When Is Joint Source-Channel Coding Worthwhile: An Information Theoretic Perspective[1]

Yangfan Zhong, Fady Alajaji and L. Lorne Campbell
Dept. of Mathematics and Statistics
Queen's University, Kingston, Ontario, Canada, K7L 3N6
{yangfan, fady, campblll}@mast.queensu.ca

Abstract — **We study, from an information theoretic perspective, the merits of joint source-channel (JSC) coding versus traditional tandem coding, which consists of separately performing and concatenating source and channel coding. Specifically, we provide a systematic comparison of the JSC coding error exponent $E_J(Q,W)$ with the tandem coding error exponent $E_T(Q,W)$ for communication systems with discrete memoryless source and channel pairs (Q,W). We establish sufficient conditions under which $E_J(Q,W) > E_T(Q,W)$, which are satisfied for a large class of (Q,W) pairs. We also show that $E_J(Q,W)$ can sometimes be twice as large as $E_T(Q,W)$, hence illustrating the substantial gain that joint coding can achieve over tandem coding.**

I. Introduction

In [3], we investigate the analytical computation of Csiszár's [2] random-coding lower bound and sphere-packing upper bound for the lossless joint source-channel (JSC) error exponent, $E_J(Q,W)$, of a communication system consisting of a discrete memoryless source (DMS) with distribution Q and a discrete memoryless channel (DMC) with transition distribution W. We provide equivalent expressions for these bounds which are readily computable for arbitrary source-channel pairs, and we derive explicit conditions under which the bounds coincide, thereby exactly determining $E_J(Q,W)$.

In this work, we employ our results in [3] to provide a systematic comparison between the JSC coding exponent $E_J(Q,W)$ and the tandem coding exponent $E_T(Q,W)$. Since $E_J(Q,W) \geq E_T(Q,W)$ in general (as tandem coding is a special case of JSC coding), we are particularly interested to investigate the situation for which $E_J(Q,W) > E_T(Q,W)$. Indeed, 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 will yield a faster exponential rate of decay for the error probability, which often translates into substantial reductions in complexity and delay for real world communication systems. We establish sufficient computable conditions (see Theorems 1, 2 and 3 in Section III) for which $E_J(Q,W) > E_T(Q,W)$ for any given source-channel pair (Q,W). Furthermore, for some communication systems, we can show that joint coding design can improve the error exponent by a factor of two – i.e., $E_J(Q,W)$ can be twice as large as $E_T(Q,W)$. Typical numerical examples show that our conditions hold for a large class of (Q,W) pairs.

[1]This work was supported in part by NSERC and PREA.

II. Preliminaries: JSC Error Exponent

Definition 1 A JSC code with blocklength n for a DMS with finite alphabet $\mathcal{S}$ and distribution Q, and a DMC with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$ and transition probability $W \triangleq P_{Y|X}$ is a pair of mappings[2] $f_n : \mathcal{S}^n \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^n$. The code's average error probability is

$$P_e^{(n)}(Q,W) \triangleq \sum_{\{(s^n,y^n):\varphi_n(y^n)\neq s^n\}} Q(s^n) P_{Y|X}(y^n \mid f_n(s^n)).$$

Definition 2 The JSC error exponent $E_J(Q,W)$ for source $\{Q : \mathcal{S}\}$ and channel $\{W : \mathcal{X} \to \mathcal{Y}\}$ is defined as the largest number E for which there exists a sequence of JSC codes (f_n, φ_n) with

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

We know from the JSC coding theorem that $E_J(Q,W)$ can be positive if and only if $H(Q) < C$ (otherwise $E_J(Q,W) = 0$), where $H(Q)$ is the source entropy and C is the channel capacity. Note that $E_J(Q,W)$ depends only on the source and channel distributions.

Proposition 1 [2] The JSC error exponent $E_J(Q,W)$ satisfies $\min_R[e(R,Q) + E_r(R,W)] \leq E_J(Q,W) \leq \min_R[e(R,Q) + E_{sp}(R,W)]$, where $e(R,Q)$ is the source error exponent. $E_r(R,W)$ and $E_{sp}(R,W)$ are the random-coding lower bound and the sphere-packing upper bound for the channel error exponent $E(R,W)$, respectively.[3]

Proposition 2 [3] For a DMS $\{Q : \mathcal{S}\}$ and DMC $\{W : X \to Y\}$ pair with $H(Q) < C$, the JSC random-coding and sphere-packing bounds of Proposition 1 can be written as

$$\max_{0\leq\rho\leq 1}[E_o(\rho) - E_s(\rho)] \leq E_J(Q,W) \leq \max_{\rho\geq 0}[E_o(\rho) - E_s(\rho)],$$

where

$$E_o(\rho) \triangleq \max_{P_X}\left[-\log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y \mid x)\right)^{1+\rho}\right]$$

and $E_s(\rho) \triangleq (1+\rho)\log\sum_{s\in\mathcal{S}} Q(s)^{\frac{1}{1+\rho}}$ are Gallager's channel and source functions, respectively.

[2]We assume that the lengths of messages and codewords are identical for the sake of convenience. The results easily follow when k source symbols are mapped to n channel symbols, with k/n converging to an arbitrary positive constant.

[3]We hence call the lower bound the "JSC random-coding bound" and the upper bound the "JSC sphere-packing bound."

III. Joint versus Tandem Coding Exponents

The tandem coding exponent, which is the exponent resulting from separately performing and concatenating optimal source and channel coding, is given by [2]

$$E_T(Q,W) = \max_R \min\{e(R,Q), E(R,W)\},$$

where $e(R,Q)$ and $E(R,W)$ are the source and channel error exponents, respectively. Hence, the random coding (respectively sphere packing) bound to $E_T(Q,W)$ is obtained by replacing $E(R,W)$ in the expression of $E_T(Q,W)$ by $E_r(R,W)$ (respectively $E_{sp}(R,W)$). We note that $E_T(Q,W)$ is positive if and only if $H(Q) < C$. Otherwise, $E_T(Q,W)$ is zero. In the following, R_{cr} is the channel critical rate [4], and $Q^{(\alpha)}$, $\alpha \geq 0$, is a distribution defined by

$$Q^{(\alpha)}(s) \triangleq \frac{Q^{\frac{1}{1+\alpha}}(s)}{\sum_{s'\in\mathcal{S}} Q^{\frac{1}{1+\alpha}}(s')}, \quad s \in \mathcal{S}.$$

Theorem 1 For a DMS $\{Q : \mathcal{S}\}$ and DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ pair with $H(Q) < C$ and $\log|\mathcal{S}| > R_{cr}$, if

$$\max\left(H(Q^{(1)}), E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q)\right) \geq R_{cr},$$

then $E_J(Q,W) > E_T(Q,W)$, where $\widehat{\beta}$ is the root of $H(Q^{(\beta)}) = R_{cr}$ if $H(Q) < R_{cr} < \log|\mathcal{S}|$ and 0 if $H(Q) \geq R_{cr}$. Here, $D(\cdot \parallel \cdot)$ is the Kullback-Leibler divergence [6]. Furthermore, if we denote $\widehat{\rho} = \arg\max_{\rho>0} E(\rho)$, where $E(\rho) \triangleq E_o(\rho) - E_s(\rho)$, then we obtain the following.

1) If $\min\left(H(Q^{(1)}), E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q)\right) \geq R_{cr}$, then

$$\begin{aligned} 0 &< \frac{1}{2}E(\widehat{\rho}) - \left|\frac{1}{2}E(\widehat{\rho}) - D(Q^{(\widehat{\rho})} \parallel Q)\right| \\ &\leq E_J(Q,W) - E_T(Q,W) \leq \frac{1}{2}E_J(Q,W), \end{aligned}$$

where $E_J(Q,W)$ is exactly given by $E_J(Q,W) = E(\widehat{\rho})$;

2) If $H(Q^{(1)}) \leq R_{cr} \leq E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q)$, then

$$\begin{aligned} 0 &< R_{cr} - E_s(1) \leq E_J(Q,W) - E_T(Q,W) \\ &\leq E(\widehat{\rho}) - \frac{1}{2}[E_o(1) - R_{cr} + D(Q^{(\widehat{\beta})} \parallel Q)], \end{aligned}$$

where $E_J(Q,W)$ satisfies $E(1) \leq E_J(Q,W) \leq E(\widehat{\rho})$;

3) If $E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q) \leq R_{cr} \leq H(Q^{(1)})$, then

$$0 < E(\widehat{\rho}) - D((Q^{(\widehat{\beta})} \parallel Q)) \leq E_J(Q,W) - E_T(Q,W) \leq E(\widehat{\rho}) - \tfrac{1}{2}[E_o(1) - R_{cr} + D(Q^{(\widehat{\beta})} \parallel Q)] \leq \tfrac{1}{2}E_J(Q,W),$$

where $E_J(Q,W)$ is exactly given by $E_J(Q,W) = E(\widehat{\rho})$.

Observation 1 $E_o(\rho)$ and $R_{cr} \triangleq (\partial E_o(\rho)/\partial\rho)\big|_{\rho=1}$ do not admit analytical expressions for arbitrary DMCs; however, $\widehat{\rho}$, $E_o(\widehat{\rho})$ and R_{cr} can be easily obtained numerically via Arimoto's algorithm in [1]. Furthermore, for the class of DMCs satisfying a certain *symmetry* (as defined in [4, p. 94]), $E_o(\rho)$ can be analytically expressed, and then $\widehat{\rho}$, $E_o(\widehat{\rho})$ and R_{cr} can be solved from exact parametric expressions.

We know that if the minimum of the sum of $e(R,Q)$ and $E(R,W)$ is attained at R_m not less than R_{cr}, then the JSC exponent is exactly determined; similarly, if $e(R_o,Q) = E(R_o,W)$ for some $R_o \geq R_{cr}$, then the tandem exponent is also exactly determined. In this case, if the minimum of the sum is attained at the intersection of $e(R,Q)$ and $E(R,W)$, i.e., if $R_m = R_o$, then the source-channel exponent is twice the tandem exponent. An equivalent condition to the sufficient condition for $E_J(Q,W) = 2E_T(Q,W)$ is: $\frac{1}{2}E(\widehat{\rho}) = D(Q^{(\widehat{\rho})} \parallel Q) \leq D(Q^{(1)} \parallel Q) \iff R_o = R_m \geq R_{cr}$, which easily follows from condition 1 of Theorem 1 and the fact that $D(Q^{(x)} \parallel Q) = e(H(Q^{(x)}),Q)$ is increasing in x.

When condition 1 is satisfied, we note that the difference between $E_J(Q,W)$ and $E_T(Q,W)$ can be half of $E_J(Q,W)$; moreover, as we will see in the examples, the quantity $\frac{1}{2}E_J(Q,W) - D(Q^{(\widehat{\rho})} \parallel Q)$ can be fairly small in practice. In this case, $E_J(Q,W)$ is around twice as large as $E_T(Q,W)$ for a large class of source-channel pairs, i.e., the rate of decay of the error probability for the JSC coding system can be twice that for the tandem coding system. For conditions 2 and 3, we also have precise bounds for $E_J(Q,W) - E_T(Q,W)$. In these cases, joint coding can still substantially outperforms tandem coding, since both the lower and upper bounds for $E_J(Q,W) - E_T(Q,W)$ are close to $\frac{1}{2}E(\widehat{\rho})$ (in condition 2) or $\frac{1}{2}E_J(Q,W)$ (in condition 3).

Lemma 1 If there exists an R_1 such that $e(R_1,Q) = E_{sp}(R_1,W)$, then when $E(1) > e(R_1,Q)$, we have $E_J(Q,W) > E_T(Q,W)$; if such R_1 does not exist, then when $E(1) > -\log\left(|\mathcal{S}|\overline{Q(s)}\right)$, we also have $E_J(Q,W) > E_T(Q,W)$, where $\overline{Q(s)}$ is the geometric mean of the source probabilities, i.e.,

$$\overline{Q(s)} \triangleq \left(\prod_{s\in|\mathcal{S}|} Q(s)\right)^{\frac{1}{|\mathcal{S}|}} \leq 1/|\mathcal{S}|.$$

Lemma 2 For any DMC such that $E_{ex}(0,W) \triangleq \lim_{R\downarrow 0} E_{ex}(R,W) < \infty$, if there exists an intersection R_2 such that $e(R_2,Q) = E_{sl}(R_2,W)$, then when $E(1) > e(R_2,Q)$, we have $E_J(Q,W) > E_T(Q,W)$, where $E_{ex}(R,W)$ and $E_{sl}(R,W)$ are the expurgated lower bound and the straight-line upper bound for the channel error exponent, respectively [6].

Using Lemmas 1 and 2, we obtain the following result.

Theorem 2 For a DMS Q and a DMC W for which $E_{ex}(0,W) < \infty$, if $E(1) \geq E_{R_l}$, where

$$E_{R_l} \triangleq \frac{k_2 \log|\mathcal{S}|\left(k_1 + \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\log Q(s)\right) + k_1 E_{ex}(0,W)}{k_1 - k_2},$$

$$k_1 = \frac{D\left(Q^{(1)} \parallel Q\right) + \log|\mathcal{S}| + \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\log Q(s)}{H\left(Q^{(1)}\right) - \log|\mathcal{S}|}$$

and

$$k_2 = \frac{E_o(1) - R_{cr} - E_{ex}(0,W)}{R_{cr}},$$

then $E_J(Q,W) > E_T(Q,W)$.

Observation 2 We point out that E_{R_l} in Theorem 2 can actually be easily computed since it only requires the knowledge of R_{cr} and $E_{ex}(0, W)$. Note that the condition $E_{ex}(0, W) < \infty$ in Lemma 2 and Theorem 2 implies that the DMC has zero-error capacity equal to 0, see [6, p. 187]. Thus, Lemma 2 and Theorem 2 apply to equidistant channels ([5, p. 231]), in particular, to every channel with binary input alphabet. An expression of $E_{ex}(0, W)$ for the DMC with 0 zero-error capacity is given in [4, Problem 5.24].

Theorem 3 If $E(1) > D\left(Q^{(\widetilde{\beta})} \parallel Q\right)$, where

$$\widetilde{\beta} = \begin{cases} \text{root of } H(Q^{(\beta)}) = R_{cr} & \text{if } H(Q) < R_{cr} < \log|\mathcal{S}| \\ 0 & \text{if } H(Q) \geq R_{cr} \\ +\infty & \text{if } \log|\mathcal{S}| \leq R_{cr} \end{cases}$$

then $E_J(Q, W) > E_T(Q, W)$.

IV. Examples

Example 1 When Does the JSC Exponent Outperform the Tandem Exponent?

Consider a binary DMS $\{q, 1-q\}$ and a binary erasure channel (BEC) with parameter α, then $E(1) = 1 - \log(1+\alpha) - 2\log(\sqrt{q} + \sqrt{1-q})$ and $E_{ex}(0, W) = -\log(\alpha)/2$. If any of the conditions in Theorems 1, 2 and 3 holds, then $E_J(Q, W) > E_T(Q, W)$. The above conditions are summarized by Region **F** in Fig. 1:

$$\begin{aligned} \mathbf{F} &\triangleq (\mathbf{A} \cup \mathbf{B} \cup \mathbf{C} \cup \mathbf{D}) \cap \mathbf{E} \\ &= \{(\alpha, q) : E_J(Q, W) > E_T(Q, W)\}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{A} &\triangleq \left\{(\alpha, q) : H(Q^{(1)}) \geq R_{cr}\right\}, \\ \mathbf{B} &\triangleq \left\{(\alpha, q) : E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q) \geq R_{cr}\right\}, \\ \mathbf{C} &\triangleq \left\{(\alpha, q) : E(1) > D(Q^{(\widehat{\beta})} \parallel Q)\right\}, \\ \mathbf{D} &\triangleq \{(\alpha, q) : E(1) \geq E_{R_l}\}, \\ \mathbf{E} &\triangleq \{(\alpha, q) : H(Q) < C\}. \end{aligned}$$

Indeed, Region **F** shows that $E_J(Q, W) > E_T(Q, W)$ for a wide range of (α, q) pairs. Region **G** consists of the pairs (α, q) such that $H(Q) \geq C$; in this case, $E_J(Q, W) = E_T(Q, W) = 0$. Finally, note that when (α, q) falls in Region **H** (i.e., when both α and q are small), we are not sure whether $E_J(Q, W)$ is still strictly bigger than $E_T(Q, W)$.

Example 2 By How Much Can the JSC Exponent Be Larger Than the Tandem Exponent?

We next compare the JSC exponent with the tandem exponent for a binary DMS $\{q, 1-q\}$ and a binary symmetric channel (BSC) with parameter ε. Fig. 2 illustrates condition 1 of Theorem 1 by showing the lower and upper bounds for $E_J(Q, W) - E_T(Q, W)$ versus ε for different values of q, when $\min[H(Q^{(1)}), E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q)] \geq R_{cr}$ is satisfied. For example, when $q = 0.1$, $H(Q^{(1)}) \geq R_{cr}$ is satisfied for $\varepsilon \geq 0.0005$, while $E_o(1) - D(Q^{(\widehat{\beta})} \parallel Q) \geq R_{cr}$ is satisfied for $\varepsilon \geq 0.0012$. Thus, plotting the lower and upper bounds for $\varepsilon \geq 0.0012$, we note that the two bounds are very close (tight). Given that the upper bound is actually equal to half of $E_J(Q, W)$, we conclude that $E_J(Q, W)$ is nearly twice as large as $E_T(Q, W)$ for all $\varepsilon \geq 0.0012$. The same behavior is observed for $q = 0.2$; in this case, condition 1 holds for $\varepsilon \geq 0.001$.

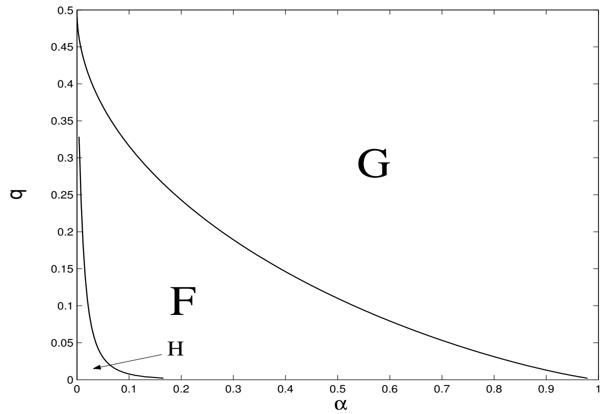

Figure 1: The regions for the (α,q) pairs in the binary DMS and BEC system.

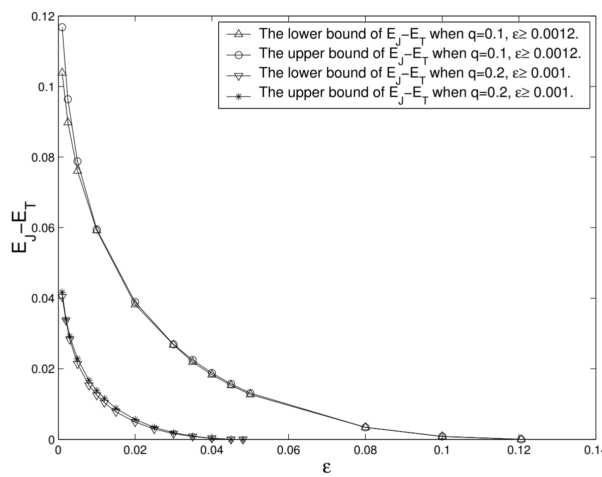

Figure 2: The lower and upper bounds of Theorem 1 (condition 1) for $E_J(Q, W) - E_T(Q, W)$ for the binary DMS and BSC system.

References

[1] S. Arimoto, "Computation of random coding exponent functions," *IEEE Trans. Inform. Theory*, vol. 22, pp. 665–671, Nov. 1976.

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

[3] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the computation of the joint source-channel error exponent for memoryless systems," *Proc. IEEE ISIT'04,* June-July 2004.

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

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

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