Random Coding Bounds that Attain the Joint Source-Channel Exponent

Adrià Tauste Campo[1], Gonzalo Vazquez-Vilar[2], Albert Guillén i Fàbregas[123], Tobias Koch[1] and Alfonso Martinez[2]
[1]University of Cambridge, [2]Universitat Pompeu Fabra, [3]Institució Catalana de Recerca i Estudis Avançats (ICREA)
Email: {atauste,gvazquez,guillen,alfonso.martinez}@ieee.org, tobi.koch@eng.cam.ac.uk

***Abstract*—This paper presents a random-coding upper bound on the average error probability of joint source-channel coding that attains Csiszár's error exponent. The bound is based on a code construction for which source messages are assigned to disjoint subsets (classes), and codewords generated according to a distribution that depends on the class of the source message. For a single class, the bound recovers Gallager's exponent; identifying the classes with source type classes, it recovers Csiszár's exponent. Moreover, it is shown that as a two appropriately designed classes are sufficient to attain Csiszár's exponent.**

I. Introduction

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

In this joint source-channel coding (JSCC) setup, the encoder maps the source message $\boldsymbol{v}$ to a length-n codeword $\boldsymbol{x}(\boldsymbol{v})$, which is then transmitted over the channel. We refer to the ratio $t \triangleq k/n$ as the transmission rate. Based on the channel output $\boldsymbol{y}$, the decoder guesses a source message $\hat{\boldsymbol{v}}$ according to the maximum a posteriori (MAP) criterion, i.e.,

$$\hat{\boldsymbol{v}} = \arg\max_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v})\big). \tag{1}$$

When clear from the context, we shall simplify notation by writing $\boldsymbol{x}$ instead of $\boldsymbol{x}(\boldsymbol{v})$, making the message $\boldsymbol{v}$ implicit.

We study the random-coding average error probability $\bar{\epsilon}$ by means of the random-coding union (RCU) bound [1], [2]:

$$\bar{\epsilon} \le \mathbb{E}\left[\min\left\{1, \sum_{\boldsymbol{v}' \neq \boldsymbol{V}} \Pr\left\{\frac{P_{\boldsymbol{V}}(\boldsymbol{v}') P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\bar{\boldsymbol{X}})}{P_{\boldsymbol{V}}(\boldsymbol{V}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})} \ge 1 \,\middle|\, \boldsymbol{V}\boldsymbol{X}\boldsymbol{Y}\right\}\right\}\right], \tag{2}$$

where the expectation is taken according to the joint distribution $P_{\boldsymbol{V}} P_{\boldsymbol{X}|\boldsymbol{V}} P_{\boldsymbol{Y}|\boldsymbol{X}}$ and the probability computed with respect to the distribution $P_{\boldsymbol{X}|\boldsymbol{V}}$ for each $\boldsymbol{v}'$ in the summation.

Optimization over the conditional distributions $P_{\boldsymbol{X}|\boldsymbol{V}}$, necessary to obtain the tightest possible bound in (2), quickly becomes computationally unfeasible as the block length grows large. In this paper, we focus instead on the exponential decay of (2) with respect to n and slightly loosen the RCU bound into a convenient form that is proven to attain the JSCC exponent found by Csiszár [3].

II. Previous work

In [4, Prob. 5.16] Gallager provided an upper bound on $\bar{\epsilon}$ when the codewords corresponding to different source messages are drawn independently according to a distribution $P_{\boldsymbol{X}}$:

$$\bar{\epsilon} \le e^{-E_0(\rho, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}) + E_{\mathrm{s}}(\rho, P_{\boldsymbol{V}})}, \quad \text{for every } \rho \in [0,1], \tag{3}$$

where $E_0(\rho, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}})$ denotes Gallager's *channel function*,

$$E_0(\rho, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}) \triangleq -\log \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{x}} P_{\boldsymbol{X}}(\boldsymbol{x}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}, \tag{4}$$

and where $E_{\mathrm{s}}(\rho, P_{\boldsymbol{V}})$ denotes Gallager's *source function*,

$$E_{\mathrm{s}}(\rho, P_{\boldsymbol{V}}) \triangleq \log \left(\sum_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v})^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{5}$$

While the derivation of (3) assumes that $P_{\boldsymbol{X}}$ is independent of the source message, it may be proved that considering an arbitrary distribution $P_{\boldsymbol{X}|\boldsymbol{V}}$ does not improve the bound.

When $P_{\boldsymbol{X}}$ is a product distribution, i.e., $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^n P_X(x_i)$, the bound in (3) becomes

$$\bar{\epsilon} \le e^{-n\left(E_0(\rho, P_{Y|X}, P_X) - tE_{\mathrm{s}}(\rho, P_V)\right)}, \tag{6}$$

thus proving that $\bar{\epsilon}$ exponentially vanishes with respect to n. By maximizing over P_X and ρ, this bound provides a lower bound on the JSCC exponent E_{J}

$$E_{\mathrm{J}} \ge E_{\mathrm{J}}^{\mathrm{G}} \triangleq \max_{\rho \in [0,1]} \big\{E_0(\rho, P_{Y|X}) - tE_{\mathrm{s}}(\rho, P_V)\big\}, \tag{7}$$

where we define $E_0(\rho, P_{Y|X}) \triangleq \max_{P_X} E_0(\rho, P_{Y|X}, P_X)$.

Csiszár refined Gallager's result using a code construction based on *fixed composition codes* [3]. Specifically, he showed

This work has been supported in part by the European Research Council under ERC grant agreement 259663. The work by A. Tauste Campo has been funded from an EPSRC Doctoral Prize Award. The work by T. Koch has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 252663. The work by A. Martinez received funding from the Ministry of Science and Innovation (Spain) under grant RYC-2011-08150.

that for all $\delta > 0$, there exists an $n_0 \in \mathbb{N}$ such that, for $n \geq n_0$ $\bar{\epsilon}$ is upper-bounded as

$$\bar{\epsilon} \leq \sum_{i=1}^{M_k} e^{-n\left(e(R_i,P_V)+E_r(R_i,P_{Y|X})-2\delta\right)}, \tag{8}$$

where M_k is the number of source-type classes in $\mathcal{V}^k$ [5], $e(R,P_V)$ is the source reliability function [5]–[7]

$$e(R,P_V) \triangleq \sup_{\rho \geq 0}\bigl\{\rho R - E_{\mathrm{s}}(\rho,P_V)\bigr\}, \tag{9}$$

and where $E_{\mathrm{r}}(R,P_{Y|X})$ is the channel random-coding exponent, given by [4]

$$E_{\mathrm{r}}(R,P_{Y|X}) \triangleq \max_{\rho\in[0,1]}\bigl\{E_0(\rho,P_{Y|X}) - \rho R\bigr\}. \tag{10}$$

Eq. (8) leads to another lower bound on the exponent E_{J},

$$\begin{aligned} E_{\mathrm{J}} &\geq E_{\mathrm{J}}^{\mathrm{Cs}} \qquad (11)\\ &\triangleq \min_{H(V)\leq R\leq R_{\mathcal{V}}}\left\{te\left(\frac{R}{t},P_V\right) + E_{\mathrm{r}}(R,P_{Y|X})\right\}, \qquad (12)\end{aligned}$$

where $R_{\mathcal{V}} \triangleq t\log|\mathcal{V}|$. Comparing (12) with the sphere packing exponent [8], Csiszár verified that $E_{\mathrm{J}}^{\mathrm{Cs}}$ gives the actual JSCC error exponent E_{J} when the minimum in (12) is attained at a rate $R > R_{\mathrm{cr}}$, where R_{cr} is the critical rate of the channel [3].

Zhong *et al.* [9] quantified the improvement of Csiszár's exponent (12) over Gallager's (7) via Fenchel's duality theorem [10, Thm. 31.1], which allows one to rewrite (12) as

$$E_{\mathrm{J}}^{\mathrm{Cs}} = \max_{\rho\in[0,1]}\bigl\{\bar{E}_0(\rho,P_{Y|X}) - tE_{\mathrm{s}}(\rho,P_V)\bigr\}, \tag{13}$$

where $\bar{E}_0(\rho,P_{Y|X})$ denotes the concave hull of $E_0(\rho,P_{Y|X})$, defined as the pointwise infimum over the family of affine functions that upper-bound $E_0(\rho,P_{Y|X})$, as a function of ρ in $\rho\in[0,1]$ [10, Cor. 12.1.1]. It follows that $E_{\mathrm{J}}^{\mathrm{Cs}} \geq E_{\mathrm{J}}^{\mathrm{G}}$, with the inequality possibly strict, as shown in an example in [9].

III. Random Coding Bound

We have recently proposed a random-coding upper bound which attains Csiszár's exponent [11]. The derivation of this bound involves the following steps:

1) Define a partition $\mathcal{P}_k$ of the message set $\mathcal{V}^k$ into N_k disjoint subsets $\mathcal{A}_1,\dots,\mathcal{A}_{N_k}$ satisfying $\bigcup_{i=1}^{N_k}\mathcal{A}_i = \mathcal{V}^k$. We shall refer to these subsets as *classes*.
2) Assign a channel input distribution $P_{\boldsymbol{X}}^{(i)}$ to each class $\mathcal{A}_i$. Then, for each source message $\boldsymbol{v}\in\mathcal{A}_i$ randomly and independently generate codewords $\boldsymbol{x}(\boldsymbol{v}) \in \mathcal{X}^n$ according to $P_{\boldsymbol{X}}^{(i)}$.
3) Upper-bound the probability of error using Gallager's bounding techniques [4].

In the following we define

$$E_{\mathrm{s}}^{(i)}(\rho,P_{\boldsymbol{V}}) \triangleq \log\left(\sum_{\boldsymbol{v}\in\mathcal{A}_i} P_{\boldsymbol{V}}(\boldsymbol{v})^{\frac{1}{1+\rho}}\right)^{1+\rho}, \tag{14}$$

for $i = 1,\dots,N_k$.

Theorem 1: For every partition $\mathcal{P}_k$, for every set of product channel input distributions $P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) = \prod_{i=1}^n P_X^{(i)}(x_i)$, $i = 1,\dots,N_k$, and for every set of parameters $\rho_1,\dots,\rho_{N_k}\in[0,1]$, the average probability of error is upper-bounded by

$$\begin{aligned}\bar{\epsilon} &\leq \bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \qquad (15)\\ &\triangleq h(k)\sum_{i=1}^{N_k} e^{-\max_{\rho_i\in[0,1]}\left\{nE_0(\rho_i,P_{Y|X})-E_{\mathrm{s}}^{(i)}(\rho_i,P_{\boldsymbol{V}})\right\}}, \qquad (16)\end{aligned}$$

where $h(k) \triangleq 2N_k(k+1)^{|\mathcal{V}|}(k/t+1)^{|\mathcal{X}||\mathcal{Y}|}$.

Proof: See [11]. ■

If we choose the partition $\mathcal{P}_k$ such that $N_k = 1$ and $\mathcal{A}_1 = \mathcal{V}^k$ for $k = 1,2,\dots$, then $E_{\mathrm{s}}^{(i)}(\rho,P_{\boldsymbol{V}}) = E_{\mathrm{s}}(\rho,P_{\boldsymbol{V}})$ and $\log h(k)/k \to 0$ as $n\to\infty$. Hence, (15) recovers Gallager's bound on the error exponent (7)

$$\lim_{n\to\infty} -\frac{1}{n}\log\bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) = E_{\mathrm{J}}^{\mathrm{G}}. \tag{17}$$

With a more judicious choice of $\mathcal{P}_k$ the upper bound (15) also recovers Csiszár's lower bound on the error exponent (12). Specifically, (12) can be achieved by identifying the classes $\mathcal{A}_1,\dots,\mathcal{A}_{N_k}$ with the source-type classes $\mathcal{T}_1,\dots,\mathcal{T}_{N_k}$. A source-type class $\mathcal{T}_i$ is defined as the set of all source messages $\boldsymbol{v}\in\mathcal{V}^k$ with type P_i [5, Def. 2.1]. Thus, for a given distribution P_i on $\mathcal{V}$, the source-type class $\mathcal{T}_i$ is the set of all source messages $\boldsymbol{v}\in\mathcal{V}^k$ satisfying

$$\mathsf{P}_i(a) = \frac{1}{k}\mathsf{N}(a|\boldsymbol{v}), \quad a\in\mathcal{V}, \tag{18}$$

where $\mathsf{N}(a|\boldsymbol{v})$ denotes the number of occurrences of $a\in\mathcal{V}$ in $\boldsymbol{v}$. With this choice of $\mathcal{P}_k$ it can be shown [11] that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \geq E_{\mathrm{J}}^{\mathrm{Cs}}. \tag{19}$$

As the number of classes used to optimize the bounds (7) and (12) ranges from one (Gallager) to a polynomial function of k (Csiszár), one may pose the natural question of how many channel input distributions are needed to attain the exponent. We next show that two classes, and therefore two associated input product distributions, suffice to attain Csiszár's exponent.

A. Attaining Csiszár's exponent with two classes

Let $\mathcal{T}(\boldsymbol{v})$ denote the source-type class associated to message $\boldsymbol{v}$. We define the partition $\hat{\mathcal{P}}_k(\lambda_0)$ as follows. For some $\lambda_0 \geq 0$ and every $k\geq 1$, we assign the source messages into two sets, $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively defined as

$$\mathcal{A}_1 \triangleq \bigl\{\boldsymbol{v} : |\mathcal{T}(\boldsymbol{v})| \geq e^{tk\lambda_0}\bigr\}, \tag{20}$$

$$\mathcal{A}_2 \triangleq \bigl\{\boldsymbol{v} : |\mathcal{T}(\boldsymbol{v})| < e^{tk\lambda_0}\bigr\}. \tag{21}$$

Then, we have the following theorem.

Theorem 2: Consider the family of partitions $\{\mathcal{P}_k = \hat{\mathcal{P}}_k(\lambda_0), \lambda_0\in[0,R_{\mathcal{V}}]\}$ for every $k\geq 1$. Then,

$$\sup_{\lambda_0\in[0,R_{\mathcal{V}}]}\lim_{n\to\infty} -\frac{1}{n}\log\bar{\epsilon}_{\mathrm{B}}(\hat{\mathcal{P}}_k(\lambda_0)) \geq E_{\mathrm{J}}^{\mathrm{Cs}}. \tag{22}$$

Proof: See Section V. ■

The division of $\mathcal{V}^k$ into the classes $\mathcal{A}_1$ and $\mathcal{A}_2$ for the best threshold $\lambda_0^\star$, together with the optimal distributions $P_X^{(1)}$ and $P_X^{(2)}$, induce a conditional distribution $P_{\boldsymbol{X}|\boldsymbol{V}}(\boldsymbol{x}|\boldsymbol{v})$ given by

$$P_{\boldsymbol{X}|\boldsymbol{V}}(\boldsymbol{x}|\boldsymbol{v}) = \begin{cases}\prod_{i=1}^n P_X^{(1)}(x_i), & \boldsymbol{v}\in\mathcal{A}_1,\\ \prod_{i=1}^n P_X^{(2)}(x_i), & \boldsymbol{v}\in\mathcal{A}_2.\end{cases} \tag{23}$$

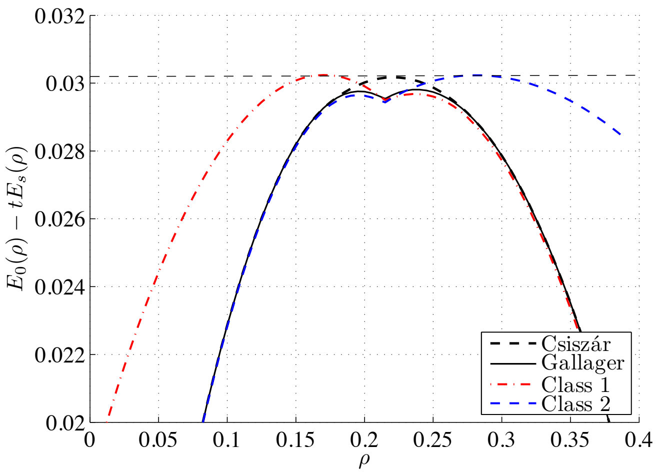

Figure 1. Error exponent bounds. Csiszár's and Gallager's curves correspond to $\bar{E}_0(\rho, P_{Y|X}) - tE_s(\rho, P_V)$ and $E_0(\rho, P_{Y|X}) - tE_s(\rho, P_V)$, respectively. Single class correspond to $E_0(\rho, P_{Y|X}) - \lim_{n\to\infty} \frac{1}{n} E_s^{(i)}(\rho, P_{\boldsymbol{V}})$, for $i = 1, 2$.

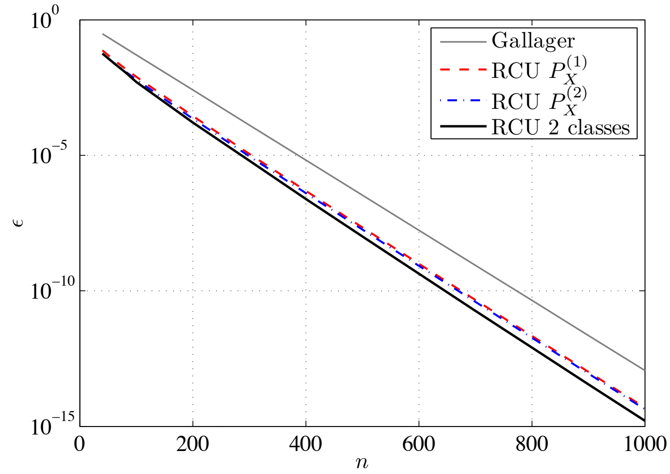

Figure 2. Random coding upper bounds to the error probability ϵ.

By construction, the new random coding bound (16) is looser with respect to the RCU bound (2) for all n and $P_{\boldsymbol{X}|\boldsymbol{V}}$. Thus, when the codewords are generated according to (23), the resulting RCU bound (2) attains an exponent at least as large as Csiszár's. Furthermore, if the minimum in (12) is attained for $R > R_{\text{cr}}$, then the resulting RCU attains the JSCC exponent.

IV. Example

Consider a binary memoryless source (BMS) and a non-symmetric memoryless channel with $|\mathcal{X}| = 6$, $|\mathcal{Y}| = 4$ and transition-probability matrix

$$P_{Y|X} = \begin{pmatrix} 1-3\xi_1 & \xi_1 & \xi_1 & \xi_1 \\ \xi_1 & 1-3\xi_1 & \xi_1 & \xi_1 \\ \xi_1 & \xi_1 & 1-3\xi_1 & \xi_1 \\ \xi_1 & \xi_1 & \xi_1 & 1-3\xi_1 \\ \frac{1}{2}-\xi_2 & \frac{1}{2}-\xi_2 & \xi_2 & \xi_2 \\ \xi_2 & \xi_2 & \frac{1}{2}-\xi_2 & \frac{1}{2}-\xi_2 \end{pmatrix}, \quad (24)$$

which is similar to the channel given in [4, Fig. 5.6.5]. Note that this channel is composed of two quaternary-output sub-channels. One sub-channel is a quaternary-input symmetric channel with parameter ξ_1 and the other one is a binary-input channel with parameter ξ_2. In this example we set $\xi_1 = 0.068$, $\xi_2 = 0.01$, $t = 2$ and $P_V(1) = 0.041$. Therefore, the source entropy is $H(V) = 0.2469$ bits/source symbol, the channel capacity is $C = 0.9468$ bits/channel use, and the critical rate is $R_{\text{cr}} = 0.4564$ bits/channel use.

As Gallager observed, optimizing the $E_0(\rho, P_{Y|X}, P_X)$ function over the input distribution may lead to a discontinuity of the derivative of the function $E_0(\rho, P_{Y|X})$ with respect to ρ. In this example, the optimal distribution abruptly changes from $P_X^{(1)} = \left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ 0\ 0\right)$ to $P_X^{(2)} = \left(0\ 0\ 0\ 0\ \frac{1}{2}\ \frac{1}{2}\right)$ for some $\rho \in [0, 1]$. In turn, this implies that $E_0(\rho, P_{Y|X})$ is not concave. Therefore, for some ρ, $\bar{E}_0(\rho, P_{Y|X}) > E_0(\rho, P_{Y|X})$.

In Fig. 1 we plot several bounds on the JSCC exponent based on the aforementioned random-coding bounds. For Gallager and Csiszár exponents we use the arguments in (7) and (13) as a function of ρ respectively. The figure shows how the non-concavity of Gallager's exponent function around the optimal ρ of Csiszár's exponent translates into a loss in exponent. For reference purposes, Fig. 1 shows the value of Csiszár's exponent with a horizontal dashed line. For the random-coding bound (16) with two classes, we apply Theorem 2 and plot the exponent of each individual class as a function of ρ. The optimum value of λ_0 that determines the partition is given by $\lambda_0^\star = 0.732$. The resulting exponent is obtained by first individually maximizing the exponent of each class over ρ, and then choosing the minimum. The figure illustrates that for the optimal partition, the exponent of both classes coincides with Csiszár's.

In Fig. 2 we compare Gallager's upper bound (6), the RCU bounds with a single input distribution (either $P_X^{(1)}$ or $P_X^{(2)}$) and the RCU bound when the two-class construction is employed and each class distribution independently optimized. While the single-class RCU bound is tighter than Gallager's bound, it attains the same asymptotic slope. The two-class construction achieves a tighter upper bound on the random coding error probability. In this example, Csiszár's exponent (12) is attained at $R^\star = 0.7273 > R_{\text{cr}}$ and thus, E_{J}^{Cs} is tight and gives the JSCC exponent.

V. Proof of Theorem 2

This section provides proof of Theorem 2. To this end we first introduce two lemmas which are then used in the derivation of the main result.

Consider the partition $\hat{\mathcal{P}}_k(\lambda_0)$ defined by (20)-(21). By noting that $E_{\text{s}}^{(i)}(\rho) \triangleq E_{\text{s}}^{(i)}(\rho, P_{\boldsymbol{V}})$ is a continuous first-order differentiable function of ρ we define, for $i = 1, 2$,

$$\lambda_i(\rho) \triangleq \frac{1}{n} \frac{\partial E_{\text{s}}^{(i)}(\rho)}{\partial \rho}. \quad (25)$$

For future reference, we also define

$$F_1(\lambda, \rho) \triangleq \max_{\rho_1 \in [0,\rho]} \left\{ E_0(\rho_1, P_{Y|X}) + \lambda(\rho - \rho_1) \right\}, \quad (26)$$

$$F_2(\lambda, \rho) \triangleq \max_{\rho_2 \in [\rho,1]} \left\{ E_0(\rho_2, P_{Y|X}) + \lambda(\rho - \rho_2) \right\}. \quad (27)$$

Lemma 1: For every $\rho \in [0,1]$, the limits $\lim_{n\to\infty} \lambda_i(\rho)$, $i = 1,2$ exist. Furthermore, for every $0 \le \lambda_0 \le R_{\mathcal{V}}$ we have

$$\lim_{n\to\infty} \lambda_2(1) \;\le\; \lambda_0 \;\le\; \lim_{n\to\infty} \lambda_1(0). \tag{28}$$

Lemma 2: For $\rho \in [\hat{\rho}, 1]$ where $\hat{\rho}$ is the smallest value $\hat{\rho}_\star$ that satisfies $\hat{\rho}_\star = \arg\max_{\rho' \in [0,1]} \left\{ E_0(\rho', P_{Y|X}) - \rho' R_{\mathcal{V}} \right\}$, it holds that

$$\max_{\lambda\in[0,R_{\mathcal{V}}]} \min_{i=1,2} \big\{ F_i(\lambda,\rho) \big\} = \min_{\lambda\in[0,R_{\mathcal{V}}]} \max_{i=1,2} \big\{ F_i(\lambda,\rho) \big\}. \tag{29}$$

Armed with the above two lemmas, we proceed to prove the lower bound (22). Since $E_{\mathrm{s}}^{(i)}(\rho)$ are convex non-decreasing functions of ρ for $i = 1,2$ we have that

1) $\lambda_i(\rho) \ge 0$ for $\rho \in [0,1]$, and
2) $\lambda_i(0) \le \lambda_i(\rho_1) \le \lambda_i(\rho_2) \le \lambda_i(1)$ for $0 \le \rho_1 \le \rho_2 \le 1$.

Then, using the first order Taylor expansion of $E_{\mathrm{s}}^{(i)}(\rho_i)$ at a given $\rho_i \in [0,1]$, we can bound $E_{\mathrm{s}}^{(i)}(\rho_0)$, for $i = 1,2$ and $\rho_0 \in [0,1]$, as

$$\begin{aligned} E_{\mathrm{s}}^{(1)}(\rho_0) &\ge E_{\mathrm{s}}^{(1)}(\rho_1) + n\lambda_1(\rho_1)(\rho_0 - \rho_1) && (30)\\ &\ge E_{\mathrm{s}}^{(1)}(\rho_1) + n\lambda_1(0)(\rho_0 - \rho_1), \quad \rho_1 \le \rho_0, && (31)\\ E_{\mathrm{s}}^{(2)}(\rho_0) &\ge E_{\mathrm{s}}^{(2)}(\rho_2) + n\lambda_2(\rho_2)(\rho_0 - \rho_2) && (32)\\ &\ge E_{\mathrm{s}}^{(2)}(\rho_2) + n\lambda_2(1)(\rho_0 - \rho_2), \quad \rho_2 \ge \rho_0. && (33) \end{aligned}$$

Since the limit $\lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho)$ exists for $\rho \in [0,1]$, Lemma 1, together with (31) and (33), shows that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(1)}(\rho_0) &\ge \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(1)}(\rho_1) + \lambda_0(\rho_0 - \rho_1), && (34)\\ \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(2)}(\rho_0) &\ge \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(2)}(\rho_2) + \lambda_0(\rho_0 - \rho_2), && (35) \end{aligned}$$

for $\rho_1 \in [0, \rho_0]$ and $\rho_2 \in [\rho_0, 1]$.

From (15) we have that, for an arbitrary dummy variable $\rho_0 \in [0,1]$,

$$\begin{aligned} &\lim_{n\to\infty} -\frac{1}{n} \log \bar{\epsilon}_{\mathrm{B}}(\hat{\mathcal{P}}_k(\lambda_0))\\ &= \lim_{n\to\infty} -\frac{1}{n} \log \Bigg(h(tn) \\ &\qquad\qquad \times \sum_{i=1,2} e^{-\max_{\rho_i\in[0,1]} \left\{ nE_0\left(\rho_i, P_{Y|X}\right) - E_{\mathrm{s}}^{(i)}(\rho_i) \right\}} \Bigg) && (36)\\ &= \lim_{n\to\infty} \min_{i=1,2} \left\{ \max_{\rho_i\in[0,1]} \left\{ E_0(\rho_i, P_{Y|X}) - \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho_i) \right\} \right\} && (37)\\ &= \min_{i=1,2} \left\{ \max_{\rho_i\in[0,1]} \left\{ E_0(\rho_i, P_{Y|X}) - \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho_i) \right\} \right\} && (38)\\ &\ge \min \Bigg\{ \max_{\rho_1\in[0,\rho_0]} \left\{ E_0(\rho_1, P_{Y|X}) - \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho_1) \right\}, \\ &\qquad\quad \max_{\rho_2\in[\rho_0,1]} \left\{ E_0(\rho_2, P_{Y|X}) - \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho_2) \right\} \Bigg\} && (39)\\ &\ge \min_{i=1,2} \left\{ F_i(\lambda_0, \rho_0) - \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho_0) \right\} && (40)\\ &\ge \min_{i=1,2} \big\{ F_i(\lambda_0, \rho_0) \big\} - tE_{\mathrm{s}}(\rho_0, P_V), && (41) \end{aligned}$$

where (37) follows by noting that $h(tn)$ is subexponential in n; in (38) we used that the limit exists for each set of the parameters over which the optimization is performed; (39) follows from restricting the intervals over which ρ_1 and ρ_2 are maximized; in (40) we applied (34)-(35) and used the definition of $F_i(\lambda, \rho)$, $i = 1,2$, cf. (26)-(27); and in (41) we used that, by definition, $E_{\mathrm{s}}^{(i)}(\rho) \le E_{\mathrm{s}}(\rho, P_{\boldsymbol{V}}) = kE_{\mathrm{s}}(\rho, P_V)$ since the sum is taken over a smaller number of terms.

As (36)-(41) hold for arbitrary $\rho_0 \in [\hat{\rho}, 1]$ (with $\hat{\rho}$ defined in Lemma 2), we obtain upon maximizing over $\lambda \in [0, R_{\mathcal{V}}]$ and $\rho_0 \in [\hat{\rho}, 1]$

$$\begin{aligned} &\sup_{\lambda_0\in[0,R_{\mathcal{V}}]} \lim_{n\to\infty} -\frac{1}{n} \log \bar{\epsilon}_{\mathrm{B}}(\hat{\mathcal{P}}_k(\lambda_0))\\ &\ge \max_{\rho_0\in[\hat{\rho},1]} \left\{ \max_{\lambda_0\in[0,R_{\mathcal{V}}]} \min_{i=1,2} \big\{ F_i(\lambda_0,\rho_0) \big\} - tE_{\mathrm{s}}(\rho_0, P_V) \right\} && (42)\\ &= \max_{\rho_0\in[\hat{\rho},1]} \left\{ \min_{\lambda_0\in[0,R_{\mathcal{V}}]} \max_{i=1,2} \big\{ F_i(\lambda_0,\rho_0) \big\} - tE_{\mathrm{s}}(\rho_0, P_V) \right\} && (43)\\ &= \max_{\rho_0\in[\hat{\rho},1]} \Bigg\{ \min_{\lambda_0\in[0,R_{\mathcal{V}}]} \max_{\rho\in[0,1]} \big\{ E_0(\rho, P_{Y|X}) + \lambda_0(\rho_0 - \rho) \big\} \\ &\qquad\qquad\qquad - tE_{\mathrm{s}}(\rho_0, P_V) \Bigg\} && (44)\\ &\ge \max_{\rho_0\in[\hat{\rho},1]} \Bigg\{ \min_{\lambda_0\ge 0} \max_{\rho\in[0,1]} \big\{ E_0(\rho, P_{Y|X}) + \lambda_0(\rho_0 - \rho) \big\} \\ &\qquad\qquad\qquad - tE_{\mathrm{s}}(\rho_0, P_V) \Bigg\} && (45)\\ &= \max_{\rho_0\in[\hat{\rho},1]} \big\{ \bar{E}_0(\rho_0, P_{Y|X}) - tE_{\mathrm{s}}(\rho_0, P_V) \big\} && (46) \end{aligned}$$

where in (42) we used that $F_i(\lambda, \rho)$, $i = 1,2$ are continuous functions of λ to write a maximum instead of a supremum; in (43) we applied Lemma 2; (44) follows from the fact that $\max_{i=1,2} \max_{x\in\mathcal{D}_i} f(x) = \max_{x\in\mathcal{D}_1\cup\mathcal{D}_2} f(x)$; in (45) we relaxed the range over which λ_0 is optimized; and finally, (46) follows from the fact that the convex hull of a function can be written as the double conjugate of the original function [10, Thm. 12.2]. The concave hull is then the negative of the double conjugate of the negative of the original function.

In order to conclude the proof it remains to show that the range of ρ_0 over which the argument of (46) is optimized can be extended to $\rho_0 \in [0,1]$ without violating the inequality chain (42)-(46). We prove it by contradiction. To this end, let us consider a $\rho_0^\star$ that satisfies

$$\rho_0^\star = \arg\max_{\rho_0\in[0,1]} \big\{ \bar{E}_0(\rho_0, P_{Y|X}) - tE_{\mathrm{s}}(\rho_0, P_V) \big\}, \tag{47}$$

and assume that $\rho_0^\star < \hat{\rho}$.

From the definition of the concave hull, it follows that the smallest value of $\rho^\star$ that satisfies $\rho^\star = \arg\max_{\rho\in[0,1]} \big\{ \bar{E}_0(\rho, P_{Y|X}) - \rho R_{\mathcal{V}} \big\}$ equals $\hat{\rho}$. Then we have that

$$\bar{E}_0(\rho_0^\star, P_{Y|X}) - tE_\mathrm{s}(\rho_0^\star, P_V)$$
$$= \bar{E}_0(\rho_0^\star, P_{Y|X}) - \rho_0^\star R_\mathcal{V} + \rho_0^\star R_\mathcal{V} - tE_\mathrm{s}(\rho_0^\star, P_V) \quad (48)$$
$$< \bar{E}_0(\hat{\rho}, P_{Y|X}) - \hat{\rho} R_\mathcal{V} + \rho_0^\star R_\mathcal{V} - tE_\mathrm{s}(\rho_0^\star, P_V) \quad (49)$$
$$\leq \bar{E}_0(\hat{\rho}, P_{Y|X}) + R_\mathcal{V}(\rho_0^\star - \hat{\rho}) - t\left(E_\mathrm{s}(\hat{\rho}, P_V) + \frac{\partial E_\mathrm{s}(\rho, P_V)}{\partial \rho}(\rho_0^\star - \hat{\rho})\right) \quad (50)$$
$$\leq \bar{E}_0(\hat{\rho}, P_{Y|X}) - tE_\mathrm{s}(\hat{\rho}, P_V) \quad (51)$$

where in (49) we used the definition of $\hat{\rho}$; (50) follows from the convexity of $E_\mathrm{s}(\rho, P_V)$; and in (51) we used that

$$\frac{\partial E_\mathrm{s}(\rho, P_V)}{\partial \rho} \leq \lim_{\rho' \to \infty} \frac{\partial E_\mathrm{s}(\rho, P_V)}{\partial \rho}\bigg|_{\rho=\rho'} = \frac{R_\mathcal{V}}{t}. \quad (52)$$

From (48)-(51) it follows that by choosing $\rho_0 = \hat{\rho}$ we would achieve an objective strictly larger than by choosing $\rho_0 = \rho_0^\star$, hence contradicting the initial assumption.

It thus follows that

$$\max_{\rho_0 \in [\hat{\rho}, 1]} \left\{ \bar{E}_0(\rho_0, P_{Y|X}) - tE_\mathrm{s}(\rho_0, P_V) \right\} = \max_{\rho_0 \in [0, 1]} \left\{ \bar{E}_0(\rho_0, P_{Y|X}) - tE_\mathrm{s}(\rho_0, P_V) \right\}. \quad (53)$$

Since (53) is equal to (13), this concludes the proof.

References

[1] 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.

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

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

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

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

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

[8] 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.

[9] 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.

[10] R. T. Rockafellar, *Convex Analysis*. Princeton, US: Princeton University Press, 1970.

[11] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martinez, "Achieving Csiszár's exponent for joint source-channel coding with product distributions," in *2012 IEEE Int. Symp. on Inf. Theory*, Boston, USA, July 2012, submitted.