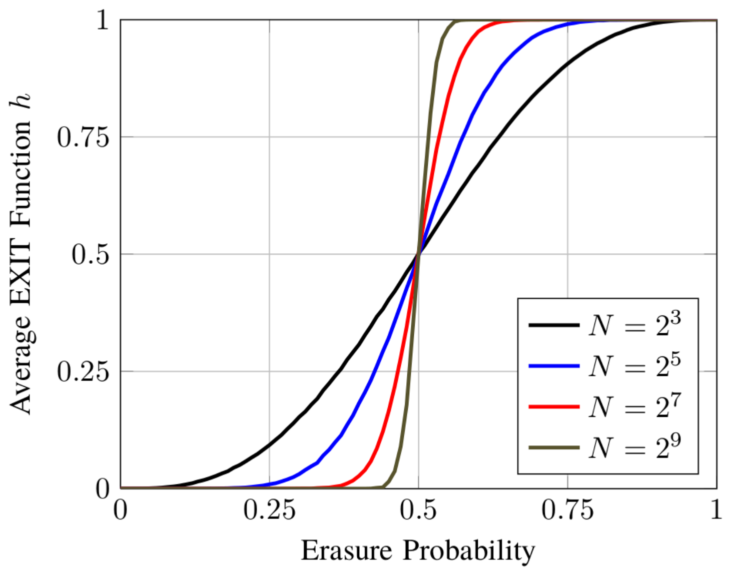

Figure 2: The average EXIT function of the rate-1/2 Reed-Muller code with blocklength N.

where $h^{(n)}(p_{\epsilon}^{(n)}) = \epsilon$.
In short S1 $\Rightarrow$ S2, due to close relationship between bit error probability and average EXIT function pointed out in propostion 4 .S2 $\Rightarrow$ S3,and S3 $\Rightarrow$ S1 by area theorem.Hence, proving S3 suffices to complete the proof.

MAIN RESULT

Doubly-transitive codes are Capacity Achieving under bit-MAP decoding.

Definition 11. We can redefine Ω_i as a set of indicator vectors of A.Let,

$$[\phi_i(A)]_l = \begin{cases} \mathbf{1}_A(l) \text{ if } l < i \\ \mathbf{1}_A(l+1) \text{ if } l \geqslant i. \end{cases}$$

$$\begin{aligned} \Omega_i' &\triangleq \{\phi_i(A) \in \{0,1\}^{N-1} : A \in \Omega_i\} \\ \partial_j \Omega_i' &\triangleq \{\phi_i(A) \in \{0,1\}^{N-1} : A \in \partial_j \Omega_i\}. \\ &= \{\underline{x} \in \{0,1\}^{N-1} | \mathbb{1}_{\Omega_i}(\underline{x}) \neq \mathbb{1}_{\Omega_i}(\underline{x}^{(j)})\} \end{aligned}$$

Here the last equality follows from definition of $\partial_j \Omega_i$.
Consider the space $\{0,1\}^M$,we can redefine measure μ_p such that

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

where the weight $|\underline{x}| = x_1 + x_2 + \ldots + x_M$ is the number of 1's in $\underline{x}$.

Definition 12. For a monotone set Ω.The influence of bit $j \in [N]$,is defined by,

$$I_j^{(p)}(\Omega) \triangleq \mu_p(\partial_j \Omega)$$

The total influence in defined by,

$$I^{(p)} \triangleq \sum_{j=1}^{N} I_j^{(p)}.$$

Using proposition 4(a) and proposition 2, we have,

$$h(p) = h_i(p) = \mu_p(\Omega_i')$$

Further,from proposition 4(b), we get,

$$I_j^p(\Omega_i') = \mu_p(\partial_j \Omega_i') = \left.\frac{\partial h_i(\underline{p})}{\partial p_j'}\right|_{\underline{p}=(p,p,\dots,p)}$$

where j' is given by

$$j' = \begin{cases} j \text{ if } j < i \\ j+1 \text{ if } j \geqslant i. \end{cases}$$

Since $\mathcal{G}$ is doubly transitive, from propostion 8,

$$I_j^p(\Omega_i') = I_k^p(\Omega_i') \text{ for all } j,k \in [N-1].$$

Hence,Ω_i is a *symmetric monotone set.*

The following theorem could be seen as a consequence of the result by Talagrand [**?**].

Theorem 1. *Let Ω be a monotone set and suppose that, for all $0 \leqslant p \leqslant 1$, the influences of all bits are equal $I_1^{(p)}(\Omega) = \ldots = I_M^{(p)}(\Omega)$. Then, for any $0 < \epsilon \leqslant 1/2$,*

$$p_{1-\epsilon} - p_\epsilon \leqslant \frac{2\log\frac{1-\epsilon}{\epsilon}}{C\log(N-1)},$$

where $p_t = \inf\{p \in [0,1] : \mu_p(\Omega) \geqslant t\}$ is well defined because $\mu_p(\Omega)$ is strictly increasing in p with $\mu_0(\Omega) = 0$ and $\mu_1(\Omega) = 1$.

Proof. Using Russo's lemma [**?**]. □

We see that Ω_i satisfies the conditions of Theorem 1. Hence,

$$\lim_{n\to\infty}(p_{1-\epsilon} - p_\epsilon) = 0.$$

Further , using proposition 10, we state,
$\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.

Doubly-transitive codes are Capacity achieving under block-MAP decoding

BLOCK ERROR TO BE EDITED BY RAMKI

APPLICATIONS

Theorem 2. *Let $\{\mathcal{C}_n\}$ be a sequence of codes where the blocklengths satisfy $N_n \to \infty$, the rates satisfy $r_n \to r$, and the permutation group $\mathcal{G}(n)$ (of $\mathcal{C}_n$) is doubly transitive for each n. If $r \in (0,1)$, then $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.*
In particular, $\mathrm{RM}(\nu, m)$ codes achieve capacity.

Proof. Let the average EXIT function of $\mathcal{C}_n$ be $h^{(n)}$. Fix some $i \in [N]$. Since $\mathcal{G}$ is transitive, from Theorem 2,

$$h(p) = h_i(p) \text{ for all } p \in [0,1].$$

Consider the sets Ω_i' from Definitions 2 and **??**, and let $M = N - 1$. Observe that, from Theorem 4,

$$h_i(p) = \mu_p(\Omega_i'), \qquad I_j^p(\Omega_i') = \left.\frac{\partial h_i(\underline{p})}{\partial p_{j}'}\right|_{\underline{p}=(p,p,\ldots,p)}$$

where j' is given by

$$j' = \begin{cases} j \text{ if } j < i \\ j+1 \text{ if } j \geqslant i. \end{cases}$$

Since $\mathcal{G}$ is doubly transitive, from Theorem 8,

$$I_j^p(\Omega_i') = I_k^p(\Omega_i') \text{ for all } j,k \in [N-1].$$

Using Theorem **??**, we have

$$p_{1-\epsilon} - p_\epsilon \leqslant \frac{2\log\frac{1-\epsilon}{\epsilon}}{C\log(N-1)},$$

where p_t is the functional inverse of h as in Theorem **??**. Since $N \to \infty$ from the hypothesis,

$$\lim_{n\to\infty}(p_{1-\epsilon} - p_\epsilon) = 0.$$

Now, using Theorem **??**, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding. Now, that $\mathrm{RM}(v,m)$ codes achieve capacity follows from Theorem **??**. □

CONCLUSION

This report is a summary of the result by S. Kudekar et al [?] wherein it is proved that Reed Muller codes achieve capacity on erasure channels under bit-MAP decoding. The key ideas used in the proof could be enumerated as follows:

1. The permutation group of Reed Muller codes is doubly transitive.
2. For an erasure channel, the average EXIT function is directly related to the average bit erasure probability.
3. The average EXIT function satisfies the *area theorem*. i.e., the area under the curve equals the rate of the code.
4. Showing that a sequence of codes is capacity achieving (the average bit erasure probabilities go to zero) is *equivalent to* showing that the sharp transition width of EXIT functions go to zero.

The interested reader may please see the original work [?, ?, ?] for complete proofs and other details.