On the Joint Source-Channel Coding Error Exponent for Systems with Memory*

Yangfan Zhong
Dept. of Mathematics & Statistics
Queen's University
Kingston, ON K7L 3N6, Canada.
Email: yangfan@mast.queensu.ca

Fady Alajaji
Dept. of Mathematics & Statistics
Queen's University
Kingston, ON K7L 3N6, Canada.
Email: fady@mast.queensu.ca

L. Lorne Campbell
Dept. of Mathematics & Statistics
Queen's University
Kingston, ON K7L 3N6, Canada.
Email: campblll@mast.queensu.ca

***Abstract*— We establish an upper bound for the joint source-channel coding (JSCC) error exponent $\mathbf{E_J(Q, W)}$ for a discrete stationary ergodic Markov (SEM) source Q and a discrete channel W with additive SEM noise. This bound, which is expressed in terms of the Rényi entropy rates of the source and noise processes, admits an identical form to Csiszár's sphere-packing upper bound for the JSCC error exponent for memoryless systems [3]. In this regard, our result is a natural extension of Csiszár's upper bound of the JSCC error exponent from the case of memoryless systems to the case of SEM systems. We also investigate the analytical computation of $\mathbf{E_J(Q, W)}$ by comparing our bound with Gallager's random-coding lower bound [5], when the latter one is specialized to the SEM source-channel system.**

I. Introduction

The lossless joint source-channel coding (JSCC) error exponent $E_J(Q, W)$, for a discrete memoryless source (DMS) Q and a discrete memoryless channel (DMC) W was thoroughly studied in [3]-[5], [13]-[15]. In [3], [4], Csiszár establishes two lower bounds and an upper bound for $E_J(Q, W)$ based on the random-coding and expurgated lower bounds and the sphere-packing upper bound for the DMC error exponent. In [13]-[15], we investigate the analytical computation of Csiszár's lower and upper bounds for $E_J(Q, W)$, and provide equivalent expressions for these bounds. As a result, we are able to systematically compare the JSCC error exponent with the traditional tandem coding error exponent $E_T(Q, W)$, the exponent resulting from separately performing and concatenating optimal source and channel coding. We show that JSCC can double the error exponent vis-a-vis tandem coding by proving that $E_J(Q, W) \leq 2E_T(Q, W)$ and give the condition for equality. Our numerical results also indicate that $E_J(Q, W)$ can be nearly twice as large as $E_T(Q, W)$ for many DMS-DMC pairs, hence illustrating the substantial gain that JSCC can achieve over tandem coding. It is also shown in [15] that this gain translates into a power saving larger than 2 dB for binary DMS sent over binary-input white Gaussian noise and Rayleigh-fading channels with finite output quantization.

As most real-world data sources (e.g., multimedia sources) and communication channels (e.g., wireless channels) exhibit statistical dependency or memory, it is of natural interest to study the JSCC error exponent for systems with memory. Furthermore, the determination of the JSCC error exponent (or its bounds), particularly in terms of computable parametric expressions, may lead to the identification of important information-theoretic design criteria for the construction of powerful JSCC techniques that fully exploit the source-channel memory. In this paper, we investigate the JSCC error exponent for a communication system with memory. Specifically, we establish an upper bound for $E_J(\mathbf{Q}, \mathbf{W})$ for a system consisting of a stationary ergodic (irreducible) Markov (SEM) source $\mathbf{Q}$ and a channel $\mathbf{W}$ with additive SEM noise $\mathbf{P_W}$ (for the sake of brevity, we hereafter refer to this channel as the SEM channel $\mathbf{W}$). Note that Markov sources are widely used to model realistic data sources, and SEM channels can approximate well discretized fading channels with memory (e.g., see [12]). The proof of the bound, which follows the standard lower bounding technique of [1] for the average probability of error, is based on the judicious construction from the original SEM source-channel pair ($\mathbf{Q}$,$\mathbf{W}$) of an artificial[1] Markov source $\widetilde{\mathbf{Q}}_{\alpha^*}$ and an artificial channel $\mathbf{V}$ with additive Markov noise $\widetilde{\mathbf{P}}_{\mathbf{W}_{\alpha^*}}$, where α^* is a parameter to be optimized, such that the stationarity and ergodicity properties are retained by $\widetilde{\mathbf{Q}}_{\alpha^*}$ and $\widetilde{\mathbf{P}}_{\mathbf{W}_{\alpha^*}}$. The proof then employs the strong converse JSCC Theorem for ergodic sources and channels with ergodic additive noise and the fact that the normalized log-likelihood ratio between n-tuples of two SEM sources asymptotically converges (as $n \to \infty$) to their Kullback-Leibler divergence rate. To the best of our knowledge, this upper bound, which is expressed in terms of the Rényi entropy rates of the source and noise processes, is new and the analytical computation of the JSCC error exponent for systems with Markovian memory has not been addressed before. The bound is also shown to admit an equivalent representation to Csiszár's sphere-packing upper bound for the case of DMS-DMC pairs [3]. In this regard, our result is a natural extension of Csiszár's upper bound from the case of memoryless systems to the case of SEM systems.

We next examine Gallager's lower bound for $E_J(\mathbf{Q}, \mathbf{W})$ [5, Problem 5.16] (which is valid for arbitrary source-channel

*This work was supported in part by PREA and NSERC of Canada.

[1]The notion of artificial (or auxiliary) Markov sources is herein adopted from [10], where Vašek employed it to study the source coding error exponent for ergodic Markov sources.

pairs with memory), when specialized to the SEM source-channel system. By comparing our upper bound with Gallager's random-coding lower bound, we provide the condition under which they coincide, hence exactly determining $E_J(\mathbf{Q},\mathbf{W})$. We note that this condition holds for a large class of SEM source-channel pairs.

II. Preliminaries

Definition 1: A joint source-channel code with blocklength n for a discrete source with finite alphabet $\mathcal{S}$ described by the sequence of $k(n)$-dimensional distributions $\mathbf{Q} \triangleq \{Q^{(k(n))} : \mathcal{S}^{k(n)}\}_{k(n)=1}^{\infty}$ and a discrete channel described by the sequence of n-dimensional transition distributions $\mathbf{W} \triangleq \{W^{(n)} : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ with common input and output alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, ..., B-1\}$ is a pair of mappings: $f_n : \mathcal{S}^{k(n)} \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{k(n)}$.

In this paper, we confine our attention to discrete channels with (modulo B) additive noise of n-dimensional distribution $\mathbf{P_W} \triangleq \{P_W^{(n)} : \mathcal{Z}^n\}_{n=1}^{\infty}$. The channels are described by

$$Y_i = X_i \oplus Z_i \pmod B,$$

where Y_i, X_i and Z_i are the output, input and noise symbols at time i, and $Z_i \in \mathcal{Z} = \{0, 1, ..., B-1\}$ is independent of X_i, $i = 1, 2, ..., n$. For the sake of convenience, we will use throughout $\oplus$ and $\ominus$ to stand for modulo-B addition and modulo-B subtraction, respectively.

Denote the transmitted source message by $\mathbf{s} \in \mathcal{S}^{k(n)}$, the corresponding n-length codeword by $f_n(\mathbf{s}) = \mathbf{x} = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ and the received codeword at the channel output by $\mathbf{y} = (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$. Then the probability of receiving $\mathbf{y}$ under the conditions that the message $\mathbf{s}$ is transmitted (i.e., the input codeword is $f_n(\mathbf{s}) = \mathbf{x}$) is given by $P(\mathbf{y}|\mathbf{s}) = W^{(n)}(\mathbf{y}|f_n(\mathbf{s})) = W^{(n)}(\mathbf{y}|\mathbf{x}) = W^{(n)}(\mathbf{y}\ominus\mathbf{x}|\mathbf{x}) = P_W^{(n)}(\mathbf{z})$, where the last equality follows by the independence of input codeword $\mathbf{x}$ and the additive noise $\mathbf{z} = \mathbf{y} \ominus \mathbf{x}$. The decoding operation φ_n is the rule decoding on a set of nonintersecting sets of output words $A_\mathbf{s}$ such that $\bigcup_\mathbf{s} A_\mathbf{s} = \mathcal{Y}^n$. If $\mathbf{y} \in A_{\mathbf{s}'}$, then we conclude that the source $\mathbf{s}'$ has been transmitted. If the source $\mathbf{s}$ has been transmitted, the conditional error probability in decoding is given by $P(\mathbf{y} \in A_\mathbf{s}^c|\mathbf{s}) \triangleq W^{(n)}(\mathbf{y} \in A_\mathbf{s}^c|f_n(\mathbf{s}))$, where $A_\mathbf{s}^c = \mathcal{Y}^n - A_\mathbf{s}$, and the average probability of error of the code (f_n, φ_n) is

$$P_e^{(n)}(Q^{(k(n))}, W^{(n)}) = \sum_{\mathbf{s}} Q^{(k(n))}(\mathbf{s}) \sum_{\mathbf{y}\in A_\mathbf{s}^c} W^{(n)}(\mathbf{y}|f_n(\mathbf{s})).$$

Since $k(n)$ source symbols are mapped to n channel symbols, $R_t \triangleq k(n)/n$ source symbols/channel use is called the code's transmission rate, which is assumed to be independent of n.

Definition 2: The JSCC error exponent $E_J(\mathbf{Q},\mathbf{W})$ for source $\mathbf{Q}$ and channel $\mathbf{W}$ is defined as the largest number E for which there exists a sequence of joint source-channel codes (f_n, φ_n) with

$$E \le \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(Q^{(k(n))}, W^{(n)}).$$

A lower bound for $E_J(\mathbf{Q},\mathbf{W})$ for arbitrary discrete source-channel pairs with memory was already obtained by Gallager [5]. In this work, we establish an upper bound for $E_J(\mathbf{Q},\mathbf{W})$ for an SEM source and an SEM channel. For a discrete source $\mathbf{Q}$, its entropy rate is defined by

$$H(\mathbf{Q}) \triangleq \limsup_{k\to\infty} \frac{1}{k} H(Q^{(k)}),$$

where $H(Q^{(k)})$ is the Shannon entropy rate of $Q^{(k)}$; $H(\mathbf{Q})$ admits an operational meaning (in the sense of the lossless fixed length source coding theorem) if $\mathbf{Q}$ is information stable [6]. The source Rényi entropy rate of order α ($\alpha \ge 0$) is defined by

$$\mathcal{R}_\alpha(\mathbf{Q}) \triangleq \limsup_{k\to\infty} \frac{1}{k} H_\alpha(Q^{(k)}),$$

where

$$H_\alpha(Q^{(k)}) \triangleq \frac{1}{1-\alpha}\log_2 \sum_{\mathbf{s}\in\mathcal{S}^k} Q^{(k)}(\mathbf{s})^\alpha$$

is the Rényi entropy, and the special case of $\alpha = 1$ should be interpreted as

$$H_1(Q^{(k)}) \triangleq \lim_{\alpha\to1} \frac{1}{1-\alpha}\log_2 \sum_{\mathbf{s}\in\mathcal{S}^k} Q^{(k)}(\mathbf{s})^\alpha = H(Q^{(k)}).$$

The channel capacity for any discrete (information stable [6], [11]) channel $\mathbf{W}$ is given by

$$C(\mathbf{W}) \triangleq \liminf_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} I(W^{(n)}; P_{X^n}),$$

where $I(\cdot;\cdot)$ denotes mutual information. If the channel $\mathbf{W}$ is an additive noise channel with noise process $\mathbf{P_W}$, then $C(\mathbf{W}) = \log_2 B - H(\mathbf{P_W})$, where $H(\mathbf{P_W})$ is the noise entropy rate.

III. Markov and Artificial Markov Sources

Without loss of generality, we only deal with first-order Markov sources since any k-th order Markov source can be converted to a first-order Markov source by k-step blocking it. For the sake of convenience (since we will apply the following results to both the SEM source and the SEM channel), we use, throughout this section, $\mathbf{P} \triangleq \left\{p^{(n)} : \mathcal{U}^n\right\}_{n=1}^{\infty}$ to denote a first-order SEM source with finite alphabet $\mathcal{U} \triangleq \{1, 2, ..., M\}$, initial distribution $p_i \triangleq Pr\{U_1 = i\}$, $i \in \mathcal{U}$ and transition distribution $p_{ij} \triangleq Pr\{U_{k+1} = j|U_k = i\}$, $i, j \in \mathcal{U}$, so that the n-tuple probability is given by

$$\begin{aligned} p^{(n)}(i^n) &\triangleq Pr\{U_1 = i_1, ..., U_n = i_n\} \\ &= p_{i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n}, \quad i_1, ..., i_n \in \mathcal{U}. \end{aligned}$$

Denote the transition matrix by $P \triangleq [p_{ij}]_{M\times M}$, we then set $P(\alpha) \triangleq \left[p_{ij}^\alpha\right]_{M\times M}$ for any $0 \le \alpha \le 1$, which is nonnegative and irreducible. The Perron-Frobenius Theorem [9] asserts that the matrix $P(\alpha)$ possesses a maximal positive eigenvalue $\lambda_\alpha(\mathbf{P})$ with positive (right) eigenvector $\mathbf{v}(\alpha) = (v_1(\alpha), ..., v_M(\alpha))^T$ such that $\sum_i v_i(\alpha) = 1$. As in [10], we define the artificial Markov source $\widetilde{\mathbf{P}}_\alpha \triangleq \left\{\widetilde{p}_\alpha^{(n)} : \mathcal{U}^n\right\}_{n=1}^{\infty}$

with respect to the original source $\mathbf{P}$ such that the transition matrix is $\widetilde{P}(\alpha) \triangleq [\widetilde{p}_{ij}(\alpha)]_{M\times M}$, where

$$\widetilde{p}_{ij}(\alpha) \triangleq \frac{p_{ij}^{\alpha} v_j(\alpha)}{\lambda_\alpha(\mathbf{P}) v_i(\alpha)}. \tag{1}$$

(It can be easily verified that $\sum_j \widetilde{p}_{ij}(\alpha) = 1$.) We emphasize that the artificial source retains the stochastic characteristics (ergodicity) of the original source because $\widetilde{p}_{ij}(\alpha) = 0$ if and only if $p_{ij} = 0$, and clearly, for all n, the n-th marginal of $\widetilde{\mathbf{P}}_\alpha$ is absolutely continuous with respect to the n-th marginal of $\mathbf{P}$. The entropy rate of the artificial Markov process is hence given by

$$H(\widetilde{\mathbf{P}}_\alpha) = -\sum_i \sum_j \pi_i(\alpha)\widetilde{p}_{ij}(\alpha)\log_2 \widetilde{p}_{ij}(\alpha),$$

where $\pi(\alpha) \triangleq (\pi(\alpha)_1, \pi(\alpha)_2, ..., \pi(\alpha)_M)$ is the stationary distribution of the stochastic matrix $\widetilde{P}(\alpha)$. We call the artificial Markov source with initial distribution $\pi(\alpha)$ the artificial SEM source. It is known [10, Lemmas 2.1-2.4] that $H(\widetilde{\mathbf{P}}_\alpha)$ is a continuous and decreasing function of $\alpha \in [0,1]$. To compare $H(\widetilde{\mathbf{P}}_0)$ with the entropy of the DMS with uniform distribution $(\frac{1}{M}, ..., \frac{1}{M})$, we have the following lemma.

Lemma 1: $H(\widetilde{\mathbf{P}}_0) \le \log_2 M$ with equality if and only if $P > [0]_{M\times M}$, i.e., $p_{ij} > 0$ for all $i, j \in \mathcal{U}$.

Lemma 2: For an SEM source $\mathbf{P}$ and the artificial SEM source $\widetilde{\mathbf{P}}_\alpha$,

$$\frac{1}{n}\log\frac{\widetilde{p}_\alpha^{(n)}(i^n)}{p^{(n)}(i^n)} \longrightarrow \frac{1-\alpha}{\alpha}H(\widetilde{\mathbf{P}}_\alpha) - \frac{1}{\alpha}\log\lambda_\alpha(\mathbf{P}), \quad i^n \in \mathcal{U}^n,$$

almost surely under $\widetilde{p}_\alpha$ as $n \to \infty$.

The proof of Lemma 2 follows by the definition of artificial SEM sources and the fact that the normalized log-likelihood ratio between n-tuples of two SEM processes converges as $n \to \infty$ to their Kullback-Leibler divergence rate (by the Ergodic Theorem [2]).

Lemma 3: [8], [10] For an SEM source $\mathbf{P}$ and any $\rho \ge 0$, we have

$$\rho\mathcal{R}_{\frac{1}{1+\rho}}(\mathbf{P}) = (1+\rho)\log\lambda_{\frac{1}{1+\rho}}(\mathbf{P}),$$

and

$$H\left(\widetilde{\mathbf{P}}_{\frac{1}{1+\rho}}\right) = \frac{\partial}{\partial\rho}(1+\rho)\log\lambda_{\frac{1}{1+\rho}}(\mathbf{P}).$$

IV. Bounds for $E_J(\mathbf{Q},\mathbf{W})$

We first derive an upper bound for $E_J(\mathbf{Q},\mathbf{W})$ for an SEM source-channel pair by employing a strong converse JSCC theorem for ergodic sources and channels with ergodic additive noise (Theorem 1) and Lemma 2.

Theorem 1: (Strong converse JSCC Theorem) For a source $\mathbf{Q}$ and a channel $\mathbf{W}$ with additive noise $\mathbf{P}_\mathbf{W}$ such that $\mathbf{Q}$ and $\mathbf{P}_\mathbf{W}$ are ergodic processes, if $C(\mathbf{W}) = \log_2 B - H(\mathbf{P}_\mathbf{W}) < R_t H(\mathbf{Q})$, then $\lim_{n\to\infty} P_e^{(n)}(Q^{(k(n))}, W^n) = 1$.

Theorem 1 is proved via a lower bounding technique [1] on the probability of error and the well known Shannon-McMillan-Breiman Theorem for ergodic processes [2].

Theorem 2: For an SEM source $\mathbf{Q}$ and a discrete channel $\mathbf{W}$ with additive SEM noise $\mathbf{P}_\mathbf{W}$ such that $R_t H(\mathbf{Q}) \le C(\mathbf{W})$, the JSCC error exponent $E_J(\mathbf{Q},\mathbf{W})$ satisfies

$$\begin{aligned} E_J(\mathbf{Q},\mathbf{W}) \quad &\le \quad \max_{\rho\ge 0}\Big\{\rho\log_2 B - (1+\rho) \\ &\qquad \times \log_2\left[\lambda_{\frac{1}{1+\rho}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{1+\rho}}(\mathbf{P}_\mathbf{W})\right]\Big\}. \end{aligned} \tag{2}$$

Remark: Using the first identity of Lemma 3, the upper bound can be equivalently represented as

$$\begin{aligned} &E_J(\mathbf{Q},\mathbf{W}) \\ \le\; &\max_{\rho\ge 0}\left\{\rho\left[\log_2 B - R_t\mathcal{R}_{\frac{1}{1+\rho}}(\mathbf{Q}) - \mathcal{R}_{\frac{1}{1+\rho}}(\mathbf{P}_\mathbf{W})\right]\right\} \end{aligned}$$

where $\mathcal{R}_{\frac{1}{1+\rho}}(\mathbf{Q})$ and $\mathcal{R}_{\frac{1}{1+\rho}}(\mathbf{P}_\mathbf{W})$ are the Rényi entropy rates of $\mathbf{Q}$ and $\mathbf{P}_\mathbf{W}$, respectively.

Sketch of Proof of Theorem 2: We introduce an auxiliary function

$$f(\rho) \triangleq R_t H\left(\widetilde{\mathbf{Q}}_{\frac{1}{1+\rho}}\right) + H\left(\widetilde{\mathbf{P}}_{\mathbf{W}_{\frac{1}{1+\rho}}}\right),$$

which is continuous and increasing in $\rho \ge 0$. We assume that

$$\lim_{\rho\to\infty} f(\rho) = R_t\log_2\lambda_0(\mathbf{Q}) + \log_2\lambda_0(\mathbf{P}_\mathbf{W}) > \log_2 B$$

such that

$$\max_{\rho\ge 0}\left\{\rho\log_2 B - (1+\rho)\log_2\left[\lambda_{\frac{1}{1+\rho}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{1+\rho}}(\mathbf{P}_\mathbf{W})\right]\right\}$$

is finite; otherwise the upper bound is trivial. Noting that $f(0) \le \log_2 B$ provided that $R_t H(\mathbf{Q}) \le C(\mathbf{W})$, we conclude that there must exist some $\rho^* \in [0, +\infty)$ such that $f(\rho^*) = \log_2 B + \varepsilon$, where $\varepsilon > 0$ is small enough. For the SEM source $\mathbf{Q}$, we introduce an artificial SEM source $\widetilde{\mathbf{Q}}_{\alpha^*}$ (as described in Section III) such that $\alpha^* \triangleq \frac{1}{1+\rho^*} \in (0,1]$. For the SEM channel $\mathbf{W}$, we introduce an artificial additive channel $\mathbf{V}$ for which the corresponding SEM noise is $\widetilde{\mathbf{P}}_{\mathbf{W}_{\alpha^*}}$.

Based on the construction of the artificial SEM source-channel pair $(\widetilde{\mathbf{Q}}_{\alpha^*}, \mathbf{V})$, we define for some δ_1 ($\delta_1 > 0$) the set

$$\begin{aligned} \widetilde{A}_\mathbf{s} \quad \triangleq \quad &\left\{\mathbf{y} : \log_2\frac{W^{(n)}(\mathbf{y}|f_n(\mathbf{s}))Q^{(k(n))}(\mathbf{s})}{V^{(n)}(\mathbf{y}|f_n(\mathbf{s}))\widetilde{Q}_\alpha^{(k(n))}(\mathbf{s})}\right. \\ &\ge -n\left(\frac{1-\alpha^*}{\alpha^*}(\log_2 B + \varepsilon)\right. \\ &\left.\left.-\frac{1}{\alpha^*}\log_2\left[\lambda_{\alpha^*}^{R_t}(\mathbf{Q})\lambda_{\alpha^*}(\mathbf{P}_\mathbf{W})\right] + \delta_1\right)\right|\mathbf{s}\right\}. \end{aligned}$$

We then have a lower bound for the average probability of error

$$\begin{aligned} &P_e^{(n)}(Q^{(k(n))}, W^n) \\ \ge\; &\sum_\mathbf{s} Q^{(k(n))}(\mathbf{s}) \sum_{\mathbf{y}\in A_\mathbf{s}^c\cap\widetilde{A}_\mathbf{s}} W^{(n)}(\mathbf{y}|f_n(\mathbf{s})) \\ \ge\; &2^{-n\left(\frac{1-\alpha^*}{\alpha^*}(\log_2 B+\varepsilon) - \frac{1}{\alpha^*}\log_2\left[\lambda_{\alpha^*}^{R_t}(\mathbf{Q})\lambda_{\alpha^*}(\mathbf{P}_\mathbf{W})\right]+\delta_1\right)} \\ &\times\sum_\mathbf{s}\widetilde{Q}_\alpha^{(k(n))}(\mathbf{s})V^{(n)}(\mathbf{y}\in A_\mathbf{s}^c\cap\widetilde{A}_\mathbf{s}|f_n(\mathbf{s})), \end{aligned} \tag{3}$$

where the last sum can be lower bounded as follows

$$\begin{aligned}&\sum_{\mathbf{s}}\widetilde{Q}_{\alpha}^{(k(n))}(\mathbf{s})V^{(n)}(\mathbf{y}\in A_{\mathbf{s}}^{c}\cap\widetilde{A}_{\mathbf{s}}|f_n(\mathbf{s}))\\ \geq\;&\sum_{\mathbf{s}}\widetilde{Q}_{\alpha}^{(k(n))}(\mathbf{s})V^{(n)}(\mathbf{y}\in A_{\mathbf{s}}^{c}|f_n(\mathbf{s}))\\ &\quad-\sum_{\mathbf{s}}\widetilde{Q}_{\alpha}^{(k(n))}(\mathbf{s})V^{(n)}(\mathbf{y}\in \widetilde{A}_{\mathbf{s}}^{c}|f_n(\mathbf{s})). \qquad (4)\end{aligned}$$

We point out that the first sum in the right hand side (RHS) of (4) is exactly the error probability of the joint source-channel system consisting of the artificial SEM source $\widetilde{\mathbf{Q}}_{\alpha^*}$ and the artificial SEM channel $\mathbf{V}$. Since by definition $f(\rho^*) > \log_2 B$, which implies

$$R_t H(\widetilde{\mathbf{Q}}_{\alpha^*}) > \log_2 B - H(\widetilde{\mathbf{P}}_{\mathbf{W}_{\alpha^*}}) = C(\mathbf{V}),$$

then applying the strong converse JSCC Theorem (Theorem 1) to $\widetilde{\mathbf{Q}}_{\alpha^*}$ and $\mathbf{V}$, the first sum in the RHS of (4) converges to 1 as n goes to infinity. On the other hand, according to Lemma 2, it can be shown that the second sum in the RHS of (4) vanishes as n goes to infinity. On account of these facts along with (3), and noting that ε and δ_1 are arbitrary, we obtain that

$$\begin{aligned}&\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(Q^{(k(n))}, W^{(n)})\\ \leq\;& \frac{1-\alpha^*}{\alpha^*}\log_2 B - \frac{1}{\alpha^*}\log_2\left[\lambda_{\alpha^*}^{R_t}(\mathbf{Q})\lambda_{\alpha^*}(\mathbf{P_W})\right].\end{aligned}$$

Finally, replacing α^* by $\frac{1}{1+\rho^*}$ in the above RHS term and taking the maximum over ρ^* complete the proof. □

We next introduce Gallager's random-coding lower bound for $E_J(\mathbf{Q},\mathbf{W})$, and specialize it for SEM source-channel pairs by using Lemma 3.

Proposition 1: [5, Problem 5.16] The JSCC error exponent $E_J(\mathbf{Q},\mathbf{W})$ for a discrete source $\mathbf{Q}$ and a discrete channel $\mathbf{W}$ with transmission rate R_t admits the following lower bound

$$E_J(\mathbf{Q},\mathbf{W}) \geq \max_{0\leq\rho\leq 1} E(\rho), \qquad (5)$$

where $E(\rho) \triangleq E_o(\rho) - R_t E_s(\rho)$, in which

$$E_s(\rho) \triangleq \limsup_{k(n)\to\infty}\frac{(1+\rho)}{k(n)}\log_2\sum_{\mathbf{s}\in\mathcal{S}^n} Q^{(k(n))}(\mathbf{s})^{\frac{1}{1+\rho}} \qquad (6)$$

is Gallager's source function for $\mathbf{Q}$ and

$$E_o(\rho) \triangleq \liminf_{n\to\infty}\sup_{P_{X^n}}\frac{1}{n}E_o(\rho, P_{X^n}) \qquad (7)$$

with

$$E_o(\rho, P_{X^n}) \triangleq -\log_2\sum_{\mathbf{y}\in\mathcal{Y}^n}\left(\sum_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x})W^{(n)}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}$$

is Gallager's channel function for $\mathbf{W}$.

The proof of Proposition 1 follows from the random-coding argument [5], and we thereby call this bound the random-coding lower bound for $E_J(\mathbf{Q},\mathbf{W})$. We remark that this bound is suitable for arbitrary discrete source and channel pairs with memory. It can be seen that the function $E(\rho)$ is concave over $\rho \geq 0$ since $E_o(\rho)$ and $-R_t E_s(\rho)$ are both concave functions of ρ. It can also be easily verified that $E(0) = 0$ and

$$\lim_{\rho\downarrow 0^+}\frac{\partial}{\partial\rho}E(\rho) \geq C(\mathbf{W}) - R_t H(\mathbf{Q}), \qquad (8)$$

where $H(\mathbf{Q})$ is the entropy rate of $\mathbf{Q}$ and $C(\mathbf{W})$ is the channel capacity of $\mathbf{W}$. This implies that Gallager's random-coding lower bound given in (5) is positive if $R_t H(\mathbf{Q}) < C(\mathbf{W})$ for arbitrary source-channel pairs. Particularly, when the channel is symmetric (in the Gallager sense [5]), which directly applies to channels with additive noise, the maximum in (7) is achieved by the uniform distribution: $P_{X^n}(\mathbf{x}) = 1/|\mathcal{X}|^n$ for all $\mathbf{x}\in\mathcal{X}^n$. Thus for our (modulo B) additive noise channels, $E_o(\rho)$ reduces to

$$E_o(\rho) = \rho\log_2 B - \limsup_{n\to\infty}\frac{(1+\rho)}{n}\log_2\left(\sum_{\mathbf{z}\in\mathcal{Z}^n} P_W^{(n)}(\mathbf{z})^{\frac{1}{1+\rho}}\right).$$

Consequently, it follows by Lemma 3 that for our SEM source-channel pair,

$$\begin{aligned}E_J(\mathbf{Q},\mathbf{W}) \;\geq\; &\max_{0\leq\rho\leq 1}\Big\{\rho\log_2 B - (1+\rho)\\ &\times\log_2\left[\lambda_{\frac{1}{1+\rho}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{1+\rho}}(\mathbf{P}_W)\right]\Big\}. \qquad (9)\end{aligned}$$

Note that for an SEM source-channel pair $(\mathbf{Q},\mathbf{W})$, the lower bound (9) is positive if and only if $R_t H(\mathbf{Q}) < C(\mathbf{W})$ since the limsup and liminf in (6) and (7) become limits and equality holds in (8).

Comparing (9) and (2), we obtain the following results regarding the computation of $E_J(\mathbf{Q},\mathbf{W})$.

Theorem 3: For an SEM source $\mathbf{Q}$ and an SEM channel $\mathbf{W}$ with noise $\mathbf{P_W}$ such that $R_t H(\mathbf{Q}) + H(\mathbf{P_W}) < \log_2 B$, $E_J(\mathbf{Q},\mathbf{W})$ is positive and determined exactly by

$$E_J(\mathbf{Q},\mathbf{W}) = \rho^*\log_2 B - (1+\rho^*)\log_2[\lambda_{\frac{1}{1+\rho^*}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{1+\rho^*}}(\mathbf{P_W})]$$

if $\rho^* \leq 1$, where ρ^* satisfies the equation

$$R_t H\left(\widetilde{\mathbf{Q}}_{\frac{1}{1+\rho^*}}\right) + H\left(\widetilde{\mathbf{P}}_{\mathbf{W}_{\frac{1}{1+\rho^*}}}\right) = \log_2 B.$$

Otherwise (if $\rho^* > 1$), the following bounds hold

$$E_J(\mathbf{Q},\mathbf{W}) \leq \rho^*\log_2 B - (1+\rho^*)\log_2[\lambda_{\frac{1}{1+\rho^*}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{1+\rho^*}}(\mathbf{P_W})],$$

and

$$E_J(\mathbf{Q},\mathbf{W}) \geq \log_2 B - 2\log_2\left[\lambda_{\frac{1}{2}}^{R_t}(\mathbf{Q})\lambda_{\frac{1}{2}}(\mathbf{P_W})\right].$$

Remark: For a source-channel pair ($\mathbf{Q}$,$\mathbf{W}$) with $R_t H(\mathbf{Q}) + H(\mathbf{P_W}) \geq \log_2 B$, $E_J(\mathbf{Q},\mathbf{W}) = 0$.

To illustrate Theorem 3, we consider the simple example of a binary SEM source $\mathbf{Q}$ and a binary SEM channel $\mathbf{W}$ with transmission rate $R_t = 1$, both with symmetric transition matrices given by

$$Q = \begin{bmatrix} q & 1-q \\ 1-q & q\end{bmatrix} \quad\text{and}\quad P_W = \begin{bmatrix} p & 1-p \\ 1-p & p\end{bmatrix},$$

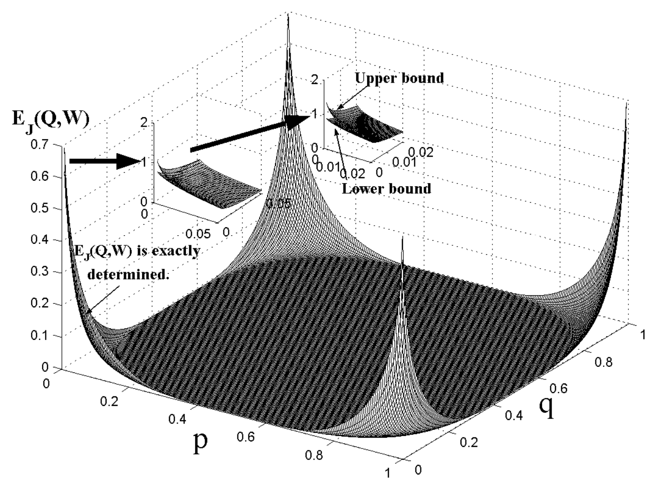

Fig. 1. Binary SEM source and binary SEM channel.

such that $0 < p, q < 1$. The upper and lower bounds for $E_J(\mathbf{Q}, \mathbf{W})$ are plotted as a function of parameters p and q in Fig. 1. We note that for this source-channel pair, the bounds are tight for a large class of (p, q) pairs. Only when p or q is extremely close to 0 or 1, is $E_J(\mathbf{Q}, \mathbf{W})$ not exactly known.

One may next ask if the lower and upper bounds for SEM source-channel pairs enjoy a form that is similar to Csiszár's bounds for DMS-DMC pairs, which are expressed as the minimum of the sum of the source error exponent and the lower/upper bound of the channel error exponent. The answer is indeed affirmative. To elucidate this point, let us first introduce the following quantities for a discrete source-channel pair ($\mathbf{Q}$,$\mathbf{W}$).

$$\widehat{e}(R, \mathbf{Q}) \triangleq \liminf_{n\to\infty} \frac{1}{n} \min_{P^{(n)}: \frac{1}{n}H(P^{(n)}) \geq R} D(P^{(n)} \parallel Q^{(n)}),$$

where $D(\cdot \parallel \cdot)$ denotes the Kullback-Leibler divergence,

$$E_{sp}(R, \mathbf{W}) \triangleq \liminf_{n\to\infty} \frac{1}{n} \max_{P_{X^n}} \min_{V^{(n)}} \left[D(V^{(n)} \parallel W^{(n)}|P_{X^n}) : \frac{1}{n} I(P_{X^n}; V^{(n)}) \leq R \right],$$

and

$$E_r(R, \mathbf{W}) \triangleq \liminf_{n\to\infty} \frac{1}{n} \max_{P_{X^n}} \min_{V^{(n)}} \left(D(V^{(n)} \parallel W^{(n)}|P_{X^n}) + \left| \frac{1}{n} I(P_{X^n}; V^{(n)}) - R \right|^+ \right),$$

where $|x|^+ \triangleq \max\{x, 0\}$ and $D(\cdot \parallel \cdot|P_{X^n})$ denotes the conditional divergence given P_{X^n}.

Theorem 4: The following equivalent representations hold

$$\min_R \left[R_t \widehat{e}\left(\frac{R}{R_t}, \mathbf{Q}\right) + E_{sp}(R, \mathbf{W}) \right] = \max_{\rho \geq 0} E(\rho), \quad (10)$$

$$\min_R \left[R_t \widehat{e}\left(\frac{R}{R_t}, \mathbf{Q}\right) + E_r(R, \mathbf{W}) \right] = \max_{0 \leq \rho \leq 1} E(\rho). \quad (11)$$

The proof of (10) and (11) is based on the Fenchel-Legendre Duality Theorem [7], [15]. We remark that Theorem 4 applies to any discrete source-channel pairs. When the source $\mathbf{Q}$ and channel $\mathbf{W}$ are discrete memoryless, the left hand sides of (10) and (11) reduce to Csiszár's lower and upper bounds for $E_J(Q, W)$ [3]. For the SEM source $\mathbf{Q}$ and the SEM channel $\mathbf{W}$, we note that the lower bound (9) and the upper bound (2) are exactly the same as the right hand sides of (10) and (11), respectively.

V. Conclusion

In this work, we establish the sphere-packing upper bound for the JSCC error exponent $E_J(\mathbf{Q}, \mathbf{W})$ of SEM source-channel systems. We also examine Gallager's random-coding lower bound for $E_J(\mathbf{Q}, \mathbf{W})$ for the same systems. We hence are able to investigate the analytical computation of $E_J(\mathbf{Q}, \mathbf{W})$ for a large class of SEM source-channel pairs. In [16], we apply these results to study the advantages of JSCC over traditional tandem coding by providing a systematic comparison of $E_J(\mathbf{Q}, \mathbf{W})$ with the tandem coding exponent $E_T(\mathbf{Q}, \mathbf{W})$ for systems with Markovian memory. We obtain sufficient conditions for which $E_J(\mathbf{Q}, \mathbf{W}) > E_T(\mathbf{Q}, \mathbf{W})$, which are satisfied by many SEM source-channel pairs. Finally, note that our results directly carry over for SEM source-channel pairs of arbitrary Markovian order.

References

[1] E. A. Arutyunyan, "Lower bound for error probability in channels with feedback," *Probl. Contr. Inform. Theory*, vol. 13, no. 2, pp. 36-44, 1977.
[2] P. Billingsley, *Ergodic Theory and Information*, Huntington, New York, 1978.
[3] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inform. Theory*, vol. 9, pp. 315–328, 1980.
[4] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Trans. Inform. Theory*, vol. 28, pp. 823–828, Nov. 1982.
[5] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.
[6] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2004.
[7] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1969.
[8] Z. Rached, F. Alajaji and L. L. Campbell, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1553–1561, May. 2001.
[9] E. Seneta, *Non-negative Matrices and Markov Chains*, New York: Springer-Verlag, 1981.
[10] Karol Vašek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, pp. 318-329, Nov. 1980.
[11] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
[12] L. Zhong, F. Alajaji, and G. Takahara, "A queue-based model for wireless Rayleigh fading channels with memory," *Proc. of IEEE 62nd Semi-annual Vehicular Technology Conference*, Dallas, TX, USA, Sept. 2005.
[13] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the computation of the joint source-channel error exponent for memoryless systems," *Proc. 2004 IEEE Int. Symp. Inform. Theory*, p. 477, June-July, 2004.
[14] Y. Zhong, F. Alajaji, and L. L. Campbell, "When is joint source-channel coding worthwhile: an information theoretic perspective," *Proc. 22nd Bienn. Symp. Commun.*, Canada, pp. 121-123 June 2004.
[15] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems: computations and applications," submitted to *IEEE Trans. Inform. Theory*, Oct. 2004.
[16] Y. Zhong, F. Alajaji, and L. L. Campbell, "When does joint source-channel coding outperform tandem coding for systems with Markovian memory?" *Proc. 2005 Canadian Workshop on Information Theory*, June 2005.