Mismatched Decoding Reliability Function at Zero Rate

Marco Bondaschi, *Student Member, IEEE,* Albert Guillén i Fàbregas, *Senior Member, IEEE,*
and Marco Dalai, *Senior Member, IEEE*

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-decoding metric pairs.

Index Terms—Error exponents, mismatched decoding.

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 $\mathcal{M} = \{1, 2, \ldots, M\}$ 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 $\boldsymbol{x}_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,n})$. The *rate of transmission* is defined as

$$R \triangleq \frac{\log M}{n}. \tag{1}$$

When message m is sent, an output sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is received with probability

$$W^n(\boldsymbol{y}|\boldsymbol{x}_m) = \prod_{i=1}^{n} W(y_i|x_{m,i}). \tag{2}$$

A decoder is a function $\mathcal{C}^{-1} : \mathcal{Y}^n \to \mathcal{M}$ whose task is to map each possible output sequence to a message in $\mathcal{M}$ which hopefully is equal to the message that was originally sent with high probability. In this paper, we consider a decoder that follows the rule

$$\mathcal{C}^{-1}(\boldsymbol{y}) \in \{m \in \mathcal{M} : q^n(\boldsymbol{x}_m, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_{m'}, \boldsymbol{y}) \; \forall m' \in \mathcal{M}\}, \tag{3}$$

M. Bondaschi is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: marco.bondaschi@epfl.ch).

A. Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K., and with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: guillen@ieee.org).

M. Dalai is with the Department of Information Engineering at the University of Brescia, Via Branze 38 I-25123 Brescia, Italy (e-mail: marco.dalai@unibs.it).

where

$$q^n(\boldsymbol{x}_m, \boldsymbol{y}) = \prod_{i=1}^{n} q(x_{m,i}, y_i) \tag{4}$$

for a fixed function $q : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ called *decoding metric*. We assume for now that when there is a tie, that is, when for a certain output sequence $\boldsymbol{y}$ the maximal $q^n(\boldsymbol{x}_m, \boldsymbol{y})$ is attained by more than one message, the decoder selects one of them with an arbitrary rule. However, as we will discuss in more detail later on, most of the results in this work are valid only for a decoder that breaks ties equiprobably among the messages that maximize $q^n(\boldsymbol{x}_m, \boldsymbol{y})$. We will distinguish this case from the general one when necessary.

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]. The mismatched decoding problem encompasses a number of important problems such as channel uncertainty, fading channels, reduced-complexity decoding, bit-interleaved coded modulation, optical communications, and zero-error and zero-undetected error capacity. See [3] for a recent survey of the information theoretic foundations of mismatched decoding.

When message m is sent, a decoding error occurs if $\mathcal{C}^{-1}(\boldsymbol{y}) \neq m$. The probability of this event is

$$P_{e,m}^{(n)} \triangleq \sum_{\boldsymbol{y} \in \mathcal{Y}_m^c} W^n(\boldsymbol{y}|\boldsymbol{x}_m), \tag{5}$$

where $\mathcal{Y}_m \subset \mathcal{Y}^n$ is the subset of output sequences that are decoded to m. The average probability of error of the code is

$$P_e^{(n)} \triangleq \frac{1}{M} \sum_{m=1}^{M} P_{e,m}^{(n)}. \tag{6}$$

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

$$E^q(R) \triangleq \limsup_{n \to \infty} -\frac{\log P_e^q(R, n)}{n} \tag{7}$$

and represents the asymptotic exponent with which the probability of error goes to zero for a given channel and decoding metric, when an optimal code with blocklength n and rate at least R is used. The supremum of the information rates R for which the error probability tends to zero is called *mismatched capacity*.

In general, there is no single-letter expression for the mismatched capacity. A number of achievable rate and error exponent results based on random coding are available in the literature [1], [2], [4]–[6]. In terms of upper bounds on the mismatched capacity or on the reliability function, there are fewer results 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 derived in [9], yielding an improved upper bound on the mismatched capacity.

In this paper, 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 $E^q(0^+)$. In the following, we focus only on decoding metrics that are meaningful for our problem. In particular, we restrict our attention to decoding metrics such that

$$W(y|x) > 0 \implies q(x,y) > 0 \tag{8}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. In fact, channels with a decoding metric that does not meet this condition for some input x have a mismatched capacity equal to 0 if that input is used [1], and so they are of little interest.

As we already mentioned, several lower bounds on the mismatched reliability function exist. The one that is most relevant for this work is a generalization of Gallager's classical expurgated bound to the case of mismatched decoding obtained by Scarlett *et al.* [10]; when the rate approaches zero, their bound takes the form

$$E^q(0^+) \geq$$

$$\max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} -\sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{X}} Q(a)Q(b) \log \sum_{y \in \mathcal{Y}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s. \tag{9}$$

In the following, we derive an upper bound on $E^q(0^+)$ for a wide class of channels and decoding metrics, under the assumption that ties are broken equiprobably. Such an upper bound will turn out to be equal to the lower bound (9); therefore, for such a class of channels, the bound (9) is tight.

In order to prove our bound, in Section II we study conditions that channels and decoding metrics must satisfy in order to have a finite reliability function at rate $R = 0^+$. Then, in Section III we derive a lower bound on the mismatched probability of error for two codewords, and in section IV we prove the tightness of (9) using the lower bound of Section III and a probabilistic result by Komlós (obtained using Ramsey theory) on the existence of a subset of random variables from a larger set that (asymptotically) have pairwise symmetric distributions. The application of these ideas in coding theory originated in works of Blinovsky, see for example [15]–[17]. See also [19] for a recent revisitation of the maximum likelihood case.

II. MISMATCHED ZERO-ERROR CAPACITY

In the following we assume that condition (8) is satisfied. It is also useful to restrict our attention only to channels and decoding metrics such that at all rates $R > 0$ the minimum probability of error is strictly positive; if this is not the case, then at $R = 0^+$ the reliability function is infinite and no finite upper bound is possible. Thus, we introduce a new quantity, the mismatched zero-error capacity C_0^q for a channel $W(y|x)$ and a decoding metric $q(x,y)$, defined as the supremum of the rates R for which there exist codes with probability of error exactly equal to zero. If C_0^q is positive for some channel and decoding metric, then the reliability function at $R = 0^+$ is infinite. Hence, we would like to restrict our attention to channels and decoding metrics with $C_0^q = 0$.

We now proceed to analyze more closely the conditions for a positive mismatched zero-error capacity. Notice that 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) $W^n(\boldsymbol{y}|\boldsymbol{x}_1) > 0 \implies q^n(\boldsymbol{x}_1, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_2, \boldsymbol{y})$
   $W^n(\boldsymbol{y}|\boldsymbol{x}_2) > 0 \implies q^n(\boldsymbol{x}_2, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})$.

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. Notice that up to now we are still assuming that ties can be decoded with an arbitrary rule. That is why Condition 2 admits cases such as $q^n(\boldsymbol{x}_1, \boldsymbol{y}) = q^n(\boldsymbol{x}_2, \boldsymbol{y})$: there always exists a tie-breaking rule that decodes each output sequence correctly in these cases (the one that associates to each $\boldsymbol{y}$ the only $\boldsymbol{x}$ with $W^n(\boldsymbol{y}|\boldsymbol{x}) > 0$).

As we stated earlier on, our upper bound on $E^q(0^+)$ only holds for a decoder that breaks ties equiprobably. Therefore, even if the definition of mismatched zero-error capacity given above is the most general, since it admits any decoding strategy for breaking ties, it is nonetheless meaningful to us to introduce a second definition of mismatched zero-error capacity, that we denote by $\bar{C}_0^q$, that is the supremum of the rates R for which there exist codes with probability of error exactly equal to zero, given that ties are broken equiprobably. Since this decoding strategy is not necessarily the best one, that is, the one that achieves the minimum probability of error, it follows that, in general, $\bar{C}_0^q \leq C_0^q$.

As for the conditions stated above, the only difference, when $\bar{C}_0^q$ is considered instead of C_0^q, is that condition 2 becomes:

2b) $W^n(\boldsymbol{y}|\boldsymbol{x}_1) > 0 \implies q^n(\boldsymbol{x}_1, \boldsymbol{y}) > q^n(\boldsymbol{x}_2, \boldsymbol{y})$
   $W^n(\boldsymbol{y}|\boldsymbol{x}_2) > 0 \implies q^n(\boldsymbol{x}_2, \boldsymbol{y}) > q^n(\boldsymbol{x}_1, \boldsymbol{y})$

since in this case ties are not allowed, given that, with ties broken equiprobably, there is always a positive probability of decoding the output sequence incorrectly.

Finally, when the chosen decoding metric is the maximum-likelihood one, that is, $q(x,y) = W(y|x)$, the two zero-error capacities introduced above are equal and coincide with the classical zero-error capacity C_0; also, since the maximum-likelihood decoding metric is the one that minimizes the probability of error, we have that, for any decoding metric, $\bar{C}_0^q \leq C_0^q \leq C_0$.

The next objective of this section is to find conditions for the mismatched zero-error capacity to be zero that depend only on the single-letter channel probabilities $W(y|x)$ and decoding

metric $q(x, y)$. This can be done using the same tools that we will need to study $E^q(0^+)$. Therefore, we introduce now a real-valued function that will be useful to both ends. For any two sequences $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathcal{X}^n$, we define

$$\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \triangleq -\log \sum_{\boldsymbol{y} \in \hat{\mathcal{Y}}^n_{\boldsymbol{x}_1, \boldsymbol{x}_2}} W^n(\boldsymbol{y}|\boldsymbol{x}_1) \left(\frac{q^n(\boldsymbol{x}_2, \boldsymbol{y})}{q^n(\boldsymbol{x}_1, \boldsymbol{y})} \right)^s, \quad (10)$$

where

$$\hat{\mathcal{Y}}^n_{\boldsymbol{x}_1, \boldsymbol{x}_2} \triangleq \left\{ \boldsymbol{y} \in \mathcal{Y}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) q^n(\boldsymbol{x}_2, \boldsymbol{y}) > 0 \right\}. \quad (11)$$

When $n = 1$, (10) becomes, for any $a, b \in \mathcal{X}$,

$$\mu_{a,b}(s) \triangleq -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) \left(\frac{q(b, y)}{q(a, y)} \right)^s, \quad (12)$$

with

$$\hat{\mathcal{Y}}_{a,b} \triangleq \left\{ y \in \mathcal{Y} : q(a, y) q(b, y) > 0 \right\}. \quad (13)$$

An additional quantity that will be useful to us is the limit of the derivative of $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ when $s \to \infty$; for this reason, we introduce a compact symbol for it:

$$\mu'_{\boldsymbol{x}_1, \boldsymbol{x}_2} \triangleq \lim_{s \to \infty} \mu'_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \quad (14)$$

$$\mu'_{a,b} \triangleq \lim_{s \to \infty} \mu'_{a,b}(s) \quad (15)$$

and we set by definition $\mu'_{\boldsymbol{x}_1, \boldsymbol{x}_2} = +\infty$ if $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = +\infty$, and the same for $\mu'_{a,b}$.[1]

Lemma 1. *The following properties hold for all $a, b \in \mathcal{X}$ and all sequences $\boldsymbol{x}_1, \boldsymbol{x}_2$ in $\mathcal{X}^n$.*

1) Let $P_{\boldsymbol{x}_1, \boldsymbol{x}_2}$ be the joint type of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Then,

$$\frac{1}{n} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = \sum_{a \in \mathcal{X}} \sum_{b \in \mathcal{X}} P_{\boldsymbol{x}_1, \boldsymbol{x}_2}(a, b) \mu_{a,b}(s). \quad (16)$$

2) $\mu_{a,b}(s)$ and $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ are concave.
3)

$$\mu_{a,a}(s) = 0. \quad (17)$$

4)

$$\mu'_{a,b} = \min_{y : W(y|a) > 0} \log \frac{q(a, y)}{q(b, y)}. \quad (18)$$

Proof. To prove property 1, notice that $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ is additive, in the sense that it can be rewritten coordinate-by-coordinate as

$$\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = \sum_{c=1}^{n} \mu_c(s), \quad (19)$$

where

$$\mu_c(s) \triangleq -\log \sum_{y \in \hat{\mathcal{Y}}_c} W(y|x_{1,c}) \left(\frac{q(x_{2,c}, y)}{q(x_{1,c}, y)} \right)^s \quad (20)$$

and

$$\hat{\mathcal{Y}}_c \triangleq \left\{ y \in \mathcal{Y} : q(x_{1,c}, y) q(x_{2,c}, y) > 0 \right\}. \quad (21)$$

Each term of the sum in (19) only depends on the pair of input symbols $(x_{1,c}, x_{2,c})$. Since every pair $(a, b) \in \mathcal{X}^2$ appears in

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

$n P_{\boldsymbol{x}_1, \boldsymbol{x}_2}(a, b)$ coordinates, grouping together the equal terms in (19) leads to (16).

Property 2 can be proved by computing the first and second derivatives of $\mu_{a,b}(s)$, that is

$$\mu'_{a,b}(s) = -\frac{\sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s \log \frac{q(b,y)}{q(a,y)}}{\sum_{\bar{y} \in \hat{\mathcal{Y}}_{a,b}} W(\bar{y}|a) \left(\frac{q(b,\bar{y})}{q(a,\bar{y})} \right)^s} \quad (22)$$

and

$$\mu''_{a,b}(s) = -\frac{\sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s \left(\log \frac{q(b,y)}{q(a,y)} \right)^2}{\sum_{\bar{y} \in \hat{\mathcal{Y}}_{a,b}} W(\bar{y}|a) \left(\frac{q(b,\bar{y})}{q(a,\bar{y})} \right)^s}$$
$$+ \left[\mu'_{a,b}(s) \right]^2. \quad (23)$$

Now, for any $s \geq 0$, these two quantities can be seen respectively as the expected value and the variance (with a minus sign) of a random variable; in fact, define the following probability distribution over the set of sequences $\hat{\mathcal{Y}}_{a,b}$,

$$Q_s(y) \triangleq \frac{W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s}{\sum_{\bar{y} \in \hat{\mathcal{Y}}_{a,b}} W(\bar{y}|a) \left(\frac{q(b,\bar{y})}{q(a,\bar{y})} \right)^s}, \quad (24)$$

and define the random variable

$$D(y) \triangleq -\log \frac{q(b, y)}{q(a, y)}. \quad (25)$$

Then, we can rewrite the two derivatives as

$$\mu'_{a,b}(s) = \mathbb{E}_{Q_s}[D] = \sum_{y \in \hat{\mathcal{Y}}_{a,b}} Q_s(y) D(y) \quad (26)$$

$$\mu''_{a,b}(s) = -\mathrm{Var}_{Q_s}[D] = -\mathbb{E}_{Q_s}[D^2] + \left(\mathbb{E}_{Q_s}[D] \right)^2 \quad (27)$$

$$= -\sum_{y \in \hat{\mathcal{Y}}_{a,b}} Q_s(y) D^2(y) + \left[\mu'_{a,b}(s) \right]^2. \quad (28)$$

Since the variance of a random variable is always non-negative, it follows that $\mu''_{a,b}(s) \leq 0$ and therefore that $\mu_{a,b}(s)$ is concave. By property 1, $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ is also concave, since it can be rewritten as a (weighted) sum of concave functions.

Property 3 follows from the fact that, from definition (12),

$$\mu_{a,a}(s) = -\log \sum_{y \in \hat{\mathcal{Y}}_{a,a}} W(y|a) = 0, \quad (29)$$

where the last equality follows from (8), since

$$\left\{ y \in \mathcal{Y} : W(y|a) > 0 \right\} \subset \left\{ y \in \mathcal{Y} : q(a, y) > 0 \right\} = \hat{\mathcal{Y}}_{a,a}. \quad (30)$$

To prove property 4, first notice that $\mu_{a,b}(s) = +\infty$ if and only if

$$\left\{ y \in \mathcal{Y} : W(y|a) q(b, y) > 0 \right\} = \varnothing. \quad (31)$$

In such a case, the right-hand side of (18) equals $+\infty$, in accordance to what we set by definition. If instead the set on the left-hand side of (31) is not empty, then the property follows directly by taking the limit $s \to +\infty$ of the right-hand side of (22), since both numerator and denominator are dominated by the exponentials with the largest base, which is equal to

$$\max_{y : W(y|a) > 0} \frac{q(b, y)}{q(a, y)}. \qquad \square$$

We are now ready to prove the following theorem on the mismatched zero-error capacities C_0^q and $\bar{C}_0^q$.

Theorem 1. *For any given discrete memoryless channel $W(y|x)$ and decoding metric $q(x,y)$:*

1) $C_0^q = 0$ *if and only if*

$$\min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} \leq \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} \tag{32}$$

for all $a,b \in \mathcal{X}$, and for all $a,b \in \mathcal{X}$ such that

$$\min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} \tag{33}$$

there exists some $y \in \mathcal{Y}$ such that

$$W(y|a)W(y|b) > 0. \tag{34}$$

2) $\bar{C}_0^q = 0$ *if and only if*

$$\min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} \leq \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} \tag{35}$$

for all $a,b \in \mathcal{X}$.

Corollary 1.

$$\bar{C}_0^q = 0 \implies \max_{Q\in\mathcal{P}(\mathcal{X})} \sup_{s\geq 0} \sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) < +\infty. \tag{36}$$

Proof. We first show that the quantities defined in (10) and (14) — and consequently also (12) and (15) — satisfy the following properties, for every $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:

$$\lim_{s\to+\infty} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = +\infty \iff \mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} > 0 \tag{37}$$

$$\lim_{s\to+\infty} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \in [0,+\infty) \iff \mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} = 0 \tag{38}$$

$$\lim_{s\to+\infty} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = -\infty \iff \mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} < 0. \tag{39}$$

In fact, consider the function

$$f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \triangleq e^{-\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)}$$
$$= \sum_{\boldsymbol{y}\in\hat{\mathcal{Y}}^n_{\boldsymbol{x}_1,\boldsymbol{x}_2}} W^n(\boldsymbol{y}|\boldsymbol{x}_1)\left(\frac{q^n(\boldsymbol{x}_2,\boldsymbol{y})}{q^n(\boldsymbol{x}_1,\boldsymbol{y})}\right)^s. \tag{40}$$

Since $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ is the sum of non-negative quantities, when $s \to \infty$ only three alternatives are possible: $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ tends to infinity, $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ tends to a finite positive number, or $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ tends to zero. In the first case, $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \to \infty$ (and consequently $\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \to -\infty$) if and only if at least one non-zero term of the sum goes to infinity, which in turn happens if and only if

$$\max_{\boldsymbol{y}:W^n(\boldsymbol{y}|\boldsymbol{x}_1)>0} \log\frac{q^n(\boldsymbol{x}_2,\boldsymbol{y})}{q^n(\boldsymbol{x}_1,\boldsymbol{y})} = -\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} > 0. \tag{41}$$

In the second case, $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ tends to a finite positive real number if and only if at least one term of the sum tends to a finite positive number and all the other terms tend to zero, which happens if and only if

$$\max_{\boldsymbol{y}:W^n(\boldsymbol{y}|\boldsymbol{x}_1)>0} \log\frac{q^n(\boldsymbol{x}_2,\boldsymbol{y})}{q^n(\boldsymbol{x}_1,\boldsymbol{y})} = -\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} = 0; \tag{42}$$

in such a case, the limit is

$$\lim_{s\to\infty} f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = \sum_{\text{some }\boldsymbol{y}} W^n(\boldsymbol{y}|\boldsymbol{x}_1), \tag{43}$$

which is strictly positive and at most 1, consequently

$$\lim_{s\to\infty} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = -\log\sum_{\text{some }\boldsymbol{y}} W^n(\boldsymbol{y}|\boldsymbol{x}_1) \tag{44}$$

is finite and greater than or equal to 0. Finally, $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)$ tends to zero (and consequently $\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) \to \infty$) if and only if all terms of the sum tend to zero, which happens if and only if

$$\max_{\boldsymbol{y}:W^n(\boldsymbol{y}|\boldsymbol{x}_1)>0} \log\frac{q^n(\boldsymbol{x}_2,\boldsymbol{y})}{q^n(\boldsymbol{x}_1,\boldsymbol{y})} = -\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} < 0. \tag{45}$$

Notice that the same properties hold also for $\mu_{a,b}(s)$ and $\mu'_{a,b}$ for any $a,b \in \mathcal{X}$, since one can choose $\boldsymbol{x}_1 = a$ and $\boldsymbol{x}_2 = b$.

Next, we analyze more closely 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 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 $C_0^q > 0$ is that

$$\mathcal{A} \triangleq \left\{(a,b)\in\mathcal{X}^2 : W(y|a)W(y|b) = 0 \text{ for all } y\in\mathcal{Y}\right\} \neq \varnothing. \tag{46}$$

Notice that this is also the condition for the classical C_0 to be positive, which is of course a necessary condition to have $C_0^q > 0$, since, as we already pointed out, we have in general $C_0^q \leq C_0$.

Instead, thanks to (18), condition 2 is satisfied if and only if both

$$\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2} \geq 0 \qquad \text{and} \qquad \mu'_{\boldsymbol{x}_2,\boldsymbol{x}_1} \geq 0. \tag{47}$$

Hence, using (16), there exists a pair of codewords satisfying condition 2 if and only if there exists a joint type P such that both

$$\sum_a \sum_b P(a,b)\mu'_{a,b} \geq 0 \quad \text{and} \quad \sum_a \sum_b P(a,b)\mu'_{b,a} \geq 0. \tag{48}$$

Any pair of codewords with a joint type P satisfying (48) satisfies Condition 2 for a positive C_0^q. Now, condition (48) is true if and only if

$$\sup_{P\in\mathcal{P}(\mathcal{X}^2)} \min\left\{\sum_a \sum_b P(a,b)\mu'_{a,b}, \sum_a \sum_b P(a,b)\mu'_{b,a}\right\} > 0 \tag{49}$$

or

$$\min\left\{\sum_a \sum_b P(a,b)\mu'_{a,b}, \sum_a \sum_b P(a,b)\mu'_{b,a}\right\} = 0 \tag{50}$$

for some $P \in \mathcal{P}_{\mathbb{Q}}(\mathcal{X}^2)$, where $\mathcal{P}_{\mathbb{Q}}(\mathcal{X}^2)$ denotes the set of probability vectors over $\mathcal{X}^2$ with rational entries. 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 distribution such that $P(a,b) = P(b,a)$ for all a,b. In such a case, the two functions are both equal to

$$\sum_{a \leq b} P(a,b)(\mu'_{a,b} + \mu'_{b,a}) \tag{51}$$

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 and so it belongs to $\mathcal{P}_{\mathbb{Q}}(\mathcal{X}^2)$. Hence, thanks to (18), conditions (49) and (50) become

$$\max_{a,b} \left(\min_{y:W(y|a)>0} \log \frac{q(a,y)}{q(b,y)} + \min_{y:W(y|b)>0} \log \frac{q(b,y)}{q(a,y)} \right) \geq 0. \tag{52}$$

Thus, if (52) is true, then we can find at least one joint type P that satisfies (48), and with it a set of pairs of codewords that satisfy Condition 2 for $C_0^q > 0$. 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 investigate this issue. If the maximum in (52) is strictly positive, then, thanks to the fact that the argument of the $\max$ in (51) is linear in P, in the neighborhood of the joint type 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 (51), still returns a positive value. Hence, the two codewords with that joint type satisfy both conditions 1 and 2, and C_0^q is positive. If, instead, the maximum in (52) is exactly zero, then, a joint type satisfying also condition 1 exists only if one of the joint types achieving the maximum already has a positive entry corresponding to a pair of symbols $(a,b) \in \mathcal{A}$, that is, there exists a pair of symbols (a,b) such that

$$\min_{y:W(y|a)>0} \log \frac{q(a,y)}{q(b,y)} + \min_{y:W(y|b)>0} \log \frac{q(b,y)}{q(a,y)} = 0 \tag{53}$$

and for all y, $W(y|a)W(y|b) = 0$.

To summarize, $C_0^q > 0$ if and only if $\mathcal{A} \neq \varnothing$ and either

$$\max_{a,b} \left(\min_{y:W(y|a)>0} \log \frac{q(a,y)}{q(b,y)} + \min_{y:W(y|b)>0} \log \frac{q(b,y)}{q(a,y)} \right) > 0 \tag{54}$$

or there exists a pair $(a,b) \in \mathcal{A}$ such that

$$\min_{y:W(y|a)>0} \log \frac{q(a,y)}{q(b,y)} + \min_{y:W(y|b)>0} \log \frac{q(b,y)}{q(a,y)} = 0. \tag{55}$$

The complementary conditions give the first part of the theorem.

The second part is a bit more straightforward. Condition 1 remains identical; as for condition 2b, condition (47) is replaced by

$$\mu'_{\boldsymbol{x}_1, \boldsymbol{x}_2} > 0 \quad \text{and} \quad \mu'_{\boldsymbol{x}_2, \boldsymbol{x}_1} > 0. \tag{56}$$

Following the same steps as before, we get that $\bar{C}_0^q > 0$ if and only if $\mathcal{A} \neq \varnothing$ and

$$\max_{a,b} \left(\min_{y:W(y|a)>0} \log \frac{q(a,y)}{q(b,y)} + \min_{y:W(y|b)>0} \log \frac{q(b,y)}{q(a,y)} \right) > 0. \tag{57}$$

The complementary conditions give the second part of the theorem.

Finally, regarding the corollary, the implication follows from the fact that

$$\max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s)$$
$$= \frac{1}{2} \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_a \sum_b Q(a)Q(b)\big(\mu_{a,b}(s) + \mu_{b,a}(s)\big) \tag{58}$$

$$\leq \frac{1}{2} \max_{Q \in \mathcal{P}(\mathcal{X})} \sum_a \sum_b Q(a)Q(b) \sup_{s \geq 0} \big(\mu_{a,b}(s) + \mu_{b,a}(s)\big), \tag{59}$$

where the equality follows from the fact that

$$\sum_a \sum_b Q(a)Q(b)\mu_{a,b}(s) = \sum_a \sum_b Q(a)Q(b)\mu_{b,a}(s). \tag{60}$$

The quantity in (59) is finite if $\bar{C}_0^q = 0$, since inequality (35) can be rewritten as

$$\mu'_{a,b} + \mu'_{b,a} \leq 0, \tag{61}$$

which by (37) is equivalent to

$$\lim_{s \to +\infty} \big(\mu_{a,b}(s) + \mu_{b,a}(s)\big) < +\infty, \tag{62}$$

which in turn implies that

$$\sup_{s \geq 0} \big(\mu_{a,b}(s) + \mu_{b,a}(s)\big) < +\infty \tag{63}$$

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

III. LOWER BOUND ON THE PROBABILITY OF ERROR

We now proceed to derive a lower bound on the minimum probability of error of any discrete memoryless channel and mismatched metric, under the assumption that $\bar{C}_0^q = 0$ and that ties are decoded equiprobably. In order to achieve this, we first 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.

Following (5), the probabilities of error for the two messages 1 and 2 satisfy

$$P_{e,1}^{(n)} \geq \sum_{\boldsymbol{y} \notin \mathcal{Y}_1^n} W^n(\boldsymbol{y}|\boldsymbol{x}_1) \tag{64}$$

$$P_{e,2}^{(n)} \geq \sum_{\boldsymbol{y} \notin \mathcal{Y}_2^n} W^n(\boldsymbol{y}|\boldsymbol{x}_2), \tag{65}$$

where

$$\mathcal{Y}_1^n = \{\boldsymbol{y} \in \mathcal{Y}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_2, \boldsymbol{y})\} \tag{66}$$

$$\mathcal{Y}_2^n = \{\boldsymbol{y} \in \mathcal{Y}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) \leq q^n(\boldsymbol{x}_2, \boldsymbol{y})\}. \tag{67}$$

Notice that the lower bounds are due to the fact that we consider all sequences that are tied between the two messages as correctly decoded. Also, we can restrict our attention only to sequences such that $q^n(\boldsymbol{x}_1, \boldsymbol{y}) \, q^n(\boldsymbol{x}_2, \boldsymbol{y}) > 0$, that is, we can substitute $\mathcal{Y}^n$ with the set

$$\hat{\mathcal{y}}^n = \{\boldsymbol{y} \in \mathcal{Y}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) \, q^n(\boldsymbol{x}_2, \boldsymbol{y}) > 0\}. \tag{68}$$

In fact, thanks to the condition in (8), if $q^n(\boldsymbol{x}_1, \boldsymbol{y})$ and $q^n(\boldsymbol{x}_1, \boldsymbol{y})$ are both zero for some sequence $\boldsymbol{y}$, then also

$W^n(\boldsymbol{y}|\boldsymbol{x}_1)$ and $W^n(\boldsymbol{y}|\boldsymbol{x}_2)$ are zero, and the sequence contributes neither to $P_{e,1}$ nor to $P_{e,2}$. If instead only one of the two is zero, for example $q^n(\boldsymbol{x}_1, \boldsymbol{y})$, then $q^n(\boldsymbol{x}_1, \boldsymbol{y}) < q^n(\boldsymbol{x}_2, \boldsymbol{y})$ and the sequence would only contribute to $P_{e,1}$; however, by (8) we have $W^n(\boldsymbol{y}|\boldsymbol{x}_1) = 0$, and so also its contribution to $P_{e,1}$ is zero.

We now introduce some tools from the method of types developed by Csiszár and Körner [11]. We define the *conditional type* of the sequence $\boldsymbol{y}$ given the codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ for any $a, b \in \mathcal{X}$ and $y \in \mathcal{Y}$ as

$$V_{\boldsymbol{y}}(y|a,b) \triangleq \frac{P_{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}}(a,b,y)}{P_{\boldsymbol{x}_1,\boldsymbol{x}_2}(a,b)}, \tag{69}$$

where $P_{\boldsymbol{x}_1,\boldsymbol{x}_2}$ and $P_{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}}$ are the joint types of the pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and the triple $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ respectively. In order to lighten the notation, from now on we let $P_{1,2}(a,b) = P_{\boldsymbol{x}_1,\boldsymbol{x}_2}(a,b)$. We also define the *conditional Kullback-Leibler divergence* as

$$D(V\|Z|P_{1,2}) \triangleq \sum_{a\in\mathcal{X}}\sum_{b\in\mathcal{X}} P_{1,2}(a,b)D\big(V(\cdot|a,b)\|Z(\cdot|a,b)\big) \tag{70}$$

for any two conditional distributions $V, Z : \mathcal{X}^2 \to \mathcal{Y}$.

Now, all sequences $\boldsymbol{y} \in \mathcal{Y}^n$ with the same conditional type $V_{\boldsymbol{y}}$ also have the same values for $W^n(\boldsymbol{y}|\boldsymbol{x}_1)$, $W^n(\boldsymbol{y}|\boldsymbol{x}_2)$, $q^n(\boldsymbol{x}_1, \boldsymbol{y})$ and $q^n(\boldsymbol{x}_2, \boldsymbol{y})$, so they all give the same contribution to the probabilities of error in (64) and (65). Hence, we can group them together and reformulate the probability of error as a function of conditional types instead of sequences:

$$P_{e,1}^{(n)} = \sum_{V \notin \mathcal{V}_1^n} W^n(V|\boldsymbol{x}_1) \tag{71}$$

$$P_{e,2}^{(n)} = \sum_{V \notin \mathcal{V}_2^n} W^n(V|\boldsymbol{x}_2), \tag{72}$$

where

$$\mathcal{V}_1^n = \{V \in \mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2) : q^n(\boldsymbol{x}_1, V) \geq q^n(\boldsymbol{x}_2, V)\} \tag{73}$$

$$= \Big\{V \in \mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2) :$$

$$\sum_{a,b} P_{1,2}(a,b) \sum_y V(y|a,b) \log \frac{q(a,y)}{q(b,y)} \geq 0\Big\}, \tag{74}$$

$$\mathcal{V}_2^n = \Big\{V \in \mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2) :$$

$$\sum_{a,b} P_{1,2}(a,b) \sum_y V(y|a,b) \log \frac{q(a,y)}{q(b,y)} \leq 0\Big\}, \tag{75}$$

and $\mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the set of all conditional types given $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Furthermore, if we define the two conditional distributions

$$W_1(y|a,b) = W(y|a) \quad \text{for all} \quad b \in \mathcal{X} \tag{76}$$

$$W_2(y|a,b) = W(y|b) \quad \text{for all} \quad a \in \mathcal{X} \tag{77}$$

then from classical results of the method of types (see [11]) we can derive the following lemma.

Lemma 2. *For any conditional type $V \in \mathcal{V}^n(\boldsymbol{x}_1, \boldsymbol{x}_2)$ we have*

$$W^n(V|\boldsymbol{x}_m) \geq \frac{1}{(n+1)^{|\mathcal{X}|^2|\mathcal{Y}|}} e^{-nD(V\|W_m|P_{1,2})} \tag{78}$$

for any $m \in \mathcal{M} = \{1, 2\}$.

Hence, we can lower bound the probabilities of error in (71) and (72) as

$$P_{e,m}^{(n)} \geq \sum_{V \notin \mathcal{V}_m^n} e^{-n[D(V\|W_m|P_{1,2})+\delta_1(n)]}, \tag{79}$$

where

$$\delta_1(n) = |\mathcal{X}|^2|\mathcal{Y}|\frac{\log(n+1)}{n}. \tag{80}$$

Also, using (10), it can be verified by substitution that

$$\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) - s\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = nD(V_s\|W_1|P_{1,2}) \tag{81}$$

for the conditional distribution

$$V_s(y|a,b) = \frac{W(y|a)\big(\frac{q(b,y)}{q(a,y)}\big)^s}{\sum_{\bar{y}\in\hat{y}} W(\bar{y}|a)\big(\frac{q(b,\bar{y})}{q(a,\bar{y})}\big)^s}. \tag{82}$$

We also need the following lemma about approximating a probability distribution with a type.

Lemma 3 (Shannon [13]). *For any distribution $Q \in \mathcal{P}(\mathcal{Y})$, for any $n \in \mathbb{N}$, there exists a type $\hat{Q} \in \mathcal{T}^n(\mathcal{Y})$ such that*

$$|\hat{Q}(y) - Q(y)| \leq \frac{1}{n} \quad \text{for all} \quad y \in \mathcal{Y} \tag{83}$$

and $\hat{Q}(y) = 0$ if $Q(y) = 0$.

We can now prove the following theorem on the probability of error.

Theorem 2. *For n large enough, the probabilities of error $P_{e,1}^{(n)}$ and $P_{e,2}^{(n)}$ are lower-bounded by*

$$P_{e,1}^{(n)} \geq e^{-\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)+s\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)-\delta(n)} \tag{84}$$

for every s such that $\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) < 0$, and

$$P_{e,2}^{(n)} \geq e^{-\mu_{\boldsymbol{x}_2,\boldsymbol{x}_1}(s)+s\mu'_{\boldsymbol{x}_2,\boldsymbol{x}_1}(s)-\delta(n)} \tag{85}$$

for every s such that $\mu'_{\boldsymbol{x}_2,\boldsymbol{x}_1}(s) < 0$, where

$$\delta(n) = |\mathcal{X}|^2|\mathcal{Y}|\left(1 + 2\log(n+1) + \log\frac{1}{W_{\min}}\right) \tag{86}$$

and $W_{\min} = \min_{x,y} W(y|x)$, where the minimum is over all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $W(y|x) > 0$.

Proof. We prove the bound for $P_{e,1}^{(n)}$; the bound for $P_{e,2}^{(n)}$ follows similarly. Notice that we can rewrite

$$\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) = \sum_{a,b} P_{1,2}(a,b) \sum_y V_s(y|a,b) \log \frac{q(a,y)}{q(b,y)} \tag{87}$$

for V_s as defined in (82). Hence, for every s such that $\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) < 0$ we have

$$\sum_{a,b} P_{1,2}(a,b) \sum_y V_s(y|a,b) \log \frac{q(a,y)}{q(b,y)} < 0. \tag{88}$$

Intuitively, for n large enough we can approximate V_s with a conditional type $\hat{V}_s$ that still satisfies (88). In fact, thanks to Lemma 3 we have

$$\left| \sum_{a,b} P_{1,2}(a,b) \sum_y [\hat{V}_s(y|a,b) - V_s(y|a,b)] \log \frac{q(a,y)}{q(b,y)} \right|$$

$$\leq \sum_{a,b} P_{1,2}(a,b) \sum_y |\hat{V}_s(y|a,b) - V_s(y|a,b)| \left| \log \frac{q(a,y)}{q(b,y)} \right| \tag{89}$$

$$\leq \sum_{a,b} P_{1,2}(a,b) \sum_y \frac{1}{nP_{1,2}(a,b)} \left| \log \frac{q(a,y)}{q(b,y)} \right| \tag{90}$$

$$\leq \frac{1}{n} \sum_{a,b,y} \left| \log \frac{q(a,y)}{q(b,y)} \right|, \tag{91}$$

that goes to 0 as $n \to \infty$. Now, $\hat{V}_s$ does not belong to $\mathcal{V}_1^n$, so we can lower bound (79) with

$$P_{e,1}^{(n)} \geq \sum_{V \notin \mathcal{V}_1^n} e^{-n[D(V\|W_1|P_{1,2})+\delta_1(n)]} \tag{92}$$

$$\geq e^{-n[D(\hat{V}_s\|W_1|P_{1,2})+\delta_1(n)]}. \tag{93}$$

Again, since $\hat{V}_s$ and V_s are close to each other, also $D(\hat{V}_s\|W_1|P_{1,2})$ and $D(V_s\|W_1|P_{1,2})$ are close to each other. As a matter of fact, we can write

$$|D(\hat{V}_s\|W_1|P_{1,2}) - D(V_s\|W_1|P_{1,2})|$$

$$\leq \sum_{a,b} P_{1,2}(a,b) \sum_y \left| \hat{V}_s(y|a,b) \log \frac{\hat{V}_s(y|a,b)}{W(y|a)} \right.$$
$$\left. - V_s(y|a,b) \log \frac{V_s(y|a,b)}{W(y|a)} \right|. \tag{94}$$

Now for each a, b and y there are two possibilities: if $V_s(y|a,b) > 0$ and $\hat{V}_s(y|a,b) = 0$, then thanks to Lemma 3 we have

$$\left| \hat{V}_s(y|a,b) \log \frac{\hat{V}_s(y|a,b)}{W(y|a)} - V_s(y|a,b) \log \frac{V_s(y|a,b)}{W(y|a)} \right|$$

$$= V_s(y|a,b) \left| \log \frac{V_s(y|a,b)}{W(y|a)} \right| \tag{95}$$

$$\leq \frac{1}{nP_{1,2}(a,b)} \left| \log \frac{V_s(y|a,b)}{W(y|a)} \right|, \tag{96}$$

where the term in absolute value is finite and independent of n. If instead both $\hat{V}_s(y|a,b)$ and $V_s(y|a,b)$ are positive, then we can apply Lagrange's mean value theorem to the function

$$g(x) = x \log \frac{x}{W(y|a)}, \tag{97}$$

whose derivative is

$$g'(x) = \log \frac{x}{W(y|x)} + 1, \tag{98}$$

to get

$$\left| \hat{V}_s(y|a,b) \log \frac{\hat{V}_s(y|a,b)}{W(y|a)} - V_s(y|a,b) \log \frac{V_s(y|a,b)}{W(y|a)} \right|$$

$$\leq |\hat{V}_s(y|a,b) - V_s(y|a,b)| \left| \log \frac{\bar{V}(y|a,b)}{W(y|a)} + 1 \right| \tag{99}$$

$$\leq \frac{1}{nP_{1,2}(a,b)} \left(1 + |\log W(y|a)| + |\log \bar{V}(y|a,b)| \right) \tag{100}$$

for some $\bar{V}(y|a,b) \in \left(V_s(y|a,b), \hat{V}_s(y|a,b) \right)$ (the interval endpoints might be switched).

Notice that both $\hat{V}_s(y|a,b)$ and $\bar{V}(y|a,b)$ depend implicitly on n. In order to make this dependence explicit, we study two cases. If $V_s(y|a,b) < \hat{V}_s(y|a,b)$, then $V_s(y|a,b) < \bar{V}(y|a,b)$ and

$$\left| \hat{V}_s(y|a,b) \log \frac{\hat{V}_s(y|a,b)}{W(y|a)} - V_s(y|a,b) \log \frac{V_s(y|a,b)}{W(y|a)} \right|$$

$$\leq \frac{1}{nP_{1,2}(a,b)} \left(1 + |\log W(y|a)| + |\log V_s(y|a,b)| \right). \tag{101}$$

If instead $\hat{V}_s(y|a,b) < V_s(y|a,b)$, then $\frac{1}{n} \leq \hat{V}_s(y|a,b) < \bar{V}(y|a,b)$ and

$$\left| \hat{V}_s(y|a,b) \log \frac{\hat{V}_s(y|a,b)}{W(y|a)} - V_s(y|a,b) \log \frac{V_s(y|a,b)}{W(y|a)} \right|$$

$$\leq \frac{1}{nP_{1,2}(a,b)} \left(1 + \log n + |\log W(y|a)| \right). \tag{102}$$

Putting it all together, for n large enough we have

$$|D(\hat{V}_s\|W_1|P_{1,2}) - D(V_s\|W_1|P_{1,2})|$$

$$\leq \frac{1}{n} \sum_{a,b,y} \left(1 + \log n + |\log W(y|a)| \right) \tag{103}$$

$$\leq \frac{|\mathcal{X}|^2|\mathcal{Y}|}{n} \left(1 + \log n + \log \frac{1}{W_{\min}} \right) \triangleq \delta_2(n) \tag{104}$$

that again goes to 0 as $n \to \infty$. Hence, equations (81), (93) and (104) lead to

$$P_{e,1}^{(n)} \geq e^{-n[D(\hat{V}_s\|W_1|P_{1,2})+\delta_1(n)]} \tag{105}$$

$$\geq e^{-nD(V_s\|W_1|P_{1,2})-n\delta_1(n)-n\delta_2(n)} \tag{106}$$

$$\geq e^{-\mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)+s\mu'_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s)-\delta(n)}, \tag{107}$$

which concludes the proof. $\square$

Notice that Theorem 2 holds for arbitrary tie-breaking rules. The following corollary, instead, holds only under the assumption that ties are broken equiprobably (or in the case where ties are always counted as errors).

Corollary 2. *If ties are broken equiprobably, then:*

$$P_{e,1}^{(n)} \geq \exp\left\{ -\sup_{s \geq 0} \mu_{\boldsymbol{x}_1,\boldsymbol{x}_2}(s) - \delta(n) \right\} \tag{108}$$

$$P_{e,2}^{(n)} \geq \exp\left\{ -\sup_{s \geq 0} \mu_{\boldsymbol{x}_2,\boldsymbol{x}_1}(s) - \delta(n) \right\}. \tag{109}$$

Proof. We prove again only the bound for $P_{e,1}^{(n)}$. There are three possibilities:

1) $\lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = +\infty$;
2) $\lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \in (-\infty, +\infty)$;
3) $\lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = -\infty$.

In the first case, we have $\sup_{s \geq 0} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = +\infty$ and the bound simply becomes $P_{e,1}^{(n)} \geq 0$, which is trivial.

In the second case, due to the concavity of $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$, we have

$$\sup_{s \geq 0} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = \lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s). \tag{110}$$

Since the limit is a finite real number, then from the definition of $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ in (10) we can deduce that for all sequences $\boldsymbol{y} \in \hat{\mathcal{Y}}^n$ such that $W^n(\boldsymbol{y}|\boldsymbol{x}_1) > 0$, that is, for all sequences $\boldsymbol{y}$ that can possibly contribute to $P_{e,1}^{(n)}$, we must have $\frac{q^n(\boldsymbol{x}_2, \boldsymbol{y})}{q^n(\boldsymbol{x}_1, \boldsymbol{y})} \leq 1$, or equivalently, $q^n(\boldsymbol{x}_2, \boldsymbol{y}) \leq q^n(\boldsymbol{x}_1, \boldsymbol{y})$. Since all sequences such that $q^n(\boldsymbol{x}_2, \boldsymbol{y}) < q^n(\boldsymbol{x}_1, \boldsymbol{y})$ do not contribute to $P_{e,1}^{(n)}$, this means that all sequences that appear in the sum (64) are those that satisfy $q^n(\boldsymbol{x}_1, \boldsymbol{y}) = q^n(\boldsymbol{x}_2, \boldsymbol{y})$. Hence, in this case we can write

$$P_{e,1}^{(n)} = \sum_{\boldsymbol{y} \in \hat{\mathcal{Y}}_{\mathrm{t}}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_1), \tag{111}$$

where

$$\hat{\mathcal{Y}}_{\mathrm{t}}^n = \{\boldsymbol{y} \in \hat{\mathcal{Y}}^n : q^n(\boldsymbol{x}_1, \boldsymbol{y}) = q^n(\boldsymbol{x}_2, \boldsymbol{y})\}. \tag{112}$$

But in this particular case we also have

$$\lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) = \lim_{s \to \infty} -\log \sum_{\boldsymbol{y} \in \hat{\mathcal{Y}}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_1) \left(\frac{q^n(\boldsymbol{x}_2, \boldsymbol{y})}{q^n(\boldsymbol{x}_1, \boldsymbol{y})} \right)^s \tag{113}$$

$$= -\log \sum_{\boldsymbol{y} \in \hat{\mathcal{Y}}_{\mathrm{t}}^n} W^n(\boldsymbol{y}|\boldsymbol{x}_1) = -\log P_{e,1}^{(n)}, \tag{114}$$

or equivalently,

$$P_{e,1}^{(n)} = \exp\left\{ -\lim_{s \to \infty} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \right\} = \exp\left\{ -\sup_{s \geq 0} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \right\}. \tag{115}$$

In the third case, let $\hat{s} = \arg\max_{s \in \mathbb{R}} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$. If $\hat{s} \geq 0$, then thanks to the continuity of $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ and its derivative, we can apply Theorem 2 for $s \to \hat{s}$, so that $\mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \to \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(\hat{s}) = \sup_{s \geq 0} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s)$ and $\mu'_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s) \to 0$. If instead $\hat{s} < 0$, we can apply Theorem 2 with $s = 0$. $\square$

Corollary 2 leads to the fact that

$$P_e^{(n)} \geq \exp\left\{ -n D_{\boldsymbol{x}_1, \boldsymbol{x}_2}^{(n)} + o(n) \right\}, \tag{116}$$

where

$$D_{\boldsymbol{x}_1, \boldsymbol{x}_2}^{(n)} = \min\left\{ \sup_{s \geq 0} \frac{1}{n} \mu_{\boldsymbol{x}_1, \boldsymbol{x}_2}(s), \sup_{s \geq 0} \frac{1}{n} \mu_{\boldsymbol{x}_2, \boldsymbol{x}_1}(s) \right\}. \tag{117}$$

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

$$P_{e,m}^{(n)} \geq \exp\left\{ -n D_{\min}(\mathcal{C}) + o(n) \right\}, \tag{118}$$

where

$$D_{\min}(\mathcal{C}) \triangleq \min_{m \neq m' \in \mathcal{C}} D_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}^{(n)}, \tag{119}$$

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

$$P_e^{(n)} \geq \frac{P_{e,m}^{(n)}}{M} \geq \exp\left\{ -n\left(D_{\min}(\mathcal{C}) + R + o(1) \right) \right\}. \tag{120}$$

IV. UPPER BOUND ON THE RELIABILITY FUNCTION

Equation (120) shows that the problem of upper-bounding the exponent of the probability of error reduces to upper-bounding $D_{\min}(\mathcal{C})$. Specifically, for any rate $R > 0$, the number of codewords M of every code of rate R goes to infinity when the blocklength n goes to infinity; hence, if we can find an upper bound on $D_{\min}(\mathcal{C})$ which is valid for all codes whose size M tends to infinity, then thanks to (120), that bound will also be a valid bound on

$$E(0^+) = \lim_{R \to 0} \limsup_{n \to \infty} -\frac{\log P_e(R, n)}{n}. \tag{121}$$

Before going into the formal details, we give a brief outline of the proof of our bound and the intuition behind it. First of all, the minimum distance $D_{\min}(\mathcal{C})$ of any code $\mathcal{C}$ can be upper-bounded by the minimum distance of any subcode extracted from $\mathcal{C}$. Furthermore, the minimum distance $D_{\min}(\mathcal{C})$ is upper-bounded by the average distance $D_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}^{(n)}$ over all pairs of codewords in $\mathcal{C}$. Therefore, one natural way to upper bound the minimum distance of a code is to upper bound the average distance over a carefully selected subcode, that is,

$$D_{\min}(\mathcal{C}) \leq D_{\min}(\hat{\mathcal{C}}) \leq \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m \neq m' \in \hat{\mathcal{C}}} D_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}^{(n)} \tag{122}$$

$$= \frac{1}{\hat{M}(\hat{M}-1)} \sum_{m \neq m' \in \hat{\mathcal{C}}} \min\left\{ \sup_{s \geq 0} \frac{1}{n} \mu_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}(s), \sup_{s \geq 0} \frac{1}{n} \mu_{\boldsymbol{x}_{m'}, \boldsymbol{x}_m}(s) \right\} \tag{123}$$

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

The choice of the subcode $\hat{\mathcal{C}}$ is crucial, since, in general, the average in (123) may be too difficult to evaluate, for two reasons: for two generic codewords $\boldsymbol{x}_m$ and $\boldsymbol{x}_{m'}$, the functions $\mu_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}(s)$ and $\mu_{\boldsymbol{x}_{m'}, \boldsymbol{x}_m}(s)$ 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. Luckily, we are able to overcome both these difficulties thanks to the following result, which is essentially by Komlós [14], and that was first employed in a coding setting by Blinovsky [15], in the maximum likelihood case. We first state the following fundamental lemma. In order to lighten the notation, from now on, in all subscripts, a pair of codewords $\boldsymbol{x}_m, \boldsymbol{x}_{m'}$ will be denoted just by m, m'.

Lemma 4 (Komlós [14]). *Consider a code $\mathcal{C}$ with M codewords. If for each pair $(a, b) \in \mathcal{X}^2$ there exists a number $r_{a,b}$ such that for all $m < m'$,*

$$\left| P_{m,m'}(a, b) - r_{a,b} \right| \leq \delta, \tag{124}$$

then, for all $m \neq m'$ and $(a,b) \in \mathcal{X}^2$,

$$\left| P_{m,m'}(a,b) - P_{m,m'}(b,a) \right| \leq \frac{6}{\sqrt{M}} + 4\sqrt{\delta} + 4\delta. \quad (125)$$

Then, using Ramsey's theorem on the edge coloring of graphs (see for example [18]), the following result can be proved.

Theorem 3. *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 $(a,b) \in \mathcal{X}^2$,*

$$\left| P_{m,m'}(a,b) - P_{\bar{m},\bar{m}'}(a,b) \right| \leq \Delta(\hat{M}, t), \quad (126)$$

where

$$\Delta(\hat{M}, t) \triangleq \frac{6}{\sqrt{\hat{M}}} + 2\sqrt{\frac{2}{t}} + \frac{3}{t}. \quad (127)$$

A proof of the two previous results in the more general list-decoding setting can be found in [19]. Komlós' result shows that for any positive integer $\hat{M}$, any code with an appropriately large number of codewords contains a subcode of size $\hat{M}$, whose codewords satisfy certain symmetry properties, namely:

(i) all pairs of codewords have approximately the same joint type;

(ii) the joint types are also approximately symmetrical, that is, $P(a,b) \simeq P(b,a)$ for all a, b.

Thanks to property (16), the fact that all pairs of codewords 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 (123), 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 $D_{\boldsymbol{x}_m, \boldsymbol{x}_{m'}}^{(n)}$ 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 $\bar{C}_0^q = 0$ and for every pair $(a,b) \in \mathcal{X}^2$ belonging to the set

$$\mathcal{B} \triangleq \left\{ (a,b) \in \mathcal{X}^2 : \min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} \right\} \quad (128)$$

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

$$\frac{q(a,y)}{q(b,y)} = B(a,b) \quad (129)$$

for all $y \in \hat{\mathcal{Y}}_{a,b}$ such that $W(y|a) + W(y|b) > 0$.

Notice that all channels and decoding metrics such that $\bar{C}_0^q = 0$ and

$$W(y|x) > 0 \iff q(x,y) > 0 \quad (130)$$

are balanced pairs, and indeed represent a very important special case. To see this, consider a channel-metric pair
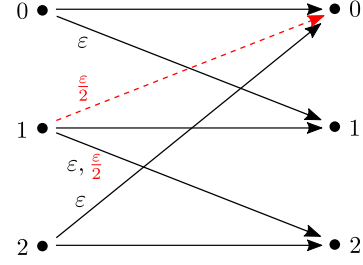

Fig. 1. Unbalanced channel-decoding metric pair of Example 1.

satisfying (130); for any $(a,b) \in \mathcal{B}$, we can partition the set of possible output symbols in $\hat{\mathcal{Y}}_{a,b}$ into three subsets:

$$\mathcal{S}_a = \{y : W(y|a) > 0 \quad \text{and} \quad W(y|b) = 0\} \quad (131)$$
$$\mathcal{S}_b = \{y : W(y|a) = 0 \quad \text{and} \quad W(y|b) > 0\} \quad (132)$$
$$\mathcal{S}_{ab} = \{y : W(y|a) > 0 \quad \text{and} \quad W(y|b) > 0\}. \quad (133)$$

For all $y \in \mathcal{S}_a$, $q(a,y) > 0$ and $q(b,y) = 0$, therefore $q(a,y)/q(b,y) = +\infty$. Similarly, $q(a,y)/q(b,y) = 0$ for all $y \in \mathcal{S}_b$. Hence, we have that

$$\min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} = \min_{y:W(y|a)W(y|b)>0} \frac{q(a,y)}{q(b,y)} \quad (134)$$

$$\max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|a)W(y|b)>0} \frac{q(a,y)}{q(b,y)}, \quad (135)$$

and since $(a,b) \in \mathcal{B}$, these two quantities must be equal, that is,

$$\min_{y:W(y|a)W(y|b)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|a)W(y|b)>0} \frac{q(a,y)}{q(b,y)}, \quad (136)$$

which means that the ratio $q(a,y)/q(b,y)$ must be equal for all possible $y \in \hat{\mathcal{Y}}_{a,b}$. This proves that the channel-metric pair is indeed balanced. Furthermore, for this particular subclass,

$$C_0 = 0 \iff C_0^q = 0 \iff \bar{C}_0^q = 0, \quad (137)$$

where C_0 is the classical zero-error capacity. An example of a non-balanced channel-decoding metric pair is the following.

Example 1. Consider the three-input typewriter channel with $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$ and crossover probabilities $W(1|0) = W(2|1) = W(0|2) = \varepsilon$, with $0 < \varepsilon < 2 - \sqrt{2}$. Furthermore, consider a decoding metric such that $q(a,y) = W(y|a)$ for all a and y with the exception of $q(1,0) = q(1,2) = \frac{\varepsilon}{2}$ (Fig. 1).

As one can check, this channel-decoding metric satisfies the condition for $\bar{C}_0^q = 0$ in Theorem 1, since for all $a, b \in \mathcal{X}$,

$$\min_{y:W(y|a)>0} \frac{q(a,y)}{q(b,y)} = \max_{y:W(y|b)>0} \frac{q(a,y)}{q(b,y)}. \quad (138)$$

Notice that for this channel also the classical zero-error capacity C_0 is zero. However, this channel-decoding metric pair does not satisfy the second condition in Definition 1 for a balanced pair, since

$$\frac{q(0,0)}{q(1,0)} = \frac{2(1-\varepsilon)}{\varepsilon} \neq \frac{q(0,1)}{q(1,1)} = \frac{\varepsilon}{1-\varepsilon}. \quad (139)$$

One last lemma that will be useful in bounding the average in (123) is the following, which is a standard trick employed, for example, in the derivation of the Plotkin bound.

Lemma 5. *For any code with $\hat{M}$ codewords of blocklength n, for any $a, b \in \mathcal{X}$, with $a \neq b$,*

$$\sum_{m \neq m'} P_{m,m'}(a,b) = \frac{1}{n} \sum_{c=1}^{n} \hat{M}_c(a) \hat{M}_c(b), \quad (140)$$

where $\hat{M}_c(a)$ is the number of times the symbol a occurs in the coordinate c in all the codewords.

Proof. Imagine the code as an $\hat{M} \times n$ matrix, with each codeword as a row. Then, $\sum_{m \neq m'} n P_{m,m'}(a,b)$ is the number of times the pair (a,b) can be found by selecting any two entries of the matrix belonging to the same column. The same computation can be performed column by column: for a generic column c, that number is simply the number of times a occurs in that column, multiplied by the number of times b occurs. Finally, summing over all columns returns the same number of the first computation, thus proving the lemma. $\square$

We are ready to proceed and prove the upper bound on the reliability function at $R = 0^+$ for any balanced channel-metric pair. In particular, we will show that for this class, the $D_{m,m'}^{(n)}$ are all close to each other for all pairs of codewords in the symmetric subcode $\hat{\mathcal{C}} \subset \mathcal{C}$, whose existence is guaranteed by Theorem 3. In order to show this, first of all, for any concave function $f(s)$, let[2]

$$\mathcal{S} \triangleq \left\{ 0 \leq s \leq +\infty : f(s) = \sup_{s \geq 0} f(s) \right\} \quad (141)$$

and define

$$\arg\sup_{s \geq 0} f(s) \triangleq \inf \mathcal{S}. \quad (142)$$

In the following lemma we prove that in the case of balanced pairs, for all pairs of codewords, the concave functions $\mu_{m,m'}(s) + \mu_{m',m}(s)$ achieve their suprema at an s in a bounded interval, determined only by the channel and the decoding metric.

Lemma 6. *For any balanced pair, for any pair of codewords m, m',*

$$\arg\sup_{s \geq 0} \left(\mu_{m,m'}(s) + \mu_{m',m}(s) \right) \in [0, \hat{s}], \quad (143)$$

where

$$\hat{s} \triangleq \max_{a,b} \left\{ \arg\sup_{s \geq 0} \left(\mu_{a,b}(s) + \mu_{b,a}(s) \right) \right\} < +\infty. \quad (144)$$

Proof. We first show that $\hat{s}$ is finite. We already pointed out in the proof of Theorem 1, that equation (35) can be rewritten as

$$\mu'_{a,b} + \mu'_{b,a} \leq 0. \quad (145)$$

For the $(a,b) \in \mathcal{X}^2$ such that $\mu'_{a,b} + \mu'_{b,a} < 0$, we have that

$$\lim_{s \to +\infty} \mu_{a,b}(s) + \mu_{b,a}(s) = -\infty, \quad (146)$$

which in turn implies that there exists a finite $\hat{s}_{a,b} \geq 0$ such that

$$\mu_{a,b}(\hat{s}_{a,b}) + \mu_{b,a}(\hat{s}_{a,b}) = \max_{s \geq 0} \left(\mu_{a,b}(s) + \mu_{b,a}(s) \right) \quad (147)$$

---

[2]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.

since $\mu_{a,b}(s) + \mu_{b,a}(s)$ is concave. For the $(a,b) \in \mathcal{X}^2$ such that $\mu'_{a,b} + \mu'_{b,a} = 0$, instead, equation (129) implies that

$$\mu_{a,b}(s) = -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) \left(\frac{q(b,y)}{q(a,y)} \right)^s \quad (148)$$

$$= -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) B(a,b)^{-s} \quad (149)$$

$$= s \log B(a,b) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a), \quad (150)$$

which is a straight line. Furthermore, since $B(b,a) = 1/B(a,b)$, we have that

$$\mu_{a,b}(s) + \mu_{b,a}(s)$$
$$= s \log B(a,b) B(b,a) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a)$$
$$\qquad\qquad - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|b) \quad (151)$$

$$= -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|b), \quad (152)$$

which is a constant. Hence,

$$\sup_{s \geq 0} \left(\mu_{a,b}(s) + \mu_{b,a}(s) \right) = \mu_{a,b}(0) + \mu_{b,a}(0) \quad (153)$$

and we can set $\hat{s}_{a,b} = 0$ for these (a,b). Thus, the $\hat{s}$ defined by (144) can be rewritten as

$$\hat{s} = \max_{a,b} \hat{s}_{a,b}, \quad (154)$$

which is finite, since all $\hat{s}_{a,b}$ are finite.

Equation (143) follows from the fact that

$$\mu'_{m,m'}(\hat{s}) + \mu'_{m',m}(\hat{s})$$
$$= n \sum_a \sum_b P_{m,m'}(a,b) \left(\mu'_{a,b}(\hat{s}) + \mu'_{b,a}(\hat{s}) \right) \leq 0, \quad (155)$$

where the equality is due to (16), while the inequality is due to the fact that for all (a,b),

$$\mu'_{a,b}(\hat{s}) + \mu'_{b,a}(\hat{s}) \leq \mu'_{a,b}(\hat{s}_{a,b}) + \mu'_{b,a}(\hat{s}_{a,b}) \leq 0, \quad (156)$$

since $\hat{s}_{a,b} \leq \hat{s}$ and $\mu_{a,b}(s) + \mu_{b,a}(s)$ is concave. $\square$

The previous lemma and the symmetry properties of the codewords in $\hat{\mathcal{C}}$ lead to the following fundamental result, that shows that the $D_{m,m'}^{(n)}$ 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 (123) possible.

Lemma 7. *For any balanced pair, for any pair of codewords $m, m' \in \hat{\mathcal{C}}$, let*

$$\bar{s}_{m,m'} \triangleq \min \left\{ \arg\sup_{s \geq 0} \mu_{m,m'}(s), \arg\sup_{s \geq 0} \mu_{m',m}(s) \right\}. \quad (157)$$

Then, $0 \leq \bar{s}_{m,m'} \leq \hat{s}$, with $\hat{s}$ defined by (144), and

$$D_{m,m'}^{(n)} \leq \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) + K\Delta(\hat{M}, t) \quad (158)$$

with $\Delta(\hat{M}, t)$ as defined by (127), and

$$K \triangleq \max_{0 \le s \le \hat{s}} \sum_a \sum_b |\mu_{a,b}(s)|. \tag{159}$$

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

$$\left| \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right| \le 4K\Delta(\hat{M}, t). \tag{160}$$

Proof. To prove that $\bar{s}_{m,m'} \le \hat{s}$, notice that from equation (143) we get

$$\mu'_{m,m'}(\hat{s}) + \mu'_{m',m}(\hat{s}) \le 0, \tag{161}$$

which is possible only if

$$\mu'_{m,m'}(\hat{s}) \le 0 \quad \text{or} \quad \mu'_{m',m}(\hat{s}) \le 0, \tag{162}$$

which in turn implies that

$$\arg\sup_{s \ge 0} \mu_{m,m'}(s) \le \hat{s} \quad \text{or} \quad \arg\sup_{s \ge 0} \mu_{m',m}(s) \le \hat{s}. \tag{163}$$

This proves that

$$\bar{s}_{m,m'} \triangleq \min \left\{ \arg\sup_{s \ge 0} \mu_{m,m'}(s), \arg\sup_{s \ge 0} \mu_{m',m}(s) \right\} \le \hat{s}. \tag{164}$$

Next, definition (157) implies that

$$\sup_{s \ge 0} \mu_{m,m'}(s) = \mu_{m,m'}(\bar{s}_{m,m'}) \tag{165}$$

or

$$\sup_{s \ge 0} \mu_{m',m}(s) = \mu_{m',m}(\bar{s}_{m,m'}). \tag{166}$$

Hence, we have that, thanks to (117),

$$D_{m,m'}^{(n)} \le \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) \quad \text{or} \quad D_{m,m'}^{(n)} \le \frac{1}{n} \mu_{m',m}(\bar{s}_{m,m'}). \tag{167}$$

In the first case, equation (158) follows immediately; in the second case, we have that

$$\left| \frac{1}{n} \mu_{m',m}(\bar{s}_{m,m'}) - \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) \right|$$
$$= \left| \sum_a \sum_b \left(P_{m,m'}(a,b) - P_{m,m'}(b,a) \right) \mu_{a,b}(\bar{s}_{m,m'}) \right| \tag{168}$$

$$\le \sum_a \sum_b |P_{m,m'}(a,b) - P_{m',m}(a,b)| |\mu_{a,b}(\bar{s}_{m,m'})| \tag{169}$$

$$\le \Delta(\hat{M}, t) \sum_a \sum_b |\mu_{a,b}(\bar{s}_{m,m'})| \tag{170}$$

$$\le \Delta(\hat{M}, t) \max_{0 \le s \le \hat{s}} \sum_a \sum_b |\mu_{a,b}(s)| \tag{171}$$

$$= K\Delta(\hat{M}, t) \tag{172}$$

and therefore,

$$D_{m,m'}^{(n)} \le \frac{1}{n} \mu_{m',m}(\bar{s}_{m,m'}) \le \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) + K\Delta(\hat{M}, t). \tag{173}$$

Finally, in order to prove (160), first notice that

$$\left| \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right|$$
$$= \frac{1}{n} \left| \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \mu_{m,m'}(\bar{s}_{m,m'}) \right.$$
$$\left. + \mu_{m,m'}(\bar{s}_{m,m'}) - \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right| \tag{174}$$

$$\le \frac{1}{n} \left| \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \mu_{m,m'}(\bar{s}_{m,m'}) \right|$$
$$+ \frac{1}{n} \left| \mu_{m,m'}(\bar{s}_{m,m'}) - \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right|. \tag{175}$$

The second absolute value can be bounded as follows:

$$\frac{1}{n} \left| \mu_{m,m'}(\bar{s}_{m,m'}) - \mu_{\bar{m},\bar{m}'}(\bar{s}_{m,m'}) \right|$$
$$= \left| \sum_a \sum_b \left(P_{m,m'}(a,b) - P_{\bar{m},\bar{m}'}(a,b) \right) \mu_{a,b}(\bar{s}_{m,m'}) \right| \tag{176}$$

$$\le \sum_a \sum_b |P_{m,m'}(a,b) - P_{\bar{m},\bar{m}'}(a,b)| |\mu_{a,b}(\bar{s}_{m,m'})| \tag{177}$$

$$\le \Delta(\hat{M}, t) \sum_a \sum_b |\mu_{a,b}(\bar{s}_{m,m'})| \tag{178}$$

$$\le K\Delta(\hat{M}, t), \tag{179}$$

which also holds for every $0 \le s \le \hat{s}$. The first absolute value, instead, can be bounded in the following way. Suppose that

$$\frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) \ge \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}); \tag{180}$$

the other case can be proved in the same way. Then, we can write that

$$\frac{1}{n} \left| \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \mu_{m,m'}(\bar{s}_{m,m'}) \right|$$
$$= \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}). \tag{181}$$

Furthermore, thanks to (165) and (166), we have two alternatives. If

$$\sup_{s \ge 0} \mu_{m,m'}(s) = \mu_{m,m'}(\bar{s}_{m,m'}),$$

then

$$\frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) \ge \frac{1}{n} \mu_{m,m'}(\bar{s}_{\bar{m},\bar{m}'}) \tag{182}$$

and we can bound (181) by

$$\frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'})$$
$$\le \frac{1}{n} \mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n} \mu_{m,m'}(\bar{s}_{\bar{m},\bar{m}'}) \tag{183}$$

$$\le K\Delta(\hat{M}, t) \tag{184}$$

as in (179). If instead

$$\sup_{s \ge 0} \mu_{m',m}(s) = \mu_{m',m}(\bar{s}_{m,m'}),$$

then

$$\frac{1}{n} \mu_{m,m'}(\bar{s}_{m,m'}) \ge \frac{1}{n} \mu_{m',m}(\bar{s}_{m,m'}) - K\Delta(\hat{M}, t) \tag{185}$$

$$\ge \frac{1}{n} \mu_{m',m}(\bar{s}_{\bar{m},\bar{m}'}) - K\Delta(\hat{M}, t) \tag{186}$$

$$\ge \frac{1}{n} \mu_{m,m'}(\bar{s}_{\bar{m},\bar{m}'}) - 2K\Delta(\hat{M}, t) \tag{187}$$

using (172) twice. Hence, we can bound (181) by

$$\frac{1}{n}\,\mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n}\,\mu_{m,m'}(\bar{s}_{m,m'})$$

$$\leq \frac{1}{n}\,\mu_{\bar{m},\bar{m}'}(\bar{s}_{\bar{m},\bar{m}'}) - \frac{1}{n}\,\mu_{m,m'}(\bar{s}_{\bar{m},\bar{m}'}) + 2K\Delta(\hat{M},t) \tag{188}$$

$$\leq 3K\Delta(\hat{M},t), \tag{189}$$

again as in (179). Putting this and (179) into (175) leads to (160). □

Finally, thanks to this lemma, we can prove our upper bound on the reliability function at $R = 0^+$, which coincides with the lower bound (9).

Theorem 4. *For any balanced pair,*

$$E^q(0^+) =$$

$$\max_{Q\in\mathcal{P}(\mathcal{X})}\sup_{s\geq 0} -\sum_{a\in\mathcal{X}}\sum_{b\in\mathcal{X}} Q(a)Q(b)\log\sum_{y\in\hat{\mathcal{Y}}_{a,b}} W(y|a)\left(\frac{q(b,y)}{q(a,y)}\right)^s. \tag{190}$$

Proof. We already pointed out that for any subcode of $\mathcal{C}$, and in particular for the subcode $\hat{\mathcal{C}}$ of Theorem 3, we have

$$D_{\min}(\mathcal{C}) \leq D_{\min}(\hat{\mathcal{C}}) \leq \frac{1}{\hat{M}(\hat{M}-1)}\sum_{m\neq m'} D^{(n)}_{m,m'} \tag{191}$$

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 maximum likelihood setting for the particular class of pairwise reversible channels [20]. Fix any pair of codewords $\hat{m} \neq \hat{m}' \in \hat{\mathcal{C}}$. Then,

$$D_{\min}(\hat{\mathcal{C}})$$

$$\leq \frac{1}{\hat{M}(\hat{M}-1)}\sum_{m\neq m'} D^{(n)}_{m,m'} \tag{192}$$

$$\leq \frac{1}{\hat{M}(\hat{M}-1)}\sum_{m\neq m'}\frac{1}{n}\mu_{m,m'}(\bar{s}_{m,m'}) + K\Delta(\hat{M},t) \tag{193}$$

$$\leq \frac{1}{\hat{M}(\hat{M}-1)}\sum_{m\neq m'}\frac{1}{n}\mu_{m,m'}(\bar{s}_{\hat{m},\hat{m}'}) + 5K\Delta(\hat{M},t) \tag{194}$$

$$= \frac{1}{\hat{M}(\hat{M}-1)}\sum_a\sum_b\sum_{m\neq m'} P_{m,m'}(a,b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'})$$
$$+ 5K\Delta(\hat{M},t) \tag{195}$$

$$= \frac{1}{n}\frac{1}{\hat{M}(\hat{M}-1)}\sum_{c=1}^n\sum_a\sum_b \hat{M}_c(a)\hat{M}_c(b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'})$$
$$+ 5K\Delta(\hat{M},t) \tag{196}$$

$$= \frac{1}{n}\frac{\hat{M}}{\hat{M}-1}\sum_{c=1}^n\sum_a\sum_b \frac{\hat{M}_c(a)}{\hat{M}}\frac{\hat{M}_c(b)}{\hat{M}}\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'})$$
$$+ 5K\Delta(\hat{M},t) \tag{197}$$

$$\leq \frac{\hat{M}}{\hat{M}-1}\max_{Q\in\mathcal{P}(\mathcal{X})}\sum_a\sum_b Q(a)Q(b)\mu_{a,b}(\bar{s}_{\hat{m},\hat{m}'})$$
$$+ 5K\Delta(\hat{M},t) \tag{198}$$

$$\leq \frac{\hat{M}}{\hat{M}-1}\sup_{s\geq 0}\max_{Q\in\mathcal{P}(\mathcal{X})}\sum_a\sum_b Q(a)Q(b)\mu_{a,b}(s)$$
$$+ 5K\Delta(\hat{M},t), \tag{199}$$

where (193) is due to (158), (194) is due to (160), (196) is due to Lemma 5, and (198) is due to the fact that for every c,

$$\left\{\frac{\hat{M}_c(a)}{\hat{M}}, \quad a\in\mathcal{X}\right\}$$

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 (191) are close to each other. Then, letting $M\to\infty$ (so that we may also let $\hat{M}\to\infty$, by Theorem 3) and $t\to\infty$ we obtain, thanks to the fact that $\Delta(\hat{M},t)\to 0$ in (199),

$$D_{\min}(\mathcal{C}) \leq \sup_{s\geq 0}\max_{Q\in\mathcal{P}(\mathcal{X})}\sum_a\sum_b Q(a)Q(b)\mu_{a,b}(s), \tag{200}$$

which is independent of the code $\mathcal{C}$. Finally, thanks to equation (120), since we let $R\to 0$ *after* $n\to\infty$, we obtain an upper bound on the reliability function at $R = 0^+$:

$$E^q(0^+) \leq$$

$$\max_{Q\in\mathcal{P}(\mathcal{X})}\sup_{s\geq 0} -\sum_a\sum_b Q(a)Q(b)\log\sum_{y\in\hat{\mathcal{Y}}_{a,b}} W(y|a)\left(\frac{q(b,y)}{q(a,y)}\right)^s, \tag{201}$$

which equals the expurgated lower bound given by (9), proving the theorem. □

If the channel and decoding metric are not a balanced pair, our method fails in that for some pair $(a,b)\in\mathcal{X}^2$ belonging to the set $\mathcal{B}$ defined in (128), the function $\mu_{a,b}(s) + \mu_{b,a}(s)$ is concave and has a horizontal asymptote at $s\to +\infty$, but it is not a straight line; because of this, a finite $\hat{s}$ as in Lemma 6 cannot be determined. A partial solution to this problem is to upper-bound these functions by their horizontal asymptote. This strategy leads to a similar upper bound as the one for balanced pairs; however, in this case, the bound is larger than the expurgated bound at $R = 0^+$. To obtain this bound, define for any pair $(a,b)\in\mathcal{B}$,

$$A(a,b) \triangleq \min_{y:W(y|a)>0}\frac{q(a,y)}{q(b,y)} = \max_{y:W(y|b)>0}\frac{q(a,y)}{q(b,y)}, \tag{202}$$

and let

$$\hat{\mathcal{Y}}^A_{a,b} \triangleq \left\{y\in\hat{\mathcal{Y}}_{a,b} : \frac{q(a,y)}{q(b,y)} = A(a,b)\right\}. \tag{203}$$

Then, we can upper-bound $\mu_{a,b}(s)$ and $\mu_{b,a}(s)$ as follows.

$$\mu_{a,b}(s) \triangleq -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a)\left(\frac{q(b,y)}{q(a,y)}\right)^s \quad (204)$$

$$\leq -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|a)A(a,b)^{-s} \quad (205)$$

$$= s \log A(a,b) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|a) \quad (206)$$

and in the same way,

$$\mu_{b,a}(s) \leq -s \log A(a,b) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|b). \quad (207)$$

Now, if we define

$$\hat{\mu}_{a,b}(s) \triangleq s \log A(a,b) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|a) \quad (208)$$

$$\hat{\mu}_{b,a}(s) \triangleq -s \log A(a,b) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|b) \quad (209)$$

we have $\mu_{a,b}(s) \leq \hat{\mu}_{a,b}(s)$ and $\mu_{b,a}(s) \leq \hat{\mu}_{b,a}(s)$, and

$$\hat{\mu}_{a,b}(s) + \hat{\mu}_{b,a}(s) = -\log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|a) - \log \sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|b) \quad (210)$$

which is constant. Finally, if we set $\hat{\mu}_{a,b}(s) \triangleq \mu_{a,b}(s)$ for all pairs $(a,b) \in \mathcal{B}^c$, one can readily check that Lemma 6, Lemma 7 and Theorem 4 (the upper bound part) still hold for any discrete memoryless channel and decoding metric if the $\mu_{a,b}(s)$ are replaced with $\hat{\mu}_{a,b}(s)$. Hence, for a generic pair of channel and decoding metric, the following theorem can be proved.

Theorem 5. *For any discrete memoryless channel and decoding metric with $\bar{C}_0 = 0$,*

$$E^q(0^+) \leq \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_a \sum_b Q(a)Q(b)\hat{\mu}_{a,b}(s) \triangleq E_{\text{up}}^q(0^+). \quad (211)$$

In such a case, the maximum distance between the expurgated lower bound and our upper bound on $E^q(0^+)$ can be estimated as follows:

$$\left| E_{\text{up}}^q(0^+) - E_{\text{ex}}^q(0^+) \right| \quad (212)$$

$$\leq \frac{1}{2} \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_a \sum_b Q(a)Q(b)$$

$$\left| \hat{\mu}_{a,b}(s) + \hat{\mu}_{b,a}(s) - \mu_{a,b}(s) - \mu_{b,a}(s) \right| \quad (213)$$

$$= \frac{1}{2} \max_{Q \in \mathcal{P}(\mathcal{X})} \sup_{s \geq 0} \sum_{(a,b) \in \mathcal{B}} Q(a)Q(b)$$

$$\left| \hat{\mu}_{a,b}(s) + \hat{\mu}_{b,a}(s) - \mu_{a,b}(s) - \mu_{b,a}(s) \right| \quad (214)$$

$$= \frac{1}{2} \max_{Q \in \mathcal{P}(\mathcal{X})} \sum_{(a,b) \in \mathcal{B}} Q(a)Q(b)$$

$$\left| \hat{\mu}_{a,b}(0) + \hat{\mu}_{b,a}(0) - \mu_{a,b}(0) - \mu_{b,a}(0) \right| \quad (215)$$

$$\leq \frac{1}{2} \max_{(a,b) \in \mathcal{B}} \left| \hat{\mu}_{a,b}(0) + \hat{\mu}_{b,a}(0) - \mu_{a,b}(0) - \mu_{b,a}(0) \right| \quad (216)$$

$$= \frac{1}{2} \max_{(a,b) \in \mathcal{B}} \left(\log \frac{\sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|a)}{\sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|a)} + \log \frac{\sum_{y \in \hat{\mathcal{Y}}_{a,b}} W(y|b)}{\sum_{y \in \hat{\mathcal{Y}}_{a,b}^A} W(y|b)} \right). \quad (217)$$

Notice that for balanced pairs, definitions (129) and (203) show that the sets $\hat{\mathcal{Y}}_{a,b}$ and $\hat{\mathcal{Y}}_{a,b}^A$ are equal, and therefore the quantity in (217) is zero, as expected.

Example 2. Consider the non-balanced channel-decoding metric pair of Example 1. For the pair of inputs $(0,1)$ one has $\hat{\mathcal{Y}}_{a,b} = \{0,1\} \neq \hat{\mathcal{Y}}_{a,b}^A = \{1\}$. Therefore, the upper bound on the gap $\left| E_{\text{up}}^q(0^+) - E_{\text{ex}}^q(0^+) \right|$ in equation (217) for this channel-decoding pair is equal to

$$\left| E_{\text{up}}^q(0^+) - E_{\text{ex}}^q(0^+) \right| \leq \frac{1}{2}\left(\log \frac{1}{\varepsilon} + \log \frac{1-\varepsilon}{1-\varepsilon} \right) = \frac{1}{2}\log\frac{1}{\varepsilon}. \quad (218)$$

Remark. Converse bounds for codes at rate $R = 0$ can often be used to also deduce bounds at $R > 0$ through appropriate code coverings. Our bound, as most of the bounds on zero-rate codes, is based on the Plotkin double counting trick of Lemma 5, which is used in (196). In the same way as the Plotkin bound on the minimum Hamming distance of codes at $R = 0$ can be extended to $R > 0$ to deduce the Singleton bound and the Elias bound, our result can also be applied to derive bounds at $R > 0$. For maximum likelihood decoding, the idea was initially presented by Blahut, although with a technical error in the proof which can however be corrected by means of the Ramsey-theoretic procedure also used here (see [19]). A similar extension can be derived for mismatched decoding. We do not expand on this point here since it would in large part repeat the discussion in [19] while at the same time requiring a significant technical digression on constant composition codes and the maximization over s, which would take us far from the main focus of this paper.

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*, to appear.

[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] C. E. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8-19, 1956.

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

[15] V. M. Blinovsky, "New Approach to Estimation of the Decoding Error Probability," *Prob. Inf. Trans.*, vol. 38, no. 1, pp. 16-19, 2002.

[16] V. M. Blinovsky, "Code bounds for multiple packings over a nonbinary finite alphabet," *Prob. Inf. Trans.*, vol. 41, no. 1, pp. 23-32, 2005.

[17] R. Ahlswede, V. Blinovsky, "Multiple packing in sum-type metric spaces," *Discrete Applied Mathematics*, vol. 156, no. 9, pp. 1469-1477, 2008.

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

[19] M. Bondaschi and M. Dalai, "A Revisitation of Low-Rate Bounds on the Reliability Function of Discrete Memoryless Channels for List Decoding," *IEEE Trans. Inf. Theory*, submitted.

[20] 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.

Marco Bondaschi is a PhD student in the Laboratory for Information in Networked Systems at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He received the Bachelor's degree in Electronics Engineering and the Master's degree in Communication Sciences and Multimedia from the University of Brescia in 2017 and 2019 respectively. His main research interests are in information theory and learning theory.

Albert Guillén i Fàbregas (S–01, M–05, SM–09, F–22) received the Telecommunications Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively in 1999, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

In 2020, he returned to a full-time faculty position at the Department of Engineering, University of Cambridge, where he had been a full-time faculty and Fellow of Trinity Hall from 2007 to 2012. Since 2011 he has been an ICREA Research Professor at Universitat Pompeu Fabra (currently on leave), where he is now an adjunct researcher. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, Universitat Pompeu Fabra, University of Cambridge, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, statistical inference.

Dr. Guillén i Fàbregas is a Member of the Young Academy of Europe, and received the Starting and Consolidator Grants from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. Since 2013 he has been an Editor of the Foundations and Trends in Communications and Information Theory, Now Publishers and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2020) AND IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).

Marco Dalai (S'05-A'06-M'11-SM'17) received the degree in Electronic Engineering (cum laude) and the PhD in Information Engineering in 2003 and 2007 respectively from the University of Brescia Italy, where he is now an associate professor with the Department of Information Engineering. He is a member of the IEEE Information Theory Society, recipient of the 2014 IEEE Information Theory Society Paper Award and currently an Associate Editor of the IEEE Transactions on Information Theory.