Mathematics and Engineering Communications Laboratory

Technical Report

On the Joint Source-Channel Coding Error Exponent for Discrete Memoryless Systems: Computation and Comparison with Separate Coding

Y. Zhong, *F. Alajaji*, and *L. L. Campbell*

December 2005

On the Joint Source-Channel Coding Error Exponent for Discrete Memoryless Systems: Computation and Comparison with Separate Coding*

Yangfan Zhong   Fady Alajaji   L. Lorne Campbell

Abstract

We investigate the computation of Csiszár's bounds for the joint source-channel coding (JSCC) error exponent, E_J, of a communication system consisting of a discrete memoryless source and a discrete memoryless channel. We provide equivalent expressions for these bounds and derive explicit formulas for the rates where the bounds are attained. These equivalent representations can be readily computed for arbitrary source-channel pairs via Arimoto's algorithm. When the channel's distribution satisfies a symmetry property, the bounds admit closed-form parametric expressions. We then use our results to provide a systematic comparison between the JSCC error exponent E_J and the tandem coding error exponent E_T, which applies if the source and channel are separately coded. It is shown that $E_T \leq E_J \leq 2E_T$. We establish conditions for which $E_J > E_T$ and for which $E_J = 2E_T$. Numerical examples indicate that E_J is close to $2E_T$ for many source-channel pairs. This gain translates into a power saving larger than 2 dB for a binary source transmitted over additive white Gaussian noise channels and Rayleigh fading channels with finite output quantization. Finally, we study the computation of the lossy JSCC error exponent under the Hamming distortion measure.

Index Terms: Joint source-channel coding, tandem source and channel coding, error exponent, reliability function, Fenchel's Duality, Hamming distortion measure, random-coding exponent, sphere-packing exponent, symmetric channels, discrete memoryless sources and channels.

*This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and the Premier's Research Excellence Award of Ontario. The authors are with the Dept. of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

1 Introduction

Traditionally, source and channel coding have been treated independently, resulting in what we call a *tandem (or separate)* coding system. This is because Shannon in 1948 [45] showed that separate source and channel coding incurs no loss of optimality (in terms of reliable transmissibility) provided that the coding blocklength goes to infinity. In practical implementations, however, there is a price to pay in delay and complexity, for extremely long blocklength. To begin, we note that *joint source-channel coding* (JSCC) might be expected to offer improvements for the combination of a source with significant redundancy and a channel with significant noise, since, for such a system, tandem coding would involve source coding to remove redundancy and then channel coding to insert redundancy. It is a natural conjecture that this is not the most efficient approach (even if the blocklength is allowed to grow without bound). Indeed, Shannon [45] made this point as follows:

> $\cdots$ *However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise. For example, in a noiseless telegraph channel one could save about* 50% *in time by proper encoding of the messages. This is not done and most of the redundancy of English remains in the channel symbols. This has the advantage, however, of allowing considerable noise in the channel. A sizable fraction of the letters can be received incorrectly and still reconstructed by the context. In fact this is probably not a bad approximation to the ideal in many cases* $\cdots$

The study of JSCC dates back to as early as the 1960's. Over the years, many works have introduced JSCC techniques and illustrated (analytically or numerically) their benefits (in terms of both performance improvement and increased robustness to variations in channel noise) over tandem coding for given source and channel conditions and fixed complexity and/or delay constraints. In JSCC systems, the designs of the source and channel codes are either well coordinated or combined into a single step. Examples of (both constructive and theoretical) previous lossless and lossy JSCC investigations include:

(a) JSCC theorems and the separation principle [6], [10], [15], [20], [23], [26], [28], [29], [32], [51];

(b) source codes that are robust against channel errors such as optimal (or sub-optimal) quantizer design for noisy channels [4], [9], [21], [22], [25], [33]–[35], [39], [41], [47], [48], [50];

(c) channel codes that exploit the source's natural redundancy (if no source coding is applied) or its residual redundancy (if source coding is applied) [3], [27], [38], [44], [57];

(d) zero-redundancy channel codes with optimized codeword assignment for the transmission of source encoder indices over noisy channels (e.g., [21], [54]);

(e) unequal error protection source and channel codes where the rates of the source and channel codes are adjusted to provide various levels of protection to the source data depending on its level of importance and the channel conditions (e.g., [30], [40]);

(f) uncoded source-channel matching where the source is uncoded, directly matched to the channel and optimally decoded (e.g., [2], [24], [46], [53]).

The above references are far from exhaustive as the field of JSCC has been quite active, particularly over the last 20 years.

In order to learn more about the performance of the best codes as a function of blocklength, much research has focused on the error exponent or reliability function for source or channel coding (see, e.g., [13], [19], [23], [31], [37], [52]). Roughly speaking, the error exponent E is a number with the property that the probability of decoding error of a good code is approximately 2^{-En} for codes of large blocklength n. Thus the error exponent can be used to estimate the trade-off between error probability and blocklength. In this paper we use the error exponent as a tool to compare the performance of tandem coding and JSCC. While jointly coding the source and channel offers no advantages over tandem coding in terms of reliable transmissibility of the source over the channel (for the case of memoryless systems as well as the wider class of stationary information stable [15, 28] systems), it is possible that the same error performance can be achieved for smaller blocklengths via optimal JSCC coding.

The first quantitative result on error exponents for lossless JSCC was a lower bound on the error exponent derived in 1964 by Gallager [23, pp. 534–535]. This result also indicates that JSCC can lead to a larger exponent than the tandem coding exponent, the exponent resulting from separately performing and concatenating optimal source and channel coding. In 1980, Csiszár [17] established a lower bound (based on the random-coding channel error exponent) and an upper bound for the JSCC error exponent $E_J(Q, W, t)$ of a communication system with transmission rate t source symbols/channel symbol and consisting of a discrete memoryless source (DMS) with distribution Q and a discrete memoryless channel (DMC) with transition distribution W. He showed that the upper bound, which is expressed as the minimum of the sum of $te(R/t, Q)$ and $E(R, W)$ over R, i.e.,

$$\min_R \left[te\left(\frac{R}{t}, Q\right) + E(R, W)\right], \tag{1}$$

where $e(R, Q)$ is the source error exponent [13], [17], [31] and $E(R, W)$ is the channel error exponent [17], [23], [31], is tight if the latter minimum is attained for an R strictly larger than the critical rate of the channel. Another (looser) upper bound to $E_J(Q, W, t)$ directly results from (1) by replacing $E(R, W)$ by the sphere-packing channel error exponent. He extended this work in 1982 [18] to obtain a new expurgated lower bound (based on the expurgated channel exponent) for the above system under some conditions, and to deal with lossy coding relative to a distortion threshold. Our first objective in this work is to recast Csiszár's results in a form more suitable for computation and to examine the connection between Csiszár's upper and lower bounds, and also the relation between the lower bounds of Gallager and Csiszár. After this, we go on to compare the tandem coding and joint coding error exponents in order to discover how much potential for improvement there is via JSCC. Since error exponents give only asymptotic expressions for system performance, our results do not have direct application to the construction of good codes. Rather, they point out certain systems for which a search for good joint codes might prove fruitful.

We first investigate the analytical computation of Csiszár's random-coding lower bound and sphere-packing upper bound for the JSCC error exponent. By applying Fenchel's Duality Theorem [36] regarding the optimization of the sum of two convex functions, we provide equivalent expressions for these bounds which involve a maximization over a non-negative parameter of the difference between the concave hull of Gallager's channel function and Gallager's source function [23]; hence, they can be readily computed for arbitrary source-channel pairs by applying Arimoto's algorithm [8]. When the channel's distribution is symmetric [23], our bounds admit closed-form parametric expressions. We also provide formulas of the rates for which the bounds are attained and establish explicit computable conditions in terms of Q and W under which the upper and lower bounds coincide; in this case, E_J can be determined exactly. A byproduct of our results is the observation that Csiszár's JSCC random-coding lower bound can be larger than Gallager's earlier lower bound obtained in [23]. Using a similar approach, we obtain the equivalent expression of Csiszár's expurgated lower bound [18] and establish the condition when the random-coding lower bound can be improved by the expurgated bound. As an example, we give closed-form parametric expressions of the improved lower bound and the corresponding condition for equidistant DMCs.

We next employ our results to provide a systematic comparison of the JSCC exponent $E_J(Q,W,t)$ and the tandem coding exponent $E_T(Q,W,t)$ for a DMS-DMC pair (Q,W) with the same transmission rate t. Since $E_J \geq E_T$ in general (as tandem coding is a special case of JSCC), we are particularly interested in investigating the situation where $E_J > E_T$. Indeed, this inequality, when it holds, provides a theoretical underpinning and justification for JSCC design as opposed to the widely used tandem approach, since the former method will yield a faster exponential rate of decay for the error probability, which may translate into substantial reductions in complexity and delay for real-world communication systems. We establish sufficient (computable) conditions for which $E_J > E_T$ for any given source-channel pair (Q,W), which are satisfied for a large class of memoryless source-channel pairs. Furthermore, we show that $E_J \leq 2E_T$. Numerical examples show that E_J can be nearly twice as large as E_T for many DMS-DMC pairs. Thus, for the same error probability, JSCC would require around half the delay of tandem coding. This potential benefit translates into more than 2 dB power gain for binary DMS sent over binary-input quantized-output additive white Gaussian noise and memoryless Rayleigh-fading channels.

We also partially address the computation of Csiszár's lower and upper bounds for the lossy JSCC exponent with distortion threshold Δ, $E_J^{\Delta}(Q,W,t)$. Under the case of the Hamming distortion measure, and for a binary DMS and an arbitrary DMC, we express the bounds for $E_J^{\Delta}(Q,W,t)$ and the rates for which the bounds are attained as in the lossless case.

The rest of this paper is arranged as follows. In Section 2 we describe the system, define the terminologies and introduce some material on convexity and Fenchel duality. Section 3 is devoted to study the analytical computation of E_J based on Csiszár's work [17], [18]. In Section 4, we assess the merits of JSCC by comparing E_J with E_T. The computation of the lossy JSCC exponent is partially studied in Section 5. Finally, we state our conclusions in Section 6.

2 Definitions and System Description

2.1 System

We consider throughout this paper a communication system consisting of a DMS $\{Q : \mathcal{S}\}$ with finite alphabet $\mathcal{S}$ and distribution Q, and a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition probability $W \triangleq P_{Y|X}$. Without loss of generality we assume that $Q(s) > 0$ for each $s \in \mathcal{S}$. Also, if the source distribution is uniform, optimal (lossless) JSCC amounts to optimal channel coding which is already well-studied. Therefore, we assume throughout that Q is not the uniform distribution on $\mathcal{S}$ except in Section 5 where we deal with JSCC under a fidelity criterion.

A joint source-channel (JSC) code with blocklength n and transmission rate $t > 0$ (measured in source symbols/channel use) is a pair of mappings $f_n : \mathcal{S}^{tn} \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{tn}$. That is, blocks $s^{tn} \triangleq (s_1, s_2, ..., s_{tn})$ of source symbols of length tn are encoded as blocks $x^n \triangleq (x_1, x_2, ..., x_n) = f_n(s^{tn})$ of symbols from $\mathcal{X}$ of length n, transmitted, received as blocks $y^n \triangleq (y_1, y_2, ..., y_n)$ of symbols from $\mathcal{Y}$ of length n and decoded as blocks of source symbols $\varphi_n(y^n)$ of length tn. The probability of erroneously decoding the block is

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

Here, Q_{tn} and $P_{n,Y|X}$ are the tn- and n-dimensional product distributions corresponding to Q and $P_{Y|X}$ respectively.

Throughout the paper, log will denote a base 2 logarithm, $|\mathcal{S}|$ will mean the number of elements in $\mathcal{S}$ and similarly for the other alphabets, C will denote the capacity of the DMC given by

$$C = \max_{P_X} I(P_X; W),$$

where $I(P_X; W)$ is the mutual information between the channel input and the channel output [23]. Finally, $H(\cdot)$ will denote the entropy of a discrete probability distribution.

2.2 Error Exponents

Definition 1 The JSCC error exponent $E_J(Q, W, t)$ is defined as the largest number E for which there exists a sequence of JSC codes (f_n, φ_n) with transmission rate t and blocklength n such that

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

When there is no possibility of confusion, $E_J(Q,W,t)$ will be written as E_J. We know from the JSCC theorem (e.g., [16, p. 216], [23]) that E_J can be positive if and only if $tH(Q) < C$.

For future use, we recall the source and channel functions used by Gallager [23] in his treatment of the JSCC theorem. We also introduce some useful notation and some elementary relations among these functions. Let Gallager's source function be

$$E_s(\rho,Q) \triangleq (1+\rho)\log\sum_{s\in\mathcal{S}} Q(s)^{\frac{1}{1+\rho}}, \qquad \rho \geq 0. \tag{2}$$

Let

$$\tilde{E}_0(\rho,P_X,W) \triangleq -\log\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}^{\frac{1}{1+\rho}}(y|x)\right)^{1+\rho}, \qquad \rho \geq 0, \tag{3}$$

and

$$\tilde{E}_x(\rho;P_X,W) \triangleq -\rho\log\sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} P_X(x)P_X(x')\left(\sum_{y\in\mathcal{Y}}\sqrt{P_{Y|X}(y\mid x)P_{Y|X}(y\mid x')}\right)^{1/\rho}, \qquad \rho \geq 1. \tag{4}$$

P_X in (3) and (4) is an unspecified probability distribution on $\mathcal{X}$. Connected with these functions are the source error exponent,

$$e(R,Q) = \sup_{0\leq\rho<\infty}[\rho R - E_s(\rho,Q)], \tag{5}$$

and three intermediate channel error exponents

$$\tilde{E}_r(R,P_X,W) \triangleq \max_{0\leq\rho\leq 1}[\tilde{E}_0(\rho,P_X,W) - \rho R], \tag{6}$$

$$\tilde{E}_{ex}(R,P_X,W) \triangleq \sup_{\rho\geq 1}[\tilde{E}_x(\rho,P_X,W) - \rho R], \tag{7}$$

and

$$\tilde{E}_{sp}(R,P_X,W) \triangleq \sup_{0\leq\rho<\infty}[\tilde{E}_0(\rho,P_X,W) - \rho R]. \tag{8}$$

From these, we can form the random-coding lower bound for the channel error exponent $E(R,W)$,

$$E_r(R,W) \triangleq \max_{P_X}\tilde{E}_r(R,P_X,W), \tag{9}$$

the expurgated lower bound

$$E_{ex}(R,W) \triangleq \max_{P_X} \tilde{E}_{ex}(R,P_X,W), \tag{10}$$

and the sphere-packing upper bound

$$E_{sp}(R,W) \triangleq \max_{P_X} \tilde{E}_{sp}(R,P_X,W). \tag{11}$$

In other words, $\max\{E_r(R,W), E_{ex}(R,W)\} \le E(R,W) \le E_{sp}(R,W)$. Also, we can form Gallager's channel functions

$$E_0(\rho,W) \triangleq \max_{P_X} \tilde{E}_0(\rho,P_X,W) \tag{12}$$

and

$$E_x(\rho,W) \triangleq \max_{P_X} \tilde{E}_x(\rho,P_X,W). \tag{13}$$

It should be noted that maximization over P_X means maximization over the closed bounded set $\{(p_1,\ldots,p_{|\mathcal{X}|}) : p_i \ge 0, \sum p_i = 1\}$. Thus, if the function involved is continuous, the maximum is achieved for some distribution $\overline{P}_X$.

The functions $\tilde{E}_r(R,P_X,W)$ and $\tilde{E}_{sp}(R,P_X,W)$ in (6) and (8) are equal if the maximizing $\rho \le 1$ in (8) or equivalently, if $R \ge R_{cr}(P_X,W)$, where $R_{cr}(P_X,W)$ is the critical rate of the channel W under distribution P_X, defined by

$$R_{cr}(P_X,W) \triangleq \left.\frac{\partial \tilde{E}_0(\rho,P_X,W)}{\partial \rho}\right|_{\rho=1}. \tag{14}$$

For all P_X, $\tilde{E}_r(R,P_X,W)$ and $\tilde{E}_{sp}(R,P_X,W)$ vanish for all $R \ge C$. Consequently, their maxima over P_X, $E_r(R,W)$ and $E_{sp}(R,W)$, vanish for $R \ge C$ and are equal on some interval $[R_{cr}(W), C]$ where $R_{cr}(W)$ is the critical rate of the channel and is defined by

$$R_{cr}(W) \triangleq \inf\{R : E_r(R,W) = E_{sp}(R,W)\}. \tag{15}$$

Furthermore, it is known that $E_{sp}(R,W)$ meets $E_r(R,W)$ on its supporting line of slope -1 [19, p. 171], which means that $E_r(R,W)$ is a straight line with slope -1 for $R \le R_{cr}(W)$ and hence

$$E_r(R,W) = E_0(1,W) - R, \qquad R \le R_{cr}(W). \tag{16}$$

For all P_X, the function $\tilde{E}_{ex}(R,P_X,W)$ is a decreasing convex curve with a straight-line section of slope -1 for $R \ge R_{ex}(P_X,W)$, and $\tilde{E}_{ex}(R,P_X,W) > \tilde{E}_r(R,P_X,W)$ for

$R < R_{ex}(P_X, W)$, where $R_{ex}(P_X, W)$ is the "expurgated" rate of the channel W under distribution P_X, defined by

$$R_{ex}(P_X, W) \triangleq \frac{\partial \tilde{E}_x(\rho, P_X, W)}{\partial \rho}\bigg|_{\rho=1}. \tag{17}$$

Since the above are satisfied for all P_X, we then obtain the following relation between the two lower bounds: $E_r(R, W) < E_{ex}(R, W)$ for $R < R_{ex}(W)$ and $E_r(R, W) \geq E_{ex}(R, W)$ otherwise, where

$$R_{ex}(W) \triangleq \inf\{R : E_r(R, W) = E_{ex}(R, W)\} \tag{18}$$

is the expurgated rate of the channel. Furthermore, it is known that $E_{ex}(R, W)$ and $E_r(R, W)$ meet their supporting line of slope -1 (according to the fact that $E_0(1, W) = E_x(1, W)$) [23, p. 154]. This geometric relation implies that $R_{ex}(W) \leq R_{cr}(W)$ and $E_r(R, W) = E_{ex}(R, W)$ is a straight line in the region $[R_{ex}(W), R_{cr}(W)]$.

We remark that Csiszár [17] defines $e(R, Q)$, $\tilde{E}_r(R, P_X, W)$, and $\tilde{E}_{sp}(R, P_X, W)$ using expressions involving constrained minima of Kullback-Leibler divergences. He also defines $\tilde{E}_{ex}(R, P_X, W)$ in terms of the Bhattacharya distance and the mutual information between two channel inputs. Our expressions are equivalent, as can be shown by the Lagrange multiplier method; see also [19, pp. 192–193] and [13].

2.3 Tilted Distributions

We associate with the source distribution Q a family of tilted distributions $Q^{(\rho)}$ defined by

$$Q^{(\rho)}(s) \triangleq \frac{Q^{\frac{1}{1+\rho}}(s)}{\sum_{s' \in \mathcal{S}} Q^{\frac{1}{1+\rho}}(s')}, \qquad s \in \mathcal{S}, \qquad \rho \geq 0. \tag{19}$$

Lemma 1 [19, p. 44] The entropy $H(Q^{(\rho)})$ is a strictly increasing function of ρ except in the case that $Q(s) = 1/|\mathcal{S}|$ for all $s \in \mathcal{S}$. Moreover, for $H(Q) \leq R \leq \log |\mathcal{S}|$, the equation $H(Q^{(\rho)}) = R$ is satisfied by a unique value ρ^* (where we define $\rho^* \triangleq \infty$ if $R = \log |\mathcal{S}|$ and define $H(Q^{(\infty)}) \triangleq \log |\mathcal{S}|$).

The proof that $H(Q^{(\rho)})$ is increasing follows easily from differentiation with respect to ρ and a use of the Cauchy-Schwarz inequality. The remainder of the proof follows from the facts that $H(Q^{(0)}) = H(Q)$, $\lim_{\rho \to \infty} H(Q^{(\rho)}) = \log |\mathcal{S}|$ and that $H(Q^{(\rho)})$ is a continuous function of ρ.

It is easily seen that

$$H(Q^{(\rho)}) = \frac{\partial E_s(\rho, Q)}{\partial \rho}, \tag{20}$$

where $E_s(\rho, Q)$ is defined by (2). From this we see that for $R \geq H(Q)$ the supremum in (5) is achieved at ρ^*.

2.4 Fenchel Duality

Although many of our results can be obtained by the use of the Lagrange multiplier method, the Fenchel Duality Theorem gives more succinct proofs and seems particularly well-adapted to the elucidation of the connection between error exponents on the one hand, and source and channel functions on the other.[1] We present here a simplified one-dimensional version which is adequate for our purposes. For more detailed discussion, the reader may consult [36, pp. 190–202], [12, Chapter 7], or [42].

For any function f defined on $F \subset \mathbb{R}$, define its convex Fenchel transform (conjugate function, Legendre transform) f^* by

$$f^*(y) \triangleq \sup_{x \in F} [xy - f(x)]$$

and let F^* be the set $\{y : f^*(y) < \infty\}$. It is easy to see from its definition that f^* is a convex function on F^*. Moreover, if f is convex and continuous, then $(f^*)^* = f$. More generally, $f^{**} \leq f$ and f^{**} is the convex hull of f, *i.e.* the largest convex function that is bounded above by f [42, Section 3], [12, Section 7.1].

Similarly, for any function g defined on $G \subset \mathbb{R}$, define its concave Fenchel transform g_* by

$$g_*(y) \triangleq \inf_{x \in G} [xy - g(x)]$$

and let G_* be the set $\{y : g_*(y) > -\infty\}$. It is easy to see from its definition that g_* is a concave function on G_*. Moreover, if g is concave and continuous, then $(g_*)_* = g$. More generally, $g_{**} \geq g$ and g_{**} is the concave hull of g, *i.e.* the smallest concave function that is bounded below by g.

Fenchel Duality Theorem [36, p. 201] Assume that f and g are, respectively, convex and concave functions on the non-empty intervals F and G in $\mathbb{R}$ and assume that $F \cap G$

[1]Another related application of Fenchel duality is carried out in [5] in the context of guessing subject to distortion, where it is shown that the guessing exponent is the Fenchel transform of the error exponent for source coding with a fidelity criterion.

has interior points. Suppose further that $\mu = \inf_{x \in F \cap G}[f(x) - g(x)]$ is finite. Then

$$\mu = \inf_{x \in F \cap G}[f(x) - g(x)] = \max_{y \in F^* \cap G_*}[g_*(y) - f^*(y)], \tag{21}$$

where the maximum on the right is achieved by some $y_0 \in F^* \cap G_*$. If the infimum on the left is achieved by some $x_0 \in F \cap G$, then

$$\max_{x \in F}[xy_0 - f(x)] = x_0 y_0 - f(x_0) \tag{22}$$

and

$$\min_{x \in G}[xy_0 - g(x)] = x_0 y_0 - g(x_0). \tag{23}$$

2.5 Properties of the Source and Channel Functions

Lemma 2 The source function $E_s(\rho, Q)$ defined by (2) is a strictly convex function of ρ.

Convexity follows directly from (20) and Lemma 1. Strict convexity is a consequence of our general assumption that Q is not the uniform distribution. It will be seen from (5) that $e(R, Q)$ is the convex Fenchel transform of $E_s(\rho, Q)$. In fact, it is easily checked that (e.g., cf. [19, pp. 44–45])

$$e(R, Q) = \begin{cases} 0 & \text{if } R \leq H(Q), \\ D(Q^{(\rho^*)} \| Q) & \text{if } H(Q) \leq R \leq \log|\mathcal{S}|\ , \\ \infty & \text{if } R > \log|\mathcal{S}|\ , \end{cases} \tag{24}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler divergence and ρ^* is the solution of $H(Q^{(\rho)}) = R$. Note that (24) implies that $e(R, Q)$ is strictly convex in R on $[H(Q), \log|\mathcal{S}|]$ when the source is nonuniform; otherwise $H(Q) = \log|\mathcal{S}|$.

The relation between the Gallager's channel function $E_0(\rho, W)$ and the random-coding and sphere-packing bounds is more complicated. First of all, recall that for each P_X, $\tilde{E}_r(R, P_X, W)$ as defined in (6) is a convex non-increasing function for all R, and is a linear function of R with slope -1 for $R \leq R_{cr}(P_X, W)$ [23, p. 143]. It will be convenient to regard this linear function as defining $\tilde{E}_r(R, P_X, W)$ for all negative R. The random coding bound $E_r(R, W)$, which is the maximum of this family of convex functions, is a convex strictly decreasing function of R for $R < C$, and is a linear function of R with slope -1 for all R below the critical rate $R_{cr}(W)$. For $R \geq C$, $E_r(R, W) = 0$. Since

$E_r(R,W)$ is convex, then $-E_r(R,W)$ is concave. Let $T_r(\rho,W)$ be the concave transform of $-E_r(R,W)$, *i.e.*

$$T_r(\rho,W) \triangleq \inf_{R\in\mathbb{R}}[\rho R + E_r(R,W)]. \tag{25}$$

It follows from the properties of $E_r(R,W)$ noted above that $T_r(\rho,W) = -\infty$ for $\rho < 0$ and $\rho > 1$ and that $T_r(\rho,W)$ is finite for $\rho \in [0,1]$.

Lemma 3 The function $T_r(\rho,W)$ defined by (25) is the concave hull on the interval $[0,1]$ of the channel function $E_0(\rho,W)$ defined in (12). Thus, $E_0(\rho,W) \le T_r(\rho,W)$ for $0 \le \rho \le 1$.

Proof: We form the concave transform of $E_0(R,W)$ on the interval $[0,1]$ to get

$$(E_0(\rho,W))_* = \inf_{0\le\rho\le1}[\rho R - E_0(\rho,W)] = -\sup_{0\le\rho\le1}[E_0(\rho,W) - \rho R].$$

Now use, in succession, (12), (6), and (9) to get

$$\begin{aligned}
(E_0(\rho,W))_* &= -\sup_{0\le\rho\le1}\max_{P_X}[\tilde{E}_0(\rho,P_X,W) - \rho R] \\
&= -\max_{P_X}\sup_{0\le\rho\le1}[\tilde{E}_0(\rho,P_X,W) - \rho R] \\
&= -\max_{P_X}\tilde{E}_r(R,P_X,W) \\
&= -E_r(R,W).
\end{aligned}$$

Since $T_r(\rho,W)$ is the concave transform of the concave function, $-E_r(R,W)$, we have that

$$(-E_r(R,W))_* = T_r(\rho,W) \quad \text{and so} \quad (E_0(\rho,W))_{**} = T_r(\rho,W).$$

Hence, $T_r(\rho,W)$ is the concave hull on $[0,1]$ of $E_0(\rho,R)$. ■

Similarly to the above, recall that $E_{sp}(R,W)$, defined in (11) is convex, zero for $R \ge C$, positive for $R < C$, and finite if $R > R_\infty(W)$ [19], [23], where $R_\infty(W)$ is given by

$$R_\infty(W) \triangleq \lim_{\rho\to\infty}\frac{E_0(\rho,W)}{\rho}. \tag{26}$$

A computable expression for $R_\infty(W)$ is given in [23, p. 158]. The normal situation is $R_\infty(W) = 0$. (As shown by Gallager, $R_\infty(W) = 0$ unless each channel output symbol is unreachable from at least one input. In the latter case, $R_\infty(W) > 0$.) We now let $T_{sp}(\rho,W)$ be the concave transform of the concave function $-E_{sp}(R,W)$, *i.e.*

$$T_{sp}(\rho,W) \triangleq \inf_{R_\infty(W)<R<\infty}[\rho R + E_{sp}(R,W)]. \tag{27}$$

It follows that $T_{sp}(\rho,W) = -\infty$ for $\rho < 0$ and that $0 \le T_{sp}(\rho,W) < \infty$ for $\rho \ge 0$.

Lemma 4 The function $T_{sp}(\rho, W)$ defined by (27) is the concave hull on $[0, \infty)$ of the channel function $E_0(\rho, W)$ defined in (12).

Proof: We now form the concave transform of $E_0(\rho, W)$ on the interval $[0, \infty)$ to get

$$(E_0(\rho, W))_* = \inf_{0\le\rho<\infty}[\rho R - E_0(\rho, W)] = -\sup_{0\le\rho<\infty}[E_0(\rho, W) - \rho R].$$

Now use (12), (8), and (11) to get

$$\begin{aligned}(E_0(\rho, W))_* &= -\sup_{0\le\rho<\infty}\max_{P_X}[\tilde{E}_0(\rho, P_X, W) - \rho R] \\ &= -\max_{P_X}\sup_{0\le\rho<\infty}[\tilde{E}_0(\rho, P_X, W) - \rho R] \\ &= -\max_{P_X}\tilde{E}_{sp}(R, P_X, W) \\ &= -E_{sp}(R, W).\end{aligned}$$

As in the previous proof, $(E_0(\rho, W))_{**} = T_{sp}(\rho, W)$. Hence, $T_{sp}(\rho, W)$ is the concave hull on $[0, \infty)$ of $E_0(\rho, R)$. ■

Observation 1 Note that the function $\tilde{E}_0(\rho, P_X, W)$ is concave in ρ for each P_X [23, p. 142]. Hence, if the maximizing P_X in (12) is *independent* of ρ, $E_0(\rho, W)$ is concave and thus $T_r(\rho, W)$ and $T_{sp}(\rho, W)$ are equal to $E_0(\rho, W)$. This situation holds if the channel is symmetric in the sense of Gallager [23, p. 94] (also see Example 2). For this case, the maximizing distribution is the uniform distribution $P_X(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$. However, there are channels for which $E_0(\rho, W)$ is not concave. One example of such a channel is provided by Gallager [23, Fig. 5.6.5]. For this particular (6-ary input, 4-ary output) channel, we plot $E_0(\rho, W)$ against ρ in Fig. 1. It is noted that the derivative of $E_0(\rho, W)$ has a positive jump increase at around $\rho = 0.51$ (see [23, Fig. 5.6.5]), and its concave hull $T_r(\rho, W)$ is strictly larger than $E_0(\rho, W)$ in the interval $\rho \in (0.41, 0.62)$.

3 Bounds on the JSCC Error Exponent

3.1 Csiszár's Random-Coding and Sphere-Packing Bounds

Csiszár [17] proved that for a DMS and a DMC the JSCC error exponent in Definition 1 satisfies

$$\underline{E}_r(Q, W, t) \le E_J(Q, W, t) \le \overline{E}_{sp}(Q, W, t), \tag{28}$$

where

$$\underline{E}_r(Q, W, t) \triangleq \min_{tH(Q) \le R \le t \log |\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_r(R, W) \right], \tag{29}$$

and

$$\overline{E}_{sp}(Q, W, t) \triangleq \inf_{tH(Q) \le R \le t \log |\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right] \tag{30}$$

are called the source-channel random-coding lower bound and the source-channel sphere-packing upper bound, since they respectively contain $E_r(R, W)$ and $E_{sp}(R, W)$ in their expressions. These bounds can be expressed in a form more adapted to calculation as follows.

Theorem 1 Let $tH(Q) < C$ and let $t \log |\mathcal{S}| > R_\infty(W)$. Then

$$\underline{E}_r(Q, W, t) = \max_{0 \le \rho \le 1} [T_r(\rho, W) - tE_s(\rho, Q)] \tag{31}$$

and

$$\overline{E}_{sp}(Q, W, t) = \max_{0 \le \rho < \infty} [T_{sp}(\rho, W) - tE_s(\rho, Q)] \tag{32}$$

where $T_r(\rho, W)$ and $T_{sp}(\rho, W)$ are the concave hulls of $E_0(\rho, W)$ on $[0, 1]$ and $[0, \infty)$ defined in (25) and (27), respectively. If the maximizing P_X in (12) is independent of ρ, $T_r(\rho, W)$ and $T_{sp}(\rho, W)$ can be replaced by $E_0(\rho, W)$.

Remark 1 When $tH(Q) \ge C$, $\underline{E}_r(Q, W, t) = \overline{E}_{sp}(Q, W, t) = 0$.

Observation 2 According to Lemma 3, $E_0(\rho, W) \le T_r(\rho, W)$. Thus the lower bound $\underline{E}_r(Q, W, t)$ can be replaced by the *possibly looser* lower bound[2]

$$\max_{0 \le \rho \le 1} [E_0(\rho, W) - tE_s(\rho, Q)]. \tag{33}$$

This is the lower bound implied by Gallager's work [23, p. 535]. As noted earlier, if the maximizing P_X in (12) is independent of ρ (e.g., for symmetric channels, see Example 2), the two lower bounds are identical.

[2] In [56], [55], we incorrectly stated that Csiszár's random-coding lower bound $\underline{E}_r(Q, W, t)$ given in (29) and Gallager's lower bound given in (33) are identical. This is indeed not always true; it is true if $E_0(\rho, W)$ is a concave function of ρ (e.g., for symmetric channels) or $tH(Q^{(1)}) \le R_{cr}(W)$ (see Corollary 3). Thus, although both lower bounds are "random coding" type bounds, Csiszár's bound is in general tighter.

Proof of Theorem 1: We first apply Fenchel's Duality Theorem (21) to the lower bound $\underline{E}_r(Q,W,t)$. From Lemma 2, (5), and (24), $te(R/t,Q)$ is convex on $(-\infty, t\log|\mathcal{S}|]$ and has convex transform $tE_s(\rho,Q)$ on the set $[0,\infty)$. Also, from the discussion preceding Lemma 3, $-E_r(R,W)$ is concave on $\mathbb{R}$ and has concave transform $T_r(\rho,W)$ which is bounded on $[0,1]$. Thus, by Fenchel's Duality Theorem,

$$\inf_{-\infty\le R\le t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+E_r(R,W)\right]=\max_{0\le\rho\le 1}[T_r(\rho,W)-tE_s(\rho,Q)]. \tag{34}$$

Now the convex function $te(R/t,Q)+E_r(R,W)$ is non-increasing for $R\le tH(Q)$ since $te(R/t,Q)=0$ in this region. This implies that the infimum on the left side of (34) can be restricted to the interval $tH(Q)\le R\le t\log|\mathcal{S}|$. Since this is now the infimum of a continuous function on a finite interval this will be a minimum. Hence, (31) is an equivalent representation of $\underline{E}_r(Q,W,t)$.

Similarly, for the upper bound, recall from the discussion preceding Lemma 4 that $-E_{sp}(R,W)$ is concave and finite for $R>R_\infty(W)$ and has a concave transform $T_{sp}(\rho,W)$, which is finite on $0\le\rho<\infty$. Thus, by Fenchel's Duality Theorem,

$$\inf_{R_\infty(W)<R\le t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+E_{sp}(R,W)\right]=\max_{0\le\rho<\infty}[T_{sp}(\rho,W)-tE_s(\rho,Q)]. \tag{35}$$

The assumption $R_\infty(W)<t\log|\mathcal{S}|$ ensures that the infimum on the left of (35) is taken over a set with interior points. If $R_\infty(W)<tH(Q)$, the infimum can be replaced by a minimum on the interval $tH(Q)\le R\le t\log|\mathcal{S}|$ by the same argument as for the lower bound. If $R_\infty(W)\ge tH(Q)$, we no longer form the infimum of a continuous function, but it can still be shown that there is a minimum point which lies in the interval $tH(Q)\le R\le t\log|\mathcal{S}|$. Hence, (35) is an equivalent representation of $\overline{E}_{sp}(Q,W,t)$. ■

Observation 3 The parametric form of the lower and upper bounds (31) and (32) indeed facilitates the computation of Csiszár's bounds. In order to compute the bounds for general non-symmetric channels (when $tH(Q)<C$ and $t\log|\mathcal{S}|>R_\infty$), one could employ Arimoto's algorithm [8] to find the maximizing distribution and thus $E_0(\rho,W)$. We then can immediately obtain the concave hulls of $E_0(\rho,W)$, $T_r(\rho,W)$ and $T_{sp}(\rho,W)$, numerically (e.g., using Matlab) and thus the maxima of $T_r(\rho,W)-tE_s(\rho,Q)$ and $T_{sp}(\rho,W)-tE_s(\rho,Q)$. This significantly reduces the computational complexity since to compute (29) and (30), we need to first compute $E_r(R,W)$ and $E_{sp}(R,W)$ for *each* R, which requires almost the same complexity as above, and then we need to find the minima by searching over *all* R's. For symmetric channels, (31) and (32) are analytically solved; see Example 2.

Example 1 Consider a communication system with a binary DMS with distribution $Q = \{q, 1-q\}$ and a DMC with $|\mathcal{X}| = 6$, $|\mathcal{Y}| = 4$, and transition probability matrix

$$W = \begin{bmatrix} 1-18\varepsilon & 6\varepsilon & 6\varepsilon & 6\varepsilon \\ 6\varepsilon & 1-18\varepsilon & 6\varepsilon & 6\varepsilon \\ 6\varepsilon & 6\varepsilon & 1-18\varepsilon & 6\varepsilon \\ 6\varepsilon & 6\varepsilon & 6\varepsilon & 1-18\varepsilon \\ 0.5-\varepsilon & 0.5-\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0.5-\varepsilon & 0.5-\varepsilon \end{bmatrix}, \qquad 0 \le \varepsilon \le \frac{1}{18}.$$

We then compute Csiszár's random-coding and sphere-packing bounds, $\underline{E}_r(Q, W, t)$ and $\overline{E}_{sp}(Q, W, t)$. For fixed Q and transmission rate t, we plot these bounds in terms of ε in Fig. 2. Our numerical results show that E_J could be determined exactly for a large class of $(q,\ \varepsilon,\ t)$ triplets: when source $Q = \{0.1, 0.9\}$ and rate $t = 0.75$, E_J is exactly known for $\varepsilon \ge 0.0025$; when $Q = \{0.1, 0.9\}$ and $t = 1$, E_J is known for $\varepsilon \ge 0.002$; and when $Q = \{0.2, 0.8\}$ and $t = 1.25$, E_J is known for $\varepsilon \ge 0.001$. Since for this channel $E_o(\rho, W)$ might not be concave (e.g., when $\varepsilon = 0.01$, W reduces to the DMC discussed in Observation 1 at the end of Section 2), our results indicate that Csiszár's lower bound is slightly but strictly larger (by ≈ 0.0001) than Gallager's lower bound (33) for $q = 0.1$, $t = 1$, and ε around 0.02. This is illustrated in Fig. 3.

3.2 When Does $\underline{E}_r(Q, W, t) = \overline{E}_{sp}(Q, W, t)$?

One important objective in investigating the bounds for the JSCC error exponent E_J is to ascertain when the bounds are tight so that the exact value of E_J is obtained. According to Csiszár's result (28), we note that if the minimum in the expressions of $\underline{E}_r(Q, W, t)$ or $\overline{E}_{sp}(Q, W, t)$ is attained for a rate (strictly) larger than the critical rate $R_{cr}(W)$, then the two bounds coincide and thus E_J is determined exactly. This raises the following question: how can we check whether the minimum in $\underline{E}_r(Q, W, t)$ or $\overline{E}_{sp}(Q, W, t)$ is attained for a rate larger than $R_{cr}(W)$? One may indeed wonder if there exist explicit conditions for which $\underline{E}_r(Q, W, t) = \overline{E}_{sp}(Q, W, t)$. The answer is affirmative; furthermore, we can verify whether the two bounds are tight in two ways: one is to compare $tH(Q^{(1)})$ with $R_{cr}(W)$, and the other is to compare the minimizer of $\overline{E}_{sp}(Q, W, t)$ in (32), $\overline{\rho}^*$ say, with 1. Before we present these conditions, we first define the following quantities which achieve the bounds

$\underline{E}_r(Q, W, t)$ and $\overline{E}_{sp}(Q, W, t)$ under the assumptions $tH(Q) < C$ and $t \log |\mathcal{S}| > R_\infty$:

$$
\begin{aligned}
\underline{R}_m &\triangleq \arg \min_{tH(Q) \le R \le t \log |\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_r(R, W) \right], && (36)\\
\overline{R}_m &\triangleq \arg \min_{tH(Q) \le R \le t \log |\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right], && (37)\\
\underline{\rho}^* &\triangleq \arg \max_{0 \le \rho \le 1} [T_r(\rho, W) - tE_s(\rho, Q)], && (38)\\
\overline{\rho}^* &\triangleq \arg \max_{0 \le \rho < \infty} [T_{sp}(\rho, W) - tE_s(\rho, Q)]. && (39)
\end{aligned}
$$

Since the functions between brackets to be minimized (or maximized) in (36)-(39) are strictly convex (or concave) functions of R (or ρ), $\underline{R}_m$, $\overline{R}_m$, $\underline{\rho}^*$ and $\overline{\rho}^*$ are well-defined and unique. We then have the following relations.

Lemma 5 Let $tH(Q) < C$ and let $t \log |\mathcal{S}| > R_\infty(W)$. Then:
(1). $\overline{\rho}^*$ and $\underline{\rho}^*$ are positive and finite.
(2). $\overline{R}_m = tH(Q^{(\overline{\rho}^*)})$.
(3). $\underline{R}_m = tH(Q^{(\underline{\rho}^*)})$ if $\underline{\rho}^* < 1$; $\underline{R}_m \ge tH(Q^{(1)})$ if $\underline{\rho}^* = 1$.

Proof: We first prove (1). Since $T_{sp}(\rho, W)$ is the concave hull of $E_0(\rho, W)$, we have the following relation

$$\lim_{\rho \downarrow 0} \frac{T_{sp}(\rho, W)}{\rho} \ge \lim_{\rho \downarrow 0} \frac{E_0(\rho, W)}{\rho} = C.$$

where the last equality follows from [7, Lemma 2]. Since $\lim_{\rho \downarrow 0} E_s(\rho, Q)/\rho = H(Q)$ by (20) and Lemma 1, we have

$$\lim_{\rho \downarrow 0} \frac{T_{sp}(\rho, W) - tE_s(\rho, Q)}{\rho} \ge C - tH(Q) > 0.$$

Note that the right-derivative of $T_{sp}(\rho, W)$ (at $\rho = 0$) must exist due to its concavity [43, pp. 113–114], and hence $\lim_{\rho \downarrow 0} T_{sp}(\rho, W)/\rho$ exists. Next we denote $\varepsilon = t \log |\mathcal{S}| - R_\infty(W) > 0$. It follows from the definition of $T_{sp}(\rho, W)$ that

$$\lim_{\rho \to \infty} \frac{T_{sp}(\rho, W)}{\rho} \le \lim_{\rho \to \infty} \frac{\rho(R_\infty(W) + \varepsilon/2) + E_{sp}(R_\infty(W) + \varepsilon/2, W)}{\rho} = R_\infty(W) + \varepsilon/2$$

because of the finiteness of $E_{sp}(R, W)$ for $R > R_\infty(W)$. This together with $\lim_{\rho \to \infty} E_s(\rho, Q)/\rho = \log |\mathcal{S}|$ implies

$$\lim_{\rho \to \infty} \frac{T_{sp}(\rho, W) - tE_s(\rho, Q)}{\rho} \le R_\infty(W) + \varepsilon/2 - t \log |\mathcal{S}| < 0.$$

Since $T_{sp}(\rho, W) - tE_s(\rho, Q)$ is 0 and has a positive right-slope at $\rho = 0$ and is negative for ρ sufficiently large, by the strict concavity of $T_{sp}(\rho, W) - tE_s(\rho, Q)$, the maximum in (39) must be achieved by a positive finite $\overline{\rho}^*$. The positivity of $\underline{\rho}^*$ can be shown in the same way and $\underline{\rho}^*$ is finite by its definition.

We next prove (2). If we now regard $te(R/t, Q)$ as $f^*(y)$ and $tE_s(\rho, Q)$ as $f(x)$ (by noting that $f^{**} = f$), then according to (22) in Fenchel's Duality Theorem,

$$\max_{0\le\rho<\infty}[\rho\overline{R}_m - tE_s(\rho, Q)] = \overline{\rho}^*\overline{R}_m - tE_s(\overline{\rho}^*, Q).$$

Setting the derivative of $\rho\overline{R}_m - tE_s(\rho, Q)$ equal to 0, we can solve for the stationary point[3] $\overline{\rho}^*$, which gives $\overline{R}_m = tH(Q^{(\overline{\rho}^*)})$.

For the lower bound, using a similar argument, we obtain the relation

$$\max_{0\le\rho\le 1}[\rho\underline{R}_m - tE_s(\rho, Q)] = \underline{\rho}^*\underline{R}_m - tE_s(\underline{\rho}^*, Q).$$

Recalling that the function between the brackets to be maximized is strictly concave, if the above maximum is achieved by $\underline{\rho}^* \in (0, 1)$, then we can solve for the stationary point as above and obtain $\underline{R}_m = tH(Q^{(\underline{\rho}^*)})$. If the maximum is achieved at $\underline{\rho}^* = 1$, then the stationary point is beyond (at least equal to) 1, and hence $\underline{R}_m \ge tH(Q^{(1)})$. Thus (3) follows. ■

In order to summarize the explicit conditions for the calculation of E_J it is convenient to define a critical rate for the source by

$$R_{cr}^{(s)}(Q) \triangleq \left.\frac{\partial E_s(\rho, Q)}{\partial \rho}\right|_{\rho=1} = H(Q^{(1)}), \tag{40}$$

recalling that $Q^{(1)}(s) = \sqrt{Q(s)}/(\sum_{s'\in\mathcal{S}}\sqrt{Q(s')})$, $s \in \mathcal{S}$.

Theorem 2 Let $tH(Q) < C$ and let $t\log|\mathcal{S}| > R_\infty(W)$. Then

- $tR_{cr}^{(s)}(Q) \ge R_{cr}(W) \Longleftrightarrow \overline{\rho}^* \le 1 \Longleftrightarrow tR_{cr}^{(s)}(Q) \ge \overline{R}_m = \underline{R}_m \ge R_{cr}(W)$. In this case,

$$E_J(Q, W, t) = T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q).$$

- $tR_{cr}^{(s)}(Q) < R_{cr}(W) \Longleftrightarrow \overline{\rho}^* > 1 \Longleftrightarrow R_{cr}(W) \ge \overline{R}_m > \underline{R}_m = tR_{cr}^{(s)}(Q)$. In this case,

$$E_0(1, W) - tE_s(1, Q) \le E_J(Q, W, t) \le T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q).$$

[3]The stationary points of a differentiable function $f(x)$ are the solutions of $f'(x) = 0$.

Remark 2 Under the condition $tR_{cr}^{(s)}(Q) > R_{cr}(W)$, $\overline{\rho}^* = 1$ is possible. However, if $tR_{cr}^{(s)}(Q) = R_{cr}(W)$, then we definitely have $\overline{\rho}^* = 1$ and $tR_{cr}^{(s)}(Q) = \overline{R}_m = \underline{R}_m = R_{cr}(W)$.

Remark 3 It can be shown that $T_{sp}(1, W) = E_0(1, W)$ and thus when $\overline{\rho}^* = 1$, the JSCC exponent is determined by

$$E_J(Q, W, t) = E_0(1, W) - tE_s(1, Q).$$

Corollary 1 Let $tH(Q) < C$ and let $t \log |\mathcal{S}| > R_\infty(W)$. Then $\underline{\rho}^* = \min\{1, \overline{\rho}^*\}$ and $\underline{R}_m = tH(Q^{(\underline{\rho}^*)})$.

The proof of Theorem 2 involves a geometric argument involving the left- and right-slopes of the convex functions $E_r(R, W)$ and $E_{sp}(R, W)$ and is deferred to Appendix A. Corollary 1 could be regarded as a complement of Lemma 5 (3) and it is also proved in Appendix A.

Corollary 2 If $\underline{R}_m \geq R_{cr}(W)$ or $\overline{R}_m > R_{cr}(W)$, then $tR_{cr}^{(s)}(Q) \geq \underline{R}_m = \overline{R}_m \geq R_{cr}(W)$, and the other equivalent conditions in Theorem 2 hold.

Proof: If $\underline{R}_m \geq R_{cr}(W)$ or $\overline{R}_m > R_{cr}(W)$, then $\underline{R}_m = \overline{R}_m$ by Lemma 9 in Appendix A. $tR_{cr}^{(s)}(Q) \geq \underline{R}_m$ immediately follows from Corollary 1. ■

Remark 4 Corollary 2 states that if $\underline{R}_m \geq R_{cr}(W)$ or $\overline{R}_m > R_{cr}(W)$, then E_J is determined exactly. Note that when $\overline{R}_m = R_{cr}(W)$, the upper and lower bounds of E_J may not be tight. In that case $\underline{R}_m < R_{cr}(W) = \overline{R}_m$ is possible. The relation between $\underline{R}_m$ and $\overline{R}_m$ is summarized in Lemma 9 in Appendix A.

We point out that, in both the computation and analysis aspects, the above conditions play an important role in verifying whether E_J can be determined exactly or not. For the class of symmetric DMCs, we can use the conditions $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ and $tR_{cr}^{(s)}(Q) < R_{cr}(W)$ to derive explicit formulas for E_J, see Example 2. In Section 4, we apply Theorem 2 to establish the conditions for which the JSCC exponent is larger than the tandem coding exponent. Note that when $tR_{cr}^{(s)}(Q) \leq R_{cr}(W)$, the source-channel random-coding bound admits a simple expression

$$\underline{E}_r(Q, W, t) = E_0(1, W) - tE_s(1, Q). \tag{41}$$

Consequently, we have the following statement.

Corollary 3 If $tR_{cr}^{(s)}(Q) \leq R_{cr}(W)$, then Csiszár's random-coding bound and Gallager's lower bound (33) are identical.

Proof: Recall Gallager's lower bound to E_J given by (33)

$$\max_{0\leq\rho\leq 1}[E_0(\rho, W) - tE_s(\rho, Q)] \geq E_0(1, W) - tE_s(1, Q).$$

Since in general Gallager's lower bound cannot be larger than Csiszár's random-coding bound, they must be equal when $tR_{cr}^{(s)}(Q) \leq R_{cr}(W)$. ■

Example 2 (DMS and Symmetric DMC) Consider a DMS $\{Q : \mathcal{S}\}$ and a *symmetric*[4] DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with rate t, where the channel transition matrix W can be partitioned along its columns into sub-matrices $W_1, W_2, \cdots, W_s$, such that in each W_i with size $|\mathcal{X}| \times |\mathcal{Y}_i|$, each row is a permutation of each other row and each column is a permutation of each other column. Denote the transition probabilities in any column of sub-matrix W_i, $i = 1, 2, \cdots, s$, by $\left\{p_{i1}, p_{i2}, ..., p_{i|\mathcal{X}|}\right\}$. Then both $E_0(\rho, W)$ and the channel capacity are achieved by the uniform distribution $P_X = 1/|\mathcal{X}|$ and have the form

$$E_0(\rho, W) = (1+\rho)\log|\mathcal{X}| - \log\left\{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\} \tag{42}$$

and

$$C = \log|\mathcal{X}| - \frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H(P_i^{(0)}),$$

where the tilted distribution $P_i^{(\alpha)}$, $\alpha \geq 0$, for each $i = 1, 2, \cdots, s$, is defined on $I_{\mathcal{X}} \triangleq \{1, 2, \cdots, |\mathcal{X}|\}$ by

$$P_i^{(\alpha)}(j) \triangleq \frac{p_{ij}^{\frac{1}{1+\alpha}}}{\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\alpha}}\right)}, \quad j \in I_{\mathcal{X}}.$$

Since now $E_0(\rho, W)$ is a concave and differentiable function of ρ, the bounds $\underline{E}_r(Q, W, t)$ and $\overline{E}_{sp}(Q, W, t)$ can be analytically obtained. If

$$\frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H(P_i^{(0)}) + tH(Q) < \log|\mathcal{X}| \tag{43}$$

[4]Here symmetry is defined in the Gallager sense [23, p. 94]; it is a generalization of the standard notion of symmetry [16] (which corresponds to $s = 1$ above).

and

$$\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2 H(P_i^{(1)})}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2}+tH(Q^{(1)})\geq\log|\mathcal{X}|, \tag{44}$$

then the source-channel exponent is positive and is exactly determined by

$$E_J(Q,W,t)=(1+\overline{\rho}^*)\log|\mathcal{X}|-\log\left\{\left[\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\overline{\rho}^*}}\right)^{1+\overline{\rho}^*}\right]\left(\sum_{s\in\mathcal{S}}Q^{\frac{1}{1+\overline{\rho}^*}}(s)\right)^{t(1+\overline{\rho}^*)}\right\}, \tag{45}$$

where $\overline{\rho}^*$ is the unique root of the equation

$$\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho} H(P_i^{(\rho)})}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}}+tH(Q^{(\rho)})=\log|\mathcal{X}|. \tag{46}$$

In the case when (43) does not hold, which means $tH(Q)\geq C$, $E_J(Q,W,t)=0$. When (43) holds but (44) does not hold, the right-hand side of (45) becomes the upper bound $\overline{E}_{sp}(Q,W,t)$ and meanwhile, E_J is lower bounded by $E_0(1,W)-tE_s(1,Q)$, where $E_0(\rho,W)$ is given by (42).

Now we apply the conditions (43) and (44) to a communication system with a binary source with distribution $\{q,1-q\}$, a binary symmetric channel (BSC) with crossover probability ε and transmission rates t =0.5, 0.75, 1, and 1.25. Note that

$$R_{cr}(W)=1-h_b\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right)$$

and

$$R_{cr}^{(s)}(Q)=h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right),$$

where $h_b(\cdot)$ is the binary entropy function. In Fig. 4, we partition the set of possible points for the (ε,q) pairs into three regions: **A**, **B** and **C**. If $(\varepsilon,q)\in\mathbf{B}$, where conditions (43) and (44) hold, i.e., $tH(Q)<C$ and $tR_{cr}^{(s)}(Q)\geq R_{cr}(W)$, then the corresponding E_J is positive and exactly known.[5] Furthermore, if $(\varepsilon,q)\in\mathbf{C}$, then E_J is bounded above (below,

[5]In light of the recent work in [11], where the random coding exponent $E_r(R,W)$ of the BSC is shown to be indeed the true value of the channel error exponent $E(R,W)$ for code rates R in some interval directly below the channel critical rate (in other words, it is shown that for the BSC with its ε above a certain threshold, $E_r(R,W)=E(R,W)$ for $R_1\leq R\leq C$ where R_1 can be *less* than $R_{cr}(W)$ [11]), we note via (1) and the lower bound in (28)-(29) that region **B** where E_J is exactly known can be enlarged.

respectively) by the right-hand side of (45) ($E_0(1,W) - tE_s(1,Q)$, respectively). When $(\varepsilon, q) \in \mathbf{A}$, where $tH(Q) > C$, E_J is zero, and the error probability of this communication system converges to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in \mathbf{B} \cup \mathbf{C}$.

3.3 Csiszár's Expurgated Lower Bound

In [18], Csiszár extended his work and obtained another lower bound to E_J for a class of source-channel pairs: for a DMS and a DMC with zero-error capacity equal to 0,

$$E_J(Q,W,t) \geq \underline{E}_{ex}(Q,W,t) \tag{47}$$

if $E_{ex}(R,W) = \max_{P_X} \tilde{E}_{ex}(R,P_X,W)$ is attained for a P_X not depending on R, where

$$\underline{E}_{ex}(Q,W,t) \triangleq \min_{tH(Q) \leq R \leq t \log|\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_{ex}(R,W) \right] \tag{48}$$

is called the source-channel expurgated lower bound since it contains $E_{ex}(R,W)$ in its expression. We then use Fenchel's Duality Theorem to derive an equivalent expression of $\underline{E}_{ex}(R,W,t)$.

Theorem 3 For a DMS and a DMC with zero-error capacity equal to 0, if $E_{ex}(R,W) = \max_{P_X} \tilde{E}_{ex}(R,P_X,W)$ is attained for a P_X not depending on R, then

$$\underline{E}_{ex}(Q,W,t) = \sup_{\rho \geq 1} [E_x(\rho,W) - tE_s(\rho,Q)]. \tag{49}$$

Proof: Recall that $\tilde{E}_x(\rho,P_X,W)$ is concave in ρ on the interval $G = [1,+\infty)$ [23, pp. 153–154]. Note that

$$-\tilde{E}_{ex}(R,P_X,W) \triangleq -\sup_{\rho \in G}[E_x(\rho,P_X,W) - \rho R] = \inf_{\rho \in G}[\rho R - \tilde{E}_x(\rho; P_X,W)]$$

is the concave transform of $\tilde{E}_x(\rho,P_X,W)$ on $R \in G^* = \{R : -\tilde{E}_{ex}(R,P_X,W) > -\infty\} = [0,+\infty)$ for DMCs with zero-error capacity equal to 0. Also recall that $tE_s(\rho,Q)$ is strictly convex in ρ on the interval $F = [0,+\infty)$. Its convex transform

$$\sup_{\rho \in F}[\rho R - tE_s(\rho,Q)] = te\left(\frac{R}{t}, Q\right)$$

is a function of R on $F^* = \{R : te(R/t,Q) < +\infty\} = (-\infty, t\log|\mathcal{S}|]$. Fenchel's Duality Theorem states that

$$\inf_{\rho \in F \cap G}[tE_s(\rho,Q) - \tilde{E}_x(\rho,P_X,W)] = \max_{R \in F^* \cap G^*} \left[-\tilde{E}_{ex}(R,P_X,W) - te\left(\frac{R}{t}, Q\right) \right]$$

or

$$\sup_{\rho\geq 1}[\tilde{E}_x(\rho,P_X,W)-tE_s(\rho,Q)]=\min_{0<R\leq t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+\tilde{E}_{ex}(R,P_X,W)\right].$$

We can now maximize over P_X and get the two equivalent lower bounds:

$$\begin{aligned}
\sup_{\rho\geq 1}[E_x(\rho,W)-tE_s(\rho,Q)] &= \max_{P_X}\min_{0<R\leq t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+\tilde{E}_{ex}(R,P_X,W)\right]\\
&\overset{(a)}{=} \min_{0<R\leq t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+\max_{P_X}\tilde{E}_{ex}(R,P_X,W)\right]\\
&\overset{(b)}{=} \min_{tH(Q)\leq R\leq t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+E_{ex}(R,W)\right]\\
&= \underline{E}_{ex}(Q,W,t),
\end{aligned}$$

where (a) follows by assumption that the maximizing P_X does not depend on R and (b) holds since the convex function $te(R/t,Q)+E_{ex}(R,W)$ is either infinity or strictly decreasing for $R<tH(Q)$. ■

In the following lemma we note that the supremum in (49) can be replaced by a maximum, and the relation between the maximizer $\underline{\rho}_x$ and its dual minimizer $\underline{R}_{xm}$ is given.

Lemma 6 For DMC with zero-error capacity equal to 0, the function $E_x(\rho,W)-tE_s(\rho,Q)$ has a global maximum at a finite $\rho\geq 1$. Let

$$\underline{\rho}_x \triangleq \arg\max_{\rho\geq 1}[E_x(\rho,W)-tE_s(\rho,Q)] \tag{50}$$

and

$$\underline{R}_{xm} \triangleq \arg\min_{tH(Q)\leq R\leq t\log|\mathcal{S}|}\left[te\left(\frac{R}{t},Q\right)+E_{ex}(R,W)\right]. \tag{51}$$

Then $\underline{R}_{xm}=tH(Q^{(\underline{\rho}_x)})$ if $\underline{\rho}_x>1$; $\underline{R}_{xm}\leq tR_{cr}^{(s)}(Q)$ if $\underline{\rho}_x=1$.

Remark 5 Since the function between brackets to be optimized in (50) (or (51)) is strictly concave (or convex), $\underline{\rho}_x$ and $\underline{R}_{xm}$ are well-defined and unique.

Proof: We first show that $\underline{\rho}_x$ is finite. Recall that for any P_X, Gallager's source and channel functions $E_s(\rho,Q)$ and $\tilde{E}_x(\rho;P_X,W)$ given in (4) at $\rho=1$ reduce to

$$E_s(1,Q)=\log\left(\sum_{s\in\mathcal{S}}\sqrt{Q(s)}\right)^2$$

and

$$\tilde{E}_x(1; P_X, W) = -\log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} P_X(x) \sqrt{P_{Y|X}(y|x)} \right)^2 .$$

Using Jensen's inequality [16] on the convex function x^2, we obtain

$$E_s(1, Q) \leq \log \sum_{s\in\mathcal{S}} (Q(s)Q(s)^{-1}) = \log |\mathcal{S}|$$

with equality if and only if Q is uniform, and

$$\tilde{E}_x(1; P_X, W) \geq -\log \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x) = 0.$$

Therefore,

$$E_x(1, W) - tE_s(1, Q) > -\log |\mathcal{S}|$$

because of the nonuniform source assumption. On the other hand, because the zero-error capacity is 0 we know that $\lim_{\rho\to\infty} \frac{E_x(\rho,W)}{\rho} = 0$ (from [23, p. 155]) and hence

$$\lim_{\rho\to\infty} \frac{E_x(\rho, W) - tE_s(\rho, Q)}{\rho} \leq -t \log_2 |\mathcal{S}|.$$

Clearly, since the concave function $E_x(\rho, W) - tE_s(\rho, Q)$ is finite (bounded below) at $\rho = 1$, and approaches to $-\infty$ as $\rho \to \infty$, there exists a global maximum at a finite $\underline{\rho}_x$. We next show the relation between $\underline{\rho}_x$ and $\underline{R}_{xm}$. Following the proof of Theorem 3, let $f^*(y)$ be $te(R/t, Q)$ and let $f(x)$ be $E_s(\rho, Q)$. Fenchel's Duality Theorem (22) says that $\underline{\rho}_x$ and $\underline{R}_{xm}$ should satisfy

$$\max_{\rho\geq 1} [\rho \underline{R}_{xm} - tE_s(\rho, Q)] = \underline{\rho}_x \underline{R}_{xm} - tE_s(\rho, Q).$$

If $\underline{\rho}_x > 1$, then $\underline{\rho}_x$ is the stationary point of the concave function $\rho\underline{R}_{xm} - tE_s(\rho, Q)$, and hence

$$\underline{R}_{xm} = tH(Q^{(\underline{\rho}_x)}).$$

Otherwise (if $\underline{\rho}_x = 1$), which means that the stationary point is less than or equal to 1, $\underline{R}_{xm} \leq tR_{cr}^{(s)}(Q)$.

■

Analogously to Theorem 2, we have the following explicit conditions regarding the expurgated lower bound to the JSCC exponent.

Theorem 4 For the expurgated lower bound in Theorem 3, the following conditions are equivalent.

- $tR_{cr}^{(s)}(Q) < R_{ex}(W) \Longleftrightarrow \underline{\rho}_x > 1 \Longleftrightarrow tR_{cr}^{(s)}(Q) < \underline{R}_{xm} \le R_{ex}(W)$. Thus,
$$E_J(Q,W,t) \ge E_x(\underline{\rho}_x, W) - tE_s(\underline{\rho}_x, Q).$$
- $tR_{cr}^{(s)}(Q) \ge R_{ex}(W) \Longleftrightarrow \underline{\rho}_x = 1 \Longleftrightarrow \underline{R}_{xm} = tR_{cr}^{(s)}(Q) \ge R_{ex}(W)$. Thus,
$$E_J(Q,W,t) \ge E_x(1,W) - tE_s(1,Q).$$

The proof of Theorem 4 is similar to that of Theorem 2 and is hence omitted. We next use Theorems 2 and 4 to compare Csiszár's random-coding and expurgated lower bounds. Of clear interest is the case when the expurgated bound improves upon the random-coding bound.

Corollary 4 The source-channel random-coding bound is improved by the expurgated bound (i.e., $\underline{E}_r(Q,W,t) < \underline{E}_{ex}(Q,W,t)$) if and only if $tR_{cr}^{(s)}(Q) < R_{ex}(W)$.

Proof: When $tR_{cr}^{(s)}(Q) < R_{ex}(W)$, we must have that $tR_{cr}^{(s)}(Q) < R_{cr}(W)$, since $R_{ex}(W)$ is never larger than $R_{cr}(W)$. It follows from Theorem 2 that the random-coding lower bound is attained at $\underline{R}_m = tR_{cr}^{(s)}(Q)$. By Theorem 4 the expurgated lower bound is attained at $R_{ex}(W) \ge \underline{R}_{xm} > tR_{cr}^{(s)}(Q)$. On account of Lemma 6, this must happen if $\underline{R}_{xm} = tH(Q^{(\underline{\rho}_x)})$ with $\underline{\rho}_x > 1$. Thus, $\underline{R}_{xm} > \underline{R}_m$ and

$$\begin{aligned}
\underline{E}_r(Q,W,t) &= E_r(\underline{R}_m, W) + te\left(\frac{\underline{R}_m}{t}, Q\right) \\
&< E_r(\underline{R}_{xm}, W) + te\left(\frac{\underline{R}_{xm}}{t}, Q\right) \\
&\le E_{ex}(\underline{R}_{xm}, W) + te\left(\frac{\underline{R}_{xm}}{t}, Q\right) \\
&= \underline{E}_{ex}(Q,W,t).
\end{aligned}$$

In this case, the source-channel expurgated lower bound is tighter than the random-coding lower bound. We then show that $\underline{E}_r(Q,W,t) \ge \underline{E}_{ex}(Q,W,t)$ if $tR_{cr}^{(s)}(Q) \ge R_{ex}(W)$.

When $R_{ex}(W) \le tR_{cr}^{(s)}(Q) \le R_{cr}(W)$, it follows from Theorems 2 and 4 that

$$\begin{aligned}
\underline{E}_r(Q,W,t) &= E_0(1,W) - tE_s(1,Q) \\
&= E_x(1,W) - tE_s(1,Q) \\
&= \underline{E}_{ex}(Q,W,t),
\end{aligned}$$

where the second equality follows from the fact that, for any P_X, Gallager's channel functions $\tilde{E}_0(1, P_X, W)$ and $\tilde{E}_x(1, P_X, W)$ are equal [23], and hence their maxima are equal. In this case, the source-channel random-coding lower bound is identical to the expurgated lower bound.

When $tR_{cr}^{(s)}(Q) > R_{cr}(W)$, we must have $tR_{cr}^{(s)}(Q) > R_{ex}(W)$. Then the expurgated lower bound is attained at $\underline{R}_{xm} = tR_{cr}^{(s)}(Q)$ by Theorem 4. On account of Theorems 2 and Corollary 1, the random-coding lower bound is attained at $\underline{R}_m = tH(Q^{(\underline{\rho}^*)}) \geq R_{cr}(W)$ with $\underline{\rho}^* \leq 1$. Consequently,

$$\begin{aligned}\underline{E}_r(Q, W, t) &= E_r(\underline{R}_m, W) + te\left(\frac{\underline{R}_m}{t}, Q\right) \\ &\geq E_{ex}(\underline{R}_m, W) + te\left(\frac{\underline{R}_m}{t}, Q\right) \\ &\geq E_{ex}(\underline{R}_{xm}, W) + te\left(\frac{\underline{R}_{xm}}{t}, Q\right) \\ &= \underline{E}_{ex}(Q, W, t).\end{aligned}$$

In this case, the source-channel random-coding lower bound is tighter than or equal to the expurgated lower bound. ■

Example 3 (DMS and Equidistant DMC) A DMC $W = P_{Y|X}$ is called equidistant if there exists a number $\beta > 0$ such that for all pairs of inputs $x \neq \widetilde{x}$,

$$\sum_y \sqrt{P_{Y|X}(y|x) P_{Y|X}(y|\widetilde{x})} = \beta.$$

Note that equidistant DMCs have 0 zero-error capacity, and every DMC with binary input alphabet is equidistant. It is shown in [31] that for an equidistant channel, $E_x(\rho, W)$ is achieved in the range $\rho \geq 1$ by a uniform input distribution $P_X(x) = 1/|\mathcal{X}|$. Therefore, we can write $E_x(\rho, W)$ as

$$E_x(\rho, W) = -\rho \log\left(\frac{|\mathcal{X}| - 1}{|\mathcal{X}|}\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|}\right) \qquad \text{for} \qquad \rho \geq 1.$$

Now we apply Theorems 3 and 4 to DMS Q and equidistant DMC W with transmission rate t. We then see that if

$$tH(Q^{(1)}) + \log\left(\frac{|\mathcal{X}| - 1}{|\mathcal{X}|}\beta + \frac{1}{|\mathcal{X}|}\right) \leq \frac{\beta \log \beta}{\beta + \frac{1}{|\mathcal{X}|-1}}, \tag{52}$$

the expurgated JSCC lower bound is tighter than the random-coding lower bound and is given by

$$E_J(Q,W,t) \geq -\underline{\rho}_x \log\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\underline{\rho}_x}} + \frac{1}{|\mathcal{X}|}\right) - t(1+\underline{\rho}_x)\log\sum_{s\in\mathcal{S}} Q^{\frac{1}{1+\underline{\rho}_x}}(s), \tag{53}$$

where $\underline{\rho}_x$ is the unique root of the equation

$$tH(Q^{(\rho)}) + \log\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|}\right) = \frac{\rho^{-1}\beta^{\frac{1}{\rho}}\log\beta}{\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|-1}}.$$

Consider a communication system with a binary source with distribution $\{q, 1-q\}$, a binary erasure channel (BEC) with erasure probability α and transmission rate $t = 1$ (similar results hold for other cases, as in the last example). Using the conditions (43), (44) in Example 2, and together with (52), we present in Fig. 5 the set of (α, q) points, partitioned into four regions. If the pair (α, q) is located in region $\mathbf{B}$, then the system E_J is positive and exactly known. If $(\alpha, q) \in \mathbf{C} = \mathbf{C}_1 \cup \mathbf{C}_2$, then upper and lower bounds for E_J are known. Here, region $\mathbf{C}_2$ consists of the values of (α, q) for which the source-channel expurgated lower bound given in (53) is tighter than the source-channel random-coding lower bound. Finally, when $(\alpha, q) \in \mathbf{A}$, $E_J(Q,W,t) = 0$. In Fig. 6, we plot the random-coding and expurgated lower bounds for different source and BEC pairs. We observe that when the source distribution is Q={0.1,0.9} (respectively Q={0.2,0.8}), the expurgated lower bound for E_J is tighter than the random-coding lower bound if $\alpha < 0.0297$ (respectively if $\alpha < 0.0102$).

4 When is JSCC Worthwhile: JSCC vs Tandem Coding Exponents

4.1 Tandem Coding Error Exponent

A tandem code $(f_n^*, \varphi_n^*) \triangleq (f_{cn} \circ f_{sn}, \varphi_{sn} \circ \varphi_{cn})$ for a DMS $\{Q : \mathcal{S}\}$ and a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with blocklength n and transmission rate t (source symbols/channel use) is composed independently by a (tn, M) block source code (f_{sn}, φ_{sn}) defined by $f_{sn} : \mathcal{S}^{tn} \longrightarrow \{1, 2, ..., M\}$ and $\varphi_{sn} : \{1, 2, ..., M\} \longrightarrow \mathcal{S}^{tn}$ with source code rate

$$R_s \triangleq \frac{\log M}{tn} \qquad \text{source code bits/source symbol,}$$

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

$$R_c \triangleq \frac{\log M}{n} \qquad \text{source code bits/channel use,}$$

where "$\circ$" means composition and R_s and R_c are independent of n. That is, blocks s^{tn} of source symbols of length tn are encoded as integers (indices) $f_{sn}(s^{tn})$ from $\{1, 2, ..., M\}$, and these integers are further encoded as blocks $x^n = f_{cn}\left[f_{sn}(s^{tn})\right]$ of symbols from $\mathcal{X}$ of length n, transmitted, received as blocks y^n of symbols from $\mathcal{Y}$ of length n. These received blocks y^n are decoded as integers $\varphi_{cn}(y^n)$ from $\{1, 2, ..., M\}$, and finally, these integers are decoded as blocks of source symbols $\varphi_n^*(y^n) = \varphi_{sn}\left[\varphi_{cn}(y^n)\right]$ of length tn. Thus, the probability of erroneously decoding the block is

$$P_{e^*}^{(n)}(Q, W, t) \triangleq \sum_{\{(s^{tn}, y^n): \varphi_{sn}[\varphi_{cn}(y^n)] \neq s^{tn}\}} Q_{tn}(s^{tn}) P_{n,Y|X}\left(y^n \left| f_{cn}\left[f_{sn}(s^{tn})\right]\right.\right),$$

where Q_{tn} and $P_{n,Y|X}$ are the tn- and n-dimensional product distributions corresponding to Q and $P_{Y|X}$. respectively.

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

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

When there is no possibility of confusion, $E_T(Q, W, t)$ will often be written as E_T. In general, we know that $E_J \geq E_T$ since by definition tandem coding is a special case of JSCC. We are hence interested in determining the conditions for which $E_J > E_T$ for the same transmission rate t. Meanwhile, it immediately follows (from the JSCC theorem) that E_T can be positive if and only if $tH(Q) < C$; otherwise, both E_J and E_T are zero.

By definition, the tandem coding exponent results from separately performing and concatenating optimal source and channel coding, which can be expressed by (e.g., see [17])

$$\begin{aligned} E_T(Q, W, t) &= \sup_{R_s, R_c: R_c = tR_s} \min\left\{te(R_s, Q), E(R_c, W)\right\} \\ &= \sup_R \min\left\{te\left(\frac{R}{t}, Q\right), E(R, W)\right\}, \end{aligned} \tag{54}$$

where $e(R,Q)$ and $E(R,W)$ are the source and channel error exponents, respectively. Note that

$$\sup_{R\le t\log|\mathcal{S}|} te\left(\frac{R}{t},Q\right) = te(\log|\mathcal{S}|,Q) = -t\log(|\mathcal{S}|\overline{Q(s)}),$$

where $\overline{Q(s)}$ is the geometric mean of the source probabilities, i.e. $\overline{Q(s)} \triangleq \left(\prod_{s\in\mathcal{S}} Q(s)\right)^{1/|\mathcal{S}|} \le 1/|\mathcal{S}|$. If $-t\log(|\mathcal{S}|\overline{Q(s)}) \ge E(t\log|\mathcal{S}|,W)$, then the graphs of $te(R/t,Q)$ and $E(R,W)$ must have exactly one intersection R_o and by (54)

$$E_T(Q,W,t) = te\left(\frac{R_o}{t},Q\right) = E(R_o,W), \tag{55}$$

since $te(R/t,Q)$ is strictly increasing in $R\in[tH(Q),t\log|\mathcal{S}|]$ and $E(R,W)$ is non-increasing in R. If $-t\log(|\mathcal{S}|\overline{Q(s)}) < E(t\log|\mathcal{S}|,W)$, then there is no intersection between $te(R/t,Q)$ and $E(R,W)$. Recall (24) that $te(R/t,Q)$ is infinite in the open interval $(t\log|\mathcal{S}|,\infty)$. In this case, we have that

$$E_T(Q,W,t) = E(t\log|\mathcal{S}|,W) \tag{56}$$

by (54). Without loss of generality, we denote

$$R_o \triangleq \begin{cases} \text{the rate satisfying} \qquad te(\frac{R_o}{t},Q) = E(R_o,W) \\ \qquad\qquad\qquad \text{if } -t\log(|\mathcal{S}|\overline{Q(s)}) \ge E(t\log|\mathcal{S}|,W), \\ t\log|\mathcal{S}| \\ \qquad\qquad\qquad \text{if } -t\log(|\mathcal{S}|\overline{Q(s)}) < E(t\log|\mathcal{S}|,W), \end{cases} \tag{57}$$

so that we can always write that $E_T(Q,W,t) = E(R_o,W)$.

When the DMS is uniform, the optimal source coding operation reduces to the trivial enumerating (identity) function with $M = |S|^{tn}$ as the source is incompressible. Hence only channel coding is performed in both JSCC and tandem coding and $E_J(Q,W,t) = E_T(Q,W,t) = E(t\log|\mathcal{S}|,W)$. Thus, our comparison of the two exponents is nontrivial only if the source is nonuniform and $tH(Q) < C$. Even though we know that E_J is never worse than E_T, the following theorem gives a limit on how much E_J can outperform E_T.

Theorem 5 JSCC exponent can at most be equal to double the tandem coding exponent, i.e.,

$$E_J(Q,W,t) \le 2E_T(Q,W,t),$$

with equality if $tR_{cr}^{(s)}(Q) \ge R_{cr}(W)$ and $T_{sp}(\overline{\rho}^*,W) = tE_s(\overline{\rho}^*,Q) + 2tD(Q^{(\overline{\rho}^*)} \parallel Q)$.

Remark 6 Equivalently, this upper bound also implies that E_J can at most exceed E_T by $E_J/2$, i.e.,

$$E_J(Q, W, t) - E_T(Q, W, t) \leq \frac{1}{2} E_J(Q, W, t). \tag{58}$$

Proof: We first refer to the upper bound of $E_J(Q, W, t)$ given by Csiszár [17, Lemma 2]

$$E_J(Q, W, t) \leq \min_{tH(Q) \leq R \leq t \log |\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E(R, W) \right], \tag{59}$$

where $te(R/t, W)$ is the source error exponent, which is strictly convex and increasing in $[tH(Q), t \log |\mathcal{S}|]$, and $E(R, W)$ is the channel error exponent, which is a positive and non-increasing in $[0, C)$. Unlike the source exponent, the behavior of $E(R, W)$ is unknown for $R < R_{cr}(W)$. Let C_0 be the zero-error capacity of the channel W, i.e., $E(R, W) = \infty$ if and only if $R < C_0$ [23]. If $C_0 > t \log |\mathcal{S}|$, obviously, we have

$$E_J(Q, W, t) = E_T(Q, W, t) = +\infty.$$

If $C_0 \leq t \log |\mathcal{S}|$, the upper bound in (59) is finite and the minimum must be achieved by some rate, say R_m, in the interval $[C_0, t \log |\mathcal{S}|]$. Then

$$\begin{aligned}
E_J(Q, W, t) &\overset{(a)}{\leq} te\left(\frac{R_m}{t}, Q\right) + E(R_m, W) \\
&\overset{(b)}{\leq} te\left(\frac{R_o}{t}, Q\right) + E(R_o, W) \\
&\overset{(c)}{\leq} 2E(R_o, W) \\
&= 2E_T(Q, W, t).
\end{aligned}$$

Here, the equality in (a) holds if our computable upper and lower bounds, $\overline{E}_{sp}(Q, W, t)$ and $\underline{E}_r(Q, W, t)$, are equal. To ensure this, we need the condition $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ by Theorem 2. The equality in (b) holds if $R_m = R_o$ by definition of R_m. The equality (c) holds if and only if there is an intersection between $te(R/t, W)$ and $E(R, W)$, i.e., $te(R_o/t, Q) = E(R_o, W)$. Now taking these considerations together, and applying Theorem 2 again, we conclude that $E_J = 2E_T$ if $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ and $T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q) = 2te(\overline{R}_m/t, Q) = 2tD(Q^{(\overline{\rho}^*)} \parallel Q)$. ■

Observation 4 The condition for the equality states that, if the minimum in the expression of $\underline{E}_r(Q, W, t)$ given in (29) is attained at the intersection of $te(\frac{R}{t}, W)$ and $E_r(R, W)$ which is no less than the critical rate of the channel, then the JSCC exponent is *twice* as large as the tandem coding exponent. In that case, the rate of decay of the error probability for the JSCC system is *double* that for the tandem coding system. In other words, for the same probability of error P_e, the delay of (optimal) JSCC is approximately *half* of the delay of (optimal) tandem coding,

$$P_e \approx 2^{-nE_T(Q,W,t)} = 2^{-\frac{n}{2}E_J(Q,W,t)} \qquad \text{for } n \text{ sufficiently large.}$$

4.2 Sufficient Conditions for which $E_J > E_T$

In the following we will use our previous results to derive computable sufficient conditions for which $E_J > E_T$. We first define

$$\gamma \triangleq \begin{cases} \text{the root of} \qquad tH(Q^{(\gamma)}) = R_{cr}(W) & \text{if } tH(Q) \le R_{cr}(W) \le t\log|\mathcal{S}|, \\ 0 & \text{if } tH(Q) > R_{cr}(W). \end{cases} \tag{60}$$

such that the source error exponent $te(R/t, Q)$ has a parametric expression at $R_{cr}(W)$

$$te\left(\frac{R_{cr}(W)}{t}, Q\right) = tD(Q^{(\gamma)} \parallel Q). \tag{61}$$

Note that γ is well defined only if $R_{cr}(W) \le t\log|\mathcal{S}|$. Denote

$$T(\overline{\rho}^*) \triangleq T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q). \tag{62}$$

Theorem 6 Let $R_{cr}(W) \le t\log|\mathcal{S}|$. If

$$\max\left\{tR_{cr}^{(s)}(Q), E_o(1,W) - tD(Q^{(\gamma)} \parallel Q)\right\} \ge R_{cr}(W), \tag{63}$$

then

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

More precisely, we have the following bounds.
(a) If $\min\left\{tR_{cr}^{(s)}(Q), E_o(1,W) - tD(Q^{(\gamma)} \parallel Q)\right\} \ge R_{cr}(W)$, then

$$E_J(Q,W,t) - E_T(Q,W,t) \ge \frac{1}{2}T(\overline{\rho}^*) - \left|\frac{1}{2}T(\overline{\rho}^*) - tD(Q^{(\overline{\rho}^*)} \parallel Q)\right| \ge 0, \tag{64}$$

where the two equalities in (64) cannot hold simultaneously.
(b) If $tR_{cr}^{(s)}(Q) \ge R_{cr}(W) > E_o(1,W) - tD(Q^{(\gamma)} \parallel Q)$, then

$$E_J(Q,W,t) - E_T(Q,W,t) > T(\overline{\rho}^*) - tD(Q^{(\gamma)} \parallel Q) \ge 0. \tag{65}$$

(c) If $E_o(1,W) - tD(Q^{(\gamma)} \parallel Q) \ge R_{cr}(W) > tR_{cr}^{(s)}(Q)$, then

$$E_J(Q,W,t) - E_T(Q,W,t) \ge R_{cr}(W) - tE_s(1,Q) > 0. \tag{66}$$

Proof: We shall show that, in each of the three cases, (a), (b), and (c), we have $E_J > E_T$.

(a). Assume $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ and $E_o(1,W) - tD(Q^{(\gamma)} \parallel Q) \geq R_{cr}(W)$. By definition of γ, we have $tD(Q^{(\gamma)} \parallel Q) = te(R_{cr}(W)/t, Q)$, see (24) and (61). Thus, the latter condition is equivalent to $E(R_{cr}(W), W) \geq te(R_{cr}(W)/t, Q)$ and by (16) and the related discussion it guarantees that $R_o \geq R_{cr}(W)$, where R_o is defined in (57). According to Theorem 2, when $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$, $\overline{E}_{sp}(Q,W,t)$ is attained by $\overline{R}_m \geq R_{cr}(W)$ and E_J is determined by

$$E_J(Q,W,t) = te\left(\frac{\overline{R}_m}{t}, Q\right) + E_{sp}(\overline{R}_m, W).$$

Since $R_o \geq R_{cr}(W)$, E_T is determined by $E_{sp}(R_o, W)$. If $R_o \neq \overline{R}_m$, we must have

$$E_T(Q,W,t) < \max\left\{te\left(\frac{\overline{R}_m}{t}, Q\right), E_{sp}(\overline{R}_m, W)\right\},$$

because $te(R/t, Q)$ is strictly increasing and $E_{sp}(R,W)$ is strictly decreasing at $\overline{R}_m$. Thus,

$$E_J(Q,W,t) - E_T(Q,W,t) > \min\left\{te\left(\frac{\overline{R}_m}{t}, Q\right), E_r(\overline{R}_m, W)\right\} \geq 0, \tag{67}$$

where equality holds if $\overline{R}_m = C$. If $R_o = \overline{R}_m$, then immediately,

$$E_J(Q,W,t) - E_T(Q,W,t) = te\left(\frac{\overline{R}_m}{t}, Q\right) = tD(Q^{(\overline{\rho}^*)} \parallel Q), \tag{68}$$

where the above is positive since $\overline{\rho}^* > 0$ by Lemma 5 (1). Note also that in this case $te(\overline{R}_m/t, Q) = E_r(\overline{R}_m, W)$, so (67) and (68) can be summarized by (64).

(b). In this case, we have $\overline{R}_m \geq R_{cr}(W) > R_o$. We can upper bound E_T by

$$E_T(Q,W,t) = te\left(\frac{R_o}{t}, Q\right) < te\left(\frac{R_{cr}(W)}{t}, Q\right) = tD(Q^{(\gamma)} \parallel Q)$$

and hence

$$E_J(Q,W,t) - E_T(Q,W,t) > T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q) - tD(Q^{(\gamma)} \parallel Q).$$

The above lower bound must be nonnegative since

$$\begin{aligned} T_{sp}(\overline{\rho}^*, W) - tE_s(\overline{\rho}^*, Q) - tD(Q^{(\gamma)} \parallel Q) &= E_r(\overline{R}_m, W) + t\left[e\left(\frac{\overline{R}_m}{t}, Q\right) - e\left(\frac{R_{cr}(W)}{t}, Q\right)\right] \\ &\geq E_r(\overline{R}_m, W) \\ &\geq 0, \end{aligned}$$

and it is equal to 0 if $R_{cr}(W) = \overline{R}_m = C$.

(c). In this case, we have $R_o \geq R_{cr}(W) > \underline{R}_m$ and from (41) E_J is bounded by

$$E_J(Q, W, t) \geq E_0(1, W) - tE_s(1, Q).$$

On the other hand, by the monotonicity of $E_r(R, W)$, we can upper bound E_T by

$$E_T(Q, W, t) = E_r(R_o, W) \leq E_r(R_{cr}(W), W) = E_0(1, W) - R_{cr}(W).$$

Thus we obtain

$$E_J(Q, W, t) - E_T(Q, W, t) \geq R_{cr}(W) - tE_s(1, Q).$$

The above is positive since

$$\begin{aligned} E_0(1, W) - tE_s(1, Q) &= te\left(\frac{\underline{R}_m}{t}, Q\right) + E_r(\underline{R}_m, W) \\ &> E_r(\underline{R}_m, W) \\ &> E_r(R_{cr}(W), W) \\ &= E_0(1, W) - R_{cr}(W), \end{aligned}$$

where the first inequality follows from the fact that $\underline{R}_m > tH(Q)$ by Lemma 5 and Corollary 1.

■

As pointed out in the proof, the condition $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ means that the JSCC exponent E_J is achieved at a rate no less than $R_{cr}(W)$. The second condition, $E_o(1, W) - tD(Q^{(\gamma)} \parallel Q) \geq R_{cr}(W)$ means that the tandem coding exponent E_T is achieved at a rate no less than $R_{cr}(W)$. Hence (63) in Theorem 6 states that E_J would be strictly larger than E_T if either E_J or E_T is determined exactly. Conversely, if the conditions in Theorem 6 are not satisfied, then neither E_J nor E_T are exactly known. Nevertheless, if the lower bound of E_J is strictly larger than the upper bound of E_T, then we must have $E_J > E_T$. Hence we obtain the following sufficient conditions.

Theorem 7 Let $E_{ex}(0, W) < \infty$ and let $t \log |\mathcal{S}| \geq R_{cr}(W)$, where $E_{ex}(R, W)$ is the expurgated channel error exponent [23]. If

$$E_0(1, W) - tE_s(1, Q) \geq E_{R_l} \triangleq \frac{k_1 k_2 t \log |\mathcal{S}| + k_2 t \log(|\mathcal{S}|\overline{Q(s)}) + k_1 E_{ex}(0, W)}{k_1 - k_2},$$

where

$$k_1 = \frac{D\left(Q^{(1)} \parallel Q\right) + \log(|\mathcal{S}|\overline{Q(s)})}{H\left(Q^{(1)}\right) - \log|\mathcal{S}|} \qquad \text{and} \qquad k_2 = \frac{E_0(1,W) - E_{ex}(0,W)}{R_{cr}(W)} - 1,$$

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

Theorem 8 Let $t\log|\mathcal{S}| \geq R_{cr}(W)$. If $E_0(1,W) - tE_s(1,Q) \geq tD\left(Q^{(\gamma)} \parallel Q\right)$, where γ is defined in (60), then $E_J(Q,W,t) > E_T(Q,W,t)$.

In Theorems 7 and 8, we establish the sufficient conditions by comparing the source-channel random-coding bound derived in Theorem 2, with the upper bound of tandem coding exponent obtained by using the geometric characteristics of $e(R,W)$ and $E(R,W)$. The proofs of Theorems 7 and 8 are given in Appendices B and C, respectively. These conditions can be readily computed since it only requires the knowledge of $R_{cr}(W)$ and $E_{ex}(0,W)$. Note that the condition $E_{ex}(0,W) < \infty$ in Theorem 7 is satisfied by the DMCs with zero-error capacity equal to 0, see [19, p. 187]. Thus, Theorem 7 applies to equidistant channels, 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 [23, Problem 5.24].

Example 4 (When Does the JSCC Exponent Outperform the Tandem Coding Exponent?) We apply Theorems 6, 7 and 8 to the binary DMS with distribution $\{q, 1-q\}$ and BSC with crossover probability ε, and the binary DMS $\{q, 1-q\}$ and the binary erasure channel (BEC) with erasure probability α, under different transmission rates t. If any one of the conditions in these theorems holds, then $E_J > E_T$. The above conditions are summarized by Region **F** in Fig. 7. Indeed, Region **F** shows that $E_J > E_T$ for a wide range of (ε, q) or (α, q) pairs. Region **G** consists of the pairs (ε, q) or (α, q) such that $tH(Q) \geq C$; in this case, $E_J = E_T = 0$. Finally, when (ε, q) or (α, q) falls in Region **H**, we are not sure whether E_J is still strictly larger than E_T.

Example 5 (By How Much Can the JSCC Exponent Be Larger Than the Tandem Coding Exponent?) In the last example we have seen that $E_J > E_T$ holds for a wide large class of source-channel pairs. Now we evaluate the performance of E_J over E_T by looking at the ratio of the two quantities. Recall that when Theorem 6 (a) is satisfied, both E_J and E_T are exactly determined. In this case we can directly compute E_J (using the results of Section 3) and E_T (using (55) and (56)). When E_J (respectively, E_T) is not known, i.e., when $tR_{cr}^{(s)}(Q) < R_{cr}(W)$ (respectively, $E_o(1,W) - tD(Q^{(\gamma)} \parallel Q) < R_{cr}(W)$), we can calculate the lower bound of E_J (respectively, the upper bound of E_T) instead and thus obtain a lower bound for E_J/E_T. For general DMCs, we lower bound E_J by its random-coding lower bound $\underline{E}_r(Q,W,t)$. For equidistant DMCs, particularly for binary DMCs, when $tR_{cr}^{(s)}(Q) < R_{ex}(W)$, we use the expurgated lower bound $\underline{E}_{ex}(Q,W,t)$; when $tR_{cr}^{(s)}(Q) \geq R_{ex}(W)$, we use the random-coding lower bound $\underline{E}_r(Q,W,t)$. To calculate the upper bound of E_T, when $E_o(1,W) - tD(Q^{(\gamma)} \parallel Q) < R_{cr}(W) \leq R_{cr}^{(s)}(Q)$, or equivalently when $R_o < R_{cr}(W) \leq \overline{R}_m$, we can bound E_T by

$$E_T(Q,W,t) \leq \min\left\{tD\left(Q^{(\gamma)} \parallel Q\right), E_{sp}(R_s,W)\right\},$$

where R_s is the intersection of $E_{sp}(R,W)$ and $te(R/t,Q)$ if any; otherwise $R_s = t\log|S|$. When $E_o(1,W) - tD(Q^{(\gamma)} \parallel Q) < R_{cr}(W)$ and $R_{cr}^{(s)}(Q) < R_{cr}(W)$, we bound E_T by

$$E_T(Q,W,t) \leq E_{sp}(R_s,W).$$

Table 1 exhibits E_J/E_T (or its lower bound, which must be no less than 1) for the binary DMS $\{q, 1-q\}$ and BSC (ε) system under transmission rates $t = 0.5$, 0.75 and 1. It is seen that the ratio E_J/E_T can be very close to 2 (its upper bound) for many (q,ε) pairs. For other systems, we have similar results: E_J substantially outperforms E_T. For instance, for binary DMS $\{q, 1-q\}$ and BEC (α) with $t = 1$, we note that $E_J/E_T \geq 1.4$ for a wide range of (q,α)'s; for ternary DMS and BSC or for DMS and ternary symmetric channel, if transmission rate t is chosen suitably (such that $tH(Q) < C$), we obtain that $E_J/E_T \geq 1.5$ for many source-channel pairs.

4.3 Power Gain Due to JSCC for DMS over Binary-input AWGN and Rayleigh-Fading Channels with Finite Output Quantization

It is well known that M-ary modulated additive white Gaussian noise (AWGN) and memoryless Rayleigh-fading channels can be converted to a DMC when finite quantization is applied at their output. For example, as illustrated in [4], [41], we know that the concatenation of a binary phase-shift keying (BPSK) modulated AWGN or Rayleigh-fading channel with m-bit soft-decision demodulation is equivalent to a binary-input, 2^m-output DMC (cf. Fig. 8). We next study the JSCC and tandem coding exponent for a system involving such channels to assess the potential benefits of JSCC over tandem coding in terms of power or channel signal-to-noise ratio (SNR) gains.

We assume that the BPSK signal $U_n \in \{-1, +1\}$ corresponding to the signal input X_n is of unit energy, and V_n is a zero-mean independent and identically distributed (i.i.d.) Gaussian random process with variance $N_o/2$. The channel SNR is defined by SNR $\triangleq E[U_n^2]/E[V_n^2] = 2/N_o$ and the received signal is

$$Z_n = A_n U_n + V_n, \qquad n = 1, 2, ...,$$

where A_n is 1 for the AWGN channel (no fading), and for the Rayleigh-fading channel, $\{A_n\}$ is the amplitude fading process assumed to be i.i.d. with probability density function (pdf)

$$f_A(a) = \begin{cases} 2ae^{-a^2}, & \text{if } a > 0, \\ 0, & \text{otherwise,} \end{cases}$$

such that $E[A_n^2] = 1$. We also assume for the Rayleigh-fading channel that A_n, U_n and V_n are independent of each other, and the values of A_n are not available at the receiver. At the receiver, as shown in Fig. 8, each $Z_n \in \mathbb{R}$ is demodulated via an m-bit uniform scalar quantizer with quantization step Δ to yield $Y_n \in \{0, 1\}^m$. If the channel input alphabet is $\mathcal{X} = \{0, 1\}$ and the channel output alphabet is $\mathcal{Y} = \{0, 1, 2, ..., 2^m - 1\}$, then the transition probability matrix Π is given by

$$\Pi = [\pi_{ij}], \qquad i \in \mathcal{X}, \qquad j \in \mathcal{Y},$$

where

$$\pi_{ij} \triangleq P(Y = j | X = i) = \mathcal{Q}\left((T_{j-1} - (2i - 1))\sqrt{\text{SNR}}\right) - \mathcal{Q}\left((T_j - (2i - 1))\sqrt{\text{SNR}}\right)$$

for the AWGN channel [41], and

$$\pi_{ij} \triangleq P(Y = j | X = i) = F_{Z|X}(T_j | i) - F_{Z|X}(T_{j-1} | i)$$

for the Rayleigh-fading channel [4]. Here $F_{Z|X}(z|i) = Pr\{Z \leq z | Z = i\}$ is given by [4], [49]

$$F_{Z|X}(z|1) = 1 - F_{Z|X}(-z|0) = 1 - \mathcal{Q}\left(\frac{z}{\sqrt{N_o/2}}\right) - \frac{e^{-(z^2/(N_o+1))}}{\sqrt{N_o + 1}} \times \left[1 - \mathcal{Q}\left(\frac{z}{\sqrt{N_o(N_o + 1)/2}}\right)\right],$$

where $\mathcal{Q}(x)$ is the complementary error function

$$\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-t^2/2\right\} dt,$$

and $\{T_j\}$ are the thresholds of the receiver's soft-decision quantizer given by

$$T_j = \begin{cases} -\infty, & \text{if } j = -1, \\ (j+1-2^{m-1})\Delta, & \text{if } j = 0, 1, ..., 2^m - 2, \\ +\infty, & \text{if } j = 2^m - 1 \end{cases} \tag{69}$$

with uniform step-size Δ. For each channel SNR, the suitable quantization step Δ is chosen as in [41], [4] to yield the maximum capacity of the binary-input 2^m-output DMC.

We compute the JSCC and tandem coding exponents for the binary source and the binary-input 2^m-output DMC converted from the AWGN (Rayleigh-fading, respectively) channel under transmission rate $t = 0.75$ ($t = 1$, respectively), and illustrate the power gain due to JSCC. In Figs. 9 and 10, we plot E_J and E_T for binary DMS $Q = \{0.1, 0.9\}$ and $m = 1, 2, 3$ by varying the channel SNR (in dB). We point out that in both the two figures, when SNR ≤ 6 dB for $m = 2, 3$ and when SNR ≤ 8 dB for $m = 1$, E_J and E_T are determined exactly. We observe that for the same SNR, E_J is almost twice as large as E_T. Furthermore, for the same exponent and the same (asymptotic) encoding length, JSCC would yield the same probability of error as tandem coding with a power gain of more than 2 dB. A similar behavior was noted for other values of transmission rate t.

5 JSCC Error Exponent with Hamming Distortion Measure

Let $\mathcal{S}$ be a finite set and $d(\cdot, \cdot)$ be a distortion measure, i.e., a nonnegative valued function d defined on $\mathcal{S} \times \mathcal{S}$ and extended to $\mathcal{S}^n \times \mathcal{S}^n$ by setting

$$d(s^n, \widetilde{s}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(s_i, \widetilde{s}_i).$$

A JSC code with blocklength n and transmission rate $t > 0$ for a tn-length DMS $\{Q : \mathcal{S}\}$ and a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with a threshold Δ of tolerated distortion is a pair of mappings $f_n : \mathcal{S}^{tn} \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{tn}$. The probability of the code exceeding the threshold Δ is given by

$$P_{\Delta}^{(n)}(Q, W, t) \triangleq \sum_{\{(s^{tn}, y^n) : d(s^{tn}, \varphi_n(y^n)) > \Delta\}} Q_{tn}(s^{tn}) P_{n,Y|X}(y^n \mid f_n(s^{tn})),$$

where Q_{tn} and $P_{n,Y|X}$ are the tn- and n-dimensional product distributions corresponding to Q and $P_{Y|X}$ respectively. $P_{\Delta}^{(n)}(Q, W, t)$ is also called the probability of excess distortion.

We remark that for the JSCC with a distortion threshold, we allow that the source has a uniform distribution.

Definition 3 The JSCC error exponent $E_J^{\Delta}(Q, W, t)$ is defined as the largest number E^{Δ} for which there exists a sequence of JSC codes (f_n, φ_n) with blocklength n and transmission rate t such that

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

When there is no possibility of confusion, $E_J^{\Delta}(Q, W, t)$ will often be written E_J^{Δ}. In [18], Csiszár proved that for a DMS Q and a DMC W, the JSCC error exponent under distortion threshold Δ satisfies

$$\underline{E}_r^{\Delta}(Q, W, t) \leq E_J^{\Delta}(Q, W, t) \leq \overline{E}_{sp}^{\Delta}(Q, W, t), \tag{70}$$

where

$$\underline{E}_r^{\Delta}(Q, W, t) \triangleq \inf_{R>0} \left[tF\left(\frac{R}{t}, Q, \Delta\right) + E_r(R, W) \right] \tag{71}$$

and

$$\overline{E}_{sp}^{\Delta}(Q, W, t) \triangleq \inf_{R>0} \left[tF\left(\frac{R}{t}, Q, \Delta\right) + E_{sp}(R, W) \right]. \tag{72}$$

In the above,

$$F(R, Q, \Delta) = \inf_{P: R(P, \Delta) > R} D(P \parallel Q) \tag{73}$$

is the source error exponent with a fidelity criterion [37] and $R(P, \Delta)$ is the rate distortion function (e.g., [16], [19]). $E_r(R, W)$ and $E_{sp}(R, W)$ are the random-coding and sphere-packing bounds to the channel error exponent. Likewise, if the infimum in (71) or (72) is attained for a rate larger than the channel critical rate, then the lower and upper bounds coincide, and we can determine E_J^{Δ} exactly. Of course, the two bounds are nontrivial if and only if $tR(Q, \Delta) < C$ by the JSCC theorem.

It can be shown that $F(R, Q, \Delta)$ is a nondecreasing function in R. However, unlike $e(R, Q)$, $F(R, Q, \Delta)$ is not necessarily convex or even continuous in R [1], [37]. Therefore, it is hard to analytically compute the JSCC exponent E_J^{Δ} in general. In this section we only address the computation of E_J^{Δ} for a binary DMS and an arbitrary DMC under the Hamming distortion measure $d_H(\cdot, \cdot)$, given by

$$d_H(s, \widetilde{s}) = \begin{cases} 1, & \text{if } s \neq \widetilde{s}, \\ 0, & \text{if } s = \widetilde{s}. \end{cases} \tag{74}$$

We first need to derive a parametric form of $F(R, Q, \Delta)$. Define

$$E_s^{\Delta}(\rho, Q) \triangleq (1+\rho) \log\left(q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}} \right) - \rho h_b(\Delta). \tag{75}$$

Lemma 7 For binary DMS $Q \triangleq \{q, 1-q\}$ $(q \le 1/2)$ under the Hamming distortion measure (74) and distortion threshold Δ such that $\Delta \le 1/2$, the following hold.

$$F(R, Q, \Delta) = \begin{cases} +\infty, & R > 1 - h_b(\Delta), \\ \sup_{\rho \ge \rho_0} [\rho R - E_s^{\Delta}(\rho, Q)], & R(Q, \Delta) < R \le 1 - h_b(\Delta), \\ 0, & R \le R(Q, \Delta), \end{cases} \tag{76}$$

where the rate-distortion function $R(Q, \Delta) = h_b(q) - h_b(\Delta)$ and $\rho_0 = 0$ if $q \ge \Delta$; otherwise $R(Q, \Delta) = 0$ and ρ_0 is the unique root of equation $H(Q^{(\rho)}) = h_b(\Delta)$ such that $\rho_0 > 0$.

The proof of this lemma is given in Appendix D. It can be easily verified that $F(R, Q, \Delta)$ is continuous and convex in $R \in (-\infty, 1 - h_b(\Delta)]$ if $q \ge \Delta$ and $F(R, Q, \Delta)$ is continuous and convex in $R \in (0, 1 - h_b(\Delta)]$ and has a jump at $R = R(Q, \Delta) = 0$ if $q < \Delta$. According to Lemma 7, the source error exponent $tF(R/t, Q, \Delta)$ is the convex transform of $tE_s^{\Delta}(\rho, Q)$ in $[\rho_0, +\infty)$. Define the binary divergence by

$$\widetilde{D}(\Delta \parallel q) \triangleq \Delta \log \frac{\Delta}{q} + (1 - \Delta) \log \frac{1 - \Delta}{1 - q}. \tag{77}$$

Adopting the approach of Section 3, we can apply Fenchel's Duality Theorem to $\underline{E}_r^{\Delta}(Q, W, t)$ and $\overline{E}_{sp}^{\Delta}(Q, W, t)$ and obtain equivalent computable bounds.

Theorem 9 Given a binary DMS $(q \le 1/2)$ and a DMC W under the Hamming distortion measure and distortion threshold Δ $(\Delta \le 1/2)$, the JSCC exponent satisfies the following.

1) *Lower Bound:* If $0 \le \Delta < \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, then $\rho_0 < 1$ and

$$\underline{E}_r^{\Delta}(Q, W, t) = \max_{\rho_0 \le \rho \le 1} [T_r(\rho, W) - tE_s^{\Delta}(\rho, Q)], \tag{78}$$

Otherwise, if $\Delta \ge \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, then

$$\underline{E}_r^{\Delta}(Q, W, t) = t\widetilde{D}(\Delta \parallel q) + E_0(1, W). \tag{79}$$

2) *Upper Bound:*

$$\overline{E}_{sp}^{\Delta}(Q, W, t) = \sup_{\rho \ge \rho_0} [T_{sp}(\rho, W) - tE_s^{\Delta}(\rho, Q)]. \tag{80}$$

Since the above result is a simple extension of the results in Section 3, the proof is omitted and we hereby only provide the following remarks.

(a) Similar to the lossless case, if $t(h_b(q)-h_b(\Delta)) \geq C$, then $\underline{E}_r^{\Delta}(Q,W,t) = \overline{E}_{sp}^{\Delta}(Q,W,t) = 0$. If $R_\infty(W) > t(1-h_b(\Delta))$, then $\overline{E}_{sp}^{\Delta}(Q,W,t) = +\infty$.

(b) Note that when $\Delta \geq \sqrt{q}/(\sqrt{q}+\sqrt{1-q})$, $\underline{E}_r^{\Delta}(Q,W,t)$ in (71) is achieved at $R \downarrow 0^+$, and

$$\begin{aligned}
\underline{E}_r^{\Delta}(Q,W,t) &= \lim_{R\downarrow 0^+} \left[tF(\frac{R}{t}, Q, \Delta) + E_r(R,W) \right] \\
&= \lim_{R\downarrow 0^+} \left[t \inf_{P:R(P,\Delta)>\frac{R}{t}} D(P \parallel Q) + E_0(1,W) - R \right] \\
&= t\widetilde{D}(\Delta \parallel q) + E_0(1,W).
\end{aligned}$$

(c) In the special case where the binary source is uniform, i.e., $q = 1/2$, Theorem 9 reduces to

$$\max_{0\leq\rho\leq 1} \left[-\rho t(1-h_b(\Delta)) + T_r(\rho,W)\right] \leq E_J^{\Delta}(Q,W,t) \leq \sup_{\rho\geq 0} \left[-\rho t(1-h_b(\Delta)) + T_{sp}(\rho,W)\right].$$

This is clearly equivalent to

$$E_r\left(t(1-h_b(\Delta)), W\right) \leq E_J^{\Delta}(Q,W,t) \leq E_{sp}\left(t(1-h_b(\Delta)), W\right) \qquad (81)$$

by the definition of $T_r(\rho,W)$ and $T_{sp}(\rho,W)$. In other words, E_J^{Δ} is bounded by the channel random-coding and sphere-packing bounds at rate $t(1-h_b(\Delta))$. If $t(1-h_b(\Delta)) \geq R_{cr}(W)$, then E_J^{Δ} is exactly determined.

(d) When the source is nonuniform, $E_s^{\Delta}(\rho,Q) = E_s(\rho,Q) - \rho t h_b(\Delta)$ is strictly concave in ρ. In this case, the maximizer

$$\overline{\rho}^{\Delta} \triangleq \arg\sup_{\rho\geq\rho_0} [T_{sp}(\rho,W) - tE_s^{\Delta}(\rho,Q)]$$

is strictly larger than ρ_0 if $t(h_b(q)-h_b(\Delta)) < C$ and $R_\infty(W) \leq t(1-h_b(\Delta))$. Particularly, $\overline{\rho}^{\Delta} < \infty$ if $R_\infty(W) < t(1-h_b(\Delta))$. As counterparts of Lemma 5 and Corollary 1, it can be shown that the upper bound $\overline{E}_{sp}^{\Delta}(Q,W,t)$ in (72) is attained at $\overline{R}_m^{\Delta} = H(Q^{(\overline{\rho}^{\Delta})}) - h_b(\Delta)$ and the lower bound in (71) is attained at $\underline{R}_m^{\Delta} = H(Q^{(\underline{\rho}^{\Delta})}) - h_b(\Delta)$, where $\underline{\rho}^{\Delta} = \min\{\overline{\rho}^{\Delta}, 1\}$. Consequently, other similar results to the lossless case regarding these optimizers can be obtained.

Example 6 For a binary DMS $\{q, 1-q\}$ ($q \leq 0.5$) and a BSC (ε) under transmission rate $t = 1$, we compute the JSCC error exponent under the Hamming distortion measure with distortion threshold Δ ($\Delta < \frac{1}{2}$). In Fig. 11, if the pair (ε, q) is located in region $\mathbf{B}$, then the corresponding JSCC exponent can be determined exactly (the lower and upper bounds are equal). If (ε, q) is located in region $\mathbf{C}_1$, then E_J^{Δ} is bounded by (78) and (80). If (ε, q) is located in region $\mathbf{C}_2$, then E_J^{Δ} is bounded by (79) and (80). When $(\varepsilon, q) \in \mathbf{A}$, E_J^{Δ} is zero, and the error probability of this communication system converges to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in \mathbf{B} \cup \mathbf{C}_1 \cup \mathbf{C}_2$.

Fig. 12 shows the JSCC error exponent lower bound of the binary DMS $\{q, 1-q\}$ ($q \leq 0.5$) and BSC (ε) pairs under different distortion thresholds. We fix the BSC parameter $\varepsilon = 0.2$, and vary q from 0 to 0.5. In Fig. 12, Segment 1 is determined by (79), and Segments 2 and 3 are determined by (78). Furthermore, the lower bound coincides with the upper bound (80) in Segment 3; i.e., the JSCC exponent is exactly determined in Segment 3.

6 Conclusions

In this work, we establish equivalent parametric representations of Csiszár's lower and upper bounds for the JSCC exponent E_J of a communication system with a DMS and a DMC, and we obtain explicit conditions for which the JSCC exponent is exactly determined. As a result, the computation of the bounds for E_J is facilitated for arbitrary DMS-DMC pairs. Furthermore, the bounds enjoy closed-form expressions when the channel is symmetric. A byproduct of our result is the fact that Csiszár's random-coding lower bound for E_J is in general larger than Gallager's lower bound [23].

We also provide a systematic comparison between E_J and E_T, the tandem coding error exponent. We show that JSCC can at most double the error exponent vis-a-vis tandem coding by proving that $E_J \leq 2E_T$ and we provide the condition for achieving this doubling effect. In the case where this upper bound is not tight, we also establish sufficient explicit conditions under which $E_J > E_T$. Numerical results indicate that $E_J \approx 2E_T$ for a large class of DMS-DMC pairs, hence illustrating the substantial potential benefit of JSCC over tandem coding. This benefit is also shown to result into a power saving gain of more than 2 dB for a binary DMS and a BPSK-modulated AWGN/Rayleigh channel with finite output quantization. Finally, we partially investigate the computation of Csiszár's lower and upper bounds for the lossy JSCC exponent under the Hamming distortion measure, and obtain equivalent representations for these bounds using the same approach as for the lossless JSCC exponent.

A Proof of Theorem 2 and Corollary 1

Theorem 2 can be shown by a left- and right- derivatives argument combined with the results of Lemma 5. Let $s_l(R)$ and $s_r(R)$ be the left and right-slopes (or left- and right-derivatives) of $E_{sp}(R, W)$ at each $R > R_\infty(W)$. Let $r_l(R)$ and $r_r(R)$ be the left and right slopes of $E_r(R, W)$ at each $R \geq 0$. Let $\rho(R)$ be the slope of $te(R/t, Q)$ for any $R \in [tH(Q), t \log |\mathcal{S}|]$. It is easy to verify that these slopes have the following properties (cf. [13], [23], [43]):

(a) $s_l(R)$ and $s_r(R)$ exist for every $R > R_\infty(W)$ and are nondecreasing in R.

(b) $r_l(R)$ and $r_r(R)$ exist for every $R \geq 0$ and are nondecreasing in R.

(c) $s_l(R) \leq s_r(R) < -1$ for $R < R_{cr}(W)$, $-1 \leq s_l(R) \leq s_r(R) \leq 0$ for $R_{cr}(W) < R < C$, and $s_l(R) = s_r(R) = 0$ for $R > C$. $s_l(R_{cr}(W)) \leq -1 \leq s_r(R_{cr}(W))$ and $s_l(C) \leq 0 = s_r(C)$.

(d) $r_l(R) = r_r(R) = -1$ for $R < R_{cr}(W)$, $r_l(R) = s_l(R)$ for $R > R_{cr}(W)$, and $r_r(R) = s_r(R)$ for $R \geq R_{cr}(W)$. $r_l(R_{cr}(W)) = -1 \leq r_r(R_{cr}(W))$.

(e) $\rho(R)$ is a strictly increasing function of R and is determined by $R = tH\left(Q^{(\rho(R))}\right)$ for $tH(Q) \leq R \leq t \log |\mathcal{S}|$. Specifically, $\rho(tH(Q)) = 0$ and $\rho(t \log |\mathcal{S}|) = \infty$.

(f) $\overline{\rho}^* = \rho(\overline{R}_m)$, where $\overline{\rho}^*$ and $\overline{R}_m$ are defined in (37) and (39), respectively.

(a) and (b) follows from the convexity of $E_{sp}(R, W)$ for $R > R_\infty(W)$ and $E_r(R, W)$ for $R \geq 0$, see [43, pp. 113–114]. Recalling that $E_r(R, W)$ involves a straight-line section with slope -1 for $R \in [0, R_{cr}(W)]$ and $E_r(R, W) = E_{sp}(R, W)$ only for $R \geq R_{cr}(W)$, where they both are equal to 0 for $R \geq C$, we obtain (c) and (d) from (a) and (b). From (24), we know that $te(R/t, Q) = tD\left(Q^{(\rho^*)} \parallel Q\right)$ for $tH(Q) \leq R \leq t \log |\mathcal{S}|$, where ρ^* is the unique root of $tH(Q^{(\rho)}) = R$. Also, it is easy to verify [13] that such ρ^* is exactly the slope of $te(R/t, Q)$ at R, i.e.,

$$\frac{\partial te(R/t, Q)}{\partial R} = \rho^*.$$

Thus (e) follows. Recalling also that in Lemma 5 we have shown the relation $\overline{R}_m =$

$tH(Q^{(\overline{\rho}^*)})$, since there is unique ρ satisfying this equation, we obtain (f).

Based on the above setup, the following lemma illustrates the geometric conditions for which $\underline{E}_r(Q, W, t)$ and $\overline{E}_{sp}(Q, W, t)$ are attained.

Lemma 8 Let $tH(Q) < C$ and let $R_\infty(W) < t \log |\mathcal{S}|$. The minimum in (30) is attained at $\overline{R}_m$ if and only if $-s_l(\overline{R}_m) \geq \rho(\overline{R}_m) \geq -s_r(\overline{R}_m)$, and the minimum in (29) is attained at $\underline{R}_m$ if and only if $-r_l(\underline{R}_m) \geq \rho(\underline{R}_m) \geq -r_r(\underline{R}_m)$.

Proof:
1. *Forward part*: We only show the case for the upper bound $\overline{E}_{sp}(Q, W, t)$, since the case for the lower bound can be shown in a similar manner. We first show that a rate $R_1 \in [tH(Q), t \log |\mathcal{S}|]$ satisfying $-s_l(R_1) \geq \rho(R_1) \geq -s_r(R_1)$ must achieve the minimum in $\overline{E}_{sp}(Q, W, t)$. Define functions

$$f_1(R) \triangleq \begin{cases} E_{sp}(R, W) & \text{if} \quad R \leq R_1, \\ E_{sp}(R_1, W) - \frac{|s_l(R_1)| + |\rho(R_1)|}{2}(R - R_1) & \text{if} \quad R \geq R_1. \end{cases}$$

and

$$g_1(R) \triangleq \begin{cases} te\left(\frac{R}{t}, Q\right) & \text{if} \quad R \leq R_1, \\ te\left(\frac{R_1}{t}, Q\right) + \frac{|\rho(R_1)| + |s_l(R_1)|}{2}(R - R_1) & \text{if} \quad R \geq R_1. \end{cases}$$

Since $-s_l(R_1) \geq \rho(R_1)$ implies $s_l(R_1) \leq -(|s_l(R_1)| + |\rho(R_1)|)/2$ and $\rho(R_1) \leq (|\rho(R_1)| + |s_l(R_1)|)/2$, we claim that $f_1(R)$ and $g_1(R)$ are both convex functions and hence their sum is convex,

$$f_1(R) + g_1(R) = \begin{cases} te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) & \text{if} \quad R \leq R_1, \\ te\left(\frac{R_1}{t}, Q\right) + E_{sp}(R_1, W) & \text{if} \quad R \geq R_1. \end{cases}$$

Since the convex function $f_1(R) + g_1(R)$ is constant for $R \geq R_1$ (noting that the convexity is strict in the interval $[tH(Q), R_1]$), we may write

$$\min_{tH(Q) \leq R \leq R_1} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right] = te\left(\frac{R_1}{t}, Q\right) + E_{sp}(R_1, W).$$

Similarly, using the relation $\rho(R_1) \geq -s_r(R_1)$ we can construct convex functions

$$f_2(R) \triangleq \begin{cases} E_{sp}(R, W) & \text{if} \quad R \geq R_1, \\ E_{sp}(R_1, W) + \frac{s_r(R_1) - \rho(R_1)}{2}(R - R_1) & \text{if} \quad R \leq R_1. \end{cases}$$

and

$$g_2(R) \triangleq \begin{cases} te\left(\frac{R}{t}, Q\right) & \text{if} \quad R \geq R_1, \\ te\left(\frac{R_1}{t}, Q\right) + \frac{\rho(R_1) - s_r(R_1)}{2}(R - R_1) & \text{if} \quad R \leq R_1, \end{cases}$$

and use them to show that the minimum

$$\min_{R_1 \leq R \leq t \log|\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right]$$

is attained at R_1. Thus, R_1 is the minimizer of $\overline{E}_{sp}(Q, W, t)$, i.e.,

$$\min_{tH(Q) \leq R \leq t \log|\mathcal{S}|} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right] = te\left(\frac{R_1}{t}, Q\right) + E_{sp}(R_1, W).$$

2. *Converse part:* We assume $\overline{R}_m \in (R_\infty(W), t\log|\mathcal{S}|)$ achieves the minimum in (30) but $\rho(\overline{R}_m) < -s_r(\overline{R}_m)$. Note that $\rho(t\log|\mathcal{S}|) = \infty > -s_r(t\log|\mathcal{S}|)$ provided that $t\log|\mathcal{S}| > R_\infty(W)$. Now let R_1 be the smallest rate in $[R_\infty(W), t\log|\mathcal{S}|]$ satisfying $\rho(R_1) \geq -s_r(R_1)$. According to our assumption together with (a) and (e), $R_1 > \overline{R}_m$. However, using our previous method, we can construct two convex functions $f_1(R)$ and $g_1(R)$ associated with R_1 to show

$$\min_{tH(Q) \leq R \leq R_1} \left[te\left(\frac{R}{t}, Q\right) + E_{sp}(R, W) \right] = te\left(\frac{R_1}{t}, Q\right) + E_{sp}(R_1, W).$$

This is clearly contradicted with the assumption that the minimum is attained at $\overline{R}_m$, a rate smaller than R_1, since there is unique minima due to the strict convexity. Thus, at $\overline{R}_m$ we must have $\rho(\overline{R}_m) \geq -s_r(\overline{R}_m)$. Consequently, we can show in a similar manner that $\rho(\overline{R}_m) \leq -s_l(\overline{R}_m)$. ■

The following facts immediately follow from Lemma 8.

Lemma 9 We have the following relations between $\overline{R}_m$ and $\underline{R}_m$:
(1). If $\overline{R}_m > R_{cr}(W)$ or $\underline{R}_m > R_{cr}(W)$, then $\underline{R}_m = \overline{R}_m > R_{cr}(W)$ and $\overline{E}_{sp}(Q, W, t) = \underline{E}_r(Q, W, t)$.
(2). If $\overline{R}_m = R_{cr}(W)$, then $\underline{R}_m \leq R_{cr}(W)$.
(3). $\overline{R}_m \geq \underline{R}_m$.

Proof: (1) is trivial since $E_r(R, W) = E_{sp}(R, W)$ for $R \geq R_{cr}(W)$. If $\overline{R}_m = R_{cr}(W)$, then by Lemma 8 and (d), $\rho(R_{cr}(W)) \geq -s_r(R_{cr}(W)) = -r_r(R_{cr}(W))$. Using Lemma 8 again we obtain (2). To show (3), we only need to show the case when $\overline{R}_m < R_{cr}(W)$.

According to Lemma 8 together with (c) and (d), we see $\rho(\overline{R}_m) > 1$ and $\rho(\underline{R}_m) = 1$. It follows from (e) that $\overline{R}_m > \underline{R}_m$.

■

This lemma emphasizes that when the JSCC error exponent upper bound is achieved at a rate equal to the channel critical rate $R_{cr}(W)$, the lower bound could be achieved at a rate smaller than $R_{cr}(W)$.

In the sequel we shall use properties (c)-(f), and Lemmas 5, 8 and 9 to prove Theorem 2. To show $A \Longleftrightarrow B \Longleftrightarrow C$, we only need to show: $A \Longrightarrow B$ (Forward) and $B \Longrightarrow C \Longrightarrow A$ (Converse).

1. *Converse Part.* We start from

$$
\begin{aligned}
\overline{\rho}^* < 1 \Longrightarrow & \quad \rho(\overline{R}_m) < 1 && \text{(by (f))} \\
\Longrightarrow & \quad \overline{R}_m < tR_{cr}^{(s)}(Q) && \text{(by (e))} \\
\text{and} & \quad s_r(\overline{R}_m) > -1 && \text{(by Lemma 8)} \\
\Longrightarrow & \quad \overline{R}_m \geq R_{cr}(W) && \text{(by (c))} \\
\Longrightarrow & \quad tR_{cr}^{(s)}(Q) > \underline{R}_m = \overline{R}_m > R_{cr}(W) && \text{(by Lemma 9 (1))} \qquad (82) \\
\text{or} & \quad tR_{cr}^{(s)}(Q) > \overline{R}_m = R_{cr}(W) \geq \underline{R}_m && \text{(by Lemma 9 (2))} \qquad (83) \\
\Longrightarrow & \quad 0 < \underline{\rho}^* = \overline{\rho}^* < 1 && \qquad (84) \\
\text{and} & \quad tR_{cr}^{(s)}(Q) > \overline{R}_m = \underline{R}_m \geq R_{cr}(W), && \qquad (85)
\end{aligned}
$$

where (84) and (85) are explained as follows. We first claim $\underline{\rho}^* < 1$, because $\underline{\rho}^* = 1$ would yield $\underline{R}_m \geq tR_{cr}^{(s)}(Q)$ by Lemma 5 (3), which is contradicted with (82) and (83). Since now $\underline{\rho}^* < 1$, from Lemma 8 and (d) we know $\underline{R}_m \geq R_{cr}(W)$. Thus in (83) we must have $\underline{R}_m = R_{cr}(W)$ and consequently (82) and (83) can both be summarized by (85).

Meanwhile, $\underline{\rho}^* = \overline{\rho}^*$ follows by Lemma 5. If now

$$
\begin{aligned}
\overline{\rho}^* = 1 \Longrightarrow \quad & \rho(\overline{R}_m) = 1 && \text{(by (f))} \\
\Longrightarrow \quad & \overline{R}_m = tR_{cr}^{(s)}(Q) && \text{(by (e))} \\
\text{and} \quad & s_l(\overline{R}_m) \le -1 \le s_r(\overline{R}_m) && \text{(by Lemma 8)} \\
\Longrightarrow \quad & \overline{R}_m \ge R_{cr}(W) && \text{(by (c))} \\
\Longrightarrow \quad & tR_{cr}^{(s)}(Q) = \underline{R}_m = \overline{R}_m > R_{cr}(W) && \text{(by Lemma 9 (1))} \qquad (86) \\
\text{or} \quad & tR_{cr}^{(s)}(Q) = \overline{R}_m = R_{cr}(W) \ge \underline{R}_m && \text{(by Lemma 9 (2))} \qquad (87) \\
\Longrightarrow \quad & \underline{\rho}^* = \overline{\rho}^* = 1 && \qquad (88) \\
\text{and} \quad & tR_{cr}^{(s)}(Q) = \underline{R}_m = \overline{R}_m \ge R_{cr}(W), && \qquad (89)
\end{aligned}
$$

where (88) and (89) are explained as follows. We first claim that $\underline{\rho}^* = 1$. If $\underline{\rho}^* < 1$, then by Lemma 5 (3) we have $\underline{R}_m < tR_{cr}^{(s)}(Q)$. In (86), we see $\underline{R}_m = tR_{cr}^{(s)}(Q)$, contradicted. In (87), it is still impossible that $\underline{R}_m < tR_{cr}^{(s)}(Q) = R_{cr}(W)$, because in that case we have $\rho(\underline{R}_m) < \rho(tR_{cr}^{(s)}(Q)) = 1$ by (e), which violates Lemma 8 since $\underline{R}_m < R_{cr}(W)$ implies $\rho(\underline{R}_m) = 1$. Thus we must have $\underline{\rho}^* = 1$ and (88) follows. According to Lemma 5 (3) again, $\underline{\rho}^* = 1$ implies $\underline{R}_m \ge tR_{cr}^{(s)}(Q)$. Hence in (87) we must have $\underline{R}_m = tR_{cr}^{(s)}(Q)$. (86) and (87) can both be summarized by (89). Next if

$$
\begin{aligned}
\overline{\rho}^* > 1 \Longrightarrow \quad & \rho(\overline{R}_m) > 1 && \text{(by (f))} \\
\Longrightarrow \quad & \overline{R}_m > tR_{cr}^{(s)}(Q) && \text{(by (e))} \qquad (90) \\
\text{and} \quad & s_l(\overline{R}_m) < -1 && \text{(by Lemma 8)} \\
\Longrightarrow \quad & \overline{R}_m \le R_{cr}(W) && \text{(by (c))} \\
\Longrightarrow \quad & \underline{R}_m \le \overline{R}_m \le R_{cr}(W) && \text{(by Lemma 9 (1) and (3))} \\
\Longrightarrow \quad & \underline{R}_m < R_{cr}(W) && \qquad (91) \\
\Longrightarrow \quad & r_l(\underline{R}_m) = -1 = r_r(\underline{R}_m) && \text{(by (d))} \\
\Longrightarrow \quad & \rho(\underline{R}_m) = 1 && \text{(by Lemma 8)} \\
\Longrightarrow \quad & \underline{R}_m = tR_{cr}^{(s)}(Q) && \text{(by (e))} \qquad (92) \\
\Longrightarrow \quad & \underline{\rho}^* = 1 && \text{(by Lemma 5 (3))} \\
\text{and} \quad & \overline{R}_m > \underline{R}_m. && \text{(by (90) and (92))}.
\end{aligned}
$$

To see (91), we let $\underline{R}_m = \overline{R}_m = R_{cr}(W)$. Then using (d) and Lemma 8 yields $\rho(\underline{R}_m) \le 1$, which is contradicted with the assumption $\rho(\underline{R}_m) = \rho(\overline{R}_m) > 1$. To show the last step, we assume $\underline{\rho}^* < 1$, then Lemma 5 (3) ensures $\underline{R}_m = tH(Q^{(\underline{\rho}^*)}) < tR_{cr}^{(s)}(Q)$, which is contradicted with the last second step.

2. *Forward Part.* First recall that $\rho(tR_{cr}^{(s)}(Q)) = 1$ by (e). Now if $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$, then $\overline{R}_m$ cannot be strictly larger than $tR_{cr}^{(s)}(Q)$ because in that case $\rho(\overline{R}_m) > \rho(tR_{cr}^{(s)}(Q)) = 1$, $-s_l(\overline{R}_m) \leq 1$ by (c), which violates Lemma 8. It then follows $\overline{R}_m \leq tR_{cr}^{(s)}(Q)$ and hence $\overline{\rho}^* \leq 1$ by (e). Conversely, if $tR_{cr}^{(s)}(Q) < R_{cr}(W)$, then $\overline{R}_m$ cannot be less than (or equal to) $tR_{cr}^{(s)}(Q)$ because in that case $\rho(\overline{R}_m) \leq \rho(tR_{cr}^{(s)}(Q)) = 1$, $-s_r(\overline{R}_m) > 1$ by (c), which violates Lemma 8. It then follows $\overline{R}_m > tR_{cr}^{(s)}(Q)$ and hence $\overline{\rho}^* > 1$ by (e).

Finally, we should note that when $tR_{cr}^{(s)}(Q) < R_{cr}(W)$, or $\overline{\rho}^* > 1$, the lower bound is achieved by $\underline{R}_m = tR_{cr}^{(s)}(Q) < R_{cr}(W)$ and $\underline{\rho}^* = 1$. Thus

$$\begin{aligned}
\underline{E}_r(Q,W,t) &= te\left(\frac{\underline{R}_m}{t}, Q\right) + E_r(\underline{R}_m, W) \\
&= \left[\underline{\rho}^*\underline{R}_m - tE_s(\underline{\rho}^*, Q)\right] + \left[E_0(1,W) - \underline{\rho}^*\underline{R}_m\right] \\
&= E_0(1,W) - tE_s(1,Q).
\end{aligned}$$

Meanwhile, Corollary 1 immediately follows by the above argument. ■

B Proof of Theorem 7

We first recall that if $-t\log(|\mathcal{S}|\overline{Q(s)}) < E(t\log|\mathcal{S}|, W)$, then there is no intersection between $te(R/t,Q)$ and $E(R,W)$. Clearly, the tandem coding exponent satisfies

$$\begin{aligned}
E_T(Q,W,t) &= E(t\log|\mathcal{S}|, W) \\
&= E_r(t\log|\mathcal{S}|, W) && (93)\\
&< E_r(\underline{R}_m, W) && (94)\\
&\leq E_J(Q,W,t),
\end{aligned}$$

Here, (93) follows by hypothesis $R_{cr}(W) \leq t\log|\mathcal{S}|$. (94) holds since $\underline{R}_m$ must be a quantity smaller than $t\log|\mathcal{S}|$ by Corollary 1.

We hence assume that $-t\log(|\mathcal{S}|\overline{Q(s)} \geq E(t\log|\mathcal{S}|, W)$, i.e., we assume that $te(R/t,Q)$ and $E(R,W)$ intersect at rate R_o. If $R_o \geq R_{cr}(W)$, which means that $E_o(1,W) - R_{cr}(W) \geq te(R_{cr}(W)/t, Q)$, then Theorem 6 guarantees that $E_J > E_T$. If $\underline{R}_m \geq R_{cr}(W)$, which implies $tR_{cr}^{(s)}(Q) \geq R_{cr}(W)$ by Corollary 2. This ensures $E_J > E_T$ by Theorem 6.

Furthermore, if $R_{cr}(W) > \underline{R}_m \geq R_o$, then

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

In the remaining, we assume that $te(R/t, Q)$ and $E(R,W)$ intersect at rate R_o and that $\underline{R}_m < R_o < R_{cr}$.

For a DMC with $E_{ex}(0,W) < \infty$, we may define the upper bound of the channel error exponent by

$$E_s(R,W) \triangleq \begin{cases} E_{sl}(R,W), & 0 \leq R \leq R_s, \\ E_{sp}(R,W), & R_s \leq R \leq C, \end{cases}$$

where $E_{sl}(R,W)$ is the straight-line upper bound for the channel error exponent, and R_s is the rate where the straight-line upper bound is tangent to the sphere-packing bound and $R_s \leq R_{cr}(W)$ [19], [23]. Clearly, $E_s(R,W)$ is also convex in $0 \leq R \leq C$, and it is shown in [19], [23] that

$$E_s(0,W) = E_{sl}(0,W) = E_{ex}(0,W).$$

Now connect $(0, E_s(0,W))$ and $(R_{cr}(W), E_s(R_{cr}(W),W))$ with a straight line, denoted by l_1, where

$$E_s(R_{cr}(W),W) = E_r(R_{cr}(W),W) = E_0(1,W) - R_{cr}(W).$$

Again, connect $(\underline{R}_m, te(\underline{R}_m/t, Q))$ and $(t \log|\mathcal{S}|,\ te(\log|\mathcal{S}|, Q))$ with a straight line, denoted by l_2, where

$$te\left(\frac{\underline{R}_m}{t}, Q\right) = tD(Q^{(1)} \parallel Q),$$

and

$$te(\log|\mathcal{S}|, Q) = -t\log(|\mathcal{S}|\overline{Q(s)}).$$

Suppose that the intersection of $E_s(R,W)$ and $te(R/t,Q)$ is $(R_1, te(R_1/t,Q))$, and that the intersection of l_1 and l_2 is (R_l, E_{R_l}). By assumption, R_o, the intersection of $te(R/t,W)$ and $E(R,W)$, is strictly larger than $\underline{R}_m$ and strictly less than $R_{cr}(W)$; hence by definition, R_1, the intersection of $te(R/t,W)$ and $E_s(R,W)$, must be strictly larger than $\underline{R}_m$ and strictly less than $R_{cr}(W)$, i.e., $\underline{R}_m < R_1 \leq R_o < R_{cr}(W)$. Likewise, it is easily seen that $\underline{R}_m < R_l < R_{cr}(W)$. Furthermore, because of the convexity of $te(R/t, Q)$ and $E_s(R, W)$ in the region $[\underline{R}_m, R_{cr}(W)]$, E_{R_l} must be strictly larger than $te(R_1/t, Q)$ (as $te(R/t, W)$ is strictly convex in this interval). It follows that

$$E_J(Q, W, t) \geq E_0(1, W) - tE_s(1, Q) \geq E_{R_l} > te\left(\frac{R_1}{t}, Q\right) \geq te\left(\frac{R_o}{t}, Q\right) = E_T(Q, W, t).$$

■

C Proof of Theorem 8

As in the previous proof, we only consider the case $-t\log_2(|\mathcal{S}|\overline{Q(s)}) \geq E(t\log_2|\mathcal{S}|, W)$ and $\underline{R}_m < R_o < R_{cr}(W)$. Thus, we can upper bound E_T by

$$\begin{aligned} E_T(Q, W, t) &= te(\frac{R_o}{t}, Q) \\ &< te\left(\frac{R_{cr}(W)}{t}, Q\right) \\ &= tD\left(Q^{(\gamma)} \parallel Q\right)\right) \end{aligned}$$

by the strict monotonicity of the source error exponent. On the other hand, Theorem 2 gives that

$$E_J(Q, W, t) \geq E_0(1, W) - tE_s(1, Q).$$

By assumption, if $E_0(1, W) - tE_s(1, Q) \geq tD\left(Q^{(\gamma)} \parallel Q\right)\right)$, then $E_J > E_T$. ■

D Proof of Lemma 7

Recall that the rate-distortion function $R(Q, \Delta)$ for a binary DMS $Q = \{q, 1-q\}$ under the Hamming distortion measure is given by (e.g., [16])

$$R(Q, \Delta) = \begin{cases} h_b(q) - h_b(\Delta), & 0 \leq \Delta \leq q, \\ 0, & \Delta > q. \end{cases} \tag{95}$$

Clearly, $F(R, Q, \Delta) = 0$ for $R \leq 0$ since the infimum in (73) is attained at $P = Q$. Similarly, since $R(P, \Delta) \leq 1 - h_b(\Delta)$ for all P, $F(R, Q, \Delta) = \infty$ for $R > 1 - h_b(\Delta)$. For the remainder of the proof, we assume $0 < R \leq 1 - h_b(\Delta)$.

(1) *Case of* $0 \leq \Delta \leq q$. For $R \leq R(Q,\Delta) = h_b(q) - h_b(\Delta)$, we have

$$F(R,Q,\Delta) = \inf_{P:R(P,\Delta)>R} D(P \parallel Q) = D(P \parallel Q)\Big|_{P=Q} = 0.$$

For $h_b(q) - h_b(\Delta) < R \leq 1 - h_b(\Delta)$, we have

$$\begin{aligned}
F(R,Q,\Delta) &= \inf_{P:R(P,\Delta)>R} D(P \parallel Q) \\
&= \min_{P \triangleq \{p,1-p\}:R(P,\Delta)=R} D(P \parallel Q) &(96)\\
&= \min_{p:h_b(p)-h_b(\Delta)=R} D(P \parallel Q) \\
&= e(R+h_b(\Delta),Q), \qquad \text{for } H(Q) \leq R + h_b(\Delta) \leq \log|\mathcal{S}| &(97)\\
&= \sup_{\rho\geq 0}[\rho(R+h_b(\Delta)) - E_s(\rho)] &(98)\\
&= \sup_{\rho\geq 0}[\rho R - E_s^{\Delta}(\rho,Q)].
\end{aligned}$$

Here (96) follows from the facts that the continuous function $\theta(p) \triangleq p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ is increasing for $p \geq q$ and $R(P,\Delta)$ given in (95) is continuous and increasing in p for $\Delta \leq p \leq \frac{1}{2}$. In (97), we note that $H(Q) = h_b(q)$ and that $\log|\mathcal{S}| = 1$ as the source is binary. (98) follows by the well known parametric form of source exponent function introduced by Blahut [13] and noting that $R' \triangleq R + h_b(\Delta) \in [H(Q), \log|\mathcal{S}|]$.

(2) *Case of* $\Delta > q$. For $0 < R \leq 1 - h_b(\Delta)$, similarly as (97), we have

$$F(R,Q,\Delta) = e(R',Q) = \sup_{\rho\in A}[\rho R' - E_s(\rho)],$$

where $R' = R + h_b(\Delta)$ such that $H(Q) < h_b(\Delta) < R' \leq 1 = \log|\mathcal{S}|$ and

$$\begin{aligned}
A &= \left\{\rho^* : \frac{\partial[\rho R' - E_s(\rho)]}{\partial\rho}\Big|_{\rho=\rho^*} = 0, \qquad h_b(\Delta) \leq R' \leq 1\right\}\\
&= \left\{\rho^* : h_b(\Delta) \leq R' = H(Q^{(\rho^*)}) \leq 1\right\}\\
&= \{\rho^* : \rho_0 \leq \rho^* < \infty\}, &(99)
\end{aligned}$$

where ρ_0 is the unique root of equation $H(Q^{(\rho)}) = h_b(\Delta)$ and $\rho_0 > 0$. Here (99) follows from the monotone property of $H(Q^{(\rho)})$. Therefore, we write

$$F(R,Q,\Delta) = \sup_{\rho\geq\rho_0}[\rho R - E_s^{\Delta}(\rho,Q)].$$

In fact, it can be shown that ρ_0 is the right slope of $F(R,Q,\Delta)$ at $R = R(Q,\Delta)$. ■

References

[1] R. Ahlswede, "Extremal properties of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 36, pp. 166–171, Jan. 1990.

[2] F. Alajaji, N. Phamdo, N. Farvardin, and T. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inform. Theory*, vol. 42, No. 1, pp. 230–239, Jan. 1996.

[3] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech and Audio Processing*, vol. 4, no. 5, pp. 325–336, Sept. 1996.

[4] F. Alajaji and N. Phamdo, "Soft-decision COVQ for Rayleigh-fading channels," *IEEE Commun. Lett.*, vol. 2, pp. 162–164, June 1998.

[5] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1041–1056, May 1998.

[6] E. Arikan and N. Merhav, "Joint source-channel coding and guessing with application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1756–1769, Sept. 1998.

[7] S. Arimoto, "'On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 19, pp. 357–359, May 1973.

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

[9] E. Ayanoğlu and R. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, vol. 33, pp. 855–865, Nov. 1987.

[10] V. B. Balakirsky, "Joint source-channel coding with variable length codes," *Probl. Inform. Transm.*, vol. 1, no. 37, pp. 10-23, Jan.–Mar. 2001.

[11] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4237-4246, Dec. 2005.

[12] D. P. Bertsekas, with A. Nedić and A. E. Ozdagler, *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA, 2003.

[13] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, July 1974.

[14] Brian D. Bunday, *Basic Optimisation Methods*. London: Arnold, 1984.

[15] P.-N. Chen and F. Alajaji, “Optimistic Shannon coding theorems for arbitrary single-user systems,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 2623–2629, Nov. 1999.

[16] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

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

[18] I. Csiszár, “On the error exponent of source-channel transmission with a distortion threshold,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 823–828, Nov. 1982.

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

[20] J. G. Dunham and R. M. Gray, “Joint source and noisy channel trellis encoding,” *IEEE Trans. Inform. Theory*, vol. 27, pp. 516–519, July 1981.

[21] N. Farvardin, “A study of vector quantization for noisy channels,” *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799–809, July 1990.

[22] T. Fine, “Properties of an optimum digital system and applications,” *IEEE Trans. Inform. Theory*, vol. 10, pp. 443–457, Oct. 1964.

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

[24] M. Gastpar, B. Rimoldi and M. Vetterli, “To code, or not to code: lossy source-channel communication revisited,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1147–1158, May 2003.

[25] J. D. Gibson and T. R. Fisher, “Alphabet-constrained data compression,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 443–457, May 1982.

[26] R. M. Gray and D. S. Ornstein, “Sliding-block joint source/noisy-channel coding theorems,” *IEEE Trans. Inform. Theory*, vol. 22, pp. 682–690, Nov. 1976.

[27] J. Hagenauer, “Source-controlled channel decoding,” *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sep. 1995.

[28] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2003.

[29] M. E. Hellman, “Convolutional source encoding,” *IEEE Trans. Inform. Theory*, vol. 21, pp. 651–656, Nov. 1975.

[30] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sep. 1997.

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

[32] V. N. Koshelev, "Direct sequential encoding and decoding for discrete sources," *IEEE Trans. Inform. Theory,* vol. 19, pp. 340–343, May 1973.

[33] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Eng. Jpn.*, vol. 67-B, no. 4, pp. 39–47, 1984.

[34] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Transactions on Communication Technology*, vol. COM-17, pp. 291–302, Apr. 1969.

[35] J. Lim and D. L. Neuhoff, "Joint and tandem source-channel coding with complexity and delay constraints," *IEEE Trans. Commun.*, vol. 51, pp. 757–766, May 2003.

[36] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1969.

[37] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

[38] J. L. Massey, "Joint source and channel coding," in *Communications and Random Process Theory,* J. K. Skwirzynski, ed., The Netherlands: Sijthoff and Nordhoff, pp. 279–293, 1978.

[39] D. Miller and K. Rose, "Combined source-channel vector quantization using deterministic annealing," *IEEE Trans. Commun.*, vol. 42, pp. 347–356, Feb.-Apr. 1994.

[40] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. 27, pp. 1644-1659, Nov. 1979.

[41] Nam Phamdo and Fady Alajaji, "Soft-decision demodulation design for COVQ over white, colored, and ISI Gaussian channels," *IEEE Trans. Commun.*, vol. 46, No. 9, pp. 1499–1506, Sep. 2000.

[42] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[43] H. L. Royden, *Real Analysis*, Third Edition, New York, 1988.

[44] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Commun.*, vol. 39, pp. 838–846, June 1991.

[45] C. E. Shannon,"A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and pp. 623-656, Jul. and Oct. 1948.

[46] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[47] M. Skoglund and P. Hedelin, "Hadamard-based soft decoding for vector quantization over noisy channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 515–532, Mar. 1999.

[48] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1293–1307, May 1999.

[49] G. Taricco, "On the capacity of the binary input Gaussian and Rayleigh fading channels," *Eur. Trans. Telecommun.*, vol. 7, no. 2, Mar.-Apr. 1996.

[50] V. A. Vaishampayan and N. Farvardin, "Joint design of block source codes and modulation signal sets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1230–1248, July 1992.

[51] S. Vembu, S. Verdú and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, pp. 44–54, Jan. 1995.

[52] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, Inc., 1979.

[53] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. J. Weinberger, "Universal discrete denoising: known channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 5–28, Jan. 2005.

[54] K. A. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.

[55] Y. Zhong, F. Alajaji, and L. L. Campbell, "When is joint source-channel coding worthwhile: an information theoretic perspective," *Proc. 22nd Bienn. Symp. Commun.*, Canada, pp. 121–123, June 2004.

[56] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the computation of the joint source-channel error exponent for memoryless systems," *Proc. 2004 IEEE Int'l. Symp. Inform. Theory,* p. 477, June-July 2004.

[57] G.-C. Zhu, F. Alajaji, J. Bajcsy and P. Mitran, "Transmission of non-uniform memoryless sources via non-systematic Turbo codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1344–1354, Aug. 2004.

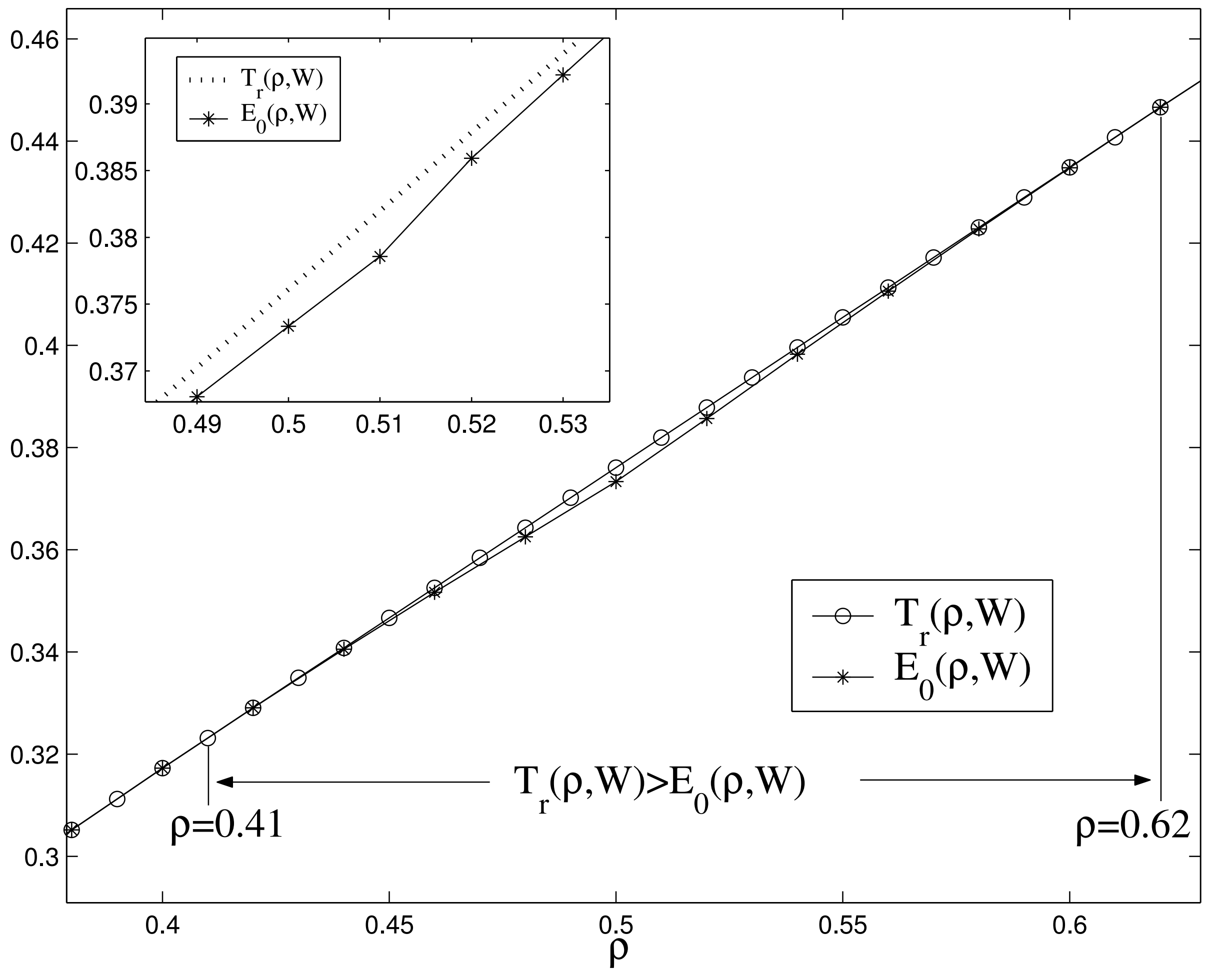

Figure 1: Example of a 6-ary input, 4-ary output DMC (see [23, Fig. 5.6.5]) for which $E_0(\rho, W)$ is not concave.

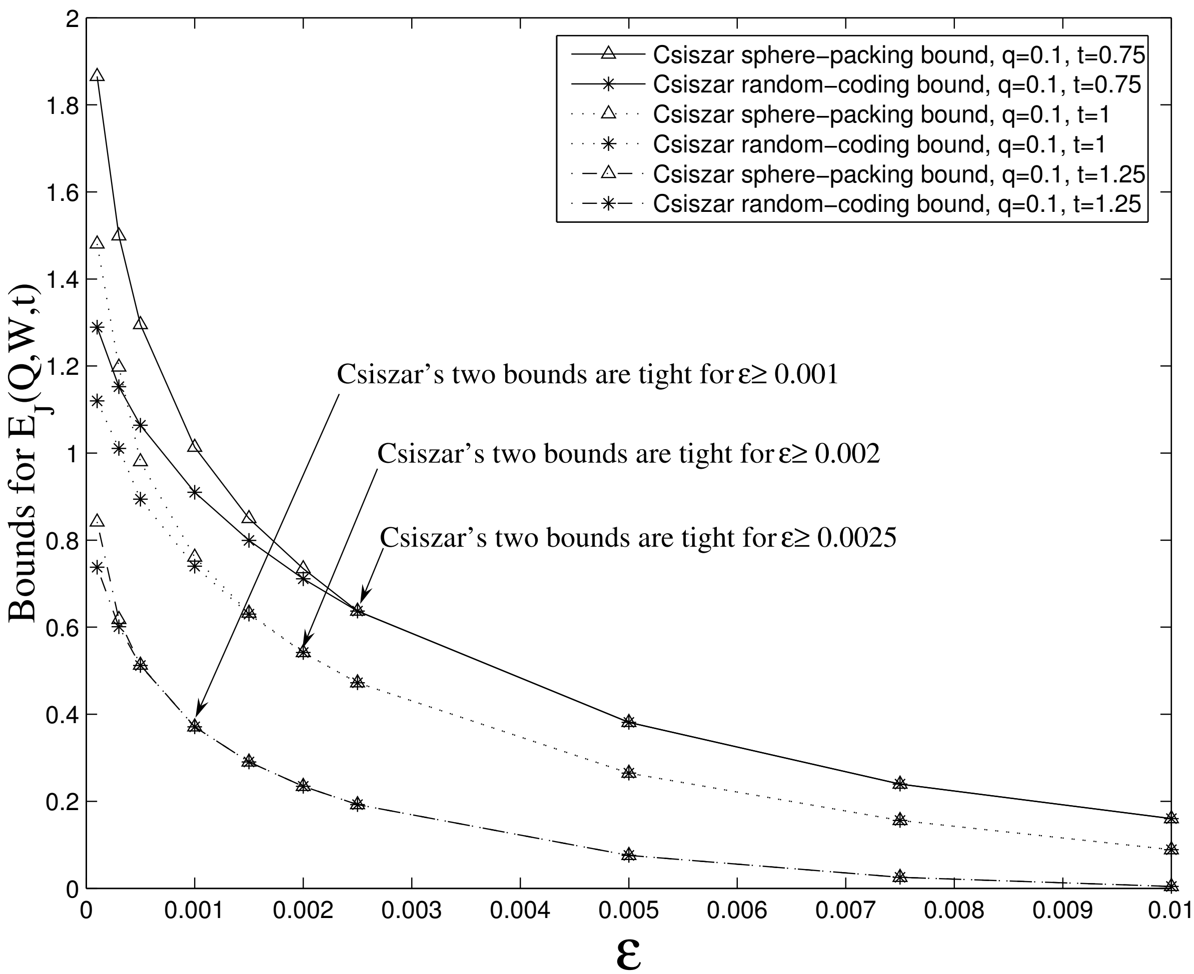

Figure 2: Csiszár's random-coding and sphere-packing bounds for the system of Example 1.

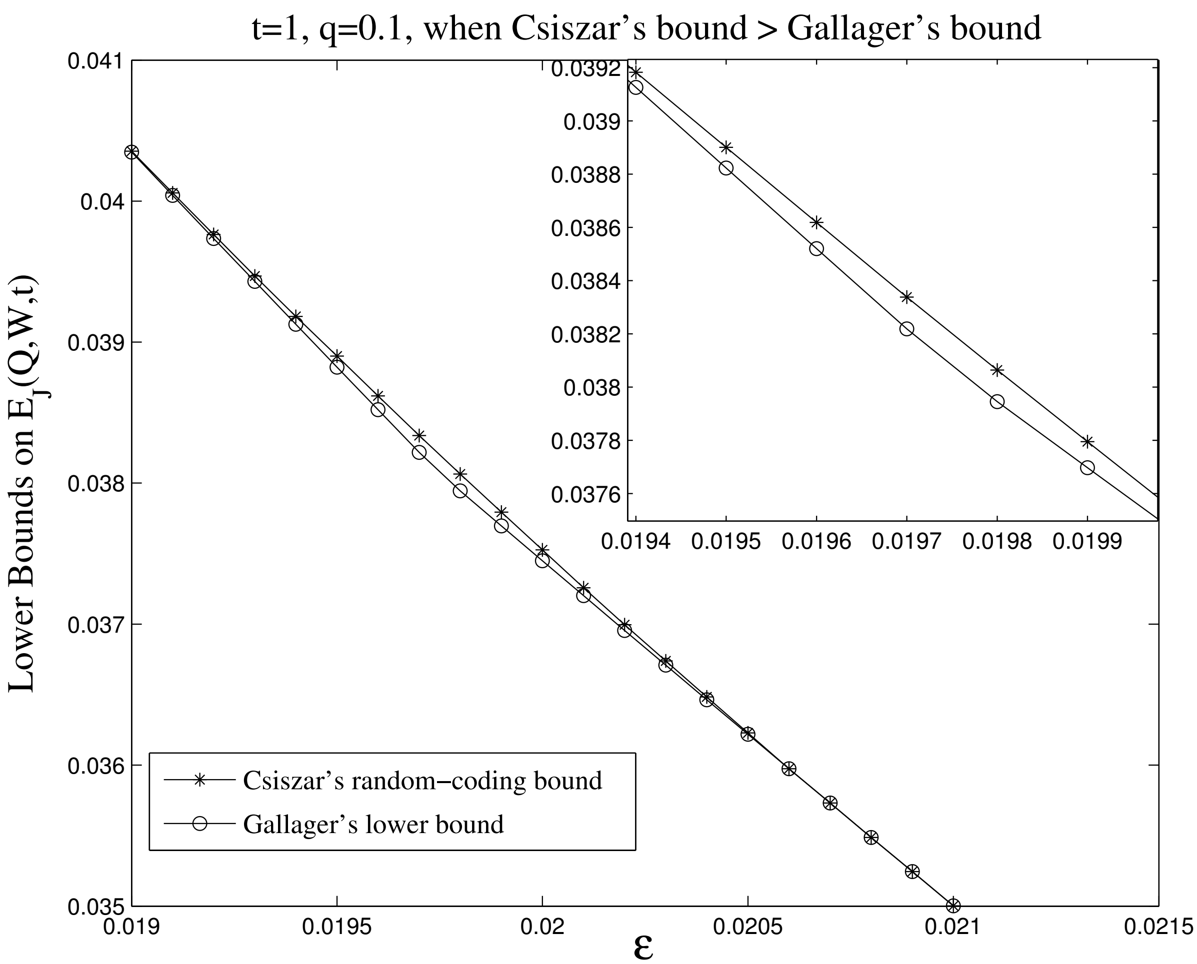

Figure 3: Csiszár's random-coding bound vs Gallager's lower bound for the system of Example 1.

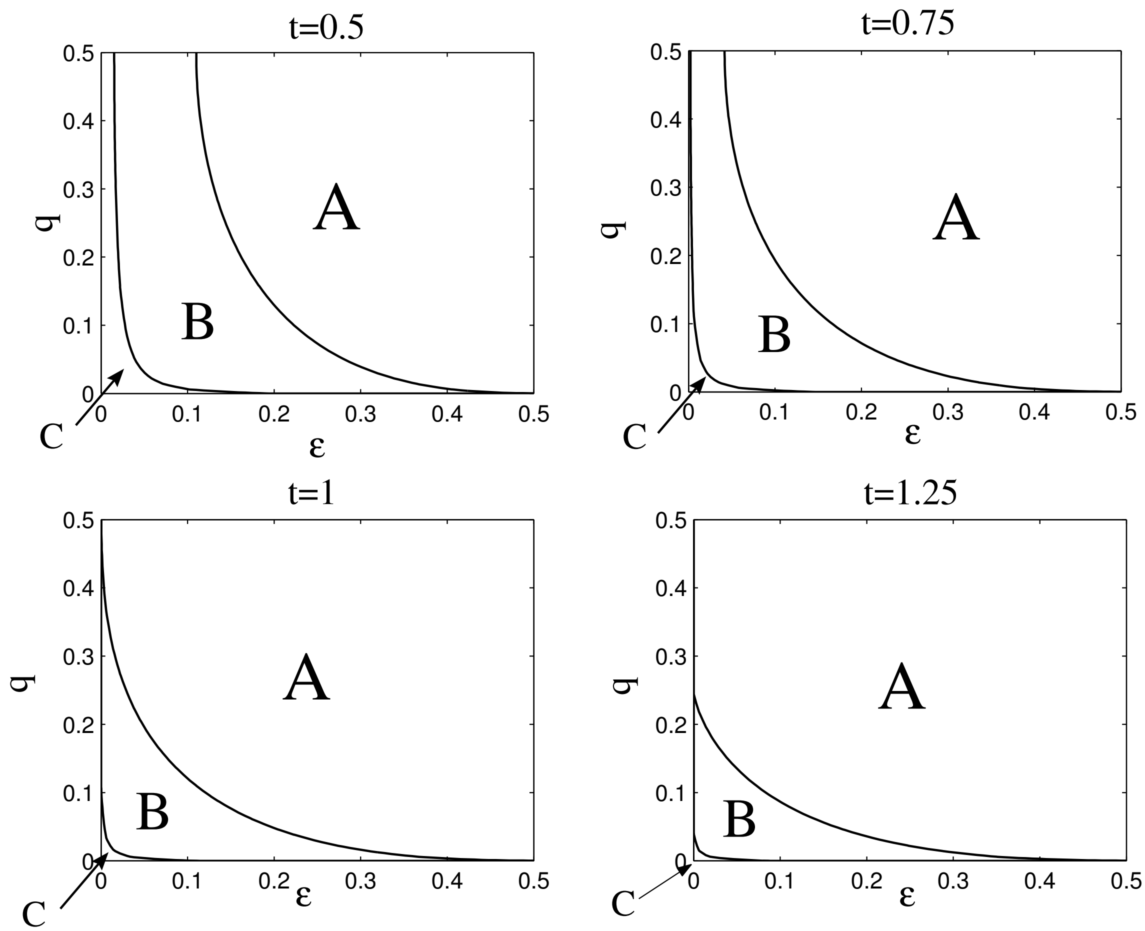

Figure 4: The regions for the (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 2 for different transmission rates t. Note that $E_J = 0$ on the boundary between **A** and **B**; E_J is exactly determined on the boundary between **B** and **C**. In **A**, $E_J = 0$. In **B**, E_J is positive and known exactly. In **C**, E_J is positive and can be bounded above and below.

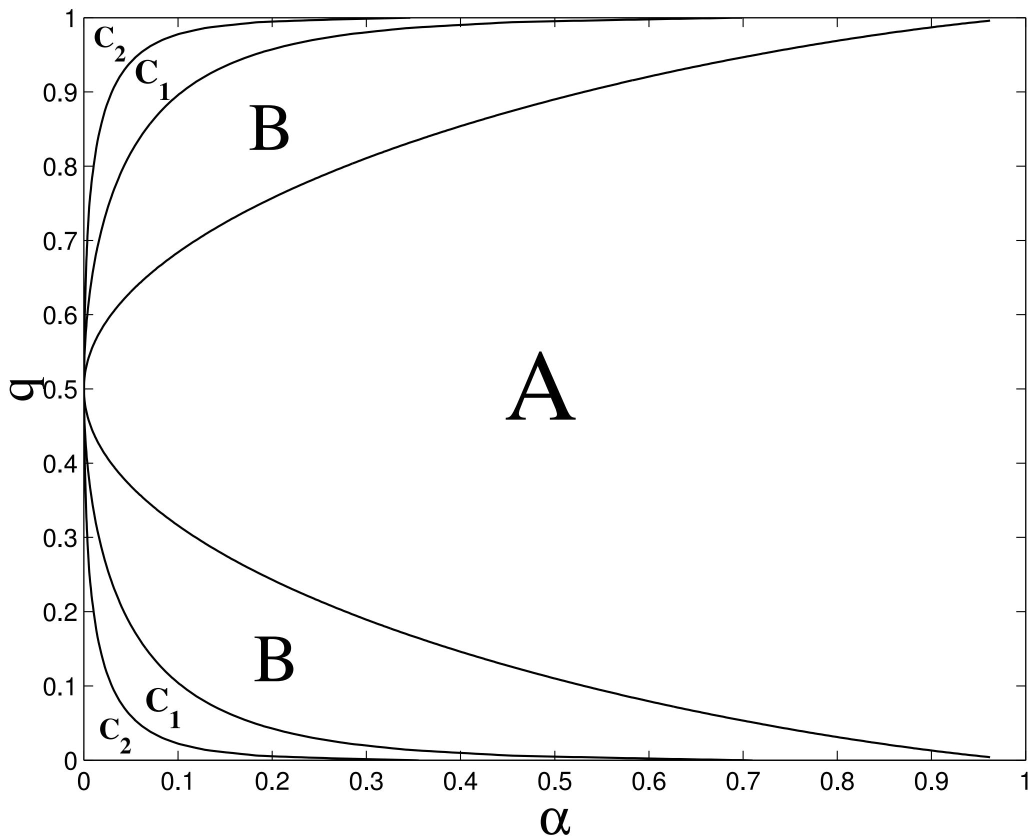

Figure 5: The regions for the (α, q) pairs in the binary DMS $\{q, 1-q\}$ and BEC (α) system of Example 3 with $t = 1$. Note that $E_J = 0$ on the boundary between **A** and **B**; E_J is determined on the boundary between **B** and $\mathbf{C}_1$; The random-coding bound and expurgated bound to E_J are equal on the boundary between $\mathbf{C}_1$ and $\mathbf{C}_2$.

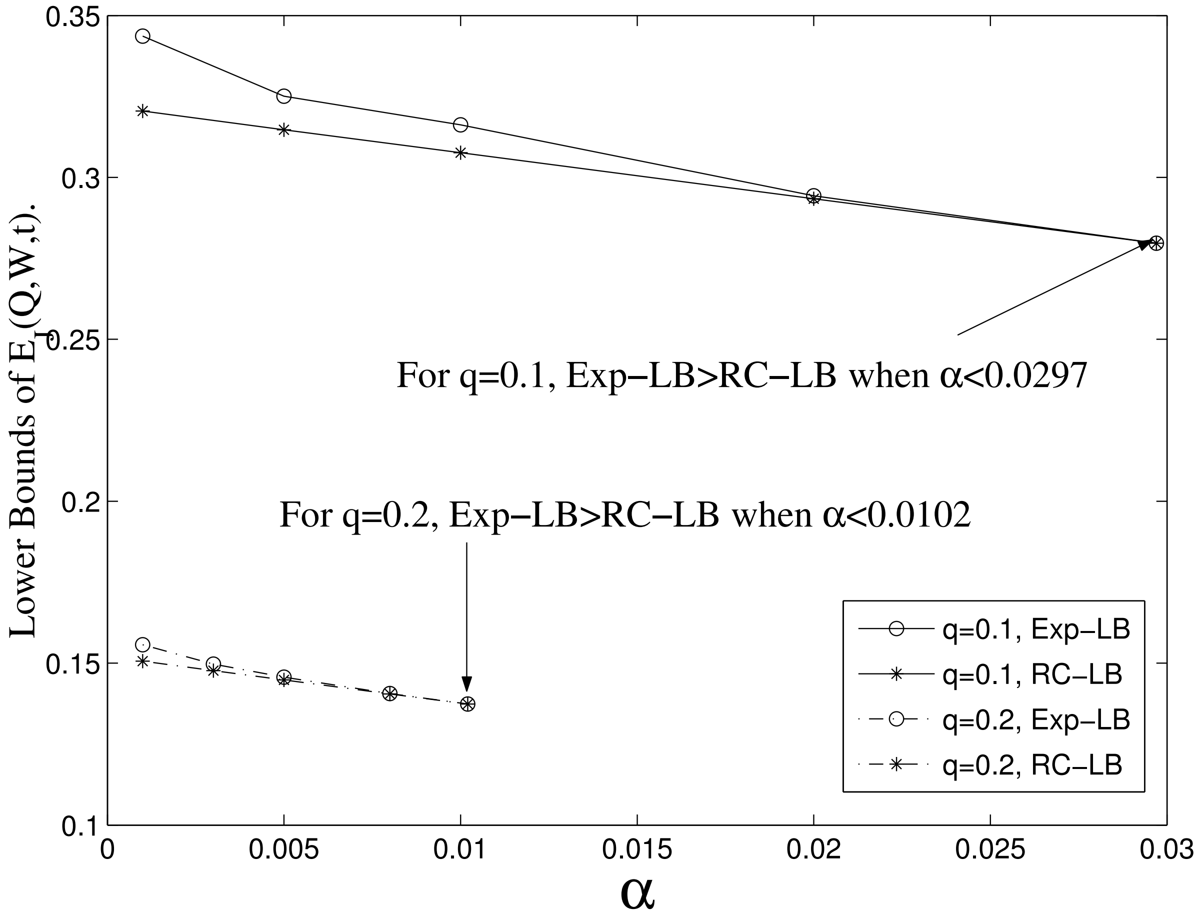

Figure 6: Improvement due to the expurgated lower bound for the binary DMS (α, q) and BEC (α) system with $t = 1$. Exp-LB and RC-LB stand for the expurgated and random-coding lower bounds, respectively.

E_J/E_T	t=0.5, q=0.1	t=0.75, q=0.1	t=0.75, q=0.15	t=1, q=0.05
$\varepsilon = 0.0005$	$1.0^\dagger$	$1.60^\dagger$	$1.58^\dagger$	$1.87^\dagger$
$\varepsilon = 0.001$	$1.0^\dagger$	$1.70^\dagger$	$1.68^\dagger$	$1.93^\dagger$
$\varepsilon = 0.005$	$1.36^\dagger$	$1.94^\dagger$	1.89	1.99
$\varepsilon = 0.01$	$1.70^\dagger$	1.95	1.91	2.0
$\varepsilon = 0.04$	1.85	1.97	1.95	2.0
$\varepsilon = 0.08$	1.91	1.99	1.96	2.0
$\varepsilon = 0.12$	1.95	1.97	2.0	2.0
$\varepsilon = 0.16$	1.96	1.95	N/A	2.0
$\varepsilon = 0.2$	1.86	N/A	N/A	N/A

Table 1: E_J/E_T for the binary DMS and BSC pairs of Example 5. "N/A" means that $tH(Q) > C$ such that $E_J = E_T = 0$. "†" means that this quantity is only a lower bound for E_J/E_T.

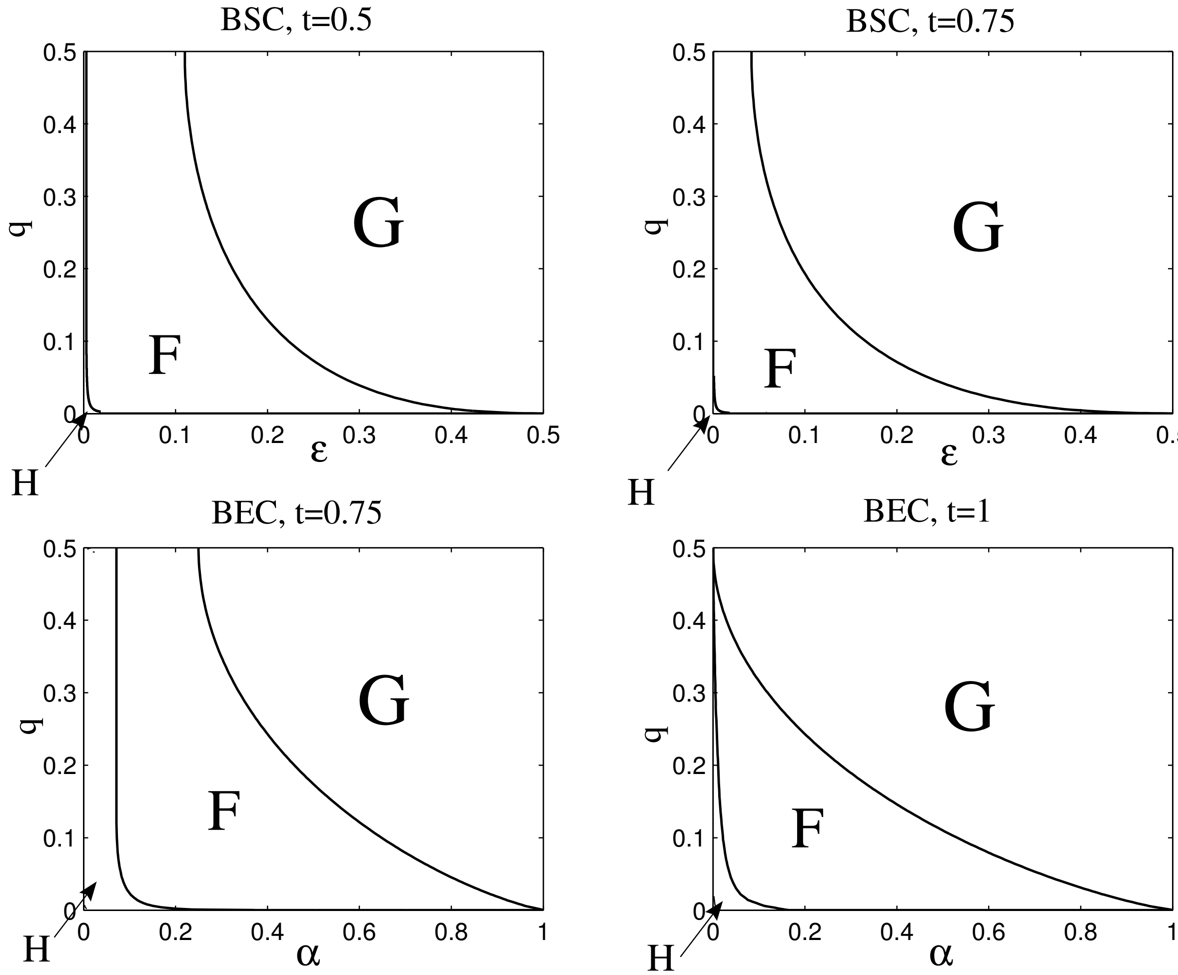

Figure 7: The regions for binary DMS-BSC (q, ε) pairs and binary DMS-BEC (q, α) pairs under different transmission rates t. In region **F** (including the boundary between **F** and **H**), $E_J > E_T > 0$; in region **G** (including the boundary between **G** and **F**), $E_J = E_T = 0$; and in region **H**, $E_J \geq E_T > 0$.

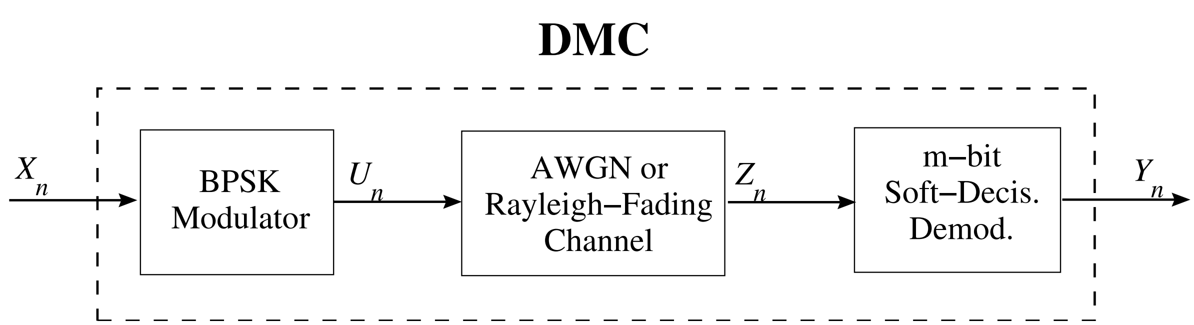

Figure 8: Binary-input AWGN or Rayleigh-fading channel with finite output quantization.

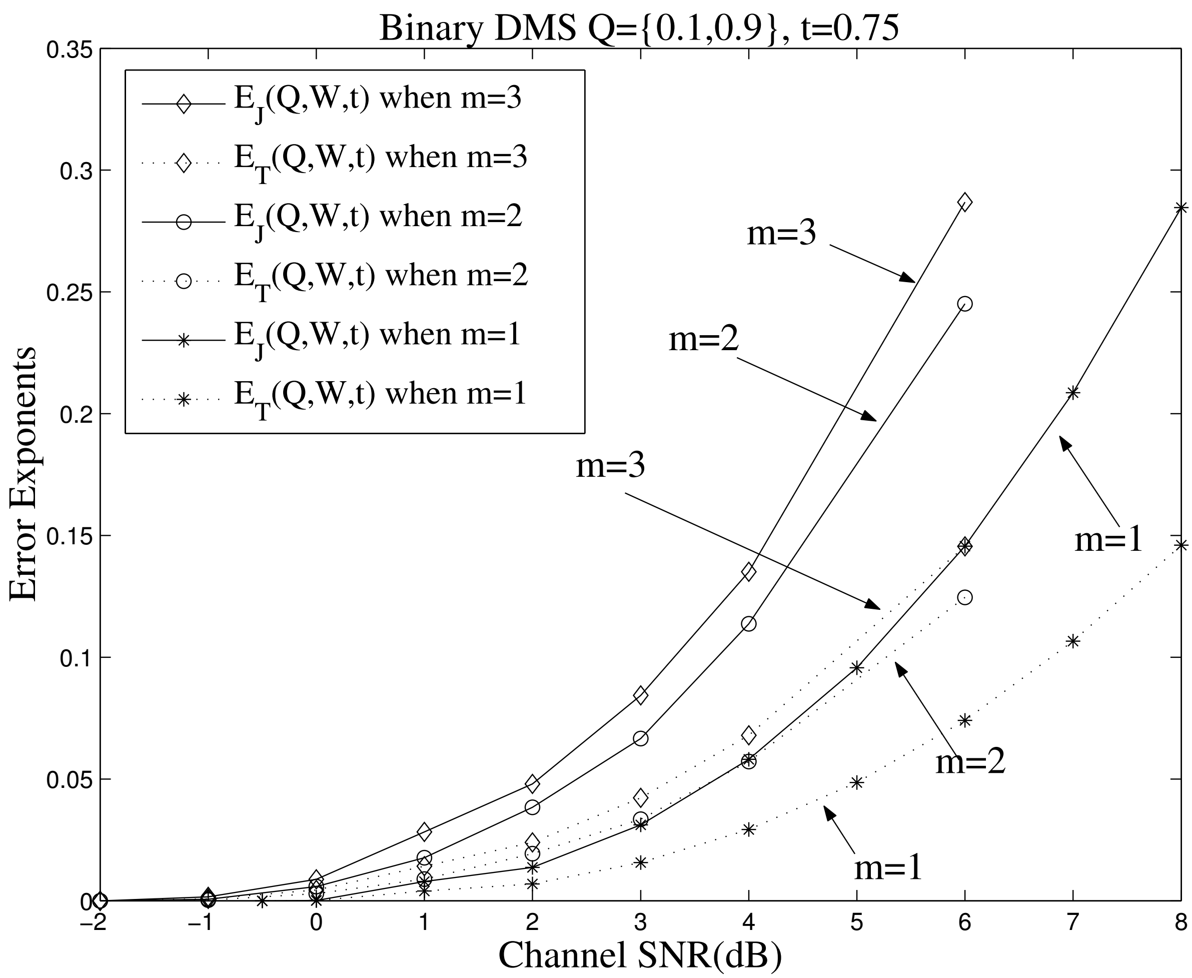

Figure 9: The power gain due to JSCC for binary DMS and binary-input 2^m-output DMC (AWGN channel) with $t = 0.75$.

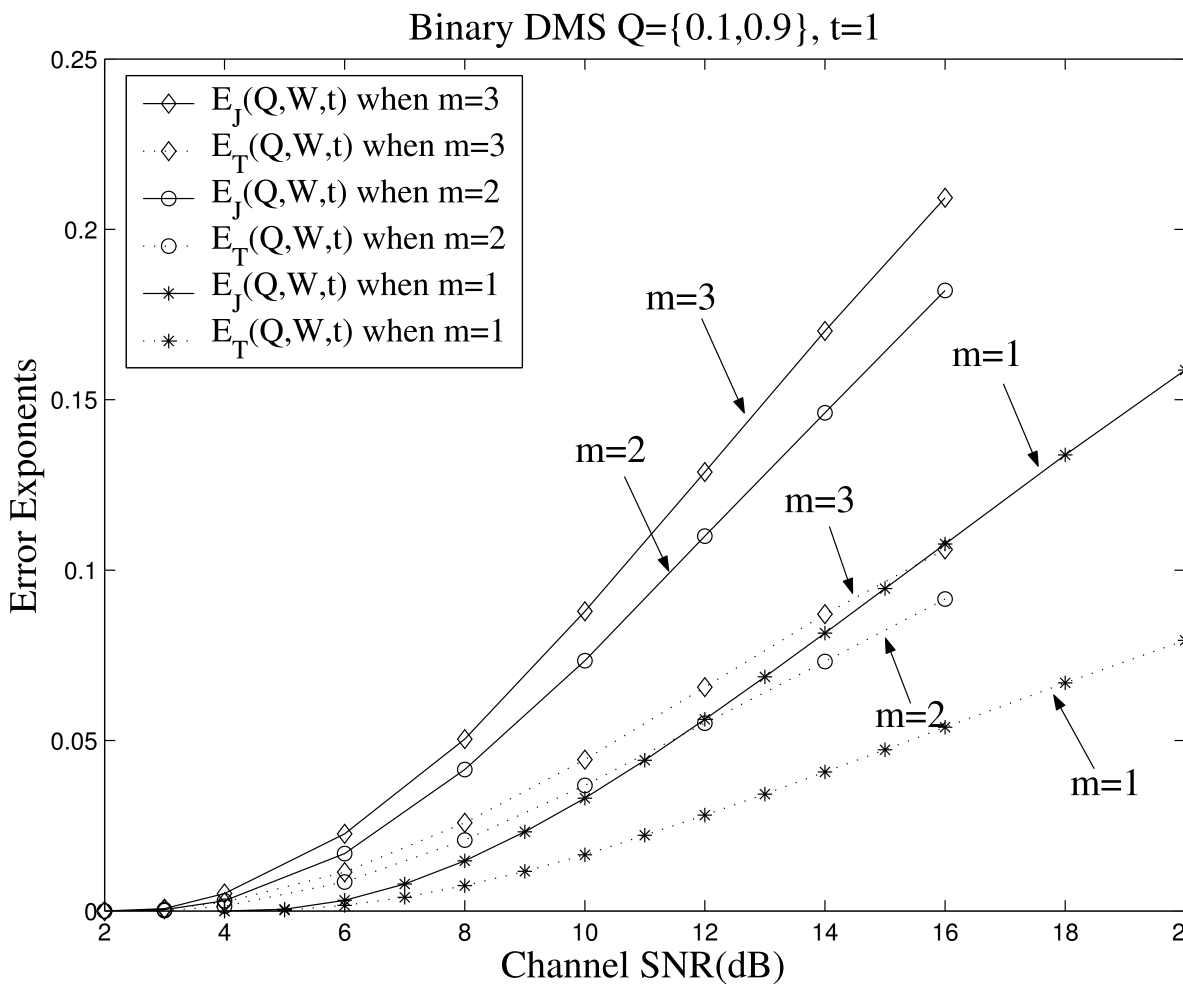

Figure 10: The power gain due to JSCC for binary DMS and binary-input 2^m-output DMC (Rayleigh-fading channel) with $t = 1$.

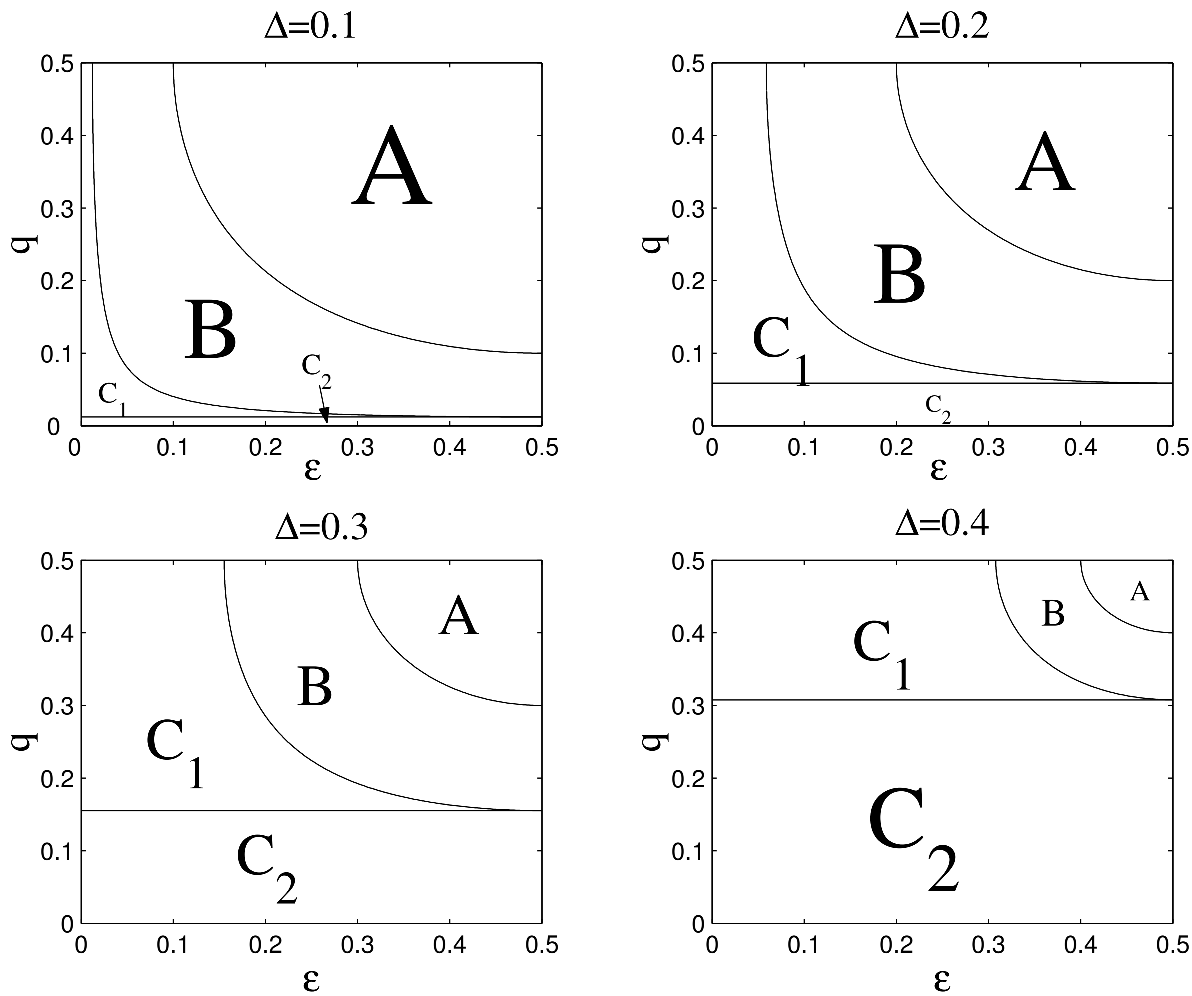

Figure 11: The regions for the (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 6 with Hamming distortion for different values of the distortion threshold Δ with $t = 1$. Note that $E_J^\Delta = 0$ on the boundary between **A** and **B**, and $E_J^\Delta > 0$ is determined on the boundary between **B** and $\mathbf{C}_1$.

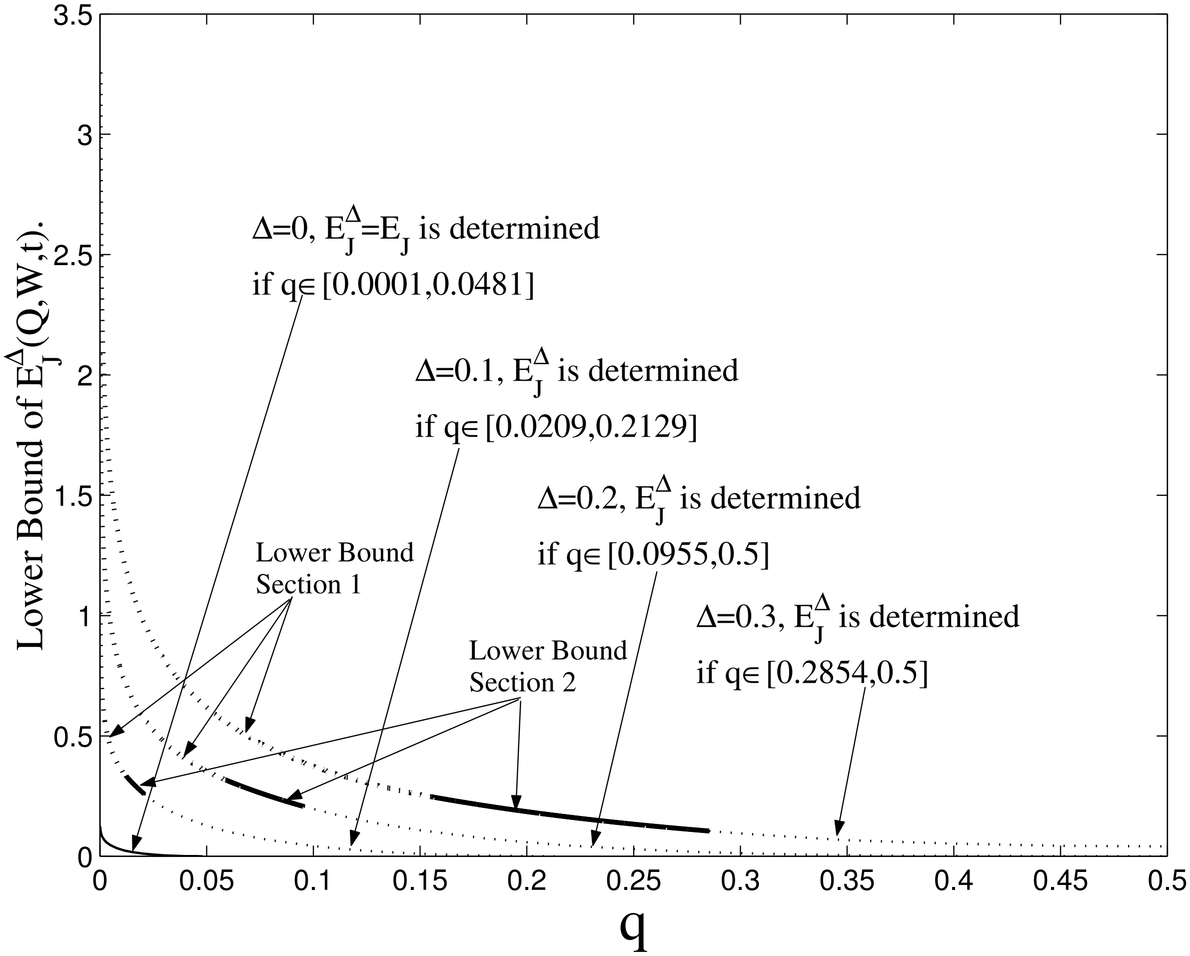

Figure 12: Fix $\varepsilon = 0.2$. The JSCC exponent lower bound of the binary DMS $\{q, 1-q\}$ ($q \leq 0.5$) and BSC (ε) pairs under Hamming distortion with $t = 1$. For $\Delta = 0$, E_J^Δ is determined if $q \in [0.0001, 0.0481]$, which is the same as the random-coding lower bound for the lossless JSCC error exponent. For $\Delta = 0.1$, E_J^Δ is determined if $q \in [0.0209, 0.2129]$. For $\Delta = 0.2$, E_J^Δ is determined if $q \in [0.0955, 0.5]$. For $\Delta = 0.3$, E_J^Δ is determined if $q \in [0.2854, 0.5]$.