

Random Gilbert-Varshamov Codes for Joint Source-Channel Coding

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—We propose a random coding technique for joint source-channel coding of discrete memoryless sources and channels. The approach builds on the random Gilbert-Varshamov code construction of Somekh-Baruch *et al.* and extends it to the joint source-channel setting. We show that the resulting ensemble attains the maximum of the random-coding and expurgated error exponents.

I. INTRODUCTION

We study joint source-channel coding (JSCC) for the transmission of non-equiprobable messages generated from a discrete memoryless source with distribution $P^k(\mathbf{v}) = \prod_{i=1}^k P_V(v_i)$, where $\mathbf{v} = (v_1, \dots, v_k) \in \mathcal{V}^k$ denotes the source message and \mathcal{V} is a finite alphabet. The channel is a discrete memoryless channel characterized by $W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n W(y_i|x_i)$, where $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{Y}^n$ are the channel input and output sequences, respectively, and \mathcal{X} and \mathcal{Y} are discrete alphabets.

In JSCC, an encoder maps each length- k source message \mathbf{v} to a length- n codeword \mathbf{x}_v , which is then transmitted over the channel. Based on the channel output \mathbf{y} the decoder estimates which source message was transmitted. We refer to $t \triangleq k/n$ as the transmission rate. An error exponent $E > 0$ is said to be *achievable* if there exists a sequence of codes of blocklength n such that the error probability p_e satisfies

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

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

Two classical achievable error exponents in channel coding are the random-coding and the expurgated exponents. In [1], Csiszár and Körner proved that there exists a code that attains the maximum of these two exponents. More recently, [2] introduced a random-coding construction in which codewords are generated sequentially from a fixed type class, subject to a minimum-distance constraint with respect to previously generated codewords, and showed that the resulting ensemble achieves an exponent that is no smaller than that of [1].

The study of error exponents for JSCC was initiated by Gallager [3, Prob. 5.16] and Csiszár [4]. In [5], Csiszár

introduced two achievable expurgated error exponents, but was unable to compare them. In [6], we compared the two exponents and showed that one is a particular case of the other. In this work we refer to the expurgated exponent as the most general of Csiszár's two error exponents. To the best of our knowledge, no JSCC random coding construction has previously been shown to attain the maximum of the random-coding and expurgated exponents. In this paper, building on the idea of [2], we extend the recursive distance-based construction to the JSCC setting. We show that the resulting ensemble simultaneously achieves both the random-coding and expurgated error exponents.

Notation: Scalar random variables are denoted by uppercase letters, their realizations by lowercase letters, and their alphabets by calligraphic letters. Random vectors are written in boldface. We use the notation $[a]$ to denote all integers from 0 to a . For two positive sequences $\{f_n\}$ and $\{g_n\}$, we write $f_n \doteq g_n$ if $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, and we write $f_n \leq g_n$ if $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} \leq 0$. The type of a sequence $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ is its empirical distribution, defined by $\hat{P}_{\mathbf{x}}(x) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i = x\}$. The set of types for vectors in \mathcal{X}^n is denoted by $\mathcal{P}_n(\mathcal{X})$. For $Q \in \mathcal{P}_n(\mathcal{X})$, the corresponding type class $\mathcal{T}^n(Q)$ is the set of all sequences in \mathcal{X}^n with type Q . For a joint distribution P on $\mathcal{X} \times \mathcal{Y}$, we write $I(P)$, or equivalently $I_P(X; Y)$, for the mutual information between X and Y under P . Finally, N_k denotes the number of types in \mathcal{V}^k . The analysis in this paper is largely based on the method of types [7, Ch. 2].

A. Preliminaries

Following [8], we assume a finite collection of codeword distributions $\mathcal{Q}_m = \{Q_1, \dots, Q_m\}$. Each source type class $\mathcal{T}^k(P_i)$ is assigned a distribution $Q_{\mu(i)} \in \mathcal{Q}_m$, and the corresponding codewords are generated from $\mathcal{T}^n(Q_{\mu(i)})$ according to the proposed random construction. Here, $\mu(i) \in [m]$ denotes the index of the codeword distribution associated with the i -th source type. This setting may be viewed as partitioning the source sequences into m random-coding classes in line with [9]. Moreover, for a source type P_i , we define $R_i \triangleq tH(P_i)$, since $|\mathcal{T}^k(P_i)| \doteq \exp(nR_i)$ [7, Lemma 2.3].

This work was supported in part by the European Research Council under Grant 101142747 and in part by the Spanish Government under Grant PID2020-116683GB-C22.

We now summarize existing results in JSCC that are relevant to this work. The average error probability \bar{p}_e of the constant-composition ensemble satisfies [4], [10]

$$\bar{p}_e \leq \sum_{i=1}^{N_k} \exp\left(-n \left[te\left(\frac{R_i}{t}, P_V\right) + E_r(Q_{\mu(i)}, R_i) \right]\right), \quad (2)$$

where

$$e(R, P_V) \triangleq \min_{Q: H(Q) \geq R} D(Q \| P_V) \quad (3)$$

$$E_r(Q, R) \triangleq \min_{P_{XY}: P_X=Q} D(P \| Q \times W) + |I_P(X; Y) - R|^+ \quad (4)$$

are the source reliability function and random coding error exponent for channel coding, respectively [7]. This bound corresponds to the random-coding bound for JSCC. Moreover, it can be shown that there exists a code whose error probability satisfies [5], [6]

$$p_e \leq \sum_{i=1}^{N_k} \exp\left(-n \left[te\left(\frac{R_i}{t}, P_V\right) + E'_{\text{ex}}(Q_{\mu(i)}, Q_m, R_i) \right]\right) \quad (5)$$

where

$$E'_{\text{ex}}(Q, Q, R) \triangleq \min_{\substack{\Lambda_X, \Lambda_{\bar{X}} \in \mathcal{Q} \\ I_\Lambda(X; \bar{X}) \leq R}} \mathbb{E}_\Lambda [d_W(X, \bar{X})] + I(\Lambda) - R, \quad (6)$$

and $d_W(x, \bar{x}) := -\log \sum_y \sqrt{W(y|x)W(y|\bar{x})}$. This bound corresponds to the expurgated bound in JSCC.

II. RANDOM CODEBOOK GENERATION

The construction considered here closely follows that of [2], and is designed to ensure that the generated codewords satisfy certain prescribed minimum-distance constraints.

Let $d(\cdot, \cdot) : \mathcal{X}^n \times \mathcal{X}^n \rightarrow \mathbb{R}$ be a symmetric, type-dependent function that depends on $(\mathbf{x}, \mathbf{x}')$ only through their joint type. We therefore use $d(\mathbf{x}, \mathbf{x}')$ and $d(\hat{P}_{\mathbf{x}\mathbf{x}'})$ interchangeably throughout this work. Although we refer to $d(\cdot, \cdot)$ as a distance, it is not necessarily a proper metric.

Since the codeword distribution assigned to a source message depends on its type, we likewise allow the required minimum distance to depend on the source type. We therefore specify a collection of real numbers $\{\Delta_i\}_{i=1}^{N_k}$, not necessarily positive, which determine the required minimum distances. In particular, if $\mathbf{v} \in \mathcal{T}^k(P_i)$, then every other message $\bar{\mathbf{v}} \neq \mathbf{v}$ must satisfy $d(\mathbf{x}_\mathbf{v}, \mathbf{x}_{\bar{\mathbf{v}}}) > \Delta_i$. Consequently, if $\mathbf{v} \in \mathcal{T}^k(P_i)$ and $\bar{\mathbf{v}} \in \mathcal{T}^k(P_j)$, then $d(\mathbf{x}_\mathbf{v}, \mathbf{x}_{\bar{\mathbf{v}}}) > \Delta_i$ and $d(\mathbf{x}_{\bar{\mathbf{v}}}, \mathbf{x}_\mathbf{v}) > \Delta_j$. By symmetry of $d(\cdot, \cdot)$, these two conditions are equivalent to $d(\hat{P}_{\mathbf{x}_\mathbf{v}\mathbf{x}_{\bar{\mathbf{v}}}}) > \max(\Delta_i, \Delta_j)$. We fix an arbitrary ordering of the source types, and within each type class, an arbitrary ordering of the source messages. We denote the ℓ -th generated codeword of source type class $\mathcal{T}^k(P_i)$ by $\mathbf{x}_\ell^{(i)}$ ¹. We generate the codewords as follows:

¹We use \mathbf{x} to denote a codeword in two contexts: $\mathbf{x}_\mathbf{v}$ denotes the codeword corresponding to source sequence \mathbf{v} and $\mathbf{x}_\ell^{(i)}$ denotes the ℓ -th generated codeword of the i -th source type class.

- 1) Class $\mathcal{T}^k(P_1)$: Choose the initial codeword $\mathbf{x}_1^{(1)}$ equiprobably from $\mathcal{T}^n(Q_{\mu(1)})$. Each subsequent codeword $\mathbf{x}_\ell^{(1)}$ (for $\ell \geq 2$) is drawn equiprobably from the set

$$\left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(1)}) : d(\mathbf{x}_j^{(1)}, \bar{\mathbf{x}}) > \Delta_1 \text{ for } j \in [\ell - 1] \right\}. \quad (7)$$

This continues until all $|\mathcal{T}^k(P_1)|$ codewords have been generated, which we denote by \mathcal{C}_1 .

- 2) Class $\mathcal{T}^k(P_2)$: Choose the initial codeword $\mathbf{x}_1^{(2)}$ equiprobably from

$$\left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(2)}) : d(\mathbf{x}, \bar{\mathbf{x}}) > \max(\Delta_1, \Delta_2) \forall \mathbf{x} \in \mathcal{C}_1 \right\}. \quad (8)$$

For $\ell \geq 2$, the ℓ -th codeword is drawn equiprobably from

$$\left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(2)}) : d(\mathbf{x}_j^{(2)}, \bar{\mathbf{x}}) > \Delta_2 \text{ for } j \in [\ell - 1], \right. \\ \left. d(\mathbf{x}, \bar{\mathbf{x}}) > \max(\Delta_1, \Delta_2) \forall \mathbf{x} \in \mathcal{C}_1 \right\}. \quad (9)$$

This continues until all $|\mathcal{T}^k(P_2)|$ codewords have been generated, which we denote by \mathcal{C}_2 .

- 3) Class $\mathcal{T}^k(P_i)$: For $i \geq 3$, the ℓ -th codeword is drawn equiprobably from

$$\left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(i)}) : d(\mathbf{x}_j^{(i)}, \bar{\mathbf{x}}) > \Delta_i \text{ for } j \in [\ell - 1], \right. \\ \left. d(\mathbf{x}, \bar{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \forall \mathbf{x} \in \mathcal{C}_j, \text{ for } j \in [i - 1] \right\}.$$

Continue until all $|\mathcal{T}^k(P_i)|$ codewords have been generated; denote this set by \mathcal{C}_i .

By construction, each codeword in class i is at distance greater than Δ_i from all other codewords of the same class, and at distance greater than $\max(\Delta_i, \Delta_j)$ from every codeword in any class $j \neq i$. At the end of this recursive process, the code construction has discarded no more than

$$\sum_{i=1}^{N_k} \sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} \sum_{j=1}^{N_k} \left| \left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) : d(\mathbf{x}_\mathbf{v}, \bar{\mathbf{x}}) \leq \Delta_i \right\} \right| \quad (11)$$

input sequences. In other words, this expression pessimistically counts the sequences that violate the distance constraints. Since the distance function is type-dependent, this quantity does not depend on the specific realization of the codewords. To see this, observe

$$\sum_{i=1}^{N_k} \sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} \sum_{j=1}^{N_k} \left| \left\{ \bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) : d(\mathbf{x}_\mathbf{v}, \bar{\mathbf{x}}) \leq \Delta_i \right\} \right| \quad (12)$$

$$= \sum_{i=1}^{N_k} \sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} \sum_{j=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_{\mu(i)} \\ P_{\bar{X}}=Q_{\mu(j)}, d(P) \leq \Delta_i}} \sum_{\substack{\bar{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ (\mathbf{x}_\mathbf{v}, \bar{\mathbf{x}}) \in \mathcal{T}^n(P)}} 1 \quad (13)$$

$$\doteq \sum_{i=1}^{N_k} \sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} \sum_{j=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_{\mu(i)} \\ P_{\bar{X}}=Q_{\mu(j)}, d(P) \leq \Delta_i}} \exp(n H_P(\bar{X}|X)) \quad (14)$$

$$= \sum_{(i,j)=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_{\mu(i)} \\ P_{\bar{X}}=Q_{\mu(j)}, d(P)\leq\Delta_i}} |\mathcal{T}^k(P_i)| \exp(nH_P(\bar{X}|X)). \quad (15)$$

Thus, we may equivalently upper bound the number of discarded sequences by

$$\sum_{i=1}^{N_k} \sum_{c=1}^m \sum_{\substack{\bar{\mathbf{x}} \in \mathcal{T}^n(Q_c) \\ d(\bar{\mathbf{x}}, \tilde{\mathbf{x}}_i) \leq \Delta_i}} |\mathcal{T}^k(P_i)|, \quad (16)$$

where $\tilde{\mathbf{x}}_i$ is any sequence from $\mathcal{T}^n(Q_{\mu(i)})$. Then, it follows that for each class $\mathcal{T}^n(Q_c)$, $c \in [m]$, the number of sequences that violate the distance constraints is at most

$$\sum_{j=1}^{N_k} \sum_{\substack{\bar{\mathbf{x}} \in \mathcal{T}^n(Q_c) \\ d(\bar{\mathbf{x}}, \tilde{\mathbf{x}}_j) \leq \Delta_j}} |\mathcal{T}^k(P_j)|. \quad (17)$$

To ensure that the code construction produces the required number of valid codewords, we assume that for each class $c \in [m]$, non-negative values $\{\delta_c\}_{c=1}^m$ are chosen such that

$$\sum_{j=1}^{N_k} \sum_{\substack{\bar{\mathbf{x}} \in \mathcal{T}^n(Q_c) \\ d(\bar{\mathbf{x}}, \tilde{\mathbf{x}}_j) \leq \Delta_j}} |\mathcal{T}^k(P_j)| \leq \zeta_n |\mathcal{T}^n(Q_c)| e^{-n\delta_c}, \quad (18)$$

where $\zeta_n := N_k(n+1)^{|\mathcal{X}|^2+|\mathcal{X}|}$. For simplicity, and without loss of generality, we take $\delta_c = \delta$ for all $c \in [m]$ and some $\delta > 0$. Condition (18) guarantees that, for each source type class $\mathcal{T}^k(P_i)$ the set from which the codewords are drawn equiprobably has cardinality at least

$$|\mathcal{T}^n(Q_{\mu(i)})| - \zeta_n |\mathcal{T}^n(Q_{\mu(i)})| e^{-n\delta} = (1 - \zeta_n e^{-n\delta}) |\mathcal{T}^n(Q_{\mu(i)})| \quad (19)$$

at every step of the construction. Therefore, whenever (18) holds, the recursive code construction produces the required number of codewords $|\mathcal{V}|^k$.

Lemma 1. *Let $\mathbf{v} \in \mathcal{T}^k(P_i)$. Then the marginal distribution of the corresponding codeword $\mathbf{X}_{\mathbf{v}}$ is uniform over $\mathcal{T}^n(Q_{\mu(i)})$,*

$$\mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x}] = \frac{1}{|\mathcal{T}^n(Q_{\mu(i)})|} \text{ for } \mathbf{x} \in \mathcal{T}^n(Q_{\mu(i)}), \quad (20)$$

and zero otherwise.

Lemma 2. *Let $\mathbf{v} \in \mathcal{T}^k(P_i)$ and $\tilde{\mathbf{v}} \in \mathcal{T}^k(P_j)$, and let $\mathbf{x} \in \mathcal{T}^n(Q_{\mu(i)})$ and $\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)})$ denote their corresponding codewords under the proposed codebook construction. If $d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j)$, then*

$$\begin{aligned} & \mathbb{P}[(\mathbf{X}_{\mathbf{v}}, \mathbf{X}_{\tilde{\mathbf{v}}}) = (\mathbf{x}, \tilde{\mathbf{x}})] \\ & \leq \frac{1}{|\mathcal{T}^n(Q_{\mu(i)})| \cdot |\mathcal{T}^n(Q_{\mu(j)})| \cdot (1 - \zeta_n e^{-n\delta})^2}. \end{aligned} \quad (21)$$

The proof of Lemma 1 follows identically to that of [2, Lemma 4]. The proof of Lemma 2, along with the remaining proofs of the paper, is provided in Section IV.

III. ACHIEVABLE ERROR EXPONENTS

Consider a maximum metric decoder with a type-dependent decoding metric, meaning that it depends on the source and channel sequences only through their types. Specifically, for $\mathbf{v} \in \mathcal{T}^k(P_i)$, the decoding metric can be expressed as $q(\mathbf{v}, \mathbf{x}, \mathbf{y}) = q(i, \hat{P}_{\mathbf{x}\mathbf{y}})$.

Theorem 1. *Consider the code construction described with parameters $(d, \{\Delta_i\}_{i=1}^{N_k}, \delta)$ satisfying (18), together with a type-dependent decoding metric q . The ensemble average error probability satisfies*

$$\bar{p}_e \leq \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \exp\left(-nE_{\text{rgv}}^{(ij)}(P_V, W)\right), \quad (22)$$

where

$$E_{\text{rgv}}^{(ij)}(P_V, W) \triangleq te\left(\frac{R_i}{t}, P_V\right) + \min_{P \in \Gamma_{ij}} \left\{ D(P_{XY} \| Q_{\mu(i)} \times W) + |I_P(\bar{X}; X, Y) - R_j|^+ \right\}, \quad (23)$$

and

$$\Gamma_{ij} \triangleq \left\{ P_{X\bar{X}Y} : P_X = Q_{\mu(i)}, P_{\bar{X}} = Q_{\mu(j)}, d(P_{X\bar{X}}) \geq \max(\Delta_i, \Delta_j), q(i, P_{XY}) \leq q(j, P_{\bar{X}Y}) \right\}. \quad (24)$$

We next show that with an appropriate choice of the distance function and the minimum-distance parameters, the ensemble average error probability in (22) achieves both the random-coding and expurgated error exponents. To this end, following the choices in [2, Sec. VI], we set $d(P) = -I(P)$, and $\Delta_i = -(R_i + \delta)$. We will show that these choices satisfy the condition in (18). Observe that, for each $c \in [m]$,

$$\begin{aligned} & \sum_{j=1}^{N_k} \sum_{\substack{\bar{\mathbf{x}} \in \mathcal{T}^n(Q_c) \\ d(\bar{\mathbf{x}}, \tilde{\mathbf{x}}_j) \leq \Delta_j}} |\mathcal{T}^k(P_j)| \\ & \leq \sum_{j=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_c \\ P_{\bar{X}}=Q_{\mu(j)} \\ -I(P) \leq -(R_j + \delta)}} \sum_{\substack{\mathbf{x} \in \mathcal{T}^n(Q_c) \\ (\mathbf{x}, \tilde{\mathbf{x}}_j) \in \mathcal{T}^n(P)}} e^{nR_j} \\ & \leq \sum_{j=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_c \\ P_{\bar{X}}=Q_{\mu(j)} \\ -I(P) \leq -(R_j + \delta)}} \frac{e^{nH_P(X|\bar{X})} e^{nR_j}}{|\mathcal{T}^n(Q_c)|} |\mathcal{T}^n(Q_c)|. \end{aligned} \quad (25)$$

Recall from [7, Lemma 2.3] that for any type Q , we have $(n+1)^{-|\mathcal{X}|} e^{nH(Q)} \leq |\mathcal{T}^n(Q)|$. Therefore,

$$\frac{e^{n(H_P(X|\bar{X})+R_j)}}{|\mathcal{T}^n(Q_c)|} \leq (n+1)^{|\mathcal{X}|} e^{-n(I_P(X;\bar{X})-R_j)}. \quad (27)$$

Moreover, by the minimum-distance condition, any admissible distribution P satisfies $I(P) - R_j \geq \delta$. Thus,

$$\frac{e^{n(H_P(X|\bar{X})+R_j)}}{|\mathcal{T}^n(Q_c)|} \leq (n+1)^{|\mathcal{X}|} e^{-n\delta}. \quad (28)$$

Substituting this back, we obtain

$$\begin{aligned}
& \sum_{j=1}^{N_k} \sum_{\substack{\bar{x} \in \mathcal{T}^n(Q_c) \\ d(\bar{x}, \bar{x}_j) \leq \Delta_j}} |\mathcal{T}^k(P_j)| \\
& \leq \sum_{j=1}^{N_k} \sum_{\substack{P_{X\bar{X}}: P_X=Q_c \\ P_{\bar{X}}=Q_{\mu(j)} \\ -I(P) \leq -(R_j+\delta)}} |\mathcal{T}^n(Q_c)| \cdot (n+1)^{|\mathcal{X}|} e^{-n\delta} \\
& \leq |\mathcal{T}^n(Q_c)| \cdot N_k (n+1)^{|\mathcal{X}|^2+|\mathcal{X}|} e^{-n\delta}, \quad (30)
\end{aligned}$$

thereby satisfying (18). Therefore, the minimum-distance condition reduces to $I(P) < \min(R_i, R_j) + \delta$. With these choices, the ensemble is universal in the sense that it does not depend on either the source or the channel distribution, as in [2]. Moreover, the resulting ensemble achieves the random-coding and expurgated exponents simultaneously.

A. Achieving the Random Coding Error Exponent

To attain the random-coding exponent, we choose the decoding metric to be the universal generalized maximum mutual information introduced by Csiszár [4], namely,

$$q(i, P) = I(P) - R_i. \quad (31)$$

We relax the constraint set Γ_{ij} in (24) by dropping the minimum-distance requirement, which can only weaken the bound. In addition, we use the inequality $I_P(\bar{X}; X, Y) \geq I_P(\bar{X}; Y)$ to further lower bound the exponent. This yields,

$$\begin{aligned}
E_{\text{rgv}}^{(ij)}(P_V, W) & \geq te \left(\frac{R_i}{t}, P_V \right) + \min_P \left\{ D(P_{XY} \| Q_{\mu(i)} \times W) \right. \\
& \quad \left. + |I_P(\bar{X}; Y) - R_j|^+ \right\}, \quad (32)
\end{aligned}$$

where the minimization is over distributions P satisfying $P_X = Q_{\mu(i)}$, $P_{\bar{X}} = Q_{\mu(j)}$, and $I_P(X; Y) - R_i \leq I_P(\bar{X}; Y) - R_j$. Hence, for any admissible P ,

$$|I_P(\bar{X}; Y) - R_j|^+ \geq |I_P(X; Y) - R_i|^+. \quad (33)$$

Therefore, for all $j \in [N_k]$,

$$E_{\text{rgv}}^{(ij)}(P_V, W) \geq te \left(\frac{R_i}{t}, P_V \right) + E_r(Q_{\mu(i)}, R_i). \quad (34)$$

Consequently,

$$\bar{p}_e \leq \sum_{i=1}^{N_k} \exp \left(-n \left[te \left(\frac{R_i}{t}, P_V \right) + E_r(Q_{\mu(i)}, R_i) \right] \right) \quad (35)$$

which matches the JSCC random coding bound (see (2)).

B. Achieving the Expurgated Error Exponent

To attain the expurgated error exponent, we adopt the decoding metric proposed in [5]. In particular, we set

$$q(i, P) = \mathbb{E}_P [\log W(Y|X)] - 2R_i. \quad (36)$$

We lower bound the expression in (23) by removing the $|\cdot|^+$ operator. For any admissible distribution $P \in \Gamma_{ij}$,

$$\begin{aligned}
& D(P_{XY} \| Q_{\mu(i)} \times W) + I_P(\bar{X}; X, Y) \\
& = D(P \| P_{X\bar{X}} \times W) + I_P(X; \bar{X}). \quad (37)
\end{aligned}$$

Using this identity, we obtain the following lower bound,

$$\begin{aligned}
E_{\text{rgv}}^{(ij)}(P_V, W) & \geq te \left(\frac{R_i}{t}, P_V \right) + \min_{\substack{\Lambda_{X\bar{X}}: \Lambda_X=Q_{\mu(i)} \\ \Lambda_{\bar{X}}=Q_{\mu(j)} \\ I(\Lambda) \leq \min(R_i, R_j) + \delta}} I_\Lambda(X; \bar{X}) \\
& \quad + \min_{P: P_{X\bar{X}}=\Lambda} D(P \| \Lambda \times W) - R_j. \quad (38) \\
& \quad \mathbb{E}_P \left[\log \frac{W(Y|X)}{W(Y|\bar{X})} \right] \leq 2(R_i - R_j)
\end{aligned}$$

We now evaluate the inner minimization. By Lagrange duality it can be shown that,

$$\begin{aligned}
& \min_{P: P_{X\bar{X}}=\Lambda} D(P \| \Lambda \times W) - R_j \\
& \mathbb{E}_P \log \left[\frac{W(Y|X)}{W(Y|\bar{X})} \right] \leq 2(R_i - R_j) \\
& = \min_{P: P_{X\bar{X}}=\Lambda} \sup_{s \geq 0} \mathbb{E}_P \left[\log \frac{P(Y|X, \bar{X})}{W(Y|X)^{1-s} W(Y|\bar{X})^s} \right] \\
& \quad - R_j + 2s(R_j - R_i). \quad (39)
\end{aligned}$$

Choosing $s = \frac{1}{2}$, which may be suboptimal, and applying the log-sum inequality yields

$$\begin{aligned}
& \min_{P: P_{X\bar{X}}=\Lambda} D(P \| \Lambda \times W) - R_j \\
& \mathbb{E}_P \left[\log \frac{W(Y|X)}{W(Y|\bar{X})} \right] \leq 2(R_i - R_j) \\
& \geq - \sum_{(x, \bar{x})} \Lambda(x, \bar{x}) \log \sum_y \sqrt{W(y|x)W(y|\bar{x})} - R_i. \quad (40)
\end{aligned}$$

Therefore,

$$\begin{aligned}
E_{\text{rgv}}^{(ij)}(t, P_V) & \geq te \left(\frac{R_i}{t}, P_V \right) \\
& \quad + \min_{\substack{\Lambda_{X\bar{X}}: \Lambda_X=Q_{\mu(i)} \\ \Lambda_{\bar{X}}=Q_{\mu(j)} \\ I(\Lambda) \leq \min(R_i, R_j) + \delta}} I(\Lambda) + \mathbb{E}_\Lambda [d_W(X, \bar{X})] - R_i. \quad (41)
\end{aligned}$$

We can further relax the bound by replacing the constraint $I(\Lambda) \leq \min(R_i, R_j) + \delta$ with $I(\Lambda) \leq R_i + \delta$, and by relaxing the condition $\Lambda_{\bar{X}} = Q_{\mu(j)}$ to $\Lambda_{\bar{X}} \in \mathcal{Q}_m$. The bound becomes independent of j . For all $j \in [N_k]$, letting $\delta \rightarrow 0$ yields

$$E_{\text{rgv}}^{(ij)}(P_V, W) \geq te \left(\frac{R_i}{t}, P_V \right) + E'_{\text{ex}}(Q_{\mu(i)}, \mathcal{Q}_m, R_i). \quad (42)$$

Therefore,

$$p_e \leq \sum_{i=1}^{N_k} \exp \left(-n \left[te \left(\frac{R_i}{t}, P_V \right) + E'_{\text{ex}}(Q_{\mu(i)}, \mathcal{Q}_m, R_i) \right] \right) \quad (43)$$

which matches the expurgated bound in (5).

The above recursive construction naturally applies to single-class coding, in which all codewords are assigned the same distribution, i.e., $Q_c = Q$ for all $c \in [m]$. In that case, the corresponding single-class results are recovered; see [8].

IV. PROOFS

Proof of Lemma 2. The proof closely follows that of [2, Lemma 2]. Without loss of generality, assume $i < j$. Let \mathbf{v} and $\tilde{\mathbf{v}}$ be the ℓ_i -th and ℓ_j -th sequences in $\mathcal{T}^k(P_i)$ and $\mathcal{T}^k(P_j)$, respectively. By construction, the codeword $\mathbf{X}_{\mathbf{v}}$ is drawn equiprobably from the set $\mathcal{T}^n(Q_{\mu(i)}, i, \ell_i) \subseteq \mathcal{T}^n(Q_{\mu(i)})$, consisting of all sequences that satisfy the imposed distance constraints with respect to all previously generated codewords. Consequently, for any admissible \mathbf{x} ,

$$\mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x} \mid \text{past codewords}] \leq \frac{1}{|\mathcal{T}^n(Q_{\mu(i)}, i, \ell_i)|}. \quad (44)$$

An analogous argument applies to $\mathbf{X}_{\tilde{\mathbf{v}}}$. Expanding the joint probability yields

$$\begin{aligned} & \mathbb{P}[(\mathbf{X}_{\mathbf{v}}, \mathbf{X}_{\tilde{\mathbf{v}}}) = (\mathbf{x}, \tilde{\mathbf{x}})] \\ &= \sum_{(\text{before } \mathbf{v})} \sum_{(\text{after } \mathbf{v} \text{ before } \tilde{\mathbf{v}})} \mathbb{P}[\text{codewords before } \mathbf{v}] \\ & \quad \times \mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x} \mid \text{codewords before } \mathbf{v}] \quad (45) \\ & \quad \times \mathbb{P}[\text{codewords after } \mathbf{v} \text{ before } \tilde{\mathbf{v}} \mid \text{past codewords}] \\ & \quad \times \mathbb{P}[\mathbf{X}_{\tilde{\mathbf{v}}} = \tilde{\mathbf{x}} \mid \text{codewords before } \tilde{\mathbf{v}}]. \end{aligned}$$

Using (44) to upper bound the second and fourth probability terms in (45), and noting that the sums over the remaining probability terms evaluate to one, we obtain

$$\begin{aligned} & \mathbb{P}[(\mathbf{X}_{\mathbf{v}}, \mathbf{X}_{\tilde{\mathbf{v}}}) = (\mathbf{x}, \tilde{\mathbf{x}})] \\ & \leq \frac{1}{|\mathcal{T}^n(Q_{\mu(i)}, i, \ell_i)|} \frac{1}{|\mathcal{T}^n(Q_{\mu(j)}, j, \ell_j)|}. \quad (46) \end{aligned}$$

Lower bounding each cardinality by (19) yields the result.

Proof of Theorem 1. We begin with the extended RCU bound, given in [9, Eq. (26)],

$$\bar{p}_e \leq \sum_{i=1}^{N_k} \bar{\varepsilon}_i, \quad (47)$$

where

$$\begin{aligned} \bar{\varepsilon}_i &= \sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} P^k(\mathbf{v}) \sum_{(\mathbf{x}, \mathbf{y})} \mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x}] W^n(\mathbf{y}|\mathbf{x}) \\ & \min \left\{ 1, \sum_{j=1}^{N_k} \sum_{\tilde{\mathbf{v}} \in \mathcal{T}^k(P_j)} \mathbb{P} \left[q(j, \hat{P}_{\mathbf{X}_{\tilde{\mathbf{v}}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{X}_{\mathbf{v}}\mathbf{y}}) \right] \right\}. \quad (48) \end{aligned}$$

For a given $(\mathbf{v}, \tilde{\mathbf{v}}, \mathbf{x}, \mathbf{y})$, and using Lemmas 1 and 2, the pairwise error probability can be written as

$$\begin{aligned} & \mathbb{P} \left[q(P_j, \hat{P}_{\mathbf{X}_{\tilde{\mathbf{v}}}\mathbf{y}}) \geq q(P_i, \hat{P}_{\mathbf{X}_{\mathbf{v}}\mathbf{y}}) \right] \\ &= \sum_{\substack{\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \\ q(j, \hat{P}_{\tilde{\mathbf{x}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{x}\mathbf{y}})}} \mathbb{P}[\mathbf{X}_{\tilde{\mathbf{v}}} = \tilde{\mathbf{x}} \mid \mathbf{X}_{\mathbf{v}} = \mathbf{x}, \mathbf{Y} = \mathbf{y}] \quad (49) \\ &= \sum_{\substack{\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \\ q(j, \hat{P}_{\tilde{\mathbf{x}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{x}\mathbf{y}})}} \frac{\mathbb{P}[\mathbf{X}_{\tilde{\mathbf{v}}} = \tilde{\mathbf{x}}, \mathbf{X}_{\mathbf{v}} = \mathbf{x}]}{\mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x}]} \quad (50) \end{aligned}$$

$$\leq \frac{1}{(1 - \zeta_n e^{-n\delta})^2} \sum_{\substack{\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \\ q(j, \hat{P}_{\tilde{\mathbf{x}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{x}\mathbf{y}})}} \frac{1}{|\mathcal{T}^n(Q_{\mu(j)})|}. \quad (51)$$

Observe that the bound does not depend on the particular $\tilde{\mathbf{v}}$. Moreover, the factor $(1 - \zeta_n e^{-n\delta})$ can be made arbitrarily close to one for sufficiently large n . Thus, we obtain

$$\begin{aligned} & \sum_{\tilde{\mathbf{v}} \in \mathcal{T}^k(P_j)} \mathbb{P} \left[q(j, \hat{P}_{\mathbf{X}_{\tilde{\mathbf{v}}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{X}_{\mathbf{v}}\mathbf{y}}) \right] \\ & \leq \sum_{\substack{\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \\ q(j, \hat{P}_{\tilde{\mathbf{x}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{x}\mathbf{y}})}} \frac{e^{nR_j}}{|\mathcal{T}^n(Q_{\mu(j)})|}. \quad (52) \end{aligned}$$

Using the method of types, we have

$$\sum_{\substack{\tilde{\mathbf{x}} \in \mathcal{T}^n(Q_{\mu(j)}) \\ d(\mathbf{x}, \tilde{\mathbf{x}}) > \max(\Delta_i, \Delta_j) \\ q(j, \hat{P}_{\tilde{\mathbf{x}}\mathbf{y}}) \geq q(i, \hat{P}_{\mathbf{x}\mathbf{y}})}} \frac{e^{nR_j}}{|\mathcal{T}^n(Q_{\mu(j)})|} \quad (53)$$

$$= \sum_{\substack{V_{X\tilde{X}Y}: V_{XY} = \hat{P}_{\mathbf{x}\mathbf{y}}, V_{\tilde{X}} = Q_{\mu(j)} \\ d(V_{X\tilde{X}}) > \max(\Delta_i, \Delta_j) \\ q(j, V_{\tilde{X}Y}) \geq q(i, V_{XY})}} \sum_{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in \mathcal{T}^n(V)} \frac{e^{nR_j}}{|\mathcal{T}^n(Q_{\mu(j)})|} \quad (54)$$

$$\doteq \sum_{\substack{V_{X\tilde{X}Y}: V_{XY} = \hat{P}_{\mathbf{x}\mathbf{y}}, V_{\tilde{X}} = Q_{\mu(j)} \\ d(V_{X\tilde{X}}) > \max(\Delta_i, \Delta_j) \\ q(j, V_{\tilde{X}Y}) \geq q(i, V_{XY})}} \frac{e^{nH_V(\tilde{X}|X, Y)} e^{nR_j}}{e^{nH_V(\tilde{X})}} \quad (55)$$

$$\doteq \max_{\substack{V_{X\tilde{X}Y}: V_{XY} = \hat{P}_{\mathbf{x}\mathbf{y}}, V_{\tilde{X}} = Q_{\mu(j)} \\ d(V_{X\tilde{X}}) > \max(\Delta_i, \Delta_j) \\ q(j, V_{\tilde{X}Y}) \geq q(i, V_{XY})}} \exp(-n [I_V(\tilde{X}; X, Y) - R_j]). \quad (56)$$

Similarly, $\mathcal{T}^n(Q_{\mu(j)}) \doteq \exp(nH_V(\tilde{X}))$. Furthermore, observe that for given (\mathbf{x}, \mathbf{y}) where $\mathbf{x} \in \mathcal{T}^n(Q_{\mu(i)})$, we can evaluate the expectation as follows

$$\begin{aligned} & \sum_{(\mathbf{x}, \mathbf{y})} \mathbb{P}[\mathbf{X}_{\mathbf{v}} = \mathbf{x}] W^n(\mathbf{y}|\mathbf{x}) \\ & \doteq \max_{P_{XY}: P_X = Q_{\mu(i)}} \exp(-nD(P \| Q_{\mu(i)} \times W)). \quad (57) \end{aligned}$$

The proof is completed by observing, following [11, Eqs. (45)–(48)], that

$$\sum_{\mathbf{v} \in \mathcal{T}^k(P_i)} P^k(\mathbf{v}) \leq \exp\left(-ke\left(\frac{R_i}{t}, P_V\right)\right). \quad (58)$$

Therefore, overall,

$$\begin{aligned} \bar{\varepsilon}_i & \leq \sum_{j=1}^{N_k} \exp\left(-n \left[te\left(\frac{R_i}{t}, P_V\right) + \min_{P \in \Gamma_{ij}} \left\{ D(P_{XY} \| Q_{\mu(i)} \times W) \right. \right. \right. \\ & \quad \left. \left. \left. + |I_P(\tilde{X}; X, Y) - R_j|^+ \right\} \right] \right). \quad (59) \end{aligned}$$

REFERENCES

- [1] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 5–12, 1981.
- [2] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert-Varshamov codes," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3452–3469, 2019.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*. USA: John Wiley & Sons, Inc., 1968.
- [4] I. Csiszár, "Joint source-channel error exponent," *Problems of Control and Information Theory*, vol. 9, pp. 315–328, 1980.
- [5] —, "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.
- [6] S. A. Moeini and A. Guillén i Fàbregas, "Two-class joint source-channel coding: Expurgated exponents with i.i.d. distributions," in *International Zurich Seminar on Information and Communication (IZS 2026)*. ETH Zurich, 2026, pp. 33–38. [Online]. Available: <https://doi.org/10.3929/ethz-c-000796722>
- [7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
- [8] 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)*, 2025, pp. 794–797.
- [9] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martínez, "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.
- [10] S. A. Moeini and A. Guillén i Fàbregas, "Dual-domain exponent of maximum mutual information decoding," in *2025 IEEE Information Theory Workshop (ITW)*, 2025, pp. 810–814.
- [11] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martínez, "Achieving Csiszár's source-channel coding exponent with product distributions," in *2012 IEEE International Symposium on Information Theory Proceedings*, 2012, pp. 1548–1552.