A Derivation of the Source-Channel Error Exponent using Non-identical Product Distributions

Adrià Tauste Campo, *Member, IEEE,* Gonzalo Vazquez-Vilar, *Member, IEEE,* Albert Guillén i Fàbregas, *Senior Member, IEEE,* Tobias Koch, *Member, IEEE,* and Alfonso Martinez, *Senior Member, IEEE*

***Abstract*—This paper studies the random-coding exponent of joint source-channel coding for a scheme where source messages are assigned to disjoint subsets (referred to as classes), and codewords are independently generated according to a distribution that depends on the class index of the source message. For discrete memoryless systems, two optimally chosen classes and product distributions are found to be sufficient to attain the sphere-packing exponent in those cases where it is tight.**

***Index Terms*—Joint source-channel coding, reliability function, random coding, product distributions, sphere-packing bound.**

I. Introduction

Jointly designed source-channel codes may achieve a lower error probability than separate source-channel coding [1]. In fact, the error exponent of a joint design may be up to twice that of the concatenation of source and channel codes [2]. The best exponent in this setting is due to Csiszár [1], who used a construction where codewords are drawn at random from a set of sequences with a composition that depends on the source message. He also showed that the exponent coincides with an upper bound, the sphere-packing exponent, in a certain rate region. A few years earlier, Gallager [3, p. 534, Prob. 5.16] derived a random-coding exponent for an ensemble whose codewords are drawn according to a fixed product distribution, independent of the source message. This method yields a simple derivation of the channel coding exponent in discrete memoryless channels [3, Th. 5.6.2]. However, the straightforward application to source-channel coding gives a (generally) weaker achievable exponent than Csiszár's method, although this difference is typically small for the optimum choice of input distributions [2].

In this paper, we study a code ensemble for which codewords associated to different source messages are generated according to different product distributions. We derive a new random-coding bound on the error probability for this ensemble and show that its exponent attains the sphere-packing exponent in the cases where it is known to be tight. We find that it is sufficient to consider either one or two different distributions in the optimum ensemble.

The paper is structured as follows. In Section II we introduce the system model and several definitions used throughout the paper. Section III reviews related previous work on source-channel coding. Section IV, the main section of the paper, presents the new random-coding bound and its error exponent. Finally, we conclude in Section V with some final remarks. Proofs of the results can be found in the appendices.

II. System Model and Definitions

An encoder maps a source message $\boldsymbol{v}$ to a length-n codeword $\boldsymbol{x}(\boldsymbol{v})$, which is then transmitted over the channel and decoded as $\hat{\boldsymbol{v}}$ at the receiver upon observation of the output $\boldsymbol{y}$. The source is characterized by a distribution $P^k(\boldsymbol{v}) = \prod_{j=1}^k P(v_j)$, $\boldsymbol{v} = (v_1, \ldots, v_k) \in \mathcal{V}^k$, where $\mathcal{V}$ is a finite alphabet. Since P fully describes the source, we shall sometimes abuse notation and refer to P as the source. The channel law is given by a conditional probability distribution $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{j=1}^n W(y_j|x_j)$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabet, respectively. While $\mathcal{X}$ and $\mathcal{Y}$ are assumed discrete for ease of exposition, our achievability results extend in a natural way to continuous alphabets.

Based on the output $\boldsymbol{y}$, the decoder selects a source message $\hat{\boldsymbol{v}}$ according to the maximum a posteriori (MAP) criterion,

$$\hat{\boldsymbol{v}} = \arg\max_{\boldsymbol{v}} P^k(\boldsymbol{v}) W^n\big(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{v})\big). \tag{1}$$

Throughout the paper, we avoid explicitly writing the set in optimizations and summations if they are performed over the entire set. Also, where unambiguous, we shall write $\boldsymbol{x}$ instead of $\boldsymbol{x}(\boldsymbol{v})$. We study the average error probability ϵ, defined as

$$\epsilon \triangleq \Pr\{\hat{\boldsymbol{V}} \neq \boldsymbol{V}\}, \tag{2}$$

where capital letters are used to denote random variables. In addition to bounds on the average error probability ϵ for finite values of k and n, we are interested in its exponential decay.

A. Tauste Campo, G. Vazquez-Vilar and A. Martinez are with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain (emails: {atauste, gvazquez, alfonso.martinez}@ieee.org). A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain, and the Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, United Kingdom (email: guillen@ieee.org). T. Koch is with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911 Leganés, Spain (email: koch@tsc.uc3m.es).

This work has been funded in part by the European Research Council (ERC) under grant agreement 259663; by the European Union under the 7th Framework Programme, grants FP7-PEOPLE-2009-IEF no. 252663, FP7-PEOPLE-2011-CIG no. 303633, FP7-PEOPLE-2012-CIG no. 333680, FP7-PEOPLE-2013-IEF no. 329837; and by the Spanish Ministry of Economy and Competitiveness under grants CSD2008-00010, TEC2009-14504-C02-01, TEC2012-38800-C03-01, TEC2012-38800-C03-03 and RYC-2011-08150. A. Tauste Campo acknowledges funding from an EPSRC (Engineering and Physical Sciences Research Council, UK) Doctoral Prize Award.

This work was presented in part at the 46th Conference on Information Sciences and Systems, Princeton, NJ, March 21–23, 2012 and at the IEEE Symposium on Information Theory, Cambridge, MA, July 1–6, 2012.

Consider a sequence of sources with length $k = 1, 2, \dots$ and a corresponding sequence of codes of length $n = n_1, n_2, \dots$ Assume that the ratio $\frac{k}{n}$ converges to some quantity

$$t \triangleq \lim_{k\to\infty} \frac{k}{n}, \tag{3}$$

referred to as *transmission rate*. An exponent $E(P, W, t) > 0$ is to said to be achievable if there exists a sequence of codes whose error probabilities ϵ satisfy

$$\epsilon \le e^{-nE(P,W,t)+o(n)}, \tag{4}$$

where $o(n)$ is a sequence such that $\lim_{n\to\infty} \frac{o(n)}{n} = 0$. The reliability function $E_{\mathrm{J}}(P, W, t)$ is defined as the supremum of all achievable error exponents; we sometimes shorten it to E_{J}.

We denote Gallager's source and channel functions as

$$E_{\mathrm{s}}(\rho, P) \triangleq \log \left(\sum_v P(v)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{5}$$

$$E_0(\rho, W, Q) \triangleq -\log \sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{6}$$

respectively. Here, $\log(\cdot)$ denotes the natural logarithm.

We are interested in the error exponent maximized over a subset of probability distributions on $\mathcal{X}$. Let $\mathcal{Q}$ be a non-empty proper subset of probability distributions on $\mathcal{X}$. With some abuse of notation we define

$$E_0(\rho, W, \mathcal{Q}) \triangleq \max_{Q\in\mathcal{Q}} E_0(\rho, W, Q). \tag{7}$$

When the optimization is done over the set of all probability distributions on $\mathcal{X}$ we simply write $E_0(\rho, W) \triangleq \max_Q E_0(\rho, W, Q)$.

We denote by $\bar{E}_0(\rho, W, \mathcal{Q})$ the concave hull of $E_0(\rho, W, \mathcal{Q})$, defined pointwise as the supremum over convex combinations of any two values of the function $E_0(\rho, W, \mathcal{Q})$ [4, p. 36], i.e.

$$\begin{aligned} &\bar{E}_0(\rho, W, \mathcal{Q}) \\ &\triangleq \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda E_0(\rho_1, W, \mathcal{Q}) + (1-\lambda) E_0(\rho_2, W, \mathcal{Q}) \Big\}. \end{aligned} \tag{8}$$

Similarly, $\bar{E}_0(\rho, W)$ denotes the concave hull of $E_0(\rho, W)$.

III. Gallager's and Csiszár's Exponents

For source coding (i.e., when W is the channel law of a noiseless channel), the reliability function of a source P at rate R, denoted by $e(R, P)$, is given by [5]

$$e(R, P) = \sup_{\rho\ge 0} \big\{ \rho R - E_{\mathrm{s}}(\rho, P) \big\}. \tag{9}$$

For channel coding (i.e., when P is the uniform distribution), the reliability function of a channel W at rate R, denoted by $E(R, W)$, is bounded as [3]

$$E_{\mathrm{r}}(R, W) \le E(R, W) \le E_{\mathrm{sp}}(R, W), \tag{10}$$

where $E_{\mathrm{r}}(R, W)$ denotes the random-coding exponent and $E_{\mathrm{sp}}(R, W)$ the sphere-packing exponent, respectively given by

$$E_{\mathrm{r}}(R, W) = \max_{\rho\in[0,1]} \Big\{ E_0(\rho, W) - \rho R \Big\}, \tag{11}$$

$$E_{\mathrm{sp}}(R, W) = \sup_{\rho\ge 0} \Big\{ E_0(\rho, W) - \rho R \Big\}. \tag{12}$$

We define the critical rate of the channel R_{cr} as the smallest value of R such that $E_{\mathrm{r}}(R, W)$ and $E_{\mathrm{sp}}(R, W)$ coincide.

For source-channel coding, Gallager used a random-coding argument to derive an upper bound on the average error probability by drawing the codewords independently of the source messages according to a given product distribution $Q^n(\boldsymbol{x}) = \prod_{j=1}^n Q(x_j)$. He found the achievable exponent [3, p. 534, Prob. 5.16]

$$\max_{\rho\in[0,1]} \Big\{ E_0(\rho, W, Q) - t E_{\mathrm{s}}(\rho, P) \Big\}, \tag{13}$$

which becomes, upon maximizing over Q,

$$E_{\mathrm{J}}^{\mathrm{G}}(P, W, t) \triangleq \max_{\rho\in[0,1]} \Big\{ E_0(\rho, W) - t E_{\mathrm{s}}(\rho, P) \Big\}. \tag{14}$$

Csiszár refined this result using the method of types [1]. By using a partition of the message set into source-type classes and considering fixed-composition codes that map messages within a source type onto sequences within a channel-input type, he found an achievable exponent

$$E_{\mathrm{J}}^{\mathrm{Cs}}(P, W, t) \triangleq \min_{tH(V)\le R\le R_{\mathcal{V}}} \left\{ te\left(\frac{R}{t}, P\right) + E_{\mathrm{r}}(R, W) \right\}, \tag{15}$$

where $R_{\mathcal{V}} \triangleq t \log|\mathcal{V}|$. A convenient alternative representation of $E_{\mathrm{J}}^{\mathrm{Cs}}$ was obtained by Zhong *et al.* [2] via Fenchel's duality theorem [4, Th. 31.1]:

$$E_{\mathrm{J}}^{\mathrm{Cs}}(P, W, t) = \max_{\rho\in[0,1]} \big\{ \bar{E}_0(\rho, W) - t E_{\mathrm{s}}(\rho, P) \big\}. \tag{16}$$

Since $\bar{E}_0(\rho, W) \ge E_0(\rho, W)$, it follows from (16) and (14) that $E_{\mathrm{J}}^{\mathrm{Cs}} \ge E_{\mathrm{J}}^{\mathrm{G}}$. Nonetheless, the finite-length bound implied by the exponent $E_{\mathrm{J}}^{\mathrm{Cs}}$ in [1] might be worse than the one in [3, p. 534, Prob. 5.16] due to the worse subexponential terms, which may dominate for finite values of k and n.

To validate the optimality of $E_{\mathrm{J}}^{\mathrm{Cs}}$, Csiszár derived a sphere-packing bound on the exponent [1, Lemma 2],

$$E_{\mathrm{J}}^{\mathrm{sp}}(P, W, t) \triangleq \min_{tH(V)\le R\le R_{\mathcal{V}}} \left\{ te\left(\frac{R}{t}, P\right) + E_{\mathrm{sp}}(R, W) \right\}. \tag{17}$$

When the minimum on the right-hand side (RHS) of (17) is attained for a value of R such that $E_{\mathrm{sp}}(R, W) = E_{\mathrm{r}}(R, W)$, the upper bound (17) coincides with the lower bound (15) and, hence, $E_{\mathrm{J}}^{\mathrm{Cs}} = E_{\mathrm{J}}$. This is the case for values of R above the critical rate of the channel R_{cr} [1].

IV. An Achievable Exponent

In this section, we analyze the error probability of random-coding ensembles where the codeword distribution depends on the source message. We find that ensembles generated with a pair of product distributions $\big\{Q_1^n, Q_2^n\big\}$ may attain a better error exponent than Gallager's exponent (13) for Q being equal to either Q_1 or Q_2. Moreover, optimizing over pairs of distributions this ensemble attains the exponent $E_{\mathrm{J}}^{\mathrm{sp}}$ in those cases where it is tight.

A. Main Results

We define a partition of the source-message set $\mathcal{V}^k$ into N_k disjoint subsets $\mathcal{A}_k^{(i)}$, $i = 1, \ldots, N_k$, such that $\bigcup_{i=1}^{N_k} \mathcal{A}_k^{(i)} = \mathcal{V}^k$. We refer to these subsets as *classes*. For each source message $\boldsymbol{v}$ in the set $\mathcal{A}_k^{(i)}$, we randomly and independently generate codewords $\boldsymbol{x}(\boldsymbol{v}) \in \mathcal{X}^n$ according to a channel-input product distribution $Q_i^n(\boldsymbol{x}) = \prod_{j=1}^n Q_i(x_j)$. This definition is a generalization of Csiszár's partition in [1] where each subset corresponds to a source-type class. Since the number of source-type classes is a polynomial function of k [6], it follows that the number of classes N_k considered in [1] is also polynomial in k.

The next result extends [3, Th. 5.6.2] to codebook ensembles where codewords are independently but not necessarily identically distributed.

Theorem 1: For a given partition $\mathcal{A}_k^{(i)}$, $i = 1, \ldots, N_k$, and associated distributions Q_i, $i = 1, \ldots, N_k$, there exists a codebook satisfying

$$\epsilon \leq h(k) \sum_{i=1}^{N_k} e^{-\max_{\rho_i \in [0,1]} \left\{ E_0\left(\rho_i, W^n, Q_i^n\right) - E_{\mathrm{s}}^{(i)}(\rho_i, P^k) \right\}}, \quad (18)$$

where $h(k) \triangleq \frac{3N_k - 1}{2}$ and

$$E_{\mathrm{s}}^{(i)}(\rho, P^k) \triangleq \log \left(\sum_{\boldsymbol{v} \in \mathcal{A}_k^{(i)}} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (19)$$

Proof: See Appendix I. ∎

Theorem 1 holds for general (not necessarily memoryless) discrete sources and channels, and for Q_i^n, $i = 1, \ldots, N_k$, being non-product distributions (including cost-constrained and fixed composition ensembles). Furthermore, it naturally extends to continuous channels by following the same arguments as those extending Gallager's analysis of the exponent of channel coding.

It was demonstrated in [7] that an application of Theorem 1 to a partition where classes are identified with source-type classes attains $E_{\mathrm{J}}^{\mathrm{Cs}}$. However, compared to the bound used to derive Csiszár's exponent in [1], Theorem 1 provides a tighter bound on the average error probability for finite values of k and n [8]. Furthermore, Theorem 1 can be generalized to derive Csiszár's lower bound on the error exponent for lossy source-channel coding [9, Sec. IV].

For a single class with associated distribution Q, Theorem 1 simply recovers the exponent in (13). The following theorem shows that the exponent may be improved by considering a partition with two classes.

Theorem 2: For a pair of distributions $\{Q, Q'\}$, there exists a partition of the source message set into two classes such that the following exponent is achievable

$$\max_{\rho \in [0,1]} \left\{ \bar{E}_0\big(\rho, W, \{Q, Q'\}\big) - tE_{\mathrm{s}}(\rho, P) \right\}. \quad (20)$$

Moreover, a partition achieving this exponent with associated distributions $Q_i \in \{Q, Q'\}$, $i = 1, 2$, is given by

$$\mathcal{A}_k^{(1)}(\gamma) \triangleq \left\{ \boldsymbol{v} : \ P^k(\boldsymbol{v}) \leq \gamma^k \right\} \quad (21)$$

$$\mathcal{A}_k^{(2)}(\gamma) \triangleq \left\{ \boldsymbol{v} : \ P^k(\boldsymbol{v}) > \gamma^k \right\}, \quad (22)$$

for some $\gamma \in [0, 1]$.

Proof: See Appendix II. ∎

In Theorem 2 we considered a particular pair of distributions $\{Q, Q'\}$. A direct application of Carathéodory's theorem [4, Cor. 17.1.5] shows that any point belonging to the graph of $\bar{E}_0(\rho, W)$ can be expressed as a convex combination of two points belonging to the graph of $E_0(\rho, W)$. Consequently, there exists a pair of distributions Q, Q' such that these two points also belong to the graph of $E_0(\rho, W, \{Q, Q'\})$. By optimizing the exponent (20) over all possible pairs of distributions $\{Q, Q'\}$, the following result follows.

Corollary 1: There exists a partition of the source message set into two classes assigned to a pair of distributions such that $E_{\mathrm{J}}^{\mathrm{Cs}}$ in (16) is achievable.

In contrast to Csiszár's original analysis [1], where the number of classes used to attain the best exponent was polynomial in k, Corollary 1 shows that a two-class construction suffices to attain $E_{\mathrm{J}}^{\mathrm{Cs}}$ when the partition and associated distributions are appropriately chosen.

B. Ensemble Tightness

Since Section IV-A only considers achievability results, one may ask whether the weakness of Gallager's exponent is due to the bounding technique or to the construction itself. A partial answer to this question can be given by studying the exact random-coding exponent, namely the exponential decay of the error probability averaged over the ensemble, which we denote by $\bar{\epsilon}$.

Theorem 3: For any non-empty set $\mathcal{Q}$ of probability distributions on $\mathcal{X}$, consider a codebook ensemble for which the codewords associated to source messages with type class $\mathcal{T}_i$ are generated according to a distribution $Q_i^n(\boldsymbol{x}) = \prod_{j=1}^n Q_i(x_j)$ with $Q_i \in \mathcal{Q}$, $i = 1, \ldots, N_k'$, where N_k' is the number of source type classes. The random-coding exponent of this ensemble is upper-bounded as

$$\limsup_{n \to \infty} -\frac{\log \bar{\epsilon}}{n} \leq \max_{\rho \in [0,1]} \left\{ \bar{E}_0(\rho, W, \mathcal{Q}) - tE_{\mathrm{s}}(\rho, P) \right\}. \quad (23)$$

Proof: See Appendix III. ∎

When $\mathcal{Q}$ contains only one distribution, the concavity of $E_0(\rho, W, Q)$ as a function of ρ shows that the RHS of (23) matches (13). In other words, if the codebook is drawn according to only one distribution Q, then $E_{\mathrm{J}}^{\mathrm{G}}$ in (14) is ensemble tight.

The ensemble considered in Theorem 2 is a particular case of that of Theorem 3 with $|\mathcal{Q}| = 2$. Since the upper bound (23) and the lower bound (20) coincide for $\mathcal{Q} = \{Q, Q'\}$, the error exponent (20) is also ensemble tight. Furthermore, for any set with cardinality $\mathcal{Q}$ with $|\mathcal{Q}| > 2$, we can always choose two distributions Q and Q' belonging to $\mathcal{Q}$ such that (20) equals the RHS of (23) [4, Cor. 17.1.5]. Therefore, the random-coding

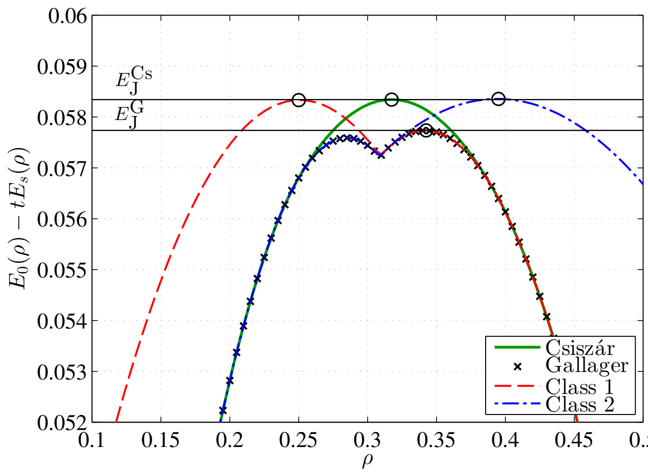

Figure 1. Error exponent bounds. Csiszár's and Gallager's curves correspond to $\bar{E}_0(\rho, W) - tE_{\mathrm{s}}(\rho, P)$ and $E_0(\rho, W) - tE_{\mathrm{s}}(\rho, P)$, respectively. Class i curve correspond to $E_0(\rho, W) - \lim_{n\to\infty} \frac{1}{n} E_{\mathrm{s}}^{(i)}(\rho, P^k)$, for $i = 1, 2$.

exponent of an ensemble with an arbitrary number of classes can be attained by a two-class partition, as in Theorem 2.

Finally, it can be shown that Theorem 3 holds for finer partitions of the source message set, not necessarily corresponding to source type classes. Since the RHS of (23) coincides with $E_{\mathrm{J}}^{\mathrm{Cs}}$ when $\mathcal{Q}$ is the set of all probability distributions on $\mathcal{X}$, we conclude that the ensembles studied in this work cannot improve Csiszár's random-coding exponent, even when the latter does not coincide with the sphere-packing exponent.

C. Example: a 6-input 4-output channel

We present an example in which the two-class partition (with their corresponding product distributions) attains the sphere-packing exponent while Gallager's one-class assignment does not. Consider the source-channel pair composed by a binary memoryless source (BMS) and a non-symmetric memoryless channel with $|\mathcal{X}| = 6$, $|\mathcal{Y}| = 4$ and transition-probability matrix

$$W = \begin{pmatrix} 1-3\xi_1 & \xi_1 & \xi_1 & \xi_1 \\ \xi_1 & 1-3\xi_1 & \xi_1 & \xi_1 \\ \xi_1 & \xi_1 & 1-3\xi_1 & \xi_1 \\ \xi_1 & \xi_1 & \xi_1 & 1-3\xi_1 \\ \frac{1}{2}-\xi_2 & \frac{1}{2}-\xi_2 & \xi_2 & \xi_2 \\ \xi_2 & \xi_2 & \frac{1}{2}-\xi_2 & \frac{1}{2}-\xi_2 \end{pmatrix}. \tag{24}$$

This channel is similar to the channel given in [3, Fig. 5.6.5] and studied in [2] for source-channel coding. It is composed of two quaternary-output sub-channels: one of them is a quaternary-input symmetric channel with parameter ξ_1, and the second one is a binary-input channel with parameter ξ_2. We set $\xi_1 = 0.065$, $\xi_2 = 0.01$, $t = 2$ (source symbols/channel use) and $P(1) = 0.028$. It follows that the source entropy is $H(V) = 0.1843$ bits/source symbol, the channel capacity is $C = 0.9791$ bits/channel use and the critical rate is $R_{\mathrm{cr}} = 0.4564$ bits/channel use. Let $R^\star$ denote the value of R minimizing (15). In this example we have $R^\star = 0.6827 > R_{\mathrm{cr}}$ and $E_{\mathrm{J}}^{\mathrm{Cs}}$ is tight.

In Fig. 1 we plot the objective functions of Gallager's exponent in (14) and Csiszár's exponent in (16) as functions of ρ, respectively. For reference purposes, we also show the values of $E_{\mathrm{J}}^{\mathrm{G}}$ and $E_{\mathrm{J}}^{\mathrm{Cs}}$ with horizontal solid lines. The distribution Q maximizing $E_0(\rho, W, Q)$ changes from $\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ 0\ 0\right)$ for $\rho \le 0.31$ to $\left(0\ 0\ 0\ 0\ \frac{1}{2}\ \frac{1}{2}\right)$ for $\rho > 0.31$. As a result, $E_0(\rho, W)$ is not concave in $\rho \in [0, 1]$. The figure shows how the non-concavity of Gallager's function around the optimal ρ of Csiszár's function translates into a loss in exponent.

Fig. 1 also shows the bracketed terms in the RHS of (18) as a function of ρ_i for the two-class partition of Theorem 2. The overall error exponent of the two-class construction is obtained by first individually maximizing the exponent of each of the curves over ρ_i, and by then choosing the minimum of the two individual maxima. In this example, the exponent of both classes coincides with $E_{\mathrm{J}}^{\mathrm{Cs}}$. The overall exponent is thus given by $E_{\mathrm{J}}^{\mathrm{Cs}}$, which is in agreement with Theorem 2.

V. Conclusions

We have studied the error probability of random-coding ensembles where different codeword distributions are assigned to different subsets of source messages. We have showed that the random-coding exponent of ensembles generated with a single distribution does not attain Csiszár's exponent in general. In contrast, ensembles with at most two appropriately chosen subsets and distributions suffice to attain the sphere-packing exponent in those cases where it is tight. One of the strengths of our achievability result is that, unlike Csiszár's approach, it does not rely on the method of types. This leads to tighter bounds on the average error probability for finite block lengths and may simplify the task of generalizing our bound to source-channel systems with non-discrete alphabets and memory.

Appendix I
Proof of Theorem 1

Generalizing the proof of the random-coding union bound for channel coding [10, Th. 16] (with earlier precedents in [3, pp. 136-137]) to joint source-channel coding we obtain

$$\bar{\epsilon} \le \sum_{\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}} P^k(\boldsymbol{v}) Q_{\boldsymbol{v}}^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \times \min\left\{1, \sum_{\bar{\boldsymbol{v}}\neq\boldsymbol{v}} \Pr\left\{P^k(\bar{\boldsymbol{v}}) W^n(\boldsymbol{y}|\bar{\boldsymbol{X}}) \ge P^k(\boldsymbol{v}) W^n(\boldsymbol{y}|\boldsymbol{x})\right\}\right\}, \tag{25}$$

where $Q_{\boldsymbol{v}}^n$ denotes the channel-input distribution corresponding to the source message $\boldsymbol{v}$ and $\bar{\boldsymbol{X}}$ is distributed according to $Q_{\bar{\boldsymbol{v}}}^n$.

When codewords are generated according to distributions that depend on the class index of the source, (25) yields

$$\epsilon \leq \sum_{i=1}^{N_k} \sum_{\boldsymbol{v}\in\mathcal{A}_k^{(i)}} P^k(\boldsymbol{v}) \sum_{\boldsymbol{x},\boldsymbol{y}} Q_i^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \times \min\left\{1, \sum_{j=1}^{N_k} \sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_k^{(j)}} \sum_{\substack{\bar{\boldsymbol{x}}: P^k(\bar{\boldsymbol{v}})W^n(\boldsymbol{y}|\bar{\boldsymbol{x}}) \\ \geq P^k(\boldsymbol{v})W^n(\boldsymbol{y}|\boldsymbol{x})}} Q_j^n(\bar{\boldsymbol{x}})\right\}. \quad (26)$$

We next use Markov's inequality for $s_j \geq 0$, $j = 1, \ldots, N_k$, to obtain [3]

$$\sum_{\substack{\bar{\boldsymbol{x}}: P^k(\bar{\boldsymbol{v}})W^n(\boldsymbol{y}|\bar{\boldsymbol{x}}) \\ \geq P^k(\boldsymbol{v})W^n(\boldsymbol{y}|\boldsymbol{x})}} Q_j^n(\bar{\boldsymbol{x}}) \leq \sum_{\bar{\boldsymbol{x}}} Q_j^n(\bar{\boldsymbol{x}}) \left(\frac{P^k(\bar{\boldsymbol{v}})W^n(\boldsymbol{y}|\bar{\boldsymbol{x}})}{P^k(\boldsymbol{v})W^n(\boldsymbol{y}|\boldsymbol{x})}\right)^{s_j}. \quad (27)$$

Using (27) and the inequality $\min\{1, A+B\} \leq A^\rho + B^{\rho'}$, $A, B \geq 0$, $\rho, \rho' \in [0,1]$ [3], (26) is upper-bounded by

$$\epsilon \leq \sum_{i,j=1}^{N_k} \sum_{\boldsymbol{v}\in\mathcal{A}_k^{(i)}} P^k(\boldsymbol{v}) \sum_{\boldsymbol{x},\boldsymbol{y}} Q_i^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \times \left(\sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_k^{(j)}} \sum_{\bar{\boldsymbol{x}}} Q_j^n(\bar{\boldsymbol{x}}) \left(\frac{P^k(\bar{\boldsymbol{v}})W^n(\boldsymbol{y}|\bar{\boldsymbol{x}})}{P^k(\boldsymbol{v})W^n(\boldsymbol{y}|\boldsymbol{x})}\right)^{s_j}\right)^{\rho_{ij}}, \quad (28)$$

where $\rho_{ij} \in [0,1]$ and $s_j \geq 0$, $i, j = 1, \ldots, N_k$.

For $s_i, s_j \in \left[\frac{1}{2}, 1\right]$ and $\rho_{ij} = \frac{1-s_i}{s_j}$, (28) yields

$$\epsilon \leq \sum_{i,j=1}^{N_k} \sum_{\boldsymbol{y}} G_i(\boldsymbol{y})^{s_i} G_j(\boldsymbol{y})^{1-s_i} \quad (29)$$

where

$$G_i(\boldsymbol{y}) \triangleq \left(\sum_{\boldsymbol{v}\in\mathcal{A}_k^{(i)}} P^k(\boldsymbol{v})^{s_i}\right)^{\frac{1}{s_i}} \left(\sum_{\boldsymbol{x}} Q_i^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x})^{s_i}\right)^{\frac{1}{s_i}}. \quad (30)$$

This choice of ρ_{ij} allows us to decompose the probability of the "inter-class" error event between classes i and j as the product of two terms corresponding to "intra-class" error events. The RHS of (30) is further upper-bounded by

$$\begin{aligned}
\epsilon &\leq \sum_{i,j=1}^{N_k} \left(\sum_{\boldsymbol{y}} G_i(\boldsymbol{y})\right)^{s_i} \left(\sum_{\boldsymbol{y}} G_j(\boldsymbol{y})\right)^{1-s_i} && (31)\\
&\leq \sum_{i,j=1}^{N_k} \left(s_i \left(\sum_{\boldsymbol{y}} G_i(\boldsymbol{y})\right) + (1-s_i)\left(\sum_{\boldsymbol{y}} G_j(\boldsymbol{y})\right)\right) && (32)\\
&\leq \sum_{i=1}^{N_k} \sum_{\boldsymbol{y}} G_i(\boldsymbol{y}) + \sum_{\substack{i,j=1\\ i\neq j}}^{N_k} \left(\sum_{\boldsymbol{y}} G_i(\boldsymbol{y}) + \frac{1}{2}\sum_{\boldsymbol{y}} G_j(\boldsymbol{y})\right) && (33)\\
&= \frac{3N_k - 1}{2} \sum_{i=1}^{N_k} \sum_{\boldsymbol{y}} G_i(\boldsymbol{y}), && (34)
\end{aligned}$$

where in (31) we applied Hölder's inequality $\|fg\|_1 \leq \|f\|_p \|g\|_q$ with $p = \frac{1}{s_i}$ and $q = \frac{1}{1-s_i}$ (with $\|\cdot\|_p$ denoting the p-norm); (32) follows from the relation between arithmetic and geometric means; and (33) follows because $\frac{1}{2} \leq s_i \leq 1$. By identifying

$$\sum_{\boldsymbol{y}} G_i(\boldsymbol{y}) = e^{-E_0(\rho_i, W^n, Q_i^n) + E_{\mathrm{s}}^{(i)}(\rho_i, P^k)}, \quad (35)$$

with $\rho_i = \frac{1-s_i}{s_i}$, and optimizing over $\rho_i \in [0,1]$, $i = 1, \ldots, N_k$, (31)-(34) yields

$$\epsilon \leq \frac{3N_k - 1}{2} \sum_{i=1}^{N_k} e^{-\max_{\rho_i\in[0,1]} \left\{E_0\left(\rho_i, W^n, Q_i^n\right) - E_{\mathrm{s}}^{(i)}(\rho_i, P^k)\right\}}. \quad (36)$$

This concludes the proof.

Appendix II
Proof of Theorem 2

The proof of the Theorem 2 is based on the next preliminary result.

Lemma 1: For any $\rho_0 \in [0,1]$ and $\gamma' \geq 0$, the partition (21)-(22) with $\gamma = \min\{1, \gamma'\}$ satisfies

$$\begin{aligned}
\frac{1}{k} E_{\mathrm{s}}^{(1)}(\rho, P^k) &\leq E_{\mathrm{s}}(\rho, P)\mathbb{1}\{\rho > \rho_0\} + r(\rho, \rho_0, \gamma')\mathbb{1}\{\rho \leq \rho_0\}\\
&\triangleq \bar{E}_{\mathrm{s}}^{(1)}(\rho, \rho_0, \gamma'), && (37)\\
\frac{1}{k} E_{\mathrm{s}}^{(2)}(\rho, P^k) &\leq E_{\mathrm{s}}(\rho, P)\mathbb{1}\{\rho < \rho_0\} + r(\rho, \rho_0, \gamma')\mathbb{1}\{\rho \geq \rho_0\}\\
&\triangleq \bar{E}_{\mathrm{s}}^{(2)}(\rho, \rho_0, \gamma'), && (38)
\end{aligned}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function, and where

$$r(\rho, \rho_0, \gamma) \triangleq E_{\mathrm{s}}(\rho_0, P) + \frac{E_{\mathrm{s}}(\rho_0, P) - \log\gamma}{1+\rho_0}(\rho - \rho_0). \quad (39)$$

Proof: For the choice $\gamma = \min\{1, \gamma'\}$ it holds that

$$\mathbb{1}\left\{P^k(\boldsymbol{v}) \leq \gamma^k\right\} = \mathbb{1}\left\{P^k(\boldsymbol{v}) \leq (\gamma')^k\right\} \quad (40)$$

since $P^k(\boldsymbol{v}) \leq 1$ for all $\boldsymbol{v}$. Using (40), the monotonicity of $\log(\cdot)$, and the bound $\mathbb{1}\{a \leq b\} \leq a^{-s}b^s$ for $s \geq 0$, the function $\frac{1}{k}E_{\mathrm{s}}^{(1)}(\rho, P^k)$ can be upper-bounded as

$$\begin{aligned}
&\frac{1}{k}E_{\mathrm{s}}^{(1)}(\rho, P^k)\\
&\quad\leq \frac{1}{k}\log\left(\sum_{\boldsymbol{v}} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}} \mathbb{1}\left\{P^k(\boldsymbol{v}) \leq (\gamma')^k\right\}\right)^{1+\rho} && (41)\\
&\quad\leq \frac{1}{k}\log\left(\sum_{\boldsymbol{v}} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}} P^k(\boldsymbol{v})^{-s}(\gamma')^{ks}\right)^{1+\rho} && (42)\\
&\quad= \log\left(\sum_{v} P(v)^{\frac{1}{1+\rho}-s}(\gamma')^s\right)^{1+\rho}, && (43)
\end{aligned}$$

for any $s \geq 0$. Here we used that $P^k(\boldsymbol{v})$ is memoryless. We continue by choosing

$$s = \max\left(0, \frac{\rho_0 - \rho}{(1+\rho_0)(1+\rho)}\right). \quad (44)$$

For $\rho > \rho_0$, it follows that $s = 0$, and (43) gives (cf. (5))

$$\frac{1}{k}E_{\mathrm{s}}^{(1)}(\rho, P^k) \leq E_{\mathrm{s}}(\rho, P). \quad (45)$$

For $\rho \le \rho_0$, the choice (44) yields $s = \frac{\rho_0-\rho}{(1+\rho_0)(1+\rho)}$, which together with (43) yields

$$\begin{aligned}
&\frac{1}{k}E_{\mathrm{s}}^{(1)}(\rho,P^k) \\
&\le (1+\rho)\log\left(\sum_v P(v)^{\frac{1}{1+\rho_0}}\right) - \frac{\rho-\rho_0}{1+\rho_0}\log\gamma' && (46)\\
&= (1+\rho_0)\log\left(\sum_v P(v)^{\frac{1}{1+\rho_0}}\right) \\
&\quad + (\rho-\rho_0)\log\left(\sum_v P(v)^{\frac{1}{1+\rho_0}}\right) - \frac{\rho-\rho_0}{1+\rho_0}\log\gamma' && (47)\\
&= E_{\mathrm{s}}(\rho_0,P) + \frac{E_{\mathrm{s}}(\rho_0,P)-\log\gamma'}{1+\rho_0}\left(\rho-\rho_0\right), && (48)
\end{aligned}$$

where in (47) we added and subtracted the term $\rho_0\log\left(\sum_v P(v)^{\frac{1}{1+\rho_0}}\right)$; and (48) follows from the definition of $E_{\mathrm{s}}(\rho,P)$ in (5). The inequality (37) follows by combining (45) and (46)-(48) for $\rho > \rho_0$ and $\rho \le \rho_0$, respectively.

In an analogous way, the inequality (38) can be proved using that $\mathbb{1}\{P^k(\boldsymbol{v}) > \gamma^k\} = \mathbb{1}\{P^k(\boldsymbol{v}) > (\gamma')^k\}$ and $\mathbb{1}\{a > b\} \le a^s b^{-s}$ with $s \ge 0$. ■

By applying Theorem 1 to the two-class partition (21)-(22) with associated product distributions Q_i^n, $i = 1,2$, for the optimal threshold γ we obtain

$$\begin{aligned}
E_{\mathrm{J}}^{\mathrm{B}} &\triangleq \max_{\gamma\in[0,1]}\left\{\liminf_{n\to\infty}\left\{-\frac{1}{n}\log\left(h(k)\right.\right.\right. \\
&\qquad \left.\left.\left.\times\sum_{i=1,2} e^{-\max_{\rho_i\in[0,1]}\left\{nE_0(\rho_i,W,Q_i)-E_{\mathrm{s}}^{(i)}(\rho_i)\right\}}\right)\right\}\right\} && (49)\\
&= \max_{\gamma\in[0,1]}\left\{\liminf_{n\to\infty}\min_{i=1,2}\left\{\max_{\rho_i\in[0,1]}\left\{E_0(\rho_i,W,Q_i)\right.\right.\right. \\
&\qquad \left.\left.\left. - \frac{1}{n}E_{\mathrm{s}}^{(i)}(\rho_i)\right\}\right\}\right\} && (50)\\
&\ge \max_{\gamma'\ge 0}\max_{\rho_0,\rho_1,\rho_2\in[0,1]}\min_{i=1,2}\Big\{E_0(\rho_i,W,Q_i) \\
&\qquad - t\bar{E}_{\mathrm{s}}^{(i)}(\rho_i,\rho_0,\gamma')\Big\} && (51)\\
&\ge \max_{\substack{\rho_0,\rho_1,\rho_2\in[0,1]:\\ \rho_1\le\rho_0\le\rho_2}}\max_{\gamma'\ge 0}\min_{i=1,2}\Big\{E_0(\rho_i,W,Q_i) \\
&\qquad - t\bar{E}_{\mathrm{s}}^{(i)}(\rho_i,\rho_0,\gamma')\Big\}, && (52)
\end{aligned}$$

where (50) follows by noting that $h(k)$ is subexponential in k; in (51) we have applied Lemma 1 with $\rho_0 \in [0,1]$ and $\gamma' \ge 0$ and have used that $\liminf_{n\to\infty}\max_x\{f_n(x)\} \ge \max_x\{\lim_{n\to\infty} f_n(x)\}$; and in (52) we have restricted the range over which we maximize ρ_i, $i = 0,1,2$, and interchanged the maximization order.

By substituting (37)-(38) with $0 \le \rho_1 \le \rho_0 \le \rho_2 \le 1$, the minimization in (52) becomes

$$\min_{i=1,2}\bigg\{E_0(\rho_i,W,Q_i) + t\frac{E_{\mathrm{s}}(\rho_0,P)-\log\gamma'}{1+\rho_0}(\rho_0-\rho_i) - tE_{\mathrm{s}}(\rho_0,P)\bigg\}. \quad (53)$$

We define $\gamma_0 \ge 0$ as the value satisfying

$$t\frac{E_{\mathrm{s}}(\rho_0,P)-\log\gamma_0}{1+\rho_0} = \frac{E_0(\rho_2,W,Q_2)-E_0(\rho_1,W,Q_1)}{\rho_2-\rho_1}. \quad (54)$$

The existence of such a γ_0 follows from the continuity of the logarithm function. Choosing $\gamma' = \gamma_0$ equalizes the two terms in the minimization in (53), thus maximizing the lower bound (52). As a result, substituting (53) into (52) we obtain

$$E_{\mathrm{J}}^{\mathrm{B}} \ge \max_{\rho_0\in[0,1]}\left\{\max_{\substack{\rho_1,\rho_2\in[0,1]:\\ \rho_1\le\rho_0\le\rho_2}}\left\{\frac{\rho_2-\rho_0}{\rho_2-\rho_1}E_0(\rho_1,W,Q_1) + \frac{\rho_0-\rho_1}{\rho_2-\rho_1}E_0(\rho_2,W,Q_2)\right\} - tE_{\mathrm{s}}(\rho_0,P)\right\}. \quad (55)$$

We now optimize the RHS of (55) over the assignments $(Q_1,Q_2) = (Q,Q')$ and $(Q_1,Q_2) = (Q',Q)$. By denoting by ρ (resp. ρ') the variable ρ_i, $i = 1,2$, associated to Q (resp. Q') and defining λ such that $\lambda\rho + (1-\lambda)\rho' = \rho_0$, the optimal assignment leads to

$$E_{\mathrm{J}}^{\mathrm{B}} \ge \max_{\rho_0\in[0,1]}\left\{\max_{\substack{\rho,\rho',\lambda\in[0,1]:\\ \lambda\rho+(1-\lambda)\rho'=\rho_0}}\Big\{\lambda E_0(\rho,W,Q) + (1-\lambda)E_0(\rho',W,Q')\Big\} - tE_{\mathrm{s}}(\rho_0,P)\right\}. \quad (56)$$

Theorem 2 follows from (56) by noting that [4, Th. 5.6]

$$\bar{E}_0(\rho_0,W,\{Q,Q'\}) = \max_{\substack{\rho,\rho',\lambda\in[0,1]:\\ \lambda\rho+(1-\lambda)\rho'=\rho_0}}\Big\{\lambda E_0(\rho,W,Q) + (1-\lambda)E_0(\rho',W,Q')\Big\}. \quad (57)$$

By inspection of the proof we conclude that the threshold γ in (21)-(22) is given by $\gamma = \min(1,\gamma_0^\star)$ where $\gamma_0^\star$ is computed from (54) for the values of $\rho_0^\star$, $\rho_1^\star$, $\rho_2^\star$ optimizing (55) and $(Q_1^\star,Q_2^\star)$ leading to (56).

Appendix III
Proof of Theorem 3

Before proving the result, we introduce some definitions that will ease the exposition. Let $\mathcal{A}$ be an arbitrary non-empty discrete set. We denote the set of all probability distributions on $\mathcal{A}$ by $\mathcal{D}(\mathcal{A})$ and the set of types in $\mathcal{A}^n$ by $\mathcal{D}_n(\mathcal{A})$. We further denote by $\mathcal{T}(\mathsf{P}_{XY})$ the type-class of sequences $(\boldsymbol{x},\boldsymbol{y})$ with joint type P_{XY}. The set $\mathcal{L}_n(P_{XY})$ is given by

$$\mathcal{L}_n(P_{XY}) \triangleq \Big\{\bar{\mathsf{P}}_{XY} \in \mathcal{D}_n(\mathcal{X}\times\mathcal{Y}): \bar{\mathsf{P}}_Y = P_Y,\ \mathbb{E}\big[\log W(\bar{Y}|\bar{X})\big] \geq \mathbb{E}\big[\log W(Y|X)\big]\Big\}, \tag{58}$$

where $(\bar{X},\bar{Y}) \sim \bar{\mathsf{P}}_{XY}$ and $(X,Y)\sim P_{XY}$, and P_Y denotes the marginal distribution of P_{XY}. Here, and throughout this appendix, we indicate that $\boldsymbol{A}$ is distributed according to the distribution $P_{\boldsymbol{A}}$ by writing $\boldsymbol{A}\sim P_{\boldsymbol{A}}$. Analogously, we define the set $\mathcal{L}(P_{XY})$ as

$$\mathcal{L}(P_{XY}) \triangleq \Big\{\bar{P}_{XY} \in \mathcal{D}(\mathcal{X}\times\mathcal{Y}): \bar{P}_Y = P_Y, \mathbb{E}\big[\log W(\bar{Y}|\bar{X})\big] \geq \mathbb{E}\big[\log W(Y|X)\big]\Big\}, \tag{59}$$

with $(\bar{X},\bar{Y})\sim\bar{P}_{XY}$ and $(X,Y)\sim P_{XY}$.

Extending [11, Th. 1] to source-channel coding, we find that

$$\bar{\epsilon} \geq \frac{1}{4}\sum_{i=1}^{N'_k}\sum_{\boldsymbol{v}\in\mathcal{T}_i} P(\boldsymbol{v})\mathbb{E}\Bigg[\min\Big\{1, \sum_{\bar{\boldsymbol{v}}\in\mathcal{T}_i}\Pr\Big\{P^k(\bar{\boldsymbol{v}})W^n(\boldsymbol{Y}|\bar{\boldsymbol{X}}_i) \geq P^k(\boldsymbol{v})W^n(\boldsymbol{Y}|\boldsymbol{X}_i)\big|\boldsymbol{X}_i\boldsymbol{Y}\Big\}\Big\}\Bigg], \tag{60}$$

where $(\boldsymbol{X}_i,\boldsymbol{Y})\sim Q_i^n\times W^n$ and $\bar{\boldsymbol{X}}_i\sim Q_i^n$. Here we lower-bound $\bar{\epsilon}$ by considering in the inner sum only those $\bar{\boldsymbol{v}}$ that are in the source type class $\mathcal{T}_i$, $i=1,\dots,N'_k$.

We rewrite this bound in terms of summations over types,

$$\bar{\epsilon} \geq \frac{1}{4}\sum_{i=1}^{N'_k}\sum_{\mathsf{P}_{XY}} \Pr\{\boldsymbol{V}\in\mathcal{T}_i\}\Pr\big\{(\boldsymbol{X}_i,\boldsymbol{Y})\in\mathcal{T}(\mathsf{P}_{XY})\big\} \times\min\Bigg\{1, \sum_{\bar{\mathsf{P}}_{XY}\in\mathcal{L}_n(\mathsf{P}_{XY})}|\mathcal{T}_i| \times\Pr\Big\{(\bar{\boldsymbol{X}}_i,\boldsymbol{y})\in\mathcal{T}(\bar{\mathsf{P}}_{XY})\ \big|\ \boldsymbol{y}\in\bar{\mathsf{P}}_Y\Big\}\Bigg\}, \tag{61}$$

where $\boldsymbol{V}\sim P^k$.

Applying [12, Lemma 2.3] and [12, Lemma 2.6], we obtain

$$\bar{\epsilon} \geq \sum_{i=1}^{N'_k}\sum_{\mathsf{P}_{XY}} e^{-kD(\mathsf{P}_i\|P)-nD(\mathsf{P}_{XY}\|Q_i\times W)+\delta'_{k,n}-\log 4} \times\min\Bigg\{1, \sum_{\bar{\mathsf{P}}_{XY}\in\mathcal{L}_n(\mathsf{P}_{XY})} e^{kH(V_i)-nD(\bar{\mathsf{P}}_{XY}\|Q_i\times\bar{\mathsf{P}}_Y)+\delta'_{k,n}}\Bigg\}, \tag{62}$$

where $V_i\sim\mathsf{P}_i$ and $\delta'_{k,n}\triangleq\log(k+1)^{-|\mathcal{V}|}(n+1)^{-|\mathcal{X}||\mathcal{Y}|}$.

The error probability can be further bounded by keeping only the leading exponential term in each summation in (62). Taking logarithms on both sides of (62), multiplying the result by $-\frac{1}{n}$, and using the notation $[x]^+=\max(x,0)$ we obtain

$$-\frac{\log\bar{\epsilon}}{n} \leq \min_{i=1,\dots,N'_k}\min_{\mathsf{P}_{XY}}\min_{\bar{\mathsf{P}}_{XY}\in\mathcal{L}_n(\mathsf{P}_{XY})}\Bigg\{\frac{k}{n}D(\mathsf{P}_i\|P) + D(\mathsf{P}_{XY}\|Q_i\times W) + \Big[D(\bar{\mathsf{P}}_{XY}\|Q_i\times\bar{\mathsf{P}}_Y)-\frac{k}{n}H(V_i)\Big]^+\Bigg\} - \frac{\delta_{k,n}}{n}, \tag{63}$$

where we define $\delta_{k,n}\triangleq 2\delta'_{k,n}+\log 4$. Here we use that $[nx]^+=n[x]^+$, for $n>0$, that $[x]^+=\max(0,x)$ is monotonically non-decreasing, and that $[x+a]^+\leq[x]^++a$, $a>0$.

Any distribution in $\mathcal{D}(\mathcal{A})$ can be written as the limit of a sequence of types in $\mathcal{D}_n(\mathcal{A})$ [6, Sec. IV]. Hence, the uniform continuity of $D(A\|B)$ over the pair (A,B) [6] ensures that for every P_{XY}, and every $\xi_1>0$, there exists a sufficiently large n such that

$$-\frac{\log\bar{\epsilon}}{n} \leq \min_{i=1,\dots,N'_k}\min_{P_{XY}}\min_{\bar{P}_{XY}\in\mathcal{L}(P_{XY})}\Bigg\{\frac{k}{n}D(\mathsf{P}_i\|P) + D(P_{XY}\|Q_i\times W) + \Big[D(\bar{P}_{XY}\|Q_i\times\bar{P}_Y)-\frac{k}{n}H(V_i)\Big]^+\Bigg\} - \frac{\delta_{k,n}}{n}+\xi_1, \tag{64}$$

where we have replaced $\mathcal{L}_n(\mathsf{P}_{XY})$ by $\mathcal{L}(P_{XY})$, and used that $[x+a]^+\leq[x]^++a$, $a>0$.

It follows from [11, Th. 4] that

$$\min_{P_{XY}}\min_{\bar{P}_{XY}\in\mathcal{L}(P_{XY})}\big\{D(P_{XY}\|Q\times W) + [D(\bar{P}_{XY}\|Q\times\bar{P}_Y)-R]^+\big\} = \max_{\rho\in[0,1]}\big\{E_0(\rho,W,Q)-\rho R\big\}, \tag{65}$$

so (64) is equivalent to

$$-\frac{\log\bar{\epsilon}}{n} \leq \min_{i=1,\dots,N'_k}\Bigg\{\frac{k}{n}D(\mathsf{P}_i\|P) + \max_{\rho\in[0,1]}\Big\{E_0(\rho,W,Q_i)-\rho\frac{k}{n}H(V_i)\Big\}\Bigg\} - \frac{\delta_{k,n}}{n}+\xi_1. \tag{66}$$

Maximizing (66) over $Q_i\in\mathcal{Q}$ for each $i=1,\dots,N'_k$ yields

$$-\frac{\log\bar{\epsilon}}{n} \leq \min_{i=1,\dots,N'_k}\Bigg\{\frac{k}{n}D(\mathsf{P}_i\|P) + \max_{\rho\in[0,1]}\Big\{E_0(\rho,W,\mathcal{Q})-\rho\frac{k}{n}H(V_i)\Big\}\Bigg\} - \frac{\delta_{k,n}}{n}+\xi_1. \tag{67}$$

It follows from (3) that for every $\xi_2>0$ there exists a sufficiently large n_0 such that $\left|\frac{k}{n}-t\right|<\xi_2$ for all $n\geq n_0$.

Consequently, we can upper-bound (67) by

$$-\frac{\log \bar{\epsilon}}{n} \leq \min_{i=1,\dots,N'_k} \Bigg\{ tD(\mathsf{P}_i \| P) + \max_{\rho\in[0,1]} \left\{ E_0(\rho, W, \mathcal{Q}) - \rho t H(V_i) \right\} \Bigg\} - \frac{\delta_{k,n}}{n} + \xi_1 + \xi_2. \tag{68}$$

Using now the uniform continuity of the RHS of (68) as a function of P_i [1, p. 323] and that any distribution in $\mathcal{D}(\mathcal{V})$ can be written as the limit of a sequence of source types in k, it follows that for every $\xi_3 > 0$ there exists a sufficiently large n such that

$$-\frac{\log \bar{\epsilon}}{n} \leq \min_{P'} \Bigg\{ tD(P' \| P) + \max_{\rho\in[0,1]} \left\{ E_0(\rho, W, \mathcal{Q}) - \rho t H(V') \right\} \Bigg\} - \frac{\delta_{k,n}}{n} + \xi_1 + \xi_2 + \xi_3, \tag{69}$$

where $V' \sim P'$. By taking the limit superior in n, we obtain

$$\limsup_{n\to\infty} -\frac{\log \bar{\epsilon}}{n} \leq \min_{P'} \Big\{ tD(P' \| P) + \max_{\rho\in[0,1]} \left\{ E_0(\rho, W, \mathcal{Q}) - \rho t H(V') \right\} \Big\} + \xi_1 + \xi_2 + \xi_3 \tag{70}$$

$$= \min_{0\leq R\leq t\log|\mathcal{V}|} \left\{ te\left(\frac{R}{t}, P\right) + \max_{\rho\in[0,1]} \left\{ E_0(\rho, W, \mathcal{Q}) - \rho R \right\} \right\} + \xi_1 + \xi_2 + \xi_3 \tag{71}$$

$$= \max_{\rho\in[0,1]} \left\{ \bar{E}_0(\rho, W, \mathcal{Q}) - tE_{\mathrm{s}}(\rho, P) \right\} + \xi_1 + \xi_2 + \xi_3, \tag{72}$$

where (71) follows from the definition of the source reliability function [1, eq. (7)] with $R = tH(V')$; and (72) can be proved by the same methods that relate (15) and (16). Finally, letting ξ_1, ξ_2 and ξ_3 tend to zero from above yields the desired result.

References

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

[2] 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.

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

[4] R. T. Rockafellar, *Convex Analysis*, 2nd ed. Princeton, US: Princeton University Press, 1972.

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

[6] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[7] 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.

[8] A. Tauste Campo, G. Vazquez-Vilar, A. Guillen i Fabregas, T. Koch, and A. Martinez, "Random coding bounds that attain the joint source-channel exponent," in *46th Annual Conference on Information Sciences and Systems (CISS 2012)*, Princeton, USA, March 2012, invited.

[9] A. Tauste Campo and G. Vazquez-Vilar and A. Guillén i Fàbregas and T. Koch and A. Martínez, "Joint source-channel coding revisited: Random-coding bounds and error exponents," *Arxiv preprint arXiv:1303.6249v1*, March 2013.

[10] 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, May 2010.

[11] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Ensemble tight error exponent for mismatched decoders," in *50th Allerton Conf. on Comms. and Control*, Monticello, IL, Oct. 1-5 2012.

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

Adrià Tauste Campo (S'08, M'12) is a Marie Curie Intra-European Research Fellow with Universitat Pompeu Fabra, Barcelona, Spain. He received the Mathematics degree in 2005 and the Telecommunications Engineering degree in 2006, both from Universitat Politècnica de Catalunya, Barcelona, Spain, the M. Sc. in Electrical Engineering from Stanford University, U.S., in 2009 and the Ph.D in Engineering from University of Cambridge, U.K., in 2011. From September 2011 until February 2012 he was a Research Associate with the Department of Engineering, University of Cambridge, under the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Prize programme.

He has held research appointments at the University of South Australia, Adelaide, Australia, Telefonica Research Labs, Barcelona, Spain, and visiting appointments at the Universitá degli Studi di Cassino, Cassino, Italy, and Universidad Nacional Autónoma de México, México D.F., México. His research interests lie in the fields of Shannon theory and statistical learning with applications to computational neuroscience.

Gonzalo Vazquez-Vilar (S'08, M'12) received the Telecommunication Engineering degree from the University of Vigo, Spain, in 2004, the Master of Science degree from Stanford University, U.S., in 2008 and the Ph.D. in Communication Systems from the University of Vigo, Spain, in 2011.

Since 2011 he has been a Research Associate in the Department of Information and Communication Technologies at Universitat Pompeu Fabra, Spain. He has held appointments as project engineer at Siemens AG, Germany, and as visiting researcher at Stanford University, U.S., and University of Cambridge, U.K. His research interests lie in the field of Shannon theory, with emphasis on finite-length information theory and communications.

Albert Guillén i Fàbregas (S '01 – M '05 – SM'09) was born in Barcelona, Catalunya, Spain, in 1974. In 1999 he received the Telecommunication Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

Since 2011 he has been a Research Professor of the Institució Catalana de Recerca i Estudis Avançats (ICREA) hosted at the Department of Information and Communication Technologies, Universitat Pompeu Fabra. He is also an Adjunct Researcher at the Department of Engineering, University of Cambridge. He has held appoinments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, University of Cambridge where he was a Reader and a Fellow of Trinity Hall, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His specific research interests are in the area of information theory, communication theory, coding theory, digital modulation and signal processing techniques.

Dr. Guillén i Fàbregas received the Starting Grant from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is a Member of the Young Academy of Europe. He is a co-author of the monograph book "Bit-Interleaved Coded Modulation". He is also an Associate Editor of the IEEE Transactions on Information Theory, an Editor of the Foundations and Trends in Communications and Information Theory, Now Publishers and was an Editor of the IEEE Transactions on Wireless Communications (2007-2011).

Tobias Koch (S'02, M'09) is a Visiting Professor with the Signal Theory and Communications Department of Universidad Carlos III de Madrid (UC3M), Spain. He received the M.Sc. degree in electrical engineering (with distinction) in 2004 and the Ph.D. degree in electrical engineering in 2009, both from ETH Zurich, Switzerland. From June 2010 until May 2012 he was a Marie Curie Intra-European Research Fellow with the University of Cambridge, UK. He was also a research intern at Bell Labs, Murray Hill, NJ in 2004 and at Universitat Pompeu Fabra (UPF), Spain, in 2007. He joined UC3M in 2012. His research interests include digital communication theory and information theory.

Dr. Koch is serving as Vice Chair of the Spain Chapter of the IEEE Information Theory Society in 2013–2014.

Alfonso Martinez (SM '11) was born in Zaragoza, Spain, in October 1973. He is currently a Ramón y Cajal Research Fellow at Universitat Pompeu Fabra, Barcelona, Spain. He obtained his Telecommunications Engineering degree from the University of Zaragoza in 1997. In 1998–2003 he was a Systems Engineer at the research centre of the European Space Agency (ESA-ESTEC) in Noordwijk, The Netherlands. His work on APSK modulation was instrumental in the definition of the physical layer of DVB-S2. From 2003 to 2007 he was a Research and Teaching Assistant at Technische Universiteit Eindhoven, The Netherlands, where he conducted research on digital signal processing for MIMO optical systems and on optical communication theory. Between 2008 and 2010 he was a post-doctoral fellow with the Information-theoretic Learning Group at Centrum Wiskunde & Informatica (CWI), in Amsterdam, The Netherlands. In 2011 he was a Research Associate with the Signal Processing and Communications Lab at the Department of Engineering, University of Cambridge, Cambridge, U.K.

His research interests lie in the fields of information theory and coding, with emphasis on digital modulation and the analysis of mismatched decoding; in this area he has coauthored a monograph on "Bit-Interleaved Coded Modulation". More generally, he is intrigued by the connections between information theory, optical communications, and physics, particularly by the links between classical and quantum information theory.