Achieving Csiszár's Source-Channel Coding Exponent with Product Distributions

Adrià Tauste Campo[12], 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*—We derive a random-coding upper bound on the average probability of error of joint source-channel coding that recovers Csiszár's error exponent when used with product distributions over the channel inputs. Our proof technique for the error probability analysis employs a code construction for which source messages are assigned to subsets and codewords are generated with a distribution that depends on the subset.**

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 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.

An encoder maps the length-k 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 length-n channel output $\boldsymbol{y}$ the decoder guesses which source message was transmitted.

We say that an error exponent $E > 0$ is *achievable* if there exists a sequence of codes of length n such that the average error probability ϵ (averaged over all source messages $\boldsymbol{v}$) is upper-bounded as

$$\epsilon \leq e^{-nE+o(n)}, \tag{1}$$

where $o(n)$ satisfies $\lim_{n\to\infty} o(n)/n = 0$. The error exponent E_{J} of joint source-channel coding is defined as the supremum of all achievable error exponents E.

Lower bounds on the error exponent are often derived by drawing an ensemble of codebooks at random, and by then analyzing the average error probability $\bar{\epsilon}$, averaged over all codebooks in the ensemble. This computation ensures the existence of at least one codebook in the ensemble whose average error probability ϵ is at most $\bar{\epsilon}$ [1, Sec. 5.5].

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

In [1, Prob. 5.16], Gallager provides an upper bound on $\bar{\epsilon}$ using maximum a posteriori (MAP) decoding when the codewords corresponding to different source messages are drawn independently according to some distribution $P_{\boldsymbol{X}}$:

$$\bar{\epsilon} \leq e^{-E_0(\rho, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}) + E_{\text{s}}(\rho, P_{\boldsymbol{V}})}, \quad \rho \in [0, 1], \tag{2}$$

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

$$E_0(\rho, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}) \triangleq -\log \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{x}} P_{\boldsymbol{X}}(\boldsymbol{x}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{3}$$

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

$$E_{\text{s}}(\rho, P_{\boldsymbol{V}}) \triangleq \log \left(\sum_{\boldsymbol{v}} P_{\boldsymbol{V}}(\boldsymbol{v})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{4}$$

The upper bound (2) is derived using similar techniques as Gallager's channel coding bound [1, p. 135]. In the context of discrete memoryless systems and for $P_{\boldsymbol{X}}$ being a product distribution $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} P_X(x_i)$, $\boldsymbol{x} \in \mathcal{X}^n$, it specializes to

$$\bar{\epsilon} \leq e^{-n\left(E_0(\rho, P_{Y|X}, P_X) - tE_{\text{s}}(\rho, P_V)\right)}. \tag{5}$$

Thus, for a fixed t, the probability of error $\bar{\epsilon}$ vanishes exponentially in n with the error exponent given by $E_0(\rho, P_{Y|X}, P_X) - tE_{\text{s}}(\rho, P_V)$. By minimizing (5) over P_X and ρ, we obtain the following lower bound on the error exponent E_J:

$$E_{\text{J}} \geq E_{\text{J}}^{\text{G}} \triangleq \max_{\rho\in[0,1]} \left\{ E_0(\rho, P_{Y|X}) - tE_{\text{s}}(\rho, P_V) \right\}, \tag{6}$$

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

Csiszár refined this result using the method of types [2]. Csiszár's approach is different from Gallager's in several ways. Firstly, Csiszár considers *fixed composition codes* rather than codes that are generated by product distributions – such codes are constructed by mapping messages within a source type onto sequences within a channel-input type. Secondly, a suboptimal maximum mutual information decoder is used at the receiver. This decoder first decides on the source type that is being transmitted and then on the source message within the type. Csiszár's code construction yields the following lower bound on the error exponent:

$$E_{\mathrm{J}} \geq E_{\mathrm{J}}^{\mathrm{Cs}} \triangleq \min_{tH(V)\leq R\leq t\log|\mathcal{V}|} \left\{ te\left(\frac{R}{t}, P_V\right) + E_{\mathrm{r}}(R, P_{Y|X}) \right\}, \quad (7)$$

where $e(R, P_V)$, referred to as the source reliability function [3]–[5], is given by

$$e(R, P_V) \triangleq \min_{Q:H(Q)\geq R} D(P_Q \| P_V) \quad (8)$$

$$= \sup_{\rho\geq 0} \big\{ \rho R - E_{\mathrm{s}}(\rho, P_V) \big\}, \quad (9)$$

with $D(\cdot\|\cdot)$ denoting the divergence, and with $E_{\mathrm{r}}(R, P_{Y|X})$ denoting the random-coding channel exponent [1]

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

A lower bound on the error probability of the best code induces an upper bound on E_{J}. One such upper bound is given by the sphere-packing expoment [2, Lemma 2]

$$E_{\mathrm{J}} \leq \min_{tH(V)\leq R\leq t\log|\mathcal{V}|} \left\{ te\left(\frac{R}{t}, P_V\right) + E_{\mathrm{sp}}(R, P_{Y|X}) \right\}. \quad (11)$$

It can be checked that $E_{\mathrm{J}} = E_{\mathrm{J}}^{\mathrm{Cs}}$ when the minimum on the right-hand side (RHS) of (11) is attained for a value of R such that $E_{\mathrm{sp}}(R, P_{Y|X}) = E_{\mathrm{r}}(R, P_{Y|X})$. This holds for values of R above the critical rate of the channel [1].

In order to obtain a clearer comparison between (6) and (7), Zhong *et al.* [6] provide a compact formulation of Csiszár's result. Specifically, the authors invoke the Fenchel duality theorem [7, Thm. 31.1] to rewrite (7) as

$$E_{\mathrm{J}}^{\mathrm{Cs}} = \max_{\rho\in[0,1]} \big\{ \bar{E}_0(\rho, P_{Y|X}) - tE_{\mathrm{s}}(\rho, P_V) \big\}, \quad (12)$$

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 $\rho \in [0, 1]$ [7, Cor. 12.1.1]. It follows from (12) that $E_{\mathrm{J}}^{\mathrm{Cs}} \geq E_{\mathrm{J}}^{\mathrm{G}}$, with the inequality possibly being strict.

In cases where the above inequality is strict, the difference between $E_{\mathrm{J}}^{\mathrm{Cs}}$ and $E_{\mathrm{J}}^{\mathrm{G}}$ is typically small [6]. The methods used to derive each exponent are conceptually different, raising a number of questions. In this paper, we address the question of whether fixed composition codes are necessary in order to achieve Csiszár's exponent. We show that random codes generated by product distributions together with MAP decoding and bounding techniques based on Markov's inequality can be used to recover Csiszár's exponent, answering the question in the negative.

II. Main Results

The derivation of the main results involves the following steps:

1) Define a partition $\mathcal{P}_k$ of the message set $\mathcal{V}^k$ into N_k disjoint subsets $\mathcal{A}_1, \ldots, \mathcal{A}_{N_k}$ 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 MAP decoding and Gallager's bounding techniques [1].

For $i = 1, \ldots, N_k$ we define

$$E_{\mathrm{s}}^{(i)}(\rho, P_{\boldsymbol{V}}) \triangleq \log \left(\sum_{\boldsymbol{v}\in\mathcal{A}_i} P_{\boldsymbol{V}}(\boldsymbol{v})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (13)$$

Theorem 1: For every partition $\mathcal{P}_k$, for every set of channel-input distributions $P_{\boldsymbol{X}}^{(1)}, \ldots, P_{\boldsymbol{X}}^{(N_k)}$, and for every set of parameters $\rho_1, \ldots, \rho_{N_k} \in [0, 1]$, the random-coding error probability is upper-bounded by

$$\bar{\epsilon} \leq h(k) \sum_{i=1}^{N_k} \exp\Big(-E_0\big(\rho_i, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}^{(i)}\big) + E_{\mathrm{s}}^{(i)}(\rho_i, P_{\boldsymbol{V}}) \Big), \quad (14)$$

where

$$h(k) \triangleq 2N_k(k+1)^{|\mathcal{V}|}(k/t+1)^{|\mathcal{X}||\mathcal{Y}|}. \quad (15)$$

Proof: See Section III-A. ∎

The bound (14) can be optimized over product distributions $P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) = \prod_{j=1}^{n} P_X^{(i)}(x_j)$, $\boldsymbol{x} \in \mathcal{X}^n$, and parameters $\rho_i \in [0, 1]$ for $i = 1, \ldots, N_k$ to obtain

$$\bar{\epsilon} \leq \bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \quad (16)$$

$$\triangleq h(k) \sum_{i=1}^{N_k} \exp\Big(-\max_{\rho_i\in[0,1]} \Big\{ nE_0\big(\rho_i, P_{Y|X}\big) - E_{\mathrm{s}}^{(i)}(\rho_i, P_{\boldsymbol{V}}) \Big\} \Big). \quad (17)$$

If the partition $\mathcal{P}_k$ only has one class, i.e., $\mathcal{A}_1 = \mathcal{V}^k$ for $k = 1, 2, \ldots$, the upper bound (16) recovers Gallager's bound on the error exponent (6):

$$\lim_{n\to\infty} -\frac{1}{n} \log \bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \geq E_{\mathrm{J}}^{\mathrm{G}}. \quad (18)$$

As we shall see next, with a more judicious choice of $\mathcal{P}_k$ the upper bound (16) also recovers Csiszár's lower bound on the error exponent (7).

Specifically, (7) can be achieved by identifying the classes $\mathcal{A}_1, \ldots, \mathcal{A}_{N_k}$ with the source-type classes $\mathcal{T}_1, \ldots, \mathcal{T}_{N_k}$. 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 [3, 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

$$\mathsf{P}_i(a) = \frac{1}{k} N(a|\boldsymbol{v}), \quad a \in \mathcal{V}, \quad (19)$$

where $N(a|\boldsymbol{v})$ denotes the number of occurrences of $a \in \mathcal{V}$ in $\boldsymbol{v}$.

Corollary 1: Let the classes $\mathcal{A}_1, \dots, \mathcal{A}_{N_k}$ of the partition $\mathcal{P}_k$ be the source-type classes $\mathcal{T}_1, \dots, \mathcal{T}_{N_k}$. Then

$$\liminf_{n\to\infty} -\frac{1}{n} \log \bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \geq E_{\mathrm{J}}^{\mathrm{Cs}}. \quad (20)$$

Proof: See Section III-B. ■

By the type counting lemma [3, Lemma 2.2], there are at most $(k+1)^{|\mathcal{V}|}$ different source-type classes. Thus, Corollary 1 demonstrates that partitions $\mathcal{P}_k$ with not more than $(k+1)^{|\mathcal{V}|}$ classes are sufficient to achieve Csiszár's error exponent. In fact, we have recently showed that Csiszár's error exponent can already be achieved with partitions $\mathcal{P}_k$ consisting of two classes [8].

III. PROOFS

A. Proof of Theorem 1

The random-coding error probability $\bar{\epsilon}$ is upper-bounded by the random-coding union (RCU) bound [9], [10]

$$\bar{\epsilon} \leq \sum_{i}^{N_k} \bar{\epsilon}(i), \quad (21)$$

where

$$\bar{\epsilon}(i) \triangleq \sum_{\boldsymbol{v}\in\mathcal{A}_i} P_{\boldsymbol{V}}(\boldsymbol{v}) \sum_{\boldsymbol{x},\boldsymbol{y}} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \times \min\left\{1, \sum_{j=1}^{N_k} \sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_j} \sum_{\substack{\bar{\boldsymbol{x}}: P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\bar{\boldsymbol{x}}) \\ \geq P_{\boldsymbol{V}}(\boldsymbol{v})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}} P_{\boldsymbol{X}}^{(j)}(\bar{\boldsymbol{x}})\right\}. \quad (22)$$

Let $\{\mathcal{T}_1, \dots, \mathcal{T}_{M_k}\}$ be the set of all source-type classes in $\mathcal{V}^k$. Furthermore, let $\{\mathcal{T}_1(\boldsymbol{y}), \dots, \mathcal{T}_{M_n}(\boldsymbol{y})\}$ be the set of all V-shells of $\boldsymbol{y}$ in $\mathcal{X}^n$, where the V-shell $\mathcal{T}_m(\boldsymbol{y})$ is defined as the set of all channel inputs $\boldsymbol{x} \in \mathcal{X}^n$ with conditional type V_m given $\boldsymbol{y} \in \mathcal{Y}^n$ [3, Def. 2.4]. Thus, for a given conditional type V_m given $\boldsymbol{y} \in \mathcal{Y}^n$, the V-shell $\mathcal{T}_m(\boldsymbol{y})$ is the set of all channel inputs $\boldsymbol{x} \in \mathcal{X}^n$ satisfying

$$N(a,b|\boldsymbol{x},\boldsymbol{y}) = N(b|\boldsymbol{y})\mathsf{V}_m(a|b), \quad a \in \mathcal{X},\, b \in \mathcal{Y}, \quad (23)$$

where $N(a,b|\boldsymbol{x},\boldsymbol{y})$ denotes the number of occurrences of $(a,b) \in \mathcal{X} \times \mathcal{Y}$ in $(\boldsymbol{x},\boldsymbol{y})$. Note that, by the type counting lemma,

$$M_k \leq (k+1)^{|\mathcal{V}|} \quad \text{and} \quad M_n \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}. \quad (24)$$

Using the above definitions, we can rewrite (22) as

$$\bar{\epsilon}(i) = \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} \sum_{\boldsymbol{v}\in\mathcal{T}_\ell\cap\mathcal{A}_i} \sum_{\boldsymbol{x}\in\mathcal{T}_m(\boldsymbol{y})} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \times \min\left\{1, \sum_{j=1}^{N_k} \sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_j} \sum_{\substack{\bar{\boldsymbol{x}}: P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\bar{\boldsymbol{x}}) \\ \geq P_{\boldsymbol{V}}(\boldsymbol{v})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}} P_{\boldsymbol{X}}^{(j)}(\bar{\boldsymbol{x}})\right\}. \quad (25)$$

It is easy to show that $P_{\boldsymbol{V}}(\cdot)$ is constant within every source-type class $\mathcal{T}_\ell$ and that $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\cdot)$ is constant within every V-shell $\mathcal{T}_m(\boldsymbol{y})$ of $\boldsymbol{y}$. This allows us to define the metric

$$d(\boldsymbol{y},\ell,m) \triangleq P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}), \quad (\boldsymbol{v},\boldsymbol{x}) \in \mathcal{T}_\ell \times \mathcal{T}_m(\boldsymbol{y}) \quad (26)$$

for every $\boldsymbol{y} \in \mathcal{Y}^n$ and $\ell = 1, \dots, M_k$, $m = 1, \dots, M_n$. We further define

$$G_j(\boldsymbol{y}, a) \triangleq \sum_{\boldsymbol{v}\in\mathcal{A}_j} \sum_{\boldsymbol{x}: P_{\boldsymbol{V}}(\boldsymbol{v})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \geq a} P_{\boldsymbol{X}}^{(j)}(\boldsymbol{x}). \quad (27)$$

Using (26) and (27) in (25) yields

$$\bar{\epsilon}(i) = \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} d(\boldsymbol{y},\ell,m) \min\left\{1, \sum_{j=1}^{N_k} G_j\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\right\} \times \sum_{\boldsymbol{v}\in\mathcal{T}_\ell\cap\mathcal{A}_i} \sum_{\boldsymbol{x}\in\mathcal{T}_m(\boldsymbol{y})} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}). \quad (28)$$

We now focus on the last double sum term in (28), which can be upper-bounded as

$$\begin{aligned}
\sum_{\boldsymbol{v}\in\mathcal{T}_\ell\cap\mathcal{A}_i} \sum_{\boldsymbol{x}\in\mathcal{T}_m(\boldsymbol{y})} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x})
&\leq \sum_{\boldsymbol{v}\in\mathcal{T}_\ell\cap\mathcal{A}_i} \sum_{\substack{\boldsymbol{x}: P_{\boldsymbol{V}}(\boldsymbol{v})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \\ = d(\boldsymbol{y},\ell,m)}} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) && (29)\\
&\leq \sum_{\boldsymbol{v}\in\mathcal{A}_i} \sum_{\substack{\boldsymbol{x}: P_{\boldsymbol{V}}(\boldsymbol{v})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \\ \geq d(\boldsymbol{y},\ell,m)}} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) && (30)\\
&= G_i\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big) && (31)
\end{aligned}$$

for every $\boldsymbol{y} \in \mathcal{Y}^n$ and $\ell = 1, \dots M_k$, $m = 1, \dots, M_n$. Here (29) follows from the fact that not every $\boldsymbol{x}$ satisfying

$$P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = d(\boldsymbol{y},\ell,m), \quad \boldsymbol{v} \in \mathcal{T}_\ell \quad (32)$$

must be in $\mathcal{T}_m(\boldsymbol{y})$, and (30) follows from summing over the entire message set $\mathcal{A}_i$ and a larger codeword set. Combining (31) with (28) and (21) yields

$$\begin{aligned}
\bar{\epsilon} &\leq \sum_{i=1}^{N_k} \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} d(\boldsymbol{y},\ell,m) \times \min\left\{1, \sum_{j=1}^{N_k} G_j\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\right\} G_i\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big) && (33)\\
&\leq \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} d(\boldsymbol{y},\ell,m) \times \min\left\{1, G_j\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\right\} G_i\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big) && (34)\\
&\leq \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} d(\boldsymbol{y},\ell,m) \times \Big(\min\big\{1, G_j\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\big\} G_j\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big) \\
&\qquad + \min\big\{1, G_i\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\big\} G_i\big(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\big)\Big), && (35)
\end{aligned}$$

where in (34) we have used the inequality $\min\{1, x+y\} \leq \min\{1,x\}+\min\{1,y\}$, $x, y \geq 0$, and in (35) we have used the inequality $\min\{1,x\}y \leq \min\{1,x\}x + \min\{1,y\}y$, $x, y \geq 0$. By rearranging the terms in (35), we obtain

$$\bar{\epsilon} \leq 2N_k \sum_{i=1}^{N_k} \sum_{\boldsymbol{y}} \sum_{\ell=1}^{M_k} \sum_{m=1}^{M_n} d(\boldsymbol{y},\ell,m) \times \min\bigl\{1, G_i\bigl(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\bigr)\bigr\} G_i\bigl(\boldsymbol{y}, d(\boldsymbol{y},\ell,m)\bigr). \quad (36)$$

We next use Markov's inequality to upper-bound

$$G_i(\boldsymbol{y},a) = \sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_i} \sum_{\bar{\boldsymbol{x}}:\left(P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\bar{\boldsymbol{x}})\right)^{s_i} \geq a^{s_i}} P_{\boldsymbol{X}}^{(i)}(\bar{\boldsymbol{x}}) \quad (37)$$

$$\leq \sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_i} \sum_{\bar{\boldsymbol{x}}} P_{\boldsymbol{X}}^{(i)}(\bar{\boldsymbol{x}}) \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\bar{\boldsymbol{x}})}{a}\right)^{s_i}, \quad (38)$$

for $a, s_i > 0$. Using that $\min\{1,x\} \leq x^{\rho_i}$ for $x \geq 0$ and $\rho_i \in [0,1]$, and choosing $s_i = \frac{1}{1+\rho_i}$ in (38), we obtain

$$\min\bigl\{1, G_i(\boldsymbol{y},a)\bigr\} G_i(\boldsymbol{y},a) \leq \bigl(G_i(\boldsymbol{y},a)\bigr)^{1+\rho_i} \quad (39)$$

$$\leq \left(\sum_{\bar{\boldsymbol{v}}\in\mathcal{A}_i} \sum_{\bar{\boldsymbol{x}}} P_{\boldsymbol{X}}^{(i)}(\bar{\boldsymbol{x}}) \left(\frac{P_{\boldsymbol{V}}(\bar{\boldsymbol{v}})P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\bar{\boldsymbol{x}})}{a}\right)^{\frac{1}{1+\rho_i}} \right)^{1+\rho_i}. \quad (40)$$

By applying (40) with $a = d(\boldsymbol{y},\ell,m)$ to (36), we finally obtain

$$\bar{\epsilon} \leq 2N_k M_k M_n \sum_{i=1}^{N_k} \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{x}} P_{\boldsymbol{X}}^{(i)}(\boldsymbol{x}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho_i}} \right)^{1+\rho_i} \times \left(\sum_{\boldsymbol{v}\in\mathcal{A}_i} P_{\boldsymbol{V}}(\boldsymbol{v})^{\frac{1}{1+\rho_i}} \right)^{1+\rho_i} \quad (41)$$

$$= 2N_k M_k M_n \sum_{i=1}^{N_k} e^{-E_0\left(\rho_i, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}^{(i)}\right) + E_{\mathrm{s}}^{(i)}(\rho_i, P_{\boldsymbol{V}})}. \quad (42)$$

Theorem 1 follows then by upper-bounding M_k and M_n using (24).

B. Proof of Corollary 1

Let the classes $\mathcal{A}_1, \ldots, \mathcal{A}_{N_k}$ of the partition $\mathcal{P}_k$ be the source-type classes $\mathcal{T}_1, \ldots, \mathcal{T}_{N_k}$. We first note that, by the type counting lemma,

$$N_k \leq (k+1)^{|\mathcal{V}|}. \quad (43)$$

Furthermore, $E_{\mathrm{s}}^{(i)}(\rho, P_{\boldsymbol{V}})$, $i = 1, \ldots, N_k$, can be rewritten as

$$E_{\mathrm{s}}^{(i)}(\rho, P_{\boldsymbol{V}}) = \rho_i \log|\mathcal{T}_i| + \log\left(\sum_{\boldsymbol{v}\in\mathcal{T}_i} P_{\boldsymbol{V}}(\boldsymbol{v}) \right), \quad (44)$$

since $P_{\boldsymbol{V}}(\cdot)$ is constant within each source-type class $\mathcal{A}_i = \mathcal{T}_i$.

Let V_i be a random variable whose distribution is the type P_i associated with $\mathcal{T}_i$. Then, if we define $R_i \triangleq tH(V_i)$ (where $H(V_i)$ denotes the entropy of V_i), we have the following inequalities [3, Lemmas 2.3 & 2.6]:

$$\frac{\log|\mathcal{T}_i|}{n} \leq R_i, \quad (45)$$

$$\log\left(\sum_{\boldsymbol{v}\in\mathcal{T}_i} P_{\boldsymbol{V}}(\boldsymbol{v}) \right) \leq -kD\left(\mathsf{P}_i \| P_V\right) \quad (46)$$

$$\leq -k \min_{\substack{j=1,\ldots,N_k: \\ H(V_j)\geq H(V_i)}} D\left(\mathsf{P}_j \| P_V\right) \quad (47)$$

$$\leq -ke\left(\frac{R_i}{t}, P_V\right), \quad (48)$$

where in (48) we have used the definitions of R_i and of the source reliability function (8).

Using (44)–(48), and using that $t = k/n$, we can upper-bound (17) as

$$\bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \leq h(k) \sum_{i=1}^{N_k} e^{-n\left(\max_{\rho_i\in[0,1]}\left\{E_0(\rho_i, P_{Y|X}) - \rho_i R_i\right\} + te\left(\frac{R_i}{t}, P_V\right)\right)} \quad (49)$$

$$= h(k) \sum_{i=1}^{N_k} e^{-n\left(E_{\mathrm{r}}(R_i, P_{Y|X}) + te\left(\frac{R_i}{t}, P_V\right)\right)} \quad (50)$$

$$\leq h(k) N_k e^{-n \min_{0<R\leq t\log|\mathcal{V}|}\left\{E_{\mathrm{r}}(R, P_{Y|X}) + te\left(\frac{R}{t}, P_V\right)\right\}}, \quad (51)$$

where in (50) we have used the definition of the random-coding channel exponent (10), and where (51) follows from minimizing the exponent over all possible values of R_i.

Using the definition of $h(k)$ in (14) and the bound (43), we have that

$$h(k)N_k \leq (k+1)^{3|\mathcal{V}|}(k/t+1)^{|\mathcal{X}||\mathcal{Y}|}, \quad (52)$$

which is polynomial in k. Hence, (51) yields

$$\liminf_{n\to\infty} -\frac{1}{n}\log\bar{\epsilon}_{\mathrm{B}}(\mathcal{P}_k) \geq \min_{0<R\leq t\log|\mathcal{V}|}\left\{E_{\mathrm{r}}\left(R, P_{Y|X}\right) + te\left(\frac{R}{t}, P_V\right)\right\}. \quad (53)$$

Since $E_{\mathrm{r}}\left(R, P_{Y|X}\right) + te\left(R/t, P_V\right)$ is a decreasing function of R for $0 < R \leq tH(V)$, it follows that the RHS of (53) is equal to the RHS of (7), thus proving Corollary 1.

IV. Extension to General Alphabets

One of the strengths of Gallager's error bound (2) is that it can be easily generalized to nondiscrete channels without resorting to limiting arguments applied to ever-finer quantizations of $\mathcal{X}$ and $\mathcal{Y}$.

While in the derivation of our new bound we mostly used the same techniques as Gallager, there are some steps that rely on the method of types. In particular, to analyze (22), we partitioned $\mathcal{V}^k$ into source-type classes and $\mathcal{X}^n$ into V-shells of $\boldsymbol{y} \in \mathcal{Y}^n$. Nevertheless, these partitions were merely introduced to simplify the analysis and are not essential. Indeed, Csiszár's error exponent can also be obtained by partitioning $\mathcal{V}^k \times \mathcal{X}^n$ into $\kappa+1$ sets of the form

$$\mathcal{S}_\ell(\boldsymbol{y}) = \left\{(\boldsymbol{v},\boldsymbol{x}) \in \mathcal{V}^k \times \mathcal{X}^n \colon P_{\boldsymbol{V}}(\boldsymbol{v}) P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \in [\alpha_\ell, \beta_\ell)\right\} \tag{54}$$

for every $\boldsymbol{y} \in \mathcal{Y}^n$, where κ is a linear function of n, and where $\alpha_0 = 0$, $\alpha_\ell = e^{-n\gamma} e^{\ell-1}$, $\ell = 1, \dots, \kappa$, and $\beta_\ell = e^{-n\gamma} e^{\ell}$, $\ell = 0, \dots, \kappa-1$, $\beta_\kappa = \infty$ for some $\gamma > 0$.

Following the arguments in Sections III-A while treating $\mathcal{S}_0$ and $\mathcal{S}_\kappa$ separately we obtain the upper bound

$$\begin{aligned} \bar{\epsilon} \leq 2N_k \kappa e \sum_{i=1}^{N_k} e^{-\max_{\rho_i \in [0,1]} \left\{ E_0\left(\rho_i, P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}}^{(i)}\right) - E_{\mathrm{s}}^{(i)}(\rho_i, P_{\boldsymbol{V}}) \right\}} \\ + N_k \sum_{i=1}^{N_k} \Pr\Big(P_{\boldsymbol{V}}(\boldsymbol{V}) P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\boldsymbol{Y}|\boldsymbol{X}_i\big) < e^{-n\gamma} \Big) \\ + N_k \sum_{i=1}^{N_k} \Pr\Big(P_{\boldsymbol{V}}(\boldsymbol{V}) P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\boldsymbol{Y}|\boldsymbol{X}_i\big) \geq e^{-n\gamma+\kappa-1} \Big), \end{aligned} \tag{55}$$

where $\boldsymbol{X}_i$ is a random variable with distribution $P_{\boldsymbol{X}}^{(i)}$, for $i = 1, \dots, N_k$. By judiciously choosing γ and κ, and by maximizing (55) over product distributions, we recover Csiszár's error exponent.

A lower bound on E_J for nondiscrete channels follows along the same lines by replacing the channel law $P_{\boldsymbol{Y}|\boldsymbol{X}}$ in (54) with the corresponding Radon-Nikodym derivative $f_{\boldsymbol{Y}|\boldsymbol{X}}$.

V. Concluding Remarks

We have presented an upper bound on the random-coding error probability for joint source-channel coding that recovers Gallager's and Csiszár's lower bounds on the error exponent for discrete memoryless systems. Thus, the new expression gives the actual error exponent at least in the cases where Csiszár's exponent is tight.

The method to obtain the new bound uses a specific random-coding construction with MAP decoding. Specifically, we partition the message set into disjoints classes and assign to each class an input distribution according to which the codewords are randomly generated.

By partitioning the message set into source-type classes, and by choosing for each class the input distribution to be a product distribution, the new bound on the error probability recovers Csiszár's lower bound on the error exponent, answering the question of whether fixed composition codes are required to achieve Csiszár's exponent in the negative.

References

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

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

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

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

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

[6] 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.

[7] R. T. Rockafellar, *Conjugate duality and optimization*. SIAM, 1974.

[8] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martinez, "Random coding bounds that attain the joint source-channel exponent," in *Proc. 46th Ann. Conf. Inf. Sciences and Systems (CISS)*, Princeton University, NJ, March 2012.

[9] 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.

[10] 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.