# Almost Optimal Scaling of Reed-Muller Codes on BEC and BSC Channels

Hamed Hassani*, Shrinivas Kudekar†, Or Ordentlich‡, Yury Polyanskiy§ and Rüdiger Urbanke¶
*University of Pennsylvania, Email: hassani@seas.upenn.edu
†Email: kudekar@gmail.com
‡Hebrew University of Jerusalem, Israel, Email: or.ordentlich@mail.huji.ac.il
§MIT EECS, USA, Email: yp@mit.edu
¶School of Computer and Communication Sciences, EPFL, Switzerland, Email: ruediger.urbanke@epfl.ch

***Abstract*—Consider a binary linear code of length $N$, minimum distance $d_{\text{min}}$, transmission over the binary erasure channel with parameter $0 < \epsilon < 1$ or the binary symmetric channel with parameter $0 < \epsilon < \frac{1}{2}$, and block-MAP decoding. It was shown by Tillich and Zemor that in this case the error probability of the block-MAP decoder transitions "quickly" from $\delta$ to $1-\delta$ for any $\delta > 0$ if the minimum distance is large. In particular the width of the transition is of order $O(1/\sqrt{d_{\text{min}}})$. We strengthen this result by showing that under suitable conditions on the weight distribution of the code, the transition width can be as small as $\Theta(1/N^{\frac{1}{2}-\kappa})$, for any $\kappa > 0$, even if the minimum distance of the code is not linear. This condition applies e.g., to Reed-Mueller codes. Since $\Theta(1/N^{\frac{1}{2}})$ is the smallest transition possible for any code, we speak of "almost" optimal scaling. We emphasize that the width of the transition says nothing about the location of the transition. Therefore this result has no bearing on whether a code is capacity-achieving or not. As a second contribution, we present a new estimate on the derivative of the EXIT function, the proof of which is based on the Blowing-Up Lemma.**

## I. Introduction

Consider a binary linear code of length $N$ and minimum distance $d_{\min}$. Assume that we transmit over the binary erasure channel (BEC) with parameter $\epsilon$, $0 < \epsilon < 1$, or the binary symmetric channel (BSC) with parameter $\epsilon$, $0 < \epsilon < \frac{1}{2}$. Assume further that the receiver performs block maximum-a posteriori (block-MAP) decoding. It was shown by Tillich and Zemor [1] that in this case the error probability transitions "quickly" from $\delta$ to $1-\delta$ for any $\delta > 0$ if the minimum distance is large. In particular they showed that the width of the transition is of order $O(1/\sqrt{d_{\min}})$. For codes whose minimum distance is proportional to the blocklength, this gives a transition width of $O(1/\sqrt{N})$) and this is the best possible. But for codes whose minimum distance is sublinear the width "guaranteed" by this result is sub-optimal. E.g, if we consider Reed-Mueller (RM) codes of fixed rate and increasing length, then their minimum distance grows only like $\Theta(\sqrt{N})$.

In this paper, we strengthen this scaling result. We show that under suitable conditions on the weight distribution of the code, the transition width will be nearly optimal, i.e., it will be as small as $\Theta(1/N^{\frac{1}{2}-\kappa})$, for any $\kappa > 0$. The required condition applies e.g., to RM codes, and hence we see that RM codes have an almost optimal scaling of their block error probability under block-MAP decoding.

It is important to note that the width of the transition has no bearing on *where* this transition happens. This is analogous to concentration results in probability (think of Azuma's inequality) where one can prove that a random variable is concentrated around its mean without determining the value of the mean. Therefore this result has no bearing on whether a code is capacity-achieving or not. In particular, our result does not resolve the question whether RM codes are capacity-achieving over any channel other than the BEC, see [2], [3].

Moreover, even though RM codes are known to achieve capacity over the BEC, our results do not imply that the gap to capacity of these codes at a fixed error probability scales like $O(1/\sqrt{N})$. The reason being that [2] only shows that a sharp transition occurs at $\epsilon^* > 1 - C - O(1/\log N)$, where $\epsilon^*$ denotes the channel parameter at which the block-MAP error is equal to $1/2$ and $C$ denote the channel capacity. However, under the assumption that $\epsilon^* = 1 - C$, our results imply that for RM codes the blocklength $N$ scales wrt the gap to capacity $C - R$ in the (almost) optimal way, i.e. $N = \Theta(1/(C-R)^{2+\kappa})$ or equivalently $C - R = O(1/N^{1/2-\kappa})$, for any $\kappa > 0$.

To establish the desired $O(1/\sqrt{N})$ gap-to-capacity result for RM over the BEC, one would need to obtain tighter bounds on $\epsilon^*$. As a first step in this direction, we develop a new tool for estimating the derivative of the EXIT function. Roughly speaking, we show that, for a transitive code, if for most pairs of erased locations $(i, j)$ for which $H(X_i, X_j | Y_{\sim i,j}) = 1$ bit, the conditional probability of the event $X_i = X_j = 1$ does not decrease with $N$, then the EXIT function transitions sharply with a transition width of $O(1/\sqrt{N})$. While we are currently unable to verify this condition for RM codes analytically, numerical indications suggest that this might indeed be the case. Our estimate on the EXIT function derivative is based on the Blowing-Up Lemma.

## II. Preliminaries

**Linear Codes.** Let $\mathcal{C}$ be a binary linear code of length $N$, dimension $K$, and minimum distance $d_{\min}$. We let $A(w)$ denote the weight distribution function of $\mathcal{C}$, i.e., for any $w \leq N$ we have

$$A(w) := |\{x \in \mathcal{C} : w_H(x) = w\}|, \tag{1}$$

where $w_H(x)$ denotes the Hamming weight of vector $x$. Let us also define the function $A(W, z)$ as follows:

$$A(W, z) \triangleq \sum_{w=1}^{W} A(w) z^w. \tag{2}$$

**Transmission Channel and Block-MAP Decoding** We consider transmission over two types of channel families: the binary erasure channel with parameter $\epsilon$ (BEC($\epsilon$)) and the binary symmetric channel with cross-over probability $\epsilon$ (BSC($\epsilon$)). Let $X$ be the codeword, chosen uniformly at random from $\mathcal{C}$, and let $Y$ be the received word. When transmission is over the BEC we have $Y \in \{0, 1, ?\}^N$, and when it is over the BSC we have $Y \in \{0, 1\}^N$. Let $\hat{x}(y)$ be the block-MAP decoding function

$$\hat{x}(y) = \text{argmax}_{x\in\mathcal{C}} p(y|x), \tag{3}$$

where ties are resolved in an arbitrary fashion.

We let $P_{\text{MAP}}(\epsilon)$ denote the probability of error for the block-MAP decoder, i.e., $P_{\text{MAP}}(\epsilon) = P\{\hat{x}(Y) \neq X\}$. Here, to simplify notation, we have used the same notation (i.e., $P_{\text{MAP}}(\epsilon)$) for transmission over both the BEC($\epsilon$) and the BSC($\epsilon$), and in the sequel, the choice of the transmission channel will be clear from the context.

**Sharp Transition for the Block-MAP Error.** Let us view $P_{\text{MAP}}(\epsilon)$ as a function of the channel parameter $\epsilon$. Consider first transmission over the BEC. In this case, it is not hard to see that $P_{\text{MAP}}(\epsilon)$ is an increasing function of $\epsilon$ for $\epsilon \in [0, 1]$ with $P_{\text{MAP}}(\epsilon = 0) = 0, P_{\text{MAP}}(\epsilon = 1) = 1$. Furthermore, the function $P_{\text{MAP}}$ exhibits a sharp transition behaviour [1]: Let $\epsilon^*$ be such that $P_{\text{MAP}}(\epsilon^*) = \frac{1}{2}$. Then, around $\epsilon = \epsilon^*$, the value of $P_{\text{MAP}}$ jumps from “almost zero" to “almost one" and the transition width is of oder $O(1/\sqrt{d_{\min}})$. We refer to Fig. 1 for a schematic illustration. The same picture holds true when the transmission channel is a BSC($\epsilon$). More precisely, we have the following theorem from [1].

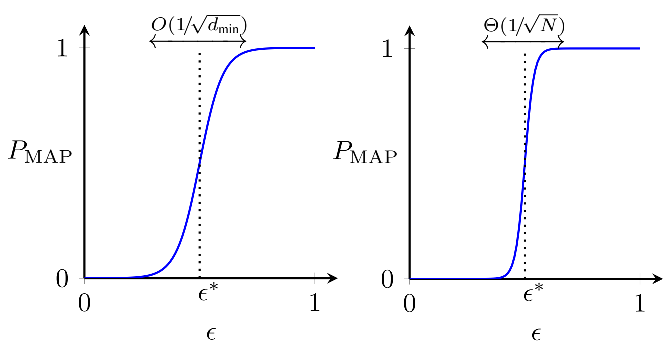


Figure 1: Left: Sharp transition of the block-MAP error. From [1] we know that the transition width is $O(1/\sqrt{d_{\min}})$. Right: The *optimal* transition width is $\Theta(1/\sqrt{N})$.

**Theorem 1.** *Let $P_{\text{MAP}}$ be the block-MAP error for transmission of a linear code $\mathcal{C}$ with minimum distance $d_{\min}$ over BEC($\epsilon$). We have*

$$P_{\text{MAP}}(\epsilon) \leq \Phi(\sqrt{2d_{\min}}(\sqrt{-\ln\epsilon^*} - \sqrt{-\ln\epsilon})) \quad \textit{for } 0 < \epsilon < \epsilon^*,$$
$$P_{\text{MAP}}(\epsilon) \geq \Phi(\sqrt{2d_{\min}}(\sqrt{-\ln\epsilon^*} - \sqrt{-\ln\epsilon})) \quad \textit{for } \epsilon^* < \epsilon < 1,$$

*where $\epsilon^*$ is defined by $P_{\text{MAP}}(\epsilon^*) = \frac{1}{2}$ and $\Phi$ stands for the Gaussian cumulative distribution, i.e., $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$.*

*Furthermore, when the transmission is over BSC($\epsilon$), we have for $\epsilon < \epsilon^*$:*

$$P_{\text{MAP}}(\epsilon) \leq \Phi(\sqrt{d_{\min}}(\sqrt{-\ln(1-\epsilon^*)} - \sqrt{-\ln(1-\epsilon)})),$$

*and for $\epsilon^* < \epsilon < \frac{1}{2}$:*

$$P_{\text{MAP}}(\epsilon) \geq \Phi(\sqrt{d_{\min}}(\sqrt{-\ln(1-\epsilon^*)} - \sqrt{-\ln(1-\epsilon)})).$$

**Optimal Transition Width and its Implications.** Theorem 1 implies that when the code has linear minimum distance (e.g., random codes or most LDPC codes) then the transition width is $O(1/\sqrt{N})$. For RM codes, which have minimum distance $O(\sqrt{N})$ if we consider elements of fixed rate and increasing length, the implied transition width is $O(N^{-\frac{1}{4}})$.

This suggests the following question: What is the optimal scaling of the transition width (i.e., how “fast” can the transition be) in terms of the blocklength $N$? It is not hard to see that the optimal transition width is $\Theta(1/\sqrt{N})$, see Fig. 1. An intuitive argument for this (e.g., for the BEC) is that, for any $\epsilon$, the number of channel erasures is with high probability smeared out over the window $[N\epsilon - \Theta(\sqrt{N}), N\epsilon + \Theta(\sqrt{N})]$. As a result, one cannot expect a drastic change in $P_{\text{MAP}}$ between $\epsilon^*$ and $\epsilon^* + o(\frac{1}{\sqrt{N}})$. Let us formally state and prove this result in the following proposition.

**Proposition 1.** *Let $P_{\text{MAP}}$ be the block-MAP error for transmission of a linear code $\mathcal{C}$ with minimum distance $d_{\min}$ over the BEC($\epsilon$) (or the BSC($\epsilon$)). For an arbitrary $\delta \in (0, 1/2)$ let $\epsilon_1(\epsilon_2)$ be such that $P_{\text{MAP}}(\epsilon_1) = \delta$ ($P_{\text{MAP}}(\epsilon_2) = 1-\delta$). Then, there exists a constant $B(\delta) > 0$, independent of the choice of the code, such that $\epsilon_2 - \epsilon_1 \geq B(\delta)/\sqrt{N}$.*

*Proof:* The proof follows by the fact that the derivative of the product measure for a monotone property (e.g. $dP_{\text{MAP}}(\epsilon)/d\epsilon$) is at most $O(\sqrt{N})$ (see [4, Corollary 9.16]). ■

## III. Main Statement

**Theorem 2.** *Let $\mathcal{C}$ be a binary linear code of length $N$, dimension $K$, and with weight distribution $A(w)$. Consider transmission over the binary erasure channel with parameter $0 < \epsilon^* < 1$, where $\epsilon^*$ is such that*

$$P_{\text{MAP}}(\epsilon^*) = P\{\hat{x}(Y) \neq X\} = \frac{1}{2}.$$

*Then, for any $1 \leq W \leq N$ the following holds. For $\epsilon < \epsilon^*$:*

$$\begin{aligned} P_{\text{MAP}}(\epsilon) \leq &\Phi\left(\sqrt{W}(\sqrt{-\ln(\epsilon^*)} - \sqrt{-\ln(\epsilon)})\right) \\ &+ 2\sqrt{-\log\epsilon}\sqrt{W}A(W, \epsilon^*), \end{aligned}$$

*and for $\epsilon^* < \epsilon < 1$:*

$$\begin{aligned} P_{\text{MAP}}(\epsilon) \geq &\Phi\left(\sqrt{W}(\sqrt{-\ln(\epsilon^*)} - \sqrt{-\ln(\epsilon)})\right) \\ &- 2\sqrt{-\log\epsilon^*}\sqrt{W}A(W, \epsilon). \end{aligned}$$

**Theorem 3.** *Let $\mathcal{C}$ be a binary linear code of length $N$, dimension $K$, and with weight distribution $A(w)$. Consider transmission over the binary symmetric channel with parameter $0 < \epsilon^* < 1$, where $\epsilon^*$ is such that*

$$P_B(\epsilon^*) = P\{\hat{x}(Y) \neq X\} = \frac{1}{2}.$$

*Then, for any $1 \le W \le N$ following holds. For $\epsilon < \epsilon^*$:*

$$P_{\text{MAP}}(\epsilon) \le \Phi\left(\frac{\sqrt{W}}{2}(\sqrt{-\ln(1-\epsilon^*)} - \sqrt{-\ln(1-\epsilon)})\right) + 4\sqrt{-\log\epsilon}\sqrt{W}A(W,\epsilon^*),$$

*and for $\epsilon^* < \epsilon < \frac{1}{2}$:*

$$P_{\text{MAP}}(\epsilon) \ge \Phi\left(\frac{\sqrt{W}}{2}(\sqrt{-\ln(1-\epsilon^*)} - \sqrt{-\ln(1-\epsilon)})\right) - 4\sqrt{-\log\epsilon^*}\sqrt{W}A(W,\epsilon).$$

**Fast Transition for RM Codes.** One immediate implication of Theorems 2 and 3 is that the transition width of a code $\mathcal{C}$ is at most $O(1/\sqrt{W})$ provided that $\sqrt{W}A(W,z)$ is small (i.e. if $\sqrt{W}A(W,z)$ vanishes as the length grows). For RM codes, we use the following result from [3, Lemma 4] to conclude that the transition width is at most $\Theta(1/N^{\frac{1}{2}-\kappa})$, for any $\kappa > 0$.

**Lemma 1.** *For any $z \in [0,1)$ and any $\kappa > 0$ the following holds for RM codes. Let $W = N^{1-\kappa}$, then*

$$A(W,z) \le e^{-W\beta(\kappa,z)},$$

*where $\beta(\kappa,z)$ is strictly positive for any $z \in (0,1)$ and $\kappa > 0$.*

*Proof of Theorem 2:* Consider transmission over the BEC($\epsilon$) with a linear binary code $\mathcal{C}$ of blocklength $N$. Since the code is linear and the channel is symmetric we can assume without loss of generality (for MAP decoding) that we transmit the all-zero codeword. Given two sequences $x, y \in \{0,1\}^N$, we say that $y$ covers $x$ if the support of $x$ is included in the support of $y$, i.e., $x_i \le y_i$ for any $1 \le i \le N$. Recall that we assume that the block length is $N$. It is therefore natural to assign to the $N$ channel actions a binary $N$-tuple, henceforth called the *erasure pattern*, which has value 1 in its $i$-th position if and only if the $i$-th channel erases its input and 0 otherwise. In this way, the set of all the erasure patterns $\{0,1\}^N$ is endowed with the product measure as its corresponding probability measure. We use $\mu_\epsilon(\cdot)$ to denote such a probability measure, i.e., for an erasure pattern $\omega$ we have $\mu_\epsilon(\omega) = \epsilon^{w_H(\omega)}(1-\epsilon)^{N-w_H(\omega)}$. Furthermore, assuming the all-zero transmission, an erasure pattern $\omega \in \{0,1\}^N$ causes a block-MAP error if and only if there exists at least one non-zero codeword $x \in \mathcal{C}$ which is covered by $\omega$. We define $\Omega$ to be the set of erasure patterns which cause a block-error[1], i.e.,

$$\Omega = \{\omega \in \{0,1\}^N : \omega \text{ covers at least one non-zero codeword}\}. \quad (4)$$

As a result, we have

$$P_{\text{MAP}}(\epsilon) = \mu_\epsilon(\Omega). \quad (5)$$

Also, let us define the boundary of $\Omega$ to be

$$\partial\Omega = \{\omega \in \Omega : \exists \omega' \notin \Omega, d_H(\omega,\omega') = 1\}, \quad (6)$$

where $d_H$ denotes the Hamming distance. By definition, if $\omega \in \partial\Omega$ then $\omega$ covers at least one non-zero codeword. We argue from [1] that it covers in fact *exactly one* non-zero codeword, call this codeword $x$. This is true since if $\omega$ covers two distinct non-zero codewords, call them $x$ and $x'$, then by linearity of the code, it also covers the codeword $x'' = x + x'$. Now note that for every position $i$, $1 \le i \le N$, at least one of $x_i$, $x'_i$, and $x''_i$ must be 0 (since by construction each of these values is the XOR of the other two). Therefore, no matter what position of $\omega$ we set from erasure to non-erasure, at least one of these three codewords will still be covered. In other words, $\omega$ does not have a neighbour at distance 1 in $\Omega^c$, i.e., $\omega \notin \partial\Omega$.

Given the product measure on the erasure patterns, the Margulis-Russo formula expresses the derivate of $\mu_\epsilon(\Omega)$ in terms of the measure of the boundary of $\Omega$:

$$\frac{d\mu_\epsilon(\Omega)}{d\epsilon} = \frac{1}{\epsilon}\int_{\omega\in\Omega} h_\Omega(\omega)d\mu_\epsilon(\omega), \quad (7)$$

where

$$\begin{aligned} &h_\Omega(\omega) = 0, && \text{if } \omega \notin \Omega, \\ &h_\Omega(\omega) = \left|\{\omega' \notin \Omega : d_H(\omega,\omega') = 1\}\right|, && \text{if } \omega \in \Omega. \end{aligned} \quad (8)$$

Let us now see how the quantity $h_\Omega$ can be lower-bounded for boundary patterns $\omega \in \Omega$. Let $x$ be the unique non-zero codeword that is covered by the boundary point $\omega$. We write $x(\omega)$. We know that the weight of $x(\omega)$ is at least $d_{min}$, and every erasure pattern $\omega'$ which is equal to $\omega$ except at one position $i$ where $x_i = 1$ is an element of $\bar{\Omega}$ and $d_H(\omega,\omega') = 1$. Hence, $\omega$ has at least $d_{\min}$ neighbours in $\bar{\Omega}$ as claimed, or in other words $h_\Omega(\omega) \ge d_{\min}$ [1]. We will now strengthen this bound and show that for most boundary points $\omega$, $h_\Omega(\omega)$ is considerably larger.

Let us define the set $\Gamma_W \subseteq \partial\Omega$,

$$\Gamma_W \triangleq \{\omega \in \partial\Omega : w_H(x(\omega)) \ge W\}. \quad (9)$$

Note that

$$\forall \omega \in \Gamma_W : h_\Omega(\omega) \ge W. \quad (10)$$

We can then write

$$\int_{\omega\in\Omega} h_\Omega(\omega)d\mu_\epsilon(\omega) \ge \sqrt{W}\int_{\omega\in\Omega}\sqrt{h_\Omega}d\mu_\epsilon - \sqrt{W}\int_{\omega\in\Gamma_W^c}\sqrt{h_\Omega}d\mu_\epsilon, \quad (11)$$

where $\Gamma_W^c$ denotes the set complement of $\Gamma_W$. By [1, Theorem 2.1] for *monotone increasing sets* $\Omega$ we have

$$\int_{\omega\in\Omega}\sqrt{h_\Omega(\omega)}d\mu_\epsilon(\omega) \ge \frac{1}{\sqrt{-2\log\epsilon}}\gamma(\mu_\epsilon(\Omega)), \quad (12)$$

where $\gamma(x) = \phi(\Phi^{-1}(x))$, where $\phi$ and $\Phi$ are the pdf and the CDF of standard normal distribution. Let us now bound the right-most term in (11). We can write

$$\int_{\omega\in\Gamma_W^c}\sqrt{h_\Omega}d\mu_\epsilon = \sum_{w=1}^{W-1}\sqrt{w}\mu_\epsilon\left(\{\omega \in \partial\Omega : w_H(x(\omega)) = w\}\right).$$

Also,

$$\begin{aligned} &\mu_\epsilon(\{\omega \in \partial\Omega : w_H(x(\omega)) = w\}) \\ &= \mu_\epsilon(\{\omega \in \partial\Omega : \exists x \in \mathcal{C} \ni (\omega \succ x) \wedge (w_H(x) = w)\}) \\ &\le \mu_\epsilon(\{\omega : \exists x \in \mathcal{C} \ni (\omega \succ x) \wedge (w(x) = w)\}) \\ &= \mu_\epsilon(\bigcup_{x\in\mathcal{C}:w(x)=w}\{\omega : \omega \succ x\}) \le \sum_{x\in\mathcal{C}:w(x)=w}\mu_\epsilon(\{\omega : \omega \succ x\}). \end{aligned}$$

[1] We assume here that block error happens if we have at least one none-zero covered codeword. Another version of the MAP decoder chooses one of the covered codewords uniformly at random (i.e. it breaks ties randomly). All the results in this paper are valid if we use the other version of the MAP decoder.

When the channel is a BEC($\epsilon$) the last step of the above expression can be bounded by $A(w)\epsilon^w$ and thus we obtain

$$\int_{\omega\in\Gamma_W^c}\sqrt{h_\Omega}d\mu_\epsilon \le \sum_{w=1}^{W}\sqrt{w}A(w)\epsilon^w. \quad (13)$$

Now, by using (11), (12), and (13) we obtain

$$\begin{aligned}\int_{\omega\in\Omega} & h_\Omega(\omega)d\mu_\epsilon(\omega) \\ &\ge \sqrt{W}\left(\frac{1}{\sqrt{-2\log\epsilon}}\gamma(\mu_\epsilon(\Omega)) - \sum_{w=1}^{W}\sqrt{w}A(w)\epsilon^w\right) \\ &\ge \sqrt{W}\left(\frac{1}{\sqrt{-2\log\epsilon}}\gamma(\mu_\epsilon(\Omega)) - \sqrt{W}A(W,\epsilon)\right).\end{aligned} \quad (14)$$

Combining (7), and (14), we obtain that for any $1 < W < N$

$$\frac{d\mu_\epsilon(\Omega)}{d\epsilon} \ge \frac{\sqrt{W}}{\epsilon}\left(\frac{1}{\sqrt{-2\log\epsilon}}\gamma(\mu_\epsilon(\Omega)) - \sqrt{W}A(W,\epsilon)\right). \quad (15)$$

Now, consider a channel parameter $\epsilon > \epsilon^*$. We have

$$\begin{aligned}\mu_\epsilon(\Omega) - \mu_{\epsilon^*}(\Omega) &= \int_{\epsilon^*}^{\epsilon}\frac{d\mu_{\bar\epsilon}(\Omega)}{d\bar\epsilon}d\bar\epsilon \\ &\ge \sqrt{W}\int_{\epsilon^*}^{\epsilon}\frac{1}{\bar\epsilon}\left(\frac{\gamma(\mu_{\bar\epsilon}(\Omega))}{\sqrt{-2\log\bar\epsilon}} - \sqrt{W}A(W,\bar\epsilon)\right)d\bar\epsilon\end{aligned} \quad (16)$$

Define $c(x) = \sqrt{-\log x}$. For computing the above integral, we consider two cases: (i) If $\sqrt{W}A(W,\epsilon) \ge \frac{1}{2}\gamma(\mu_\epsilon(\Omega))/c(\epsilon^*)$, then by using the inequality $\gamma(x) \ge x(1-x)$ we obtain that $\mu_\epsilon(\Omega) \ge 1 - 2c(\epsilon^*)\sqrt{W}A(W,\epsilon)$. Hence the result of the theorem holds for this case. (ii) If $\sqrt{W}A(W,\epsilon) < \frac{1}{2}\gamma(\mu_\epsilon(\Omega))/c(\epsilon^*)$, then as $A(W,\bar\epsilon)$ is an increasing function in $\bar\epsilon$ and $\gamma(x)$ is concave and symmetric around $x = 1/2$, then for any $\bar\epsilon \in [\epsilon^*, \epsilon]$ we have $\sqrt{W}A(W,\bar\epsilon) \le 1/2\gamma(\mu_{\bar\epsilon}(\Omega))/c(\bar\epsilon)$. As a result, the quantity inside the integral in (16) will be lower bounded by $\frac{1}{2\bar\epsilon}\gamma(\mu_{\bar\epsilon}(\Omega))/\sqrt{-2\log(\bar\epsilon)}$. Now, by integrating this new lower bound we obtain the result of the Theorem (for more details see [1]).

The result of the Theorem for $\epsilon < \epsilon^*$ follows similarly.

*Proof of Theorem 3:* Consider now transmission over the BSC($\epsilon$) with a linear binary code $\mathcal{C}$ of blocklength $N$. Similar to the proof of Theorem 2, we can assume the all-zero transmission. Also, we can naturally map the set of $N$ channel usages to an *error pattern* $\omega \in \{0,1\}^n$, where a 1 at position $i$ means that the $i$-th channel has flipped its input. In this way, the set of error patterns is endowed with the product measure, i.e., i.e. for an error pattern $\omega$ we have $\mu_\epsilon(\omega) = \epsilon^{w_H(\omega)}(1-\epsilon)^{N-w_H(\omega)}$. . We let $\Omega$ to be the set of error patterns which cause a block-error, i.e.,

$$\Omega = \{\omega \in \{0,1\}^N : \exists x \in \mathcal{C} : x \ne 0, w_H(\omega + x) < w_H(\omega)\}. \quad (17)$$

In this regard, we have $P_{\text{MAP}} = \mu_\epsilon(\Omega)$. Also, let us define the boundary of $\Omega$ to be

$$\partial\Omega = \{\omega \in \Omega : \exists\omega' \notin \Omega, d_H(\omega,\omega') = 1\}. \quad (18)$$

The Margulis-Russo formula (7) expresses the derivate of $\mu_\epsilon(\Omega)$ in terms of the function $h_\Omega$ (defined in (8)) over the boundary of $\Omega$. Now consider an error pattern $\omega$ in the boundary, i.e., $\omega \in \partial\Omega$. Then there exists at least one codeword, call it $x(\omega)$, for which $w_H(\omega + x) < w_H(\omega)$. From [1], we know that

$$h_\Omega(\omega) \ge \frac{w_H(x(\omega))}{2}. \quad (19)$$

Hence, considering the set $\Gamma_W$ as in (9), we have

$$\forall\omega \in \Gamma_W : h_\Omega(\omega) \ge \frac{W}{2}. \quad (20)$$

We can now use the similar steps as for the derivation of (15) to show that for the case of the BSC we have

$$\frac{d\mu_\epsilon(\Omega)}{d\epsilon} \ge \sqrt{W}\left(\frac{1}{2\sqrt{-2\log\epsilon}}\gamma(\mu_\epsilon(\Omega)) - 2\sqrt{W}A(W,\epsilon)\right). \quad (21)$$

The rest of the proof now follows similarly to the case of the BEC.

## IV. Estimating EXIT derivative via the Blowing-Up Lemma

As above, we consider a linear code $\mathcal{C}$, $X \sim \text{Uniform}(\mathcal{C})$ and denote by $Y^\epsilon$ be the result of passing $X$ through a BEC($\epsilon$), $0 \le \epsilon \le 1$. For $\{i,j\} \in [n]$, define

$$\begin{aligned}\mathcal{C}_{00}^{ij} &\triangleq \{c \in \mathcal{C} \ : \ (c_i, c_j) = (0,0)\}, \\ \mathcal{C}_{01}^{ij} &\triangleq \{c \in \mathcal{C} \ : \ (c_i, c_j) = (0,1)\}, \\ \mathcal{C}_{10}^{ij} &\triangleq \{c \in \mathcal{C} \ : \ (c_i, c_j) = (1,0)\}, \\ \mathcal{C}_{11}^{ij} &\triangleq \{c \in \mathcal{C} \ : \ (c_i, c_j) = (1,1)\},.\end{aligned}$$

For a vector $z \in \{0,1\}^n$ and a code $\mathcal{A} \subset \{0,1\}^n$ we define

$$\mathcal{A}(z) \triangleq \{a \in \mathcal{A} \ : \ z \succ a\}. \quad (22)$$

We now define the following partition of $\{0,1\}^n$ w.r.t. the codebook $\mathcal{C}$ and the coordinates $\{i,j\}$:

$$\begin{aligned}\mathcal{B}_1^{ij} &\triangleq \{z \in \{0,1\}^n \ : \ \mathcal{C}_{01}^{ij}(z) = \mathcal{C}_{10}^{ij}(z) = \mathcal{C}_{11}^{ij}(z) = \emptyset\}, \\ \mathcal{B}_2^{ij} &\triangleq \{z \in \{0,1\}^n \ : \ \mathcal{C}_{01}^{ij}(z) \ne \emptyset, \mathcal{C}_{10}^{ij}(z) = \mathcal{C}_{11}^{ij}(z) = \emptyset\}, \\ \mathcal{B}_3^{ij} &\triangleq \{z \in \{0,1\}^n \ : \ \mathcal{C}_{10}^{ij}(z) \ne \emptyset, \mathcal{C}_{01}^{ij}(z) = \mathcal{C}_{11}^{ij}(z) = \emptyset\}, \\ \mathcal{B}_4^{ij} &\triangleq \{z \in \{0,1\}^n \ : \ \mathcal{C}_{11}^{ij}(z) \ne \emptyset, \mathcal{C}_{01}^{ij}(z) = \mathcal{C}_{10}^{ij}(z) = \emptyset\}, \\ \mathcal{B}_5^{ij} &\triangleq \{z \in \{0,1\}^n \ : \ \mathcal{C}_{01}^{ij}(z) \ne \emptyset, \mathcal{C}_{10}^{ij}(z) \ne \emptyset, \mathcal{C}_{11}^{ij}(z) \ne \emptyset\}.\end{aligned}$$

Note that indeed $\cup_{k=1}^5 \mathcal{B}_k^{ij} = \{0,1\}^n$ due to the linearity of the code. To see this, note that if $z$ covers $c_1 \in \mathcal{C}_{01}^{ij}$ and also $c_2 \in \mathcal{C}_{10}^{ij}$, then it must also cover $c_3 = c_1 + c_2 \in \mathcal{C}_{11}^{ij}$, since $\text{supp}(c_1 + c_2) \subset (\text{supp}(c_1) \cup \text{supp}(c_2))$. Using the same reasoning, and recalling that $\mathbf{0} \in \mathcal{C}_{00}^{ij}$ is covered by all $z \in \{0,1\}^n$, we see that each $z \in \{0,1\}^n$ can either 1)cover only codewords from $\mathcal{C}_{00}^{ij}$; 2)cover codewords from $\mathcal{C}_{00}^{ij}$ and one of the codebooks $\mathcal{C}_{01}^{ij}$, $\mathcal{C}_{10}^{ij}$, or $\mathcal{C}_{11}^{ij}$; 3)cover codewords from all four codebooks $\mathcal{C}_{00}^{ij}$, $\mathcal{C}_{01}^{ij}$,$\mathcal{C}_{10}^{ij}$, $\mathcal{C}_{11}^{ij}$.

Define the $n$-dimensional random vector $Z^\epsilon = Z^\epsilon(ij)$ such that $Z_i^\epsilon = Z_j^\epsilon = 1$, and $Z_k^\epsilon \sim \text{Bernoulli}(\epsilon)$ i.i.d, for $k \in [n] \setminus \{i,j\}$, and define the quantity

$$\alpha_{ij}^\epsilon \triangleq \Pr\left(Z^\epsilon \in \mathcal{B}_4^{ij} | Z^\epsilon \in \mathcal{B}_2^{ij} \cup \mathcal{B}_3^{ij} \cup \mathcal{B}_4^{ij}\right). \quad (23)$$

In the sequel, for a vector $x \in \mathcal{X}^n$ and a subset $A \subset [n]$, we denote $x_{\sim A} = x_{[n]\setminus A}$.

**Theorem 4.** *Assume $I(X_i; Y^\epsilon_{\sim i}) \in (\delta, 1-\delta)$ for some $0 < \delta < 1/2$. Then*

$$I(X_i; X_j | Y^\epsilon_{\sim i,j}) \geq \frac{c(\delta)}{\sqrt{n}} \alpha^\epsilon_{ij}, \tag{24}$$

*where $c(\delta)$ is positive if $\delta$ is bounded away from 0.*

Before proving Theorem 4, let us demonstrate its implication. Define the EXIT function

$$g(\epsilon) = -\frac{1}{n}\frac{d}{d\epsilon} I(X; Y^\epsilon) = \frac{1}{n}\sum_{i=1}^n H(X_i|Y^\epsilon_{\sim i}). \tag{25}$$

We can further compute

$$\begin{aligned} g'(\epsilon) &= \frac{d}{d\epsilon} g(\epsilon) \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \frac{\partial}{\partial \epsilon_j} H(X_i|Y^{\epsilon_j}_j, Y^\epsilon_{\sim i,j})\Big|_{\epsilon_j=\epsilon} \\ &= -\frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} \frac{\partial}{\partial \epsilon_j} I(X_i, Y^{\epsilon_j}_j | Y^\epsilon_{\sim i,j})\Big|_{\epsilon_j=\epsilon} \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} I(X_i, X_j | Y^\epsilon_{\sim i,j}). \end{aligned} \tag{26}$$

**Theorem 5.** *Let $\mathcal{C}$ be a 1-transitive code. Then if $g(\epsilon) \in (\delta, 1-\delta)$ for some $0 < \delta \leq 1/2$, then*

$$g'(\epsilon) \geq \frac{c(\delta)}{n^{3/2}} \sum_{i=1}^n \sum_{j\neq i} \alpha^\epsilon_{ij}, \tag{27}$$

*where $c(\delta)$ is positive if $\delta$ is bounded away from 0.*

*Proof:* For transitive codes $g(\epsilon) = 1 - I(X_i; Y_{\sim i})$ for all $i \in [n]$. Combining (26) with Theorem 4, gives the result. ∎

*Proof of Theorem 4:* For any $y \in \{0, 1, ?\}^{n-2}$ set

$$\begin{aligned} \alpha_A(y) &\triangleq \Pr(X_i = 0, X_j = 0 | Y^\epsilon_{\sim i,j} = y) \\ \alpha_B(y) &\triangleq \Pr(X_i = 0, X_j = 1 | Y^\epsilon_{\sim i,j} = y) \\ \alpha_C(y) &\triangleq \Pr(X_i = 1, X_j = 0 | Y^\epsilon_{\sim i,j} = y) \\ \alpha_D(y) &\triangleq \Pr(X_i = 1, X_j = 1 | Y^\epsilon_{\sim i,j} = y), \end{aligned}$$

and $P(y) \triangleq [\alpha_A(y)\ \alpha_B(y)\ \alpha_C(y)\ \alpha_D(y)]$. Let $S = \{k_1, \ldots, k_{|S|}\} \subset [n]\setminus\{i,j\}$ be the locations of non-erased bits within $y$, and let $x^*$ be a codeword in $\mathcal{C}$ for which $x^*_S = y_S$ (such a codeword must always exist). Let $\bar{S} \triangleq [n] \setminus S$ be the erased bits, and note that in vector representation $\bar{S}$ is a random vector with distribution $Z^\epsilon$. Thus, given $y$, we have that the transmitted codeword $x$ is uniformly distributed on $x^* + \mathcal{C}(z)$, where $z$ is the vector representation of $\bar{S}$. Without loss of generality, we may assume that $x^*_i = x^*_j = 0$. Thus, since $\mathcal{C}(z)$ is a subspace for any $z$, we have that

$$P(y) = P(z) = \begin{cases} P_1 = [1\ 0\ 0\ 0] & z \in \mathcal{B}^{ij}_1 \\ P_2 = \left[\frac{1}{2}\ \frac{1}{2}\ 0\ 0\right] & z \in \mathcal{B}^{ij}_2 \\ P_3 = \left[\frac{1}{2}\ 0\ \frac{1}{2}\ 0\right] & z \in \mathcal{B}^{ij}_3 \\ P_4 = \left[\frac{1}{2}\ 0\ 0\ \frac{1}{2}\right] & z \in \mathcal{B}^{ij}_4 \\ P_5 = \left[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right] & z \in \mathcal{B}^{ij}_5 \end{cases} \tag{28}$$

Note that $I(X_i; X_j | Y^\epsilon_{\sim i,j} = y) = 1$ if $P(y) = P(z) = P_4$ and $I(X_i; X_j | Y^\epsilon_{\sim i,j} = y) = 0$ otherwise. Thus, defining $Q_k = \Pr(Z^\epsilon \in \mathcal{B}^{ij}_k)$, $k \in [5]$, we have

$$I(X_i; X_j | Y^\epsilon_{\sim i,j}) = Q_4. \tag{29}$$

By inspection of the 5 different possibilities for $P(y)$, we observe that

$$H(X_i|Y^\epsilon_{\sim i}) = Q_1 \cdot 0 + Q_2 \cdot 0 + Q_3 \cdot 1 + Q_4 \cdot \epsilon + Q_5 \cdot 1.$$

Consequently,

$$\begin{aligned} I(X_i; Y^\epsilon_{\sim i}) &= 1 - (Q_3 + \epsilon \cdot Q_4 + Q_5) \\ &= Q_1 + Q_2 + (1-\epsilon) \cdot Q_4, \end{aligned}$$

and by the theorem's assumption, we therefore have that

$$Q_1 + Q_2 + (1-\epsilon) \cdot Q_4 \in (\delta, 1-\delta). \tag{30}$$

We proceed by using the Blowing Up Lemma (see e.g. [5, Theorem 5.3]) to show that (30) implies that $\eta \triangleq Q_2 + Q_3 + Q_4 \geq \frac{c(\delta)}{\sqrt{n}}$ for some constant $c(\delta)$.

Define the set $\Omega^{ij} = \mathcal{B}^{ij}_2 \cup \mathcal{B}^{ij}_3 \cup \mathcal{B}^{ij}_4 \cup \mathcal{B}^{ij}_5$ and let $\partial\Omega^{ij}$ be its boundary. Further, let $\mathcal{D}^{ij} = \mathcal{B}^{ij}_2 \cup \mathcal{B}^{ij}_3 \cup \mathcal{B}^{ij}_4$. The crucial observation is that $\partial\Omega^{ij} \subset D^{ij}$, as erasure of a single additional coordinate, which corresponds to changing the Hamming weight of the erasure pattern by 1, can increase the conditional entropy of $(X_i, X_j)$ by at most one bit. Applying the blowing-up lemma (see e.g. [5, Theorem 5.3]), we therefore have that

$$\begin{aligned} \eta = \Pr(Z^\epsilon \in \mathcal{D}^{ij}) &\geq \Pr(Z^\epsilon \in \partial\Omega^{ij}) \\ &\geq \frac{a}{\sqrt{n}} \gamma\left(\Pr(Z^\epsilon \in \Omega^{ij})\right), \end{aligned} \tag{31}$$

where $a = a(\epsilon)$ is a positive constant and $\gamma$ is as defined after (12). Invoking (30), we see that either $Q_2 + Q_3 + Q_4 > \delta/2$ or $\Pr(Z^\epsilon \in \Omega^{ij}) = 1 - Q_1 \in (\delta/2, 1-\delta/2)$. Thus, $\eta \geq c(\delta)/\sqrt{n}$ where $c(\delta) = a\gamma(\delta/2)$. The result now follows as $Q_4 = \alpha^\epsilon_{ij} \cdot \eta$.


## Acknowledgement

The research of H. Hassani was supported by NSF-CRII award CCF-1755707. The work of O. Ordentlich was supported by ISF under Grant 1791/17. The research of Y. Polyanskiy was supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, NSF CAREER award CCF-12-53205, and NSF grant CCF-17-17842. The work of R. Urbanke was supported by SNSF grant No. 200021–166106.


∎

## References


[1] J.-P. Tillich and G. Zémor, "Discrete isoperimetric inequalities and the probability of a decoding error," *Combinatorics, Probability and Computing*, vol. 9, no. 5, pp. 465–479, 2000.

[2] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Sasoglu, and R. Urbanke, "Reed-Muller codes achieve capacity on erasure channels," *IEEE Transactions on Information Theory*, vol. 63, no. 6, 2017.

[3] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbanke, "Comparing the bit-MAP and block-MAP decoding thresholds of Reed-Muller codes on BMS channels," in *Information Theory (ISIT), 2016 IEEE International Symposium on*, 2016, pp. 1755–1759.

[4] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[5] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.