Joint Source-Channel Coding Excess Distortion Exponent for Some Memoryless Continuous-Alphabet Systems*

Yangfan Zhong Fady Alajaji L. Lorne Campbell

Abstract

We investigate the joint source-channel coding (JSCC) excess distortion exponent E_J (the exponent of the probability of exceeding a prescribed distortion level) for some memoryless communication systems with continuous alphabets. We first establish upper and lower bounds for E_J for systems consisting of a memoryless Gaussian source under the squared-error distortion fidelity criterion and a memoryless additive Gaussian noise channel with a quadratic power constraint at the channel input. A necessary and sufficient condition for which the two bounds coincide is provided, thus exactly determining the exponent. This condition is observed to hold for a wide range of source-channel parameters. As an application, we study the advantage in terms of the excess distortion exponent of JSCC over traditional tandem (separate) coding for Gaussian systems. A formula for the tandem exponent is derived in terms of the Gaussian source and Gaussian channel exponents, and numerical results show that JSCC often substantially outperforms tandem coding. The problem of transmitting memoryless Laplacian sources over the Gaussian channel under the magnitude-error distortion is also carried out. Finally, we establish a lower bound for E_J for a certain class of continuous source-channel pairs when the distortion measure is a metric.

Index Terms: Continuous memoryless sources and channels, memoryless Gaussian and Laplacian sources, memoryless Gaussian channels, joint source-channel coding, tandem separate source and channel coding, probability of excess distortion, squared/magnitude-error distortion, excess distortion exponent, error exponent, Fenchel transform, Fenchel duality.

**IEEE Transactions on Information Theory:* Submitted April 2007; Revised July 2008. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Seattle, Washington, July 2006.

The authors are with the Dept. of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

1 Introduction

In [9], Csiszár studies the joint source-channel coding (JSCC) excess distortion exponent under a fidelity criterion for discrete memoryless systems – i.e., the largest rate of asymptotic decay of the probability that the distortion resulting from transmitting the source over the channel via a joint source-channel (JSC) code exceeds a certain tolerated threshold. Specifically, given a discrete memoryless source (DMS) Q and a discrete memoryless channel (DMC) W (both with finite alphabets), a transmission rate t and a distortion measure, Csiszár shows that the lower (respectively upper) bound of the JSCC excess distortion exponent $E_J(Q, W, \Delta, t)$ under a distortion threshold Δ is given by the minimum of the sum of $tF(R/t, Q, \Delta)$ and $E_r(R, W)$ (respectively $E_{sp}(R, W)$) over R, where $F(R, Q, \Delta)$ is the source excess distortion exponent with distortion threshold Δ [18], and $E_r(R, W)$ and $E_{sp}(R, W)$ are respectively the random-coding and sphere-packing channel error exponents [13]. If the minimum of the lower (or upper) bound is attained for an R larger than the critical rate of the channel, then the two bounds coincide and E_J is determined exactly. The analytical computation of these bounds has been partially addressed in [26], where the authors use Fenchel duality [17] to provide equivalent bounds for a binary DMS and an arbitrary DMC under the Hamming distortion measure.

Since many real-world communication systems deal with analog signals, it is important to study the JSCC excess distortion exponent for the compression and transmission of a continuous alphabet source over a channel with continuous input/output alphabets. For instance, it is of interest to determine the best performance (in terms of the excess distortion probability) that a source-channel code can achieve if a stationary memoryless Gaussian source (MGS) is coded and sent over a stationary memoryless Gaussian channel (MGC), i.e., an additive white Gaussian noise channel. To the best of our knowledge, the JSCC excess distortion exponent for continuous-alphabet systems has not been addressed before. In this work, we study the JSCC excess distortion exponent for the following classes of memoryless communication systems:

1. MGS-MGC systems with squared-error distortion measure;
2. Laplacian-source and MGC systems with magnitude-error distortion measure;
3. A certain class of continuous source-channel systems when the distortion is a metric.

For a Gaussian communication system consisting of an MGS P_S with the squared-error distortion and an MGC W with additive noise P_Z and the power input constraint, we show that the JSCC excess distortion exponent $E_J(P_S, W, \Delta, \mathcal{E}, t)$ with transmission rate t, under a distortion threshold Δ and power constraint $\mathcal{E}$, is upper bounded by the minimum of the sum of the Gaussian source excess distortion exponent $tF(R/t, P_S, \Delta)$ and the sphere-packing upper bound of the Gaussian channel error exponent $E_{sp}(R, W, \mathcal{E})$; see Theorem 1. The proof of the upper bound relies on a strong converse JSCC theorem (Theorem 8) and the judicious construction of an auxiliary MGS and an auxiliary MGC to lower bound the probability of excess distortion. We also establish a lower bound for $E_J(P_S, W, \Delta, \mathcal{E}, t)$; see Theorem 2. In fact, we derive the lower bound for MGS's and general continuous MC's with an input cost constraint. To prove the lower bound, we employ a concatenated "quantization – lossless JSCC" scheme as in [2], use the type covering lemma [10] for the MGS [1], and then bound the probability of error for the lossless JSCC part, which involves a memoryless source with a countably infinite alphabet and the memoryless continuous channel, by using a modified version of Gallager's random-coding bound for the JSCC error exponent for DMS-DMC pairs [13, Problem 5.16] (the modification is made to allow for input cost constrained channels with countably-infinite input alphabets and continuous output alphabets). This lower bound is expressed by the maximum of the difference of Gallager's constrained-input channel function $E_0(W, \mathcal{E}, \rho)$ and the source function $tE(P_S, \Delta, \rho)$. Note that when the channel is an MGC with an input power constraint, a computable but somewhat looser lower bound is obtained by replacing $E_0(W, \mathcal{E}, \rho)$ by Gallager's Gaussian-input channel function $\widetilde{E}_0(W, \mathcal{E}, \rho)$. Also note that the source function $E(P_S, \Delta, \rho)$ for the MGS is equal to the guessing exponent [1] and admits an explicit analytic form.

As in our previous work for discrete systems [26, 27], we derive equivalent expressions for the lower and upper bounds by applying Fenchel's Duality Theorem [17]. We show (in Theorem 3) that the upper bound, though proved in the form of a minimum of the sum of source and channel exponents, can also be represented as a (dual) maximum of the difference of Gallager's channel function $\widetilde{E}_0(W, \mathcal{E}, \rho)$ and the source function $tE(P_S, \Delta, \rho)$. Analogously, the lower bound, which is established in Gallager's form, can also be represented in Csiszár's form, as the minimum of the sum of the source exponent and the lower bound of the channel exponent. In this regard, our upper and lower bounds are natural extensions of Csiszár's upper and lower bounds from the case of (finite alphabet) discrete memoryless systems to the case of memoryless Gaussian systems. We then compare the upper and lower bounds using their equivalent forms and derive an explicit condition under which the two bounds coincide; see Theorem 4. We observe numerically that the condition is satisfied for a large class of source-channel parameters. We proceed by investigating the advantage of JSCC over traditional tandem (separate) coding in terms of the excess distortion exponent. We first derive a formula for the tandem coding excess distortion exponent when the distortion threshold is less than 1/4 of the source variance. Numerical results indicate that the JSCC exponent can be strictly superior to the tandem exponent for many MGS-MGC pairs.

We next observe that Theorems 1 and 2 can also be proved for memoryless Laplacian sources (MLS's) under the magnitude-error distortion measure. Using a similar approach, we establish upper and lower bounds for the JSCC excess distortion exponent for the lossy transmission of MLS's over MGC's (see Theorem 6). Finally, we considerably modify our approach in light of the result of [16] to prove a lower bound for some continuous source-channel pairs when the distortion measure is a metric. We show that the lower bound for MGS's and continuous memoryless channels (given in Theorem 2), expressed by the maximum of the difference of source and channel functions, still holds for a continuous source-channel pair if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E} \exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, where the expectation is taken over the source distribution defined on $\mathbb{R}$ (see Theorem 7). Although this condition does not hold for both MGS's with the squared-error distortion and MLS's with the magnitude-error distortion, it holds for generalized MGS's with parameters (α, σ) under the distortion $d(x, y) = |x - y|^p$, $p < \alpha$, and $p \leq 1$.

The rest of the paper is organized as follows. In Section 2, we summarize prior results on the source excess distortion and the channel error exponents, and we define the JSCC excess distortion exponent. In Section 3, we establish upper and lower bounds for E_J for Gaussian systems. A sufficient and necessary condition for which the upper and lower bounds coincide is provided. We also derive the tandem coding exponent and numerically compare it with the JSCC exponent for Gaussian systems. In Section 4, we extend our results for other source-channel pairs. Direct extensions without proof of the bounds for coding MLS's over MGC's are given in Section 4.1. In Section 4.2, we show a lower bound for E_J for a class of continuous source-channel pairs with a metric distortion measure and satisfying a finiteness condition. Finally, we draw conclusions in Section 5.

2 Notation and Definitions

All logarithms and exponentials throughout this paper are in the natural base. In the sequel $o(n)$ serves as a generic notation for a vanishing quantity with respect to n such that $\lim_{n\to\infty} o(n)/n = 0$. Likewise, $\zeta(\epsilon)$ serves as a generic notation for a vanishing quantity with respect to ϵ such that $\lim_{\epsilon\to 0}\zeta(\epsilon) = 0$. The expectation of the random variable (RV) X is denoted by $\mathbb{E}(X)$.

2.1 Source Excess Distortion Exponent

Let P_S be a (stationary) memoryless source (MS) with alphabet $\mathcal{S}$. If the source has a continuous alphabet, P_S stands for the probability density function (pdf) of the source (we only consider continuous sources for which a pdf exists). If an MS P_S is a DMS (with a countable alphabet $\mathcal{S}$), then P_S denotes the probability mass function (pmf) of the source. Consequently, the pdf (pmf) of a k-length source sequence $\mathbf{s} \triangleq (s_1, s_2, ..., s_k) \in \mathcal{S}^k$ is hence given by $P_{S^k}(\mathbf{s}) = \prod_{i=1}^k P_S(s_i)$. Let $d : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ be a single-letter distortion function. The distortion measure on $\mathcal{S}^k$ is defined as

$$d^{(k)}(\mathbf{s}, \mathbf{s}') \triangleq \frac{1}{k}\sum_{i=1}^{k} d(s_i, s_i')$$

for any $\mathbf{s} \triangleq (s_1, ..., s_k) \in \mathcal{S}^k, \mathbf{s}' \triangleq (s_1', ..., s_k') \in \mathcal{S}^k$. Given a distortion threshold $\Delta > 0$, the rate-distortion function for the MS P_S is given by (e.g., [5])

$$R(P_S, \Delta) = \inf_{P_{S'|S}:\mathbb{E}d(S,S')\le\Delta} I(S; S'), \tag{1}$$

where $I(S; S')$ is the mutual information between the source input and its representation, and the infimum is taken over all the conditional distributions $P_{S'|S}(\cdot|s)$ defined on $\mathcal{S}$ for any $s \in \mathcal{S}$ subject to the constraint $\mathbb{E}d(S, S') \le \Delta$.

A (k, M_k) block source code for an MS P_S is a pair of mappings: $f_{sk} : \mathcal{S}^k \longrightarrow \{1, 2, ..., M_k\}$ and $\varphi_{sk} : \{1, 2, ..., M_k\} \longrightarrow \mathcal{S}^k$. The code rate is defined by

$$R_k \triangleq \frac{1}{k}\ln M_k \quad \text{nats/source symbol.}$$

The probability of exceeding a given distortion threshold $\Delta > 0$ for the code $(f_{sk}, \varphi_{sk}, \Delta)$ is given by

$$P_\Delta^{(k)}(P_S, R_k) \triangleq \int_{\mathbf{s}:d^{(k)}(\mathbf{s},\varphi_{sk}(f_{sk}(\mathbf{s})))>\Delta} P_{S^k}(\mathbf{s})d\mathbf{s}. \tag{2}$$

Note that the integral should be replaced with a summation if P_S is a DMS. We call $P_\Delta^{(k)}(P_S, R_k)$ the probability of excess distortion for coding the MS P_S.

Definition 1 For any $R > 0$ and $\Delta > 0$, the excess distortion exponent $F(R, P_S, \Delta)$ of the MS P_S is defined as the supremum of the set of all numbers e for which there exists a sequence of (k, M_k) block codes (f_k, φ_k, Δ) with

$$e \leq \liminf_{k\to\infty} -\frac{1}{k} \ln P_{\Delta}^{(k)}(P_S, R_k)$$

and

$$R \geq \limsup_{k\to\infty} R_k.$$

It has been shown in [15, 16, 28] that the excess distortion exponent for some particular sources can be expressed in Marton's form [18]. In other words, we know that

$$F(R, P_S, \Delta) = \inf_{Q_S: R(Q_S,\Delta)>R} D(Q_S \parallel P_S), \tag{3}$$

where $D(Q_S \parallel P_S)$ is the Kullback-Leibler divergence between distributions Q_S and P_S, and the infimum is taken over all distributions Q_S defined on $\mathcal{S}$, holds for the following cases:

1. Finite-alphabet DMS's with arbitrary distortion measures [18];
2. MGS's with squared-error distortion measure [15];
3. MLS's with magnitude-error distortion measure [28];
4. (Stationary) MS's whose alphabets are complete metric spaces with a metric distortion measure $d(\cdot,\cdot)$ satisfying the condition that there exists an element $s_o \in \mathcal{S}$ with $\mathbb{E}\exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$ [16].

Note that Cases 2 and 3 are not included in Case 4 (since the squared-error distortion is not a metric, and the condition in Case 4 on the metric and the source distribution does not hold for both MGS's with squared-error distortion measure and MLS's with magnitude-error distortion measure). When P_S is an MGS (respectively MLS) with a squared-error (respectively magnitude-error) distortion measure, the explicit analytical form of $F(R, P_S, \Delta)$ will be given in Section 3 (respectively Section 4.1).

2.2 Channel Error Exponent

Let W be a (stationary) MC with continuous input and output alphabets $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \mathbb{R}$ and transition pdf $W \triangleq P_{Y|X}$. The conditional pdf of receiving $\mathbf{y} \triangleq (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$ at the channel output given that the codeword $\mathbf{x} \triangleq (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ is transmitted is given by $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$.

Given an input cost function $g : \mathcal{X} \to [0, \infty)$ such that $g(x) = 0$ if and only if $x = 0$, and a constraint $\mathcal{E} > 0$, the channel capacity of the MC W is given by

$$C(W, \mathcal{E}) = \sup_{P_X:\mathbb{E}g(X)\leq\mathcal{E}} I(X; Y), \tag{4}$$

where $I(X; Y)$ is the mutual information between the channel input and channel output, the supremum is taken over all channel input distributions P_X subject to the constraint $\mathbb{E}g(X) \leq \mathcal{E}$.

An (n, M_n) block channel code for an MC W with an input cost constraint $\mathcal{E}$ is a pair of mappings: $f_{cn} : \{1, 2, ..., M_n\} \longrightarrow \mathcal{X}^n$ and $\varphi_{cn} : \mathcal{Y}^n \longrightarrow \{1, 2, ..., M_n\}$, where f_{cn} is subject to an (arithmetic average) cost constraint:

$$f_{cn} \in \mathcal{F}_{cn}^{\mathcal{E}} \triangleq \left\{ f_{cn} : \frac{1}{n} \sum_{j=1}^{n} g(x_j) \leq \mathcal{E} \quad \text{for all} \quad \mathbf{x} = f_{cn}(i), \quad i \in \{1, 2, ..., M_n\} \right\}.$$

The code rate is defined as

$$R_n \triangleq \frac{1}{n} \ln M_n \quad \text{nats/channel use.}$$

The (average) probability of decoding error for the $(f_{cn}, \varphi_{cn}, \mathcal{E})$ code is given by

$$P_{ec}^{(n)}(W, R_n, \mathcal{E}) \triangleq \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j \neq i}^{M_n} \int_{\mathbf{y}:\varphi_{cn}(\mathbf{y})=j} P_{Y^n|X^n}(\mathbf{y}|f_{cn}(i)) d\mathbf{y}. \tag{5}$$

Definition 2 For any $R > 0$, the channel error exponent $E(R, W, \mathcal{E})$ of the channel W is defined as the supremum of the set of all numbers E for which there exists a sequence of (n, M_n) block codes $(f_{cn}, \varphi_{cn}, \mathcal{E})$ with

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \ln P_{ec}^{(n)}(W, R_n, \mathcal{E})$$

and

$$R \leq \liminf_{n\to\infty} R_n.$$

In contrast to the source excess distortion exponent, the channel error exponent is not known for general MC's (not even for the binary symmetric channels); it is partially determined for high rates for several families of MC's, such as DMC's with no input constraints ($\mathcal{E} = \infty$) and MGC's with an input quadratic power constraint. For the continuous MC W with a transition pdf $P_{Y|X}$, only a lower bound for $E(R, W, \mathcal{E})$ due to Gallager [12], [13, Section 7.3] is known, which we refer to as Gallager's random-coding lower bound for the channel error exponent $E(R, W, \mathcal{E})$,

$$E(R, W, \mathcal{E}) \geq E_r(R, W, \mathcal{E}) \triangleq \max_{0\leq\rho\leq1} [-\rho R + E_0(W, \mathcal{E}, \rho)], \tag{6}$$

where

$$E_0(W, \mathcal{E}, \rho) \triangleq \sup_{P_X:\mathbb{E}g(X)\leq\mathcal{E}, \mathbb{E}g(X)^3<\infty} \max_{r\geq0} E_0(\rho, r, W, P_X, \mathcal{E}) \tag{7}$$

is Gallager's constrained channel function with

$$E_0(\rho, r, W, P_X, \mathcal{E}) \triangleq -\ln \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} P_X(x) e^{r(g(x)-\mathcal{E})} P_{Y|X}(y|x)^{\frac{1}{1+\rho}} dx \right]^{1+\rho} dy,$$

and where the supremum in (7) is taken over all pdfs $P_X(x)$ defined on $\mathcal{X}$ subject to $\mathbb{E}g(X) \leq \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$. The constraints are satisfied, for example, when $g(x) = x^2$ and P_X is a Gaussian distribution with mean zero and variance $\mathcal{E}$. The integrals should be replaced with summations if W has discrete alphabets. Note that in general we do not have an explicit formula for this bound, because it is not known whether the supremum in (7) is achievable or not, and under what distribution it is achievable.

2.3 JSCC Excess Distortion Exponent

Given a source distribution measure $d(\cdot,\cdot)$ and a channel input function $g(\cdot)$, a joint source-channel (JSC) code $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$ with blocklength n and transmission rate t (source symbols/channel use) for the MS P_S, and the MC W with input cost constraint $\mathcal{E}$ is a pair of mappings: $f_n : \mathcal{S}^{tn} \longrightarrow \mathcal{X}^n$ and $\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{tn}$, where $f_n \in \mathcal{F}_n^{\mathcal{E}}$, and

$$\mathcal{F}_n^{\mathcal{E}} \triangleq \left\{ f_n : \frac{1}{n}\sum_{i=1}^{n} g(x_i) \leq \mathcal{E} \quad \text{for all} \quad \mathbf{x} = f_n(\mathbf{s}) \right\}. \tag{8}$$

Here $\mathbf{s} \in \mathcal{S}^{tn}$ is the transmitted source message and $\mathbf{x} = f_n(\mathbf{s}) \in \mathcal{X}^n$ is the corresponding n-length codeword. The conditional pdf of receiving $\mathbf{y} \in \mathcal{Y}^n$ at the channel output given that the message $\mathbf{s}$ is transmitted is given by

$$P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) = \prod_{i=1}^{n} W(y_i|x_i).$$

The probability of failing to decode the JSC code $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$ within a prescribed distortion level $\Delta > 0$ is called the probability of excess distortion and defined by

$$P_\Delta^{(n)}(P_S, W, \mathcal{E}, t) \triangleq \int_{\mathcal{S}^{tn}} P_{S^{tn}}(\mathbf{s}) \int_{\mathbf{y}: d^{(tn)}(\mathbf{s}, \varphi_n(\mathbf{y})) > \Delta} P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{y} d\mathbf{s}.$$

Definition 3 The JSCC excess distortion exponent $E_J(P_S, W, \Delta, \mathcal{E}, t)$ for the above MS P_S and MC W is defined as the supremum of the set of all numbers E for which there exists a sequence of source-channel codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$ with blocklength n such that

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

When there is no possibility of confusion, throughout the sequel the JSCC excess distortion exponent $E_J(P_S, W, \Delta, \mathcal{E}, t)$ will be written as E_J.

3 JSCC Excess Distortion Exponent for Gaussian Systems

We now focus on the communication system consisting of an MGS with alphabet $\mathcal{S} = \mathbb{R}$, mean zero, variance σ_S^2, and pdf $P_S(s) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left\{-\frac{s^2}{2\sigma_S^2}\right\}$, $s \in \mathcal{S}$, denoted by $P_S \sim \mathcal{N}(0, \sigma_S^2)$, and an MGC W with common input, output, and additive noise alphabets $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ and described by $Y_i = X_i + Z_i$, where Y_i, X_i and Z_i are the channel's output, input and noise symbols at time i. We assume that X_i and Z_i are independent from each other. The noise admits a zero-mean σ_Z^2-variance Gaussian pdf, denoted by $P_Z \sim \mathcal{N}(0, \sigma_Z^2)$ and thus the transition pdf of the channel is given by

$$W(y|x) = P_Z(z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left\{-\frac{z^2}{2\sigma_Z^2}\right\}, \qquad z = y - x \in \mathcal{Z}.$$

Let the distortion measure be squared-error distortion $d(s, s') = (s - s')^2$ and let the an input cost function be a power cost constraint $g(x) = x^2$.

Given a distortion threshold $\Delta > 0$, the rate-distortion function for MGS P_S is given by (e.g., [13])

$$R(P_S, \Delta) = \inf_{P_{S'|S}: \mathbb{E}d(S,S') \leq \Delta} I(S; S') = \begin{cases} \frac{1}{2} \ln \frac{\sigma_S^2}{\Delta} & 0 < \Delta < \sigma_S^2, \\ 0 & \sigma_S^2 \leq \Delta. \end{cases} \tag{9}$$

For the MGS P_S with a squared-error distortion measure, the explicit analytical form of $F(R, P_S, \Delta)$ is given by [16]

$$F(R, P_S, \Delta) = \begin{cases} \frac{1}{2} \left(\frac{\Delta \beta}{\sigma_S^2} - \ln \frac{\Delta \beta}{\sigma_S^2} - 1 \right) & \text{if} \quad R > R(P_S, \Delta), \\ 0 & \text{otherwise}, \end{cases} \tag{10}$$

where $\beta = e^{2R}$. Since $F(R, P_S, \Delta)$ is not meaningful at $R = 0$, we let

$$F(0, P_S, \Delta) \triangleq \lim_{R \downarrow 0} F(R, P_S, \Delta) = \begin{cases} \frac{1}{2} \left(\frac{\Delta}{\sigma_S^2} - \ln \frac{\Delta}{\sigma_S^2} - 1 \right) & \text{if} \quad R(P_S, \Delta) = 0, \\ 0 & \text{if} \quad R(P_S, \Delta) > 0, \end{cases}$$

Consequently, $F(R, P_S, \Delta)$ is convex strictly increasing in $R \geq 0$.

Given a power constraint $\mathcal{E} > 0$, the channel capacity of MGC W is given by

$$C(W, \mathcal{E}) = \sup_{P_X: \mathbb{E}X^2 \leq \mathcal{E}} I(X; Y) = \frac{1}{2} \ln (1 + \text{SNR}), \tag{11}$$

where $\text{SNR} \triangleq \mathcal{E}/\sigma_Z^2$ is the signal-to-noise ratio.

As mentioned before, the error exponent for the MGC $E(R, W, \mathcal{E})$ is only partially known. In the last fifty years, the error exponent for the MGC was actively studied and several lower and upper bounds were established (see, e.g., [3, 13, 22]). The most familiar upper bound is obtained by Shannon [22], called the sphere-packing upper bound and given by

$$\begin{aligned} E_{sp}(R, W, \mathcal{E}) \quad \triangleq \quad & \frac{\text{SNR}}{4\beta} \left[(\beta + 1) - (\beta - 1) \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right] \\ & + \frac{1}{2} \ln \left\{ \beta - \frac{\text{SNR}(\beta - 1)}{2} \left[\sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} - 1 \right] \right\}, \end{aligned} \tag{12}$$

where $\beta = e^{2R}$, $R \leq C(W, \mathcal{E})$. It can be shown (see Appendix A for a direct proof) that $E_{sp}(R, W, \mathcal{E})$ is convex strictly decreasing in $R \leq C(W, \mathcal{E})$ and vanishes for $R \geq C(W, \mathcal{E})$. It can also be easily verified that $E_{sp}(R, W, \mathcal{E}) \to \infty$ as $R \downarrow 0$. For the lower bound, we specialize Gallager's random-coding lower bound for the MGC W as follows: choosing the channel input distribution $P_X(x)$ as the Gaussian distribution $P_X^*(x) \sim \mathcal{N}(0, \mathcal{E})$, and replacing $g(x)$ by our square cost function x^2 yield the following lower bound for $E_0(W, \mathcal{E}, \rho)$

$$\begin{aligned} & E_0(W, \mathcal{E}, \rho) \geq \widetilde{E}_o(W, \mathcal{E}, \rho) \triangleq \max_{r \geq 0} E_0(W, \mathcal{E}, \rho, r, P_X^*) \\ = \quad & \max_{0 \leq r \leq 1/2\mathcal{E}} \left\{ r(1 + \rho)\mathcal{E} + \frac{1}{2} \ln(1 - 2r\mathcal{E}) + \frac{\rho}{2} \ln \left[1 - 2r\mathcal{E} + \frac{\mathcal{E}}{(1 + \rho)\sigma_Z^2} \right] \right\}. \end{aligned} \tag{13}$$

We hereby call $\widetilde{E}_o(W, \mathcal{E}, \rho)$ Gallager's Gaussian-input channel function. Note also that

$$E_{sp}(R, W, \mathcal{E}) = \max_{\rho \geq 0} [-\rho R + \widetilde{E}_o(W, \mathcal{E}, \rho)],$$

and the inner function is concave in ρ. Thus, the random-coding lower bound $E_r(R, W, \mathcal{E})$ can be further lower bounded by [13, pp. 339–340]

$$
\begin{aligned}
E_{\dagger}(R, W, \mathcal{E}) &= \max_{0 \le \rho \le 1} [-\rho R + \widetilde{E}_o(W, \mathcal{E}, \rho)] \\
&= \begin{cases} E_{sp}(R, W, \mathcal{E}), & R_{cr}(W) \le R \le C(W, \mathcal{E}), \\ 1 - \gamma + \frac{\text{SNR}}{2} + \frac{1}{2} \ln \left(\gamma - \frac{\text{SNR}}{2}\right) + \frac{1}{2} \ln \gamma - R, & 0 \le R \le R_{cr}(W), \end{cases}
\end{aligned} \tag{14}
$$

where

$$
\gamma \triangleq \frac{1}{2} \left[1 + \frac{\text{SNR}}{2} + \sqrt{1 + \frac{\text{SNR}^2}{4}} \right],
$$

and

$$
R_{cr}(W) \triangleq \frac{1}{2} \ln \left[\frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right]
$$

is the critical rate of the MGC (obtained by solving for the R where the straight-line of slope -1 is tangent to $E_{\dagger}(R, W, \mathcal{E})$). It is easy to show that $E_{\dagger}(R, W, \mathcal{E})$ is convex strictly decreasing in $0 < R \le C(W, \mathcal{E})$ with a straight-line section of slope -1 for $R \le R_{cr}(W)$. It has to be pointed out [13] that $E_{\dagger}(R, W, \mathcal{E})$ is not the real random-coding bound (as given in (6)) for $R < R_{cr}(W)$, but it admits a computable parametric form and it coincides with the upper bound $E_{sp}(R, W, \mathcal{E})$ for $R \ge R_{cr}(W)$. Thus, the channel coding error exponent $E(R, W, \mathcal{E})$ is determined for high rates ($R \ge R_{cr}(W)$).[1]

In the following we establish an upper and a lower bound for the JSCC excess distortion exponent for the Gaussian system in Sections 3.1 and 3.2. As will be seen in Section 3.3, the upper bound coincides with the lower bound for a large class of MGS-MGC pairs, and hence determines the exponent exactly.

3.1 The Upper Bound for E_J

Theorem 1 For an MGS P_S and an MGC W such that $tR(P_S, \Delta) < C(W, \mathcal{E})$, the JSCC excess distortion exponent satisfies

$$
E_J(P_S, W, \Delta, \mathcal{E}, t) \le \overline{E}_J(P_S, W, \Delta, \mathcal{E}, t), \tag{15}
$$

where

$$
\overline{E}_J(P_S, W, \Delta, \mathcal{E}, t) \triangleq \min_{tR(P_S, \Delta) \le R \le C(W, \mathcal{E})} \left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{sp}(R, W, \mathcal{E}) \right], \tag{16}
$$

where $F(R, P_S, \Delta)$ is the MGS excess distortion exponent given in (10) and $E_{sp}(R, W, \mathcal{E})$ is the sphere-packing bound of the MGC channel error exponent given in (12).

Proof: See Appendix B.

Since the MGS excess distortion exponent $tF(R/t, P_S, \Delta)$ is convex increasing for $R \ge tR(P_S, \Delta)$ and the sphere-packing bound $E_{sp}(R, W, \mathcal{E})$ is convex decreasing in $R \le C(W, \mathcal{E})$, their sum is also convex and

[1] In the recent work of [4], the lower bound $E_{\dagger}(R, W, \mathcal{E})$ is improved and is shown to be tight in a interval slightly below the critical rate, i.e., it is shown that the error exponent of the MGC is determined by $E_{\dagger}(R, W, \mathcal{E})$ for rates $R \ge R_1$ and R_1 can be less than $R_{cr}(W)$.

there exists a global minimum in the interval $[tR(P_S, \Delta), C(W, \mathcal{E})]$ for the upper bound given in (15). For $R \in [tR(P_S, \Delta), C(W, \mathcal{E})]$, setting

$$t \frac{\partial F\left(\frac{R}{t}, P_S, \Delta\right)}{\partial R} + \frac{\partial E_{sp}(R, W, \mathcal{E})}{\partial R} = 0,$$

gives (cf. Appendix A)

$$\frac{\beta^{\frac{1}{t}}}{\text{SDR}} = \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}}\right), \tag{17}$$

where $\text{SDR} \triangleq \sigma_S^2/\Delta$ is called the source-to-distortion ratio (i.e., the source variance to distortion threshold ratio), and $\beta = e^{2R}$. Thus, the minimum of the upper bound is achieved by the R which is the (unique) root of (17).

3.2 The Lower Bound for E_J

Given $\rho \geq 0$, for the continuous MS P_S, define source function

$$E(P_S, \Delta, \rho) \triangleq \sup_{Q_S} [\rho R(Q_S, \Delta) - D(Q_S \parallel P_S)], \tag{18}$$

where the supremum is taken over all the probability distributions Q_S defined on $\mathcal{S}$ such that $R(Q_S, \Delta)$ and $D(Q_S \parallel P_S)$ are well-defined and finite. We remark that (18) is equal to the guessing exponent for MGS's [1] under the squared-error distortion measure and admits an explicit form

$$E(P_S, \Delta, \rho) = \max\left\{0, \frac{1}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1 + \rho)\ln(1 + \rho) - \rho\right]\right\}. \tag{19}$$

Theorem 2 For an MGS P_S and a continuous MC W with a cost constraint $\mathcal{E}$ at the channel input, the JSCC excess distortion exponent satisfies

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \geq E_{RC}(P_S, W, \Delta, \mathcal{E}, t), \tag{20}$$

where

$$E_{RC}(P_S, W, \Delta, \mathcal{E}, t) \triangleq \max_{0 \leq \rho \leq 1} [E_o(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)], \tag{21}$$

where $E_o(W, \mathcal{E}, \rho)$ is Gallager's constrained channel function given by (7) and $E(P_S, \Delta, \rho)$ is the source function for the MGS P_S given by (19). Furthermore, if W is an MGC, we have

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \geq \underline{E}_J(P_S, W, \Delta, \mathcal{E}, t), \tag{22}$$

where

$$\underline{E}_J(P_S, W, \Delta, \mathcal{E}, t) \triangleq \max_{0 \leq \rho \leq 1} [\widetilde{E}_o(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)], \tag{23}$$

where $\widetilde{E}_o(W, \mathcal{E}, \rho)$ is Gallager's Gaussian-input channel function given by (13).

Proof: See Appendix C.

3.3 Tightness of the Lower and Upper Bounds: When Does $\overline{\boldsymbol{E}}_{\boldsymbol{J}} = \underline{\boldsymbol{E}}_{\boldsymbol{J}}$?

In order to evaluate the upper and lower bounds given in Theorems 1 and 2, we need to briefly review some concepts about Fenchel transforms. 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 $\{y : f^*(y) < \infty\}$.[2] 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 [21, Sec. 3], [11, Sec. 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.

Lemma 1 $E(P_S, \Delta, \rho)$ and $F(R, P_S, \Delta)$ are a pair of convex Fenchel transforms $\rho \geq 0$ and $R \geq 0$, i.e.,

$$E(P_S, \Delta, \rho) = F(R, P_S, \Delta)^* \qquad \text{for all } \rho \geq 0$$

and

$$F(R, P_S, \Delta) = E(P_S, \Delta, \rho)^* \qquad \text{for all } R \geq 0.$$

Proof: See Appendix E.

Lemma 2 $-E_{sp}(R, W, \mathcal{E})$ and $\widetilde{E}_o(W, \mathcal{E}, \rho)$ are a pair of concave Fenchel transforms for $\rho \geq 0$ and $R > 0$, i.e.,

$$-E_{sp}(R, W, \mathcal{E}) = \widetilde{E}_0(W, \mathcal{E}, \rho)_* \qquad \text{for all } R > 0$$

and

$$\widetilde{E}_0(W, \mathcal{E}, \rho) = (-E_{sp}(R, W, \mathcal{E}))_* \qquad \text{for all } \rho \geq 0.$$

Proof: See Appendix F.

Lemma 3 $-E_\dagger(R, W, \mathcal{E})$ and $\widetilde{E}_o(W, \mathcal{E}, \rho)$ are a pair of concave Fenchel transforms for $0 \leq \rho \leq 1$ and $R \geq 0$, i.e.,

$$-E_\dagger(R, W, \mathcal{E}) = \widetilde{E}_0(W, \mathcal{E}, \rho)_* \qquad \text{for all } R \geq 0$$

and

$$\widetilde{E}_0(W, \mathcal{E}, \rho) = (-E_\dagger(R, W, \mathcal{E}))_* \qquad \text{for all } 0 \leq \rho \leq 1.$$

[2] With a slight abuse of notation, both $f^*(y)$ and $f(y)^*$ refer to the Fenchel transform except when indicated otherwise.

Proof: See Appendix F.

Now 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 $F \cap G$ has interior points. Suppose further that $\mu = \inf_{x \in F \cap G}[f(x) - g(x)]$ is finite. Then *Fenchel's Duality Theorem* [17] asserts that

$$\mu = \inf_{x \in F \cap G}[f(x) - g(x)] = \max_{y \in F^* \cap G_*}[g_*(y) - f^*(y)]. \tag{24}$$

Applying Fenchel's Duality (24) to our source and channel functions $E(P_S, \Delta, \rho)$ and $\widetilde{E}_0(W, \mathcal{E}, \rho)$ with respect to their Fenchel transforms in Lemmas 1, 2 and 3, we obtain the following equivalent bounds.

Theorem 3 Let $tR(P_S, \Delta) < C(W, \mathcal{E})$. Then

$$\min_{tR(P_S,\Delta) \le R \le C(W,\mathcal{E})} \left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{sp}(R, W, \mathcal{E}) \right] = \max_{0 \le \rho < \infty} [\widetilde{E}_0(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)], \tag{25}$$

$$\min_{tR(P_S,\Delta) \le R \le C(W,\mathcal{E})} \left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{\dagger}(R, W, \mathcal{E}) \right] = \max_{0 \le \rho \le 1} [\widetilde{E}_0(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)]. \tag{26}$$

The proof of the theorem follows from the above argument regarding Fenchel transforms and Fenchel's Duality Theorem (24); for more details, readers may consult [26]. We next provide a necessary and sufficient condition under which $\overline{E}_J = \underline{E}_J$ for the MGS-MGC pair.

Theorem 4 Let $tR(P_S, \Delta) < C(W, \mathcal{E})$. The upper and lower bounds for $E_J(P_S, W, \Delta, \mathcal{E}, t)$ given in Theorem 1 and (22) of Theorem 2 are equal if and only if

$$2(2\text{SDR})^t - \frac{2(2\text{SDR})^t}{2(2\text{SDR})^t - 1} \ge \text{SNR}. \tag{27}$$

Remark 1 For $tR(P_S, \Delta) \ge C(W, \mathcal{E})$, $E_J(P_S, W, \Delta, \mathcal{E}, t) = 0$.

Proof: See Appendix G.

Example 1 In Fig. 1, we partition the SDR-SNR plane into three parts for transmission rate $t = 0.5$, 1, 1.5 and 2: in region **A** (including the boundary between **A** and **B**) $tR(P_S, \Delta) \ge C(W, \mathcal{E})$ and $E_J = 0$; in region **B** (including the boundary between **B** and **C**), $\overline{E}_J = \underline{E}_J$ and hence E_J is determined exactly; and in region **C**, $E_J > 0$ is bounded by $\overline{E}_J$ and $\underline{E}_J$. Fig. 2 shows the two bounds $\overline{E}_J$ and $\underline{E}_J$ for different SDR-SNR pairs and transmission rate $t = 1$. We observe from the two figures that the two bounds coincide for a large class of SDR-SNR pairs.

3.4 JSCC vs Tandem Coding Exponents for Gaussian Systems

We herein study the advantage of JSCC over tandem coding in terms of the excess distortion exponent for Gaussian systems. A tandem code $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t) \triangleq (f_{cn} \circ \pi_m \circ f_{sn}, \varphi_{sn} \circ \pi_m^{-1} \circ \varphi_{cn}, \Delta, \mathcal{E}, t, P)$ with blocklength n and transmission rate t (source symbols/channel use) for the MGS and the MGC W is composed (see Fig. 3) of two "separately" designed codes: a (tn, M_n) block source code $(f_{sn}, \varphi_{sn}, \Delta)$ with codebook $\mathcal{C} \triangleq \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{M_n}\} \subseteq \mathcal{S}^{tn}$ and source code rate $R_{s,n} = \ln M_n/tn$ source code nats/source symbol, and an (n, M_n) block channel code $(f_{cn}, \varphi_{cn}, \mathcal{E})$ with channel code rate $R_{c,n} = \ln M_n/n$ source code nats/channel use, where $f_{cn} \in \mathcal{F}_{cn}^{\mathcal{E}}$ with $g(x) = x^2$, assuming that the limit $\lim_{n\to\infty} \frac{\ln M_n}{n}$ exists, i.e., $\limsup_{n\to\infty} \frac{\ln M_n}{n} = \liminf_{n\to\infty} \frac{\ln M_n}{n}$. Here "separately" means that the source code is designed without the knowledge of the channel statistics, and the channel code is designed without the knowledge of the source statistics. However, as long as the source encoder is directly concatenated by a channel encoder, the source statistics would be automatically brought into the channel coding stage. Thus common randomization is needed to decouple source and channel coding (e.g., [14]). We assume that the source coding index $i = f_{sn}(\mathbf{s})$ is mapped to a channel index through a permutation mapping $\pi_m : \{1, 2, ..., M_n\} \to \{1, 2, ..., M_n\}$ (the index assignment π_m is assumed to be known at both the transmitter and the receiver). Furthermore, the choice of π_m is assumed random and equally likely from all the $M_n!$ different possible index assignments, so that the indices fed into the channel encoder have a uniform distribution. Hence common randomization achieves statistical separation between the source and channel coding operations.

The (overall) excess distortion probability of the tandem code $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t)$ is given by

$$P_{\Delta^*}^{(n)}(P_S, W, \mathcal{E}, t) \triangleq \Pr\left(d^{(tn)}\left(\mathbf{s}, \varphi_{sn}\left\{\pi_m^{-1}[\varphi_{cn}(\mathbf{y})]\right\}\right) > \Delta\right)$$

In order to facilitate the evaluation of the tandem excess distortion probability $P_{\Delta^*}^{(n)}(P_S, W, \mathcal{E}, t)$, we simplify the problem by making some (natural) assumptions on the component channel and source codes (which are statistically decoupled from each other via common randomization).

1. We assume that the channel codes $(f_{cn}, \varphi_{cn}, \mathcal{E})$ in the tandem system are "good channel codes (in the weak sense)," i.e., $(f_{cn}, \varphi_{cn}, \mathcal{E}) \in \Xi(W, \mathcal{E})$, where

$$\Xi(W, \mathcal{E}) \triangleq \left\{(f_{cn}, \varphi_{cn}, \mathcal{E}) : \limsup_{n\to\infty} P_{ec}^{(n)}(W, R_{c,n}, \mathcal{E}) < \gamma \quad \text{for all } \gamma > 0\right\}$$

   and $P_{ec}^{(n)}(W, R_{c,n}, \mathcal{E})$ is the channel coding probability of error given by (5).

2. We assume that the source codes $(f_{sn}, \varphi_{sn}, \Delta)$ in the tandem system are "good source codes (in the strong sense)," i.e., $(f_{sn}, \varphi_{sn}, \Delta) \in \Omega(P_S, \Delta)$, where

$$\Omega(P_S, \Delta) \triangleq \left\{(f_{sn}, \varphi_{sn}, \Delta) : \liminf_{n\to\infty} -\frac{1}{tn} \ln P_{\Delta}^{(n)}(P_S, R_{s,n}) \geq F(R, P_S, \Delta) > 0, \text{ where } R = \lim_{n\to\infty} R_{s,n}\right\},$$

   and $P_{\Delta}^{(n)}(P_S, R_{s,n})$ is the source coding excess distortion probability given by (2).

The converse JSCC theorem (Theorem 8) states that the MGS cannot be reliably transmitted over the MGC if $tR(P_S, \Delta) > C(W, \mathcal{E})$, and also note that if $tR(P_S, \Delta) > C(W, \mathcal{E})$ then either $\Xi(W, \mathcal{E}) = \phi$ or $\Omega(P_S, \Delta) = \phi$. Thus, we are only interested in the case $tR(P_S, \Delta) < C(W, \mathcal{E})$ as before. In order to guarantee the existence of good source and channel codes, we focus on the sequences of tandem codes with $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t) \in \Lambda(P_S, W, \Delta, \mathcal{E}, t)$, where

$$\Lambda(P_S, W, \Delta, \mathcal{E}, t) \triangleq \left\{(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t) : tR(P_S, \Delta) < \lim_{n\to\infty} \frac{\ln M_n}{n} < C(W, \mathcal{E})\right\}.$$

Definition 4 The tandem coding excess distortion exponent $E_T(P_S, W, \Delta, \mathcal{E}, t)$ for the MGS P_S and the MGC W is defined as the supremum of the set of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t)$ composed by good source and channel codes with blocklength n provided $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, t) \in \Lambda(P_S, W, \Delta, \mathcal{E}, t)$, such that

$$\widehat{E} \leq \liminf_{n\to\infty} -\frac{1}{n} \ln P_{\Delta^*}^{(n)}(P_S, W, \mathcal{E}, t).$$

When there is no possibility of confusion, throughout the sequel, the tandem coding excess distortion exponent $E_T(P_S, W, \Delta, \mathcal{E}, t)$ will be written as E_T. It can be easily shown by definition that $E_J \geq E_T$; however, we are particularly interested in investigating the situation where a strict inequality holds. 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 excess distortion probability, which may translate into substantial reductions in complexity and delay for real-world communication systems.

We obtain the following formula for the tandem excess distortion exponent for SDR $\geq 4 (\approx 6\text{dB})$. Note that this condition is not too restrictive, since a large distortion threshold is useless in practice.

Theorem 5 For the tandem MGS-MGC system provided $tR(P_S, \Delta) < C(W, \mathcal{E})$ and SDR ≥ 4,

$$E_T(P_S, W, \Delta, \mathcal{E}, t) = \sup_{tR(P_S,\Delta)<R<C(W,\mathcal{E})} \min\left\{ tF\left(\frac{R}{t}, P_S, \Delta\right), E(R, W, \mathcal{E}) \right\}$$

where $F(R, P_S, \Delta)$ is the MGS excess distortion exponent given by (10) and $E(R, W, \mathcal{E})$ is the MGC error exponent.

Proof: See the proof of [29, Theorem 8].

Remark 2 Since $tF(R/t, P_S, \Delta)$ is a strictly increasing function of R for $R \geq R(P_S, \Delta) > 0$, and $E(R, W, \mathcal{E})$ is decreasing function of R for $0 < R \leq C(W, \mathcal{E})$, the supremum must be achieved at their intersection[3]

$$E_T(P_S, W, \Delta, \mathcal{E}, t) = tF\left(\frac{R_o}{t}, P_S, \Delta\right) = E(R_o, W, \mathcal{E}),$$

with $tR(P_S, \Delta) < R_o < C(W, \mathcal{E})$.

We next numerically compare the lower bound of joint exponent $\underline{E}_J$ and the upper bound of tandem exponent $\overline{E}_T$ given by

$$\overline{E}_T(P_S, W, \Delta, \mathcal{E}, t) \triangleq \sup_{tR(P_S,\Delta)<R<C(W,\mathcal{E})} \min\left\{ tF\left(\frac{R}{t}, P_S, \Delta\right), E_{sp}(R, W, \mathcal{E}) \right\}.$$

Example 2 For transmission rate $t = 1$, we plot the SNR-SDR region for which $E_J > E_T$ in Fig. 4 obtained from the inequality $\underline{E}_J > \overline{E}_T$. It is seen that $E_J > E_T$ for many SNR-SDR pairs. For example, when SDR = 7 dB, $E_J > E_T$ holds for 10 dB $\leq$ SNR $\leq$ 24 dB (approximately). We also compute the two

[3]Unlike the discrete case in [26], the intersection always exists since source exponent is continuous and increasing in $R > 0$.

bounds of E_J and E_T, and we see from Fig. 5 that when SDR = 8 dB, E_J (or its lower bound) almost double E_T (or its upper bound) for 8dB $\leq$ SNR $\leq$ 15dB. It is also observed that for the same exponent (e.g. $0.2 \sim 1.1$), the gain of JSCC over tandem coding could be as large as 2dB in SNR. Similar results are observed for other parameters, see Figs. 6 and 7 for $t = 1.5$. We conclude that JSCC considerably outperforms tandem coding in terms of excess distortion exponent for a large class of MGS-MGC pairs.

4 Extensions

In this section, we provide extensions of the upper and/or lower bounds for the JSCC excess distortion exponent for other memoryless continuous source-channel pairs.

4.1 Laplacian Sources with the Magnitude-Error Distortion over MGC's

In image coding applications, the Laplacian distribution is well known to provide a good model to approximate the statistics of transform coefficients such as discrete cosine and wavelet transform coefficients [20, 24]. Thus, it is of interest to study the theoretical performance for the lossy transmission of MLS's, say, over an MGC. Due to the striking similarity between the Laplacian source and the Gaussian source, the results of the previous section (especially regarding the bounds for $E_J(P_S, W, \Delta, \mathcal{E}, t)$) can be easily extended to a system composed by an MLS under the magnitude-error distortion measure and an MGC.

Consider an MLS P_S with alphabet $\mathcal{S} = \mathbb{R}$, mean zero, $\mathbb{E}|s| = \alpha$, and pdf

$$P_S(s) = \frac{1}{2\alpha} \exp\left\{-\frac{|s|}{\alpha}\right\}, \quad s \in \mathcal{S},$$

denoted by $P_S \backsim L(0, \alpha)$. We assume that the distortion measure is the magnitude-error distortion given by $d(s, s') \triangleq |s - s'|$ for any $s, s' \in \mathbb{R}$. For the MLS $P_S \sim L(0, \alpha)$ and distortion threshold Δ, the source excess distortion exponent is given by [28]

$$F(R, P_S, \Delta) = \begin{cases} \frac{e^R \Delta}{\alpha} - \ln \frac{e^R \Delta}{\alpha} - 1 & \text{if} \quad R > R(P_S, \Delta) = \max\{0, \ln \frac{\alpha}{\Delta}\}, \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

The upper and lower bounds for E_J can be derived in an analogous method to the one used for the Gaussian systems.

Theorem 6 For the MLS P_S and the MGC W with transmission rate t,

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \leq \min_R \left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{sp}(R, W, \mathcal{E}) \right]$$

and

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \geq \min_R \left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{\dagger}(R, W, \mathcal{E}) \right],$$

where $E_{sp}(R, W, \mathcal{E})$ and $E_{\dagger}(R, W, \mathcal{E})$ are given by (12) and (14) respectively.

Proof: See Appendix H.

4.2 Memoryless Systems with a Metric Source Distortion

In this section we consider the transmission of a class of continuous MS's with alphabet $\mathcal{S} = \mathbb{R}$ over continuous MC's when the source distortion function is a metric; i.e., for $s, s' \in \mathcal{S}$ (1) $d(s, s') \geq 0$ with equality if and only if $s = s'$; (2) $d(s, s') = d(s', s)$; (3) the triangle inequality holds, i.e., for any $s_1, s_2, s_3 \in \mathcal{S}$, $d(s_1, s_2) + d(s_2, s_3) \geq d(s_1, s_3)$. We still assume that for any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}^k$,

$$d^{(k)}(\mathbf{s}, \mathbf{s}') \triangleq \frac{1}{k} \sum_{i=1}^{k} d(s_i, s_i').$$

Theorem 7 For the continuous MS P_S with a distortion being a metric and the continuous MC W with a cost constraint $\mathcal{E}$ at the channel input, if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E} \exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, the JSCC excess distortion exponent satisfies

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \geq \max_{0 \leq \rho < 1} [E_0(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)], \tag{29}$$

where $E_o(W, \mathcal{E}, \rho)$ is Gallager's constrained channel function given by (7) and $E(P_S, \Delta, \rho)$ is the source function for P_S given by (18). Furthermore, if W is an MGC, we have

$$E_J(P_S, W, \Delta, \mathcal{E}, t) \geq \max_{0 \leq \rho < 1} [\widetilde{E}_0(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)], \tag{30}$$

where $\widetilde{E}_0(W, \mathcal{E}, \rho)$ is Gallager's Gaussian-input channel function given by (13).

Proof: See Appendix I.

Although Theorem 7 does not apply to MGS's under the squared-error distortion (which is not a metric) and MLS's under the magnitude-error distortion (which does not satisfy the finiteness condition), it applies to MGS's under the magnitude-error distortion, and more generally, it applies to generalized MGS's with parameters (α, σ) under the distortion function $d(s, s') \triangleq |s - s'|^p$ for any $s, s' \in \mathbb{R}$, whenever $0 < p \leq 1$ and $p < \alpha$; see the following example.

Example 3 The Gaussian and Laplacian distributions belong to the class of generalized Gaussian distributions, which are widely used in image coding applications. It is well known that the distribution of image subband coefficients is well approximated by the generalized Gaussian distribution [6, 24]. A generalized MGS P_S with parameters (α, σ) has alphabet $\mathcal{S} = \mathbb{R}$, mean zero, variance σ^2, and pdf

$$P_S(s) = \frac{\alpha \eta(\alpha, \sigma)}{2\Gamma(1/\alpha)} \exp\left\{-(\eta(\alpha, \sigma)|s|)^\alpha\right\}, \quad s \in \mathcal{S},$$

where $\Gamma(\cdot)$ is the Gamma function and

$$\eta(\alpha, \sigma) \triangleq \frac{1}{\sigma} \left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right)^{\frac{1}{2}} \quad \alpha > 0.$$

Note that the pdf reduces to the Gaussian and Laplacian pdf's for $\alpha = 2$ and 1, respectively. When $0 < p \leq 1$, the distortion $d(s, s') \triangleq |s - s'|^p$ is a metric. If we choose $s_o = 0$, then $\mathbb{E}\exp[td(s, s_o)]$ would have the form

$$\mathbb{E}\exp[td(s, s_o)] = \int_{-\infty}^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)} ds = 2\int_{0}^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)} ds$$

where $A > 0$, $B > 0$, and C are independent of s. Clearly, the above integral is finite for any $Ct \geq 0$. If $Ct < 0$, and $\alpha > p$ is provided, the integral can be bounded by

$$\int_{0}^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)} ds \leq \int_{0}^{x} Ae^{-BCt|s|^p} ds + \int_{x}^{+\infty} Ae^{-B|s|^\alpha} ds$$

which is also finite, where $x > 0$ satisfies $x^{\alpha-p} + Ct = 0$.

5 Conclusion

In this work, we investigate the JSCC excess distortion exponent E_J for some memoryless communication systems with continuous alphabets. For the Gaussian system with the squared-error source distortion measure and a power channel input constraint, we derive upper and lower bounds for the excess distortion exponent. The bounds extend our earlier work for discrete systems [26] in such a way that the lower/upper bound can be expressed by Csiszár's form [8] in terms of the sum of source and channel exponents. They can also be expressed in equivalent parametric forms as differences of source and channel functions. We then extend these bounds to Laplacian-Gaussian source-channel pairs with the magnitude-error distortion. By employing a different technique, we also derive a lower bound (of similar parametric form) for E_J for a class of memoryless source-channel pairs under a metric distortion measure and some finiteness condition.

For the Gaussian system, a sufficient and necessary condition for which the two bounds of E_J coincide is provided. It is observed that the two bounds are tight in many cases, thus exactly determining E_J. We also derive an expression for the tandem coding exponent for Gaussian source-channel pairs provided that SDR ≥ 4 ($\approx$ 6dB). The tandem Gaussian exponent has a similar form as the discrete tandem error exponent. As in the discrete cases, the JSCC exponent is observed to be considerable larger than the tandem exponent for a large class of Gaussian source-channel pairs.

A The Properties of $E_{sp}(R, W, \mathcal{E})$

Proof of Monotonicity: Since $E_{sp}(R, W, \mathcal{E})$ is a differentiable function for $R > 0$, we have

$$\begin{aligned}
\frac{\partial E_{sp}(R, W, \mathcal{E})}{\partial R} &= \frac{\beta\left[-\mathrm{SNR}\beta^2 - 4\mathrm{SNR}\beta + \mathrm{SNR}^2 + (\mathrm{SNR}+2)\Psi\right]}{\Psi\left[2\beta + \mathrm{SNR}\beta - \mathrm{SNR} - \Psi\right]} \\
&= \frac{\left[-\mathrm{SNR}^2\beta - 4\mathrm{SNR}\beta + \mathrm{SNR}^2 + \Psi(\mathrm{SNR}+2)\right](2\beta + \mathrm{SNR}\beta - \mathrm{SNR} + \Psi)}{4\beta\Psi} \\
&= \frac{2\mathrm{SNR}^2 - 2\mathrm{SNR}^2\beta - 8\mathrm{SNR}\beta + (4\beta - 2\mathrm{SNR})\Psi}{4\beta\Psi} \\
&= 1 - \frac{\mathrm{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\mathrm{SNR}(\beta-1)}}\right), \qquad (31)
\end{aligned}$$

where $\beta = e^{2R}$ and

$$\Psi = \sqrt{(\text{SNR}\beta - \text{SNR} + 4\beta)\text{SNR}(\beta - 1)}.$$

Now solving

$$1 - \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}\right) \le 0$$

yields

$$R \le \frac{1}{2}\ln(1 + \text{SNR}) = C(W, \mathcal{E}).$$

Particularly, we have

$$\lim_{R \to C(W,\mathcal{E})} \frac{\partial E_{sp}(R, W, \mathcal{E})}{\partial R} = 0 \quad \text{and} \quad \lim_{R \downarrow 0} \frac{\partial E_{sp}(R, W, \mathcal{E})}{\partial R} = -\infty.$$

Hence, $E_{sp}(R, W, \mathcal{E})$ is a strictly decreasing function in $R \in (0, C(W, \mathcal{E})]$ with a slope ranging from $-\infty$ to 0.

Proof of Convexity: It follows from (31) that for $R \in (0, C(W, \mathcal{E})]$,

$$\frac{\partial^2 E_{sp}(R, W, \mathcal{E})}{\partial R^2} = \frac{\text{SNR}}{\beta}\left[1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}\right] + \frac{2}{\text{SNR}^2(\beta-1)^2\sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}} > 0. \tag{32}$$

This demonstrates the (strict) convexity of $E_{sp}(R, W, \mathcal{E})$. ■

B Proof of Theorem 1

We first derive a strong converse JSCC theorem under the probability of excess distortion criterion for the Gaussian system. We use later this result to obtain an upper bound for the excess distortion exponent E_J.

Theorem 8 (Strong Converse JSCC Theorem) For an MGS P_S and an MGC W, if $tR(P_S, \Delta) > C(W, \mathcal{E})$, then $\lim_{n\to\infty} P_\Delta^{(n)}(P_S, W, \mathcal{E}, t) = 1$ for any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$.

Proof: Assume that $C(W, \mathcal{E}) = tR(P_S, \Delta) - \varepsilon$, where ε is a positive number. For some δ $(0 < \delta < \varepsilon)$, define

$$\widetilde{A} = \left\{ (\mathbf{s}, \mathbf{y}) : \ln \frac{P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))P^*_{S'^{tn}}(\varphi_n(\mathbf{y}))}{P^*_{Y^n}(\mathbf{y})P^*_{S'^{tn}|S^{tn}}((\varphi_n(\mathbf{y}))|\mathbf{s})} \le n\left(C(W, \mathcal{E}) - tR(P_S, \Delta) + \delta\right) \right\},$$

where $P^*_{S'^{tn}|S^{tn}}$ and $P^*_{S'^{tn}}$ are the $tn-$dimensional product distributions corresponding to

$$P^*_{S'|S}(s'|s) = \frac{1}{\sqrt{2\pi\frac{\Delta(\sigma_S^2 - \Delta)}{\sigma_S^2}}} \exp\left\{ -\frac{\left(s' - \frac{\sigma_S^2 - \Delta}{\sigma_S^2}s\right)^2}{\frac{2\Delta(\sigma_S^2-\Delta)}{\sigma_S^2}} \right\}, \tag{33}$$

and

$$P^*_{S'}(s') = \int P_S(s)P^*_{S'|S}(s'|s)ds = \frac{1}{\sqrt{2\pi(\sigma_S^2 - \Delta)}} \exp\left\{ -\frac{s'^2}{2(\sigma_S^2 - \Delta)} \right\}, \tag{34}$$

respectively, and $P^*_{Y^n}$ is the $n-$dimensional product distribution corresponding to

$$P^*_Y(y) = \frac{1}{\sqrt{2\pi(\mathcal{E}+\sigma_Z^2)}} \exp\left\{-\frac{y^2}{2(\mathcal{E}+\sigma_Z^2)}\right\}. \tag{35}$$

Here, note that $P^*_{S'|S}$ is the pdf that achieves the infimum of (9) provided that $R(P_S,\Delta) > 0$. $P^*_{S'}$ is the marginal pdf of $P_S P^*_{S'|S}$. P^*_Y is the marginal pdf of $P^*_X P_{Y|X}$ where P^*_X achieves the channel capacity (11).

Recalling that

$$P_\Delta^{(n)}(P_S, W, \mathcal{E}, t) = 1 - \Pr\left(d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right), \tag{36}$$

where the probability is with respect to the joint distribution $P_{S^{tn}}(\cdot)P_{Y^n|X^n}(\cdot|\cdot)$, it suffices to show that the probability $\Pr\left(d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right)$ approaches 0 asymptotically for any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$. We first decompose $\Pr\left(d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right)$ as follows

$$\begin{aligned} & \Pr\left(d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right) \\ = \; & \Pr\left(\left\{d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right\} \bigcap \widetilde{A}\right) + \Pr\left(\left\{d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right\} \bigcap \widetilde{A}^c\right), \end{aligned} \tag{37}$$

where $\widetilde{A}^c$ stands for the complement of $\widetilde{A}$. For the first probability in (37), we can bound it by using the property of set $\widetilde{A}$

$$\begin{aligned} & \Pr\left(\left\{d^{(tn)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \le \Delta\right\} \bigcap \widetilde{A}\right) \\ = \; & \int_{\left\{(\mathbf{s},\mathbf{y}): d^{(tn)}(\mathbf{s},\varphi_n(\mathbf{y}))\le\Delta\right\}\bigcap \widetilde{A}} P_{S^{tn}}(\mathbf{s}) P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{s} d\mathbf{y} \\ \le \; & \int_{\left\{(\mathbf{s},\mathbf{y}): d^{(tn)}(\mathbf{s},\varphi_n(\mathbf{y}))\le\Delta\right\}\bigcap \widetilde{A}} e^{n(C(W,\mathcal{E})-tR(P_S,\Delta)+\delta)} P_{S^{tn}}(\mathbf{s}) \frac{P^*_{Y^n}(\mathbf{y}) P^*_{S'^{tn}|S^{tn}}((\varphi_n(\mathbf{y}))|\mathbf{s})}{P^*_{S'^{tn}}(\varphi_n(\mathbf{y}))} d\mathbf{s} d\mathbf{y} \\ \le \; & e^{-n(\varepsilon-\delta)} \int_{\mathcal{Y}^n} \frac{P^*_{Y^n}(\mathbf{y})}{P^*_{S'^{tn}}(\varphi_n(\mathbf{y}))} \underbrace{\int_{\mathbf{s}: d^{(tn)}(\mathbf{s},\varphi(\mathbf{y}))\le\Delta} P_{S^{tn}}(\mathbf{s}) P^*_{S'^{tn}|S^{tn}}(\varphi_n(\mathbf{y})|\mathbf{s}) d\mathbf{s}}_{\le P^*_{S'^{tn}}(\varphi_n(\mathbf{y}))} d\mathbf{y} \\ \le \; & e^{-n(\varepsilon-\delta)} \int_{\mathcal{Y}^n} P^*_{Y^n}(\mathbf{y}) d\mathbf{y} \\ = \; & e^{-n(\varepsilon-\delta)}. \end{aligned} \tag{38}$$

It remains to bound the second probability in (37). Using the expressions of the pdf's, we have

$$\begin{aligned} \frac{1}{n} \ln \frac{P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) P^*_{S'^{tn}}(\varphi_n(\mathbf{y}))}{P^*_{Y^n}(\mathbf{y}) P^*_{S'^{tn}|S^{tn}}((\varphi_n(\mathbf{y}))|\mathbf{s})} \;=\; & C(W,\mathcal{E}) + \frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)} - \frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2} \\ - \; & tR(P_S,\Delta) + \frac{td^{(tn)}(\varphi_n(\mathbf{y}),\mathbf{s})}{2\Delta} - \frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}. \end{aligned}$$

Hence,

$$
\begin{aligned}
& \Pr\left(\left\{d^{(tn)}\left(\mathbf{s},\varphi_n(\mathbf{y})\right)\le\Delta\right\}\bigcap\widetilde{A}^c\right) \\
= \; & \Pr\left(\left\{d^{(tn)}\left(\mathbf{s},\varphi_n(\mathbf{y})\right)\le\Delta\right\}\bigcap\left\{\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}+\frac{td^{(tn)}(\varphi_n(\mathbf{y}),\mathbf{s})}{2\Delta}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right\}\right) \\
\le \; & \Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}+\frac{t}{2}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right) \\
\le \; & \Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right)+\Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2}-1<-\frac{2\delta}{3}\right)+\Pr\left(\frac{\mathbf{s}^T\mathbf{s}}{n\sigma_S^2}-t<-\frac{2\delta}{3}\right). \qquad (39)
\end{aligned}
$$

It suffices to show

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right)=0, \qquad (40)$$

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2}-1<-\frac{2\delta}{3}\right)=0, \qquad (41)$$

and

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{s}^T\mathbf{s}}{n\sigma_S^2}-t<-\frac{2\delta}{3}\right)=0. \qquad (42)$$

Clearly, (41) and (42) follow by the weak law of large numbers (WLLN), noting that $\mathbf{s}$ and $\mathbf{z}$ are memoryless sequences. To derive (40), we write, as in the proof of [19, Lemma 4])

$$
\begin{aligned}
\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right) &= \Pr\left(\frac{\mathbf{x}^T\mathbf{x}}{n}+\frac{\mathbf{z}^T\mathbf{z}}{n}+\frac{2\mathbf{x}^T\mathbf{z}}{n}-(\mathcal{E}+\sigma_Z^2)>\frac{2\delta}{3}(\mathcal{E}+\sigma_Z^2)\right) \\
&\le \Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n}+\frac{2\mathbf{x}^T\mathbf{z}}{n}-\sigma_Z^2>\frac{2\delta}{3}(\mathcal{E}+\sigma_Z^2)\right) \\
&\le \Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n}-\sigma_Z^2>\frac{\delta}{3}(\mathcal{E}+\sigma_Z^2)\right)+\Pr\left(\frac{2\mathbf{x}^T\mathbf{z}}{n}>\frac{\delta}{3}(\mathcal{E}+\sigma_Z^2)\right), \qquad (43)
\end{aligned}
$$

where the first inequality follows from the power constraint (8), the first probability in (43) converges to zero as $n\to\infty$ by the WLLN and the second probability in (43) converges to zero as $n\to\infty$ by the WLLN, the fact the $\mathbf{z}$'s have zero mean, and the independence of $\mathbf{x}$ and $\mathbf{z}$. Thus, (40), (41) and (42) yield

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}+\frac{t}{2}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right)=0. \qquad (44)$$

On account of (38), (44) and (36), we complete the proof. ■

Note that the above theorem also holds for a slightly wider class of MGCs with scaled inputs, described by $Y_i = bX_i + Z_i$ (X_i and Z_i are independent from each other), and with transition pdf

$$W(y|x)=P_Z(y-bx)=\frac{1}{\sqrt{2\pi\sigma_Z^2}}e^{-\frac{(y-bx)^2}{2\sigma_Z^2}},$$

where b is a nonzero constant. We next apply this result to prove the upper bound of E_J. It follows from Theorem 8 that the JSCC excess distortion exponent is 0 if the source rate-distortion function is larger than the channel capacity, i.e., $tR(P_S, \Delta) > C(W, \mathcal{E})$. We thus confine our attention to the case of $tR(P_S, \Delta) < C(W, \mathcal{E})$ in the following proof.

Proof of Theorem 1: For any sufficiently small $\varepsilon > 0$, fix an $R \in [tR(P_S, \Delta) + \varepsilon, C(W, \mathcal{E})]$. Define an auxiliary MGS for this R with alphabet $\mathcal{S} = \mathbb{R}$ and distribution $\widetilde{P}_S \sim \mathcal{N}(0, \widetilde{\sigma}_S^2)$, where $\widetilde{\sigma}_S^2 \triangleq \Delta e^{2R/t}$, so that the rate-distortion function of $\widetilde{P}_S$ is given by

$$R(\widetilde{P}_S, \Delta) = \frac{1}{2} \ln \max \left\{ \frac{\widetilde{\sigma}_S^2}{\Delta}, 1 \right\} = \frac{R}{t}.$$

Also, it can be easily verified that the Kullback-Leibler divergence between the auxiliary MGS $\widetilde{P}_S$ and the original source P_S is

$$D(\widetilde{P}_S \parallel P_S) = \frac{1}{2} \left(\frac{\widetilde{\sigma}_S^2}{\sigma_S^2} - \ln \frac{\widetilde{\sigma}_S^2}{\sigma_S^2} - 1 \right) = F\left(\frac{R}{t}, P_S, \Delta \right).$$

Next we define for $R' \triangleq R - \frac{\varepsilon}{2} > 0$ an auxiliary MGC with scaled inputs $\widetilde{W}$ associated with the original MGC W with the alphabets $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and transition pdf

$$\widetilde{P}_{Y|X}(y|x) \triangleq \frac{1}{\sqrt{2\pi \widetilde{\sigma}_Z^2}} e^{-\frac{(y+ax)^2}{2\widetilde{\sigma}_Z^2}}$$

where the parameter a is uniquely determined by β' ($\beta' = e^{2R'}$) and SNR as follows

$$a \triangleq \frac{-\text{SNR}(\beta' - 1) - \sqrt{\text{SNR}^2(\beta' - 1)^2 + 4\text{SNR}\beta'(\beta' - 1)}}{2\text{SNR}\beta'} < 0, \tag{45}$$

and

$$\widetilde{\sigma}_Z^2 \triangleq \frac{a^2 \mathcal{E}}{\beta' - 1}. \tag{46}$$

It can be verified that the capacity of the MGC $\widetilde{W}$ is given by

$$C(\widetilde{W}, \mathcal{E}) = \sup_{P_X : \mathbb{E}X^2 \le \mathcal{E}} I(X;Y) = \frac{1}{2} \ln \left(1 + \frac{a^2 \mathcal{E}}{\widetilde{\sigma}_Z^2} \right) = R',$$

where the supremum is achieved by the Gaussian distribution $P_X^* \sim \mathcal{N}(0, \mathcal{E})$.

For some $\delta > 0$, define the set

$$\widehat{A} \triangleq \left\{ (\mathbf{s}, \mathbf{y}) : \ln \frac{\widetilde{P}_{S^{tn}}(\mathbf{s}) \widetilde{P}_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))}{P_{S^{tn}}(\mathbf{s}) P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))} \le n \left(tF\left(\frac{R}{t}, P_S, \Delta \right) + E_{sp}(R', W, \mathcal{E}) + \delta \right) \right\}.$$

Consequently, we can use $\widehat{A}$ to lower bound the probability of excess distortion of any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$,

$$\begin{aligned} P_\Delta^{(n)}(P_S, W, \mathcal{E}, t) \;\; &\ge \int_{\left\{ (\mathbf{s},\mathbf{y}) : d^{(tn)}(\mathbf{s}, \varphi_n(\mathbf{y})) > \Delta \right\} \cap \widehat{A}} P_{S^{tn}}(\mathbf{s}) P_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{s} d\mathbf{y} \\ &\ge e^{-n\left(tF\left(\frac{R}{t}, P_S, \Delta \right) + E_{sp}(R', W, \mathcal{E}) + \delta \right)} \\ &\quad \int_{\left\{ (\mathbf{s},\mathbf{y}) : d^{(tn)}(\mathbf{s}, \varphi_n(\mathbf{y})) > \Delta \right\} \cap \widehat{A}} \widetilde{P}_{S^{tn}}(\mathbf{s}) \widetilde{P}_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{s} d\mathbf{y}, \end{aligned} \tag{47}$$

and the last integration can be decomposed as

$$\begin{aligned}
&\int_{\left\{(\mathbf{s},\mathbf{y}):d^{(tn)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta)\right\}\cap\widehat{A}} \widetilde{P}_{S^{tn}}(\mathbf{s})\widetilde{P}_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
\geq\ & \int_{(\mathbf{s},\mathbf{y}):d^{(tn)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta)} \widetilde{P}_{S^{tn}}(\mathbf{s})\widetilde{P}_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} - \int_{\widehat{A}^c} \widetilde{P}_{S^{tn}}(\mathbf{s})\widetilde{P}_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
=\ & P_{\Delta}^{(n)}(\widetilde{P}_S,\widetilde{W},\mathcal{E},t) - \Pr\left(\widehat{A}^c\right), \qquad (48)
\end{aligned}$$

where the probabilities are with respect to the joint distribution $\widetilde{P}_{S^{tn}}(\cdot)\widetilde{P}_{Y^n|X^n}(\cdot|\cdot)$. Note that the first term in the right-hand side of (48) is exactly the probability of excess distortion for the joint source-channel system consisting of the auxiliary MGS $\widetilde{P}_S$ and the auxiliary MGC $\widetilde{W}$ with transmission t, and, according to our setting, with

$$tR(\widetilde{P}_S,\Delta) = R > R' = C(\widetilde{W},\mathcal{E}).$$

Thus, this quantity converges to 1 as n goes to infinity according to the strong converse JSCC theorem. It remains to show that the second term in the right-hand side of (48) vanishes asymptotically. Note that

$$\begin{aligned}
\Pr\left(\widehat{A}^c\right) \ \leq\ & \Pr\left(\frac{1}{nt}\ln\frac{\widetilde{P}_{S^{tn}}(\mathbf{s})}{P_{S^{tn}}(\mathbf{s})} > F\left(\frac{R}{t},P_S,\Delta\right) + \frac{\delta}{2t}\right) \\
& + \Pr\left(\frac{1}{n}\ln\frac{\widetilde{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x})}{P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})} > E_{sp}(R',W,\mathcal{E}) + \frac{\delta}{2}\right). \qquad (49)
\end{aligned}$$

It follows by the WLLN that as $n\to\infty$,

$$\frac{1}{nt}\ln\frac{\widetilde{P}_{S^{tn}}(\mathbf{s})}{P_{S^{tn}}(\mathbf{s})} \longrightarrow \mathbb{E}_{\widetilde{P}_S}\left[\ln\frac{\widetilde{P}_S(s)}{P_S(s)}\right] = F\left(\frac{R}{t},P_S,\Delta\right) \qquad \text{in Prob.},$$

which implies that

$$\lim_{n\to\infty}\Pr\left(\frac{1}{nt}\ln\frac{\widetilde{P}_{S^{tn}}(\mathbf{s})}{P_{S^{tn}}(\mathbf{s})} > F\left(\frac{R}{t},P_S,\Delta\right) + \frac{\delta}{2t}\right) = 0. \qquad (50)$$

For the second term of (49), setting $\mathbf{z} = \mathbf{y} + a\mathbf{x}$, we can write

$$\frac{1}{n}\ln\frac{\widetilde{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x})}{P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})} = \frac{1}{2}\left[\ln\frac{\sigma_Z^2}{\widetilde{\sigma}_Z^2} - \frac{\mathbf{z}^T\mathbf{z}}{n\widetilde{\sigma}_Z^2} + \frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2} - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} + \frac{(a+1)^2\mathbf{x}^T\mathbf{x}}{n\sigma_Z^2}\right].$$

On the other hand, recalling that a is given in (45) and $\widetilde{\sigma}_Z^2$ is given in (46), and noting that

$$\begin{aligned}
\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} &= \frac{\mathrm{SNR}(\beta'-1) + 2\beta' + \sqrt{\mathrm{SNR}^2(\beta'-1)^2 + 4\mathrm{SNR}\beta'(\beta'-1)}}{2\beta'^2} \\
&= \frac{4\beta'^2}{2\beta'^2[\mathrm{SNR}(\beta'-1) + 2\beta' - \sqrt{\mathrm{SNR}^2(\beta'-1)^2 + 4\mathrm{SNR}\beta'(\beta'-1)}]} \\
&= \frac{2}{2\beta' + \mathrm{SNR}(\beta'-1)\left[1 - \sqrt{1 + \frac{4\beta'}{\mathrm{SNR}(\beta'-1)}}\right]},
\end{aligned}$$

where $\beta' = e^{2R'}$, we see that

$$\frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] \overset{(a)}{=} \frac{\text{SNR}}{4\beta'}\left[(\beta'+1) - (\beta'-1)\sqrt{1+\frac{4\beta'}{\text{SNR}(\beta'-1)}}\right] + \frac{1}{2}\ln\left\{\beta' - \frac{\text{SNR}(\beta'-1)}{2}\left[\sqrt{1+\frac{4\beta'}{\text{SNR}(\beta'-1)}} - 1\right]\right\},$$

which is exactly the sphere-packing bound $E_{sp}(R', W, \mathcal{E})$, and where the derivation of (a) is provided in [29, Appendix B]. Therefore, it suffices to show that

$$\Pr\left(\frac{1}{n}\ln\frac{\widetilde{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x})}{P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})} > \frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] + \frac{\delta}{2}\right)$$
$$= \Pr\left[\left(\frac{1}{\sigma_Z^2} - \frac{1}{\widetilde{\sigma}_Z^2}\right)\left(\frac{\mathbf{z}^T\mathbf{z}}{n} - \widetilde{\sigma}_Z^2\right) - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\left(\frac{\mathbf{x}^T\mathbf{x}}{n} - \mathcal{E}\right) > \delta\right]$$

converges to 0 as n goes to infinity. This is true (as before) since the above probability is less than

$$\Pr\left[\left(\frac{1}{\sigma_Z^2} - \frac{1}{\widetilde{\sigma}_Z^2}\right)\left(\frac{\mathbf{z}^T\mathbf{z}}{n} - \widetilde{\sigma}_Z^2\right) - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} > \delta\right] \tag{51}$$

by the power constraint (8), and $\mathbf{z}^T\mathbf{z}/n \to \widetilde{\sigma}_Z^2$ and $\mathbf{x}^T\mathbf{z}/n \to 0$ in probability 1. This yields

$$\lim_{n\to\infty}\Pr\left(\frac{1}{n}\ln\frac{\widetilde{P}_{Y^n|X^n}(\mathbf{y}|\mathbf{x})}{P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})} \le \frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] + \frac{\delta}{2}\right) = 0. \tag{52}$$

On account of (47), (48), (50) and (52), we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta}^{(n)}(P_S, W, \mathcal{E}, t) \le tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{sp}\left(R - \frac{\varepsilon}{2}, W, \mathcal{E}\right) + \delta.$$

Since the above inequality holds for any rate R in the region $[tR(P_S,\Delta)+\varepsilon, C(W,\mathcal{E})]$ and δ and ε can be arbitrarily small, we obtain that

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta}^{(n)}(P_S, W, \mathcal{E}, t) \le \min_{tR(P_S,\Delta)\le R\le C(W,\mathcal{E})}\left[tF\left(\frac{R}{t}, P_S, \Delta\right) + E_{sp}(R, W, \mathcal{E})\right]. \tag{53}$$

■

C Proof of Theorem 2

Before we start to prove Theorem 2, let us introduce the Gaussian-type class and the type covering lemma for MGS's [1]. For a DMS with finite alphabet $\mathcal{S}$ and a given rational pmf P_S, the type-P class of k-length sequences $\mathbf{s} \triangleq (s_1 s_2 \cdots s_k) \in \mathcal{S}^k$ is the set of sequences that have single-symbol empirical distribution equal to P. Thus, the probability of a particular event (the probability of error, say) can be obtained by summing the probabilities of intersections of various type classes which decay exponentially as the length of sequence approaches infinity [10]. Unfortunately, most of the properties of type classes, as well as the bounding technique of types, do not hold any more for sequences with continuous alphabets. When $\mathcal{S}$ is continuous, we need to find a counterpart to the type classes which partition the whole source space $\mathcal{S}^k$, while keeping an exponentially small probability in the length of sequence.

In [1, Sec. VI. A], a continuous-alphabet analog to the method of types was studied for the MGS by introducing the notion of Gaussian-type classes. Given $\sigma^2 > 0$ and $\epsilon \in (0, \sigma^2)$, the Gaussian-type class, denoted by $\mathcal{T}^\epsilon(\sigma^2)$, is the set of all k-length sequences $\mathbf{s} \in \mathbb{R}^k$ such that

$$|\mathbf{s}^T\mathbf{s} - k\sigma^2| \le k\epsilon, \tag{54}$$

where T is the transpose operation. Based on a sequence of positive parameters $\{\sigma_i^2\}_{i=1}^\infty$, the Euclidean space $\mathbb{R}^k$ can be partitioned using (54), and it can be shown that for the zero-mean MGS, the probability of each type defined by (54) decays exponentially in k [1]. Specifically, the probability of the type $\mathcal{T}^\epsilon(\widehat{\sigma}_S^2)$ under the Gaussian distribution P_S decays exponentially in k at the rate of $D(\widehat{P}_S \parallel P_S)$ within a term that tends to zero as $\epsilon \to 0$, where $\widehat{P}_S \backsim \mathcal{N}(0, \widehat{\sigma}_S^2)$, i.e.,

$$P_{S^k}\left(\mathcal{T}^\epsilon(\widehat{\sigma}_S^2)\right) \le \exp\left\{-k\left(D(\widehat{P}_S \parallel P_S) + \zeta_1(\epsilon)\right)\right\}, \tag{55}$$

where

$$D(\widehat{P}_S \parallel P_S) = \frac{1}{2}\left(\frac{\widehat{\sigma}_S^2}{\sigma_S^2} - \ln\frac{\widehat{\sigma}_S^2}{\sigma_S^2} - 1\right) \tag{56}$$

is the Kullback-Leibler divergence between the two MGS's $\widehat{P}_S$ and P_S, and $\zeta_1(\epsilon) = -\epsilon/\sigma_S^2 - \ln(1 + \epsilon/\widehat{\sigma}_S^2)$. The following type covering lemma is an important tool which we will later employ to derive the lower bound for the JSCC excess distortion exponent.

Lemma 4 (**Covering Lemma for Gaussian-Type Classes** [1]) Given $\sigma_S^2 > \Delta$ and $\mu > 0$, for sufficiently small ϵ and for sufficiently large k, there exists a set $\mathcal{C} \subset \mathbb{R}^k$ of size $|\mathcal{C}| \le \exp\{k[R(P_S, \Delta) + \zeta_2(\epsilon)] + \mu\}$ with

$$\zeta_2(\epsilon) = \frac{1}{2}\ln\frac{\Delta}{(\sqrt{\Delta} - \epsilon)^2 - \epsilon\Delta\left(1 + 4\sqrt{\frac{\Delta}{\sigma_S^2}}\right)} + 2\epsilon + 2\ln\left[1 + \epsilon\left(1 + 4\sqrt{\frac{\Delta}{\sigma_S^2 - \Delta}}\right)\right]$$

if $\sigma_S^2 > \Delta$ and $\zeta_2(\epsilon) = 0$ otherwise, such that every sequence $\mathbf{s} \in \mathcal{T}^\epsilon(\sigma_S^2)$ is contained, for some $\mathbf{c} \in \mathcal{C}$, in the ball of size Δ

$$B(\mathbf{c}, \Delta) \triangleq \left\{\mathbf{s} : \frac{1}{k}\sum_{i=1}^k (s_i - c_i)^2 \le \Delta\right\},$$

where $R(P_S, \Delta)$ is the rate-distortion function of MGS $P_S \sim \mathcal{N}(0, \sigma_S^2)$.

Now we are ready to prove Theorem 2.

Proof of Theorem 2: Fix $t > 0$. In the sequel we let $k = tn$ and assume that k (and hence n) is sufficiently large. For a given $\epsilon \in (0, \Delta)$ small enough, we construct a sequence of Gaussian-type classes $\mathcal{T}_i \triangleq \mathcal{T}^\epsilon(\sigma^2(i))$ by $\sigma^2(i) = \Delta + (2i - 1)\epsilon$, $i = 1, 2, \cdots$. That is,

$$\begin{aligned} \mathcal{T}_i &\triangleq \left\{\mathbf{s} : \left|\mathbf{s}^T\mathbf{s} - k(\Delta + (2i-1)\epsilon)\right| \le k\epsilon\right\} \\ &= \left\{\mathbf{s} : k(\Delta + (2i-2)\epsilon) \le \mathbf{s}^T\mathbf{s} \le k(\Delta + 2i\epsilon)\right\}, \qquad i = 1, 2, \cdots. \end{aligned} \tag{57}$$

Also, we define the set $\mathcal{T}_0 \triangleq \{\mathbf{s} : \mathbf{s}^T\mathbf{s} \le k\Delta\}$ such that all these type classes $(\mathcal{T}_1, \mathcal{T}_2, \cdots)$ together with $\mathcal{T}_0$ partition the whole space $\mathbb{R}^k$. For this special set $\mathcal{T}_0$, we shall use the trivial bound $P_{S^k}(\mathcal{T}_0) \le 1$ and by definition $\mathcal{T}_0$ is covered by the ball $B(\mathbf{0}, \Delta)$; thus, we say that $\mathcal{T}_0$ satisfies the type covering lemma in the sense that there exists a set $\mathcal{C} \triangleq \{\mathbf{0}\}$ of size $|\mathcal{C}| = 1 \le \exp\{k[R(\widehat{P}_S, \Delta)]\}$ such that every $\mathbf{s} \in \mathcal{T}_0$ is covered by the the ball of size Δ, where we let $\widehat{P}_S \sim \mathcal{N}(0, \Delta)$ and hence $R(\widehat{P}_S, \Delta) = 0$.

Based on the above setup, we claim that, first, for all $i = 1, 2, \cdots$, the probability of $\mathcal{T}_i$ under the k-dimensional Gaussian pdf P_{S^k}, denoted by $P_{S^k}(\mathcal{T}_i)$, decays exponentially at the rate of $D(P_S^{(i)} \parallel P_S) + \widetilde{\zeta}_1(\epsilon)$ in k, where $P_S^{(i)}$ is a zero-mean Gaussian source with variance $\sigma^2(i) = \Delta + (2i-1)\epsilon$, and

$$\widetilde{\zeta}_1(\epsilon) = -\frac{\epsilon}{\sigma_S^2} - \ln\left(1 + \frac{\epsilon}{\Delta}\right) \tag{58}$$

is a vanishing term independent of i (cf. (55)). Second, the type covering lemma is applicable for all $\mathcal{T}_i$, $i = 1, 2, \cdots$. Note that when $\sigma^2(i) > \Delta$, $\zeta_2(\epsilon)$ in the type covering lemma can be bounded by

$$\zeta_2(\epsilon) \le \widetilde{\zeta}_2(\epsilon) \triangleq \frac{1}{2} \ln \frac{\Delta}{(\sqrt{\Delta} - \epsilon)^2 - 5\epsilon\Delta} + 2\epsilon + 2\ln[1 + \epsilon + 4\sqrt{\Delta\epsilon}] \tag{59}$$

and is also independent of i. In the sequel, we will denote, without loss of generality, that all these vanishing terms $\widetilde{\zeta}_1(\epsilon)$ and $\widetilde{\zeta}_2(\epsilon)$ by $\zeta(\epsilon)$.

We next employ a concatenated "quantization – lossless JSCC" scheme [2] to show the existence of a sequence of JSC codes for the source-channel pair (P_S, W) such that its probability of excess distortion is upper bounded by

$$\exp[-nE_{RC}(P_S, W, \Delta, \mathcal{E}, t) + o(n)]$$

for n sufficiently large.

First Stage Coding: Δ-admissible Quantization.

It follows from the above setup and the type covering lemma (Lemma 4) that for each $\mathcal{T}_i$ $(i = 1, 2, \cdots)$, there exists a code $\mathcal{C}_i = \{\mathbf{c}^{(i)}\}$ with codebook size $|\mathcal{C}_i| \le \exp\{k[R(P_S^{(i)}, \Delta) + \zeta(\epsilon)] + o(k)\}$ that covers $\mathcal{T}_i$. Recall that we also have, trivially, that a code $\mathcal{C}_0 = \{\mathbf{0}\}$ with $|\mathcal{C}_0| = 1$ which covers $\mathcal{T}_0$. Therefore, we can employ a Δ-admissible quantizer via the sets $\mathcal{C}_i$, $i = 0, 1, 2, ...$ as follows:

$$f_{\Delta,k} : \mathbb{R}^k \longrightarrow \bigcup_{i=0}^{\infty} \mathcal{C}_i$$

such that for every $\mathbf{s} \in \mathbb{R}^k$, the output of $f_{\Delta,k}$ with respect to $\mathbf{s}$ has a distortion less than Δ. We denote the DMS at the output of $f_{\Delta,k}$ by P with alphabet $\bigcup_{i=0}^{\infty} \mathcal{C}_i$ and pmf

$$P(\mathbf{c}^{(i)}) = \int_{\mathbf{s} \in \mathcal{T}_i : f_{\Delta,k}(\mathbf{s}) = \mathbf{c}^{(i)}} P_{S^k}(\mathbf{s}) d\mathbf{s}, \quad \forall \quad \mathbf{c}^{(i)} \in \mathcal{C}_i, \quad i = 0, 1, 2, ...$$

Second Stage Coding and Decoding: Lossless JSCC with Power Constraint $\mathcal{E}$.

For the DMS P and the continuous MC W, a pair of (asymptotically) lossless JSC code

$$\widetilde{f}_n : \bigcup_{i=0}^{\infty} \mathcal{C}_i \longrightarrow \mathcal{X}^n \quad \text{and} \quad \widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \bigcup_{i=0}^{\infty} \mathcal{C}_i$$

is applied, where the encoder is subject to a cost constraint $\mathcal{E}$, i.e., $\widetilde{f}_n \in \mathcal{F}_n^{\mathcal{E}}$. Note that the decoder $\widetilde{\varphi}_n$ creates an approximation $\widehat{\mathbf{c}} = \widetilde{\varphi}_n(\mathbf{y})$ of $\mathbf{c}^{(i)}$ based upon the sequence $\mathbf{y}$ received at the channel output. According to a modified version of Gallager's JSCC random-coding bound (which is derived in Appendix D), there exists a sequence of lossless JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n, \mathcal{E})$ with bounded probability of error

$$\begin{aligned} P_e^{(n)}(P, W, \mathcal{E}) &\triangleq \Pr(\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}) \\ &= \sum_{i=0}^{\infty} \sum_{\mathbf{c}^{(i)} \in \mathcal{C}_i} P(\mathbf{c}^{(i)}) \int_{\mathbf{y}: \widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{c}^{(i)}} P_{Y^n|X^n}\left(\mathbf{y} \,\middle|\, \widetilde{f}_n(\mathbf{c}^{(i)})\right) d\mathbf{y} \\ &\leq \exp\left\{-n \max_{0 \leq \rho \leq 1} \left[E_o(W, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P)\right] + o(n)\right\}, \end{aligned}$$

where $E_0(W, \mathcal{E}, \rho)$ is Gallager's constrained channel function given in (7) and $E_s^{(n)}(\rho, P)$ is Gallager's source function (see Appendix D)) here given by

$$E_s^{(n)}(\rho, P) = \frac{1+\rho}{n} \ln \left\{ \sum_{i=0}^{\infty} \sum_{\mathbf{c}^{(i)} \in \mathcal{C}_i} P(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \right\}.$$

Probability of Excess Distortion.

According to the Δ-admissible quantization rule, if the distortion between the source message $\mathbf{s}$ and the reproduced sequence $\widehat{\mathbf{c}}$ is larger than Δ, then we must have $\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}$. This implies that

$$\begin{aligned} P_{\Delta}^{(n)}(P_S, W, \mathcal{E}, t) &= \Pr\left(d^{(k)}(\widehat{\mathbf{c}}, \mathbf{s}) > \Delta\right) \\ &\leq \Pr\left(\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}\right) \\ &\leq \exp\left\{-n \max_{0 \leq \rho \leq 1} \left[E_o(W, \mathcal{E}, \rho) - t E_s^{(n)}(\rho, P)\right] + o(n)\right\}. \end{aligned} \tag{60}$$

Next we bound $E_s^{(n)}(\rho, P)$ in terms of P_S for k (also n) sufficiently large and when ϵ goes to zero (when N goes to infinity). Rewrite

$$\begin{aligned} E_s^{(n)}(\rho, P) &= \frac{1+\rho}{n} \ln \left\{ \sum_{i=0}^{\infty} \sum_{\mathbf{c} \in \mathcal{C}_i} \left[P_{S^k}(\mathcal{T}_i) P_{S^k}^{(i)}(\mathbf{c}^{(i)})\right]^{\frac{1}{1+\rho}} \right\} \\ &= \frac{1+\rho}{n} \ln \left\{ \sum_{i=0}^{\infty} P_{S^k}(\mathcal{T}_i)^{\frac{1}{1+\rho}} \sum_{\mathbf{c} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \right\} \\ &\leq \frac{1+\rho}{n} \ln \left\{ 1 + \sum_{i=1}^{\infty} P_{S^k}(\mathcal{T}_i)^{\frac{1}{1+\rho}} \sum_{\mathbf{c} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \right\} \end{aligned}$$

where

$$P_{S^k}^{(i)}(\mathbf{c}^{(i)}) \triangleq \frac{P(\mathbf{c}^{(i)})}{P_{S^k}(\mathcal{T}_i)}$$

is the normalized probability over $\mathcal{T}_i$ for each $i = 0, 1, \ldots$. By Jensen's inequality [7] and the type covering lemma, the sum over each $\mathcal{C}_i$ $(i \geq 1)$ can be bounded by

$$\sum_{\mathbf{c}^{(i)} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \leq |\mathcal{C}_i|^{\frac{\rho}{1+\rho}} \leq \exp\left\{ \frac{\rho}{1+\rho} [kR(P_S^{(i)}, \Delta) + \zeta(\epsilon)] + o(k) \right\}$$

for k sufficiently large and ϵ sufficiently small. Recalling that

$$P_{S^k}(\mathcal{T}_i) \le \exp\{-k[D(P_S^{(i)} \parallel P_S) + \zeta(\epsilon)]\},$$

we have

$$E_s^{(n)}(\rho, P) \le \frac{t(1+\rho)}{k} \ln \left\{ 1 + \sum_{i=1}^{\infty} \exp\left[\frac{k}{1+\rho} \left(\rho R(P_S^{(i)}, \Delta) - D(P_S^{(i)} \parallel P_S) + \zeta(\epsilon) \right) + o(k) \right] \right\} \tag{61}$$

for k sufficiently large and ϵ sufficiently small, by noting that $k = tn$. Recall that $P_S^{(i)}$ denotes the Gaussian source with mean zero and variance $\sigma^2(i) = \Delta + (2i-1)\epsilon$. Consequently, using the fact [1] that if the exponential rate of each term, as a function of i, is of the form $U_i = \ln(Ai + B) - Ci$, where A, B, and C are positive reals, then the term with the largest exponent dominates the exponential behavior of the summation, i.e.,

$$\lim_{k\to\infty} \frac{1}{k} \ln \left\{ 1 + \sum_{i=1}^{\infty} \exp\left[k(\ln(Ai+B) - Ci) + o(k) \right] \right\} = \max_{i\ge 1} [\ln(Ai+B) - Ci], \tag{62}$$

we obtain

$$\lim_{n\to\infty} E_s^{(n)}(\rho, P) \le t \max_{i\ge 1} [\rho R(P_S^{(i)}, \Delta) - D(P_S^{(i)} \parallel P_S) + \zeta(\epsilon)]. \tag{63}$$

Note also that the sequence $\left\{ \rho R(P_S^{(i)}, \Delta) - D(P_S^{(i)} \parallel P_S) \right\}_{i=1}^{\infty}$ is non-increasing after some finite i, which means the maximum of (63) is achieved for some finite $\sigma^2(i)$. Letting ϵ go to zero, it follows by the continuity of $R(P_S^{(i)}, \Delta)$ and $D(P_S^{(i)} \parallel P_S)$ as functions of $\sigma^2(i)$ that

$$\lim_{\epsilon\to 0} \sup_{\sigma^2(i)} [\rho R(P_S^{(i)}, \Delta) - D(P_S^{(i)} \parallel P_S) + \zeta(\epsilon)] \quad = \quad \max[\rho R(\widetilde{P}_S, \Delta) - D(\widetilde{P}_S \parallel P_S)]$$

where the maximum is taken over all the MGS $\widetilde{P}_S$ with mean zero and variance $\sigma^2 > \Delta$. Therefore,

$$\lim_{n\to\infty} E_s^{(n)}(\rho, P) \le \left\{ 0, \frac{t}{2} \left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho \right] \right\} = tE(P_S, \Delta, \rho). \tag{64}$$

Finally, on account of (60) and (64), we may claim that, there exists a sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$, where $f_n = \widetilde{f}_n \circ f_{\Delta,k}$ and $\varphi_n = \widetilde{\varphi}_n$, such that for n sufficiently large,

$$P_\Delta^{(n)}(P_S, W, \mathcal{E}, t) \le \exp\left\{ -n \max_{0\le\rho\le 1} [E_o(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)] + o(n) \right\},$$

by which we establish the lower bound $\underline{E}_J(P_S, W, \Delta, \mathcal{E}, t)$ given in (21). Furthermore, when W is an MGC, the bound (22) holds trivially since $\widetilde{E}_o(W, \mathcal{E}, \rho)$ is a lower bound of $E_o(W, \mathcal{E}, \rho)$. ■

D Gallager's Lower Bound for Lossless JSCC Error Exponent

In this appendix, we modify Gallager's upper bound for the error probability of JSCC for discrete memoryless systems, so that it is applicable to a JSCC system consisting of a DMS and a continuous MC with cost constraint $\mathcal{E}$.

A JSC code $(\widetilde{f}_n, \widetilde{\varphi}_n)$ [26] for a DMS P_C and a continuous MC W with transition pdf $P_{Y|X}$ is a pair of mappings $\widetilde{f}_n : \mathcal{C} \longrightarrow \mathcal{X}^n$ and $\widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \mathcal{C}$, where $\mathcal{C} \subseteq \mathcal{S}^{tn}$. That is, each source message $\mathbf{s} \in \mathcal{C}$ with pmf $P_C(\mathbf{s})$ is encoded as blocks $\mathbf{x} = \widetilde{f}_n(\mathbf{s})$ of symbols from $\mathcal{X}$ of length n, transmitted, received as blocks $\mathbf{y}$ of symbols from $\mathcal{Y}$ of length n and decoded as source symbol $\widetilde{\varphi}_n(\mathbf{y}) \in \mathcal{S}$. Denote the codebook for the codewords be $\mathcal{B} \triangleq \{\mathbf{x} = \widetilde{f}(\mathbf{s})\}$. The probability of decoding error is

$$P_e^{(n)}(P_C, W) = P_e^{(n)}(P_C, W, \mathcal{B}) \triangleq \sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s}) \int_{\mathbf{y}\in\mathcal{Y}^n} P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})\, \mathbf{1}\{\widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{s}\} d\mathbf{y}$$

where $\mathbf{1}\{\dot{}\}$ is the indicator function.

We next recast Gallager's random-coding bound for the JSCC probability of error [13, Problem 5.16] for DMS's and continuous MC's and we show the following bound.

Proposition 1 For each $n \geq 1$, given pdf P_{X^n} defined on $\mathcal{X}^n = \mathbb{R}^n$, there exists a sequence of JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n)$ such that for any $0 \leq \rho \leq 1$ the probability of error is upper bounded by

$$P_e^{(n)}(P_C, W) \leq \left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho} \int_{\mathbf{y}\in\mathcal{Y}^n} \left[\int_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x}) P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{x}. \tag{65}$$

Proof: The proof is very similar to Gallager's random-coding bound for discrete systems and appears in [29]. ■

Next, we need a small modification of (65) for the DMS P_C and the MC W to incorporate the channel input cost constraint (8). Let P_X^* be an arbitrary pdf of the channel input on $\mathcal{X}$ satisfying $\mathbb{E}g(X) \leq \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$ (these restrictions are made to make the term $\left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho}$ in (66) grow sub-exponentially with respect to n) and let $P_{X^n}^*$ be the corresponding n-dimensional pdf on sequences of n channel inputs, i.e., the product pdf of P_X^*. We then adopt the technique of Gallager [13, Chapter 7], by setting $P_{X^n}(\mathbf{x}) = \kappa^{-1}\Phi(\mathbf{x})P_{X^n}^*(\mathbf{x})$, where

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \text{if } n\mathcal{E} - \eta \leq \sum_{i=1}^n g(x_i) \leq n\mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

in which $\eta > 0$ is arbitrary, and $\kappa = \int_{\mathbf{x}} P_{X^n}^*(\mathbf{x})\Phi(\mathbf{x})d\mathbf{x}$ is a normalizing constant. Thus, P_{X^n} is a valid probability density that satisfies the constraint (8). We thus have, for any $r \geq 0$,

$$P_{X^n}(\mathbf{x}) \leq \kappa^{-1} e^{r\eta} P_{X^n}^*(\mathbf{x}) e^{r[\sum_{i=1}^n g(x_i) - n\mathcal{E}]}.$$

Substituting the above into (65) for the MC W, changing the summation to integration, and denoting the probability of error under constraint $\mathcal{E}$ by $P_e^{(n)}(P_S, W, \mathcal{E})$, we have

$$\begin{aligned} P_e^{(n)}(P_C, W, \mathcal{E}) \quad &\leq \quad \left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho} \left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho} \\ &\times \int_{\mathbf{y}\in\mathcal{Y}^n} \left[\int_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}^*(\mathbf{x}) e^{r[\sum_{i=1}^n g(x_i) - n\mathcal{E}]} P_{Y^n|X^n}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{y}. \end{aligned} \tag{66}$$

We remark that $\left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho}$ grows with n as $n^{(1+\rho)/2}$ and does not affect the exponential dependence of the bound on n [12], [13, pp. 326–333]. Thus, applying the upper bound for the DMS P_C and the MC W with cost constraint, and noting that P_X^* is an arbitrary pdf satisfying $\mathbb{E}g(X) \le \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$, we obtain

$$P_e^{(n)}(P_C, W, \mathcal{E}) \le \exp\left\{-n \max_{0\le\rho\le 1}\left[E_o(W,\mathcal{E},\rho) - E_s^{(n)}(\rho, P_C)\right] + o(n)\right\}, \tag{67}$$

where $E_o(W,\mathcal{E},\rho)$ is the Gallager's constraint channel function given by (7), $o(n)$ has the form $c_1 \ln n + c_2$ for some constants c_1 and c_2, and $E_s^{(n)}(\rho, P_S)$ is Gallager's source function

$$E_s^{(n)}(\rho, P_C) \triangleq \frac{1+\rho}{n} \ln\left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right].$$

E Proof of Lemma 1

By definition, $F(R,P_S,\Delta)^* = \sup_{R\ge 0}[\rho R - F(R,P_S,\Delta)] = \sup_{R\ge R(P_S,\Delta)} f(R)$, where

$$f(R) = \rho R - \frac{1}{2}\left(\frac{\Delta e^{2R}}{\sigma_S^2} - \ln\frac{\Delta e^{2R}}{\sigma_S^2} - 1\right).$$

Since

$$\frac{\partial f(R)}{\partial R} = 1 + \rho - \frac{\Delta e^{2R}}{\sigma_S^2},$$

it is seen that $f(R)$ is concave and

$$\sup_{R\ge R(P_S,\Delta)} f(R) = f\left(\frac{1}{2}\ln\frac{\sigma_S^2(1+\rho)}{\Delta}\right) = \frac{1}{2}\left[\rho\ln\frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right] > 0$$

if $\frac{\Delta}{\sigma_S^2} \le 1+\rho$, and $f(R)$ is concave decreasing with

$$\sup_{R\ge R(P_S,\Delta)} f(R) = \max_{R\ge 0} f(R) = f(0) = 0 > \frac{1}{2}\left[\rho\ln\frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]$$

if $\frac{\Delta}{\sigma_S^2} > 1+\rho$. The above facts imply that $E(P_S,\Delta,\rho)$ is the convex Fenchel transform of $F(R,P_S,\Delta)$, i.e.,

$$F(R,P_S,\Delta)^* = E(P_S,\Delta,\rho) = \max\left\{0, \frac{1}{2}\left[\rho\ln\frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]\right\}.$$

Finally, $F(R,P_S,\Delta)$ is also the convex Fenchel transform of $E(P_S,\Delta,\rho)$ since $F(R,P_S,\Delta)$ is convex in R.

■

F Proof of Lemmas 2 and 3

Proof of Lemma 2: Note that

$$E_{sp}(R,W,\mathcal{E}) = \max_{\rho\ge 0}[-\rho R + \widetilde{E}_0(W,\mathcal{E},\rho)] = -\inf_{\rho\ge 0}[\rho R - \widetilde{E}_0(W,\mathcal{E},\rho)],$$

which implies that $-E_{sp}(R,W,\mathcal{E})$ is the concave transform of $\widetilde{E}_0(W,\mathcal{E},\rho)$ on

$$\{R : -E_{sp}(R,W,\mathcal{E}) > -\infty\} = \mathbb{R}^+.$$

Thus, the transform

$$(-E_{sp}(R,W,\mathcal{E}))_* = \inf_{R\in\mathbb{R}^+}[\rho R + E_{sp}(R,W,\mathcal{E})]$$

is the concave hull of $\widetilde{E}_0(W,\mathcal{E},\rho)$ in $\rho \in [0,\infty)$. We next show $(-E_{sp}(R,W,\mathcal{E}))_* = \widetilde{E}_0(W,\mathcal{E},\rho)$ by definition. Now if we set

$$\frac{\partial}{\partial R}[\rho R + E_{sp}(R,W,\mathcal{E})] = 0,$$

we have (refer to Appendix A)

$$\sqrt{1+\frac{4\beta}{\mathrm{SNR}(\beta-1)}} = \frac{2\beta}{\mathrm{SNR}}(1+\rho) - 1, \tag{68}$$

where $\beta = e^{2R}$. Substituting (68) back into $(-E_{sp}(R,W,\mathcal{E}))_*$ and using (12) yield

$$(-E_{sp}(R,W,\mathcal{E}))_* = \frac{1}{2}\left[\rho\ln\beta^* + (1-\beta^*)(1+\rho) + \mathrm{SNR} + \ln\left(\beta^* - \frac{\mathrm{SNR}}{1+\rho}\right)\right], \tag{69}$$

where β^* is determined by (68), which can be equivalently written by

$$-(1+\rho) + \frac{1+\rho}{\beta(1+\rho)-\mathrm{SNR}} + \frac{\rho}{\beta} = 0, \tag{70}$$

subject to $\beta > \mathrm{SNR}/(1+\rho)$ according to (69). In this range the left-hand side of (70) is decreasing in β and ranges from $+\infty$ to the negative number $-(1+\rho)$, which means there is a unique β^* satisfying (70). Solving the function (70) for the stationary point β^* we obtain

$$\beta^* = \frac{1}{2}\left(1+\frac{\mathrm{SNR}}{1+\rho}\right)\left[1+\sqrt{1-\frac{4\mathrm{SNR}\rho}{(1+\rho+\mathrm{SNR})^2}}\right]. \tag{71}$$

On the other hand, we can replace

$$\widehat{\beta} = 1 - 2r\mathcal{E} + \frac{\mathrm{SNR}}{1+\rho}$$

in the expression of $\widetilde{E}_o(W,\mathcal{E},\rho)$ given by (13) and obtain

$$\widetilde{E}_o(W,\mathcal{E},\rho) = \max_{\frac{\mathrm{SNR}}{1+\rho}<\widehat{\beta}<1+\frac{\mathrm{SNR}}{1+\rho}} \frac{1}{2}\left[\rho\ln\widehat{\beta} + (1-\widehat{\beta})(1+\rho) + \mathrm{SNR} + \ln\left(\widehat{\beta} - \frac{\mathrm{SNR}}{1+\rho}\right)\right].$$

Maximizing the above over $\widehat{\beta}$ (see [13, p. 339] for details), we see that $\widetilde{E}_o(W,\mathcal{E},\rho)$ has the same parametric form as (69), which implies

$$(-E_{sp}(R,W,\mathcal{E}))_* = \widetilde{E}_o(W,\mathcal{E},\rho),$$

and hence $\widetilde{E}_o(W,\mathcal{E},\rho)$ is the concave transform of $-E_{sp}(R,W,\mathcal{E})$. ■

Proof of Lemma 3: Recall that by Gallager [13, Chapter 7]

$$E_\dagger(R,W,\mathcal{E}) = \max_{0\le\rho\le1}[-\rho R + \widetilde{E}_0(W,\mathcal{E},\rho)] = -\inf_{0\le\rho\le1}[\rho R - \widetilde{E}_0(W,\mathcal{E},\rho)],$$

which means that $-E_\dagger(R, W, \mathcal{E})$ is the concave transform of $\widetilde{E}_0(W, \mathcal{E}, \rho)$ on

$$\{R : -E_\dagger(R, W, \mathcal{E}) > -\infty\} = \mathbb{R}^+.$$

Thus, the transform

$$(-E_\dagger(R, W, \mathcal{E}))_* = \inf_{R \in \mathbb{R}^+} [\rho R + E_\dagger(R, W, \mathcal{E})]$$

is the concave hull of $\widetilde{E}_0(W, \mathcal{E}, \rho)$ in $\rho \in [0, 1]$. Lemma 2 implies that $\widetilde{E}_0(W, \mathcal{E}, \rho)$ is concave in $[0, \infty)$. Thus we have $(-E_\dagger(R, W, \mathcal{E}))_* = \widetilde{E}_0(W, \mathcal{E}, \rho)$ for all $\rho \in [0, 1]$. ■

G Proof of Theorem 4

By comparing (25) and (26) we observe that the two bounds are identical if and only if the minimum of (25) (or (26)) is achieved at a rate no less than the channel critical rate, i.e.,

$$R_m \geq R_{cr}(W) = \frac{1}{2} \ln \left[\frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2} \sqrt{1 + \frac{\text{SNR}^2}{4}} \right]$$

where R_m is the solution of (17). Let

$$f(R) \triangleq \frac{\beta^{\frac{1}{t}}}{\text{SDR}} - \frac{\text{SNR}}{2\beta} \left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}} \right),$$

which is a strictly increasing function of R (refer to (32)), where $\beta = e^{2R}$. In order to ensure that the root of $f(R)$, R_m, is no less than $R_{cr}(W)$, we only need $f(R_{cr}(W)) \leq 0$. This reduces to the condition (27). ■

H Proof of Theorem 6

The upper bound can be established as in a similar manner as the proof of Theorem 1 and is hence omitted. To establish the lower bound, we need to extend the Gaussian type classes and the type covering lemma to MLS's. For given $\alpha > 0$ and $0 < \epsilon < \alpha$, a Laplacian-type class $\mathcal{T}^\epsilon(\alpha)$ is defined as the set of all k-vectors $\mathbf{s} \in \mathbb{R}^k$ such that $\left| \sum_{i=1}^k |s_i| - k\alpha \right| \leq k\epsilon$, i.e.,

$$\mathcal{T}^\epsilon(\alpha) \triangleq \left\{ \mathbf{s} : \left| \sum_{i=1}^k |s_i| - k\alpha \right| \leq k\epsilon \right\}.$$

It can also be shown that the probability of the type class $\mathcal{T}^\epsilon(\widetilde{\alpha})$, for $\widetilde{\alpha} > 0$, under the Laplacian distribution $P_S \backsim L(0, \alpha)$ is bounded by the exponential function

$$P_{S^k}(\mathcal{T}^\epsilon(\widetilde{\alpha})) \leq \exp\left\{ -k \left(\frac{\widetilde{\alpha}}{\alpha} - \ln \frac{\widetilde{\alpha}}{\alpha} - 1 + \zeta(\epsilon) \right) \right\}$$

where $\zeta_3(\epsilon) = -\epsilon/\alpha - \ln(1 + \epsilon/\widetilde{\alpha})$. We next introduce the type covering lemma for Laplacian-type classes.

Lemma 5 (**Covering Lemma for Laplacian-Type Classes** [28]) Given $\alpha > \Delta$ and $\mu > 0$, for sufficiently small ϵ and for sufficiently large k, there exists a set $\mathcal{C} \subset \mathbb{R}^k$ of size $|\mathcal{C}| \leq \exp\{k[R(P_S, \Delta) + \zeta_4(\epsilon)] + \mu\}$ with

$$\zeta_4(\epsilon) = \ln \frac{\Delta}{\Delta - \epsilon} + \ln\left(1 + \frac{\epsilon}{\alpha - \Delta + \epsilon}\right) + \frac{2\alpha\epsilon}{(\alpha - \Delta + \epsilon)(\Delta - \epsilon)}$$

such that every sequence in $\mathcal{T}^{\epsilon}(\alpha)$ is contained, for some $\mathbf{c} \in \mathcal{C}$, in the ball (cube)

$$B(\mathbf{c}, \Delta) \triangleq \left\{ \mathbf{s} : \frac{1}{k} \sum_{i=1}^{k} |s_i - c_i| \leq \Delta \right\}$$

of size Δ, where $R(P_S, \Delta)$ is the rate distortion function of Laplacian source $P_S \backsim L(0, \alpha)$.

Consequently, using Lemma 5, the lower bound can be deduced by employing a similar proof of Theorem 2 and using Fenchel Duality Theorem. ■

I Proof of Theorem 7

For general continuous MS's, unfortunately, we do not have counterparts to the type class and the type covering results of Lemmas 4 and 5 (for MGS's and MLS's, respectively). Hence, to establish the lower bound for the JSCC excess distortion exponent, we need to modify the proof of Theorem 2. We will use a different approach based on the technique introduced in [16] and the type covering lemma [10] for finite alphabet DMS's.

Since the lower bound (30) immediately follows from (29), we only show the existence of a sequence of JSC codes for the source-channel pair (P_S, W) such that its probability of excess distortion is upper bounded by

$$\exp\left\{-n \max_{0 \leq \rho < 1} [E_0(W, \mathcal{E}, \rho) - tE(P_S, \Delta, \rho)] + o(n)\right\}$$

for n sufficiently large. We shall employ a concatenated "scalar discretization - vector quantization - lossless JSCC" scheme as shown in Fig. 8. Throughout the proof, we let $k = tn$, where $t > 0$ is finite, and set $0 < \epsilon < \Delta$ and $0 < \delta < \Delta - \epsilon$.

First Stage Coding: ϵ-Neighborhood Scalar Quantization.

As in [16], we approximate the continuous MS P_S by a DMS $\widetilde{P}_{\widetilde{S}}$ with countably infinite alphabet $\widetilde{S}$ via an ϵ-neighborhood scalar quantization scheme. In particular, for any given $0 < \epsilon < \Delta$, there exists a countable set $\widetilde{S} = \{s_i, i = 1, 2, ...\} \subseteq \mathbb{R}$ with corresponding mutually disjoint subsets $\mathcal{S}_i \subseteq \{s \in \mathbb{R} : d(s_i, s) \leq \epsilon\}$, $i = 1, 2, ...$, such that $\bigcup_{i=1}^{\infty} \mathcal{S}_i = \mathbb{R}$. Specifically, the subsets $\{\mathcal{S}_i\}$ partition $\mathbb{R}$; for example, a specific partition could be $\mathcal{S}_1 = \{s \in \mathbb{R} : d(s_1, s) \leq \epsilon\}$ and

$$\mathcal{S}_i = \{s \in \mathbb{R} : d(s_i, s) \leq \epsilon \quad \text{and} \quad d(s_j, s) > \epsilon \quad \text{for any } j < i\}$$

for $i \geq 2$. Consequently, we can employ a scalar quantizer $f_\epsilon : \mathcal{S} \longrightarrow \widetilde{S}$ to discretize the original MS P_S, such that $f_\epsilon(s) = s_i$ if $s \in \mathcal{S}_i$. Therefore, the first stage coding can be described as a mapping:

$$f_{\epsilon,k} : \mathcal{S}^k \longrightarrow \widetilde{S}^k$$

where $f_{\epsilon,k}(\mathbf{s}) = (f_\epsilon(s_1), f_\epsilon(s_2), ..., f_\epsilon(s_k))$. We denote the source obtained at the output of $f_{\epsilon,k}$ by $\widetilde{P}_{\widetilde{S}}$ with alphabet $\widetilde{\mathcal{S}}$ and pmf

$$\widetilde{P}_{\widetilde{S}}(s_i) = \int_{s \in \mathcal{S}_i} P_S(s) ds, \quad s_i \in \widetilde{\mathcal{S}}.$$

Lemma 6 For any $\epsilon > 0$ and $\rho > 0$, $E\left(\widetilde{P}_{\widetilde{S}}, \Delta + \epsilon, \rho\right) \le E(P_S, \Delta, \rho)$.

Proof: see [29, Appendix E]. ■

Second Stage Coding: Truncating Source Alphabet.

We next truncate the alphabet $\widetilde{S}$ to obtain a finite-alphabet source. Without loss of generality, assuming that $\widetilde{\mathcal{S}} = \{s_1, s_2, ...\}$ such that $\widetilde{P}_{\widetilde{S}}(s_1) \ge \widetilde{P}_{\widetilde{S}}(s_2) \ge \widetilde{P}_{\widetilde{S}}(s_3) \ge \cdots$, then for M sufficiently large, we take $\widehat{\mathcal{S}}$ be the set of the first M elements, i.e., $\widehat{\mathcal{S}} = \{s_1, s_2, ..., s_M\}$. For $s \in \widetilde{\mathcal{S}} = \{s_1, s_2, ...\}$ define function

$$f_M(s) = \begin{cases} s & \text{if } s \in \widehat{\mathcal{S}}, \\ s_1 & \text{otherwise.} \end{cases}$$

Then the second stage coding is a mapping $f_{M,k} : \widetilde{\mathcal{S}}^k \longrightarrow \widehat{\mathcal{S}}^k$, where $f_{M,k}(\mathbf{s}) = (f_M(s_1), f_M(s_2), ..., f_M(s_k))$. We denote the finite-alphabet DMS at the output of $f_{M,k}$ by $\widehat{P}_{\widehat{S}}$ with alphabet $\widehat{\mathcal{S}}$ and pmf

$$\widehat{P}_{\widehat{S}}(s) = \sum_{s_i \in \widetilde{\mathcal{S}} : f_M(s_i) = s} \widetilde{P}_{\widetilde{S}}(s_i) \qquad s \in \widehat{\mathcal{S}}.$$

We now have the following results (Lemma 7 is proved in a similar manner as Lemma 6 and Lemma 8 is proved in [16]).

Lemma 7 For any $\delta > 0$ and $\rho > 0$, $E\left(\widehat{P}_{\widehat{S}}, \Delta + \delta, \rho\right) \le E(\widetilde{P}_{\widetilde{S}}, \Delta, \rho)$ for M large enough.

Lemma 8 [16, Lemma 1] For any δ such that $\mathbb{E} d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\}$, if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E} \exp[t d(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, then

$$\lim_{k \to \infty} -\frac{1}{k} \ln \Pr\left\{d^{(k)}\left[f_{\epsilon,k}(\mathbf{s}), f_{M,k}(f_{\epsilon,k}(\mathbf{s}))\right] > \delta\right\} = r(M)$$

such that $r(M) \to \infty$ as $M \to \infty$, where the expectations are taken under P_S, and the probability is taken under P_{S^k}.

Remark 3 Note also that $\mathbb{E} d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] \to 0$ as $M \to \infty$. Equivalently, Lemma 8 states that for any $0 < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\}$ and $r > 0$,

$$\lim_{k \to \infty} -\frac{1}{k} \ln \Pr\left\{d^{(k)}\left[f_{\epsilon,k}(\mathbf{s}), f_{M,k}(f_{\epsilon,k}(\mathbf{s}))\right] > \delta\right\} \ge r$$

for M sufficiently large.

Third Stage Coding: $(\Delta - \epsilon - \delta)$-Admissible Quantization.

Consider transmitting the DMS $\widehat{P}_{\widehat{S}}$ over the continuous MC W. Since $\widehat{P}_{\widehat{S}}$ has a finite alphabet $\{s_1, s_2, ..., s_M\}$, we now can employ a similar method as used in the proof of Theorem 2. In the sequel we need to introduce the notation of types and the type covering lemma for DMS's with finite alphabets [10].

Let the set of all probability distributions on $\widehat{\mathcal{S}}$ be $\mathcal{P}(\widehat{\mathcal{S}})$. We say that the type of a k-length sequence $\mathbf{s} \in \widehat{\mathcal{S}}^k$ is $P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}}) \subseteq \mathcal{P}(\widehat{\mathcal{S}})$ in the sense that the empirical distribution of $\mathbf{s}$ is equal to $P_{\widehat{S}}$, where $\mathcal{P}_k(\widehat{\mathcal{S}})$ is the collection of all types of sequences in $\widehat{\mathcal{S}}^k$. For any $P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})$, the set of all $\mathbf{s} \in \widehat{\mathcal{S}}^k$ with type $P_{\widehat{S}}$ is denoted by $\mathcal{T}_{P_{\widehat{S}}}$, called type class $\mathcal{T}_{P_{\widehat{S}}}$.

Now we partition the k-dimensional source space $\widehat{\mathcal{S}}^k$ by a sequence of type classes $\left\{\mathcal{T}_{P_{\widehat{S}}} : P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})\right\}$.

Lemma 9 (**Covering Lemma for Discrete Type Classes** [10]) Given $\mu > 0$, for each sufficiently large k depending only on $d(\cdot, \cdot)$ and μ, for every type class $\mathcal{T}_{P_{\widehat{S}}}$ there exists a set $\mathcal{C}_{P_{\widehat{S}}} \subset \mathcal{S}^k$ of size $|\mathcal{C}_{P_{\widehat{S}}}| \le \exp\{k[R(P_{\widehat{S}}, \Delta') + \mu]\}$ such that every sequence $\mathbf{s} \in \mathcal{T}_{P_{\widehat{S}}}$ is contained, for some $\mathbf{c}_{P_{\widehat{S}}} \in \mathcal{C}_{P_{\widehat{S}}}$, in the ball of size Δ'

$$B(\mathbf{c}_{P_{\widehat{S}}}, \Delta') \triangleq \left\{\mathbf{s} : d^{(k)}(\mathbf{s}, \mathbf{c}_{P_{\widehat{S}}}) \le \Delta'\right\},$$

where $R(P_{\widehat{S}}, \Delta')$ is the rate-distortion function of the DMS $P_{\widehat{S}}$.

Let δ be a number satisfying $0 < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\}$. Setting $\Delta' = \Delta - \epsilon - \delta$ in the type covering lemma, we can employ a $(\Delta - \epsilon - \delta)$-admissible quantizer via the sets $\mathcal{C}_{P_{\widehat{S}}}$ as follows:

$$f_{\Delta-\epsilon-\delta,k} : \widehat{\mathcal{S}}^k \longrightarrow \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$$

such that for every $\mathbf{s} \in \widehat{\mathcal{S}}^k$, the output of $f_{\Delta-\epsilon-\delta,k}$ with respect to $\mathbf{s}$ has a distortion less that $\Delta - \epsilon - \delta$ and each $|\mathcal{C}_{P_{\widehat{S}}}|$ is bounded by $\exp\{k[R(P_{\widehat{S}}, \Delta - \epsilon - \delta) + \mu]\}$ for sufficiently large k. We denote the finite DMS at the output of $f_{\Delta-\epsilon-\delta,k}$ by P with alphabet $\bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$ and pmf

$$P(\mathbf{c}_{P_{\widehat{S}}}) = \sum_{\mathbf{s} \in \mathcal{T}_{P_{\widehat{S}}} : f_{\Delta-\epsilon-\delta,k}(s) = \mathbf{c}_{P_{\widehat{S}}}} \widehat{P}_{\widehat{S}^k}(\mathbf{s}), \qquad \mathbf{c}_{P_{\widehat{S}}} \in \mathcal{C}_{P_{\widehat{S}}}, \quad P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}}).$$

Fourth Stage Coding and Decoding: Lossless JSCC with Cost Constraint $\mathcal{E}$.

For the DMS P and the continuous MC W, a pair of (asymptotically) lossless JSC code

$$\widetilde{f}_n : \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}} \longrightarrow \mathcal{X}^n \qquad \text{and} \qquad \widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$$

is applied, where the encoder is subject to a cost constraint $\mathcal{E}$, i.e., $f_n \in \mathcal{F}_n^{\mathcal{E}}$. Note that the decoder φ_n creates an approximation $\widehat{\mathbf{c}} = \varphi_n(\mathbf{y})$ of $\mathbf{c}_{P_{\widehat{S}}}$ based on the sequence $\mathbf{y}$. According to a modified version of Gallager's JSCC random-coding bound (which is derived in Appendix D), there exists a sequence of lossless JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n, \mathcal{E})$ with bounded probability of error

$$P_e^{(n)}(P, W, \mathcal{E}) \triangleq \Pr(\widehat{\mathbf{c}} \neq \mathbf{c}_{P_{\widehat{S}}}) \le \exp\left\{-n \max_{0 \le \rho \le 1} \left[E_o(W, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P)\right] + o(n)\right\}.$$

Analysis of the Probability of Excess Distortion.

For the sake of simplicity, let (see Fig. 8)

$$\widetilde{\mathbf{s}} = f_{\epsilon,k}(\mathbf{s}), \quad \widehat{\mathbf{s}} = f_{M,k}(\widetilde{\mathbf{s}}) \in \mathcal{T}_{P_{\widehat{S}}}, \quad \mathbf{c}_{P_{\widehat{S}}} = f_{\Delta-\epsilon-\delta,k}(\widehat{\mathbf{s}}), \quad \mathbf{x} = f_n(\mathbf{c}_{P_{\widehat{S}}}), \quad \widehat{\mathbf{c}} = \varphi_n(\mathbf{y}).$$

Since $d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) \le d^{(k)}(\mathbf{s},\widetilde{\mathbf{s}}) + d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) \le \epsilon + d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}})$, we have

$$\begin{aligned}
\Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) &\le \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta - \epsilon) \\
&\le \Pr\left(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta - \epsilon, d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) < \delta\right) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \\
&\le \Pr\left(d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta - \epsilon - \delta\right) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta),
\end{aligned}$$

where the probabilities are taken under the joint distribution $P_{S^k}(\cdot)P_{Y^n|X^n}(\cdot|\cdot)$. According to the $(\Delta-\epsilon-\delta)$-admissible quantization rule, $d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta - \epsilon - \delta$ implies that $\mathbf{c}_{P_{\widehat{S}}} \ne \widehat{\mathbf{c}}$, therefore, we can further bound

$$\begin{aligned}
\Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) &< \Pr(\mathbf{c}_{P_{\widehat{S}}} \ne \widehat{\mathbf{c}}) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \\
&\le \exp\left\{-n\left[\max_{0\le\rho\le1}\left[E_o(W,\mathcal{E},\rho) - E_s^{(n)}(\rho,P)\right] + o(n)\right]\right\} + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta)
\end{aligned}$$

for k sufficiently large. It follows from Lemma 8 (also see the remark after it) that

$$\lim_{k\to\infty} -\frac{1}{k}\ln\Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \to \infty$$

as $M \to \infty$. When we take the sum of two exponential functions that both converge to 0, the one with a smaller convergence rate would dominate the exponential behavior of the sum. Therefore, for sufficiently large M which only depends on δ, noting that $k = tn$, we have

$$\liminf_{n\to\infty} -\frac{1}{n}\ln\Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) \ge \liminf_{n\to\infty}\max_{0\le\rho\le1}\left[E_o(W,\mathcal{E},\rho) - E_s^{(n)}(\rho,P)\right]. \tag{72}$$

Consequently, it can be shown by using the method of types (in a similar manner as the proof of Theorem 2) that for M sufficiently large

$$\lim_{n\to\infty} E_s^{(n)}(\rho,P) \le tE(\widehat{P}_{\widehat{S}}, \Delta - \epsilon - \delta, \rho).$$

Using Lemmas 7 and 6 successively, we can approximate $E(\widehat{P}_{\widehat{S}}, \Delta - \epsilon - \delta, \rho)$ by

$$\begin{aligned}
\lim_{n\to\infty} E_s^{(n)}(\rho,P) &\le tE(\widetilde{P}_{\widetilde{S}}, \Delta - \epsilon - 2\delta, \rho) \\
&\le tE(P_S, \Delta - 2\epsilon - 2\delta, \rho).
\end{aligned} \tag{73}$$

Finally, substituting (73) back into (72), and letting $\epsilon \to 0$ and $\delta \to 0$, we complete the proof of Theorem 7.

■

References

[1] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. 44, No. 3, pp. 1041–1056, May 1998.

[2] E. Arikan and N. Merhav, "Joint source-channel coding and guessing with application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 44, No. 5, pp. 1756–1769, Sep. 1998.

[3] A. E. Ashikhmin, A. Barg, and S. N. Litsyn, "A new upper bound on the reliability function of the Gaussian channel,"*IEEE Trans. Inform. Theory*, vol. 46, No. 6, pp. 1945–1961, Sep. 2000.

[4] 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.

[5] T. Berger, *Rate Distortion Theory*, Prentice-Hall, New Jersey, 1971.

[6] Q. Chen and T. R. Fisher, "Image coding using robust quantization for noisy digital transmission," *IEEE Trans. Image Processing*, vol. 7, pp. 496–505, Apr. 1998.

[7] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, 2nd Edition, New York: Wiley, 2006.

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

[9] 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.

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

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

[12] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no.1, pp. 3–18, Jan. 1965.

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

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

[15] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with fidelity criterion," *IEICE Tran. Fundamentals,* vol. E83-A, No. 10, pp 1891-1897, Oct. 2000.

[16] S. Ihara and M. Kubo, "Error exponent of coding for stationary memoryless sources with a fidelity criterion," *IEICE Tran. Fundamentals,* vol. E88-A, no. 5, pp 1339-1345, May 2005.

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

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

[19] U. Mittal and N. Phamdo, "Duality theorems for joint source-channel coding," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1263–1275, July 2000.

[20] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. 31, pp. 835–839, June 1983.

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

[22] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611-656, 1959.

[23] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65-103, Feb. 1967.

[24] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926-943, June 1992.

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

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

[27] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent of discrete communication systems with Markovian memory," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4457–4472, Dec. 2007.

[28] Y. Zhong, F. Alajaji, and L. L. Campbell, "A type covering lemma and the excess distortion exponent for coding memoryless Laplacian sources," in the Proceedings of *23nd Bienn. Symp. Commun.*, Kingston, Canada, pp. 100–103, June 2006.

[29] Y. Zhong, F. Alajaji, and L. L. Campbell, "On The Joint Source-Channel Coding Excess Distortion Exponent for Memoryless Continuous-Alphabet Systems," Mathematics and Engineering Technical Report, Dept. Math. and Stats., Queen's Univ., Kingston, ON, Canada, July 2008. Available [Online] at http://www.mast.queensu.ca/∼web/publications.html/

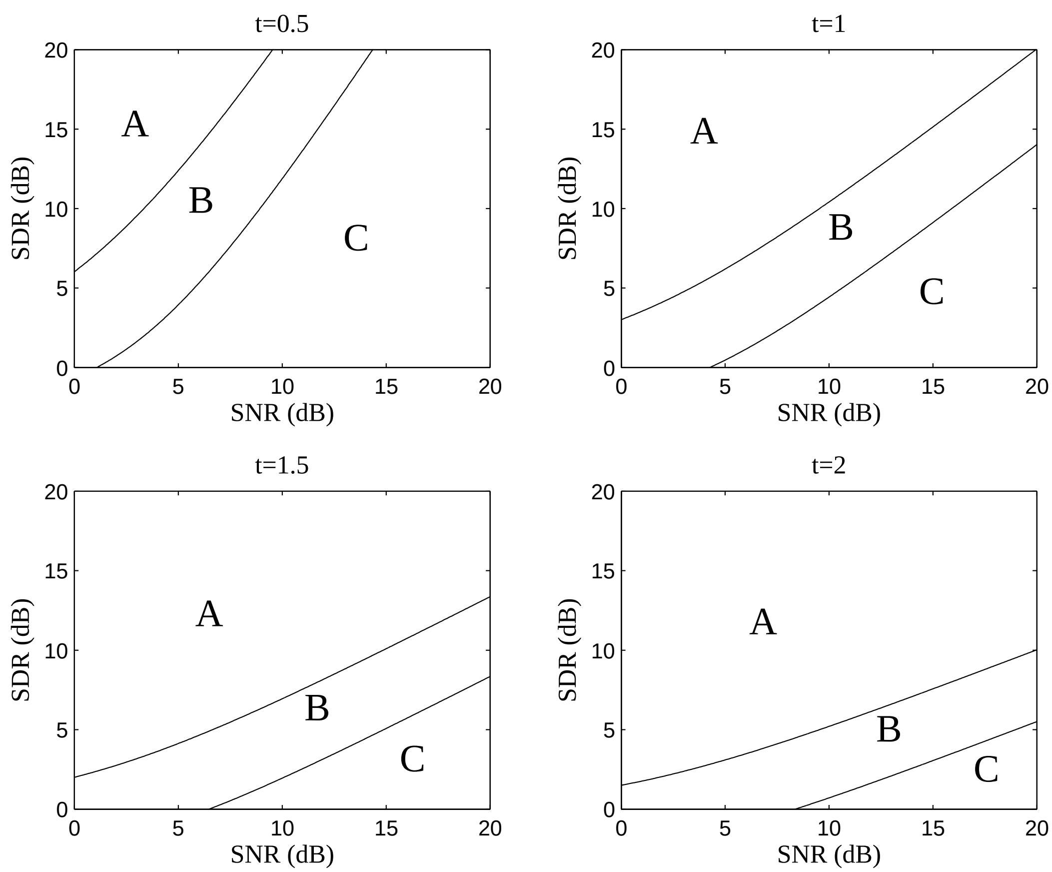

Figure 1: MGS-MGC source-channel pair: the regions for SNR and SDR pairs (both in dB) for different t. In region **A** (including the boundary between **A** and **B**) $E_J = 0$; in region **B** (including the boundary between **B** and **C**), E_J is determined exactly; and in region **C**, $E_J > 0$ is bounded by $\overline{E}_J$ and $\underline{E}_J$.

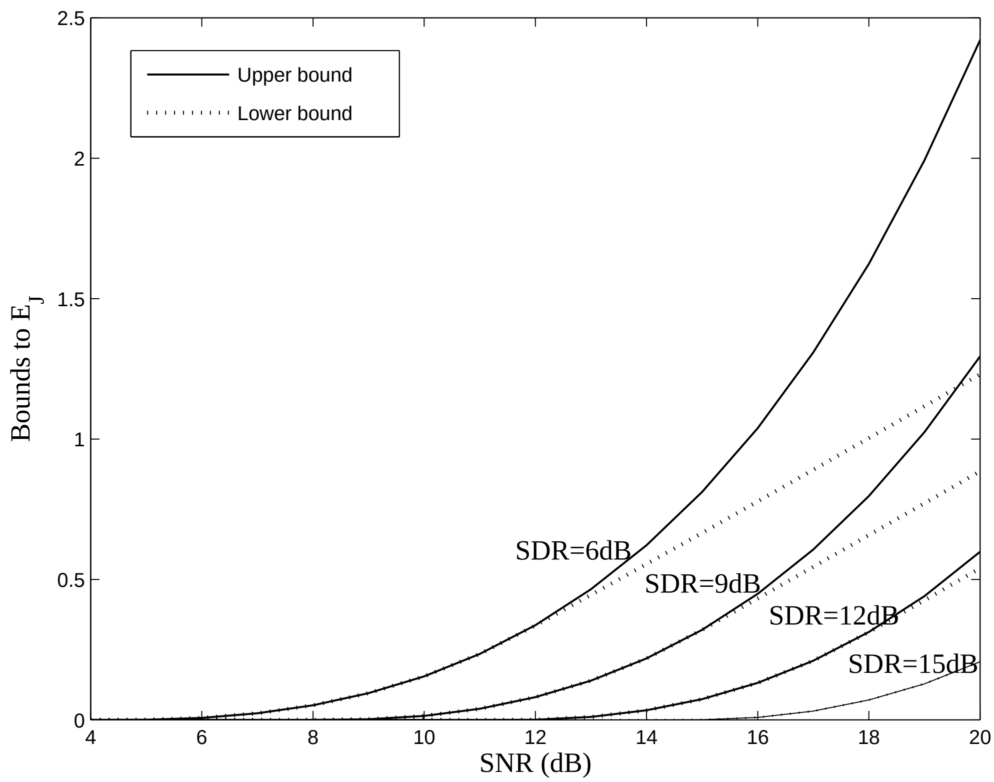

Figure 2: MGS-MGC source-channel pair: the upper and lower bounds for E_J with $t = 1$.

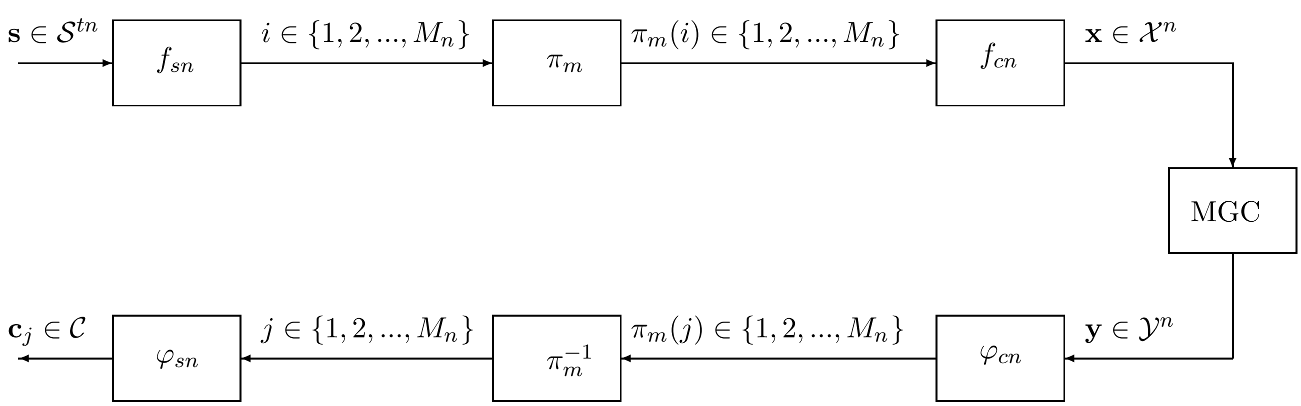

Figure 3: Tandem MGS-MGC coding system.

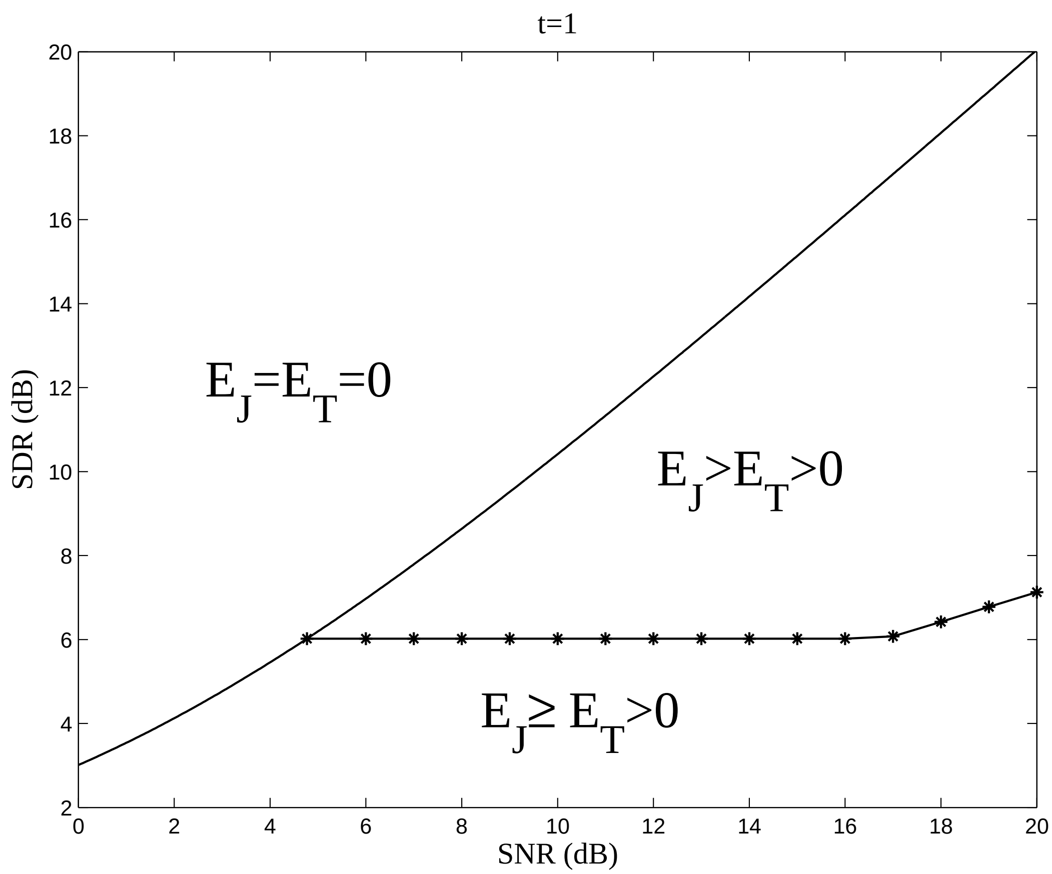

Figure 4: The regions for the MGS-MGC pairs with $t = 1$. Note that the region for $E_J > E_T$ does not include the boundary.

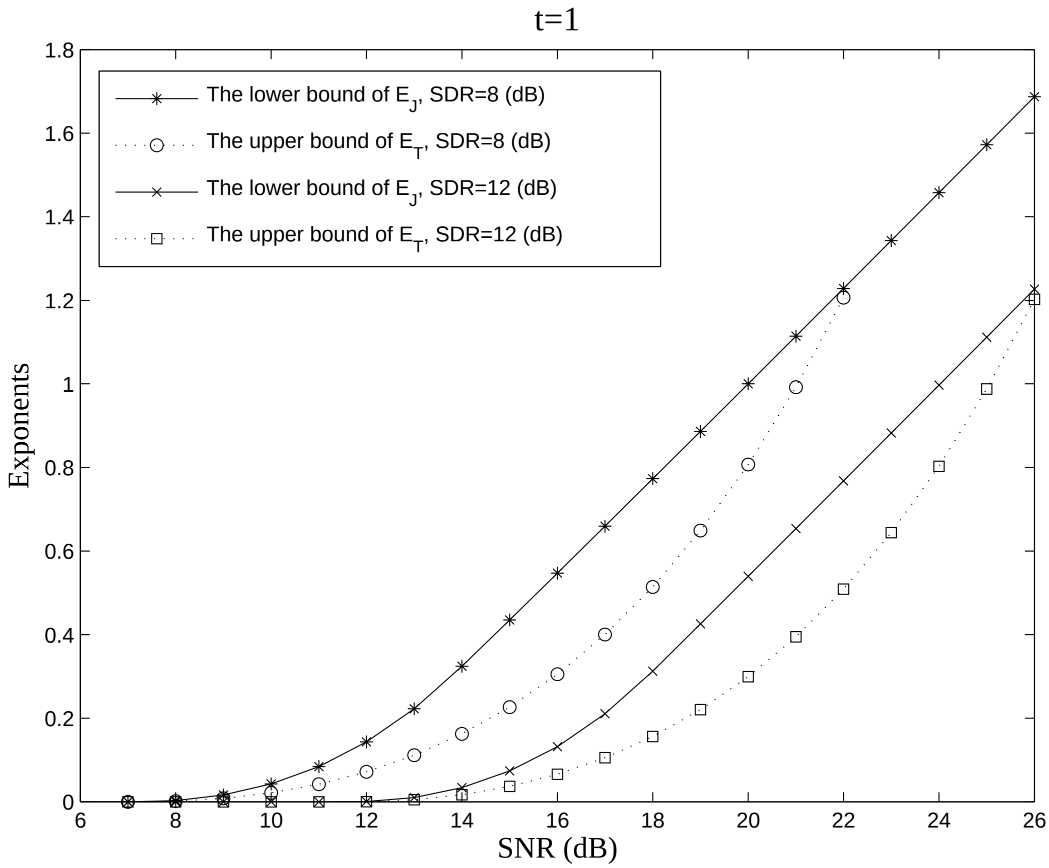

Figure 5: MGS-MGC source-channel pair: $\underline{E}_J$ vs $\overline{E}_T$ for $t = 1$.

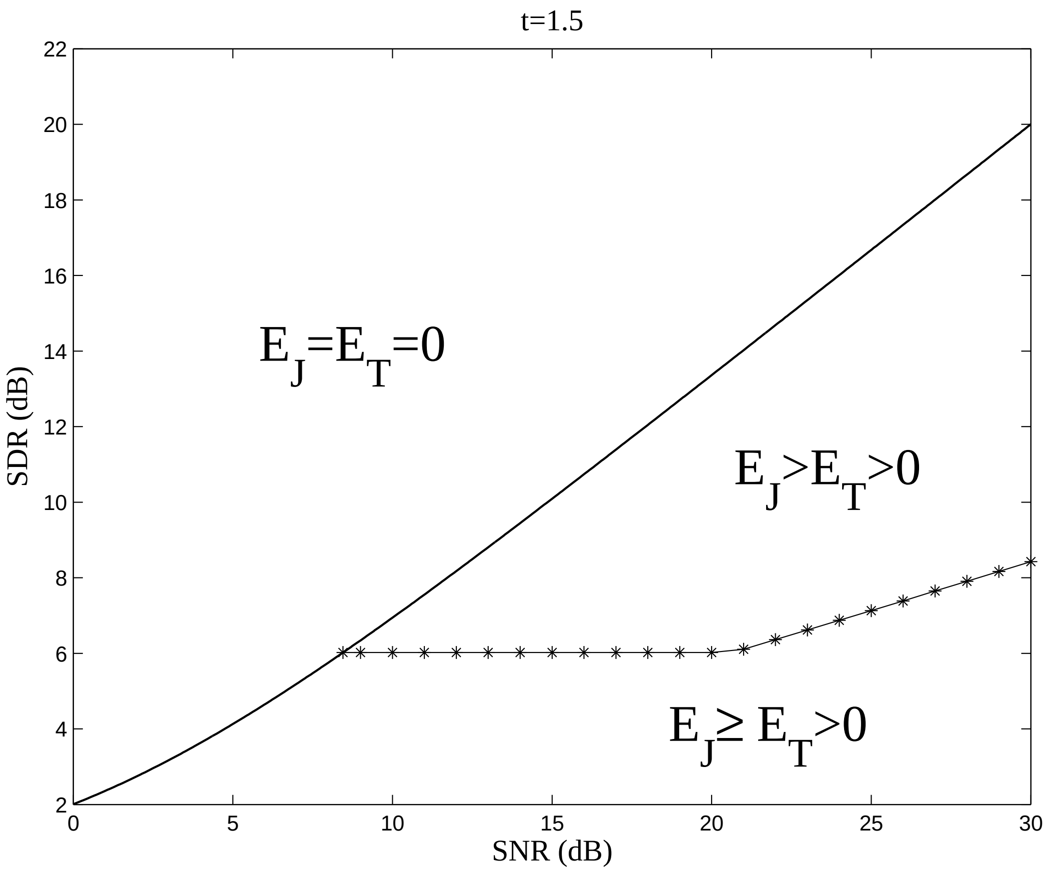

Figure 6: The regions for the MGS-MGC pairs with $t = 1.5$. Note that the region for $E_J > E_T$ does not include the boundary.

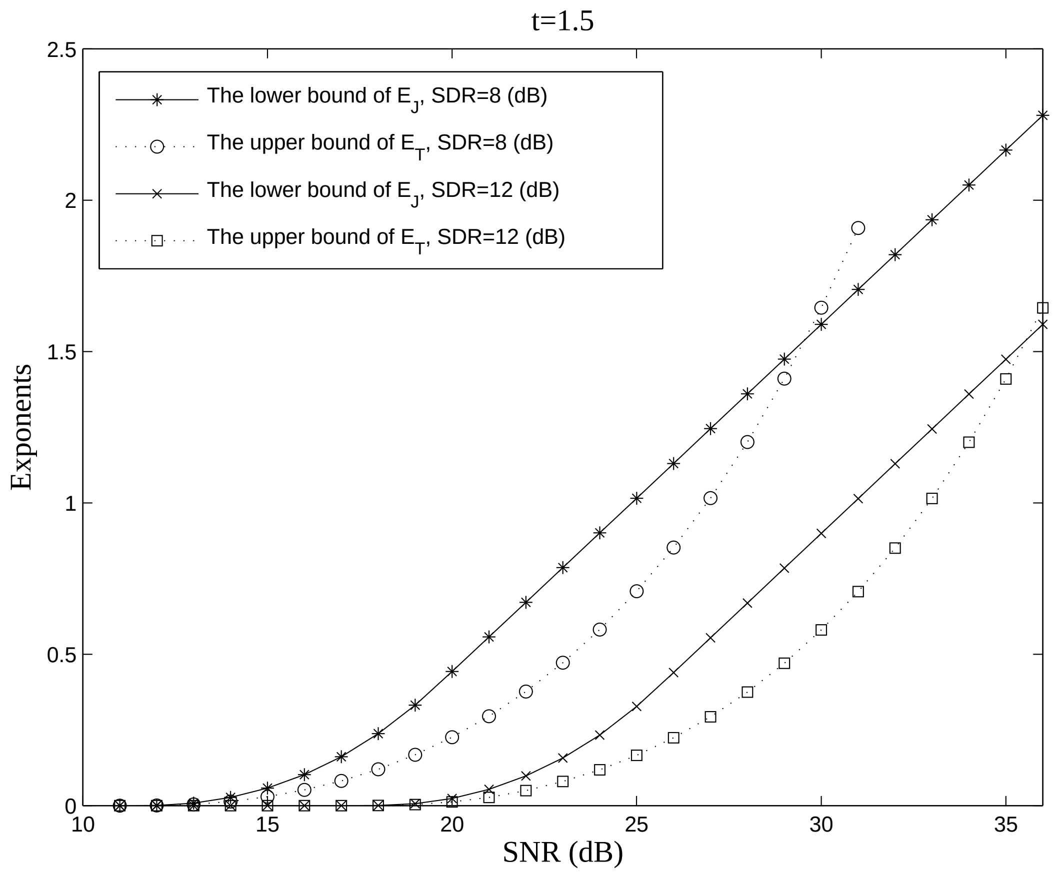

Figure 7: MGS-MGC source-channel pair: $\underline{E}_J$ vs $\overline{E}_T$ for $t = 1.5$.

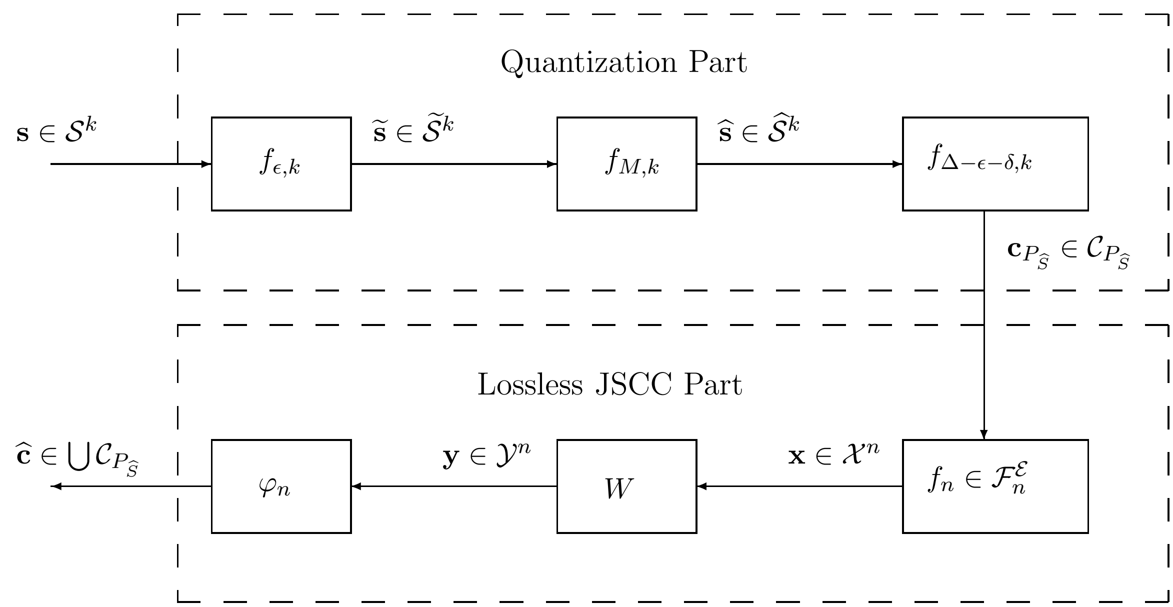

Figure 8: "Quantization plus lossless JSCC" scheme used in the proof of Theorem 7.