Two-Class Joint Source-Channel Coding: Expurgated Exponents with i.i.d. Distributions

Seyed AmirPouya Moeini
University of Cambridge
sam297@cam.ac.uk

Albert Guillén i Fàbregas
University of Cambridge
Universitat Politècnica de Catalunya
guillen@ieee.org

Abstract—**This paper studies expurgated exponents for joint source-channel coding of discrete memoryless sources and channels under i.i.d. random coding. We show that a two-class partitioning of source sequences, where the codeword distribution depends on the source type, achieves an exponent at least as high as that of optimal single-class coding, in which the codeword distribution is independent of the source message.**

I. INTRODUCTION

We study the transmission of non-equiprobable messages from a discrete memoryless source (DMS) over a discrete memoryless channel (DMC). In 1980, Csiszár [1] showed that joint source-channel coding can achieve a lower error probability than separate source and channel coding, and derived the corresponding random-coding error exponent using a construction in which codewords are drawn from a distribution that depends on the source message. In contrast, Gallager [2] had earlier derived a different error exponent where codewords are drawn from a fixed distribution, independent of the source message. Zhong et al. [3] later revisited these results and showed that Csiszár's exponent is always at least as large as Gallager's, and sometimes strictly larger.

Compared to random coding, expurgated bounds in joint source-channel coding have received relatively little attention. In the channel coding, the main techniques for obtaining expurgated exponents are due to Gallager [2, Section 5.7] and Csiszár-Körner-Marton (CKM) [4], [5]. See [6] for a more extensive study of expurgated bounds in channel coding.

In [7], Csiszár generalized the CKM method and derived two expurgated exponents for joint source-channel coding. The first corresponds to single-class coding, in which all codewords are drawn from a fixed distribution (a Gallager-like setting), while the second assigns a distinct codeword distribution to each source-type class, so that the codeword distribution of a source sequence depends on its type, following [1]. The comparison between these two exponents has remained an open problem. Scarlett et al. [8] later revisited the expurgation problem in joint source-channel coding and rederived Csiszár's exponents using a new technique based on source-type duplication and type-by-type expurgation. More recently, Moeini et al. [9] extended this line of work by considering a broader class of

message partitionings, and showed the existence of a two-class partitioning that recovers Csiszár's second exponent in [7].

To the best of our knowledge, the partitioned and non-partitioned approaches have not been directly compared in the expurgated setting, either theoretically or numerically. It has therefore remained open whether partitioning can yield a larger exponent, as it does in the random coding case.

In this work, we focus on i.i.d. random coding ensembles for convenience, in contrast to earlier studies [1], [7]–[9] that considered constant-composition coding. We establish the existence of a two-class partitioning that achieves an exponent no smaller than that of optimal single-class coding, thereby settling one side of the comparison. In our analysis, single-class coding appears naturally as a special case of partitioning, an observation aligns with intuition but not reflected in previous studies. The detailed proofs are presented in Section V.

II. PRELIMINARIES

We study the transmission of non-equiprobable messages from a DMS with distribution $P^k(\boldsymbol{v}) = \prod_{i=1}^{k} P_V(v_i)$, where $\boldsymbol{v} = (v_1, ..., v_k) \in \mathcal{V}^k$ is the source message, and $\mathcal{V}$ is a finite discrete alphabet. The channel is a DMC given by $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i)$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, 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, and decoded as $\hat{\boldsymbol{v}}$ at the receiver upon observing the channel output $\boldsymbol{y}$. We refer to $t \triangleq k/n$ as the transmission rate. We study the average error probability, defined as

$$p_e \triangleq \mathbb{P}[\boldsymbol{V} \neq \hat{\boldsymbol{V}}]. \tag{1}$$

We say that an error exponent $E > 0$ is *achievable* (for any fixed t) if there exists a sequence of codes of length n such that the error probability satisfies

$$p_e \leq e^{-nE + o(n)}, \tag{2}$$

where $\lim_{n \to \infty} o(n)/n = 0$.

In this paper, scalar random variables are denoted by upper-case letters, their realizations by lowercase letters, and their alphabets by calligraphic letters. Random vectors are written in boldface. For two positive sequences $\{f_n\}$ and $\{g_n\}$, we

write $f_n \doteq g_n$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, and we write $f_n \dot{\leq} g_n$ if $\lim\sup_{n\to\infty} \frac{1}{n} \log \frac{f_n}{g_n} \leq 0$.

The type of a sequence $\boldsymbol{x} = (x_1,...,x_n) \in \mathcal{X}^n$ is its empirical distribution, defined by $\hat{P}_{\boldsymbol{x}}(x) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{x_i = x\}$. The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, while $\mathcal{P}_n(\mathcal{X})$ represents the set of empirical distributions for vectors in $\mathcal{X}^n$. For $P_X \in \mathcal{P}_n(\mathcal{X})$, the type class $\mathcal{T}^n(P_X)$ consists of all sequences in $\mathcal{X}^n$ with type P_X. We denote by N_k the number of types in $\mathcal{V}^k$, which grows polynomially with k [10, Lemma 2.2].

Continuing the line of [9], [11], we consider partitioning the source sequences into m disjoint sets, each consisting of one or more full source-type classes. Each class c is assigned a codeword distribution $Q_c \in \mathcal{P}(\mathcal{X})$, from which the codewords for source sequences in that class are generated i.i.d. according to Q_c. This setting includes Csiszár's partitioning [1], [7], in which each class corresponds to a single source-type class (i.e., $m = N_k$). Throughout, we use the indices i and j to refer to source types, and c to indicate the class to which a source message (or a source type) belongs under a given partitioning. Moreover, for a source type P_i, we define $R_i \triangleq tH(P_i)$.

III. ACHIEVABLE EXPURGATED ERROR EXPONENTS

In this section, we present our main results. We first derive an expurgated exponent for an arbitrary partitioning, and then consider a special case of partitioning into only two classes.

Lemma 1. *Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ be a given partition of the source sequences, with each class assigned an i.i.d. codeword distribution from $\mathcal{Q}_m = \{Q_1, \ldots, Q_m\}$. Then, there exists a sequence of codebooks such that, for all choices of $\{\rho_c\}_{c=1}^m$ with $\rho_c \geq 1$,*

$$p_e \dot{\leq} \sum_{c=1}^m e^{-n E_{\mathrm{x}}'\left(Q_c, \mathcal{Q}_m, \rho_c\right) + E_s^{(c)}\left(\rho_c, P^k\right)}, \qquad (3)$$

where

$$E_{\mathrm{x}}'\left(Q, \mathcal{Q}, \rho\right) \triangleq \min_{\tilde{Q} \in \mathcal{Q}} -\rho \log \sum_{(x,\bar{x})} Q(x)\tilde{Q}(\bar{x}) e^{-\frac{d_B(x,\bar{x})}{\rho}}, \quad (4)$$

and

$$E_s^{(c)}\left(\rho, P^k\right) \triangleq \log \left(\sum_{\boldsymbol{v} \in \mathcal{A}_c} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \qquad (5)$$

with $d_B(x,\bar{x}) \triangleq -\log \sum_y \sqrt{W(y|x)W(y|\bar{x})}$ denoting the Bhattacharyya distance.

A constant-composition counterpart of this result is given in [9, Corollary 1]. The improvement in Lemma 1, however, is that the minimization in (4) is restricted to the given set of codeword distributions, whereas in [9, Corollary 1] (equivalently [7]) it is taken over all possible distributions.

Theorem 1. *For a pair of i.i.d. distributions $\mathcal{Q} = \{Q, Q'\}$, there exists a partition of the source message set into two classes such that the following exponent is achievable*

$$E_{J,\mathrm{ex}}^{\mathrm{mc}}(\mathcal{Q}, t) \triangleq \sup_{\rho \geq 1} \left\{ \overline{E}_{\mathrm{x}}'\left(\mathcal{Q}, \rho\right) - tE_s\left(\rho, P_V\right) \right\}, \qquad (6)$$

where $\overline{E}_{\mathrm{x}}'(\mathcal{Q}, \rho)$ denotes the concave hull of $E_{\mathrm{x}}'(\mathcal{Q}, \rho)$ over $\rho \in [1, \infty)$, with $E_{\mathrm{x}}'(\mathcal{Q}, \rho) \triangleq \max_{Q \in \mathcal{Q}} E_{\mathrm{x}}'(Q, \mathcal{Q}, \rho)$, and $E_s(\rho, P_V) \triangleq \log \left(\sum_{v \in \mathcal{V}} P_V(v)^{\frac{1}{1+\rho}} \right)^{1+\rho}$.

The partitioning that achieves this exponent is the same as in [12, Theorem 2], and is given by

$$\mathcal{A}_1 = \left\{ \boldsymbol{v} \in \mathcal{V}^k : P^k(\boldsymbol{v}) \leq \gamma^k \right\} \qquad (7)$$

$$\mathcal{A}_2 = \left\{ \boldsymbol{v} \in \mathcal{V}^k : P^k(\boldsymbol{v}) > \gamma^k \right\}, \qquad (8)$$

for some $\gamma \in [0,1]$. Optimizing over the pair of distributions, the optimal expurgated exponent under this partitioning is

$$E_{J,\mathrm{ex}}^{\mathrm{mc}}(t) \triangleq \max_{\{Q,Q'\}} E_{J,\mathrm{ex}}^{\mathrm{mc}}(\{Q, Q'\}, t). \qquad (9)$$

IV. COMPARISON WITH SINGLE-CLASS CODING

We first present the achievable exponent under single-class coding, closely following [9, Corollary 4].

Theorem 2. *There exist single-class codes with i.i.d. ensembles that achieve the following exponent*

$$E_{J,\mathrm{ex}}^{\mathrm{sc}}(t) \triangleq \sup_{\rho \geq 1} \left\{ E_{\mathrm{x}}(\rho) - tE_s\left(\rho, P_V\right) \right\}, \qquad (10)$$

where $E_{\mathrm{x}}(\rho) \triangleq \max_Q E_{\mathrm{x}}(Q, \rho)$, and

$$E_{\mathrm{x}}(Q, \rho) = -\rho \log \sum_{(x,\bar{x})} Q(x)Q(\bar{x}) e^{-\frac{d_B(x,\bar{x})}{\rho}}. \qquad (11)$$

We use the notation $E_{J,\mathrm{ex}}^{\mathrm{sc}}(Q, t)$ to denote the exponent corresponding to a fixed codeword distribution Q. A key observation is that this result can be viewed as a special case of the two-class partitioning when $Q = Q' = Q^*$, with Q^* denoting the optimal distribution in Theorem 2. This implicitly implies that $E_{J,\mathrm{ex}}^{\mathrm{mc}}(t) \geq E_{J,\mathrm{ex}}^{\mathrm{sc}}(t)$, as stated in the following theorem.

Theorem 3. *The optimal expurgated exponent with two-class partitioning is no smaller than that of optimal single-class coding, namely*

$$E_{J,\mathrm{ex}}^{\mathrm{mc}}(t) \geq E_{J,\mathrm{ex}}^{\mathrm{sc}}(t). \qquad (12)$$

When the optimal distribution is chosen, the i.i.d. and constant-composition channel-coding expurgated exponents coincide, as shown in [10, Prob. 10.24]. Hence, $E_{J,\mathrm{ex}}^{\mathrm{sc}}(t)$ also equals the optimal single-class expurgated exponent under constant composition.

We therefore establish the existence of a two-class partitioning that, under i.i.d. coding, achieves an exponent no smaller than that of optimal single-class coding with constant composition. This gives a one-sided comparison between partitioning and non-partitioning in expurgated joint source-channel coding, a result that, to the best of our knowledge, has not appeared previously.

A. Numerical Examples

In this section, we provide numerical examples of expurgated error exponents in joint source-channel coding, illustrating in particular the impact of partitioning. We consider a source-channel pair consisting of a binary memoryless source, and a non-symmetric memoryless channel of the form

$$W = \begin{pmatrix} 1 - 2\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 1 - 2\varepsilon & \varepsilon \\ \delta & \delta & 1 - 2\delta \end{pmatrix}, \tag{13}$$

with the parameters $\varepsilon = 0.0001$ and $\delta = 0.1$.

For each case, we are given two codeword distributions $\{Q_1, Q_2\}$, and we observe how the two-class partitioning defined in (7)-(8) for these distributions compares with single-class coding using one of them. In each figure, we plot the functions $\overline{E}'_{\mathrm{x}}(\{Q_1, Q_2\}, \rho) - tE_s(\rho, P_V)$ and $E_{\mathrm{x}}(Q_c, \rho) - tE_s(\rho, P_V)$ for $c = 1, 2$. For reference, the values $E^{\mathrm{mc}}_{J,\mathrm{ex}}(\{Q_1, Q_2\}, t)$ and $\max_{c=1,2} E^{\mathrm{sc}}_{J,\mathrm{ex}}(Q_c, t)$ are shown as horizontal solid lines, while the JSCC random-coding exponent [12, Eq. (20)] is plotted with dot-dashed lines. In all three cases, the expurgated exponents exceed the random-coding exponent, although this need not hold in general. Depending on the setup, the random-coding exponent can be strictly larger and even tight, coinciding with the sphere-packing bound, as discussed in [1], [3].

The first figure illustrates a case with non-optimal codeword distributions, where the partitioned exponent is strictly higher than the two single-class exponents.
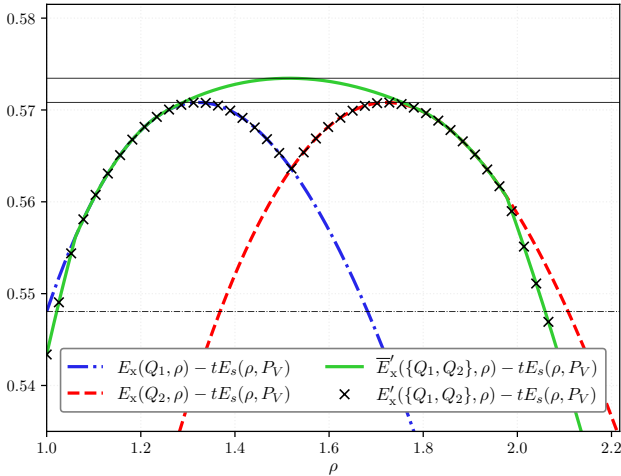

Fig. 1. Example where partitioning strictly improves the exponent, obtained with $Q_1 = (0.4, 0.4, 0.2)$, and $Q_2 = (0.5, 0.5, 0)$, $t = 0.75$ and $P_V(0) = 0.025$.

The second plot, on the other hand, illustrates a case where partitioning weakens the achievable exponent, with the codeword distributions still being non-optimal.
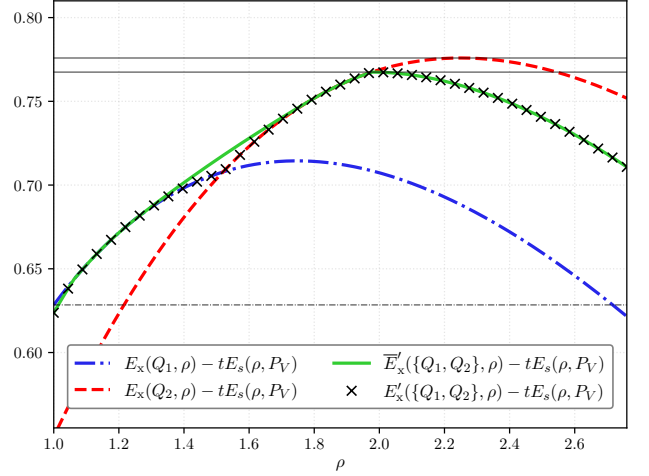

Fig. 2. Example where single-class coding yields a higher exponent, with the same setting as Fig. 1 except $t = 0.5$ and $P_V(0) = 0.02$.

Figures 1 and 2 show that when the codeword distributions are fixed and not optimized, it is not predetermined which exponent dominates; either one can exceed the other. In the final example, Q_1 and Q_2 are chosen optimally, showing that in this particular case the two exponents coincide.
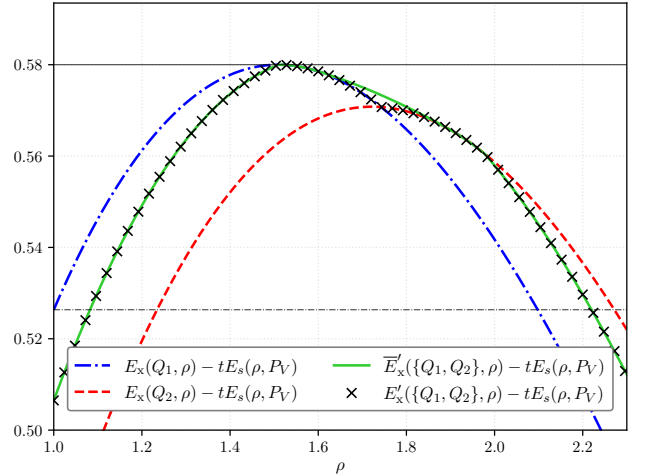

Fig. 3. The two exponents coincide under optimal distributions, with the same setting as Fig. 1 except $Q_1 = (0.4489, 0.4489, 0.1021)$ and $Q_2 = (0.5, 0.5, 0)$.

To date, we have not found any joint source-channel coding setting in which optimal partitioning yields a strictly higher exponent. In cases examined, the two exponents coincide, raising the possibility that they may always be equal. Observe that for a gap to exist, the input alphabet of the channel must have size at least three. This is because when the input alphabet size is two, it has been shown [13] that the uniform distribution $Q(0) = Q(1) = 1/2$ always maximizes $E_{\mathrm{x}}(Q, \rho)$, regardless of the value of ρ. Consequently, for binary-input channels, partitioning does not provide any improvement to the exponent.

V. Proofs

The work of [8] revisits expurgation in joint source-channel coding and rederives Csiszár's two exponents [7] using a new technique based on source-type duplication and type-by-type expurgation. The key result of [8], which forms the basis for all achievability results in this paper, is summarized in the following lemma.

Lemma 2. *There exists a joint source-channel code $\mathcal{C}$ such that for any source type $P_i \in \mathcal{P}_k(\mathcal{V})$ and any sequence $\boldsymbol{v} \in \mathcal{T}^k(P_i)$, the source message error probability satisfies*

$$p_e(\boldsymbol{v}, \mathcal{C}) \doteq \sum_{j=1}^{N_k} \mathbb{E}\Big[p_e\big(\boldsymbol{v}, j, \mathsf{C}\big)^{\frac{1}{\rho_{ij}}}\Big]^{\rho_{ij}}, \qquad (14)$$

where $\rho_{ij} > 0$, and $p_e(\boldsymbol{v}, j, \mathsf{C})$ refers to the probability that, given $\boldsymbol{V} = \boldsymbol{v}$, there exists some $\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)$ that yields a decoding metric at least as high as that of $\boldsymbol{v}$. Moreover, the expectation is taken with respect to the given (not necessarily distinct) codeword distributions $\{P_{\boldsymbol{X}}^{(j)}\}_{j=1}^{N_k}$.

A. Proof of Lemma 1

The proof closely follows the approach in [9, Lemma 3] and [8, Sec. III.C]. We consider the code from Lemma 2, and use the following decoding metric

$$q(\boldsymbol{v}, \boldsymbol{x}, \boldsymbol{y}) = W^n(\boldsymbol{y}|\boldsymbol{x}) \exp\big(-2kD\left(P_i\|P_V\right)\big), \qquad (15)$$

if $\boldsymbol{v} \in \mathcal{T}^k(P_i)$. Observe that this is a type-dependent metric, meaning that it depends on $(\boldsymbol{v}, \boldsymbol{x}, \boldsymbol{y})$ only through their types. Then, for any $s > 0$, we can upper bound $p_e(\boldsymbol{v}, j, \mathsf{C})$ as

$$p_e(\boldsymbol{v}, j, \mathsf{C}) = \mathbb{P}\left[\bigcup_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} q(\bar{\boldsymbol{v}}, \boldsymbol{x}_{\bar{\boldsymbol{v}}}, \boldsymbol{Y}) \geq q(\boldsymbol{v}, \boldsymbol{x}_{\boldsymbol{v}}, \boldsymbol{Y})\right] \quad (16)$$

$$= \mathbb{P}\left[\bigcup_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} q(\bar{\boldsymbol{v}}, \boldsymbol{x}_{\bar{\boldsymbol{v}}}, \boldsymbol{Y})^s \geq q(\boldsymbol{v}, \boldsymbol{x}_{\boldsymbol{v}}, \boldsymbol{Y})^s\right] \quad (17)$$

$$\leq \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\boldsymbol{x}_{\boldsymbol{v}}) \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \left(\frac{q(\bar{\boldsymbol{v}}, \boldsymbol{x}_{\bar{\boldsymbol{v}}}, \boldsymbol{y})}{q(\boldsymbol{v}, \boldsymbol{x}_{\boldsymbol{v}}, \boldsymbol{y})}\right)^s. \quad (18)$$

Substituting the decoding metric and setting $s = \frac{1}{2}$, we obtain

$$p_e(\boldsymbol{v}, j, \mathsf{C}) \leq \left(\frac{\exp\left(-kD\left(P_j\|P_V\right)\right)}{\exp\left(-kD\left(P_i\|P_V\right)\right)}\right) \\ \times \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \sum_{\boldsymbol{y}} \sqrt{W^n(\boldsymbol{y}|\boldsymbol{x}_{\boldsymbol{v}})W^n(\boldsymbol{y}|\boldsymbol{x}_{\bar{\boldsymbol{v}}})}. \quad (19)$$

If we assume $\rho_{ij} \geq 1$ for all (i,j), we can use Hölder's inequality [2, Problem 4.15.f] to move the summation over

$\bar{\boldsymbol{v}}$ outside, as follows

$$p_e(\boldsymbol{v}, j, \mathsf{C})^{\frac{1}{\rho_{ij}}} \leq \left[\left(\frac{\exp\left(-kD\left(P_j\|P_V\right)\right)}{\exp\left(-kD\left(P_i\|P_V\right)\right)}\right) \\ \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \sum_{\boldsymbol{y}} \sqrt{W^n(\boldsymbol{y}|\boldsymbol{x}_{\boldsymbol{v}})W^n(\boldsymbol{y}|\boldsymbol{x}_{\bar{\boldsymbol{v}}})}\right]^{\frac{1}{\rho_{ij}}} \quad (20)$$

$$\leq \left(\frac{\exp(-kD(P_j\|P_V))}{\exp(-kD(P_i\|P_V))}\right)^{\frac{1}{\rho_{ij}}} \\ \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \prod_{l=1}^{n} \left[\sum_{y} \sqrt{W(y|x_{\boldsymbol{v}}^l)W(y|x_{\bar{\boldsymbol{v}}}^l)}\right]^{\frac{1}{\rho_{ij}}}, \quad (21)$$

where $x_{\boldsymbol{v}}^l$ denotes the l-th element of $\boldsymbol{x}_{\boldsymbol{v}}$. We now take the expectation of the above expression

$$\mathbb{E}\left[p_e(\boldsymbol{v}, j, \mathsf{C})^{\frac{1}{\rho_{ij}}}\right] \leq \left(\frac{\exp(-kD(P_j\|P_V))}{\exp(-kD(P_i\|P_V))}\right)^{\frac{1}{\rho_{ij}}} \\ \mathbb{E}\left[\sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \prod_{l=1}^{n} e^{-\frac{d_B(X_l, \bar{X}_l)}{\rho_{ij}}}\right], \quad (22)$$

where the expectation is taken with respect to $P_{\boldsymbol{X}}^{(\hat{i})} \times P_{\boldsymbol{X}}^{(\hat{j})}$, with $P_{\boldsymbol{X}}^{(c)}$ denoting Q_c^n. Here, $\hat{i}$ denotes the index of the class to which source type i belongs. We can evaluate this expectation as follows

$$\mathbb{E}\left[\sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \prod_{l=1}^{n} e^{-\frac{d_B(X_l, \bar{X}_l)}{\rho_{ij}}}\right] \\ = \sum_{(\boldsymbol{x}, \bar{\boldsymbol{x}})} Q_{\hat{i}}^n(\boldsymbol{x}) Q_{\hat{j}}^n(\bar{\boldsymbol{x}}) \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \prod_{l=1}^{n} e^{-\frac{d_B(x_l, \bar{x}_l)}{\rho_{ij}}} \quad (23)$$

$$= \sum_{\bar{\boldsymbol{v}} \in \mathcal{T}^k(P_j)} \prod_{l=1}^{n} \sum_{(x, \bar{x})} Q_{\hat{i}}(x) Q_{\hat{j}}(\bar{x}) e^{-\frac{d_B(x, \bar{x})}{\rho_{ij}}} \quad (24)$$

$$\doteq e^{nR_j} \left[\sum_{(x, \bar{x})} Q_{\hat{i}}(x) Q_{\hat{j}}(\bar{x}) e^{-\frac{d_B(x, \bar{x})}{\rho_{ij}}}\right]^n. \quad (25)$$

Raising both to the power ρ_{ij}, and then taking the worst-case $Q_{\hat{i}}$ over $\mathcal{Q}_m$ follows that

$$\mathbb{E}\left[p_e(\boldsymbol{v}, j, \mathsf{C})^{\frac{1}{\rho_{ij}}}\right]^{\rho_{ij}} \leq \left(\frac{\exp(-kD(P_j\|P_V))}{\exp(-kD(P_i\|P_V))}\right) \\ \exp\left(n\left[\rho_{ij}R_j - E_{\mathrm{x}}'(Q_{\hat{j}}, \mathcal{Q}_m, \rho_{ij})\right]\right). \quad (26)$$

Having evaluated the expectation, we now obtain the following bound on the error probability of the code in Lemma 2 as

follows

$$p_e(\mathcal{C}) = \sum_{\boldsymbol{v}} P^k(\boldsymbol{v})\, p_e(\boldsymbol{v}, \mathcal{C}) \tag{27}$$

$$\dot{\leq} \sum_{(i,j)} \sum_{\boldsymbol{v} \in \mathcal{T}^k(P_i)} P^k(\boldsymbol{v}) \left(\frac{\exp(-kD(P_j\|P_V))}{\exp(-kD(P_i\|P_V))} \right) \tag{28}$$

$$\times \exp\left(-n\left[E'_{\mathrm{x}}(Q_{\hat{j}}, \mathcal{Q}_m, \rho_{ij}) - \rho_{ij}R_j \right] \right).$$

Observe that $\sum_{\boldsymbol{v} \in \mathcal{T}^k(P_i)} P^k(\boldsymbol{v}) \doteq \exp(-kD(P_i\|P_V))$. This implies that

$$\sum_{\boldsymbol{v} \in \mathcal{T}^k(P_i)} P^k(\boldsymbol{v}) \left(\frac{\exp(-kD(P_j\|P_V))}{\exp(-kD(P_i\|P_V))} \right) \doteq \sum_{\boldsymbol{v} \in \mathcal{T}^k(P_j)} P^k(\boldsymbol{v}). \tag{29}$$

Moreover, since $\prod_{l=1}^{k} \hat{P}_{\boldsymbol{v}}(v_l) = \exp(-nR_j)$ for $\boldsymbol{v} \in \mathcal{T}^k(P_j)$, we obtain

$$p_e(\mathcal{C}) \dot{\leq} \sum_{(i,j)=1}^{N_k} \sum_{\boldsymbol{v} \in \mathcal{T}^k(P_j)} \frac{P^k(\boldsymbol{v})}{\hat{P}_{\boldsymbol{v}}(\boldsymbol{v})^{\rho_{ij}}} \, e^{-nE'_{\mathrm{x}}(Q_{\hat{j}}, \mathcal{Q}_m, \rho_{ij})}. \tag{30}$$

Let us set $\rho_{ij} = \rho_{\hat{j}}$ for all i. It then follows that

$$p_e(\mathcal{C}) \dot{\leq} \sum_{j=1}^{N_k} \sum_{\boldsymbol{v} \in \mathcal{T}^k(P_j)} \frac{P^k(\boldsymbol{v})}{\hat{P}_{\boldsymbol{v}}(\boldsymbol{v})^{\rho_{\hat{j}}}} \, e^{-nE'_{\mathrm{x}}(Q_{\hat{j}}, \mathcal{Q}_m, \rho_{\hat{j}})} \tag{31}$$

$$= \sum_{c=1}^{m} \left[\sum_{\boldsymbol{v} \in \mathcal{A}_c} \frac{P^k(\boldsymbol{v})}{\hat{P}_{\boldsymbol{v}}(\boldsymbol{v})^{\rho_c}} \right] e^{-nE'_{\mathrm{x}}(Q_c, \mathcal{Q}_m, \rho_c)}. \tag{32}$$

It was shown in [9, Corollary 1] that

$$\sum_{\boldsymbol{v} \in \mathcal{A}_c} \frac{P^k(\boldsymbol{v})}{\hat{P}_{\boldsymbol{v}}(\boldsymbol{v})^{\rho}} \dot{\leq} \left[\sum_{\boldsymbol{v} \in \mathcal{A}_c} P^k(\boldsymbol{v})^{\frac{1}{1+\rho}} \right]^{1+\rho}. \tag{33}$$

Thus, we have

$$p_e \dot{\leq} \sum_{c=1}^{m} e^{-nE'_{\mathrm{x}}(Q_c, \mathcal{Q}_m, \rho_c) + E_s^{(c)}(\rho_c, P^k)}. \tag{34}$$

B. Proof of Theorem 1

To prove Theorem 1, we first state the following auxiliary lemma, given in [12, Lemma 1].

Lemma 3. *For any $\rho_0 \geq 1$ and $\gamma' \geq 0$, the partition (7)-(8) with $\gamma = \min\{1, \gamma'\}$ satisfies*

$$\frac{1}{k} E_s^{(1)}(\rho, P^k) \tag{35}$$
$$\leq E_s(\rho, P_V)\mathbb{1}\{\rho > \rho_0\} + r(\rho, \rho_0, \gamma')\mathbb{1}\{\rho \leq \rho_0\}$$
$$\triangleq \bar{E}_s^{(1)}(\rho, \rho_0, \gamma'), \tag{36}$$

$$\frac{1}{k} E_s^{(2)}(\rho, P^k) \tag{37}$$
$$\leq E_s(\rho, P_V)\mathbb{1}\{\rho < \rho_0\} + r(\rho, \rho_0, \gamma')\mathbb{1}\{\rho \geq \rho_0\}$$
$$\triangleq \bar{E}_s^{(2)}(\rho, \rho_0, \gamma'), \tag{38}$$

where

$$r(\rho, \rho_0, \gamma) \triangleq E_s(\rho_0, P_V) + \frac{E_s(\rho_0, P_V) - \log\gamma}{1+\rho_0}(\rho - \rho_0). \tag{39}$$

Using this lemma, for any $\gamma' \geq 0$ the achievable error exponent of Lemma 1 can be evaluated as

$$E_n = \liminf_{n\to\infty} -\frac{1}{n}\log\left(\sum_{c=1,2} \exp\left(-\sup_{\rho_c \geq 1}\left[nE'_{\mathrm{x}}(Q_c, \mathcal{Q}, \rho_c) \right.\right.\right.$$
$$\left.\left.\left. - E_s^{(c)}(\rho_c, P^k) \right] \right) \right) \tag{40}$$

$$= \liminf_{n\to\infty}\left\{ \min_{c=1,2}\left\{ \sup_{\rho_c \geq 1}\left\{ E'_{\mathrm{x}}(Q_c, \mathcal{Q}, \rho_c) \right.\right.\right.$$
$$\left.\left.\left. - \frac{1}{n}E_s^{(c)}(\rho_c, P^k) \right\}\right\}\right\} \tag{41}$$

$$\geq \min_{c=1,2}\left\{ \sup_{\rho_0, \rho_1, \rho_2 \geq 1}\left\{ E'_{\mathrm{x}}(Q_c, \mathcal{Q}, \rho_c) \right.\right.$$
$$\left.\left. - t\bar{E}_s^{(c)}(\rho_c, \rho_0, \gamma') \right\}\right\}, \tag{42}$$

where in (42) we applied Lemma 3 and used the fact that $\liminf_{n\to\infty} \max_x\{f_n(x)\} \geq \max_x\{\lim_{n\to\infty} f_n(x)\}$. Optimizing over $\gamma' \geq 0$, swapping the order of min and max, and restricting the maximization range yields

$$E_n \geq \sup_{\substack{\rho_0, \rho_1, \rho_2 \geq 1 \\ \rho_1 \leq \rho_0 \leq \rho_2}} \sup_{\gamma' \geq 0}\left\{ \min_{c=1,2}\left\{ E'_{\mathrm{x}}(Q_c, \mathcal{Q}, \rho_c) \right.\right.$$
$$\left.\left. - t\bar{E}_s^{(c)}(\rho_c, \rho_0, \gamma') \right\}\right\}. \tag{43}$$

By substituting the expression of $\bar{E}_s^{(c)}(\rho_c, \rho_0, \gamma')$, the minimization becomes

$$\min_{c=1,2}\left\{ E'_{\mathrm{x}}(Q_c, \mathcal{Q}, \rho_c) \right.$$
$$\left. + t\frac{E_s(\rho_0, P_V) - \log\gamma'}{1+\rho_0}(\rho_0 - \rho_c) - tE_s(\rho_0, P_V) \right\}. \tag{44}$$

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

$$t\frac{E_s(\rho_0, P_V) - \log\gamma_0}{1+\rho_0} = \frac{E'_{\mathrm{x}}(Q_2, \mathcal{Q}, \rho_2) - E'_{\mathrm{x}}(Q_1, \mathcal{Q}, \rho_1)}{\rho_2 - \rho_1}. \tag{45}$$

Choosing $\gamma' = \gamma_0$ equalizes the two terms in the minimization in (44), thereby maximizing the lower bound in (43). Consequently, we obtain

$$E_n \geq \sup_{\substack{\rho_0, \rho_1, \rho_2 \geq 1 \\ \rho_1 \leq \rho_0 \leq \rho_2}}\left\{ \left\{ \frac{\rho_2 - \rho_0}{\rho_2 - \rho_1}E'_{\mathrm{x}}(Q_1, \mathcal{Q}, \rho_1) \right.\right.$$
$$\left.\left. + \frac{\rho_0 - \rho_1}{\rho_2 - \rho_1}E'_{\mathrm{x}}(Q_2, \mathcal{Q}, \rho_2) \right\} - tE_s(\rho_0, P_V) \right\} \tag{46}$$

$$= \sup_{\rho_0 \geq 1} \left\{ \sup_{\substack{\rho_1, \rho_2 \geq 1 \\ \rho_1 \leq \rho_0 \leq \rho_2}} \left\{ \frac{\rho_2 - \rho_0}{\rho_2 - \rho_1} E'_x(Q_1, \mathcal{Q}, \rho_1) \right. \right.$$
$$\left. \left. + \frac{\rho_0 - \rho_1}{\rho_2 - \rho_1} E'_x(Q_2, \mathcal{Q}, \rho_2) \right\} - t E_s(\rho_0, P_V) \right\}. \tag{47}$$

We can further optimize the expression by assigning the distributions to the classes in the best possible way, obtaining

$$E_n \geq \sup_{\rho_0 \geq 1} \left\{ \sup_{\substack{\rho_1, \rho_2 \geq 1 \\ \rho_1 \leq \rho_0 \leq \rho_2}} \left\{ \frac{\rho_2 - \rho_0}{\rho_2 - \rho_1} E'_x(\mathcal{Q}, \rho_1) \right. \right.$$
$$\left. \left. + \frac{\rho_0 - \rho_1}{\rho_2 - \rho_1} E'_x(\mathcal{Q}, \rho_2) \right\} - t E_s(\rho_0, P_V) \right\}. \tag{48}$$

Let $\lambda \in [0,1]$ be defined such that $\lambda \rho_1 + (1 - \lambda) \rho_2 = \rho_0$, yielding

$$E_n \geq \sup_{\rho_0 \geq 1} \left\{ \sup_{\substack{\rho_1, \rho_2 \geq 1, \lambda \in [0,1] \\ \lambda \rho_1 + (1-\lambda)\rho_2 = \rho_0}} \left\{ \lambda E'_x(\mathcal{Q}, \rho_1) \right. \right.$$
$$\left. \left. + (1 - \lambda) E'_x(\mathcal{Q}, \rho_2) \right\} - t E_s(\rho_0, P_V) \right\} \tag{49}$$
$$= \sup_{\rho_0 \geq 1} \left\{ \overline{E}'_x(\mathcal{Q}, \rho_0) - t E_s(\rho_0, P_V) \right\}, \tag{50}$$

where

$$\overline{E}'_x(\mathcal{Q}, \rho) \triangleq \sup_{\substack{\rho_1, \rho_2 \geq 1, \lambda \in [0,1] \\ \lambda \rho_1 + (1-\lambda)\rho_2 = \rho}} \lambda E'_x(\mathcal{Q}, \rho_1) + (1 - \lambda) E'_x(\mathcal{Q}, \rho_2). \tag{51}$$

This corresponds to the pointwise supremum over convex combinations of any two values of the function $E'_x(\mathcal{Q}, \rho)$, which, as described in [14], is equivalent to its concave hull. This proof closely follows the argument of [12, Theorem 2].

C. Proof of Theorem 2

The proof follows the same steps as [9, Corollary 4], with the main difference being that i.i.d. coding ensembles are used instead of constant-composition.

D. Proof of Theorem 3

Let Q^* denote the optimal distribution for $E^{sc}_{J,ex}(t)$, namely the solution to

$$E^{sc}_{J,ex}(t) = \max_Q \sup_{\rho \geq 1} \left\{ E_x(Q, \rho) - t E_s(\rho, P_V) \right\}. \tag{52}$$

We can lower bound $E^{mc}_{J,ex}(t)$ by choosing the maximization pair to be the (potentially suboptimal) choice $\{Q^*, Q^*\}$, as follows

$$E^{mc}_{J,ex}(t) = \max_{\{Q, Q'\}} E^{mc}_{J,ex}(\{Q, Q'\}, t) \tag{53}$$
$$\geq E^{mc}_{J,ex}(\{Q^*, Q^*\}, t) \tag{54}$$
$$= \sup_{\rho \geq 1} \left\{ \overline{E}'_x(\{Q^*, Q^*\}, \rho) - t E_s(\rho, P_V) \right\} \tag{55}$$
$$= \sup_{\rho \geq 1} \left\{ E'_x(\{Q^*, Q^*\}, \rho) - t E_s(\rho, P_V) \right\} \tag{56}$$
$$= \sup_{\rho \geq 1} \left\{ E_x(Q^*, \rho) - t E_s(\rho, P_V) \right\} \tag{57}$$
$$= E^{sc}_{J,ex}(t). \tag{58}$$

Equality in (56) holds because for fixed Q, the function $E_x(Q, \rho)$ is concave in ρ, so $E'_x(\{Q^*, Q^*\}, \rho)$ is concave as well, and hence equal to its concave hull.

REFERENCES

[1] I. Csiszár, "Joint source-channel error exponent," *Problems of Control and Information Theory*, vol. 9, pp. 315–328, 1980.

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

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

[4] I. Csiszar and J. Korner, "Graph decomposition: A new key to coding theorems," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 5–12, 1981.

[5] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of discrete memoryless channels," in *the IEEE Int. Symp. Information Theory, Cornell Univ.*, 1977.

[6] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements, and connections," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4449–4462, 2014.

[7] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 823–828, 1982.

[8] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Expurgated joint source-channel coding bounds and error exponents," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 908–912.

[9] S. A. Moeini and A. Guillén i Fàbregas, "Class-based expurgation attains csiszár's expurgated source-channel exponent," in *2025 IEEE Information Theory Workshop (ITW)*, Sydney, Australia, Sep-Oct 2025. [Online]. Available: https://arxiv.org/abs/2507.03481

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

[11] S. A. Moeini and A. Guillén i Fàbregas, "Dual-domain exponent of maximum mutual information decoding," in *2025 IEEE Information Theory Workshop (ITW)*, Sydney, Australia, Sep-Oct 2025. [Online]. Available: https://arxiv.org/abs/2501.13724

[12] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martinez, "A derivation of the source-channel error exponent using nonidentical product distributions," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3209–3217, 2014.

[13] F. Jelinek, "Evaluation of expurgated bound exponents," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 501–505, 1968.

[14] R. T. Rockafellar, *Convex analysis*, ser. Princeton Mathematical Series. Princeton, N. J.: Princeton University Press, 1970.