Converse Bounds for Finite-Length Joint Source-Channel Coding

Adrià Tauste Campo[1], Gonzalo Vazquez-Vilar[1], Albert Guillén i Fàbregas[123], Alfonso Martinez[1]

[1]Universitat Pompeu Fabra, [2]ICREA, [3]University of Cambridge

Email: {atauste,gvazquez,guillen,alfonso.martinez}@ieee.org

Abstract—Based on the hypothesis-testing method, we derive lower bounds on the average error probability of finite-length joint source-channel coding. The extension of the meta-converse bound of channel coding to joint source-channel coding depends on the codebook and the decoding rule and thus, it is a priori computationally challenging. Weaker versions of this general bound recover known converses in the literature and provide computationally feasible expressions.

I. INTRODUCTION

Reliable communication of messages in the finite block length regime can be characterized by upper and lower bounds on the average error probability of the best possible code. In order to prove the existence of a good code, random-coding techniques are often employed to derive upper bounds on the average error probability. In contrast, the computation of lower bounds satisfied by every code is in general challenging since one must optimize the bound over each possible codebook and decoding rule.

For equiprobable messages, a number of lower bounds on the average error probability [1]–[5] lead to a proof of the converse part of Shannon's theorem [6] when the block length grows to infinity. More recently, some of these bounds have been generalized to non-equiprobable messages using information-spectrum measures [7], [8] or the hypothesis-testing method [9].

In this paper, we elaborate the hypothesis-testing method in the context of joint source-channel coding to provide lower bounds on the average error probability. Following the footsteps of [4], [5], we propose an extension of the *meta-converse* by Polyanskyi *et al.* [5, Th. 26], which states that every channel code with M codewords, block length n and average error probability ϵ satisfies

$$\inf_{P_{\boldsymbol{X}}} \sup_{Q_{\boldsymbol{Y}}} \beta_{1-\epsilon} \left(P_{\boldsymbol{X}} \times P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}} \right) \leq \frac{1}{M}, \quad (1)$$

where $\beta_\alpha \left(P_{\boldsymbol{X}} \times P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}} \right)$ is the minimum type-II error given by the Neyman-Pearson lemma [10] for a maximum type-I error of $1 - \alpha$ when testing between $P_{\boldsymbol{X}} \times P_{\boldsymbol{Y}|\boldsymbol{X}}$ and $P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}}$, where $P_{\boldsymbol{X}}$ is the input distribution

induced by the codebook, $P_{\boldsymbol{Y}|\boldsymbol{X}}$ is the channel law and $Q_{\boldsymbol{Y}}$ is an arbitrary output distribution.

The central idea of our method is to consider an independent binary hypothesis test for every source message and obtain a lower bound on the average error probability by applying the Neyman-Pearson lemma [10] to each test. This approach initially provides a converse bound involving a costly optimization over all possible codebooks and decoding rules. Moreover, we show that this bound recovers several known results, including the information-spectrum bounds [7], [8] and more importantly, it is proven to attain Csiszár's sphere-packing exponent for joint source-channel coding [11, Th. 3]. Finally, we weaken the converse result to obtain lower bounds on the average error probability that can be numerically computed for some source-channel pairs of interest.

A. Notation and System Model

We consider the transmission of a length-k discrete memoryless source over a discrete memoryless channel using length-n block codes that are known both at the transmitter and receiver. 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 a discrete 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.

Without loss of generality we assume that the source messages are indexed as $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{|\mathcal{V}|^k}$. An encoder maps the length-k source message $\boldsymbol{v}_l$ to a length-n codeword $\boldsymbol{x}_l$, which is then transmitted over the channel. We refer to the ratio $t \triangleq k/n$ as the transmission rate. Based on the length-n channel output $\boldsymbol{y}$ the decoder guesses which source message was transmitted. The decoding rule is specified by the (possibly random) transformation $P_{\boldsymbol{Z}|\boldsymbol{Y}} : \mathcal{Y}^n \to \mathcal{V}^k$. The decoded message will be denoted in the following by $\boldsymbol{z}$. The average error probability ϵ is given by

$$\epsilon = \sum_{l=1}^{|\mathcal{V}|^k} P_{\boldsymbol{V}}(\boldsymbol{v}_l) \epsilon(\boldsymbol{v}_l), \quad (2)$$

where

$$\epsilon(\boldsymbol{v}_l) \triangleq \Pr\{\boldsymbol{Z} \neq \boldsymbol{v}_l\} \tag{3}$$

$$= 1 - \sum_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_l) P_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}_l|\boldsymbol{y}), \tag{4}$$

is the error probability when message $\boldsymbol{v}_l$ is transmitted.

In this paper, we obtain tight lower bounds on the average error probability of the best code following a hypothesis-testing approach [4], [5].

II. Hypothesis-Testing Approach

For every pair $(\boldsymbol{v}_l, \boldsymbol{x}_l)$ we define a binary hypothesis-testing problem between the channel conditional distribution $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_l}$ and an arbitrary output distribution $Q_{\boldsymbol{Y}}^{(l)}$ as

$$\mathcal{H}_0: \quad \boldsymbol{Y} \sim P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_l}, \tag{5}$$

$$\mathcal{H}_1: \quad \boldsymbol{Y} \sim Q_{\boldsymbol{Y}}^{(l)}. \tag{6}$$

We can construct a sub-optimal test for the above problem from the system described in Section I-A: for a given source message $\boldsymbol{v}_l$, upon observation of the channel output $\boldsymbol{y}$, we choose $\mathcal{H}_0$ if $\boldsymbol{z} = \boldsymbol{v}_l$, and $\mathcal{H}_1$, otherwise. The performance of this test can be evaluated according to its type-I and type-II errors. Specifically, the probability of choosing $Q_{\boldsymbol{Y}}^{(l)}$ when the true distribution is $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_l}$ (type-I error) is equal to

$$\epsilon(\boldsymbol{v}_l) = 1 - \sum_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_l) P_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}_l|\boldsymbol{y}). \tag{7}$$

Similarly, the probability of choosing $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_l}$ when the true distribution is given by $Q_{\boldsymbol{Y}}^{(l)}$ (type-II error) is given by

$$Q_{\boldsymbol{Z}}^{(l)}(\boldsymbol{v}_l) \triangleq \sum_{\boldsymbol{y}} Q_{\boldsymbol{Y}}^{(l)}(\boldsymbol{y}) P_{\boldsymbol{Z}|\boldsymbol{Y}}(\boldsymbol{v}_l|\boldsymbol{y}). \tag{8}$$

The two types of error can be related via the Neyman-Pearson lemma [10]. This result states that the optimal type-II error among all possibly randomized tests $P_{W|\boldsymbol{Y}} : \mathcal{Y}^n \to \{\mathcal{H}_0, \mathcal{H}_1\}$ with a type-I error of at most $1 - \alpha$ is given by

$$\begin{aligned}
&\beta_\alpha\big(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}_l}, Q_{\boldsymbol{Y}}^{(l)}\big) \\
&\triangleq \min_{\substack{P_{W|\boldsymbol{Y}}: \\ \sum_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_l) P_{W|\boldsymbol{Y}}(\mathcal{H}_0|\boldsymbol{y}) \geq \alpha}} \sum_{\boldsymbol{y}} Q_{\boldsymbol{Y}}^{(l)}(\boldsymbol{y}) P_{W|\boldsymbol{Y}}(\mathcal{H}_0|\boldsymbol{y}).
\end{aligned} \tag{9}$$

In the rest of the paper, for ease of notation, we shall use $\beta_\alpha(\boldsymbol{x}, Q_{\boldsymbol{Y}}) \triangleq \beta_\alpha(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}, Q_{\boldsymbol{Y}})$. Consequently, the type-II error of any test for (5)–(6) is lower-bounded by $\beta_\alpha\big(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}^{(l)}\big)$ as long as the type-I error is no greater than $1 - \alpha$. In particular, by setting $1 - \alpha = \epsilon(\boldsymbol{v}_l)$ in (9) and combining it with (8) we obtain

$$\beta_{1-\epsilon(\boldsymbol{v}_l)}\big(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}^{(l)}\big) \leq Q_{\boldsymbol{Z}}^{(l)}(\boldsymbol{v}_l), \quad l = 1, \dots, |\mathcal{V}|^k, \tag{10}$$

which upon recalling (2) gives an implicit lower bound on the average error probability of our proposed coding scheme.

In order to obtain a valid converse bound from (10) one needs to perform a challenging optimization over all possible codebooks $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{|\mathcal{V}|^k}\}$ and decoding transformations $P_{\boldsymbol{Z}|\boldsymbol{Y}}$. In contrast, any choice of $Q_{\boldsymbol{Y}}^{(l)}, l = 1, \dots, |\mathcal{V}|^k$, gives

a converse bound as it is independent of the codebook and the decoder. Alternatively, a converse bound can be derived by defining $P_{\boldsymbol{Z}|\boldsymbol{Y}}$ according to the MAP decoding rule and optimizing (10) over all possible codebooks.

In both cases, the derivation of a converse bound becomes computationally unfeasible as the block length increases. Hence, in the rest of the paper we analyze the performance of lower bounds derived from (10) which are computable in several cases of interest. In particular, in the next section we weaken (10) to re-derive a generalized version of the Verdú-Han lemma for source-channel coding and we show that the lower bound induced by (2)–(10) attains Csiszár's sphere-packing exponent [11, Th. 3]. Then, in Section IV we further weaken (10) to obtain computable finite-length lower bounds on the average error probability of source-channel coding.

III. Connection with Previous Work

A. Information-Spectrum Bounds

In [8, Th. 6], we derived a lower bound on the average error probability as a generalization of the Verdú-Han lemma for channel coding [3],

$$\epsilon \geq \Pr\left\{P_{\boldsymbol{V}}(\boldsymbol{V}) P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\boldsymbol{Y}|\boldsymbol{X}(\boldsymbol{V})\big) \leq \gamma(\boldsymbol{Y})\right\} - \sum_{\boldsymbol{y}} \gamma(\boldsymbol{y}), \tag{11}$$

where $\boldsymbol{X}(\boldsymbol{V})$ denotes the mapping induced by a specific codebook and $\gamma : \mathcal{Y}^n \to \mathbb{R}^+$ is an arbitrary non-negative function. By choosing $\gamma(\boldsymbol{Y}) = \gamma Q_{\boldsymbol{Y}}(\boldsymbol{Y})$, with $Q_{\boldsymbol{Y}}(\boldsymbol{Y})$ being an arbitrary output distribution and $\gamma > 0$, optimizing the bound over γ, $Q_{\boldsymbol{Y}}$ and the $P_{\boldsymbol{X}|\boldsymbol{V}}$ induced by each codebook, one obtains the converse bound

$$\epsilon \geq \inf_{P_{\boldsymbol{X}|\boldsymbol{V}}} \sup_{\gamma>0} \left\{\sup_{Q_{\boldsymbol{Y}}} \Pr\left\{\frac{P_{\boldsymbol{V}}(\boldsymbol{V}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{Q_{\boldsymbol{Y}}(\boldsymbol{Y})} < \gamma\right\} - \gamma\right\}, \tag{12}$$

which has been independently given in [9, Eq. 34].

We next show that (12) can be seen as a consequence of (10). First, fix a given codebook $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{|\mathcal{V}|^k}$. Then, by combining the inequality [2]

$$\beta_{1-\epsilon}(\boldsymbol{x}, Q_{\boldsymbol{Y}}) \geq \sup_{\gamma>0} \frac{1}{\gamma} \left(\Pr\left\{\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{x})}{Q_{\boldsymbol{Y}}(\boldsymbol{Y})} < \gamma\right\} - \epsilon\right) \tag{13}$$

where $\Pr\{\cdot\}$ is computed according to $\boldsymbol{Y} \sim P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$, with (10) for a (message-independent) distribution $Q_{\boldsymbol{Y}}$, one obtains the set of inequalities

$$Q_{\boldsymbol{Z}}(\boldsymbol{v}_l) \geq \frac{1}{\gamma_l} \left(\Pr\left\{\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{x}_l)}{Q_{\boldsymbol{Y}}(\boldsymbol{Y})} < \gamma_l\right\} - \epsilon(\boldsymbol{v}_l)\right), \tag{14}$$

for $\gamma_l > 0$, $l = 1, \dots, |\mathcal{V}|^k$. By choosing $\gamma_l = \frac{\gamma}{P_{\boldsymbol{V}}(\boldsymbol{v}_l)}$ with $\gamma > 0$ for every $l = 1, \dots, |\mathcal{V}|^k$ such that $P_{\boldsymbol{V}}(\boldsymbol{v}_l) \neq 0$, (14) is equivalent to

$$\begin{aligned}
&Q_{\boldsymbol{Z}}(\boldsymbol{v}_l) \\
&\geq \frac{P_{\boldsymbol{V}}(\boldsymbol{v}_l)}{\gamma} \left(\Pr\left\{\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{x}_l)}{Q_{\boldsymbol{Y}}(\boldsymbol{Y})} < \frac{\gamma}{P_{\boldsymbol{V}}(\boldsymbol{v}_l)}\right\} - \epsilon(\boldsymbol{v}_l)\right), \tag{15}
\end{aligned}$$

for $l = 1, \ldots, |\mathcal{V}|^k$. Consider now the set of conditional distributions $P_{X|V}$ induced by the codebook, i.e, $P_{X|V}(\boldsymbol{x}_i|\boldsymbol{v}_j) = 1$ if $i = j$ and $P_{X|V}(\boldsymbol{x}_i|\boldsymbol{v}_j) = 0$, otherwise. By summing both sides of (15) over $l = 1, \ldots |\mathcal{V}|^k$ we finally have

$$\epsilon \geq \Pr\left\{\frac{P_V(V)P_{Y|X}(Y|X)}{Q_Y(Y)} < \gamma\right\} - \gamma, \quad (16)$$

which upon optimization over γ, Q_Y and $P_{X|V}$ yields (12).

In particular, Han's generalization of the Verdú-Han lemma derived in [7, Lemma 3.2] can be recovered from (12) by setting $Q_Y = P_Y$ to be the output distribution induced by a particular codebook and by rewriting (12) with the definitions of entropy density $h(V) \triangleq -\log P_V(V)$, and information density $i(X, Y) \triangleq \log \frac{P_{Y|X}(Y|X)}{P_Y(Y)}$ [12], as

$$\epsilon \geq \inf_{P_{X|V}} \sup_{\gamma > 0} \Pr\{i(X; Y) - h(V) < -\log \gamma\} - \frac{1}{\gamma}. \quad (17)$$

B. Csiszár's Sphere-Packing Exponent

Csiszár showed in [11, Th. 3] that the error exponent of every source-channel code is upper-bounded by

$$E_J^{\text{sp}} \triangleq \min_{R \in [tH(V), t\log|\mathcal{V}|]} te\left(\frac{R}{t}, P_V\right) + E_{\text{sp}}(R, P_{Y|X}), \quad (18)$$

where

$$e(R, P_V) \triangleq \min_{Q:H(Q)\geq R} D(P_Q\|P_V) \quad (19)$$

is the source reliability function [13] and

$$E_{\text{sp}}(R, P_{Y|X}) \triangleq \max_{P_X} \min_{\substack{P'_{Y|X}: \\ I(P_X, P'_{Y|X}) \leq R}} D(P'_{Y|X}\|P_{Y|X}|P_X) \quad (20)$$

is the channel-coding sphere-packing exponent [14].

When the minimizing R in (18) lies above the critical rate of the channel [11], [15], the bound (18) is tight and gives the actual error exponent.

We next show that using (10) with an appropriate choice of $Q_Y^{(l)}$ recovers Csiszár's result. We first decompose the average error probability using the set of source-type classes $\mathcal{T}_i^k$, $i = 1, \ldots, N_k$. Rewriting (2) we have that

$$\epsilon = \sum_{i=1}^{N_k} \Pr\{\mathcal{T}_i^k\} \epsilon_i, \quad (21)$$

where

$$\epsilon_i \triangleq \frac{1}{|\mathcal{T}_i^k|} \sum_{\boldsymbol{v} \in \mathcal{T}_i^k} \epsilon(\boldsymbol{v}). \quad (22)$$

Our re-derivation relies on the next result.

Lemma 1 ([4, Thm. 20]): For every $\boldsymbol{v}_l \in \mathcal{T}_i^k$ consider the binary hypothesis test in (5) between $P_{Y|X=\boldsymbol{x}_l}$ and the distribution $Q_Y^{(l)} = Q_Y^{(\mathcal{T}_i^k)}$. Let a decision rule have type-I error equal to $\epsilon(\boldsymbol{v}_l)$ and type-II error equal to b. Then, there exists a distribution $Q_Y^{(\mathcal{T}_i^k)}$ such that, if $\bar{R} > 0$ satisfies

$$b \leq \gamma e^{-n(\bar{R}+\eta)}, \quad \eta > 0, \quad \gamma \in (0, 1), \quad (23)$$

then

$$\epsilon(\boldsymbol{v}_l) \geq \frac{1}{2}\left(1 - \frac{A(\bar{R})}{n\eta^2} - \gamma\right) e^{-n\left(E_{\text{sp}}\left(\bar{R} - \frac{\log 2}{n}, P_{Y|X}\right) + \eta\right)} \quad (24)$$

for all $\boldsymbol{v}_l \in \mathcal{T}_i^k$, where $A(\bar{R}) > 0$ is a function of $\bar{R}$ independent of n.

For every source type-class $\mathcal{T}_i^k$, $i = 1, \ldots, N_k$, we define the probability distribution

$$\bar{Q}_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}) \triangleq \begin{cases} \frac{Q_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v})}{\sum_{\boldsymbol{v}' \in \mathcal{T}_i^k} Q_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}')} &, \quad \boldsymbol{v} \in \mathcal{T}_i^k, \\ 0 &, \quad \text{otherwise}, \end{cases} \quad (25)$$

where $Q_Z^{(l)} = Q_Z^{(\mathcal{T}_i^k)}$ for all $\boldsymbol{v}_l \in \mathcal{T}_i^k$. In view of (25) there must exist $\boldsymbol{v} \in \mathcal{T}_i^k$, such that

$$Q_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}) \leq \bar{Q}_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}) \leq \frac{1}{|\mathcal{T}_i^k|}. \quad (26)$$

Otherwise, $\sum_{\boldsymbol{v} \in \mathcal{T}_i^k} \bar{Q}_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}) > 1$ and $\bar{Q}_Z^{(\mathcal{T}_i^k)}$ would not be a probability distribution. Without loss of generality and for ease of exposition we next assume that the indexing of the message set is such that $\boldsymbol{v}_i$ is a source message fulfilling (26) for $\mathcal{T}_i^k$, $i = 1, \ldots, N_k$. Then, we rewrite (26) as

$$Q_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}_i) \leq \gamma e^{-n(\bar{R}_i(k,n)+\eta(n))}, \quad (27)$$

for $\gamma \in (0, 1)$, $\eta(n) = \frac{K}{\sqrt{n}}$, $K > 0$, and where we defined

$$\bar{R}_i(k, n) \triangleq \frac{1}{n}\log|\mathcal{T}_i^k| + \frac{1}{n}\log\gamma - \frac{K}{\sqrt{n}} \quad (28)$$

such that $\gamma e^{-n\left(\bar{R}_i(k,n)+\frac{K}{\sqrt{n}}\right)} = |\mathcal{T}_i^k|^{-1}$. We now apply Lemma 1 with $P_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = P_{Y|X}(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v}_i))$, $b = Q_Z^{(\mathcal{T}_i^k)}(\boldsymbol{v}_i)$, and $\bar{R}_i(k, n)$, which satisfies (27) for $\gamma \in (0, 1)$. Then, it follows from (24) that

$$\epsilon(\boldsymbol{v}) \geq \frac{1}{2}\left(1 - \frac{A(\bar{R}_i(k,n))}{K^2} - \gamma\right)$$
$$\times e^{-n\left(E_{\text{sp}}\left(\bar{R}_i(k,n) - \frac{\log 2}{n}, P_{Y|X}\right) + \frac{K}{\sqrt{n}}\right)} \quad (29)$$

for all $\boldsymbol{v} \in \mathcal{T}_i^k$. By plugging (29) into (22) we have that

$$\epsilon_i = \frac{1}{|\mathcal{T}_i^k|} \sum_{\boldsymbol{v} \in \mathcal{T}_i^k} \epsilon(\boldsymbol{v}) \quad (30)$$

$$\geq \frac{1}{2}\left(1 - \frac{A(\bar{R}_i(k,n))}{K^2} - \gamma\right) e^{-n\left(E_{\text{sp}}\left(R_i(k,n), P_{Y|X}\right) + \frac{K}{\sqrt{n}}\right)} \quad (31)$$

where $R_i(k, n) \triangleq \bar{R}_i(k, n) - \frac{\log 2}{n}$. We now focus on the terms $\Pr\{\mathcal{T}_i^k\}$, $i = 1, \ldots, N_k$ in (21). Using [16, Lemma 2.6] we have that $\Pr\{\mathcal{T}_i^k\}$ can be lower-bounded as

$$\Pr\{\mathcal{T}_i^k\} \geq (k+1)^{-|\mathcal{V}|} e^{-kD(\mathsf{P}_i^k\|P_V)}, \quad (32)$$

for every $i = 1, \ldots, N_k$, where P_i^k is the type associated to the class $\mathcal{T}_i^k$. Hence, combining (21), (31) and (32), we obtain

$$\epsilon \geq \sum_{i=1}^{N_k} e^{-n\left(tD(\mathsf{P}_i^{tn}\|P_V) + E_{\mathrm{sp}}\left(R_i(k,n), P_{Y|X}\right)\right) + o_1(k,n)} \quad (33)$$

where

$$o_1(k,n) \triangleq K\sqrt{n} - |\mathcal{V}|\log(k+1)$$
$$+ \log\left(1 - \frac{A\left(\bar{R}_i(k,n)\right)}{K^2} - \gamma\right) - \log 2 \quad (34)$$

on account of (28). Finally, by choosing $K > 0$ appropriately, using that $A(\cdot)$ is a continuous function and $E_{\mathrm{sp}}(R, P_{Y|X})$ is a non-increasing continuous function with respect to R, the proof follows along the same lines as in [11] to conclude that

$$\lim_{n \to \infty} -\frac{1}{n}\log(\epsilon) \leq te\left(\frac{R}{t}, P_V\right) + E_{\mathrm{sp}}(R, P_{Y|X}), \quad (35)$$

for some $R \in \left[tH(V), t\log|\mathcal{V}|\right]$, such that $R \triangleq \lim_{n \to \infty} \bar{R}_i(k,n)$, which after minimization over all $R \in [tH(V), t\log\mathcal{V}]$ yields (18).

IV. COMPUTABLE BOUNDS

The aim of this section is to show that (10) can be conveniently weakened to obtain practical converse results in several cases of interest. First, observe from (10) that if $\beta_\alpha\left(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}^{(l)}\right)$ is invertible with respect to α in an appropriate range, one may formulate (10) as an explicit lower bound on $\epsilon(\boldsymbol{v}_l)$ for every $l = 1, \ldots, |\mathcal{V}|^k$, which in turn gives a lower bound on ϵ after averaging over all source messages.

To this end, we make use of the analytical properties of $\beta_\alpha(\boldsymbol{x}, Q_{\boldsymbol{Y}})$ as a function of α. It is known that $\beta_\alpha(\boldsymbol{x}_l, Q_{\boldsymbol{Y}})$ is a piecewise-linear, convex, and non-decreasing function in $\alpha \in [0, 1]$ [17] that takes values in $[0, \beta_{\max}]$, where $\beta_{\max} \leq 1$. Then, the fact that $\beta_0(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}) = 0$ and the convexity in $\alpha \in [0, 1]$ implies there must exist $\alpha_{\min} \in [0, 1]$ such that the function takes the value 0 in $[0, \alpha_{\min})$ and it is strictly increasing in $[\alpha_{\min}, 1]$. As a consequence, the function $\beta_\alpha(\cdot)$ is invertible with respect to α in the range $(0, \beta_{\max}]$. The aforementioned arguments can be used to define the function

$$\alpha_b(\boldsymbol{x}, Q_{\boldsymbol{Y}}) \triangleq \begin{cases} \alpha_{\min}, & b = 0, \\ a \text{ such that } \beta_a = b, & b \in (0, \beta_{\max}], \\ 1, & b \in (\beta_{\max}, 1], \end{cases} \quad (36)$$

in the domain $[0, 1]$. From the above definition one can check that for given $a, b \in [0, 1]$,

$$\beta_a(\boldsymbol{x}, Q_{\boldsymbol{Y}}) \leq b \quad \Rightarrow \quad \alpha_b(\boldsymbol{x}, Q_{\boldsymbol{Y}}) \geq a. \quad (37)$$

Consequently, by applying (37) to (10) it follows that

$$\epsilon(\boldsymbol{v}_l) \geq 1 - \alpha_{Q_{\boldsymbol{Z}}^{(l)}(\boldsymbol{v}_l)}\left(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}^{(l)}\right) \quad (38)$$

for $l = 1, \ldots, |\mathcal{V}|^k$. Averaging (38) over the source messages and upon appropriate optimization we obtain the next result.

Lemma 2: The average error probability ϵ incurred by any codebook is lower-bounded by

$$\epsilon \geq$$
$$1 - \sup_{\substack{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{V}|^k} \\ P_{\boldsymbol{Z}|\boldsymbol{Y}}}} \left\{ \sum_{l=1}^{|\mathcal{V}|^k} P_{\boldsymbol{V}}(\boldsymbol{v}_l) \inf_{Q_{\boldsymbol{Y}}^{(l)}} \left\{ \alpha_{Q_{\boldsymbol{Z}}^{(l)}(\boldsymbol{v}_l)}(\boldsymbol{x}_l, Q_{\boldsymbol{Y}}^{(l)}) \right\} \right\}. \quad (39)$$

In order to provide computationally feasible bounds, we restrict our attention to channels for which, when $Q_{\boldsymbol{Y}}$ is appropriately chosen, the function $\beta_\alpha(Q_{\boldsymbol{Y}}) \triangleq \beta_\alpha(\boldsymbol{x}, Q_{\boldsymbol{Y}})$ and thus, $\alpha_b(Q_{\boldsymbol{Y}}) \triangleq \alpha_b(\boldsymbol{x}, Q_{\boldsymbol{Y}})$, is independent of $\boldsymbol{x}$. Channels of interest fulfilling this property are symmetric channels according to [18, p. 94] and $Q_{\boldsymbol{Y}}^{(l)}(\boldsymbol{y}) = Q_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_{j=1}^n Q_Y(y_j)$ with Q_Y being the capacity-achieving output distribution [5].

For this class of channels we can rewrite (39) using the decomposition of the message set into N_k source-type classes as

$$\epsilon \geq 1 - \sup_{Q_{\boldsymbol{Z}}} \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \frac{1}{|\mathcal{T}_i^k|} \sum_{\boldsymbol{v}' \in \mathcal{T}_i^k} \alpha_{Q_{\boldsymbol{Z}}(\boldsymbol{v}')}(Q_{\boldsymbol{Y}}). \quad (40)$$

Although $\alpha_b(Q_{\boldsymbol{Y}})$ is independent of $\boldsymbol{x}$, the outer sum in (40) still depends on the codebook and the decoder through $Q_{\boldsymbol{Z}}(\cdot)$. Hence, the optimization in (40) can be performed over all possible distributions $Q_{\boldsymbol{Z}}(\boldsymbol{v})$, $\boldsymbol{v} \in \mathcal{V}^k$. Given that $\alpha_b(\cdot, \cdot)$ is concave with respect to $b \in [0, 1]$ (see Appendix A) this is a convex optimization problem. However, since the minimization must be carried out over an exponentially large number of elements, the optimization soon becomes computationally infeasible as the message length increases. A possible approach to simplify the aforementioned drawback is to weaken (40) using Jensen's inequality.

Theorem 1: The average error probability of every source-channel code in a symmetric channel is lower-bounded as

$$\epsilon \geq 1 - \sup_{Q_{\mathcal{T}}} \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \alpha_{\frac{Q_{\mathcal{T}}(i)}{|\mathcal{T}_i^k|}}(Q_{\boldsymbol{Y}}), \quad (41)$$

where $Q_{\mathcal{T}}(i) \triangleq \sum_{\boldsymbol{v} \in \mathcal{T}_i^k} Q_{\boldsymbol{Z}}(\boldsymbol{v})$, $i = 1, \ldots N_k$, and $Q_{\boldsymbol{Y}}$ is the capacity-achieving output distribution.

Proof: Using Jensen's inequality, we obtain

$$\epsilon \geq 1 - \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \frac{1}{|\mathcal{T}_i^k|} \sum_{\boldsymbol{v}' \in \mathcal{T}_i^k} \alpha_{Q_{\boldsymbol{Z}}(\boldsymbol{v}')}(Q_{\boldsymbol{Y}}) \quad (42)$$

$$\geq 1 - \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \alpha_{\left\{\frac{\sum_{\boldsymbol{v} \in \mathcal{T}_i^k} Q_{\boldsymbol{Z}}(\boldsymbol{v})}{|\mathcal{T}_i^k|}\right\}}(Q_{\boldsymbol{Y}}) \quad (43)$$

$$= 1 - \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \alpha_{\frac{Q_{\mathcal{T}}(i)}{|\mathcal{T}_i^k|}}(Q_{\boldsymbol{Y}}). \quad (44)$$

$\blacksquare$

Theorem 1 depends on the codebook and the decoder only through the distribution $Q_{\mathcal{T}}(\cdot)$, and therefore it is optimized over all distributions defined over source-type classes. Since
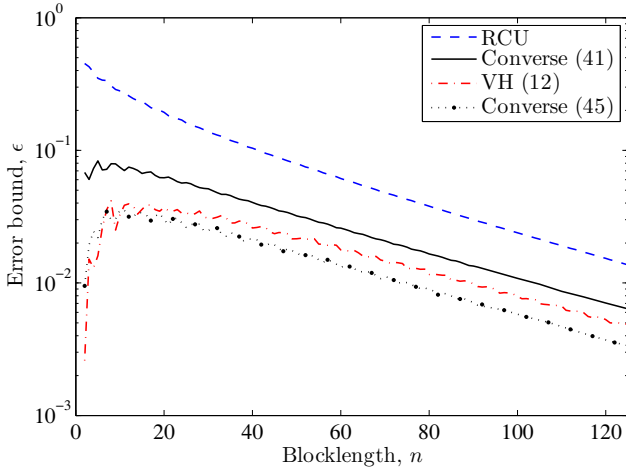
Figure 1. Upper and lower bounds for a BMS-BSC pair. Parameters: $P_V(1) = 0.05$, $P_{Y|X}(1|0) = P_{Y|X}(0|1) = 0.1$, $t = 1$.

Figure 2. Upper and lower bounds for a BMS-BEC pair. Parameters: $P_V(1) = 0.001$, $P_{Y|X}(e|0) = P_{Y|X}(e|1) = 0.95$, $t = 1$.

the dimension of the domain of $Q_{\mathcal{T}}$ grows polynomially with the block length n, this involves an (exponentially) less complex computation than that of (40).

Remark 1: It can be checked (for instance, by performing the method of the Lagrange multipliers) that the optimizing $Q_{\mathbf{Z}}(\cdot)$ of (40) is uniform over the values of $\mathbf{v}$ belonging to the same source-type class. Hence, the optimizing $Q_{\mathcal{T}}(\cdot)$ in (41) induces the optimizing $Q_{\mathbf{Z}}(\cdot)$ of (40) and as a consequence, both bounds coincide. This is tantamount to state that Theorem 1 also recovers the generalization of the Verdú-Han lemma and attains Csiszár's sphere-packing exponent.

Theorem 1 can be weakened to obtain a converse bound that does not require an optimization over the distribution $Q_{\mathcal{T}}$. Using the fact that $\alpha_b(Q_{\mathbf{Y}})$ is a non-decreasing function of $b \in [0, 1]$ and upper-bounding $Q_{\mathcal{T}}(i) \leq 1$, $i = 1, \ldots N_k$ in (41) we obtain the following result.

Corollary 1: The average error probability of every source-channel code in a symmetric channel is lower-bounded as

$$\epsilon \geq 1 - \sum_{i=1}^{N_k} \Pr\left\{\mathcal{T}_i^k\right\} \alpha_{\frac{1}{|\mathcal{T}_i^k|}}(Q_{\mathbf{Y}}), \qquad (45)$$

where $Q_{\mathbf{Y}}$ is the capacity-achieving output distribution.

Equation (45) does not depend on the decoder nor the codebook, and thus, it gives directly a computable converse result. While Corollary 1 cannot be used to recover (12) using (13), it still can be shown to attain Csiszár's exponent by applying the arguments in [5, Sec. III-F] for each source-type class.

We next compare the finite length-bounds given in Theorem 1 and Corollary 1 for two source-channel pairs: a binary memoryless source (BMS) transmitted over a binary symmetric channel (BSC) and over a binary erasure channel (BEC) respectively.
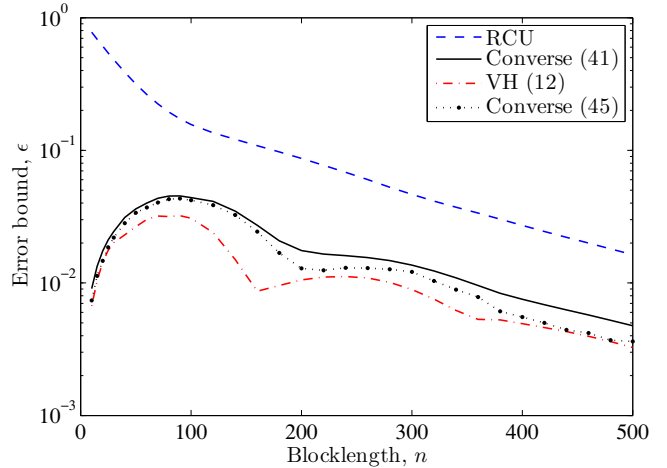
A. BMS-BSC

In this example we consider a source-channel pair given by a BMS with $P_V(1) = 0.05$ and a BSC with crossover probability $P_{Y|X}(1|0) = P_{Y|X}(0|1) = 0.1$, $t = 1$. We now compare the finite length-bounds given in (41) from Theorem 1 (computed using [19]) and (45) from Corollary 1 with respect to the Verdú-Han lemma (12), and the RCU upper bound [8] corresponding to a random-coding ensemble generated with the product uniform distribution.

For the BSC the capacity-achieving output distribution is the uniform distribution. For this choice of $Q_{\mathbf{Y}}$ and for $\mathbf{Y}$ distributed according to $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$, the random variable $\Psi_x = \frac{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{x})}{Q_{\mathbf{Y}}(\mathbf{Y})}$ is independent of $\mathbf{x}$, and so it is $\Pr\{P_{\mathbf{V}}(\mathbf{V})\Psi_x < \gamma\}$. Hence, in this case it is possible to compute a valid lower bound from the Verdú-Han lemma without resorting to an optimization over all possible codebooks.

Fig. 1 shows that in this scenario the bound (45) is looser than (12) while the bound (41) is tighter in the range of block lengths shown. This agrees with the fact that while (12) can be derived by weakening Thm. 1, this is not possible from Cor. 1.

B. BMS-BEC

We now consider an scenario of a BMS with $P_V(1) = 0.001$ which needs to be transmitted over a BEC with erasure probability $P_{Y|X}(e|0) = P_{Y|X}(e|1) = 0.95$, $t = 1$. Fig. 2 shows the same plot as Fig. 1 for this source-channel pair. From the figure we observe that in this case the lower bound (45) is tighter than (12), hence none of these two bounds dominates in general. Similarly to the previous case, the bound (41) improves the other two.

APPENDIX A
CONCAVITY OF THE FUNCTION α_b

Lemma 3: The function $\alpha_b \equiv \alpha_b(\cdot, \cdot)$ is a concave function with respect to b in $[0, 1]$.

Proof: Consider $\alpha_b, \alpha_{b'}$, where $b, b' \in [0,1]$ and denote $\beta_a \equiv \beta_a(\cdot, \cdot)$. Since β_α is convex, this implies $\forall \lambda \in [0,1]$ that

$$\beta\left(\lambda\alpha_b + (1-\lambda)\alpha_{b'}\right) \le \lambda\beta_{\alpha_b} + (1-\lambda)\beta_{\alpha_{b'}} \qquad (46)$$
$$\le \lambda b + (1-\lambda)b', \qquad (47)$$

where $\beta_{\alpha_b} \le b$ in (47) follows from (36). If $\lambda b + (1-\lambda)b' \le \beta_{\max}$, the monotonicity of β_α implies that

$$\lambda\alpha_b + (1-\lambda)\alpha_{b'} \le \alpha_{\lambda b + (1-\lambda)b'} \qquad (48)$$

on account of (36). Otherwise, if $\lambda b + (1-\lambda)b' > \beta_{\max}$, we have that

$$\lambda\alpha_b + (1-\lambda)\alpha_{b'} \le 1 \qquad (49)$$
$$= \alpha_{\lambda b + (1-\lambda)b'}, \qquad (50)$$

where (49) follows from $\alpha_b, \alpha_{b'} \in [0,1]$ and (50) from $\lambda b + (1-\lambda)b' > \beta_{\max}$ and (36). $\blacksquare$

REFERENCES

[1] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, 1973.

[2] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois J. Math.*, vol. 1, pp. 591–606, 1957.

[3] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[4] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.

[5] 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.

[6] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[7] T. S. Han, "Joint source-channel coding revisited: Information-spectrum approach," *arxiv preprint arXiv:0712.2959v1*, 2007.

[8] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Martinez, "Random-coding joint source-channel coding bounds," in *Proc. IEEE Int. Symp. on Inf. Theory*, Saint Petersburg, Russia, July-Aug. 2011.

[9] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *arXiv preprint cs/1209.1317v1*, 2012.

[10] 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, no. 694-706, p. 289, 1933.

[11] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inf. Theory*, vol. 9, pp. 315–328, 1980.

[12] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

[13] F. Jelinek, *Probabilistic Information Theory*. New York: McGraw-Hill, 1968.

[14] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Contr.*, vol. 10, no. 1, pp. 65–103, 1967.

[15] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1450–1468, April 2006.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[17] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.

[19] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip