Zero-rate Reliability Function for Mismatched Decoding

Marco Bondaschi
EPFL
marco.bondaschi@epfl.ch

Albert Guillén i Fàbregas
ICREA and Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

Marco Dalai
University of Brescia
marco.dalai@unibs.it

***Abstract*—We derive an upper bound on the reliability function of mismatched decoding for zero-rate codes. The bound is based on a result by Komlós that shows the existence of a subcode with certain symmetry properties. The bound is shown to coincide with the expurgated exponent at rate zero for a broad family of channel and decoding metric pairs.**

A full version of this paper is accessible at: https://arxiv.org/pdf/2101.10238.pdf

I. Introduction

Consider a discrete memoryless channel with finite input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, and with transition probabilities $W(y|x)$. For a message set $\mathcal{M} = \{1, 2, \dots, M\}$ and blocklength n, an encoder is a function $\mathcal{C} : \mathcal{M} \to \mathcal{X}^n$ that assigns to each message m a corresponding codeword $\boldsymbol{x}_m = (x_{m,1}, x_{m,2}, \dots, x_{m,n})$. The rate of transmission is defined as $R \triangleq \frac{\log M}{n}$. When message m is sent, an output sequence $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ is received with probability $W^n(\boldsymbol{y}|\boldsymbol{x}_m) = \prod_{i=1}^n W(y_i|x_{m,i})$. A decoder is a function $\mathcal{C}^{-1} : \mathcal{Y}^n \to \mathcal{M}$ which maps each output sequence to a message in $\mathcal{M}$ as

$$\mathcal{C}^{-1}(\boldsymbol{y}) \in \arg\max_{m \in \mathcal{M}} q^n(\boldsymbol{x}_m, \boldsymbol{y}) \tag{1}$$

where $q^n(\boldsymbol{x}_m, \boldsymbol{y}) = \prod_{i=1}^n q(x_{m,i}, y_i)$ where $q : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ is a *decoding metric*. Ties are broken uniformly at random.

When message m is sent, a decoding error occurs if $\mathcal{C}^{-1}(\boldsymbol{y}) \neq m$, that is if $\boldsymbol{y}$ is not in the subset $\mathsf{Y}_m \subset \mathcal{Y}^n$ of output sequences that are decoded in m. Letting $P_{e,m}$ be the probabilty of this event, the average probability of error of the code is $P_e \triangleq \frac{1}{M} \sum_{m=1}^M P_{e,m}$.

When $q(x, y) = W(y|x)$, the decoder is the maximum likelihood decoder, achieving the lowest probability of error. Instead, when the decoding metric $q(x, y) \neq W(y|x)$, the decoder is, in general, said to be mismatched [1], [2] (see also [3] for a recent survey on the subject).

For fixed R, n and decoding metric q, let $P_e^q(R, n)$ be the smallest probability of error over all codes with rate at least R and block length n, with decoding metric q. The reliability function is defined as

$$E^q(R) \triangleq \limsup_{n \to \infty} -\frac{1}{n} \log P_e^q(R, n) \tag{2}$$

This work was initiated while the first author was visiting UPF in Feb.-Apr. 2020. This work was supported in part by the European Research Council under Grant 725411.

is the best possible asymptotic exponential rate of decay of the error probability for a given channel and decoding metric, for codes of rate at least R and blocklength n. The supremum of the information rates R for which the error probability tends to zero is called *mismatched capacity*.

There is no single-letter expression for the mismatched capacity. Random-coding rates and error exponents are available in the literature [1], [2], [4]–[6]. There are fewer upper bounds on the mismatched capacity in the literature. Recently, single-letter upper bounds on the mismatched capacity improving over the Shannon capacity were proposed in [7], [8]. A sphere-packing upper bound on the mismatched reliability function was recently proved in [9].

We study the problem of finding an upper bound on the mismatched reliability function of any given discrete memoryless channel and decoding metric, when the rate tends to 0, that is, we are interested in upper-bounding $E^q(0^+)$. We restrict our attention to channel and decoding metric pairs such that

$$W(y|x) > 0 \implies q(x, y) > 0 \tag{3}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. In fact, whenever the decoding metric does not meet this condition for some input x, the mismatched capacity is zero if that input is used [1].

Relevant for this work is a generalization of Gallager's expurgated bound to mismatched decoding [10]; when the rate approaches zero, the bound takes the form

$$E(0^+) \geq \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} -\sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{X}} Q(a)Q(b) \log \sum_{y \in \mathcal{Y}} W(y|a) \left(\frac{q(b, y)}{q(a, y)} \right)^s. \tag{4}$$

In this paper, we derive an upper bound on $E(0^+)$ for a wide class of channel and metrics pairs that matches the above lower bound, showing that it is tight for such pairs.

II. Mismatched zero-error capacity

We restrict our attention to channels and decoding metrics that give a strictly positive probability of error at any $R > 0$; if this is not the case, $E^q(R = 0^+)$ is infinite and no finite upper bound is possible. We thus need the mismatched zero-error capacity $\bar{C}_0^q = 0$, where $\bar{C}_0^q$ is defined as the supremum of the rates R for which there exist codes with probability of error

exactly equal to zero, given that ties are broken uniformly at random, for a channel $W(y|x)$ and a decoding metric $q(x,y)$[1].

Notice that $\bar{C}_0^q$ is positive if and only if there exist two codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (of arbitrary blocklength) such that for all output sequences $\boldsymbol{y}$:

1) either $W^n(\boldsymbol{y}|\boldsymbol{x}_1) = 0$ or $W^n(\boldsymbol{y}|\boldsymbol{x}_2) = 0$;
2) $W^n(\boldsymbol{y}|\boldsymbol{x}_1) > 0 \implies q(\boldsymbol{x}_1, \boldsymbol{y}) > q(\boldsymbol{x}_2, \boldsymbol{y})$
   $W^n(\boldsymbol{y}|\boldsymbol{x}_2) > 0 \implies q(\boldsymbol{x}_2, \boldsymbol{y}) > q(\boldsymbol{x}_1, \boldsymbol{y})$.

Condition 1 states that each possible output sequence can be obtained only from one of the two codewords; condition 2 states that each sequence is always decoded correctly.

The study of the conditions which ensure $\bar{C}_0^q = 0$ can be done using the same tools that we later need for the study of $E(0^+)$. Hence, we introduce a real-valued function that will be useful to both ends. For any two sequences $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}^n$, we define, for $s \geq 0$,

$$\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \triangleq -\log \sum_{\boldsymbol{y} \in \hat{\mathcal{Y}}^n_{\boldsymbol{x}_1,\boldsymbol{x}_2}} W^n(\boldsymbol{y}|\boldsymbol{x}_1) \left(\frac{q^n(\boldsymbol{x}_2, \boldsymbol{y})}{q^n(\boldsymbol{x}_1, \boldsymbol{y})} \right)^s \quad (5)$$

where $\hat{\mathcal{Y}}^n_{\boldsymbol{x}_1,\boldsymbol{x}_2} \triangleq \left\{ \boldsymbol{y} \in \mathcal{Y}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) q^n(\boldsymbol{x}_2, \boldsymbol{y}) > 0 \right\}$. When $n = 1$, (5) becomes, for any $a, b \in \mathcal{X}$,

$$\mu_{a,b}(s) \triangleq -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s \quad (6)$$

with $\hat{\mathcal{Y}}_{a,b} \triangleq \left\{ y \in \mathcal{Y} : q(a,y) q(b,y) > 0 \right\}$. We will also use the limit of the derivatives of these functions when $s \to \infty$, and we thus define

$$\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} \triangleq \lim_{s \to \infty} \mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \quad (7)$$

$$\mu'_{a,b} \triangleq \lim_{s \to \infty} \mu'_{a,b}(s) \quad (8)$$

setting by definition $\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} = +\infty$ if $\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = +\infty$, and the same for $\mu'_{a,b}$.[2] It is easily seen that $\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ is additive, in the sense that letting $P_{\boldsymbol{x}_1,\boldsymbol{x}_2}$ denote the joint type of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we have

$$\frac{1}{n} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{X}} P_{\boldsymbol{x}_1,\boldsymbol{x}_2}(a,b) \mu_{a,b}(s) \quad (9)$$

So, important properties of $\mu_{a,b}(s)$ are thus immediately inherited by $\frac{1}{n} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$. In particular, $\mu_{a,b}(s)$ is concave in s and, due to assumption (3), it satisfies $\mu_{a,a} = 0$. Furthermore, it can be proved that

$$\mu'_{a,b} = \min_{y: W(y|a) > 0} \log \frac{q(a,y)}{q(b,y)} \quad (10)$$

and that

$$\lim_{s \to +\infty} \mu_{a,b}(s) = +\infty \iff \mu'_{a,b} > 0$$
$$\lim_{s \to +\infty} \mu_{a,b}(s) \in [0, +\infty) \iff \mu'_{a,b} = 0$$
$$\lim_{s \to +\infty} \mu_{a,b}(s) = -\infty \iff \mu'_{a,b} < 0.$$

[1]More discussion on the zero-error capacity for mismatched decoding with different ways of handling ties is given in the extended version of this paper.

[2]Throughout the paper we use the convention $\frac{\cdot}{0} = +\infty$.

We are now ready to state the following theorem on $\bar{C}_0^q$.

Theorem 1: For any discrete memoryless channel $W(y|x)$ and decoding metric $q(x,y)$, $\bar{C}_0^q = 0$ if and only if

$$\min_{y: W(y|a) > 0} \frac{q(a,y)}{q(b,y)} \leq \max_{y: W(y|b) > 0} \frac{q(a,y)}{q(b,y)} \qquad \forall\, a, b \in \mathcal{X}. \quad (11)$$

Corollary 1:

$$\bar{C}_0^q = 0 \implies \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_a \sum_b Q(a) Q(b) \mu_{a,b}(s) < +\infty. \quad (12)$$

Proof: Consider the properties that a pair of codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ must have in order to satisfy Conditions 1 and 2 above for a positive $\bar{C}_0^q$. Condition 1 is satisfied if and only if in at least a coordinate of the pair of codewords, there is a pair of input symbols (a,b) such that $W(y|a)W(y|b) = 0$ for all y, that is, the joint type $P_{\boldsymbol{x}_1,\boldsymbol{x}_2}$ of the two codewords must have $P_{\boldsymbol{x}_1,\boldsymbol{x}_2}(a,b) > 0$ for that pair of input symbols. This condition can be satisfied only if there actually exists a pair of symbols (a,b) such that $W(y|a)W(y|b) = 0$ for all y. Thus, a precondition for $\bar{C}_0^q > 0$ is that

$$\mathcal{A} \triangleq \left\{ (a,b) \in \mathcal{X}^2 : W(y|a)W(y|b) = 0 \text{ for all } y \in \mathcal{Y} \right\} \neq \varnothing. \quad (13)$$

Notice that this is also the condition for the classical C_0 to be positive, which is of course a necessary condition to have $\bar{C}_0^q > 0$, since clearly $\bar{C}_0^q \leq C_0$.

Instead, from (10), Condition 2 is satisfied if and only if

$$\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} > 0 \qquad \text{and} \qquad \mu'_{\boldsymbol{x}_2,\boldsymbol{x}_1} > 0. \quad (14)$$

Hence, using (9), there exists a pair of codewords satisfying Condition 2 if and only if there exists a joint type P such that both

$$\sum_a \sum_b P(a,b) \mu'_{a,b} > 0 \text{ and } \sum_a \sum_b P(a,b) \mu'_{b,a} > 0. \quad (15)$$

Any pair of codewords with a joint type P satisfying (15) satisfies Condition 2 for a positive $\bar{C}_0^q$. Now, condition (15) is possible for some n if and only if

$$\sup_{P \in \mathcal{P}(\mathcal{X}^2)} \min \left\{ \sum_a \sum_b P(a,b) \mu'_{a,b}, \sum_a \sum_b P(a,b) \mu'_{b,a} \right\} > 0 \quad (16)$$

This supremum can be computed easily. Notice first that the minimum of two linear functions is concave. Then, since the minimum of the two functions is invariant with respect to the transformation $P(a,b) \leftrightarrow P(b,a)$, its maximum is always attained (also) by a joint type such that $P(a,b) = P(b,a)$ for all a, b. In such a case, the two functions are both equal to

$$\sum_{a \leq b} P(a,b) (\mu'_{a,b} + \mu'_{b,a}) \quad (17)$$

and this quantity is maximized when all the weight is given to the largest term. Notice also that the P achieving this maximum has rational entries. Hence, thanks to (10), condition (16) becomes

$$\max_{a,b} \left(\min_{y: W(y|a) > 0} \log \frac{q(a,y)}{q(b,y)} + \min_{y: W(y|b) > 0} \log \frac{q(b,y)}{q(a,y)} \right) > 0. \quad (18)$$

Thus, if (18) is true, then we can find at least one joint type P that satisfies (15), and with it a whole set of pairs of codewords that satisfy Condition 2 for having $\bar{C}_0^q > 0$. However, we have no guarantees that there exists a pair of codewords in this set that satisfies also Condition 1. For this to be true, it is necessary that a pair of codewords in the set has a joint type with $P(a,b) > 0$ for some $(a,b) \in \mathcal{A}$. We now consider this issue. Since the maximum in (18) is strictly positive, then, thanks to the fact that the argument of the $\max$ in (17) is linear in P, in the neighborhood of the joint distribution achieving the maximum, there exists a (symmetric) joint type $\hat{P}$ that has $\hat{P}(a,b) > 0$ for a pair of symbols $(a,b) \in \mathcal{A}$, and that, when put into (17), still returns a positive value. Hence, the two codewords with that joint type satisfy both Conditions 1 and 2, and $\bar{C}_0^q$ is positive.

Finally, the corollary follows from the fact that

$$\begin{aligned}\max_{Q\in\mathcal{P}(\mathcal{X})} \sup_{s\geq 0} &\sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) \\ &= \frac{1}{2} \max_{Q\in\mathcal{P}(\mathcal{X})} \sup_{s\geq 0} \sum_a \sum_b Q(a)Q(b)\big(\mu_{a,b}(s)+\mu_{b,a}(s)\big) \\ &\leq \frac{1}{2} \max_{Q\in\mathcal{P}(\mathcal{X})} \sum_a \sum_b Q(a)Q(b) \sup_{s\geq 0} \big(\mu_{a,b}(s)+\mu_{b,a}(s)\big)\end{aligned} \tag{19}$$

where the equality follows from the fact that

$$\sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) = \sum_a \sum_b Q(a)Q(b)\mu_{b,a}(s).$$

The quantity in (19) is finite if $\bar{C}_0^q = 0$, since inequality (11) can be rewritten as $\mu'_{a,b} + \mu'_{b,a} \leq 0$, which is equivalent to

$$\lim_{s\to+\infty} \big(\mu_{a,b}(s)+\mu_{b,a}(s)\big) < +\infty \tag{20}$$

which in turn implies that

$$\sup_{s\geq 0} \big(\mu_{a,b}(s)+\mu_{b,a}(s)\big) < +\infty \tag{21}$$

since $\mu_{a,b}(s) + \mu_{b,a}(s)$ is concave. ■

III. Lower bound on the probability of error

Under the assumption that $\bar{C}_0^q = 0$ and that ties are resolved equiprobably, we now derive a lower bound on the probability of error of codes with two codewords, and then we generalize the result to codes with an arbitrary number of codewords.

The following derivation depends heavily on the method of types developed by Csiszár and Körner [11]. Let $P_{\boldsymbol{x}_1,\boldsymbol{x}_2}$ be the joint type of codewords $\boldsymbol{x}_1, \boldsymbol{x}_2$. We define the conditional type of an output sequence $\boldsymbol{y} \in \mathcal{Y}^n$ as

$$V_{\boldsymbol{y}}(y|a,b) \triangleq \frac{P_{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}}(a,b,y)}{P_{\boldsymbol{x}_1,\boldsymbol{x}_2}(a,b)} \tag{22}$$

for any $a, b \in \mathcal{X}$ and $y \in \mathcal{Y}$, and we denote by $\mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2)$ the set of all conditional types given the codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Since the channel is memoryless, output sequences with the same conditional type $V_{\boldsymbol{y}}$ have the same conditional probability $W^n(\boldsymbol{y}|\boldsymbol{x}_i)$ given any of the two codewords $i = 1, 2$, and also the same value of the decoding metric $q^n(\boldsymbol{x}_i, \boldsymbol{y})$. Hence, sequences with the same conditional type are decoded in the same way. Therefore, we can group them together according to their conditional type $V_{\boldsymbol{y}}$.

We lower bound the probability of error for the code composed of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and decoding metric $q(x,y)$ as

$$P_e \geq \frac{1}{4}\Bigg(\sum_{V\in\mathcal{V}_2^n} W^n(V|\boldsymbol{x}_1) + \sum_{V\in\mathcal{V}_1^n} W^n(V|\boldsymbol{x}_2)\Bigg) \tag{23}$$

where $\mathcal{V}_i^n$ is the set of conditional types decoded to message i, or tied between i and the other message. Using standard properties of types, we can bound the first sum in (23) as

$$\sum_{V\in\mathcal{V}_2^n} W^n(V|\boldsymbol{x}_1) \geq \exp\Big\{-n \min_{V\in\mathcal{V}_2^n} D(V\|W_{\boldsymbol{x}_1}|P_{\boldsymbol{x}_1,\boldsymbol{x}_2}) + o(n)\Big\} \tag{24}$$

where

$$D(V\|W|P) = \sum_x P(x) \sum_y V(y|x) \log \frac{V(y|x)}{W(y|x)} \tag{25}$$

is the conditional relative entropy. The second term in (23) can be handled similarly. It can be proved with methods similar to those used [13] (we omit the details due to space limitations), that the following expression holds

$$\sup_{s\geq 0} \frac{1}{n}\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = \begin{cases} \min\limits_{V\in\mathcal{V}_2} D(V\|W_{\boldsymbol{x}_1}|P_{\boldsymbol{x}_1,\boldsymbol{x}_2}), & \text{if } \mathcal{V}_2 \neq \varnothing \\ +\infty, & \text{if } \mathcal{V}_2 = \varnothing. \end{cases} \tag{26}$$

A similar expression holds with $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ swapped. Hence, after passing from types $\mathcal{V}_2^n$ to distributions $\mathcal{V}_2$ (details are again omitted here), equation (23) can be rewritten as

$$P_e \geq \exp\big\{ - nD_{\boldsymbol{x}_1,\boldsymbol{x}_2}^{(n)} + o(n)\big\} \tag{27}$$

where

$$D_{\boldsymbol{x}_1,\boldsymbol{x}_2}^{(n)} = \min\left\{ \sup_{s\geq 0} \frac{1}{n}\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s), \sup_{s\geq 0} \frac{1}{n}\mu_{\boldsymbol{x}_2,\boldsymbol{x}_1}(s)\right\}. \tag{28}$$

Finally, notice that if we consider a code with more than two codewords, say M, then there is one message m for which

$$P_{e,m} \geq \exp\big\{ - nD_{\min}(\mathcal{C}) + o(n)\big\} \tag{29}$$

where

$$D_{\min}(\mathcal{C}) \triangleq \min_{m\neq m'\in\mathcal{C}} D_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}}^{(n)} \tag{30}$$

and therefore, for the whole code, the average probability of error is lower-bounded by

$$P_e \geq \frac{P_{e,m}}{M} \geq \exp\big\{ - n\big(D_{\min}(\mathcal{C}) + R + o(1)\big)\big\}. \tag{31}$$

IV. Upper bound on the reliability function

Equation (31) illustrates us that the problem of upper-bounding the reliability function when the rate R approaches 0, reduces to upper-bounding $D_{\min}(\mathcal{C})$. We note here that the minimum distance $D_{\min}(\mathcal{C})$ of any code $\mathcal{C}$ can be upper-bounded by the minimum distance of any subcode $\hat{\mathcal{C}}$ extracted

from $\mathcal{C}$. Furthermore, the minimum distance $D_{\min}(\mathcal{C})$ is upper-bounded by the average distance $D^{(n)}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}}$ over all pairs of codewords in $\mathcal{C}$. Therefore, we upper bound the minimum distance of a code by upper bounding the average distance over a carefully selected subcode, that is,

$$D_{\min}(\mathcal{C}) \leq D_{\min}(\hat{\mathcal{C}}) \leq \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m \neq m' \in \hat{\mathcal{C}}} D^{(n)}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}} \quad (32)$$

for any $\hat{\mathcal{C}} \subset \mathcal{C}$ with $|\hat{\mathcal{C}}| = \hat{M}$.

The choice of the subcode $\hat{\mathcal{C}}$ is critical, since, in general, the average in (32) may be difficult to evaluate, for two reasons: for two arbitrary codewords $\boldsymbol{x}_m, \boldsymbol{x}_{m'}$, the functions $\mu_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}}(s), \mu_{\boldsymbol{x}_{m'},\boldsymbol{x}_m}(s)$ that define $D^{(n)}_{\boldsymbol{x}_1,\boldsymbol{x}_2}$ in (28) can be very different from each other, and also, different pairs of codewords have in general very different values of s at which the functions $\mu(s)$ attain their supremum. To overcome these difficulties, we apply the following result by Komlós [15], first employed by Blinovsky [16] in the classical maximum likelihood setting, which ensures the existence of a "symmetric" subcode in any large enough code.

Theorem 2 (Komlós [15]): For any positive integers t and $\hat{M}$, there exists a positive integer $M_0(\hat{M}, t)$ such that from any code $\mathcal{C}$ with $M > M_0(\hat{M}, t)$ codewords, a subcode $\hat{\mathcal{C}} \subset \mathcal{C}$ with $\hat{M}$ codewords can be extracted such that for any $m \neq m'$ and $\bar{m} \neq \bar{m}'$ (not necessarily different from m and m') in $\hat{\mathcal{C}}$, and any $(a, b) \in \mathcal{X}^2$,

$$\big|\, P_{m,m'}(a,b) - P_{\bar{m},\bar{m}'}(a,b) \,\big| \leq \Delta(\hat{M}, t) \quad (33)$$

where

$$\Delta(\hat{M}, t) \triangleq \frac{6}{\sqrt{\hat{M}}} + 2\sqrt{\frac{2}{t}} + \frac{3}{t}. \quad (34)$$

Thanks to property (9), the fact that all pairs of codewords in the subcode have similar joint types implies that they also have similar $\mu(s)$, while the fact that these types are close to symmetrical implies that $\mu_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}}(s)$ and $\mu_{\boldsymbol{x}_{m'},\boldsymbol{x}_m}(s)$ are close to each other. However, technical problems arise due to the presence of the suprema in (28), since even if the joint types are close to each other, the suprema of the functions $\mu(s)$ might be very different if they are approached as $s \to \infty$. This constrains our study only to a (very wide) class of channels and decoding metrics for which we are sure that at least one supremum in the definition of each $D^{(n)}_{\boldsymbol{x}_m,\boldsymbol{x}_{m'}}$ is attained at an s no larger than a known fixed value. The class is the following.

Definition 1: A discrete memoryless channel $W(y|x)$ and a decoding metric $q(x, y)$ form a *balanced pair* if $\bar{C}^q_0 = 0$ and for every pair $(a, b) \in \mathcal{X}^2$ belonging to the set

$$\mathcal{B} \triangleq \left\{ (a,b) \in \mathcal{X}^2 : \min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} \right\} \quad (35)$$

there exists a constant $B(a, b)$ such that

$$\frac{q(a,y)}{q(b,y)} = B(a,b) \quad \forall\, y \in \hat{\mathcal{Y}}_{a,b} : W(y|a) + W(y|b) > 0. \quad (36)$$

A careful check of the above condition shows that all channels and decoding metrics such that $\bar{C}^q_0 = 0$ and

$$W(y|x) > 0 \iff q(x,y) > 0 \quad (37)$$

are balanced pairs, and indeed represent a very important special case. Furthermore, for this particular subclass,

$$C_0 = 0 \iff \bar{C}^q_0 = 0 \quad (38)$$

where C_0 is the classical zero-error capacity.

We are ready to proceed and prove the upper bound on the reliability function at $R = 0^+$ for any balanced channel-metric pair. What we are going to prove is that, for this class, the $D^{(n)}_{m,m'}$ are all close to each other for all pairs of codewords in the symmetric subcode of $\mathcal{C}$ guaranteed by Theorem 3, that from now on we shall denote by $\hat{\mathcal{C}}$. In order to show this, first of all, for any concave function $f(s)$, let[3]

$$\mathcal{S} \triangleq \left\{ 0 \leq s \leq +\infty : f(s) = \sup_{s \geq 0} f(s) \right\} \quad (39)$$

and define

$$\arg\sup_{s \geq 0} f(s) \triangleq \inf \mathcal{S}. \quad (40)$$

The following two lemmas ensure that the $D^{(n)}_{m,m'}$ are close to each other for all pairs of codewords in $\hat{\mathcal{C}}$. This fact is what will make the computation of the average in (32) possible.

Lemma 1: For any balanced pair, for any pair of codewords m, m',

$$\arg\sup_{s \geq 0} \big(\mu_{m,m'}(s) + \mu_{m',m}(s)\big) \in [\,0,\, \hat{s}\,] \quad (41)$$

where

$$\hat{s} \triangleq \max_{a,b} \left\{ \arg\sup_{s \geq 0} \big(\mu_{a,b}(s) + \mu_{b,a}(s)\big) \right\} < +\infty. \quad (42)$$

Lemma 2: For any balanced pair, for any pair of codewords $m, m' \in \hat{\mathcal{C}}$, let

$$\bar{s}_{m,m'} \triangleq \min \left\{ \arg\sup_{s \geq 0} \mu_{m,m'}(s), \arg\sup_{s \geq 0} \mu_{m',m}(s) \right\}. \quad (43)$$

Then, $0 \leq \bar{s}_{m,m'} \leq \hat{s}$, with $\hat{s}$ defined by (42), and

$$D^{(n)}_{m,m'} \leq \frac{1}{n}\, \mu_{m,m'}(\bar{s}_{m,m'}) + K\Delta(\hat{M}, t) \quad (44)$$

with $\Delta(\hat{M}, t)$ as defined by (34), and

$$K \triangleq \max_{0 \leq s \leq \hat{s}} \sum_a \sum_b \big|\mu_{a,b}(s)\big|. \quad (45)$$

Furthermore, for any other pair of codewords $\bar{m}, \bar{m}' \in \hat{\mathcal{C}}$,

$$\left| \frac{1}{n}\, \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n}\, \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right| \leq 4K\Delta(\hat{M}, k). \quad (46)$$

Finally, thanks to this lemma, we can prove our upper bound on the reliability function at $R = 0^+$, which coincides with the lower bound (4).

[3]Here $f(+\infty)$ means $\lim_{s \to +\infty} f(s)$. If $\lim_{s \to +\infty} f(s) = +\infty$, then $\mathcal{S} = \{+\infty\}$, since $f(s)$ is concave.

Theorem 3: For any balanced pair,

$$E^q(0^+) = \max_{Q\in\mathcal{P}(\mathcal{X})} \sup_{s\geq 0} \left[-\sum_{a\in\mathcal{X}} \sum_{b\in\mathcal{X}} Q(a)Q(b) \log \sum_{y\in\mathcal{Y}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s \right].$$

Proof: We already pointed out that for any subcode of $\mathcal{C}$, and in particular for the subcode $\hat{\mathcal{C}}$ of Theorem 2, we have

$$D_{\min}(\mathcal{C}) \leq D_{\min}(\hat{\mathcal{C}}) \leq \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m\neq m'} D^{(n)}_{m,m'} \tag{47}$$

with $m, m' \in \hat{\mathcal{C}}$. Then, we can bound the average as follows, similarly as what Shannon, Gallager and Berlekamp did in the non-mismatch setting for pairwise reversible channels [18]. Fix any pair of codewords $\boldsymbol{x}_{\hat{m}} \neq \boldsymbol{x}_{\hat{m}'} \in \hat{\mathcal{C}}$. Then,

$$\begin{aligned}
&D_{\min}(\hat{\mathcal{C}}) \\
&\leq \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m\neq m'} D^{(n)}_{m,m'} && (48)\\
&\leq K\Delta(\hat{M},t) + \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m\neq m'} \frac{1}{n}\, \mu_{m,m'}(\bar{s}_{m,m'}) && (49)\\
&\leq 5K\Delta(\hat{M},t) + \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m\neq m'} \frac{1}{n}\, \mu_{m,m'}(\bar{s}_{\hat{m},\hat{m}'}) && (50)\\
&\simeq \frac{1}{\hat{M}(\hat{M}-1)} \sum_a \sum_b \sum_{m\neq m'} P_{m,m'}(a,b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'}) && (51)\\
&\simeq \frac{1}{n} \frac{1}{\hat{M}(\hat{M}-1)} \sum_{c=1}^{n} \sum_a \sum_b \hat{M}_c(a)\hat{M}_c(b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'}) && (52)\\
&\simeq \frac{1}{n} \frac{\hat{M}}{\hat{M}-1} \sum_{c=1}^{n} \sum_a \sum_b \frac{\hat{M}_c(a)}{\hat{M}} \frac{\hat{M}_c(b)}{\hat{M}} \mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'}) && (53)\\
&\lesssim \frac{\hat{M}}{\hat{M}-1} \max_{Q\in\mathcal{P}(\mathcal{X})} \sum_a \sum_b Q(a)Q(b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'}) && (54)\\
&\lesssim \frac{\hat{M}}{\hat{M}-1} \sup_{s\geq 0} \max_{Q\in\mathcal{P}(\mathcal{X})} \sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) && (55)
\end{aligned}$$

where the approximations hide the term $5K\Delta(\hat{M},t)$ in (50) (otherwise they would be equalities or inequalities), (49) is due to (44), (50) is due to (46), (52) is due to the standard Plotkin double counting trick, and (54) is due to the fact that for every c, $\{\hat{M}_c(a)/\hat{M}\}_{a\in\mathcal{X}}$ is a probability distribution over $\mathcal{X}$. As we already underlined, these steps are possible thanks to the fact that all pairs of codewords in $\hat{\mathcal{C}}$ have joint types that are both symmetrical and close to each other, and that this combined with the fact that for all balanced pairs we can focus the attention only on the s in a known bounded interval, all the $D^{(n)}_{m,m'}$ that appear in the average (47) are close to each other. Then, letting $M \to \infty$ (so that we may also let $\hat{M} \to \infty$, by Theorem 2) and $t \to \infty$ we obtain, thanks to the fact that the hidden term $5K\Delta(\hat{M},t)$ in (55) tend to zero,

$$D_{\min}(\mathcal{C}) \leq \sup_{s\geq 0} \max_{Q\in\mathcal{P}(\mathcal{X})} \sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) + o(1)$$

which is independent of the code $\mathcal{C}$. Finally, replacing $\mu_{a,b}(s)$ with its definition (6), thanks to equation (31), since we let $R \to 0$ *after* $n \to \infty$, we obtain the claimed upper bound upper bound on the reliability function at $R = 0^+$:

$$\begin{aligned}
&E^q(0^+) \\
&\leq \max_{Q\in\mathcal{P}(\mathcal{X})} \sup_{s\geq 0} - \sum_{a,b\in\mathcal{X}} Q(a)Q(b) \log \sum_{y\in\mathcal{Y}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s
\end{aligned}$$

which equals the expurgated lower bound given by (4), proving the theorem. ∎

REFERENCES

[1] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35-43, 1995.

[2] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[3] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Found. Trends Commun. Inf. Theory*, vol. 17, no. 2–3, pp. 149-401, 2020.

[4] T. R. M. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," *Trans. 8th Prague Conf. Inf. Theory*, pp. 211–226, 1978.

[5] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elek. Uber.*, vol. 47, no. 4, pp. 228–239, 1993.

[6] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert-Varshamov codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3452–3469, 2019.

[7] E. Asadi Kangarshahi and A. Guillén i Fàbregas, "A single-letter upper bound to the mismatch capacity," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2013-2033, 2021.

[8] A. Somekh-Baruch, "A single-letter upper bound on the mismatch capacity via a multicasting approach," *Proc. 2020 IEEE Inf. Theory Workshop*, 2021. Available online: https://arxiv.org/pdf/2007.14322.pdf

[9] E. Asadi Kangarshahi and A. Guillén i Fàbregas, "A sphere-packing exponent for mismatched decoding," *Proc. 2021 IEEE Int. Symp. Inf. Theory*, 2021.

[10] J. Scarlett, L. Peng, N. Merhav, A. Martinez and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: exponents, refinements, and connections," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4449-4462, 2014.

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

[12] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[13] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proc. 2011 IEEE Int. Symp. Inf. Theory*, pp. 894-898, 2011.

[14] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171-176, 1958.

[15] J. Komlós, "A strange pigeon-hole principle," *Order*, vol. 7, no. 2, pp. 107-113, 1990.

[16] V. M. Blinovsky, "New Approach to Estimation of the Decoding Error Probability," *Problems of Information transmission*, vol. 38, no. 1, pp. 16-19, 2002.

[17] R. Diestel, *Graph Theory*, Springer-Verlag Berlin Heidelberg, 2017.

[18] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Information and Control*, vol. 10, no. 5, pp. 522-552, 1967.