The Meta-Converse Bound is Tight

Gonzalo Vazquez-Vilar[1], Adrià Tauste Campo[1], Albert Guillén i Fàbregas[123] and Alfonso Martinez[1]
[1]Universitat Pompeu Fabra, [2]ICREA, [3]University of Cambridge
Email: {gvazquez,atauste,guillen,alfonso.martinez}@ieee.org

***Abstract*—We show that the meta-converse bound derived by Polyanskiy *et al.* provides the exact error probability for a fixed joint source-channel code and an appropriate choice of the bound parameters. While the expression is not computable in general, it identifies the weaknesses of known converse bounds to the minimum achievable error probability.**

I. Introduction

In the study of reliable communication, the hypothesis-testing method is a useful technique to derive converse bounds on the average error probability [1]–[4]. For channel coding, Polyanskiy *et al.* provided a meta-converse bound [2, Th. 26], which states that the average error probability $\epsilon(\mathcal{C})$ of a channel code $\mathcal{C}$ with M codewords and block length n transmitted over a channel $P_{\boldsymbol{Y}|\boldsymbol{X}}$ satisfies

$$\epsilon(\mathcal{C}) \geq \sup_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{\frac{1}{M}} \left(P_{\boldsymbol{X}}^{(\mathcal{C})} \times P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}^{(\mathcal{C})} \times Q_{\boldsymbol{Y}} \right) \right\}, \quad (1)$$

where $\alpha_\beta(P, Q)$ is the minimum type-I error[1] for a maximum type-II error $\beta \in [0,1]$ for a binary hypothesis test between distributions P and Q; and $P_{\boldsymbol{X}}^{(\mathcal{C})}$ denotes the channel-input distribution induced by the codebook $\mathcal{C}$. This result has been extended to joint source-channel coding in [3] and [4].

In this paper, we show that (1) holds with equality. First, we use the hypothesis-testing method to provide a lower-bound on the error probability for source-channel coding. Then, for a fixed codebook, this bound is shown to be equal to the error probability of a maximum a posteriori (MAP) decoder.

A. System Model and Notation

We consider the transmission of a length-k discrete memoryless source over a discrete memoryless channel (DMC) using length-n block codes. The source is distributed according to $P_{\boldsymbol{V}}(\boldsymbol{v}) = \prod_{i=1}^{k} P_V(v_i)$, $\boldsymbol{v} = (v_1, \ldots, v_k) \in \mathcal{V}^k$, where $\mathcal{V}$ is an alphabet with cardinality $|\mathcal{V}|$. The channel law is given by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are discrete alphabets with cardinalities $|\mathcal{X}|$ and $|\mathcal{Y}|$, respectively.

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

[1]Since the focus of this paper is the error probability, we refer to the minimum type-I error for a maximum type-II error $\beta \in [0,1]$ as $\alpha_\beta(\cdot,\cdot)$. This definition is the counterpart of the function $\beta_{1-\alpha}(\cdot,\cdot)$, defined in [2] as the minimum type-II error for a maximum type-I error $\alpha \in [0,1]$.

An encoder maps the length-k source message $\boldsymbol{v}$ to a length-n codeword $\boldsymbol{x}(\boldsymbol{v})$ using a codebook $\mathcal{C}$ and then, $\boldsymbol{x}(\boldsymbol{v})$ is transmitted over the channel. We consider a MAP decoder that randomly chooses one source message $\boldsymbol{z}$ (decoded message) among the source messages belonging to the set

$$\mathcal{S}_{\mathcal{C}}(\boldsymbol{y}) \triangleq \left\{ \boldsymbol{v} \mid P_{\boldsymbol{V}\boldsymbol{Y}}^{(\mathcal{C})}(\boldsymbol{v}, \boldsymbol{y}) = \max_{\boldsymbol{v}'} P_{\boldsymbol{V}\boldsymbol{Y}}^{(\mathcal{C})}(\boldsymbol{v}', \boldsymbol{y}) \right\}, \quad (2)$$

where we defined the joint distribution $P_{\boldsymbol{V}\boldsymbol{Y}}^{(\mathcal{C})} \triangleq P_{\boldsymbol{V}} \times P_{\boldsymbol{Y}|\boldsymbol{V}}^{(\mathcal{C})}$, where $P_{\boldsymbol{Y}|\boldsymbol{V}}^{(\mathcal{C})}(\boldsymbol{y}|\boldsymbol{v}) \triangleq P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v}))$, $\boldsymbol{y} \in \mathcal{Y}^n$, $\boldsymbol{v} \in \mathcal{V}^k$. This decoding rule is described by the distribution $P_{\boldsymbol{Z}|\boldsymbol{Y}}^{(\mathcal{C})}$,

$$P_{\boldsymbol{Z}|\boldsymbol{Y}}^{(\mathcal{C})}(\boldsymbol{z}|\boldsymbol{y}) \triangleq \begin{cases} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|}, & \text{if } \boldsymbol{z} \in \mathcal{S}_{\mathcal{C}}(\boldsymbol{y}), \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The error probability $\epsilon(\mathcal{C})$ of the code $\mathcal{C}$ can be expressed as

$$\epsilon(\mathcal{C}) \triangleq \Pr\{\boldsymbol{Z} \neq \boldsymbol{V}\} \quad (4)$$

$$= 1 - \sum_{\boldsymbol{v}} \sum_{\boldsymbol{y}} P_{\boldsymbol{V}\boldsymbol{Y}}^{(\mathcal{C})}(\boldsymbol{v}, \boldsymbol{y}) P_{\boldsymbol{Z}|\boldsymbol{Y}}^{(\mathcal{C})}(\boldsymbol{v}|\boldsymbol{y}). \quad (5)$$

II. Hypothesis-Testing Approach

For an observation $(\boldsymbol{v}, \boldsymbol{y})$ we define the hypotheses

$$\mathcal{H}_0: \quad (\boldsymbol{V}, \boldsymbol{Y}) \sim P_{\boldsymbol{V}\boldsymbol{Y}}, \quad (6)$$

$$\mathcal{H}_1: \quad (\boldsymbol{V}, \boldsymbol{Y}) \sim Q_{\boldsymbol{V}\boldsymbol{Y}}, \quad (7)$$

for arbitrary distributions $P_{\boldsymbol{V}\boldsymbol{Y}}$ and $Q_{\boldsymbol{V}\boldsymbol{Y}}$. Any test deciding between these two hypotheses can be defined by a (possibly random) transformation $(\mathcal{V}^k, \mathcal{Y}^n) \to \{\mathcal{H}_0, \mathcal{H}_1\}$ described by the conditional distribution $P_{W|\boldsymbol{V}\boldsymbol{Y}}$.

The performance of a test $P_{W|\boldsymbol{V}\boldsymbol{Y}}$ can be evaluated according to its type-I and type-II errors. The type-I error, the probability of choosing $\mathcal{H}_1$ when the true hypothesis is $\mathcal{H}_0$, is given by

$$\epsilon_{\text{I}}(P_{\boldsymbol{V}\boldsymbol{Y}}, P_{W|\boldsymbol{V}\boldsymbol{Y}}) = \sum_{\boldsymbol{v}} \sum_{\boldsymbol{y}} P_{\boldsymbol{V}\boldsymbol{Y}}(\boldsymbol{v}, \boldsymbol{y}) P_{W|\boldsymbol{V}\boldsymbol{Y}}(\mathcal{H}_1|\boldsymbol{v}, \boldsymbol{y}). \quad (8)$$

Similarly, the type-II error, i.e. the probability of choosing $\mathcal{H}_0$ when the true hypothesis is $\mathcal{H}_1$, is given by

$$\epsilon_{\text{II}}(Q_{\boldsymbol{V}\boldsymbol{Y}}, P_{W|\boldsymbol{V}\boldsymbol{Y}}) = \sum_{\boldsymbol{v}} \sum_{\boldsymbol{y}} Q_{\boldsymbol{V}\boldsymbol{Y}}(\boldsymbol{v}, \boldsymbol{y}) P_{W|\boldsymbol{V}\boldsymbol{Y}}(\mathcal{H}_0|\boldsymbol{v}, \boldsymbol{y}). \quad (9)$$

We define the smallest type-I error among all tests $P_{W|\boldsymbol{VY}}$ with a type-II error at most β as

$$\alpha_\beta\big(P_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big) \triangleq \min_{\substack{P_{W|\boldsymbol{VY}}:\\ \epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P_{W|\boldsymbol{VY}}) \le \beta}} \Big\{\epsilon_{\mathrm{I}}(P_{\boldsymbol{VY}}, P_{W|\boldsymbol{VY}})\Big\}. \tag{10}$$

The Neyman-Pearson (NP) lemma [5] gives the form of a test $P_{W|\boldsymbol{VY}}$ achieving this optimum performance,

$$P^{\mathrm{NP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) = \begin{cases} 1, & \text{if } \frac{P_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}{Q_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})} > \gamma,\\ p_0, & \text{if } \frac{P_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}{Q_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})} = \gamma,\\ 0, & \text{otherwise,}\end{cases} \tag{11}$$

where the threshold γ and the probability p_0 are chosen such that type-II error equals β.

This lemma is a key result to derive lower bounds on the error probability of a code $\mathcal{C}$ (see e.g. [1].) Consider the hypotheses (6)-(7) for $P_{\boldsymbol{VY}} = P^{(\mathcal{C})}_{\boldsymbol{VY}}$ and arbitrary $Q_{\boldsymbol{VY}}$. We define a (possibly suboptimum) test based on the MAP decoder in (3) as follows. Given an observation $(\boldsymbol{v},\boldsymbol{y})$, we choose $\mathcal{H}_0$ if $\boldsymbol{v}=\boldsymbol{z}$ and $\mathcal{H}_1$ otherwise, i.e. the test is defined as

$$P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) = P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}). \tag{12}$$

Particularizing (8) and (9) for this test, we obtain

$$\epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}) = 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}), \tag{13}$$

$$\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}) = \sum_{\boldsymbol{v},\boldsymbol{y}} Q_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}). \tag{14}$$

As (5) and (13) coincide, we have that $\epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}) = \epsilon(\mathcal{C})$. As a result, lower bounds on the type-I error for this binary hypothesis-testing problem with type-II error given by (14) directly provide lower bounds on the MAP decoding error probability. In particular, for a fixed distribution $Q_{\boldsymbol{VY}}$, the type-I error can be lower-bounded by the type-I error performance of the Neyman-Pearson test, i.e.,

$$\epsilon(\mathcal{C}) \ge \alpha_{\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big). \tag{15}$$

This inequality can be tightened by maximizing (15) over the set of distributions $\{Q_{\boldsymbol{VY}}\}$, yielding

$$\epsilon(\mathcal{C}) \ge \max_{Q_{\boldsymbol{VY}}} \Big\{\alpha_{\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big)\Big\}. \tag{16}$$

This lower bound is a particularization of [3, Eq. (59)] to almost lossless source-channel coding and a given codebook. In principle, the computation of (16) is at least as difficult as the computation of $\epsilon(\mathcal{C})$ since it requires the knowledge of the MAP decoding transformation $P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}$.

III. Main Result

In this section, we show that (16) holds with equality by proving that the test $P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}$ achieves the Neyman-Pearson performance for a specific choice of $Q_{\boldsymbol{VY}}$ and the type-II error given in (14).

Theorem 1: The average error probability of a given codebook $\mathcal{C}$ under MAP decoding satisfies

$$\epsilon(\mathcal{C}) = \max_{Q_{\boldsymbol{VY}}} \Big\{\alpha_{\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big)\Big\}. \tag{17}$$

Moreover, there exists a distribution $Q_{\boldsymbol{VY}}$ optimizing (17) such that $Q_{\boldsymbol{VY}} = Q^\star_{\boldsymbol{V}} \times Q_{\boldsymbol{Y}}$ with $Q^\star_{\boldsymbol{V}}(\boldsymbol{v}) = |\mathcal{V}|^{-k}$ for all $\boldsymbol{v}$. Then

$$\epsilon(\mathcal{C}) = \max_{Q_{\boldsymbol{Y}}} \Big\{\alpha_{|\mathcal{V}|^{-k}}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q^\star_{\boldsymbol{V}} \times Q_{\boldsymbol{Y}}\big)\Big\}. \tag{18}$$

Proof: To prove (17), consider the hypotheses (6)-(7) with $P_{\boldsymbol{VY}} = Q_{\boldsymbol{VY}} = P^{(\mathcal{C})}_{\boldsymbol{VY}}$. From (13) and (14), it follows that $\epsilon_{\mathrm{I}} = 1 - \epsilon_{\mathrm{II}}$. Noting that $\alpha_\beta(P,P) = 1-\beta$ and using $\epsilon(\mathcal{C}) = \epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})$ it follows that

$$\epsilon(\mathcal{C}) = \alpha_{\epsilon_{\mathrm{II}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{(\mathcal{C})}_{\boldsymbol{VY}}\big) \tag{19}$$

$$\le \max_{Q_{\boldsymbol{VY}}} \Big\{\alpha_{\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big)\Big\}. \tag{20}$$

By combining the inequalities (16) and (19)-(20) the first part of the theorem follows.

In order to prove the second part, consider the hypotheses (6)-(7) with $P_{\boldsymbol{VY}} = P^{(\mathcal{C})}_{\boldsymbol{VY}}$ and $Q_{\boldsymbol{VY}} = Q^{(\mathcal{C})}_{\boldsymbol{VY}}$, where

$$Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) \triangleq Q^\star_{\boldsymbol{V}}(\boldsymbol{v}) Q^{(\mathcal{C})}_{\boldsymbol{Y}}(\boldsymbol{y}), \quad \boldsymbol{v} \in \mathcal{V}^k, \boldsymbol{y} \in \mathcal{Y}^n, \tag{21}$$

with

$$Q^\star_{\boldsymbol{V}}(\boldsymbol{v}) \triangleq |\mathcal{V}|^{-k}, \tag{22}$$

$$Q^{(\mathcal{C})}_{\boldsymbol{Y}}(\boldsymbol{y}) \triangleq \frac{1}{\mu} \max_{\boldsymbol{v}'} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v}',\boldsymbol{y}), \tag{23}$$

and where μ is a normalization constant. The test $P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}$, according to (3) and (21)-(23), is given by

$$P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) \triangleq \begin{cases} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|}, & \text{if } \frac{P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}{Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})} = \mu|\mathcal{V}|^k,\\ 0, & \text{otherwise.}\end{cases} \tag{24}$$

Let us choose $\gamma = \mu|\mathcal{V}|^k$ and

$$p_0 = \frac{\sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}{\sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})} \tag{25}$$

$$= \frac{\sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|} Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}{\sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})}, \tag{26}$$

where equality between (25) and (26) holds since $P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) = \mu|\mathcal{V}|^k Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y})$ for all $\boldsymbol{y}$, $\boldsymbol{v} \in \mathcal{S}_{\mathcal{C}}(\boldsymbol{y})$.

We now show that the MAP test (24) achieves the same type-I and type-II error probability as the NP test (11) for this choice of parameters γ and p_0. This shows that both the MAP and the NP test achieve the optimum performance in the Neyman-Pearson sense.

The type-I error probability of the NP test (11) is given by

$$\begin{aligned}
\epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{NP}}_{W|\boldsymbol{VY}}) \\
&= 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P^{\mathrm{NP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) && (27)\\
&= 1 - \sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} p_0 P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) && (28)\\
&= 1 - \sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) && (29)\\
&= 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) && (30)\\
&= \epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}) \;=\; \epsilon(\mathcal{C}), && (31)
\end{aligned}$$

where in (28) we used the definitions of $\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})$ and $P^{\mathrm{NP}}_{W|\boldsymbol{VY}}$; (29) follows from (25), and (30) follows from the definition of $P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}$. Similarly, the type-II error probability of the NP test is

$$\begin{aligned}
\epsilon_{\mathrm{II}}(Q^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{NP}}_{W|\boldsymbol{VY}}) \\
&= \sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} p_0 Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) && (32)\\
&= \sum_{\boldsymbol{y}} \sum_{\boldsymbol{v}\in\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})} \frac{1}{|\mathcal{S}_{\mathcal{C}}(\boldsymbol{y})|} Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) && (33)\\
&= \sum_{\boldsymbol{v},\boldsymbol{y}} Q^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) && (34)\\
&= \epsilon_{\mathrm{II}}(Q^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}), && (35)
\end{aligned}$$

where (33) follows from (26); and (34) follows from the definition of $P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}$.

Then it holds that

$$\begin{aligned}
&\max_{Q_{\boldsymbol{VY}}} \left\{ \alpha_{\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big) \right\} && (36)\\
&\quad\geq \alpha_{\epsilon_{\mathrm{II}}(Q^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}})}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q^{(\mathcal{C})}_{\boldsymbol{VY}}\big) && (37)\\
&\quad= \epsilon_{\mathrm{I}}(P^{(\mathcal{C})}_{\boldsymbol{VY}}, P^{\mathrm{NP}}_{W|\boldsymbol{VY}}) \;=\; \epsilon(\mathcal{C}) && (38)
\end{aligned}$$

where in (38) we used (27)-(31) and (32)-(35).

From the inequalities (16) and (36)-(38) it follows that (37) holds with equality. Hence, $Q^{(\mathcal{C})}_{\boldsymbol{VY}}$ is a maximizer of (36). Then, in order to perform the optimization in (17) we may restrict ourselves to distributions of the form $Q_{\boldsymbol{VY}} = Q^{\star}_{\boldsymbol{V}} \times Q_{\boldsymbol{Y}}$. For this choice of $Q_{\boldsymbol{VY}}$ we have that

$$\begin{aligned}
\epsilon_{\mathrm{II}}(Q_{\boldsymbol{VY}}, P^{\mathrm{MAP}}_{W|\boldsymbol{VY}}) &= \sum_{\boldsymbol{v},\boldsymbol{y}} Q^{\star}_{\boldsymbol{V}}(\boldsymbol{v}) Q_{\boldsymbol{Y}}(\boldsymbol{y}) P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}) && (39)\\
&= |\mathcal{V}|^{-k} \sum_{\boldsymbol{v},\boldsymbol{y}} Q_{\boldsymbol{Y}}(\boldsymbol{y}) P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}) && (40)\\
&= |\mathcal{V}|^{-k}. && (41)
\end{aligned}$$

As a result, (18) follows. ■

IV. Connection with previous results

In [4], the present authors derived a lower bound on the minimum error probability of almost lossless source-channel coding. Following a variation of the method described in Section II, they considered an independent binary hypothesis test for every source message and individually applied the NP lemma to each test. In an analogous way to Section II, this setup provides lower bounds on the error probability conditioned on each transmitted message. The next result shows that this method also gives the error probability:

Corollary 1: The average error probability of a given codebook $\mathcal{C}$ under MAP decoding satisfies

$$\epsilon(\mathcal{C}) = \max_{Q_{\boldsymbol{Y}|\boldsymbol{V}}} \left\{ \sum_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v}) \alpha_{Q_{\boldsymbol{Z}}(\boldsymbol{v})}\big(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}(\boldsymbol{v})}, Q_{\boldsymbol{Y}|\boldsymbol{V}=\boldsymbol{v}}\big) \right\}, \quad (42)$$

where

$$Q_{\boldsymbol{Z}}(\boldsymbol{v}) \triangleq \sum_{\boldsymbol{y}} Q_{\boldsymbol{Y}|\boldsymbol{V}}(\boldsymbol{y}|\boldsymbol{v}) P^{(\mathcal{C})}_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}|\boldsymbol{y}). \quad (43)$$

Proof: The proof follows from the definition of $\alpha_{(\cdot)}(\cdot,\cdot)$ using convex-optimization techniques. Let us define

$$\begin{aligned}
f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \\
&\triangleq \sum_{\boldsymbol{y}} Q_{\boldsymbol{Y}|\boldsymbol{V}=\boldsymbol{v}}(\boldsymbol{y}) P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) - Q_{\boldsymbol{Z}}(\boldsymbol{v}). && (44)
\end{aligned}$$

When optimized over the auxiliary distribution $Q_{\boldsymbol{V}}$, the bracketed term in (17) becomes

$$\begin{aligned}
&\max_{Q_{\boldsymbol{V}}} \left\{ \alpha_{\left(\sum_{\boldsymbol{v}} Q_{\boldsymbol{V}}(\boldsymbol{v}) Q_{\boldsymbol{Z}}(\boldsymbol{v})\right)}\big(P^{(\mathcal{C})}_{\boldsymbol{VY}}, Q_{\boldsymbol{VY}}\big) \right\}\\
&= \max_{Q_{\boldsymbol{V}}} \min_{\substack{P_{W|\boldsymbol{VY}}:\\ \sum_{\boldsymbol{v}} Q_{\boldsymbol{V}}(\boldsymbol{v}) f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \leq 0}} \Bigg\{ 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) \Bigg\} && (45)\\
&= \max_{\lambda\geq 0, Q_{\boldsymbol{V}}} \min_{P_{W|\boldsymbol{VY}}} \Bigg\{ 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) - \lambda \sum_{\boldsymbol{v}} Q_{\boldsymbol{V}}(\boldsymbol{v}) f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \Bigg\} && (46)\\
&= \max_{\{\lambda_{\boldsymbol{v}}\geq 0\}} \min_{P_{W|\boldsymbol{VY}}} \Bigg\{ 1 - \sum_{\boldsymbol{v},\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) - \sum_{\boldsymbol{v}} \lambda_{\boldsymbol{v}} f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \Bigg\} && (47)\\
&= \sum_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v}) \min_{\substack{P_{W|\boldsymbol{VY}}:\\ f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \leq 0}} \Bigg\{ 1 - \sum_{\boldsymbol{y}} P^{(\mathcal{C})}_{\boldsymbol{Y}|\boldsymbol{V}}(\boldsymbol{y}|\boldsymbol{v}) P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) \Bigg\} && (48)\\
&= \sum_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v}) \alpha_{Q_{\boldsymbol{Z}}(\boldsymbol{v})}\big(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}(\boldsymbol{v})}, Q_{\boldsymbol{Y}|\boldsymbol{V}=\boldsymbol{v}}\big), && (49)
\end{aligned}$$

where in (46) we introduced the constraint into the objective by means of the Lagrange multiplier λ; (47) follows from the fact that λ and $Q_{\boldsymbol{V}}$ only appear in the objective as $\lambda_{\boldsymbol{v}} \triangleq \lambda Q_{\boldsymbol{V}}(\boldsymbol{v})$; and finally (48) follows from considering

$\{\lambda_{\boldsymbol{v}}\}$ as Lagrange multipliers associated with the individual constraints $f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}}) \leq 0$, $\forall \boldsymbol{v}$. Eqs. (46) and (48) hold with equality since the constrained optimization problem is convex, $f_{\boldsymbol{v}}(P_{W|\boldsymbol{VY}})$ is affine in $P_{W|\boldsymbol{VY}}$ and there exists a feasible point (e.g. by choosing $P_{W|\boldsymbol{VY}}(\mathcal{H}_0|\boldsymbol{v},\boldsymbol{y}) = 0$ for all $\boldsymbol{v}$,$\boldsymbol{y}$.) The result follows by substituting (45)-(49) into (17). ■

The performance of the MAP decoder can thus be equivalently characterized by either a bank of independent binary hypothesis tests defined over the channel outputs for each source message or by a single binary hypothesis test defined over the 2-fold space of messages and channel outputs. This equivalence holds as long as the bound (16) is maximized over $Q_{\boldsymbol{V}}$. For fixed $Q_{\boldsymbol{V}}$ the bank of independent binary hypothesis tests gives tighter bounds in general.

We now assess the weaknesses of previous converse bounds [3], [4] with respect to the minimum achievable error probability. One can obtain the converse bound [4, Lem. 2] from Corollary 1 by minimizing (42) over all codebooks and distributions Q_Z. This converse is not tight in general as the minimizing distribution $Q_Z^\star$ does not need to coincide with the distribution induced by the MAP decoder.

Also, from (18) it follows that

$$\min_{\mathcal{C}} \epsilon(\mathcal{C}) = \min_{\mathcal{C}} \max_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{|\mathcal{V}|^{-k}} \left(P_{\boldsymbol{VY}}^{(\mathcal{C})}, Q_{\boldsymbol{V}}^\star \times Q_{\boldsymbol{Y}} \right) \right\} \tag{50}$$

$$\geq \inf_{P_{\boldsymbol{X}|\boldsymbol{V}}} \sup_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{|\mathcal{V}|^{-k}} \left(P_{\boldsymbol{VY}}, Q_{\boldsymbol{V}}^\star \times Q_{\boldsymbol{Y}} \right) \right\}, \tag{51}$$

where the last step follows from relaxing the minimization to account for every $P_{\boldsymbol{X}|\boldsymbol{V}}$ such that $P_{\boldsymbol{VY}}(\boldsymbol{v},\boldsymbol{y}) = \sum_{\boldsymbol{x}} P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{X}|\boldsymbol{V}}(\boldsymbol{x}|\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. The optimum value of the minimization in (51) does not coincide in general with the distribution induced by the codebook and hence the inequality (51) is not tight in general. The bound (51) is equivalent to [3, Th. 4] when particularized to the almost lossless setting.

V. Discussion

Theorem 1 shows that the extension of the meta-converse to source-channel coding coincides with the MAP decoding error probability. This is no longer true if we weaken (50) to ignore the structure imposed by the codebooks, which yields the possibly strict inequality (51).

As an example, consider the channel-coding problem of transmitting M equiprobable messages over a DMC. Particularizing (50)-(51) with $k = 1$, $|\mathcal{V}| = M$, we obtain

$$\min_{\mathcal{C}} \epsilon(\mathcal{C}) = \min_{\mathcal{C}} \max_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{\frac{1}{M}} \left(P_{\boldsymbol{X}}^{(\mathcal{C})} \times P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}^{(\mathcal{C})} \times Q_{\boldsymbol{Y}} \right) \right\} \tag{52}$$

$$\geq \inf_{P_{\boldsymbol{X}}} \sup_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{\frac{1}{M}} \left(P_{\boldsymbol{X}} \times P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}} \right) \right\}, \tag{53}$$

where (53) coincides with the hypothesis-testing bound from [2, Thm. 27]. The exponential decay of (53) has recently been shown [6, Sec. VI.E] to be equal to the sphere-packing exponent. However, below the critical rate of the channel, the reliability function is in general bounded away [7] from the sphere-packing exponent and thus, the gap between (52) and (53) may grow exponentially with n.

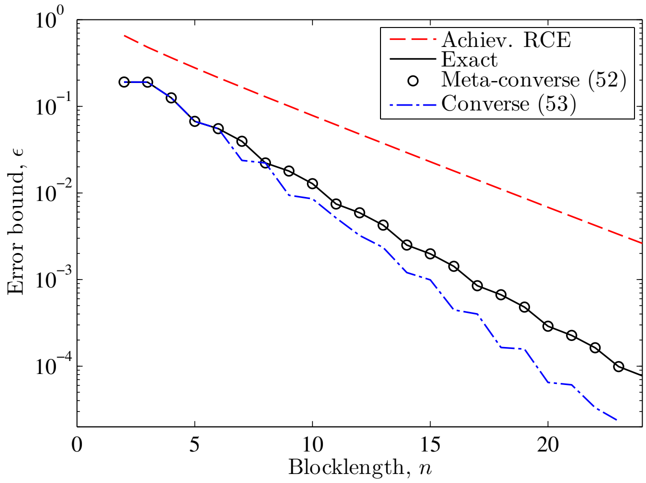

Fig. 1. Channel coding error probability bounds for the BSC with parameters $P_{Y|X}(1|0) = 0.1$, $M = 4$.

Fig. 1 shows different bounds on the error probability for $M = 4$ messages and a binary symmetric channel (BSC). In this setup, the best code can be obtained explicitly [8] and hence the exact ML decoding error probability can be computed. As upper bound, we show the exact random coding error probability when the ties are decoded in error (RCE) for a random-coding ensemble generated from the uniform distribution. As lower bounds, we show (52) (computed for the best code) and (53). The figure confirms our main result in Theorem 1, since (52) is equal to the exact error probability of the best code given in [8]. In this scenario, the reliability function of the error probability does not coincide neither with the random-coding nor with the sphere-packing error exponents, given respectively by the exponential decay of the bounds RCE and (53) in the figure. As a result, the weakening (53) yields a looser bound that incurs in a loss in exponent.

References

[1] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.
[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
[3] V. Kostina and S. Verdu, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, 2013.
[4] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Martinez, "Converse bounds for finite-length joint source-channel coding," in *Proc. 50th Allerton Conf. on Comm., Cont. and Comp.*, Monticello, IL, USA, Oct. 2012.
[5] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, vol. 231, pp. 289–337, 1933.
[6] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, 2013.
[7] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Inf. Contr.*, vol. 10, no. 5, pp. 522–552, 1967.
[8] P.-N. Chen, H.-Y. Lin, and S. Moser, "Ultra-small block-codes for binary discrete memoryless channels," in *Proc. IEEE Inf. Theory Works.*, Paraty, Brazil, Oct. 2011.