Joint Source–Channel Coding Error Exponent for Discrete Communication Systems With Markovian Memory

Yangfan Zhong, *Student Member, IEEE*, Fady Alajaji, *Senior Member, IEEE*, and L. Lorne Campbell, *Fellow, IEEE*

***Abstract*—We study the error exponent, E_J, for reliably transmitting a discrete stationary ergodic Markov (SEM) source Q over a discrete channel W with additive SEM noise via a joint source–channel (JSC) code. We first establish an upper bound for E_J in terms of the Rényi entropy rates of the source and noise processes. We next investigate the analytical computation of E_J by comparing our bound with Gallager's lower bound (1968) when the latter one is specialized to the SEM source–channel system. We also note that both bounds can be represented in Csiszár's form (1980), as the minimum of the sum of the source and channel error exponents. Our results provide us with the tools to systematically compare E_J with the tandem (separate) coding exponent E_T. We show that as in the case of memoryless source–channel pairs $E_J \leq 2E_T$ and we provide explicit conditions for which $E_J > E_T$. Numerical results indicate that $E_J \approx 2E_T$ for many SEM source–channel pairs, hence illustrating a substantial advantage of JSC coding over tandem coding for systems with Markovian memory.**

***Index Terms*—Additive noise, error probability, error exponent, joint source-channel (JSC) coding, Markov types, Rényi entropy rate, stationary ergodic Markov source–channel, tandem separate coding.**

I. Introduction

The lossless joint source–channel (JSC) coding error exponent, E_J, for a discrete memoryless source (DMS) Q and a discrete memoryless channel (DMC) W with transmission rate t was thoroughly studied in [5], [6], [10], [26]. In [5], [6], Csiszár establishes two lower bounds and an upper bound for E_J based on the random-coding and expurgated lower bounds and the sphere-packing upper bound for the DMC error exponent. In [26], we investigate the analytical computation of Csiszár's lower and upper bounds for E_J using Fenchel duality, and we provide equivalent expressions for these bounds. As a result, we are able to systematically compare the JSC coding error exponent with the traditional tandem coding error exponent E_T, the exponent resulting from separately performing and concatenating optimal source and channel coding. We show that JSC coding can double the error exponent vis-a-vis tandem coding by proving that $E_J \leq 2E_T$. Our numerical results also indicate that E_J can be nearly twice as large as E_T for many DMS-DMC pairs, hence illustrating the considerable gain that JSC coding can potentially achieve over tandem coding. It is also shown in [26] 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 JSC coding error exponent for systems with memory. Furthermore, the determination of the JSC coding 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 JSC coding techniques that fully exploit the source–channel memory. In this paper, we investigate the JSC coding error exponent for a discrete communication system with Markovian memory. Specifically, we establish a (computable) upper bound for E_J for transmitting a stationary ergodic (irreducible) Markov (SEM) source $\boldsymbol{Q}$ over a channel $\boldsymbol{W}$ with additive SEM noise $P_{\boldsymbol{W}}$ (for the sake of brevity, we hereafter refer to this channel as the SEM channel $\boldsymbol{W}$). Note that Markov sources are widely used to model realistic data sources, and binary SEM channels can approximate well binary input hard-decision demodulated fading channels with memory (e.g., see [16], [24], [25]). The proof of the bound, which follows the standard lower bounding technique for the average probability of error, is based on the judicious construction from the original SEM source–channel pair ($\boldsymbol{Q}$, $\boldsymbol{W}$) of an artificial[1] Markov source $\widetilde{\boldsymbol{Q}}_{\alpha^*}$ and an artificial channel $\boldsymbol{V}$ with additive Markov noise $\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha^*}}$, where α^* is a parameter to be optimized, such that the stationarity and ergodicity properties are retained by $\widetilde{\boldsymbol{Q}}_{\alpha^*}$ and $\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha^*}}$. The proof then employs the strong converse JSC coding Theorem[2] 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

Manuscript received May 1, 2006; revised July 12, 2007. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and the Premier's Research Excellence Award of Ontario. The material in this paper was presented in part at the Canadian Workshop on Information Theory , Montreal, QC, Canada, June 2005 and the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

The authors are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: yangfan@mast.queensu.ca; fady@mast.queensu.ca; campbell@mast.queensu.can).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2007.909092

[1]The notion of artificial (or auxiliary) Markov sources is herein adopted from [21], where Vašek employed it to study the source coding error exponent for ergodic Markov sources. However, it should be pointed out that the auxiliary source concept was first introduced by Csiszár and Longo in [4] for the memoryless case.

[2]The idea of using a strong converse coding theorem for error exponents was first initiated by Haroutunian in [12], where a strong converse channel coding theorem is used to bound the channel error exponent.

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 JSC coding error exponent for systems with Markovian memory has not been addressed before.

We also examine Gallager's lower bound for E_J [10, Problem 5.16] (which is valid for arbitrary source–channel pairs with memory), when specialized to the SEM source–channel system. By comparing our upper bound with Gallager's lower bound, we provide the condition under which they coincide, hence exactly determining E_J. We note that this condition holds for a large class of SEM source–channel pairs. Using a Fenchel-duality-based approach as in [26], we provide equivalent representations for these bounds. We show that our upper bound (respectively, Gallager's lower bound) to E_J, can also be represented by the minimum of the sum of SEM source error exponent and the upper (respectively, lower) bound of SEM channel error exponent. In this regard, our result is a natural extension of Csiszár's bounds [5] from the case of memoryless systems to the case of SEM systems.

Next, we focus our interests on the comparison of the JSC coding error exponent E_J with the tandem coding error exponent E_T under the same transmission rate. As in [26], which considers the JSC coding error exponent for discrete memoryless systems, we investigate the situation where $E_J > E_T$ for the same SEM source–channel pair. Indeed, as pointed out in [26], this inequality, when it holds, provides a theoretical underpinning and justification for JSC coding design as opposed to the widely used classical tandem or separate coding approach, since the former method provides a faster exponential rate of decay for the error probability, which often translates into improved performance and substantial reductions in complexity/delay for real-world applications. We prove that $E_J \le 2E_T$ and establish sufficient conditions for which $E_J > E_T$. We observe via numerical examples that such conditions are satisfied by a wide class of SEM source–channel pairs. Furthermore, numerical results indicate that E_J is nearly twice as large as E_T for many SEM source–channel pairs.

The rest of the paper is organized as follows. In Section II, we present preliminaries on the JSC coding error exponent and information rates for systems with memory. Some relevant results involving Markov sources and their artificial counterparts are given in Section III. In Section IV, we derive an upper bound for E_J for SEM source–channel pairs and study the computation of E_J by comparing our bound with Gallager's lower bound. Section V is devoted to a systematic comparison of E_J and E_T, and sufficient conditions for which $E_J > E_T$ are provided. In Section VI, we extend our results to SEM systems with arbitrary Markovian orders and we give an example for a system consisting of a SEM source and the queue-based channel with memory introduced in [24]. We close with concluding remarks in Section VII.

II. System Description and Definitions

A. System

We consider throughout this paper a communication system with transmission rate t (source symbols/channel use) consisting of a discrete source with finite alphabet $\mathcal{S}$ described by the sequence of tn-dimensional distributions $\boldsymbol{Q} \triangleq \left\{Q^{(tn)} : \mathcal{S}^{tn}\right\}_{tn=1}^{\infty}$, and a discrete channel described by the sequence of n-dimensional transition distributions $\boldsymbol{W} \triangleq \{W^{(n)} : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$ with common input and output alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, \ldots, B-1\}$. Given a fixed $t > 0$, a JSC code with block length n and transmission rate t is a pair of mappings: $f_n : \mathcal{S}^{tn} \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{tn}$.

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

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

where Y_i, X_i, and Z_i are the channel's output, input, and noise symbols at time i, and $Z_i \in \mathcal{Z} = \{0, 1, \ldots, B-1\}$ is independent of X_i, $i = 1, 2, \ldots, n$.

Denote the transmitted source message by $\boldsymbol{s} \triangleq (s_1, s_2, \ldots, s_{tn}) \in \mathcal{S}^{tn}$, the corresponding n-length codeword by $f_n(\boldsymbol{s}) = \boldsymbol{x} \triangleq (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, and the received codeword at the channel output by $\boldsymbol{y} \triangleq (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$. Denote $Y^n \triangleq (Y_1, Y_2, \ldots, Y_n)$ and $S^{tn} \triangleq (S_1, S_2, \ldots, S_{tn})$ as the random vectors in $\mathcal{Y}^n$ and $\mathcal{S}^{tn}$, respectively. The probability of receiving $\boldsymbol{y}$ under the conditions that the message $\boldsymbol{s}$ is transmitted (i.e., the input codeword is $f_n(\boldsymbol{s}) = \boldsymbol{x}$) is given by

$$\begin{aligned}\Pr(Y^n = \boldsymbol{y} | S^{tn} = \boldsymbol{s}) &= W^{(n)}(\boldsymbol{y} | f_n(\boldsymbol{s})) \\ &= W^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = W^{(n)}(\boldsymbol{y} \ominus \boldsymbol{x} | \boldsymbol{x}) \\ &= P_W^{(n)}(\boldsymbol{z})\end{aligned}$$

where the last equality follows by the independence of input codeword $\boldsymbol{x}$ and the additive noise $\boldsymbol{z} = \boldsymbol{y} \ominus \boldsymbol{x}$, noting that $\ominus$ is modulo-B subtraction here. The decoding operation φ_n is the rule decoding on a set of nonintersecting sets of output words $A_{\boldsymbol{s}}$ such that $\bigcup_{\boldsymbol{s}} A_{\boldsymbol{s}} = \mathcal{Y}^n$. If $\boldsymbol{y} \in A_{\boldsymbol{s}'}$, then we conclude that the source message $\boldsymbol{s}'$ has been transmitted. If the source message $\boldsymbol{s}$ has been transmitted, the conditional error probability in decoding is given by

$$\Pr(Y^n \in A_{\boldsymbol{s}}^c | S^{tn} = \boldsymbol{s}) \triangleq \sum_{\boldsymbol{y} \in A_{\boldsymbol{s}}^c} W^{(n)}\left(\boldsymbol{y} | f_n(\boldsymbol{s})\right)$$

where $A_{\boldsymbol{s}}^c = \mathcal{Y}^n - A_{\boldsymbol{s}}$, and the probability of error of the code (f_n, φ_n) is

$$P_e^{(n)}(\boldsymbol{Q}, \boldsymbol{W}, t) = \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) \sum_{\boldsymbol{y} \in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y} | f_n(\boldsymbol{s})). \quad (1)$$

B. Error Exponent and Information Rates

Roughly speaking, the error exponent E is a number with the property that the probability of decoding error is approximately 2^{-En} for codes of large block length n. The formal definition of the JSC coding error exponent is given by the following.

Definition 1: The JSC coding error exponent $E_J(\boldsymbol{Q}, \boldsymbol{W}, t)$ for source $\boldsymbol{Q}$ and channel $\boldsymbol{W}$ is defined as the supremum of all numbers E for which there exists a sequence of JSC codes (f_n, φ_n) with transmission rate t block length n such that

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

When there is no possibility of confusion, $E_J(\boldsymbol{Q}, \boldsymbol{W}, t)$ will be written as E_J (as in Section I). A lower bound for E_J for arbitrary discrete source–channel pairs with memory was already obtained by Gallager [10]. In Section IV, we establish an upper bound for E_J for SEM source–channel pairs. For a discrete source $\boldsymbol{Q}$, its (lim sup) entropy rate is defined by

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

where $H(Q^{(k)})$ is the Shannon entropy of $Q^{(k)}$; $H(\boldsymbol{Q})$ admits an operational meaning (in the sense of the lossless fixed length source coding theorem) if $\boldsymbol{Q}$ is information stable [11]. The source Rényi entropy rate of order $\alpha (\alpha \geq 0)$ is defined by

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

where

$$H_\alpha(Q^{(k)}) \triangleq \frac{1}{1-\alpha} \log_2 \sum_{\boldsymbol{s}\in\mathcal{S}^k : Q^{(k)}(\boldsymbol{s})>0} Q^{(k)}(\boldsymbol{s})^\alpha$$

is the Rényi entropy of $Q^{(k)}$, and the special case of $\alpha = 1$ should be interpreted as

$$H_1(Q^{(k)}) \triangleq \lim_{\alpha\to 1} \frac{1}{1-\alpha} \log_2 \sum_{\boldsymbol{s}\in\mathcal{S}^k : Q^{(k)}(\boldsymbol{s})>0} Q^{(k)}(\boldsymbol{s})^\alpha = H(Q^{(k)}).$$

The channel capacity for any discrete (information stable [11], [23]) channel $\boldsymbol{W}$ is given by

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

where $I(\cdot;\cdot)$ denotes mutual information. For discrete channels with finite-input finite-output alphabets, the supremum is achievable and can be replaced by maximum. If the channel $\boldsymbol{W}$ is an additive noise channel with noise process $\boldsymbol{P_W}$, then

$$C(\boldsymbol{W}) = \log_2 B - H(\boldsymbol{P_W})$$

where $H(\boldsymbol{P_W})$ is the noise entropy rate.

III. Markov Sources and Artificial Markov Sources

Without loss of generality, we consider first-order Markov sources since any Lth-order Markov source can be converted to a first-order Markov source by L-step blocking it (see Section VI). 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, $\boldsymbol{P} \triangleq \left\{p^{(n)} : \mathcal{U}^n\right\}_{n=1}^\infty$ to denote a first-order SEM source with finite alphabet $\mathcal{U} \triangleq \{1, 2, \ldots, M\}$, initial distribution

$$p_i \triangleq \Pr\{U_1 = i\}, \qquad i \in \mathcal{U}$$

and transition distribution

$$p_{ij} \triangleq \Pr\{U_{k+1} = j | U_k = i\}, \qquad i, j \in \mathcal{U}$$

so that the n-tuple probability is given by

$$p^{(n)}(i^n) \triangleq \Pr\{U_1 = i_1, \ldots, U_n = i_n\} = p_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}, \qquad i_1, \ldots, i_n \in \mathcal{U}.$$

Denote the transition (stochastic) matrix by $P \triangleq [p_{ij}]_{M\times M}$, we then set

$$P(\alpha) \triangleq \left[p_{ij}^\alpha\right]_{M\times M} \qquad (0 \leq \alpha \leq 1)$$

which is nonnegative and irreducible (here we define $0^0 = 0$). The Perron–Frobenius theorem [18] asserts that the matrix $P(\alpha)$ possesses a maximal positive eigenvalue $\lambda_\alpha(\boldsymbol{P})$ with positive (right) eigenvector $\boldsymbol{v}(\alpha) = (v_1(\alpha), \ldots, v_M(\alpha))^T$ such that $\sum_i v_i(\alpha) = 1$, where $(\cdot)^T$ denotes transposition. As in [21], we define the artificial Markov source $\widetilde{\boldsymbol{P}}_\alpha \triangleq \{\widetilde{p}_\alpha^{(n)} : \mathcal{U}^n\}_{n=1}^\infty$ with respect to the original source $\boldsymbol{P}$ such that the transition matrix is $\widetilde{P}(\alpha) \triangleq [\widetilde{p}_{ij}(\alpha)]_{M\times M}$, where

$$\widetilde{p}_{ij}(\alpha) \triangleq \frac{p_{ij}^\alpha v_j(\alpha)}{\lambda_\alpha(\boldsymbol{P}) v_i(\alpha)}. \tag{2}$$

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

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

where $\boldsymbol{\pi}(\alpha) \triangleq (\pi(\alpha)_1, \pi(\alpha)_2, \ldots, \pi(\alpha)_M)$ is the stationary distribution of the stochastic matrix $\widetilde{P}(\alpha)$. We call the artificial Markov source with initial distribution $\boldsymbol{\pi}(\alpha)$ the artificial SEM source. It is known [21, Lemmas 2.1–2.4] that $H(\widetilde{\boldsymbol{P}}_\alpha)$ is a continuous and nonincreasing function of $\alpha \in [0, 1]$. In particular, $H(\widetilde{\boldsymbol{P}}_0) = \log_2 \lambda_0(\boldsymbol{P})$ and $H(\widetilde{\boldsymbol{P}}_1) = H(\boldsymbol{P})$. The following lemma illustrates the relation between $H(\widetilde{\boldsymbol{P}}_0)$ and the entropy of the DMS with uniform distribution $\left(\frac{1}{M}, \ldots, \frac{1}{M}\right)$.

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

The following properties regarding the artificial SEM source are important in deriving the upper and lower bounds for the JSC coding exponent of SEM source–channel pairs.

Lemma 2: Let $\{U_i\}_{i=1}^\infty$ be a SEM source under $\boldsymbol{P}$ and $\widetilde{\boldsymbol{P}}_\alpha (0 < \alpha \leq 1)$, then

$$\frac{1}{n} \log_2 \frac{\widetilde{p}_\alpha^{(n)}(U^n)}{p^{(n)}(U^n)} \longrightarrow \frac{1-\alpha}{\alpha} H(\widetilde{\boldsymbol{P}}_\alpha) - \frac{1}{\alpha} \log_2 \lambda_\alpha(\boldsymbol{P})$$

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

Lemma 3: [17], [21] For a SEM source $\boldsymbol{P}$ and any $\rho \geq 0$, we have

$$\rho \mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{P}) = (1+\rho) \log_2 \lambda_{\frac{1}{1+\rho}}(\boldsymbol{P})$$

and

$$H\left(\widetilde{\boldsymbol{P}}_{\frac{1}{1+\rho}}\right) = \frac{\partial}{\partial \rho}(1+\rho)\log_2 \lambda_{\frac{1}{1+\rho}}(\boldsymbol{P}).$$

The proofs of Lemmas 1 and 2 are given in Appendix I and II. Lemma 3 follows directly from [17, Lemma 1] and [21, Lemma 2.3]. Note that there is a slight error in the expression of $H(\alpha)$ in [21, Lemma 2.3, eq. (2.11)], where a factor α is missing in the second term of the right-hand side of the equation.

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

We first prove a strong converse JSC coding theorem for ergodic sources and channels with additive ergodic noise; no Markov assumption for either the source or the channel is needed for this result.

Theorem 1: (Strong converse JSC coding Theorem) For a source $\boldsymbol{Q}$ and a channel $\boldsymbol{W}$ with additive noise $\boldsymbol{P_W}$ such that $\boldsymbol{Q}$ and $\boldsymbol{P_W}$ are ergodic processes, if $C(\boldsymbol{W}) = \log_2 B - H(\boldsymbol{P_W}) < tH(\boldsymbol{Q})$, then $\lim_{n\to\infty} P_e^{(n)}(\boldsymbol{Q}, \boldsymbol{W}, t) = 1$.

Proof: Assume $C(\boldsymbol{W}) = tH(\boldsymbol{Q}) - \varepsilon (\varepsilon > 0)$. We first recall the fact that for additive channels the channel capacity $C(\boldsymbol{W})$ is achieved by the uniform input distribution $\widehat{P}_{X^n}(\boldsymbol{x}) \triangleq 1/B^n$. Furthermore, this uniform input distribution yields a uniform distribution at the output

$$\widehat{P}_{Y^n}(\boldsymbol{y}) \triangleq \sum_{\boldsymbol{x}\in\mathcal{X}^n} \widehat{P}_{X^n}(\boldsymbol{x}) W^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{B^n}.$$

Define for some $\delta (0 < \delta < \varepsilon)$

$$\widehat{A}_{\boldsymbol{s}} = \left\{ \boldsymbol{y} : \log_2 \frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{\widehat{P}_{Y^n}(\boldsymbol{y})} \le n(C(\boldsymbol{W}) - tH(\boldsymbol{Q}) + \delta) \right\}.$$

Considering that

$$P_e^{(n)}(\boldsymbol{Q}, \boldsymbol{W}, t) = 1 - \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in A_{\boldsymbol{s}} | f_n(\boldsymbol{s})) \quad (3)$$

we need to show that $\sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in A_{\boldsymbol{s}} | f_n(\boldsymbol{s}))$ vanishes as n goes to infinity. Note that

$$\begin{aligned}\sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in A_{\boldsymbol{s}} | f_n(\boldsymbol{s})) &\le \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in A_{\boldsymbol{s}} \cap \widehat{A}_{\boldsymbol{s}} | f_n(\boldsymbol{s})) \\ &\quad + \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in \widehat{A}_{\boldsymbol{s}}^c | f_n(\boldsymbol{s})).\end{aligned}$$

For the first sum, we have

$$\begin{aligned}\sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}(\boldsymbol{y} \in A_{\boldsymbol{s}} \cap \widehat{A}_{\boldsymbol{s}} | f_n(\boldsymbol{s})) &= \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) \sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}\cap\widehat{A}_{\boldsymbol{s}}} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s})) \\ &\le \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) \sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}\cap\widehat{A}_{\boldsymbol{s}}} \frac{\widehat{P}_{Y^n}(\boldsymbol{y})}{Q^{(tn)}(\boldsymbol{s})} 2^{n(C(\boldsymbol{W})-tH(\boldsymbol{Q})+\delta)} \\ &\le 2^{n(C(\boldsymbol{W})-tH(\boldsymbol{Q})+\delta)} \sum_{\boldsymbol{s}} \sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}} \widehat{P}_{Y^n}(\boldsymbol{y}) \\ &= 2^{-n(\varepsilon-\delta)}.\end{aligned} \quad (4)$$

For the second sum, we have (5) and (6) at the bottom of the page, where $P_{Q^{(tn)}W^{(n)}}$ denotes the probability measure under the joint distribution $Q^{(tn)}(\boldsymbol{s})W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))$, and (5) follows from the fact that $P_W^{(n)}(\boldsymbol{z}) = W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))$. It follows from the well-known Shannon–McMillan–Breiman theorem for ergodic processes [1] that the above probabilities converge to 0 as n goes to infinity. On account of (4), (6), and (3), the proof is complete. □

We next establish an upper bound for E_J for SEM source–channel pairs $(\boldsymbol{Q}, \boldsymbol{W})$. Before we proceed, we define the following function for an SEM source-channel pair:

$$F(\rho) \triangleq \rho \log_2 B - (1+\rho)\log_2 \left[\lambda^t_{\frac{1}{1+\rho}}(\boldsymbol{Q}) \lambda_{\frac{1}{1+\rho}}(\boldsymbol{P_W})\right] \quad \rho \ge 0. \quad (7)$$

Lemma 4: $F(\rho)$ has the following properties.

$$\begin{aligned}\sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) W^{(n)}\left(\boldsymbol{y} \in \widehat{A}_{\boldsymbol{s}}^c | f_n(\boldsymbol{s})\right) &= P_{Q^{(tn)}W^{(n)}} \left\{ (\boldsymbol{s}, \boldsymbol{y}) : \frac{1}{n}\log_2 \frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{\widehat{P}_{Y^n}(\boldsymbol{y})} - (C(\boldsymbol{W}) - tH(\boldsymbol{Q})) > \delta \right\} \\ &\le P_{Q^{(tn)}W^{(n)}} \left\{ (\boldsymbol{s}, \boldsymbol{y}) : \left| \frac{1}{n}\log_2 \frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{\widehat{P}_{Y^n}(\boldsymbol{y})} - (C(\boldsymbol{W}) - tH(\boldsymbol{Q})) \right| > \delta \right\} \\ &= P_{Q^{(tn)}P_W^{(n)}} \left\{ (\boldsymbol{s}, \boldsymbol{z}) : \left| -\frac{1}{n}\log_2 P_W^{(n)}(\boldsymbol{z}) - \frac{1}{n}\log_2 Q^{(tn)}(\boldsymbol{s}) - H(\boldsymbol{P_W}) - tH(\boldsymbol{Q}) \right| > \delta \right\}\end{aligned} \quad (5)$$

$$\begin{aligned}&\le P_{Q^{(tn)}} \left\{ \boldsymbol{s} : \left| -\frac{1}{tn}\log_2 Q^{(tn)}(\boldsymbol{s}) - H(\boldsymbol{Q}) \right| > \frac{\delta}{2t} \right\} \\ &\quad + P_{P_W^{(n)}} \left\{ \boldsymbol{z} : \left| -\frac{1}{n}\log_2 P_W^{(n)}(\boldsymbol{z}) - H(\boldsymbol{P_W}) \right| > \frac{\delta}{2} \right\}\end{aligned} \quad (6)$$

(a) $F(0) = 0$ and

$$f(\rho) \triangleq \frac{\partial}{\partial\rho}F(\rho) = \log_2 B - \left(tH\left(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho}}\right) + H\left(\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\frac{1}{1+\rho}}}\right)\right) \quad (8)$$

is continuous non-increasing in ρ.

(b) $F(\rho)$ is concave in ρ; hence, every local maximum (stationary point) of $F(\cdot)$ is the global maximum.

(c) $\sup_{\rho\geq 0} F(\rho)$ is positive if and only if $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$; otherwise, $\sup_{\rho\geq 0} F(\rho) = 0$.

(d) $\sup_{\rho\geq 0} F(\rho)$ is finite if $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$ and infinite if $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) < B$.

Remark 1: If $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) \geq B$, then

$$\sup_{\rho\geq 0} F(\rho) = \lim_{\rho\to\infty} F(\rho)$$

no matter whether the limit is finite or not.

Proof: We start from (a). $F(0) = 0$ since the largest eigenvalue for any stochastic matrix is 1. Equation (8) follows from Lemma 3. $f(\rho)$ is a continuous nonincreasing function since $H(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho}})$ and $H(\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\frac{1}{1+\rho}}})$ are both continuous nondecreasing functions. (b) follows immediately from (a). (c) follows from the concavity of $F(\rho)$ and the facts that $F(0) = 0$ and that $f(0) = C(\boldsymbol{W}) - tH(\boldsymbol{Q})$. (d) follows from the concavity of $F(\rho)$ and the facts that $F(0) = 0$ and that

$$\lim_{\rho\to\infty} f(\rho) = \log_2 B - \log_2\left[\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W})\right]. \qquad \square$$

Theorem 2: For a SEM source $\boldsymbol{Q}$ and a discrete channel $\boldsymbol{W}$ with additive SEM noise $\boldsymbol{P_W}$ such that $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$ and $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$, the JSC coding error exponent $E_J(\boldsymbol{Q},\boldsymbol{W},t)$ satisfies

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) \leq \max_{\rho\geq 0} F(\rho). \quad (9)$$

Remark 2: We point out that the condition $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$ holds for most cases of interest. First note that the eigenvalues $\lambda_0(\boldsymbol{Q})$ and $\lambda_0(\boldsymbol{P_W})$ are no less than 1. By Lemma 1, we have that $\lambda_0(\boldsymbol{P_W}) = B$ if the noise transition matrix P_W has positive entries (i.e., $P_W > [0]_{B\times B}$); in that case, the condition $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$ is satisfied if $\lambda_0^t(\boldsymbol{Q}) > 1$ (i.e., if the source transition matrix Q is not a deterministic matrix). In fact, when $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) < B$, $\max_{\rho\geq 0} F(\rho) = +\infty$ by Lemma 4 (d), and hence it gives a trivial upper bound for E_J. When $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) = B$, we do not have an upper bound for E_J.

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

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) \leq \max_{\rho\geq 0}\left\{\rho\left[\log_2 B - t\mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{Q}) - \mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{P_W})\right]\right\}$$

where $\mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{Q})$ and $\mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{P_W})$ are the Rényi entropy rates of $\boldsymbol{Q}$ and $\boldsymbol{P_W}$, respectively. Meanwhile, the upper bound (9) holds for any one of the following source–channel pairs: DMS Q and SEM channel $\boldsymbol{W}$, SEM source $\boldsymbol{Q}$ and additive DMC W, and DMS Q and additive DMC W (note that the more general cases of DMS Q and arbitrary DMC W are investigated in [26]), all with finite alphabets. For example, when the source is DMS with distribution $\boldsymbol{q} \triangleq \{q_1, q_2, \ldots, q_M\}$ such that $q_i > 0$ for all $i = 1, 2, \ldots, M$, the source could be regarded as a SEM source $\boldsymbol{Q}$ with transition matrix

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_M \\ q_1 & q_2 & \cdots & q_M \\ \vdots & \vdots & \vdots & \vdots \\ q_1 & q_2 & \cdots & q_M \end{bmatrix}$$

and initial distribution $\boldsymbol{q}$. It is easy to verify that for such a Q, the eigenvalue $\lambda_{\frac{1}{1+\rho}}(Q)$ reduces to $\lambda_{\frac{1}{1+\rho}}(\boldsymbol{Q}) = \sum_i q_i^{1/1+\rho}$, which agrees with the results for memoryless systems given in [26]. Thus, the above bound is a sphere-packing-type upper bound for E_J for SEM source–channel systems.

Proof of Theorem 2: Under the assumption $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$ and $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$, it follows from Lemma 4 that $f(0) > 0$ and $\lim_{\rho\to\infty} f(\rho) < 0$. Since $f(\rho)$ is continuous and nonincreasing, there must exist some $\rho_o \in (0, +\infty)$ such that $f(\rho_o) + \varepsilon = 0$, where $\varepsilon > 0$ is small enough. For the SEM source $\boldsymbol{Q}$, we introduce an artificial SEM source $\widetilde{\boldsymbol{Q}}_{\alpha_o}$ (as described in Section III) such that $\alpha_o \triangleq 1/(1+\rho_o) \in (0,1)$. For the SEM channel $\boldsymbol{W}$, we introduce an artificial additive channel $\boldsymbol{V}$ for which the corresponding SEM noise is $\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha_o}}$.

Based on the construction of the artificial SEM source–channel pair $(\widetilde{\boldsymbol{Q}}_{\alpha_o}, \boldsymbol{V})$, we define for some $\delta_1 (\delta_1 > 0)$ the set

$$\widetilde{A}_{\boldsymbol{s}} = \left\{\boldsymbol{y} : \log_2 \frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{V^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\widetilde{Q}_{\alpha_o}^{(tn)}(\boldsymbol{s})} \geq -n\left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B + \varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda_{\alpha_o}^t(\boldsymbol{Q})\lambda_{\alpha_o}(\boldsymbol{P_W})\right] + \delta_1\right)\right\}$$

where we set $\widetilde{A}_{\boldsymbol{s}} = \emptyset$ for those $\boldsymbol{s}$ such that

$$W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s}) = 0, \quad \text{for some } \boldsymbol{y} \in \mathcal{Y}^n.$$

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

$$\begin{aligned} P_e^{(n)}(\boldsymbol{Q},\boldsymbol{W},t) &\geq \sum_{\boldsymbol{s}} Q^{(tn)}(\boldsymbol{s}) \sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}^c\cap\widetilde{A}_{\boldsymbol{s}}} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s})) \\ &\geq 2^{-n\left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B+\varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda_{\alpha_o}^t(\boldsymbol{Q})\lambda_{\alpha_o}(\boldsymbol{P_W})\right]+\delta_1\right)} \\ &\quad \times \sum_{\boldsymbol{s}} \widetilde{Q}_{\alpha_o}^{(tn)}(\boldsymbol{s}) V^{(n)}\left(\boldsymbol{y} \in A_{\boldsymbol{s}}^c \cap \widetilde{A}_{\boldsymbol{s}}|f_n(\boldsymbol{s})\right) \end{aligned} \quad (10)$$

where the last sum can be lower-bounded as follows:

$$\begin{aligned} &\sum_{\boldsymbol{s}} \widetilde{Q}_{\alpha_o}^{(tn)}(\boldsymbol{s}) V^{(n)}\left(\boldsymbol{y} \in A_{\boldsymbol{s}}^c \cap \widetilde{A}_{\boldsymbol{s}}|f_n(\boldsymbol{s})\right) \\ &\quad \geq \sum_{\boldsymbol{s}} \widetilde{Q}_{\alpha_o}^{(tn)}(\boldsymbol{s}) V^{(n)}\left(\boldsymbol{y} \in A_{\boldsymbol{s}}^c|f_n(\boldsymbol{s})\right) \\ &\qquad - \sum_{\boldsymbol{s}} \widetilde{Q}_{\alpha_o}^{(tn)}(\boldsymbol{s}) V^{(n)}\left(\boldsymbol{y} \in \widetilde{A}_{\boldsymbol{s}}^c|f_n(\boldsymbol{s})\right). \end{aligned} \quad (11)$$

We point out that the first sum in the right-hand side of (11) is exactly the error probability of the JSC system consisting of the artificial SEM source $\widetilde{\boldsymbol{Q}}_{\alpha_o}$ and the artificial SEM channel $\boldsymbol{V}$. Since by definition $f(\rho_o) < 0$, which implies

$$tH(\widetilde{\boldsymbol{Q}}_{\alpha_o}) > \log_2 B - H(\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha_o}}) = C(\boldsymbol{V}),$$

then applying the strong converse JSC coding theorem (Theorem 1) to $\widetilde{\boldsymbol{Q}}_{\alpha_o}$ and $\boldsymbol{V}$, the first sum in the right-hand side of (11) converges to 1 as n goes to infinity. We next show that the second term in the right-hand side of (11) vanishes asymptotically (see (12) and (13) at the bottom of the page), where $P_{\widetilde{Q}^{(tn)}_{\alpha_o}V^{(n)}}$ denotes the probability measure under the joint distribution $\widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})V^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))$, and (12) follows from the facts that

$$P_W^{(n)}(\boldsymbol{z}) = W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s})) \text{ and that } \widetilde{P}^{(n)}_{W_{\alpha_o}}(\boldsymbol{z}) = V^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s})).$$

Applying Lemma 2, the above probabilities in (13) converge to 0 as $n \to \infty$.[3] On account of (10), (11), and (13), and noting that ε and δ_1 are arbitrary, we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(Q^{(tn)}, W^{(n)}) \le \frac{1-\alpha_o}{\alpha_o}\log_2 B - \frac{1}{\alpha_o}\log_2\left[\lambda^t_{\alpha_o}(\boldsymbol{Q})\lambda_{\alpha_o}(\boldsymbol{P_W})\right].$$

Finally, replacing α_o by $1/(1+\rho_o)$ in the above right-hand side terms and taking the maximum over ρ_o completes the proof. □

[3]Convergence almost surely implies convergence in probability.

We next introduce Gallager's lower bound for E_J and specialize it for SEM source–channel pairs by using Lemma 3.

Proposition 1: [10, Problem 5.16] The JSC coding error exponent $E_J(\boldsymbol{Q}, \boldsymbol{W}, t)$ for a discrete source $\boldsymbol{Q}$ and a discrete channel $\boldsymbol{W}$ with transmission rate t admits the following lower bound:

$$E_J(\boldsymbol{Q}, \boldsymbol{W}, t) \ge \max_{0\le\rho\le 1} E(\rho) \tag{14}$$

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

$$E_s(\rho) \triangleq \limsup_{tn\to\infty} \frac{(1+\rho)}{tn}\log_2 \sum_{\boldsymbol{s}\in\mathcal{S}^{tn}} Q^{(tn)}(\boldsymbol{s})^{\frac{1}{1+\rho}} \tag{15}$$

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

$$E_o(\rho) \triangleq \liminf_{n\to\infty} \max_{P_{X^n}} \frac{1}{n} E_o(\rho, P_{X^n}) \tag{16}$$

with

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

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

We remark that this bound is suitable for arbitrary discrete source–channel pairs with memory. Particularly, when the channel is symmetric (in the Gallager sense [10]), which directly applies to channels with additive noise, the maximum in (16) is achieved by the uniform distribution: $P_{X^n}(\boldsymbol{x}) = 1/|\mathcal{X}|^n$

$$\begin{aligned}
\sum_{\boldsymbol{s}} \widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})V^{(n)}\left(\boldsymbol{y}\in \widetilde{A}^c_{\boldsymbol{s}}|f_n(\boldsymbol{s})\right) &= P_{\widetilde{Q}^{(tn)}_{\alpha_o}V^{(n)}}\Bigg\{(\boldsymbol{s},\boldsymbol{y}) : \frac{1}{n}\log_2\frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{V^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})} \\
&\qquad + \left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B+\varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda^t_{\alpha_o}(\boldsymbol{Q})\lambda_{\alpha_o}(\boldsymbol{P_W})\right]\right) < -\delta_1\Bigg\} \\
&\le P_{\widetilde{Q}^{(tn)}_{\alpha_o}V^{(n)}}\Bigg\{(\boldsymbol{s},\boldsymbol{y}) : \Bigg|\frac{1}{n}\log_2\frac{W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))Q^{(tn)}(\boldsymbol{s})}{V^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})} \\
&\qquad + \left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B+\varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda^t_{\alpha_o}(\boldsymbol{Q})\lambda_{\alpha_o}(\boldsymbol{P_W})\right]\right)\Bigg| > \delta_1\Bigg\} \\
&= P_{\widetilde{Q}^{(tn)}_{\alpha_o}\widetilde{P}^{(n)}_{W_{\alpha_o}}}\Bigg\{(\boldsymbol{s},\boldsymbol{z}) : \Bigg|\frac{1}{n}\log_2\frac{\widetilde{P}^{(n)}_{W_{\alpha_o}}(\boldsymbol{z})}{P^{(n)}_W(\boldsymbol{z})} + \frac{1}{n}\log_2\frac{\widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})}{Q^{(tn)}(\boldsymbol{s})} \\
&\qquad - \Bigg[t\left(\frac{1-\alpha_o}{\alpha_o}H(\widetilde{\boldsymbol{Q}}_{\alpha_o}) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\boldsymbol{Q})\right) \\
&\qquad + \frac{1-\alpha_o}{\alpha_o}H(\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha_o}}) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\boldsymbol{P_W})\Bigg]\Bigg| > \delta_1\Bigg\}
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\le P_{\widetilde{Q}^{(tn)}_{\alpha_o}}\left\{\boldsymbol{s} : \left|\frac{1}{tn}\log_2\frac{\widetilde{Q}^{(tn)}_{\alpha_o}(\boldsymbol{s})}{Q^{(tn)}(\boldsymbol{s})} - \left[\frac{1-\alpha_o}{\alpha_o}H(\widetilde{\boldsymbol{Q}}_{\alpha_o}) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\boldsymbol{Q})\right]\right| > \frac{\delta_1}{2t}\right\} \\
&\quad + P_{\widetilde{P}^{(n)}_{W_{\alpha_o}}}\left\{\boldsymbol{z} : \left|\frac{1}{n}\log_2\frac{\widetilde{P}^{(n)}_{W_{\alpha_o}}(\boldsymbol{z})}{P^{(n)}_W(\boldsymbol{z})} - \left[\frac{1-\alpha_o}{\alpha_o}H(\widetilde{\boldsymbol{P}}_{\boldsymbol{W}_{\alpha_o}}) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\boldsymbol{P_W})\right]\right| > \frac{\delta_1}{2}\right\}
\end{aligned} \tag{13}$$

for all $\boldsymbol{x} \in \mathcal{X}^n$. Thus. for our (modulo B) additive noise channels, $E_o(\rho)$ reduces to

$$E_o(\rho) = \rho \log_2 B - \limsup_{n\to\infty} \frac{(1+\rho)}{n} \log_2 \left(\sum_{\boldsymbol{z}\in\mathcal{Z}^n} P_W^{(n)}(\boldsymbol{z})^{\frac{1}{1+\rho}} \right). \quad (17)$$

It immediately follows by Lemma 3 that for our SEM source–channel pair

$$E(\rho) = \rho \log_2 B - \rho t \mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{Q}) - \rho \mathcal{R}_{\frac{1}{1+\rho}}(\boldsymbol{P}_W) = F(\rho). \quad (18)$$

That is, the SEM source–channel function we defined in (7) is exactly the same as the difference of Gallager's channel and source function. In light of Theorem 2 and Proposition 1, we obtain the following regarding the computation of E_J.

Theorem 3: For a SEM source $\boldsymbol{Q}$ and a SEM channel $\boldsymbol{W}$ with noise $\boldsymbol{P_W}$ such that $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$ and $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$, $E_J(\boldsymbol{Q}, \boldsymbol{W}, t)$ is positive and determined exactly by $E_J(\boldsymbol{Q}, \boldsymbol{W}, t) = F(\rho^*)$ if $\rho^* \le 1$, where ρ^* is the smallest positive number satisfying the equation $f(\rho^*) = 0$. Otherwise (if $\rho^* > 1$), the following bounds hold:

$$\log_2 B - 2\log_2 \left[\lambda_{\frac{1}{2}}^t(\boldsymbol{Q})\lambda_{\frac{1}{2}}(\boldsymbol{P_W})\right] \le E_J(\boldsymbol{Q}, \boldsymbol{W}, t) \le F(\rho^*).$$

Remark 4: If $tH(\boldsymbol{Q}) \ge C(\boldsymbol{W})$, i.e., $tH(\boldsymbol{Q}) + H(\boldsymbol{P_W}) \ge \log_2 B$, then $E_J(\boldsymbol{Q}, \boldsymbol{W}, t) = 0$.

Remark 5: According to Lemma 4 (c) and (d), there must exist a positive and finite ρ^* provided that $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$ and $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$. Using Lemma 4 (a), such ρ^* can be numerically determined.

The proof of Theorem 3 directly follows from Theorem 2 and Proposition 1 and the use of Lemma 4. The following by-product results regarding the error exponents of SEM sources and SEM channels immediately follow from Theorems 1 and 2.

Corollary 1: For any rate $0 \le R < \log_2 \lambda_0(\boldsymbol{Q})$, the source error exponent $e(R, \boldsymbol{Q})$ for a SEM source $\boldsymbol{Q}$ satisfies

$$e(R, \boldsymbol{Q}) \le \overline{e}(R, \boldsymbol{Q}) \quad (19)$$

where

$$\overline{e}(R, \boldsymbol{Q}) \triangleq \sup_{\rho \ge 0} [R\rho - (1+\rho)\log_2 \lambda_{\frac{1}{1+\rho}}(\boldsymbol{Q})]. \quad (20)$$

Particularly, for $0 \le R \le H(\boldsymbol{Q})$, $\overline{e}(R, \boldsymbol{Q}) = 0$.

Note that $\log_2 \lambda_0(\boldsymbol{Q}) = \log_2 |S|$ when the source reduces to a DMS (with alphabet $\mathcal{S}$). This upper bound is exactly the same as the one given by Vašek [21]. In fact, he shows that $\overline{e}(R, \boldsymbol{Q})$ is the real source error exponent (also see [3]) for all $R \ge 0$. We point out that $\overline{e}(R, \boldsymbol{Q})$ can be equivalently expressed in terms of a constrained minimum of Kullback–Leibler divergence [15], as the error exponent for DMS [22]; also see (35) in Appendix III.

Corollary 2: For any rate $\log_2 (B/\lambda_0(\boldsymbol{P_W})) < R < \infty$, the channel error exponent $E(R, \boldsymbol{W})$ for a SEM channel $\boldsymbol{W}$ satisfies

$$E(R, \boldsymbol{W}) \le \overline{E}(R, \boldsymbol{W}) \quad (21)$$

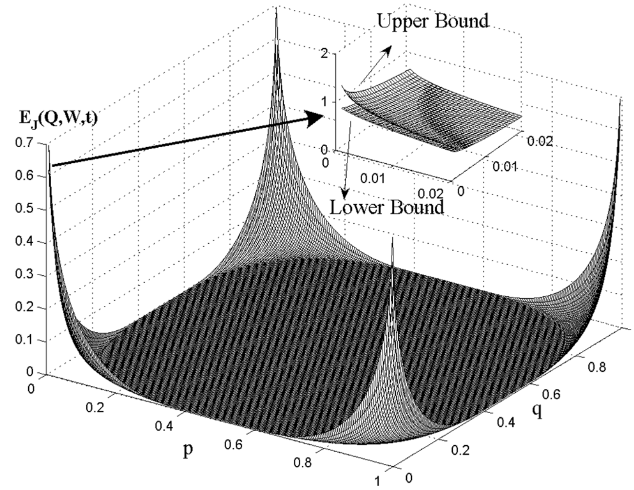

Fig. 1. The lower and upper bounds of E_J for the binary SEM source and the binary SEM channel of Example 1 with $t = 1$.

where

$$\overline{E}(R, \boldsymbol{W}) \triangleq \sup_{\rho \ge 0} \left\{ \rho(\log_2 B - R) - (1+\rho)\log_2 \lambda_{\frac{1}{1+\rho}}(\boldsymbol{P_W}) \right\}. \quad (22)$$

Particularly, for $C(\boldsymbol{W}) \le R < \infty$, $\overline{E}(R, \boldsymbol{W}) = 0$.

When the SEM channel reduces to an additive noise DMC, $\log_2 (B/\lambda_0(\boldsymbol{P_W})) = R_\infty$ [10, p. 158]. Note that the usual case (when the transition matrix is positive) is that $\log_2 (B/\lambda_0(\boldsymbol{P_W})) = 0$ (see Lemma 1). It can be shown that $\overline{E}(R, \boldsymbol{W})$ is positive, nonincreasing, and convex, and hence strictly decreasing in R. Comparing with Gallager's random-coding lower bound for $E(R, \boldsymbol{W})$[10] (when specialized for SEM channels) given by

$$E_r(R, \boldsymbol{W}) \triangleq \max_{0 \le \rho \le 1} \left\{ \rho(\log_2 B - R) - (1+\rho)\log_2 \lambda_{\frac{1}{1+\rho}}(\boldsymbol{P_W}) \right\}, \quad (23)$$

and applying the results of Section III, we note that the upper and lower bounds are equal if $R \ge R_{cr}$, where $R_{cr} \triangleq \log_2 B - H\left(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{2}}}\right)$ is the critical rate of the SEM channel. Thus, the channel error exponent for SEM channel is determined exactly for $R \ge R_{cr}$.

Example 1: We consider a system consisting of a binary SEM source $\boldsymbol{Q}$ and a binary SEM channel $\boldsymbol{W}$ with transmission rate $t = 1$, both with symmetric transition matrices given by

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

such that $0 < p, q < 1$.[4] The upper and lower bounds for $E_J(\boldsymbol{Q}, \boldsymbol{W}, t)$ are plotted as a function of parameters p and q in Fig. 1. It is observed that for this source–channel pair, the

[4]Note that P_W is not the channel probability transition matrix; it is the transition matrix of SEM channel noise.

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 not exactly known.

One may next ask if the lower and upper bounds for the SEM source–channel pair enjoy a form that is similar to Csiszár's bounds for DMS–DMC pairs [5], 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, as given in the following theorem.

Theorem 4: Let $tH(\boldsymbol{Q}) < C(\boldsymbol{W})$ and $\lambda_0^t(\boldsymbol{Q})\lambda_0(\boldsymbol{P_W}) > B$. The equivalent representations of (24) and (25) (shown at the bottom of the page) hold, where $F(\rho)$ is defined in (7), $e(R, \boldsymbol{Q}) = \overline{e}(R, \boldsymbol{Q})$ is given by (20), and $\overline{E}(R, \boldsymbol{W})$ and $E_r(R, \boldsymbol{W})$ are given by (22) and (23), respectively.

Note that we write "min" instead of "inf" in (24) and (25) because the optimizations are achievable due to the convexity of the source and channel exponents. Theorem 4 can be proved via the Lagrange multiplier method, since the functions $\log_2 \lambda_{1/(1+\rho)}(\boldsymbol{Q})$ and $\log_2 \lambda_{1/(1+\rho)}(\boldsymbol{P_W})$ are differentiable functions of ρ and their derivatives admit closed-form expressions (recall Lemma 3). Alternately, and more succinctly, we can prove (24) and (25) using Fenchel duality [14]; the reader may consult [26, Theorem 1] for details. When the source $\boldsymbol{Q}$ and channel $\boldsymbol{W}$ are discrete memoryless, the right-hand side of (24) and (25) reduce to Csiszár's lower and upper bounds for E_J[5]. In fact, Csiszár establishes the upper bound for E_J for a DMS–DMC pair (Q, W) in terms of the exact source and channel exponents $e(R, Q)$ and $E(R, W)$ [5]

$$E_J(Q, W, t) \le \min_R \left[te\left(\frac{R}{t}, Q\right) + E(R, W) \right]. \qquad (26)$$

Meanwhile, he points out that if we replace the channel exponent in (26) by its sphere-packing bound $E_{sp}(R, W)$, we can obtain a (possibly) looser but computable upper bound

$$\min_R \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right] \qquad (27)$$

which is called the sphere-packing bound to E_J by the authors in [26]. In Appendix III, we show that the bound (26) still applies for SEM source–channel pairs, i.e., E_J is upper-bounded by the minimum of the sum of the SEM source exponent $e(R, \boldsymbol{Q})$ and the SEM channel exponent $E(R, \boldsymbol{W})$, by which we prove that the JSC exponent can at most double the tandem coding exponent (see Theorem 5). However, the bound in terms of $e(R, \boldsymbol{Q})$ and $E(R, \boldsymbol{W})$, though tighter than the sphere-packing type bound (24), is not computable in general, since the behavior of the SEM channel error exponent $E(R, \boldsymbol{W})$ is unknown for rates smaller than the critical rate R_{cr}.

We point out that the parametric expressions of these bounds (the left-hand side of (24) and (25)) facilitate the computation of E_J, while the bounds in Csiszár's form are instrumental for the comparison of JSC and tandem coding exponents, a subject studied in Section V.

V. JSC Versus Tandem Coding Error Exponents

A. Tandem Coding Exponent for Systems With Memory

A tandem code $(f_n^*, \varphi_n^*) \triangleq (f_{cn} \circ f_{sn}, \varphi_{sn} \circ \varphi_{cn})$ for a discrete source $\boldsymbol{Q}$ and a discrete channel $\boldsymbol{W}$ is composed "independently" of a (tn, L_n) block source code (f_{sn}, φ_{sn}) defined by $f_{sn} : \mathcal{S}^{tn} \longrightarrow \{1, 2, \ldots, L_n\}$ and $\varphi_{sn} : \{1, 2, \ldots, L_n\} \longrightarrow \mathcal{S}^{tn}$ with source code rate

$$R_{s,n} \triangleq \frac{\log_2 L_n}{tn} \quad \text{source code bits/source symbol}$$

and an (n, L_n) block channel code (f_{cn}, φ_{cn}) defined by $f_{cn} : \{1, 2, \ldots, L_n\} \longrightarrow \mathcal{X}^n$ and $\varphi_{cn} : \mathcal{Y}^n \longrightarrow \{1, 2, \ldots, L_n\}$ with channel code rate

$$R_{c,n} \triangleq \frac{\log_2 L_n}{n} \quad \text{source code bits/channel use}$$

where "$\circ$" denotes composition, and we assume that the limit $\lim_{n\to\infty} \frac{\log_2 L_n}{n}$ exists, i.e.,

$$\liminf_{n\to\infty} \frac{\log_2 L_n}{n} = \limsup_{n\to\infty} \frac{\log_2 L_n}{n}.$$

Here "independently" means that the source code is designed without the knowledge of the channel statistics, and the channel code is designed without the knowledge of the source statistics.

The error probability of the tandem code (f_n^*, φ_n^*) is hence given by

$$P_{e^*}^{(n)}(\boldsymbol{Q}, \boldsymbol{W}, t) \triangleq \Pr\left(\varphi_{sn}\left[\varphi_{cn}(Y^n)\right] \ne S^{tn}\right). \qquad (28)$$

Definition 2: The tandem coding error exponent $E_T(\boldsymbol{Q}, \boldsymbol{W}, t)$ for source $\boldsymbol{Q}$ and channel $\boldsymbol{W}$ is defined as the supremum of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes $(f_n^*, \varphi_n^*) \triangleq (f_{cn} \circ f_{sn}, \varphi_{sn} \circ \varphi_{cn})$ with transmission rate t block length n such that

$$\widehat{E} \le \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{e^*}^{(n)}(\boldsymbol{Q}, \boldsymbol{W}, t).$$

$$\max_{\rho \ge 0} F(\rho) = \min_{\log_2(B/\lambda_0(\boldsymbol{P_W})) < R < t\log_2 \lambda_0(\boldsymbol{Q})} \left[te\left(\frac{R}{t}, \boldsymbol{Q}\right) + \overline{E}(R, \boldsymbol{W}) \right], \qquad (24)$$

$$\max_{0 \le \rho \le 1} F(\rho) = \min_{0 \le R < t\log_2 \lambda_0(\boldsymbol{Q})} \left[te\left(\frac{R}{t}, \boldsymbol{Q}\right) + E_r(R, \boldsymbol{W}) \right]. \qquad (25)$$

In the sequel, we sometimes refer to $E_T(\boldsymbol{Q},\boldsymbol{W},t)$ by E_T when there is no possibility of confusion. Since tandem coding exponent results from separately performing and concatenating optimal source and channel coding, it can be shown[5] (e.g., [5], [27]) that

$$E_T(\boldsymbol{Q},\boldsymbol{W},t)=\sup_R \min\left\{te\left(\frac{R}{t},\boldsymbol{Q}\right),E(R,\boldsymbol{W})\right\} \quad (29)$$

where $e(R,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$ are the source and channel error exponents, respectively. To evaluate E_T for a SEM source–channel pair $(\boldsymbol{Q},\boldsymbol{W})$, we recall that $e(R,\boldsymbol{Q})$ is 0 for $R\le H(\boldsymbol{Q})$, strictly increasing in $H(\boldsymbol{Q})\le R\le \log_2\lambda_0(\boldsymbol{Q})$ and infinity for $R>\log_2\lambda_0(\boldsymbol{Q})$ ([15], [21]), while $E(R,\boldsymbol{W})$ is nonincreasing and positive in $R<C(\boldsymbol{W})$, and vanishes at $R=C(\boldsymbol{W})$.

Therefore, if the graphs of $te\left(R/t,\boldsymbol{Q}\right)$ and $E(R,\boldsymbol{W})$ have an intersection at R_o, then it immediately follows from (29) that

$$E_T(\boldsymbol{Q},\boldsymbol{W},t)=te\left(\frac{R_o}{t},\boldsymbol{Q}\right)=E(R_o,\boldsymbol{W}).$$

If there is no intersection between $te\left(R/t,\boldsymbol{Q}\right)$ and $E(R,\boldsymbol{W})$ then

$$E_T(\boldsymbol{Q},\boldsymbol{W},t)=E(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{W})$$

by (29).

Note that $E_J\ge E_T$; meanwhile, $E_J=E_T=0$ if $tH(\boldsymbol{Q})\ge C(\boldsymbol{W})$ for SEM source–channel pairs. We are hence interested in determining the conditions for which $E_J>E_T$ when $tH(\boldsymbol{Q})<C(\boldsymbol{W})$. Although both E_J and E_T are not always determined, we can still provide some sufficient conditions for which $E_J>E_T$. Before we proceed, we first show that for SEM source–channel pairs, the JSC coding exponent can at most double the tandem coding exponent. Note that the same result holds for DMS–DMC pairs, as shown in [26].

Theorem 5: For a SEM source $\boldsymbol{Q}$ and a SEM channel $\boldsymbol{W}$, the JSC coding exponent is upper-bounded by twice the tandem coding exponent

$$E_J(\boldsymbol{Q},\boldsymbol{W},t)\le 2E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

To prove this result, we need two steps. The first is to establish another upper bound for E_J, as we discussed in the end of the last section, in terms of $e(R,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$ by using the technique of Markov types ([8], [9], [15]), and the second is to justify that the bound is at most equal to twice E_T. Although the approach for the first step is analogous to the one that Csiszár used for DMS–DMC pairs [5], we still give a self-contained proof in Appendix III for the sake of completeness.

B. Sufficient Conditions for Which $E_j>E_t$

When the entropy rate of the SEM source is equal to $\log_2\lambda_0(\boldsymbol{Q})$, the source error exponent would be zero for $R\le\log_2\lambda_0(\boldsymbol{Q})$ and infinity otherwise. In this case, the source is incompressible and only channel coding is performed in both JSC coding and tandem coding; as a result

$$E_J(\boldsymbol{Q},\boldsymbol{W},t)=E_T(\boldsymbol{Q},\boldsymbol{W},t)=E(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{W})$$

by (24), (25), and (29). Note that $\log_2\lambda_0(\boldsymbol{Q})$ might not be equal to $\log_2|\mathcal{S}|$ by Lemma 1, as compared with the DMS. Thus, we assume in the rest of the section that $H(\boldsymbol{Q})<\log_2\lambda_0(\boldsymbol{Q})$ (such that the source is compressible) and that $tH(\boldsymbol{Q})<C(\boldsymbol{W})$ (such that both E_J and E_T are positive). We also assume in the sequel that all the sources and channels are SEM.

Theorem 6: Let f be defined by (8). If $f(1)\le 0$, i.e.,

$$tH\left(\widetilde{\boldsymbol{Q}}_{\frac{1}{2}}\right)+H\left(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{2}}}\right)\ge\log_2 B$$

then $E_J(\boldsymbol{Q},\boldsymbol{W},t)>E_T(\boldsymbol{Q},\boldsymbol{W},t)$.

Proof: Since we assumed that $tH(\boldsymbol{Q})<C(\boldsymbol{W})$ or equivalently $f(0)>0$ (see Lemma 4), if now $f(1)\le 0$, then there exists some ρ $(0<\rho\le 1)$ such that $f(\rho)=0$ by the continuity of $f(\cdot)$. Let ρ^* be the smallest one satisfying $f(\rho^*)=0$. According to Theorem 3, the JSC coding error exponent is determined exactly by $E_J(\boldsymbol{Q},\boldsymbol{W},t)=F(\rho^*)$. On the other hand, we know from (24) that

$$F(\rho^*)=\min_{\log_2(B/\lambda_0(\boldsymbol{P_W}))<R<t\log_2\lambda_0(\boldsymbol{Q})}\left[te\left(\frac{R}{t},\boldsymbol{Q}\right)+\overline{E}(R,\boldsymbol{W})\right].$$

Suppose the above minimum is achieved by some $\overline{R}_m$, i.e.,

$$F(\rho^*)=te\left(\frac{\overline{R}_m}{t},\boldsymbol{Q}\right)+\overline{E}(\overline{R}_m,\boldsymbol{W}).$$

It can be shown (cf. [26]) that $\overline{R}_m$ is related to ρ^* as follows:

$$\overline{R}_m=tH\left(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho^*}}\right)=\log_2 B-H\left(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{1+\rho^*}}}\right).$$

Since ρ^* is positive, from the above we know $tH(\boldsymbol{Q})\le\overline{R}_m\le C(\boldsymbol{W})$ by the monotonicity of $H\left(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho}}\right)$ and $H(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{1+\rho}}})$. In the following, we first assume that $te(R/t,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$ intersect at R_o, i.e., there exists an $R_o\in(tH(\boldsymbol{Q}),C(\boldsymbol{W}))$ such that

$$E_T(\boldsymbol{Q},\boldsymbol{W},t)=te\left(\frac{R_o}{t},\boldsymbol{Q}\right)=E(R_o,\boldsymbol{W})>0.$$

If $\overline{R}_m>R_o$, then

$$E_J(\boldsymbol{Q},\boldsymbol{W},t)\ge te\left(\frac{\overline{R}_m}{t},\boldsymbol{Q}\right)>te\left(\frac{R_o}{t},\boldsymbol{Q}\right)=E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

If $\overline{R}_m=R_o$, then

$$E_J(\boldsymbol{Q},\boldsymbol{W},t)=2E_T(\boldsymbol{Q},\boldsymbol{W},t)>E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

If $\overline{R}_m<R_o$, then

$$E_J(\boldsymbol{Q},\boldsymbol{W},t)\ge\overline{E}(\overline{R}_m,\boldsymbol{W})>\overline{E}(R_o,\boldsymbol{W})=E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

We next assume that there is no intersection between $te(R/t,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$, i.e.,

$$te(R/t,\boldsymbol{Q})<E(R,\boldsymbol{W}),\ \text{for all } R<t\log_2\lambda_0(\boldsymbol{Q}).$$

[5]To prove (29), one needs to assume that the source and channel coding operations are decoupled via common randomization (by applying a randomly selected permutation map, e.g., see [13]) at their interface in both the transmitter and the receiver. This is a natural assumption needed to achieve total (statistical) separation between source and channel coding; see [27] for the details.

If $\overline{R}_m = tH(\boldsymbol{Q})$, then

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) = \overline{E}(\overline{R}_m,\boldsymbol{W}) > \overline{E}(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{Q}) = E_T(\boldsymbol{Q},\boldsymbol{W},t)$$

since $H(\boldsymbol{Q}) < \log_2\lambda_0(\boldsymbol{Q})$ is assumed. If $\overline{R}_m > tH(\boldsymbol{Q})$, then

$$\begin{aligned}E_J(\boldsymbol{Q},\boldsymbol{W},t) &\geq te\left(\frac{\overline{R}_m}{t},\boldsymbol{Q}\right) + \overline{E}(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{Q})\\ &> \overline{E}(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{Q})\\ &= E_T(\boldsymbol{Q},\boldsymbol{W},t)\end{aligned}$$

since the source error exponent is positive at $\overline{R}_m > tH(\boldsymbol{Q})$. □

Theorem 6 states that if E_J is determined exactly (i.e., its upper and lower bounds coincide), no matter whether E_T is known or not, then the JSC coding exponent is larger than the tandem exponent. Conversely, if E_T is determined exactly, irrespective of whether E_J is determined or not, the strict inequality between E_J and E_T also holds, as shown by the following results.

Theorem 7:

(a) If $tH(\boldsymbol{Q}) \geq R_{cr}$, then $E_J(\boldsymbol{Q},\boldsymbol{W},t) > E_T(\boldsymbol{Q},\boldsymbol{W},t)$.

(b) Otherwise, if $tH(\boldsymbol{Q}) < R_{cr}$ and $t\log_2\lambda_0(\boldsymbol{Q}) > R_{cr}$, there must exist some ρ satisfying $tH(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho}}) = R_{cr}$. Let ρ_m be the smallest one satisfying such equation. If

$$(1+\rho_m)t[H(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho_m}}) - \log_2\lambda_{\frac{1}{1+\rho_m}}(\boldsymbol{Q})] \leq \log_2 B - 2\log_2\lambda_{\frac{1}{2}}(\boldsymbol{P_W})$$

then $E_J(\boldsymbol{Q},\boldsymbol{W},t) > E_T(\boldsymbol{Q},\boldsymbol{W},t)$.

Remark 6: By the monotonicity of $H(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho}})$, ρ_m can be solved numerically.

Proof: Recall that $R_{cr} = \log_2 B - H\left(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{2}}}\right)$ is the critical rate of the channel $\boldsymbol{W}$ such that the channel exponent is determined for $R \geq R_{cr}$, i.e., $E(R,\boldsymbol{W}) = E_r(R,\boldsymbol{W}) = \overline{E}(R,\boldsymbol{W})$ if $R \geq R_{cr}$. We first show that $E_J > E_T$ if $te(R_{cr}/t,\boldsymbol{Q}) \leq E(R_{cr},\boldsymbol{W})$, and then we show that $te(R_{cr}/t,\boldsymbol{Q}) \leq E(R_{cr},\boldsymbol{W})$ if and only if (a) or (b) holds.

Now if $te(R_{cr}/t,\boldsymbol{Q}) \leq E(R_{cr},\boldsymbol{W})$, then $E_T(\boldsymbol{Q},\boldsymbol{W},t)$ is determined exactly. There are two cases to consider.

1) If $te(R/t,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$ intersect at R_o such that $R_{cr} \leq R_o < C(\boldsymbol{W})$, then

$$E_T(\boldsymbol{Q},\boldsymbol{W},t) = te\left(\frac{R_o}{t},\boldsymbol{Q}\right) = E_r(R_o,\boldsymbol{W}) > 0.$$

On the other hand, (17) and (18) yield

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) \geq \max_{0\leq\rho\leq 1} F(\rho) = F(\rho_*)$$

where $\rho_* = \min(1,\rho^*) > 0$ and recall that ρ^* is the smallest positive number satisfying $f(\rho^*) = 0$. It follows from (25) that

$$F(\rho_*) = \min_{0\leq R<t\log_2\lambda_0(\boldsymbol{Q})}\left[te\left(\frac{R}{t},\boldsymbol{Q}\right) + E_r(R,\boldsymbol{W})\right].$$

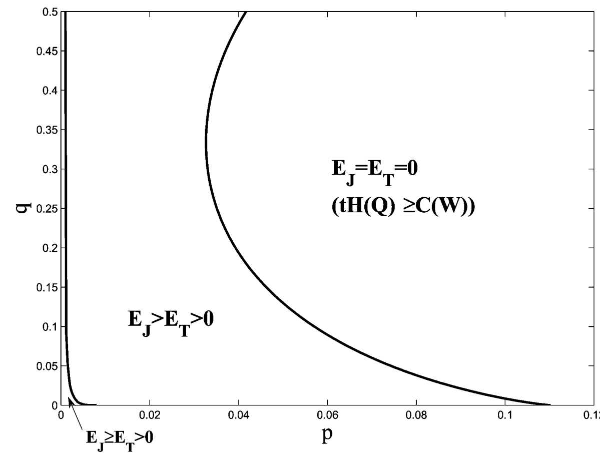

Fig. 2. The regions for the ternary SEM source and the binary SEM channel of Example 2 with $t = 0.5$.

Similar to $\overline{R}_m$ in the last proof, it can be shown (cf. [26]) that the above minimum is achieved by some $\underline{R}_m$ such that

$$\underline{R}_m = tH\left(\widetilde{\boldsymbol{Q}}_{\frac{1}{1+\rho_*}}\right) \geq tH(\boldsymbol{Q}).$$

If $\underline{R}_m > R_o$, then

$$\begin{aligned}E_J(\boldsymbol{Q},\boldsymbol{W},t) &\geq te\left(\frac{\underline{R}_m}{t},\boldsymbol{Q}\right) > te\left(\frac{R_o}{t},\boldsymbol{Q}\right)\\ &= E_T(\boldsymbol{Q},\boldsymbol{W},t).\end{aligned}$$

If $\underline{R}_m = R_o$, then

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) = 2E_T(\boldsymbol{Q},\boldsymbol{W},t) > E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

If $\underline{R}_m < R_o$, likewise, we have

$$E_J(\boldsymbol{Q},\boldsymbol{W},t) \geq E_r(\underline{R}_m,\boldsymbol{W}) > E_r(R_o,\boldsymbol{W}) = E_T(\boldsymbol{Q},\boldsymbol{W},t).$$

2) If $te(R/t,\boldsymbol{Q})$ and $E(R,\boldsymbol{W})$ have no intersection, we still have, as in the last proof, if $\underline{R}_m = tH(\boldsymbol{Q})$, then

$$\begin{aligned}E_J(\boldsymbol{Q},\boldsymbol{W},t) &\geq E_r(\underline{R}_m,\boldsymbol{W}) > E_r(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{Q})\\ &= E_T(\boldsymbol{Q},\boldsymbol{W},t);\end{aligned}$$

otherwise, if $\underline{R}_m > tH(\boldsymbol{Q})$, then

$$\begin{aligned}E_J(\boldsymbol{Q},\boldsymbol{W},t) &> E_r(\underline{R}_m,\boldsymbol{W}) \geq E_r(t\log_2\lambda_0(\boldsymbol{Q}),\boldsymbol{W})\\ &= E_T(\boldsymbol{Q},\boldsymbol{W},t).\end{aligned}$$

Finally, we point out that the sufficient and necessary conditions for $te(R_{cr}/t,\boldsymbol{Q}) \leq E(R_{cr},\boldsymbol{W})$ is that (a) $tH(\boldsymbol{Q}) \geq R_{cr}$ such that $te(R_{cr}/t,\boldsymbol{Q}) = 0$; or (b) $te(R_{cr}/t,\boldsymbol{Q}) > 0$ but $te(R_{cr}/t,\boldsymbol{Q}) \leq E(R_{cr},\boldsymbol{W})$. Using the fact that

$$E(R_{cr},\boldsymbol{W}) = H(\widetilde{\boldsymbol{P}}_{W_{\frac{1}{2}}}) - 2\log_2\lambda_{\frac{1}{2}}(\boldsymbol{P_W})$$

we obtain Condition (b) and complete the proof. □

Example 2: We next examine Theorems 6 and 7 for the following simple example. Consider a ternary SEM source $\boldsymbol{Q}$ and

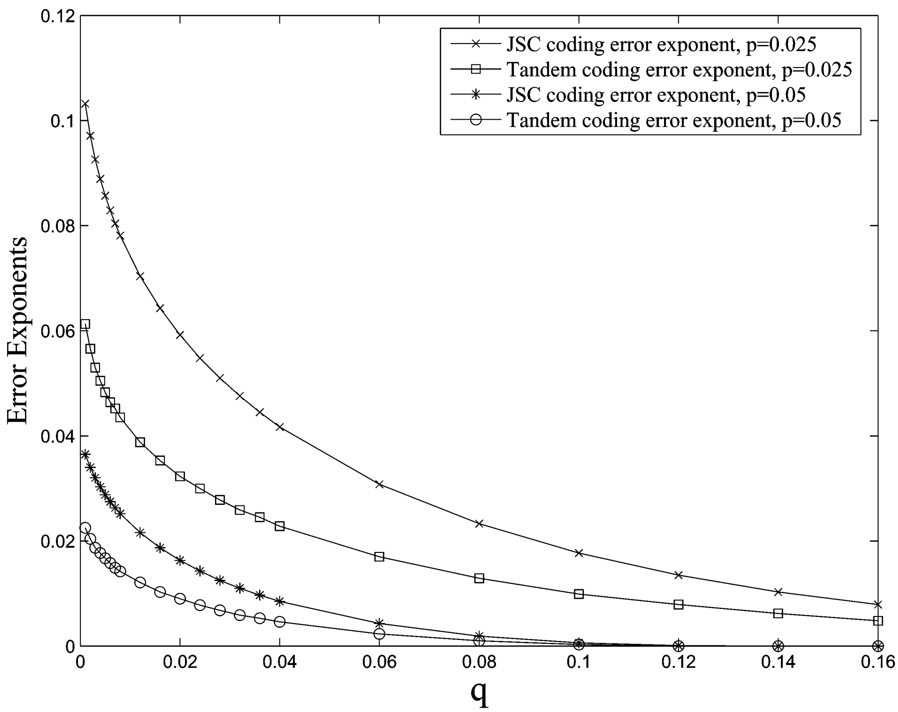

Fig. 3. Comparison of E_J and E_T for the ternary SEM source and the binary SEM channel of Example 2 with $t = 0.5$.

a binary SEM channel $\boldsymbol{W}$, both with symmetric transition matrices given by

$$Q = \begin{bmatrix} q & (1-q)/2 & (1-q)/2 \\ (1-q)/2 & q & (1-q)/2 \\ (1-q)/2 & (1-q)/2 & q \end{bmatrix} \text{ and}$$

$$P_W = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

such that $0 < p, q < 0.5$. Suppose now the transmission rate $t = 0.5$. If (q,p) satisfies any one of the conditions of Theorems 6 and 7, then $E_J(\boldsymbol{Q}, \boldsymbol{W}, t) > E_T(\boldsymbol{Q}, \boldsymbol{W}, t)$. The range for which the inequality holds is summarized in Fig. 2. For the channel with $p = 0.025$ and $p = 0.05$, we plot the JSC coding and tandem coding error exponents against the source parameter q whenever they are exactly determined, see Fig. 3. We note that for these source–channel pairs, $E_J(\boldsymbol{Q}, \boldsymbol{W})$ substantially outperforms $E_T(\boldsymbol{Q}, \boldsymbol{W})$ (indeed $E_J(\boldsymbol{Q}, \boldsymbol{W}) \approx 2E_T(\boldsymbol{Q}, \boldsymbol{W})$) for a large class of (q,p) pairs. We then plot the two exponents under the transmission rate $t = 0.75$ whenever they are determined exactly, and obtain similar results, see Fig. 4. In fact, for many other SEM source–channel pairs (not necessarily binary SEM sources or ternary SEM channels) with other transmission rates, we observe similar results; this indicates that the JSC coding exponent is strictly better than the tandem coding exponent for a wide class of SEM systems.

VI. Systems With Arbitrary Markovian Orders

Suppose that the SEM source $\{U_i\}_{n=1}^{\infty}$ with alphabet $\mathcal{U}$ has a Markovian order $K_s \geq 1$. Define process $\{S_i\}_{n=1}^{\infty}$ obtained by K_s-step blocking the Markov source $\boldsymbol{P}$; i.e.,

$$S_n \triangleq (U_n, U_{n+1}, \ldots, U_{n+K_s-1}).$$

Then

$$\begin{aligned}&\Pr(S_n = j_n | S_{n-1} = j_{n-1}, \ldots, S_1 = j_1)\\ &\quad = \Pr(S_n = j_n | S_{n-1} = j_{n-1}), \quad j_1, \ldots, j_n \in \mathcal{S} = \mathcal{U}^{K_s}\end{aligned}$$

and the source $\boldsymbol{Q}$ is a first-order SEM source with $|\mathcal{U}|^{K_s}$ states. Therefore, all the results in this paper can be readily extended to SEM systems with arbitrary order by converting the K_sth-order SEM source to a first-order SEM source of larger alphabet. Also, if the additive SEM noise $\boldsymbol{P_W}$ of the channel has Markovian order $K_c \geq 1$, we can similarly convert it to a first-order SEM noise with expanded alphabet. In the following, we present an example for the system consisting of a SEM source (of order $K_s = 1$) and the queue-based channel (QBC) [24] with memory $K_c = 2$, as the QBC approximates well for a certain range of channel conditions the Gilbert–Elliott channel [24] and hard-decision-demodulated correlated fading channels [25].

Example 3 (Transmission of a SEM Source Over the QBC [24]): A QBC is a binary additive channel whose noise process $\boldsymbol{P_W} = \{p_W^{(n)} : \mathcal{Z}^n\}_{n=1}^{\infty}$ (where $\mathcal{Z} = \{0, 1\}$) is generated according to a mixture mechanism of a finite queue and a Bernoulli process [24]. At time i, the noise symbol Z_i is chosen either from the queue described by a sequence of random variables $(Q_{i,1}, \ldots, Q_{i,K_c})$ $(Q_{i,j} \in \{0,1\}, j = 1, 2, \ldots, K_c)$ with probability ε or from a Bernoulli process with probability $1 - \varepsilon$ such that

- if Z_i is chosen from the queue process, then

$$\Pr(Z_i = Q_{i,j}) = \begin{cases} 1/(K_c - 1 + \alpha), & j = 1, 2, \ldots, K_c - 1 \\ \alpha/(K_c - 1 + \alpha), & j = K_c \end{cases}$$

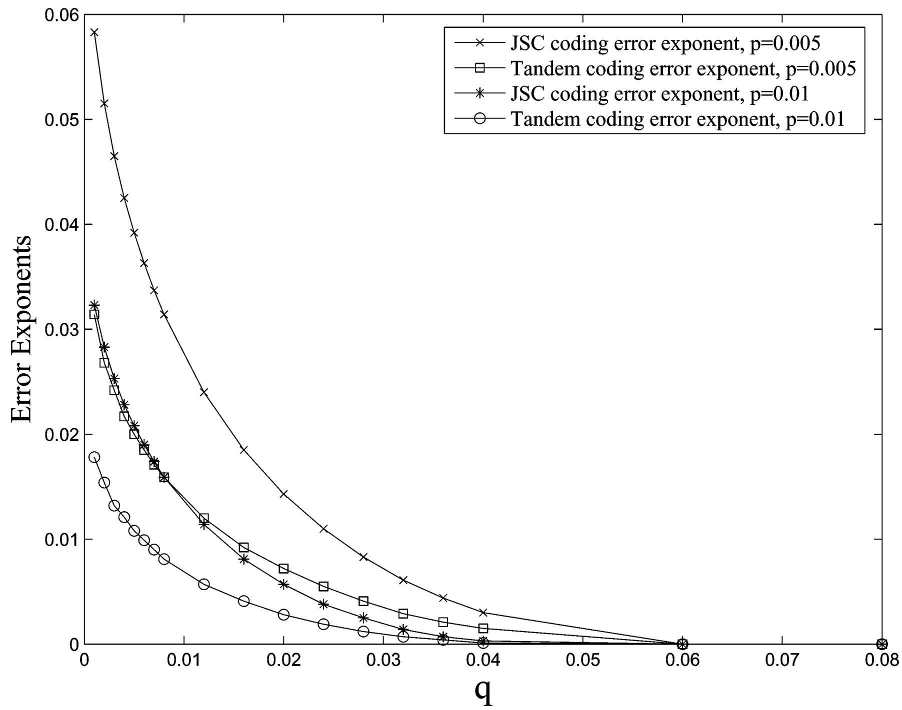

Fig. 4. Comparison of E_J and E_T for the ternary SEM source and the binary SEM channel of Example 2 with $t = 0.75$.

if $K_c > 1$ and $\alpha \geq 0$ is arbitrary; otherwise $\Pr(Z_i = Q_{i,1}) = 1$ if $K_c = 1$;

- if Z_i is chosen from the Bernoulli process, then $\Pr(Z_i = 1) = p$ $(p \ll 1/2)$ and $\Pr(Z_i = 0) = 1 - p$.

At time $i+1$, we first shift the queue from left to right by the following rule

$$(Q_{i+1,1}, \ldots, Q_{i+1,K_c}) = (Z_i, Q_{i,1}, \ldots, Q_{i,K_c-1}),$$

then we generate the noise symbol Z_{i+1} according to the same mechanism. It can be shown [24] that the QBC is actually an K_cth-order SEM channel characterized only by four parameters ε, α, p, and K_c.

Now we consider transmitting the first-order SEM source $\boldsymbol{Q}$ with transition matrix

$$Q = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.05 & 0.15 & 0.8 \end{bmatrix}$$

under transmission rate $t = 1$ over the QBC with $K_c = 2$ such that the noise process $\boldsymbol{P_W}$ is a second-order SEM process. After two-step blocking $\boldsymbol{P_W}$, we obtain a first-order SEM process $\boldsymbol{P_W^{K_c}}$ with transition matrix shown at the bottom of the page.

We next compute E_J and E_T for the ternary SEM source and the QBC given above. When $p = 0.05$, $\alpha = 1$, E_J, and E_T are both determined exactly if $\varepsilon \in [0.001, 0.992]$. We plot the two exponents by varying ε. We see from Fig. 5 that $E_J \approx 2E_T$ for all the $\varepsilon \in [0.001, 0.992]$. When we choose $p = 0.05$, $\alpha = 0.1$ for which E_J and E_T are both determined exactly if $\varepsilon \in [0.001, 0.968]$, we have similar results, see Fig. 5. It is interesting to note that when ε gets smaller, E_J and E_T approach the exponents resulting from the SEM source $\boldsymbol{Q}$ and the binary-symmetric channel (BSC) with crossover probability $p = 0.05$. This is indeed expected since the QBC reduces to the BSC when $\varepsilon = 0$[24].

VII. Concluding Remarks

In this work, we establish a computable upper bound for the JSC coding error exponent E_J of SEM source–channel systems. We also examine Gallager's lower bound for E_J for the same systems. It is shown that E_J can be exactly determined by the two bounds for a large class of SEM source–channel pairs.

As a result, we can systematically compare the JSC coding exponent with the tandem exponent for such systems with memory and study the advantages of JSC coding over the traditional tandem coding. We first show $E_J \leq 2E_T$ by deriving an upper bound for E_J in terms of the source and channel exponents. We then provide sufficient (computable) conditions for which $E_J > E_T$. Numerical results indicate that the inequality holds for most SEM source–channel pairs, and that $E_J \approx 2E_T$

$$P_W^{K_c} = \begin{bmatrix} \varepsilon + (1-\varepsilon)(1-p) & 0 & (1-\varepsilon)p & 0 \\ \frac{\varepsilon}{1+\alpha} + (1-\varepsilon)(1-p) & 0 & \frac{\varepsilon\alpha}{1+\alpha} + (1-\varepsilon)p & 0 \\ 0 & \frac{\varepsilon\alpha}{1+\alpha} + (1-\varepsilon)(1-p) & 0 & \frac{\varepsilon}{1+\alpha} + (1-\varepsilon)p \\ 0 & (1-\varepsilon)(1-p) & 0 & \varepsilon + (1-\varepsilon)p \end{bmatrix}.$$

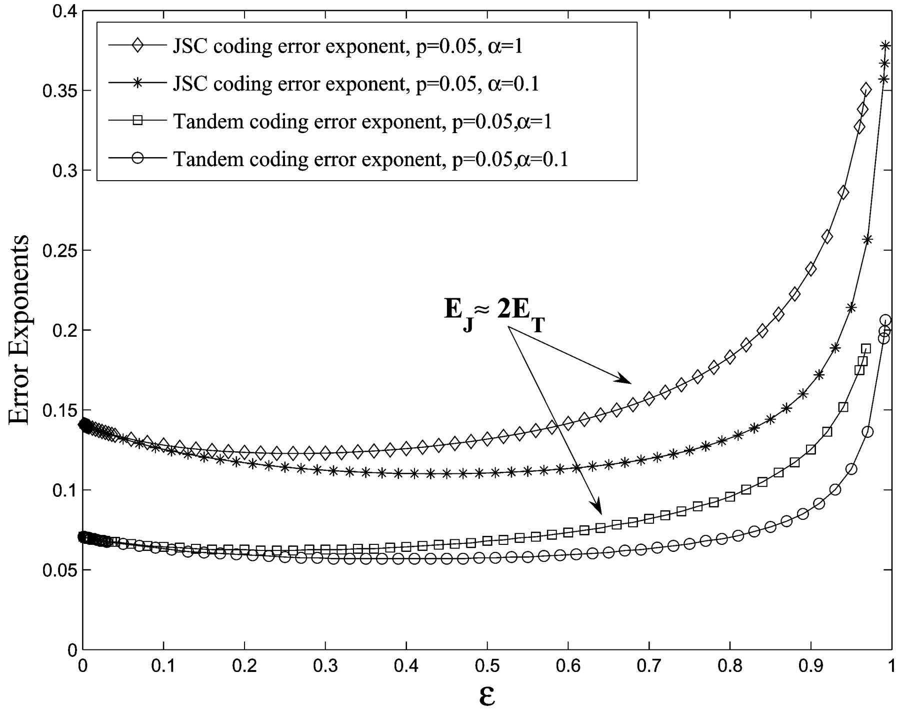

Fig. 5. Comparison of E_J and E_T for the SEM source and the QBC of Example 3 with $t = 1$.

in many cases even though E_J is upper-bounded by twice E_T, which means that for the same error probability P_e, JSC coding would require around half the delay of tandem coding, that is

$$P_e \approx 2^{-nE_T(\boldsymbol{Q},\boldsymbol{W},t)} = 2^{-\frac{n}{2}E_J(\boldsymbol{Q},\boldsymbol{W},t)}$$

for n sufficiently large. Finally, we note that our results directly carry over for SEM source–channel pairs of arbitrary Markovian order.

Appendix I
Proof of Lemma 1

Let $\mathcal{H}$ be the $M \times M$ matrix with all components equal to 1, i.e., $\mathcal{H} \triangleq_{M\times M}$. Clearly, $\boldsymbol{u} \triangleq \left[\frac{1}{M}, \ldots, \frac{1}{M}\right]$ is the unique normalized positive eigenvector (Perron vector) of $\mathcal{H}$ with associated positive eigenvalue M; thus, when $P > [0]_{M\times M}$, $\lambda_0(\boldsymbol{P}) = M$. We next show by contradiction that $\lambda_0(\boldsymbol{P}) < M$ if there are zero components in matrix P. We assume that there exist some $p_{ij} = 0$ and $\lambda_0(\boldsymbol{P}) \geq M$. Then

$$\lambda_0(\boldsymbol{P})\boldsymbol{u} \geq M\boldsymbol{u} = \mathcal{H}\boldsymbol{u} = \mathcal{H}\boldsymbol{v}(0),$$

where the last equality holds since $\boldsymbol{u}$ and $\boldsymbol{v}(0)$ are both normalized vectors. We thus have

$$(\mathcal{H} - P(0))\boldsymbol{v}(0) \leq \lambda_0(\boldsymbol{P})(\boldsymbol{u} - \boldsymbol{v}(0)).$$

Now summing all the components of the vectors on both sides, we obtain

$$\sum_{i,j} a_{ij}v_j(0) \leq 0$$

where a_{ij} is the (i,j)th component of the matrix $\mathcal{H} - P(0)$ such that $a_{ij} = 0$ if $p_{ij} > 0$ and $a_{ij} = 1$ if $p_{ij} = 0$. This contradicts with the fact that all $v_j(0)$'s are positive and thus $\lambda_0(\boldsymbol{P}) < M$ if $\boldsymbol{P}$ has zero components. We also conclude that $P > [0]_{M\times M}$ is the sufficient and necessary condition for $\lambda_0(\boldsymbol{P}) = M$. □

Appendix II
Proof of Lemma 2

Since $\{U_i\}_{i=1}^{\infty}$ is a SEM source under $\boldsymbol{P}$ and $\widetilde{\boldsymbol{P}}_\alpha$, it follows by the Ergodic Theorem [1] that the normalized log-likelihood ratio between $\boldsymbol{P}$ and $\widetilde{\boldsymbol{P}}_\alpha$ converges to their Kullback–Leibler divergence rate almost surely, i.e.,

$$\frac{1}{n}\log_2\frac{\widetilde{p}_\alpha^{(n)}(U^n)}{p^{(n)}(U^n)} \longrightarrow D(\widetilde{\boldsymbol{P}}_\alpha \parallel \boldsymbol{P})$$

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

$$D(\widetilde{\boldsymbol{P}}_\alpha \parallel \boldsymbol{P}) \triangleq \lim_{n\to\infty}\frac{1}{n}D\left(\widetilde{p}_\alpha^{(n)} \parallel p^{(n)}\right).$$

Note that for any n, we can write

$$\begin{aligned}\frac{1}{n}D(\widetilde{p}_\alpha^{(n)} \parallel p^{(n)}) = &-\frac{1}{n}H(\widetilde{p}_\alpha^{(n)})\\ &-\frac{1}{n}\sum_{i^n}\widetilde{p}_\alpha^{(n)}(i^n)\log_2 p^{(n)}(i^n),\\ &i^n = (i_1\cdots i_n) \in \mathcal{U}^n. \quad (30)\end{aligned}$$

Recalling that $\boldsymbol{P}$ is described by the initial stationary distribution $\pi = \{\pi_1, \pi_2, \ldots, \pi_M\}$ and transition matrix $P = [p_{ij}]_{M\times M}$, and that $\widetilde{\boldsymbol{P}}_\alpha$ is described by the initial stationary distribution $\boldsymbol{\pi}(\alpha) = (\pi(\alpha)_1, \pi(\alpha)_2, \ldots, \pi(\alpha)_M)$ and transition matrix $\widetilde{P}(\alpha) \triangleq [\widetilde{p}_{ij}(\alpha)]_{M\times M}$ given by (2), we have

$$\widetilde{p}_\alpha^{(n)}(i^n) = \pi(\alpha)_{i_1}\frac{p_{i_1i_2}^\alpha\cdots p_{i_{n-1}i_n}^\alpha}{\lambda_\alpha(\boldsymbol{P})^{n-1}}\frac{v_{i_n}(\alpha)}{v_{i_1}(\alpha)}$$

$$= \frac{p^{(n)}(i^n)^\alpha}{\lambda_\alpha(\boldsymbol{P})^{n-1}} \frac{\pi(\alpha)_{i_1} v_{i_n}(\alpha)}{\pi_{i_1}^\alpha v_{i_1}(\alpha)} \tag{31}$$

for all $i^n \in \mathcal{U}^n$. Consequently, using (30) and (31), we have

$$\frac{1}{n} D(\widetilde{p}_\alpha^{(n)} \parallel p^{(n)}) = \frac{1-\alpha}{\alpha} \frac{1}{n} H(\widetilde{p}_\alpha^{(n)}) - \frac{1}{\alpha} \frac{n-1}{n} \log_2 \lambda_\alpha(\boldsymbol{P}) - \frac{1}{n} \frac{1}{\alpha} \sum_{i_1, i_n} \widetilde{p}_\alpha(i_1, i_n) \log_2 \left(\frac{\pi_{i_1}^\alpha v_{i_1}(\alpha)}{\pi(\alpha)_{i_1} v_{i_n}(\alpha)} \right). \tag{32}$$

Taking the limit on both sides of (32), and noting that the last term approaches 0 since

$$\left| \frac{1}{\alpha} \sum_{i_1, i_n} \widetilde{p}_\alpha(i_1, i_n) \log_2 \left(\frac{\pi_{i_1}^\alpha v_{i_1}(\alpha)}{\pi(\alpha)_{i_1} v_{i_n}(\alpha)} \right) \right| \leq \frac{M^2}{\alpha} \max_{i_1, i_n} \left| \log_2 \left(\frac{\pi_{i_1}^\alpha v_{i_1}(\alpha)}{\pi(\alpha)_{i_1} v_{i_n}(\alpha)} \right) \right| < +\infty$$

where π, $\boldsymbol{\pi}(\alpha)$, and $\boldsymbol{v}(\alpha)$ are all positive for SEM sources (according to the Perron–Frobenius theorem [18]). We hence obtain

$$D(\widetilde{\boldsymbol{P}}_\alpha \parallel \boldsymbol{P}) = \frac{1-\alpha}{\alpha} H(\widetilde{\boldsymbol{P}}_\alpha) - \frac{1}{\alpha} \log_2 \lambda_\alpha(\boldsymbol{P}). \qquad \square$$

Appendix III
Proof of Theorem 5

Step 1: We first set up some notations and basic facts regarding Markov types adopted from [8] and [15]. Given a source sequence $\boldsymbol{s} = (s_1, s_2, \ldots, s_k) \in \mathcal{S}^k (|\mathcal{S}| = M)$, let $k_{ij}(\boldsymbol{s})$ be the number of transitions from $i \in \mathcal{S}$ to $j \in \mathcal{S}$ in $\boldsymbol{s}$ with the cyclic convention that s_1 follows s_k. We denote the matrix

$$\left[\frac{k_{ij}(\boldsymbol{s})}{k} \right]_{M \times M}$$

by $\Phi^{(k)}(\boldsymbol{s})$ and call it the Markov type (empirical matrix) of $\boldsymbol{s}$, where $\sum_{i,j} k_{ij}(\boldsymbol{s}) = k$ and it is easily seen that $\sum_j k_{ij} = \sum_j k_{ji}$ for all i. In other words, the (k-length) sequence $\boldsymbol{s}$ of type P has the empirical matrix $\Phi^{(k)}(\boldsymbol{s})$ which is equal to P. The set of all types of k-length sequences will be denoted by $\mathcal{E}_k$. Next we introduce a class of matrices that includes $\mathcal{E}_k$ for all k as a dense subset. Let

$$\mathcal{E} = \left\{ P : P = [p_{ij}]_{M \times M}, \sum_{i,j} p_{ij} = 1, \text{ and } p_{ij} \geq 0, \sum_j p_{ij} = \sum_j p_{ji} \text{ for all } i \right\}.$$

Note that $\mathcal{E}_k \to \mathcal{E}$ as $k \to \infty$ in the sense that for any $P \in \mathcal{E}$, there exists a sequence of $\{\Phi^{(k)}\} \in \mathcal{E}_k$, such that $\Phi^{(k)} \to P$ uniformly.

For $P \in \mathcal{E}$ and any $M \times M$ transition (stochastic) matrix $Q = [q_{ij}]_{M \times M}$ (such that $\sum_j q_{ij} = 1$ for all i), define

$$H_c(P) \triangleq -\sum_{i,j} p_{ij} \log_2 \frac{p_{ij}}{\sum_j p_{ij}}$$

to be the conditional entropy of P and

$$D_c(P \parallel Q) \triangleq \sum_{i,j} p_{ij} \log_2 \frac{p_{ij}}{q_{ij} \sum_j p_{ij}}$$

be the conditional divergence of P over Q. Let $P \in \mathcal{E}_k$ be a Markov type, and let $\mathcal{T}_P = \left\{ \boldsymbol{s} \in \mathcal{S}^k : \Phi^{(k)}(\boldsymbol{s}) = P \right\}$ be a Markov type class. We define

$$\mathcal{M}_P(i,j) \triangleq \{ \boldsymbol{s} = (s_1, s_2, \ldots, s_k) \in \mathcal{T}_P : s_1 = i, s_k = j \}.$$

Clearly, $\mathcal{M}_P(i,j)$ partitions the entire type class $\mathcal{T}_P$ over $(i,j) \in \mathcal{S} \times \mathcal{S}$, and all sequences in $\mathcal{M}_P(i,j)$ are equiprobable under $Q^{(k)}(\cdot)$.

Lemma 5: [8] Let $\boldsymbol{Q}$ be a first-order finite-alphabet irreducible Markov source with transition matrix $Q = [q_{ij}]_{M \times M}$ and arbitrary initial distribution $\boldsymbol{q} > 0$. Let $\alpha \triangleq \min_i q_i$. Then we have the following bounds.

1) For any $i, j \in \mathcal{S}$ and $P \in \mathcal{E}_k$ such that $\mathcal{M}_P(i,j) \neq \emptyset$, $|\mathcal{M}_P(i,j)| \geq k^{-M} (k+1)^{-M^2} 2^{k H_c(P)}$.
2) $Q^{(k)}(\mathcal{T}_P) \geq k^{-M} (k+1)^{-M^2} \alpha 2^{-k D_c(P \parallel Q)}$.

Remark 7: Note that in [8], the authors assume both irreducibility and aperiodicity for the Markov source $\boldsymbol{Q}$ and also derive an upper bound for the probability of type classes $Q^{(k)}(\mathcal{T}_P)$. Here we only need the lower bound above for $Q^{(k)}(\mathcal{T}_P)$; thus, the aperiodicity assumption is not required.

Note also that M and α are quantities independent of k, and that for SEM sources, the stationary distribution (which is the initial distribution) is unique and positive.

Step 2: Set $k = tn$. Rewrite the probability of error given in (1) as a sum of probabilities of types and lower bound it by the expression (33) at the top of the following page, where

$$P_e(\mathcal{M}_P(i,j)) \triangleq \frac{1}{|\mathcal{M}_P(i,j)|} \sum_{\boldsymbol{s} \in \mathcal{M}_P(i,j)} \sum_{\boldsymbol{y} \in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y} | f_n(\boldsymbol{s})).$$

We note that $P_e(\mathcal{M}_P(i,j))$ is actually the (average) probability of error of the n-block *channel* code (f_n, φ_n) with message set (source) $\mathcal{M}_P(i,j)$ and channel $\boldsymbol{W}$. Recall that the channel error exponent $E(R, \boldsymbol{W})$ is the largest exponential rate such that the probability of error decays to zero [7] over all channel codes of rate no larger than R. Then $P_e(\mathcal{M}_P(i,j))$ is lower-bounded by

$$P_e(\mathcal{M}_P(i,j)) \geq 2^{-nE\left(\frac{1}{n} \log_2 |\mathcal{M}_P(i,j)|, \boldsymbol{W}\right) + o(n)} \geq 2^{-nE\left(t H_c(P) - \frac{M}{n} \log_2 k - \frac{M^2}{n} \log_2 (k+1), \boldsymbol{W}\right) + o(n)}$$

where the second inequality follows from the monotonicity of $E(R, \boldsymbol{W})$ and Lemma 5 (1), and $o(n)$ is a term that tends to zero as n goes to infinity. The second equation at the top of the following page is the result of (33) and Lemma 5 (2), and it holds for any source–channel codes (f_n, φ_n), where $\alpha_\pi > 0$ denotes the smallest component in the stationary distribution, which is independent of k. Since when $n \to \infty$, $\mathcal{E}_k \to \mathcal{E}$ (recalling that $k = tn$ and t is a constant). By the definition of JSC coding error exponent, we obtain (34) also at the top of the following page . In (34), we used the facts that

$$\min_{P \in \mathcal{E} : H_c(P) \geq R/t} D_c(P \parallel Q) = \min_{P \in \mathcal{E} : H_c(P) = R/t} D_c(P \parallel Q) \tag{35}$$

$$\begin{aligned} P_e^{(n)}(\boldsymbol{Q},\boldsymbol{W},t) &= \sum_{P\in\mathcal{E}_k}\sum_{\boldsymbol{s}\in\mathcal{T}_P} Q^{(k)}(\boldsymbol{s})\sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\\ &\ge \max_{P\in\mathcal{E}_k}\sum_{\boldsymbol{s}\in\mathcal{T}_P} Q^{(k)}(\boldsymbol{s})\sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\\ &= \max_{P\in\mathcal{E}_k}\sum_{(i,j)\in\mathcal{S}\times\mathcal{S}:\mathcal{M}_P(i,j)\neq\emptyset}\ \sum_{\boldsymbol{s}\in\mathcal{M}_P(i,j)} Q^{(k)}(\boldsymbol{s})\sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\\ &= \max_{P\in\mathcal{E}_k}\sum_{(i,j)\in\mathcal{S}\times\mathcal{S}:\mathcal{M}_P(i,j)\neq\emptyset}\left(\sum_{\boldsymbol{s}'\in\mathcal{M}_P(i,j)} Q^{(k)}(\boldsymbol{s}')\right)\sum_{\boldsymbol{s}\in\mathcal{M}(i,j)}\frac{Q^{(k)}(\boldsymbol{s})}{\sum_{\boldsymbol{s}'\in\mathcal{M}_P(i,j)} Q^{(k)}(\boldsymbol{s}')}\sum_{\boldsymbol{y}\in A_{\boldsymbol{s}}^c} W^{(n)}(\boldsymbol{y}|f_n(\boldsymbol{s}))\\ &= \max_{P\in\mathcal{E}_k}\sum_{(i,j)\in\mathcal{S}\times\mathcal{S}:\mathcal{M}_P(i,j)\neq\emptyset}\ \sum_{\boldsymbol{s}'\in\mathcal{M}_P(i,j)} Q^{(k)}(\boldsymbol{s}')P_e(\mathcal{M}_P(i,j)) \end{aligned} \tag{33}$$

$$\begin{aligned} P_e^{(n)}(\boldsymbol{Q},\boldsymbol{W},t) &\ge \max_{P\in\mathcal{E}_k} Q^{(k)}(\mathcal{T}_P)2^{-nE\left(tH_c(P)-\frac{M}{n}\log_2 k-\frac{M^2}{n}\log_2(k+1),\boldsymbol{W}\right)+o(n)}\\ &\ge \max_{P\in\mathcal{E}_k} k^{-M}(k+1)^{-M^2}\alpha_\pi 2^{-kD_c(P\|Q)}2^{-nE\left(tH_c(P)-\frac{M}{n}\log_2 k-\frac{M^2}{n}\log_2(k+1),\boldsymbol{W}\right)+o(n)} \end{aligned}$$

$$\begin{aligned} E_J(\boldsymbol{Q},\boldsymbol{W},t) &\le \min_{P\in\mathcal{E}}[tD_c(P\parallel Q)+E(tH_c(P),\boldsymbol{W})]\\ &\le \min_{P\in\mathcal{E}:tH_c(P)=R\in[tH(\boldsymbol{Q}),t\log_2\lambda_0(\boldsymbol{Q})]}[tD_c(P\parallel Q)+E(R,\boldsymbol{W})]\\ &= \min_{tH(\boldsymbol{Q})\le R\le t\log_2\lambda_0(\boldsymbol{Q})}\left[\min_{P\in\mathcal{E}:tH_c(P)=R} tD_c(P\parallel Q)+E(R,\boldsymbol{W})\right]\\ &= \min_{tH(\boldsymbol{Q})\le R\le t\log_2\lambda_0(\boldsymbol{Q})}\left[te\left(\frac{R}{t},\boldsymbol{Q}\right)+E(R,\boldsymbol{W})\right]. \end{aligned} \tag{34}$$

is an equivalent representation of $\overline{e}(R,\boldsymbol{Q})$ given in Corollary 1 (cf. [15]), and that $\overline{e}(R,\boldsymbol{Q})$ actually determines the source error exponent $e(R,\boldsymbol{Q})$, where the second equality of (35) follows from the strict monotonicity of $e(R/t,\boldsymbol{Q})$ in $[tH(\boldsymbol{Q}),t\log_2\lambda_0(\boldsymbol{Q})]$.

Step 3: We recall that $te\,(R/t,\boldsymbol{Q})$ is a strictly increasing function when $tH(\boldsymbol{Q})\le R\le t\log_2\lambda_0(\boldsymbol{Q})$ and is infinity when $R > t\log_2\lambda_0(\boldsymbol{Q})$, and $E(R,\boldsymbol{W})$ is a nonincreasing function of R. We thereby denote R_o to be the rate satisfying $te(R_o/t,\boldsymbol{Q}) = E(R_o,\boldsymbol{W})$ if any; otherwise, we just let R_o be $t\log_2\lambda_0(\boldsymbol{Q})$. Thus, according to (29) we can always write that $E_T(\boldsymbol{Q},\boldsymbol{W},t)=E(R_o,\boldsymbol{W})$ and R_o is a rate in the interval $[tH(\boldsymbol{Q}),t\log_2\lambda_0(\boldsymbol{Q})]$. To avoid triviality, we assume that $E_T(\boldsymbol{Q},\boldsymbol{W},t)$ (or $E(R_o,\boldsymbol{W})$) is finite, which also implies that $E_J(\boldsymbol{Q},\boldsymbol{W},t)$ is finite by (34). Suppose now the minimum of (34) is attained at R_m. We then have

$$\begin{aligned} E_J(\boldsymbol{Q},\boldsymbol{W},t) &\le te\left(\frac{R_m}{t},\boldsymbol{Q}\right)+E(R_m,\boldsymbol{W})\\ &\le te\left(\frac{R_o}{t},\boldsymbol{Q}\right)+E(R_o,\boldsymbol{W})\\ &\le 2E(R_o,\boldsymbol{W})\\ &= 2E_T(\boldsymbol{Q},\boldsymbol{W},t). \end{aligned}$$ □

REFERENCES

[1] P. Billingsley, *Ergodic Theory and Information*. Huntington, NY: Robert E. Krieger Pub. Co., 1978.

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

[3] P. N. Chen and F. Alajaji, "Csiszár's cutoff rates for arbitrary discrete sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 330–337, Jan. 2001.

[4] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Scient. Math. Hungarica*, vol. 6, pp. 181–191, 1971.

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

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

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[8] L. D. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 431–438, Jul. 1981.

[9] L. Finesso, C. C. Liu, and P. Narayan, "The optimal error exponent for Markov order estimation," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1488–1497, Sep. 1996.

[10] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[11] T. S. Han, *Information-Spectrum Methods in Information Theory*. New York: Springe-Verlag, 2004.

[12] E. A. Haroutunian, "Estimates of the error exponent for the semi-continuous memoryless channel," (in Russian) *Probl. Pered. Inform.*, vol. 4, no. 4, pp. 37–48, 1968.

[13] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1412–1424, Sep. 1997.

[14] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[15] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 360–365, May 1985.

[16] C. Pimentel, T. H. Falk, and L. Lisbôa, "Finite-state Markov modeling of correlated Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1491–1501, Sep. 2004.

[17] Z. Rached, F. Alajaji, and L. L. Campbell, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1553–1561, May 2001.

[18] E. Seneta*, Non-Negative Matrices and Markov Chains*. New York: Springer-Verlag, 1981.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.

[20] C. E. Shannon, R. G. Gallager, and R. R. Berlecamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Contr.*, vol. 10, no. 1, Feb. 1967.

[21] K. Vašek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, pp. 318–329, Nov. 1980.

[22] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 197–199, Mar. 1974.

[23] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[24] L. Zhong, F. Alajaji, and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2815–2840, Aug. 2007.

[25] L. Zhong, F. Alajaji, and G. Takahara, "Model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, Jan. 2008, to be published.

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

[27] Y. Zhong, "Joint Source-Channel Coding Reliability Function for Single and Multi-Terminal Communication Systems," Ph.D. dissertation, Dept. Math. Statist., Queen's Univ., Kingston, ON, Canada, 2008, to be published.