RC23442 (W0411-170) November 19, 2004
Mathematics

IBM Research Report

Near Sufficiency of Random Coding for Two Descriptions

Luis Alfonso Lastras-Montaño, Vittorio Castelli
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Near sufficiency of random coding for two descriptions

Luis Alfonso Lastras-Montaño Vittorio Castelli

IBM T.J. Watson Research Center

P.O. Box 218, Yorktown Heights, NY 10598

October 12, 2005

Abstract

We give a single-letter outer bound for the two-descriptions problem for iid sources that is universally close to the El Gamal and Cover (EGC) inner bound. The gaps for the sum and individual rates using a quadratic distortion measure are upper bounded by 1.5 and 0.5 bits/sample respectively, and are universal with respect to the source being encoded and the desired distortion levels. Variants of our basic ideas are presented, including upper and lower bounds on the second channel's rate when the first channel's rate is arbitrarily close to the rate-distortion funcion; these bounds differ, in the limit as the code block length goes to infinity, by not more than 2 bits/sample. An interesting aspect of our methodology is the manner in which the matching single letter outer bound is obtained, as we eschew common techniques for constructing single-letter bounds in favor of new ideas in the field of rate loss bounds. We expect these techniques to be generally applicable to other settings of interest.

I. Introduction

A significant number of the research efforts in the field of multi-terminal Shannon theory seek to find simple descriptions of the limitations and capabilities of systems that process information in a variety of settings. It is fair to state that Shannon's discovery [1] [2] that such simple descriptions were feasible in the basic point-to-point communication scenario, where the process of interest (be it a data source or corrupting noise) is memoryless, constitutes the main momentum behind the field. Among other significant results in the field of source coding are the discoveries due to Slepian and Wolf [3], and Wyner and Ziv [4], that Shannon's basic ideas, namely, random code selection and single-letter converses, suffice to completely characterize the achievable regions in more complex and interesting settings in elegant, surprising, and ultimately relatively simple manners.

The aforementioned works pioneered an explosion of research in which a general theory of multi-terminal communication constitutes the yet-unattained ultimate goal. Implicit in most efforts in this direction is that random coding as a general technique is sufficient to demonstrate the existence of codes with parameters close to any point in the achievable region. Also implicit is the belief that for the iid setting, sharp statements regarding the limitations of communication systems are to be found within the class of single-letter converses, and in particular, expressions involving mutual informations or entropies of scalar random variables.

In spite of the past successes rooted in the beliefs exposed above, a number of important source and channel coding settings are yet to be shown to admit single-letter characterizations. In fact, the comprehensive survey of Kieffer [5] contains the following statement: "the admissible rate region for a memoryless multi-terminal source may not have the single-letter characterization that we have come to expect in dealing with memoryless single-terminal sources". A prominent example is the multiple descriptions (MD) problem, posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow at the September 1979 IEEE Information Theory Workshop, for which such characterization is not known to exist, except for the Gaussian source in the two descriptions case (Ozarow, [6]) , and in restricted settings of the main problem (Ahlswede, [7]). Ozarow and Ahlswede showed that the achievable region of El Gamal and Cover [8] is the actual complete achievable region for specific settings; the conjecture that this might hold for all choices of source statistics and distortion measures was disproved by Zhang and Berger [9]. These authors devised a coding strategy that yields an achievable region containing, for the binary memoryless source and Hamming distortion metric, points that do not belong to El Gamal and Cover's region.

The first main contribution of this paper is to demonstrate that, albeit not tight, El Gamal and Cover's region is actually "nearly" tight. We show this by devising an inner bound to the El Gamal and Cover rate region, and an outer bound to the complete achievable rate region. The inner and outer regions have the form $R_1 > A_1^{in}$, $R_2 > A_2^{in}$, $R_1 + R_2 > A_{1,2}^{in}$, and $R_1 \geq A_1^{out}$, $R_2 \geq A_2^{out}$, $R_1 + R_2 \geq A_{1,2}^{out}$, respectively. We demonstrate that $A_1^{in} - A_1^{out} \leq 1/2$, $A_2^{in} - A_2^{out} \leq 1/2$, and $A_{1,2}^{in} - A_{1,2}^{out} \leq 3/2$, for all sources with finite variance and all distortion triples (D_0, D_1, D_2) with $D_i \in (0, \sigma^2)$ for $i \in \{0, 1, 2\}$ and $0 < D_0 < D_H(D_1, D_2) = (D_1^{-1} + D_2^{-1} - \sigma^{-2})^{-1}$. To simplify the notation we shall assume that $\sigma^2 = 1$ throughout the rest of the paper; this does not result in any loss of generality as the substitution $D_i \to D_i/\sigma^2$ can always be used to retrieve the more general result. The finite variance restriction on the source does not impact the majority of current applications of lossy coding and Feng and Effros [10] have shown that if $D_0 \geq D_H(D_1, D_2)$ then the region $R_1 \geq R(D_1) + 1/2, R_2 \geq R(D_2) + 1/2$ is always achievable.

The second main contribution of this paper is a novel approach to devising single-letter outer bounds, based on Lemma 1, which we believe to be of applicability to a variety of multi-terminal information theory settings. In this light, our result on multiple descriptions serves as an example of the new approach. The root of this contribution can be traced to Zamir [11], where the author compared the minimum rate needed to represent a data source within a certain distortion when side information is present at the decoder (the Wyner-Ziv setting [4]) with the smallest rate needed when side information is present at both the encoder and decoder. Zamir's surprising result was that for squared error and universally for all iid sources, choices of side information, and distortion levels, the additional rate needed in the Wyner-Ziv setting never exceeds 1/2 bit/sample. The basic tools employed by Zamir are generally applicable and have found use in establishing results for different source coding settings, including the fact that progressive lossy source coding can be achieved within a constant distance of the standard rate-distortion function at every stage [12] [13].

As it is commonplace in Shannon theory, the El Gamal and Cover achievable region employs random code selection in order to show the existence of the associated codes; this observation, the discussion in this Introduction, and our results lead to our choice of paper title. Note that our results remain interesting even upon the eventual resolution of the full MD problem, as they reveal competitive optimality aspects of an otherwise suboptimal (in general) technique. A preliminary version of this work has appeared elsewhere [14].

This paper is organized as follows: Section II has the Preliminaries. Statement, proof, and discussion of the main result will be found in Section III. In the same section we also discuss the work of Zamir [15], Feng and Effros [10] and Frank-Dayan and Zamir [16], which to our knowledge are the closest predecessors to our contribution. In Section IV we study a setting in which we impose no-excess rate in one of the channels; the natural question is then whether one can provide necessary and sufficient conditions on the second channel's rate. We succeed in providing upper and lower bounds on said rate whose gap never exceeds 2 bits/sample. Conclusions and Appendix can be found in Sections V and VI, respectively. The Acknowledgments are in Section VII, in which Toby Berger is singled out; the present research direction in fact emerged during a discussion with him.

II. Preliminaries

The most commonly studied version of the multiple descriptions problem is one in which a memoryless stationary source $\{X_i\} = X_1, X_2, \cdots$ is fed to an encoder that produces two packets with data rates R_1 and R_2 bits/sample, respectively. Each sample X_i takes value in an alphabet $\mathcal{X}$ and is governed by a law μ_X. These packets are transmitted over noiseless channels that could fail independently. Failure of a channel results in the complete loss of a transmitted packet. The goal of the encoder is to allow a receiver to produce the best possible guess for the source sequence given the information that it actually receives. As it is often assumed in Shannon theory, the encoder produces these packets every n samples of the source, where n is called the code block-length; we shall denote vectors of length n with bold letters, so that, for example, $\mathbf{X}$ is the input to the encoder.

Formally, the encoder consists of functions

$$\begin{aligned} \mathcal{X}^n &\stackrel{f_1^n}{\to} \{1, \ldots, 2^{nR_1}\}, \\ \mathcal{X}^n &\stackrel{f_2^n}{\to} \{1, \ldots, 2^{nR_2}\}, \end{aligned}$$

and the decoder consists of functions

$$\begin{aligned} \{1, \ldots, 2^{nR_1}\} &\stackrel{g_1^n}{\to} \mathcal{X}^n, \\ \{1, \ldots, 2^{nR_2}\} &\stackrel{g_2^n}{\to} \mathcal{X}^n, \\ \{1, \ldots, 2^{nR_1}\} \times \{1, \ldots, 2^{nR_2}\} &\stackrel{g_0^n}{\to} \mathcal{X}^n. \end{aligned}$$

R_1 and R_2 are called the rates of the code and by definition only take values of the form $j^{-1} \log_2 i$, where j, i are positive integers. For a given code, define

$$\begin{aligned} \mathbf{X}_1 &= g_1^n(f_1^n(\mathbf{X})), \\ \mathbf{X}_2 &= g_2^n(f_2^n(\mathbf{X})), \\ \mathbf{X}_0 &= g_0^n(f_1^n(\mathbf{X}), f_2^n(\mathbf{X})). \end{aligned}$$

The fidelity of a decoder's reconstruction is measured by a distortion function defined as

$$d(x_1^n, y_1^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i)$$

for any n-vectors $x_1^n, y_1^n \in \mathcal{X}^n$ and for a given single-letter distortion measure $d : \mathcal{X} \times \mathcal{X} \to \Re$. In this paper, we will restrict our attention to the case where $d(x, y) = (x - y)^2$, the quadratic distortion measure. The expected distortions of a code for n-vectors are defined as

$$Ed(\mathbf{X}, \mathbf{X}_i) \qquad i \in \{0, 1, 2\}.$$

The properties of a code can be summarized in a quintuple describing the code rates and distortions. We say that a code

$$\left(f_1^n,\, f_2^n,\, g_0^n,\, g_1^n,\, g_2^n\right)$$

achieves $(R_1, R_2, D_0,\, D_1,\, D_2)$ if

$$\begin{aligned} \frac{1}{n}\log_2 \|f_i^n\| &\leq R_i, \quad i \in \{1,2\} \\ E\, d\big(\mathbf{X}, \mathbf{X}_i\big) &\leq D_i, \quad i \in \{0,1,2\} \end{aligned}$$

where the operator $\|\cdot\|$ denotes the cardinality of the range of the corresponding function. We say that the quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if for every $\epsilon > 0$, there exists a sequence of codes, indexed by n, such that for sufficiently large n the corresponding code achieves $(R_1, R_2, D_0, D_1, D_2) + \mathbf{1}^T\epsilon$. Under this asymmetric definition, every code is associated with an achievable quintuple, but the opposite may not be true except in an asymptotic sense. The mathematical results in our paper only address the question whether a code can achieve a certain quintuple, and we employ the notion of achievable quintuples only during a discussion on the relation between our result and those of El Gamal and Cover [8] (henceforth EGC).

The set of all achievable quintuples is called the achievable region. For fixed values of (D_0, D_1, D_2), the set of all (R_1, R_2) such that $(R_1, R_2, D_0, D_1, D_2)$ is achievable is called the achievable rate region for the given distortion triple.

The general goal is to find a complete, computable characterization of the achievable region (for a more thorough discussion on the desired properties of a solution, see Csiszár and Körner [17], p. 259). In the case of multiple descriptions, the general problem is still unsolved. Nevertheless, a variety of inner and outer bounds exist having varying degrees of generality. Sometimes the inner and outer bound match; for example, Ozarow [6] [18] has provided an outer bound that, when combined with the inner bound of El Gamal and Cover [8] and an appropriate forward channel, gives a complete characterization of the achievable region for the Gaussian source with a quadratic distortion measure. Another setting in which it has been shown that EGC is tight was provided by Ahlswede [7], who found the set of achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$ under the "no excess sum-rate" assumption, namely, when $R_1 + R_2 = R(D_0)$. Zhang and Berger [9] have shown via counterexample that EGC is not tight in general. The same authors [19] also found upper and lower bounds to the central decoder distortion under the "no excess marginal rate" assumption, where $R_i = R(D_i)$, $i = 1, 2$. The purpose of this paper is not to serve as a general review, and thus we have left out many other results in the field. Nevertheless the next section describes related work that is used as foundation of the current paper, and Section III-F contrasts our results to those that are closest in spirit within the current literature.

A. The El Gamal-Cover inner region

For the two descriptions problem, known achievable regions include that of El Gamal and Cover [8] and that of Zhang and Berger [9]. The former is the first known achievable region for the multiple descriptions problem, and the latter has been shown to contain points that are not in the former. In this work we only exploit the EGC theorem. Define $\mathcal{EGC}$ as the set of all quintuples $(R_1, R_2, D_0, D_1, D_2)$ such that there exist random variables Y_0, Y_1, and Y_2 jointly distributed with $X \sim \mu_X$

such that

$$\begin{aligned} E[d(X,Y_i)] &\leq D_i, \\ R_1 &> I(X;Y_1), \\ R_2 &> I(X;Y_2), \\ R_1+R_2 &> I(X;Y_0,Y_1,Y_2)+I(Y_1;Y_2). \end{aligned}$$

In some of our theorem statements we shall also refer to $\overline{\mathcal{EGC}}$, which is the defined as the closure of $\mathcal{EGC}$. The EGC theorem states that any quintuple in $\mathcal{EGC}$ is achievable. To prove the EGC theorem, both channel packets are partitioned into two segments each. The first segment of each packet contains all the information that the associated side decoder needs to obtain its reconstruction, and the second segment contains information meant solely for the center decoder. This basic idea was adopted in Frank-Dayan and Zamir's [16] work on dithered quantizers for multiple descriptions, which can be regarded as a natural generalization of earlier work on dithered quantization for point-to-point communication [20], [21].

III. Near sufficiency for two descriptions

A. Statement of main result

Our efforts will be directed towards establishing the following theorem:

Theorem 1: Let X be a generic random variable distributed according to μ_X, with variance $\sigma^2=1$, and consider a source $\{X_i\}$ generating iid samples $\sim \mu_X$.
If a code achieves (R_1,R_2,D_0,D_1,D_2) with $D_i \in (0,1)$ for $i \in \{0,1,2\}$ and $0<D_0<D_H(D_1,D_2) \triangleq \left(D_1^{-1}+D_2^{-1}-1\right)^{-1}$, then

$$\begin{aligned} R_1+R_2 &\geq -L+I(X;X+N_0,X+N_1,X+N_2)+I(X+N_1;X+N_2), \\ L &= \frac{1}{2}\log_2(2-D_0)(2-D_1)(2-D_2), \end{aligned} \tag{1}$$

where N_0, N_1, and N_2 are zero-mean Gaussian random variables with variances P_0, P_1, and P_2; and N_0, N_1, N_2, and X are independent[1]. P_0, P_1, and P_2 are the unique solution of the system of equations

$$\begin{aligned} D_i &= \frac{P_i}{1+P_i}, \qquad i \in \{1,2\}, \\ D_0 &= \frac{P_0P_1P_2}{P_0P_1P_2+P_0P_1+P_1P_2+P_0P_2} \end{aligned} \tag{2}$$

B. Result interpretation

The relevance of this result is established in this subsection; some preliminary remarks follow.

Note that $D_i/(1-D_i)=P_i>0$, for $i=1,2$, where the positivity follows from the assumption that $0<D_i<1$. Also, $P_0=\left(1+1/D_0-1/D_1-1/D_2\right)^{-1}$. It is easy to verify that the positivity of D_1,D_2 and $0<D_0<D_H(D_1,D_2)$ imply $P_0>0$. In the case where $D_0=D_H(D_1,D_2)$ the random variable N_0 and associated parameter P_0 are not needed. The

[1] Recall that, if a collection of k random variables are independent, by definition any sub-collection is also independent.

lower bound on the rate in this case is substituted by $-L + I(X; X+N_1, X+N_2) + I(X+N_1; X+N_2)$. More importantly, in the case $D_0 = D_H(D_1, D_2)$ a result of Feng and Effros [10] gives a better bound than that of Theorem 1. The threshold $D_H(D_1, D_2)$ is intimately associated with the characterization of the achievable region for the quadratic Gaussian problem; in particular, the minimal central distortion D_0 when no excess marginal rate conditions are imposed is exactly $D_H(D_1, D_2)$. This threshold has meaning beyond the Gaussian setting; see Section III-E for details.

Theorem 1, in conjunction with the observation that each of the rates of the channels cannot be smaller than the rate-distortion function of the source evaluated at the appropriate distortion, yields a single-letter outer bound to the achievable region, which in principle can be computed for every source μ_X.

The true significance of this result emerges when it is paired with the EGC theorem as illustrated in Figure 2. Specifically, consider the achievable region for two descriptions obtained by setting (see also the corresponding forward channel in Figure 1)

$$\begin{aligned} Y_i &= \alpha_i(X+N_i), \quad i \in \{1,2\}, \\ U &= X - a(X+N_1) - b(X+N_2), \\ Y_0 &= X - U + \alpha_0(U+N_0), \\ \alpha_i &= \frac{1}{1+P_i}, \quad i \in \{1,2\}, \\ \alpha_0 &= \frac{P_1P_2}{P_1P_2 + P_0(P_1P_2 + P_1 + P_2)}, \\ a &= \frac{P_2}{P_1 + P_2 + P_1P_2}, \\ b &= \frac{P_1}{P_1 + P_2 + P_1P_2}, \end{aligned} \tag{3}$$

in the EGC theorem. Here, X and N_i for $i \in \{0,1,2\}$ are defined as in Theorem 1. It can be verified through substitution that the distortions achieved by this channel are those in Equation (2). These definitions and the resulting distortion computations in (2) appear quite contrived yet they admit reasonably simple explanations [16] that are essentially optimum linear estimation arguments. We shall discuss the motivation behind these definitions at the end of this subsection.

The EGC theorem ensures that $(R_1, R_2, D_0, D_1, D_2)$ is achievable if

$$\begin{aligned} R_i &> I(X; \alpha_i(X+N_i)) = I(X; X+N_i), \qquad i \in \{1,2\}, \\ R_1 + R_2 &> I(X; X - U + \alpha_0(U+N_0), \alpha_1(X+N_1), \alpha_2(X+N_2)) \\ &\quad + I(\alpha_1(X+N_1); \alpha_2(X+N_2)), \\ &= I(X; X+N_0, X+N_1, X+N_2) + I(X+N_1; X+N_2). \end{aligned} \tag{4}$$

Comparison of the sum rate in Equation (4) with the lower bound in Equation (1) reveals that the expressions involving mutual informations are identical in both instances. Note that $L \le 3/2$ bits/sample. Furthermore, from [11] we know that

$$I(X; X+N_i) - R(D_i) \le \frac{1}{2}\log_2(2 - D_i) \quad i \in \{1,2\}.$$

We therefore conclude that, although the EGC theorem does not describe optimal coding strategies for general sources [9], the EGC rates are in fact competitive up to universal constants, namely, constants that do not depend on the source or the

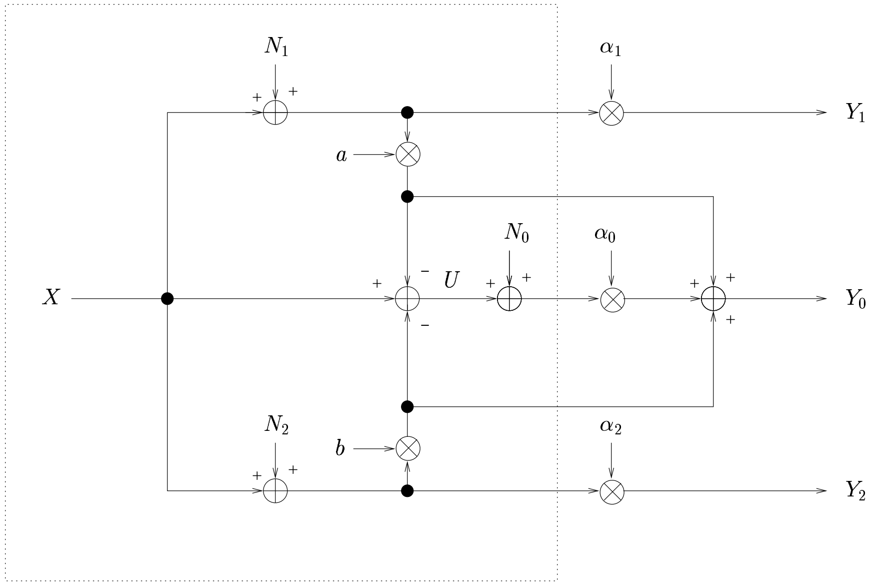

Fig. 1. Test channel used to demonstrate near sufficiency in Theorem 1. The relevant portion of the figure enclosed by the dotted box matches exactly Figure 2 in Frank-Dayan and Zamir [16]. The other portion is necessary to completely specify a test channel to be employed in conjunction with the EGC theorem, and is directly mapped from the appropriate Equations in [16] as well.

distortion triple[2].

Further discussion of the ideas offered in this work can be found in subsections III-E, III-F, and III-G, in which we address the important issue of contrasting these results with those contained in Zamir's work [15], Frank-Dayan and Zamir's [16] and in Feng and Effros' work [10], as well as an extended discussion on our choice of forward channel.

We now provide basic motivation for the definitions in the Equations (3). The values of α_i for $i \in \{1, 2\}$ are such that $\alpha_i(X + N_i)$ is the LMMSE estimator for X from $X + N_i$. The values of a and b are chosen so that the quantity $a(X + N_1) + b(X + N_2)$ is the LMMSE estimator of X from $(X + N_1, X + N_2)$. Finally, α_0 is chosen so that $Y_0 = \alpha_0(X + N_0) + a(1 - \alpha_0)(X + N_1) + b(1 - \alpha_0)(X + N_2)$ is the LMMSE estimator of X from $(X + N_0, X + N_1, X + N_2)$. Although Figure 1 and Equations (3) can be simplified in light of the observation above, we have represented this channel along the lines of Frank-Dayan and Zamir [16] to emphasize the connection to their work. For example, although not proved in this paper, we believe our results can be easily interpreted to be new and interesting performance statements about the dithered quantization technique introduced in [16].

C. Proof of Theorem 1

Let n be sufficiently large and let $(f_1^n, f_2^n, g_0^n, g_1^n, g_2^n)$ be a code for n-vectors for two descriptions that achieves quintuple $(R_1, R_2, D_0, D_1, D_2)$. Recall that

$$\begin{aligned} \mathbf{X}_0 &= g_0^n\left(f_1^n(\mathbf{X}), f_2^n(\mathbf{X})\right), \\ \mathbf{X}_1 &= g_1^n\left(f_1^n(\mathbf{X})\right), \\ \mathbf{X}_2 &= g_2^n\left(f_2^n(\mathbf{X})\right). \end{aligned}$$

[2]More specifically, the theorem requires the existence of the variance of the source, and that D_0 is not greater than $D_H(D_1, D_2)$.

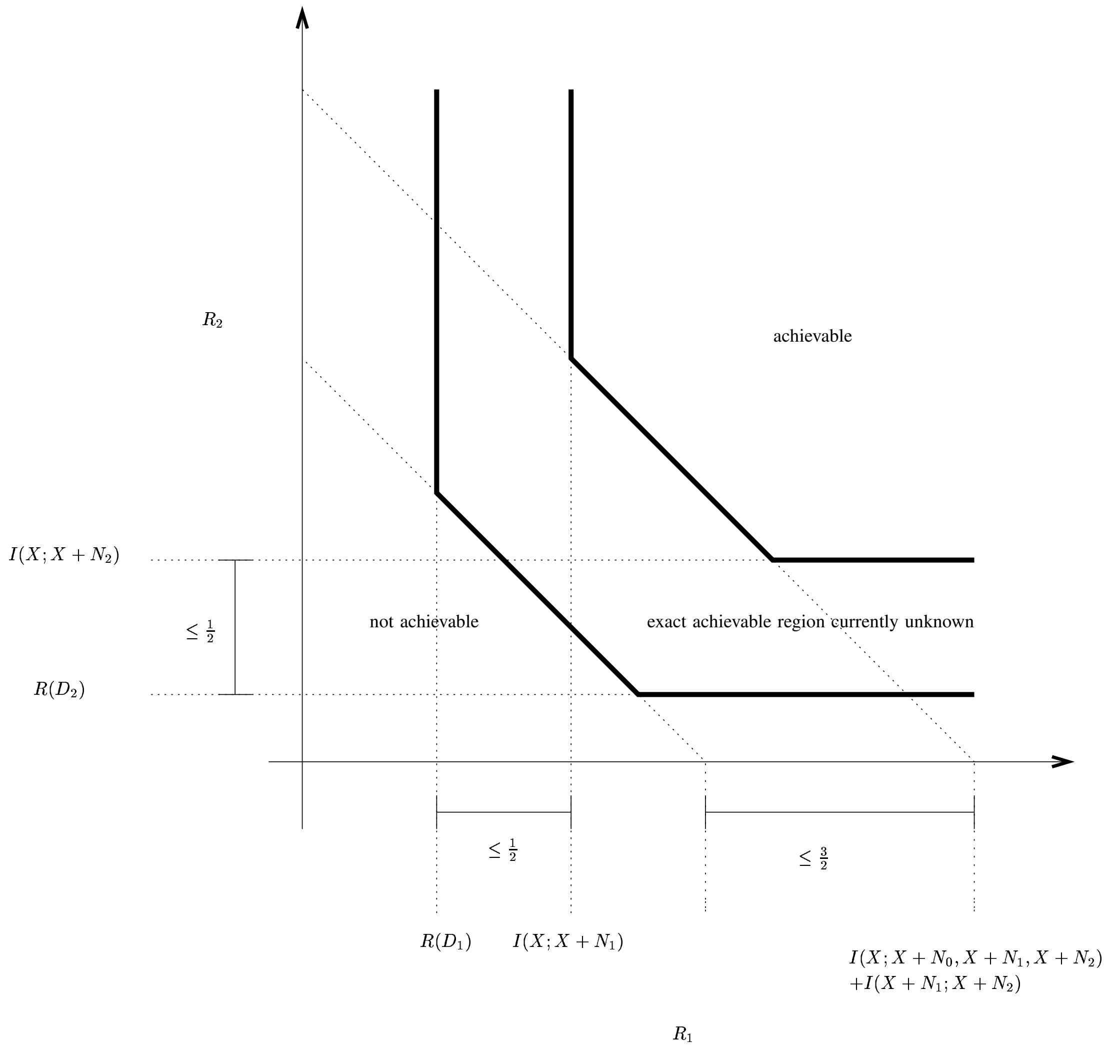

Fig. 2. Illustration of Theorem 1. The constants denoting the maximum width of the gaps can be improved; see the theorem statement and subsequent discussions.

The sum rate of this code is lower bounded by

$$\begin{aligned} n(R_1+R_2) &\geq \log_2\|f_1^n\| + \log_2\|f_2^n\| \\ &\overset{(a)}{\geq} H(f_1^n(\mathbf{X})) + H(f_2^n(\mathbf{X})) \\ &\overset{(b)}{=} H\left(f_1^n(\mathbf{X}), f_2^n(\mathbf{X})\right) + I\left(f_1^n(\mathbf{X}); f_2^n(\mathbf{X})\right) \\ &\overset{(c)}{\geq} I\left(\mathbf{X}; f_1^n(\mathbf{X}), f_2^n(\mathbf{X})\right) + I\left(f_1^n(\mathbf{X}); f_2^n(\mathbf{X})\right) \\ &\overset{(d)}{\geq} I\left(\mathbf{X}; \mathbf{X}_0\right) + I\left(\mathbf{X}_1; \mathbf{X}_2\right), \end{aligned} \tag{5}$$

where (a) follows from the fact that the entropy of a random variable is upper bounded by the logarithm of the cardinality of its alphabet, (b) follows the well-known relation between entropies and mutual information, (c) follows from the non-negativity of the conditional entropy (note that for deterministic codes, i.e., codes for which $H(f_i^n(\mathbf{X})|\mathbf{X}) = 0$, (c) is an equality), and (d) follows from applying thrice the data processing inequality to the Markov chains $\mathbf{X} \rightarrow (f_1^n(\mathbf{X}), f_2^n(\mathbf{X})) \rightarrow \mathbf{X}_0$, $\mathbf{X}_1 \rightarrow f_1^n(\mathbf{X}) \rightarrow f_2^n(\mathbf{X}) \rightarrow \mathbf{X}_2$.

We introduce auxiliary random vectors $\mathbf{N}_0$, $\mathbf{N}_1$, and $\mathbf{N}_2$ having entries independently distributed as zero mean Gaussians with variances P_0, P_1, and P_2, respectively. We assume also that the vectors $\mathbf{N}_0$, $\mathbf{N}_1$, and $\mathbf{N}_2$ are independent. To further bound (5) from below, we seek an upper bound to

$$\{I\left(\mathbf{X};\mathbf{X}+\mathbf{N}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2\right)-I(\mathbf{X};\mathbf{X}_0)\}+\{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}_1;\mathbf{X}_2)\}. \tag{6}$$

To bound the first difference in Expression (6), we apply the chain rule twice to $I(\mathbf{X};\mathbf{X}+\mathbf{N}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2,\mathbf{X}_0)$ and invoke the non-negativity of the mutual information, to obtain

$$I(\mathbf{X};\mathbf{X}+\mathbf{N}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X};\mathbf{X}_0)\leq I(\mathbf{X};\mathbf{X}+\mathbf{N}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2|\mathbf{X}_0). \tag{7}$$

We now use the following lemma, which is based upon the ideas originally published by Zamir [11] and further developed in [12] and [13]. This lemma is a non trivial extension of these results to vectors with a single distortion constraint, and it supports a novel approach to deriving single-letter outer bounds.

Lemma 1: Suppose that $\mathbf{A}$ and $\mathbf{B}$ are n-vectors with the property that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{A}[i]-\mathbf{B}[i])^2\right]\leq\Delta.$$

Further assume that, for every $i\in\{1,\cdots,k\}$, $\mathbf{G}_i$ is a vector with independent Gaussian entries, with zero mean and positive variance σ_i^2. It is also assumed that $\mathbf{A}$,$\mathbf{B}$,$\mathbf{G}_1,\cdots,\mathbf{G}_k$ are independent random vectors. Then

$$n^{-1}\left\{I(\mathbf{A};\mathbf{A}+\mathbf{G}_1,\cdots,\mathbf{A}+\mathbf{G}_k|\mathbf{B})\right\}\leq\frac{1}{2}\log_2\left(1+\sum_{j=1}^{k}\frac{\Delta}{\sigma_j^2}\right).$$

The proof is in the appendix. Recall the assumption that D_0, D_1, and D_2 are strictly positive, and that therefore $P_i>0$, as discussed at the beginning of Section III-B. The lemma can thus be applied to Equation (7), to show that

$$\begin{aligned} I(\mathbf{X};\mathbf{X}+\mathbf{N}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2|\mathbf{X}_0) &\leq \frac{n}{2}\log_2\left(1+\frac{D_0}{P_0}+\frac{D_0}{P_1}+\frac{D_0}{P_2}\right) && (8)\\ &= \frac{n}{2}\log_2\left(2-D_0\right), && (9)\end{aligned}$$

where the last equality can be verified by noting from Equation (2) that

$$\frac{D_0}{P_0} = 1-D_0\left(1+\frac{1}{P_1}+\frac{1}{P_2}\right). \tag{10}$$

The second difference in Expression (6) can be bounded by adding and subtracting $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_2)$ to obtain

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}_1;\mathbf{X}_2) &= \{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_2)\}+\\ &\quad \{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_2)-I(\mathbf{X}_1;\mathbf{X}_2)\}. && (11)\end{aligned}$$

For the first brace, the chain rule applied in its two possible manners to $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2,\mathbf{X}_2)$ and the non-negativity of the mutual information yield

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_2) &= I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2|\mathbf{X}_2)-I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_2|\mathbf{X}+\mathbf{N}_2)\\ &\leq I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2|\mathbf{X}_2). && (12)\end{aligned}$$

Now expand $I(\mathbf{X}, \mathbf{X}+\mathbf{N}_1; \mathbf{X}+\mathbf{N}_2 \mid \mathbf{X}_2)$ using the chain rule in its two possible manners to obtain

$$\begin{aligned} & I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}+\mathbf{N}_2|\mathbf{X}_2) + I(\mathbf{X}+\mathbf{N}_2; \mathbf{X}|\mathbf{X}_2, \mathbf{X}+\mathbf{N}_1) \\ = \quad & I(\mathbf{X}+\mathbf{N}_2; \mathbf{X}|\mathbf{X}_2) + I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}+\mathbf{N}_2|\mathbf{X}, \mathbf{X}_2). \end{aligned} \tag{13}$$

Note that

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}+\mathbf{N}_2|\mathbf{X}, \mathbf{X}_2) &= I(\mathbf{N}_1; \mathbf{N}_2|\mathbf{X}, \mathbf{X}_2) \\ &\overset{(a)}{=} I(\mathbf{N}_1; \mathbf{N}_2) \\ &\overset{(b)}{=} 0, \end{aligned}$$

where (a) follows from the assumption that $(\mathbf{N}_1, \mathbf{N}_2) \perp\!\!\!\perp (\mathbf{X}, \mathbf{X}_2)$ and (b) follows from the assumption that $\mathbf{N}_1 \perp\!\!\!\perp \mathbf{N}_2$. Combining the last equality with Equation (13), we bound the right hand side of Inequality (12) by

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_2; \mathbf{X}+\mathbf{N}_1|\mathbf{X}_2) &\overset{(a)}{\leq} I(\mathbf{X}+\mathbf{N}_2; \mathbf{X}|\mathbf{X}_2) \\ &\overset{(b)}{\leq} \frac{n}{2} \log_2 \left(1 + \frac{D_2}{P_2}\right) \\ &\overset{(c)}{=} \frac{n}{2} \log_2 \left(2 - D_2\right), \end{aligned} \tag{14}$$

where (a) follows from the non-negativity of $I(\mathbf{X}+\mathbf{N}_2; \mathbf{X} \mid \mathbf{X}_2, \mathbf{X}+\mathbf{N}_1)$, (b) follows from Lemma 1, and (c) follows from the definition of P_2 in Equation (2).

To bound the second brace in Equation (11), we expand $I(\mathbf{X}+\mathbf{N}_1, \mathbf{X}_1; \mathbf{X}_2)$ in two manners using the chain rule:

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2) - I(\mathbf{X}_1; \mathbf{X}_2) &= I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2 \mid \mathbf{X}_1) - I(\mathbf{X}_1; \mathbf{X}_2 \mid \mathbf{X}+\mathbf{N}_1) \\ &\leq I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2 \mid \mathbf{X}_1). \end{aligned}$$

To bound the term on the right hand side, we expand $I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}, \mathbf{X}_2|\mathbf{X}_1)$ in two different manners

$$I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2|\mathbf{X}_1) + I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}|\mathbf{X}_1, \mathbf{X}_2) = I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}|\mathbf{X}_1) + I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2|\mathbf{X}, \mathbf{X}_1).$$

From the independence of $\mathbf{N}_1$ and $(\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2)$, it follows that

$$I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2|\mathbf{X}, \mathbf{X}_1) = I(\mathbf{N}_1; \mathbf{X}_2|\mathbf{X}, \mathbf{X}_1) = 0.$$

If the code under consideration is deterministic, we can deduce that the expression on the left hand side is equal to zero from the fact that $\mathbf{X}_2$ is a deterministic function of $\mathbf{X}$. However, the chain of equalities shows that our analysis is not limited to deterministic coding systems. From the above it is deduced that

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1; \mathbf{X}_2|\mathbf{X}_1) &\leq I(\mathbf{X}; \mathbf{X}+\mathbf{N}_1|\mathbf{X}_1) \\ &\leq \frac{n}{2} \log_2 \left(2 - D_1\right), \end{aligned} \tag{15}$$

where the last step follows from an application of Lemma 1. Finally, the proof of the theorem is concluded by substituting the bounds in (9), (14), and (15) in (6) and combining the resulting expression with (5).

D. Variants on the main theorem

We now show how to improve Theorem 1 in two particular settings. In the first case, the side distortions are imbalanced; in the second D_0 is very small with respect to D_1 and D_2. Theorems 2 and 3 address these two regimes, respectively. The ideas employed in these theorems are in fact orthogonal and can be combined to further refine Theorem 1; the details are left to the interested reader.

Theorem 2: Theorem 1 also holds with

$$L = \frac{1}{2}\log_2(2-D_0)\,(2-D_1)\left(1+\frac{D_2}{D_1}-D_2\right),$$

which yields a tighter bound if $D_2 < D_1$. A symmetrical statement holds if $D_1 < D_2$.

Proof. The proof of this theorem deviates from that of Theorem 1 in Equation (11), were we instead write

$$\begin{aligned} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_1;\mathbf{X}_2) &\overset{(a)}{=} \{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1)\} + \\ &\quad\ \{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1) - I(\mathbf{X}_1;\mathbf{X}_2)\} \\ &\overset{(b)}{\leq} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2|\mathbf{X}_1) + I(\mathbf{X}_1;\mathbf{X}+\mathbf{N}_1|\mathbf{X}_2) \\ &\overset{(c)}{\leq} \frac{n}{2}\log_2(2-D_1) + I(\mathbf{X};\mathbf{X}+\mathbf{N}_1|\mathbf{X}_2) \\ &\overset{(d)}{\leq} \frac{n}{2}\log_2(2-D_1) + \frac{n}{2}\log_2\left(1+\frac{D_2}{P_1}\right) \\ &\overset{(e)}{=} \frac{n}{2}\log_2(2-D_1) + \frac{n}{2}\log_2\left(1+\frac{D_2}{D_1}-D_2\right), \end{aligned}$$

where (a) follows by adding and subtracting $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1)$; (b) is obtained by applying the chain rule twice to $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1,\mathbf{X}+\mathbf{N}_2)$ and $I(\mathbf{X}_1;\mathbf{X}_2,\mathbf{X}+\mathbf{N}_1)$ and invoking the non-negativity of the mutual information; (c) results from arguments similar to those that led to Inequality (14) and from the data processing inequality applied to the conditional Markov chain $\mathbf{X}_1 \to \mathbf{X} \to \mathbf{X}+\mathbf{N}_1$ given $\mathbf{X}_2$ (since $\mathbf{N}_1$ is independent of all the other quantities, we have $\mu(\mathbf{X}_1,\mathbf{X}+\mathbf{N}_1 \mid \mathbf{X},\mathbf{X}_2) = \mu(\mathbf{X}_1,\mathbf{N}_1 \mid \mathbf{X},\mathbf{X}_2) = \mu(\mathbf{N}_1)\mu(\mathbf{X}_1 \mid \mathbf{X},\mathbf{X}_2) = \mu(\mathbf{N}_1 \mid \mathbf{X},\mathbf{X}_2)\mu(\mathbf{X}_1 \mid \mathbf{X},\mathbf{X}_2)$, where $\mu(\cdot)$ denotes the distribution of its argument); (d) is an application of Lemma 1; and (e) follows from the definition of P_1.

The symmetric case can be handled with a similar calculation. □

As previewed, we now present a result that is useful when D_0 is small compared to D_1 and D_2.

Theorem 3: Let $D_i \in (0,1)$ for $i \in \{0,1,2\}$ and $0 < D_0 < D_H(D_1,D_2)$. There exists $(R_1^{EGC}, R_2^{EGC}, D_0, D_1, D_2) \in \overline{\mathcal{EGC}}$ such that if a code achieves (R_1,R_2,D_0,D_1,D_2), then

$$R_1+R_2 \geq R_1^{EGC}+R_2^{EGC} - \frac{1}{2}\log_2(2-D_0-D_0/P_0)\,(2-D_1)\,(2-D_2)\,.$$

Remark: From Equation (10),

$$2-D_0-D_0/P_0 = 1+\frac{D_0}{P_1}+\frac{D_0}{P_2}.$$

If D_1, D_2 are fixed (and consequently P_1 and P_2 are fixed too), then $0.5\log_2(2-D_0-D_0/P_0) \to 0$ as $D_0 \to 0$. Moreover, $2-D_0-D_0/P_0 \leq 2-D_0$, which shows that this bound is an improvement over the one found in Theorem 1.

Proof. Let U_0 be a random variable such that $E(U_0-X)^2 = D_0$ and $I(X;U_0) = R(D_0)$. Call F_{X,U_0} the joint distribution of X and U_0, and let $\Phi_{(\mu,\sigma^2)}$ be the Gaussian distribution with mean μ and variance σ^2. Let $\mathbf{X}$, $\mathbf{U_0}$, $\mathbf{N}_1$, and $\mathbf{N}_2$ be n-vectors indexed by i where the quadruples $(\mathbf{X}[i], \mathbf{U_0}[i], \mathbf{N}_1[i], \mathbf{N}_2[i])$ are independent and identically distributed according to $F_{X,U_0}(x, u_0)\Phi_{(0,P_1)}(n_1)\Phi_{(0,P_2)}(n_2)$, where the definition of P_i can be found in Equation (2).

Choose the quintuple $(R_1^{EGC}, R_2^{EGC}, D_0, D_1, D_2)$ anywhere in the region

$$\begin{aligned} R_1 &\overset{(a)}{\geq} n^{-1}I(\mathbf{X};\mathbf{X}+\mathbf{N}_1) \\ R_2 &\overset{(b)}{\geq} n^{-1}I(\mathbf{X};\mathbf{X}+\mathbf{N}_2) \\ R_1 + R_2 &\overset{(c)}{=} n^{-1}I(\mathbf{X};\mathbf{U}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) + n^{-1}I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2) \end{aligned}$$

Note that $I(\mathbf{X};\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2)+I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}+\mathbf{N}_2) = I(\mathbf{X};\mathbf{X}+\mathbf{N}_1)+I(\mathbf{X};\mathbf{X}+\mathbf{N}_2)$ and thus the sum of the right hand sides of (a) and (b) is never greater than the right hand side of (c). It follows that the region is nonempty; it is easy to see that the chosen quintuple belongs to $\overline{\mathcal{EGC}}$. The development now parallels the proof of Theorem 1. Inequality (7) now becomes:

$$\begin{aligned} I(\mathbf{X};\mathbf{U}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X};\mathbf{X}_0) &\overset{(a)}{\leq} I(\mathbf{X};\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2|\mathbf{U}_0) \\ &\overset{(b)}{\leq} \frac{n}{2}\log_2\left(1+\frac{D_0}{P_1}+\frac{D_0}{P_2}\right) \\ &\overset{(c)}{=} \frac{n}{2}\log_2\left(2-D_0-\frac{D_0}{P_0}\right). \end{aligned} \tag{16}$$

Here, inequality (a) follows from expanding $I(\mathbf{X};\mathbf{U}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2)$ using the chain rule, and then noting that $I(\mathbf{X};\mathbf{U_0}) \leq I(\mathbf{X};\mathbf{X}_0)$ by Shannon's rate-distortion converse; Inequality (b) follows from Lemma 1, which holds due to the assumptions on the distributions of $\mathbf{X}$, $\mathbf{U}_0$, $\mathbf{N}_1$, and $\mathbf{N}_2$; finally, Equality (c) can be verified by comparing Inequality (16) with the right hand side of (8) and (9).

E. Discussion of the results

For iid sources our results demonstrate that the EGC theorem suffices to give a single letter characterization of "nearly" all of the achievable region. Loosely speaking, associated with the qualifier "nearly" is one or more constants that are source- and distortion-level-independent. Said constants bound gaps of interest, such as the distance of the Wyner-Ziv rate distortion function to the conditional rate distortion function as in [11].

In this case, the problem is to compare the EGC inner region to the boundary of the exact achievable region. Our methodology has already been discussed extensively in the previous section; our conclusion is that the the gap between the EGC achievable region and the optimum outer bound is no greater than 1.5 bits for the sum rate and 0.5 bits for each of the individual rates (see Figure 2). As stated before, this result is universal for all sources with finite variance for which the EGC theorem holds, and all positive distortion levels with $D_0 < D_H(D_1, D_2)$; for the case $D_0 \geq D_H(D_1, D_2)$ see the discussion on the results of Feng and Effros below. The constants can be improved by substituting in Theorems 1, 2, and 3 specific values for the desired distortion levels, under the assumption that the source has unit variance.

The fact that our results are *valid* for all distortions does not imply that they are *interesting* for all distortions. For example, it is possible for our result to be trivial in that the lower bound that we give to the sum rate could be smaller than $R(D_1)+R(D_2)$ in some instances. This situation may arise if a sufficiently large value for D_0 is prescribed. In those situations in which this is an issue, note that by definition the EGC theorem suffices to exhibit a coding strategy with sum rate no greater than

$R(D_1) + R(D_2) + 1.5$ (to see this note that the individual sum rate thresholds cannot be combined to imply a stronger requirement for the sum rate). In light of this observation, it can be argued that nearly optimal coding strategies exist by definition whenever our result holds trivially.

F. Comparison with prior results

Zamir [15] has obtained inner and outer bounds on the two descriptions achievable region for smooth sources and the quadratic distortion measure. These bounds coincide asymptotically in the high resolution setting. When Zamir's result holds it is obviously strictly better than any of Theorems 1, 2, and 3. The inner bound is based on an application of the El Gamal and Cover inner region as in our work. A principal tool in Zamir's derivations is the Shannon Lower Bound, which states that $R(D) \geq h(X) - \frac{1}{2}\log_2 2\pi e D$, and therefore the distortion regime in which Zamir's results are interesting is related to the particular value of $h(X)$, suggesting that the gap between the inner and outer bounds is not uniformly bounded over all sources. In particular, for sources having discrete components, $h(X) = -\infty$ and thus Zamir's bounds are not applicable.

Also, preceding the present work are Hanying Feng's thesis [22], the subsequent publication of Feng and Effros [10] and the paper of Frank-Dayan and Zamir [16]. Feng and Effros' treatment of the quadratic multiple descriptions problem is based on a partition of the distortion triples in three regions, which coincide exactly with the partitioning needed to state the full set of coding strategies for the Gaussian source with a quadratic distortion measure [6], [8] (note the remarks published in [18]). The three regions $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ are defined, for $\sigma^2 = 1$, as

$$
\begin{aligned}
\mathcal{D}_1 &= \{(D_0, D_1, D_2) \,:\, 0 \leq D_0 < D_L(D_1, D_2),\ D_1 \leq 1, D_2 \leq 1\}, \\
\mathcal{D}_2 &= \{(D_0, D_1, D_2) \,:\, D_L(D_1, D_2) \leq D_0 \leq D_H(D_1, D_2), D_1, D_2 \leq 1\}, \\
\mathcal{D}_3 &= \{(D_0, D_1, D_2) \,:\, D_0 \geq D_H(D_1, D_2), D_1, D_2 \leq 1\},
\end{aligned}
$$

where

$$
\begin{aligned}
D_L(D_1, D_2) &\triangleq \max(0, D_1 + D_2 - 1), \\
D_H(D_1, D_2) &\triangleq \left(D_1^{-1} + D_2^{-1} - 1\right)^{-1}.
\end{aligned}
$$

For the quadratic Gaussian problem, $\mathcal{D}_1$ and $\mathcal{D}_3$ are degenerate regions (see our remarks in subsection III-G). Consider the two rate regions given by conditions

$$
\begin{aligned}
R_1 &\geq R(D_1) + 0.5 \\
R_2 &\geq R(D_2) + 0.5 \\
R_1 + R_2 &\geq R(D_0) + 1.0
\end{aligned}
$$

and conditions

$$
\begin{aligned}
R_1 &\geq R(D_1) + 0.5, \\
R_2 &\geq R(D_2) + 0.5,
\end{aligned}
$$

Feng and Effros [22, Chapter 4, p. 39] [10] have shown that if $D_0 \in \mathcal{D}_1$ then the first rate region is achievable for all iid sources whereas if $D_0 \in \mathcal{D}_3$ then the second rate region is achievable for all iid sources. Since the region $R_1 \geq R(D_1)$,

$R_2 \geq R(D_2)$, $R_1 + R_2 \geq R(D_0)$ is always an outer bound to the achievable rate region, these results give a nearly complete characterization of the achievable region when $D_0 \in \mathcal{D}_1 \cup \mathcal{D}_3$, in the sense of this paper. For region $\mathcal{D}_2$ Feng and Effros do not have a result that holds in comparable generality.

It is of interest to further explore the restrictions on (D_0, D_1, D_2) which define $\mathcal{D}_1$ and $\mathcal{D}_3$. Note that only the ranges $0 \leq D_i \leq 1$ for $i \in \{0, 1, 2\}$ are of interest. If $D_0 \in \mathcal{D}_1$, it is easy to see that $D_0 \leq \min(D_1, D_2)$. Assume without loss of generality that $D_0 \leq D_1 \leq D_2$, and consider an encoder that produces, for some $\epsilon > 0$:

- Packet A: The source encoded using a code with rate $R(D_2) + \epsilon$ at distortion D_2. Note that necessarily $D_2 \geq 1/2$, and thus $R(D_2) \leq 1/2 \log_2 D_2^{-1} \leq 1/2$.
- Packets B and C: A *progressively* encoded representation of the source consisting of two packets, the first of which (packet B) produces a reconstruction within distortion D_1. When packets B and C are available, a reconstruction within distortion D_0 is possible. The results in [12] imply that it is possible to find codes for any source such that the rate of the first packet is not greater than $R(D_1) + 1/2 + \epsilon$ and such that the combined rate of the two packets is not greater than $R(D_0) + 1/2 + \epsilon$.

The encoder then places packet A in channel 2, packet B in channel 1 and splits the bits of packet C arbitrarily among the two channels. This discussion then implies that the region

$$\begin{aligned} R_1 &\geq R(D_1) + 1/2 \\ R_2 &\geq R(D_2) \\ R_1 + R_2 &\geq R(D_0) + 1 \end{aligned}$$

is achievable, which is a slight improvement over Feng and Effros' result in that there is no penalty term for the condition on the rate R_2. In the symmetric case $D_1 = D_2$ an even simpler alternative which produces three independent and rate-distortion optimal representations of the source at distortions D_2, D_1 and D_0 for packets A, B and C respectively, can be used to remove the remaining $1/2$ penalty term in the condition on R_1.

If $D_0 \in \mathcal{D}_3$, note that the absence of D_0 in the description of $\mathcal{D}_3$ immediately implies that only one value of D_0 is of actual interest, namely $D_H(D_1, D_2)$. This result guarantees a universal value for the distortion D_0 for all sources and for coding systems operating in a near no-excess marginal rate setting.

It may be argued that our contribution over Feng and Effros' work is to establish a "near" characterization of the achievable rates in $\mathcal{D}_2$. However, such a narrow statement overlooks other aspects of the present work which we believe are of value. For example, in our case a single argument thread suffices to establish a result which is true for all positive distortion levels ($D_0 < D_H(D_1, D_2)$) and all sources with finite variance. Also, the idea of deriving many-letter converses for a coding problem and then employing Lemma 1 to obtain a "matching" single-letter outer bound appears to be both novel and applicable to a variety of settings.

The work of Frank-Dayan and Zamir is more closely related to the present one in various respects. In TSDQ-MD, three lattices are employed for the quantization step. To produce the side descriptions, the source vector is pre-dithered twice with two dithering vectors which are drawn independently from a uniform distribution on the respective lattice cell. The dithered vectors are then lattice quantized; the indexes of the selected lattice vectors are sent through each channel after entropy coding conditional on the value of the respective dither has been performed. The side decoders' task is to recover the respective lattice vector and to subtract the random dither previously added at the encoder side (there is an assumption that the encoder and decoder share randomness). To produce a reconstruction for the center decoder, the encoder sends additional information regarding the difference between the original source vector and the best linear estimate of the source vector that can be obtained from the data already placed in both channels. This is accomplished via similar techniques, this is, pre-dithering and lattice quantization followed with entropy coding at the encoder side. The information bits of this step are then partitioned among both channels in an arbitrary manner.

A distinct feature of the idea of pre/post dithering and lattice quantization is that in the point-to-point scenario, the rate needed (assuming very good entropy coding) can be written as a mutual information involving the source vector as viewed through an additive noise channel. In the case of TSDQ-MD, a similar statement has been shown to hold in [16], where the test channel of Figure 1 was introduced as a means of analyzing the rate requirements of the proposed technique. Comparison of said requirements with outer bounds for multiple descriptions then can yield insight into the degree of optimality that the technique admits.

One such result states that, under high resolution conditions and for smooth sources, the sum rate in TSDQ-MD is within 0.5 bits of the optimum for infinite lattice dimensions (finite lattice dimension analysis was also provided). This is in contrast with the 1.5 bound that we give in here. Frank-Dayan and Zamir's result uses results from [15], and in particular, also relies in Shannon lower bound techniques, and thus comments similar to those in the first paragraph of this subsection apply in here.

G. A comparison of forward channels

It is known from Ozarow's work [6] that the boundary of the achievable region for the Gaussian multiple descriptions problem with quadratic distortion can be found by substituting in the EGC theorem the random variables $\hat{Y}_0, \hat{Y}_1, \hat{Y}_2$ obtained through the test channel illustrated in Figure 3 (henceforth referred to as the Ozarow channel), which is to be contrasted that of Figure 1 (henceforth referred to as the Frank-Dayan/Zamir or FDZ channel for short). In Ozarow's channel, the variances of $\hat{N}_1$ and $\hat{N}_2$ are still P_1 and P_2 respectively as in the FDZ channel, nevertheless they can be correlated. Formally, the random variables $(\hat{N}_1, \hat{N}_2)$ are independent of X and are zero mean Gaussians with covariance matrix

$$\begin{bmatrix} P_1 & \rho\sqrt{P_1 P_2} \\ \rho\sqrt{P_1 P_2} & P_2 \end{bmatrix}$$

where if $i \in \{1, 2\}$, $P_i = D_i/(1 - D_i)$ (previously defined in Theorem 1), the parameters α_1 and α_2 are as before and the correlation parameter is equal to

$$\rho = -\sqrt{\frac{D_1 D_2 - \exp(-2(R_1 + R_2))}{D_1 D_2}} \tag{17}$$

The central distortion obtained by this channel is equal to

$$D_0 = \frac{\exp(-2(R_1 + R_2))}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2} \tag{18}$$

where

$$\begin{aligned} \Pi &= (1 - D_1)(1 - D_2) \\ \Delta &= D_1 D_2 - \exp(-2(R_1 + R_2)) \end{aligned}$$

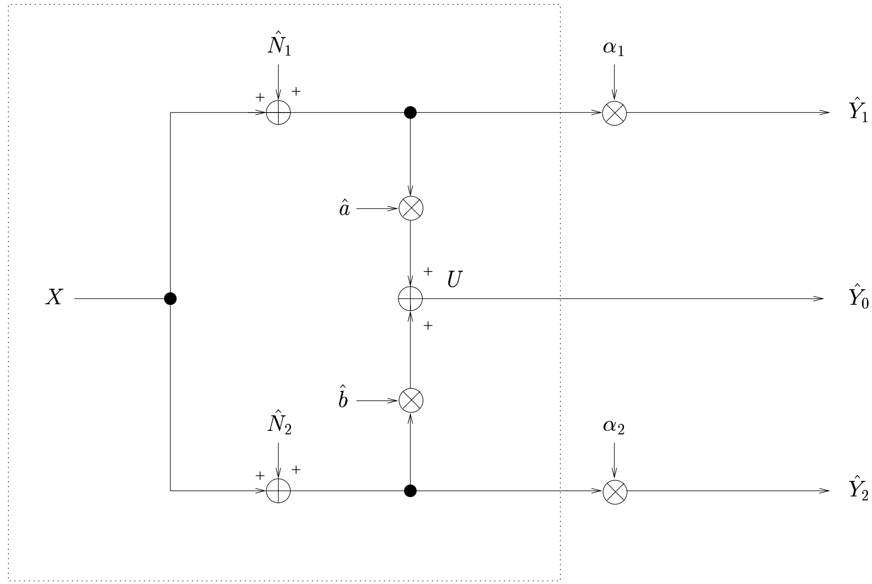

Fig. 3. The forward channel that attains the Gaussian multiple descriptions region (Ozarow [6]).

It is important to note that for the Gaussian case, only those triples (D_0, D_1, D_2) such that $D_0 \in \mathcal{D}_2$ or equivalently

$$D_L(D_1, D_2) \leq D_0 \leq D_H(D_1, D_2)$$

are of intrinsic interest and that the definitions above are meant for this setting only. The reason is that the Gaussian achievable region is such that if $D_0 \in \mathcal{D}_3$ then D_0 can be lowered to $D_H(D_1, D_2)$ without changing the rate requirements or side distortions and if $D_0 \in \mathcal{D}_1$ then at least one of D_1 and D_2 can be improved so that $D_0 = D_L(D_1, D_2)$ without changing the rate requirements; the authors thank Chao Tian for pointing the latter out, see also comments in [15] and [23] in this regard.

The scalar multiplicative factors $\hat{a}$ and $\hat{b}$ are equal to

$$\begin{aligned} \hat{a} &= \frac{P_2 - \rho\sqrt{P_1 P_2}}{(1+P_1)(1+P_2) - (1+\rho\sqrt{P_1 P_2})^2} \\ \hat{b} &= \frac{P_1 - \rho\sqrt{P_1 P_2}}{(1+P_1)(1+P_2) - (1+\rho\sqrt{P_1 P_2})^2} \end{aligned} \quad (19)$$

Note that the expression defining a is identical to that of $\hat{a}$ if one sets $\rho = 0$, with a similar statement holding for b and $\hat{b}$. The values of a and b are such that the expression

$$a(X + N_1) + b(X + N_2)$$

is the best linear estimate of the arbitrary source X (with $EX = 0$ and $EX^2 = 1$) from the random vector $(X + N_1, X + N_2)$ in the mean square sense with a parallel statement being true for $\hat{a}$ and $\hat{b}$ when estimation of X from $(X + \hat{N}_1, X + \hat{N}_2)$ is of concern.

Finally, the other difference between the Ozarow channel and the FDZ channel is that in the latter we have the presence of an alternate path (the "middle path") connecting the source to the destination. It is worth noting that in El Gamal and Cover's discussion of the Gaussian case [8] not only possible correlation between $\hat{N}_1$ and $\hat{N}_2$ is considered but also an additional noisy look at X is introduced, to take possible advantage of the "refining layer" in El Gamal and Cover's theorem. Since every point in $\mathcal{D}_2$ is attainable with the structure in Figure 3, this refinement layer is not necessary for the Gaussian case (we thank Chao Tian again for this observation).

It is interesting to understand the setting where these two channels are identical. Assume a Gaussian source with unit variance and a quadratic distortion measure. It can be easily checked that for the Ozarow channel if $\rho = 0$ then

$$R_i = \frac{1}{2}\log_2 \frac{1}{D_i} \qquad i \in \{1,2\}$$

This corresponds to the so called "no excess marginal rate" setting. It can be checked that the best central error that can be attained under these conditions is equal to $D_H(D_1, D_2)$. For other combinations of source statistics and distortion measure this choice of a channel will generally not result in optimal coding strategies and hence the rates achievable through this channel in principle do not correspond to the "no excess marginal rate" regime; the work of Feng and Effros [10] shows nevertheless that these rates are universally close to the rate-distortion function of the specific problem under consideration.

The negative correlation is a basic property of optimal coding schemes for the quadratic Gaussian problem, aside from the no excess marginal rate setting. Moreover by Zamir's work [15] it is known that the Ozarow channel is asymptotically optimal as D_0, D_1 and D_2 go to zero with the ratios D_1/D_0 and D_2/D_0 held fixed, for quadratic distortion and for generic sources with finite differential entropy.

The discussion above suggests that good MD codes should exploit negative correlation in a manner akin to Ozarow's solution to the Gaussian problem; yet in this work we have chosen the FDZ channel instead as our generic forward channel. An explanation for our decision is as follows: our initial attack we actually used the Ozarow channel but we were unable to uncover a method of proof that would eventually lead to the desired statement, namely, that there is a sense in which the two descriptions problem is "almost" solved. Yet when we switched to the FDZ channel, we were able to make faster progress to the point of obtaining a result that met our goal expectations.

Nevertheless, the question still remains of whether the Ozarow channel can yield a better universal bound. For the quadratic Gaussian problem, it can be seen that the Ozarow and FDZ regions have identical marginal rate conditions

$$R_i \quad \geq \quad \frac{1}{2}\log_2 \frac{1}{D_i} \qquad i \in \{1,2\}$$

but differ in that for the FDZ channel we have that the best central distortion is

$$D_0 = \frac{\exp(-2(R_1 + R_2))}{1 - \Pi}$$

and for the Ozarow channel we have the expression in Equation (18). Examination of these two expressions immediately show that Ozarow's central distortion is never greater than that of FDZ. Yet another reason for studying the possibility of using the Ozarow channel instead in our analysis is that the bound on the inefficiency gap of the EGC region is three times $1/2$ bit. Generally, the rate loss calculations have a curious tendency to "pile up" $1/2$ bit penalties in a manner proportional to the complexity of the forward channel being examined (see nevertheless an exception to this observation [24]). Since the FDZ channel has an additional "middle look" at X than Ozarow's channel does, one may conjecture that one can at least improve our bounds by $1/2$ bit by using Ozarow's instead.

We claim nevertheless that the case for negative correlation in "universal" forward channels is not as clear as one may suspect. Assume that the source X has *finite entropy*, zero mean and unit variance. We shall assume that the reconstruction alphabet is the entire real line, so that the Ozarow channel is a feasible forward channel. We shall also use squared error as the distortion measure.

Suppose one wishes to design a system with the property that $D_0 = \epsilon$, $R_1 = R_2 < H(X)$ and $D_1 = D_2 > 0$ is the smallest possible distortion achievable through the rates provisioned and under the ϵ distortion constraint for D_0. These parameters imply a system operating on the boundary of the achievable region. Clearly, this can be done with a total (asymptotic) sum rate no greater than $2H(X)$. We are considering designing this system via the generic forward channel in Figure 3 together with the EGC theorem. We are willing to tolerate some rate loss and our goal is to choose the best channel parameters and subsequently bound said rate loss. Under this constraint, we are forced to choose

$$P_1 = P_2 = \frac{D_1}{1 - D_1} \tag{20}$$

so as to satisfy the side distortion requirement. The value that Ozarow chooses for ρ in (17) was obtained to give maximum negative correlation (and hence smallest possible central distortion) with the condition that the rate requirements of the EGC theorem are not violated. The Gaussian assumption is obviously used at this step and hence we can no longer use the definition (17) and subsequent calculation (18) for this discussion. To avoid using information about the source statistics beyond the zero mean and unit variance assumptions, we consider ρ as a parameter of the desired distortion at the center decoder rather than a parameter of the rates R_1 and R_2. The best estimator for X still uses the values for $\hat{a}$ and $\hat{b}$ in (19). The resulting central distortion is equal to

$$D_0 = \frac{D_1(\rho + 1)}{D_1(\rho - 1) + 2}$$

From this it can be seen that to obtain $D_0 = \epsilon$ with ϵ a vanishingly small number and with a fixed $D_1 > 0$ it is necessary to set ρ close to -1. Consider the condition on the sum rate obtained by substituting on the EGC theorem the output random variables obtained through Ozarow's channel with the modified correlation parameter:

$$\begin{aligned} R_1 + R_2 &\geq I(X; X + \hat{N}_1, X + \hat{N}_2) + I(X + \hat{N}_1; X + \hat{N}_2) \\ &= I(X; X + \hat{N}_1) + I(X; X + \hat{N}_2) + I(\hat{N}_1; \hat{N}_2) \\ &= I(X; X + \hat{N}_1) + I(X; X + \hat{N}_2) + \frac{1}{1 - \rho^2} \end{aligned}$$

As we can see, the lower bound on the sum rate can be arbitrarily large, as ρ approaches -1. Yet as discussed no multiple description scheme should necessitate more than $2H(X)$ bits per sample. On the other hand, from the second variant of subsection III-D we know of the existence of a technique that does not use negative correlation that would attain for this setting a sum rate that is within $\log_2(2 - D_1)$ bits of the optimum as $\epsilon \to 0$.

We stress nevertheless that the above discussion does not rule out the possibility that some form of negative correlation (or more generally "randomness cancellation") is generally advantageous and thus would improve our results. For example, we could consider the more general channel in El Gamal and Cover's treatment of the Gaussian case [8] and optimize the rates accordingly. We do not present such derivation in this paper; our intended message is that the best way to employ correlation for general sources (and specifically, for sources with finite entropy) cannot be mapped directly from the way it is used in the Gaussian setting.

H. A comment on other forward channels

Throughout this paper we have made extensive use of the robust properties of forward channels assembled using various Gaussian additive noises as building blocks. Nevertheless, other possibilities exist; for example Zamir [11] has obtained a rate loss calculation for the binary source and Hamming distortion using binary noise instead of Gaussian noise to construct the forward channel. A reasonable question is whether the techniques presented here can be adapted to use more specialized forward channels (say, for the binary source) to obtain outer bounds that improve previous outer bounds (see for example [19]). We do not present such derivations in this paper for two reasons: a) the goal of the paper was to demonstrate the existence of a universal near tightness result, as opposed to improved bounds on specific sources (interesting as this topic may be in itself) and b) as a consequence of a), our choice of tools and approach gives general but coarse results and therefore we do not intend nor expect to be competitive with the best bounds for specific settings.

IV. No excess rate on one channel

The idea of approximating optimal forward channels with ones that use additive dithers has proved fruitful, as exemplified in [11], [12], [13], [22]. Nevertheless, additive dithers need not be postulated from the outset when proposing a forward channel to be employed in conjunction with a positive coding theorem, but rather only introduced later as a proof technique. This has the drawback of distancing the coding technique from the attractive dithered lattice quantization ideas that have been associated with additive dithers, but can sometimes yield interesting statements regarding the structure of the achievable region, as in the works of [12] and [13], where, for successive refinement, bounds to the rate loss in one stage are computed when the rate loss in the other stage is exactly zero.

Ideas in this spirit are exploited in the subsequent calculation, where we impose the condition that the rate in channel one be very close to the rate-distortion function of the source evaluated at D_1. The relevant question is how much rate is required in channel two to satisfy the desired distortion requirements D_0 and D_2. In this respect we offer the following result:

Theorem 4: Let $D_i \in (0,1)$ for $i \in \{0,1,2\}$ and $0 < D_0 < D_H(D_1, D_2)$. There exists $(R(D_1), R_2^{ECG}, D_0, D_1, D_2) \in \overline{\mathcal{EGC}}$ such that, if for some $\epsilon > 0$ a code achieves $(R(D_1)+\epsilon, R_2, D_0, D_1, D_2)$, then

$$\begin{aligned} R_2 &\geq R_2^{EGC} - \epsilon - \frac{1}{2}\log_2(2-D_0-D_0/P_0)(2-D_1)^2(2-D_2) \\ &\geq R_2^{EGC} - \epsilon - 2 \text{ bits/sample.} \end{aligned}$$

To prove Theorem 4 we shall make use of the following basic result, which in turn is proved in the Appendix:

Lemma 2: Suppose that $A \to D \to (B,C)$. Then $I(D;A|BC) \leq I(D;A|B)$.

Proof of Theorem 4. Let U_0 and U_1 be such that $I(X;U_0) = R(D_0)$ and $I(X;U_1) = R(D_1)$, $U_0 \to X \to U_1$, and let N_1 and N_2 be such that N_1, N_2 and (X, U_0, U_1) are independent and $N_1 \sim \phi(0, P_1)$, $N_2 \sim \phi(0, P_2)$, where P_1 and P_2 are defined in the system of Equations (2). Let

$$R_2^{EGC} = I(X; U_0, U_1, X+N_2) + I(U_1; X+N_2) - R(D_1).$$

It is easily seen that $(R(D_1), R_2^{EGC}, D_0, D_1, D_2) \in \overline{\mathcal{EGC}}$. Let $(\mathbf{X}, \mathbf{U}_0, \mathbf{U}_1, \mathbf{N}_1, \mathbf{N}_2)$ be $n-$vectors with iid entries distributed according to the law governing (X, U_0, U_1, N_1, N_2). From Inequality (5), we know that if a code achieves $(R(D_1) +$

$\epsilon, R_2, D_0, D_1, D_2)$ then necessarily

$$n\big(R(D_1) + R_2 + \epsilon\big) \geq I(\mathbf{X};\mathbf{X}_0) + I(\mathbf{X}_1;\mathbf{X}_2)$$

It thus follows that

$$\begin{aligned}
n\big(R_2^{EGC} - R_2 - \epsilon\big) &= n\big(I(X;U_0,U_1,X+N_2) + I(U_1;X+N_2)\big) - n\big(R(D_1) + R_2 + \epsilon\big) \\
&\leq n\big(I(X;U_0,U_1,X+N_2) + I(U_1;X+N_2)\big) - I(\mathbf{X};\mathbf{X}_0) + I(\mathbf{X}_1;\mathbf{X}_2) \\
&= \{I(\mathbf{X};\mathbf{U}_0,\mathbf{U}_1,\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X};\mathbf{X}_0)\} + \{I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_1;\mathbf{X}_2)\} \\
&\leq I(\mathbf{X};\mathbf{U}_1,\mathbf{X}+\mathbf{N}_2 \mid \mathbf{U}_0) + \{I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_1;\mathbf{X}_2)\}\,, \qquad (21)
\end{aligned}$$

where the latter inequality follows from the chain rule for mutual information, and noting that $I(\mathbf{X};\mathbf{U}_0) = nR(D_0) \leq I(\mathbf{X};\mathbf{X}_0)$ because of the choice of $\mathbf{U}_0$ and Shannon's Rate-Distortion converse. The first term on the right hand side can be upper bounded as

$$\begin{aligned}
I(\mathbf{X};\mathbf{U}_1,\mathbf{X}+\mathbf{N}_2 \mid \mathbf{U}_0) &\overset{(a)}{\leq} I(\mathbf{X};\mathbf{U}_1,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2 \mid \mathbf{U}_0) \\
&\overset{(b)}{=} I(\mathbf{X};\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2 \mid \mathbf{U}_0) + I(\mathbf{X};\mathbf{U}_1 \mid \mathbf{U}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) \\
&\overset{(c)}{\leq} \frac{n}{2}\log_2\left(1 + \frac{D_0}{P_1} + \frac{D_0}{P_2}\right) + I(\mathbf{X};\mathbf{U}_1 \mid \mathbf{U}_0,\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) \\
&\overset{(d)}{\leq} \frac{n}{2}\log_2\left(1 + \frac{D_0}{P_1} + \frac{D_0}{P_2}\right) + I(\mathbf{X};\mathbf{U}_1 \mid \mathbf{X}+\mathbf{N}_1) \\
&\overset{(e)}{\leq} \frac{n}{2}\log_2\left(1 + \frac{D_0}{P_1} + \frac{D_0}{P_2}\right) + I(\mathbf{X};\mathbf{X}+\mathbf{N}_1 \mid \mathbf{U}_1) \\
&\overset{(f)}{\leq} \frac{n}{2}\log_2\left(1 + \frac{D_0}{P_1} + \frac{D_0}{P_2}\right) + \frac{n}{2}\log_2(2 - D_1), \qquad (22)
\end{aligned}$$

where Inequality (a) follows from the chain rule applied to the right hand side and the non-negativity of the mutual information; Equality (b) follows from the chain rule; Inequality (c) follows from Inequality (16); Inequality (d) follows from Lemma 2 where we have substituted $\mathbf{X}$, $\mathbf{U}_1$, $\mathbf{X}+\mathbf{N_1}$, and $(\mathbf{U}_0,\mathbf{X}+\mathbf{N}_2)$ for D, A, B, and C, respectively; Inequality (e) is obtained by expanding $I(X;U_1,X+N_1)$ in two ways with the chain rule and noting that $R(D_1) = I(X;U_1) \leq I(X;X+N_1)$; and Inequality (f) is an application of Lemma 1.

To bound the second term on the right-hand side of Inequality (21) we consider separately the two cases $D_1 \leq D_2$ and $D_1 > D_2$. If $D_1 \leq D_2$ we add and subtract $I(\mathbf{X}_2;\mathbf{X}+\mathbf{N}_2)$ to obtain

$$\begin{aligned}
&I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_1;\mathbf{X}_2) \\
&\quad = \{I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_2;\mathbf{X}+\mathbf{N}_2)\} + \{I(\mathbf{X}_2;\mathbf{X}+\mathbf{N}_2) - I(\mathbf{X}_1;\mathbf{X}_2)\}\,.
\end{aligned}$$

We bound the first brace by

$$\begin{aligned}
I(\mathbf{X}+\mathbf{N}_2;\mathbf{U}_1) - I(\mathbf{X}+\mathbf{N}_2;\mathbf{X}_2) &\overset{(a)}{\leq} I(\mathbf{X}+\mathbf{N}_2;\mathbf{U}_1 \mid \mathbf{X}_2) \\
&\overset{(b)}{\leq} I(\mathbf{X}+\mathbf{N}_2;\mathbf{X} \mid \mathbf{X}_2) \\
&\overset{(c)}{\leq} \frac{n}{2}\log_2(2 - D_2)\,,
\end{aligned}$$

where (a) is obtained by expanding $I(\mathbf{X}+\mathbf{N}_2;\mathbf{U}_1,\mathbf{X}_2)$ in two ways using the chain rule and by invoking the non-negativity of the mutual information; (b) is similarly obtained by expanding in two ways $I(\mathbf{X}+\mathbf{N}_2;\mathbf{U}_1,\mathbf{X}\mid\mathbf{X}_2)$ and noting that $I(\mathbf{X}+\mathbf{N}_2;\mathbf{U}_1\mid\mathbf{X},\mathbf{X}_2)=0$; and (c) follows from Inequality (15) and symmetry considerations.

As for the second brace,

$$\begin{aligned}
I(\mathbf{X}_2;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}_2;\mathbf{X}_1) &\overset{(a)}{\leq} I(\mathbf{X}_2;\mathbf{X}+\mathbf{N}_2\mid\mathbf{X}_1)\\
&\overset{(b)}{\leq} I(\mathbf{X};\mathbf{X}+\mathbf{N}_2\mid\mathbf{X}_1)\\
&\overset{(c)}{\leq} \frac{n}{2}\log_2\left(1+\frac{D_1}{D_2}-D_1\right) &(23)\\
&\overset{(d)}{\leq} \frac{n}{2}\log_2\left(2-D_1\right), &(24)
\end{aligned}$$

where the derivations of Inequalities (a) and (b) are analogous to those of the first brace, Inequality (c) follows from Lemma 1 by substituting Δ with D_1 and σ_1^2 with $D_2/(1-D_2)$, and Inequality (d) follows from the assumption that $D_1\leq D_2$.

If $D_2<D_1$ we write

$$\begin{aligned}
&I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}_1;\mathbf{X}_2)\\
&\overset{(a)}{\leq} I(\mathbf{U}_1;\mathbf{X})-I(\mathbf{X}_1;\mathbf{X}_2)\\
&\overset{(b)}{\leq} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X})-I(\mathbf{X}_1;\mathbf{X}_2)\\
&\overset{(c)}{=} \{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X})-I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1)\}+\{I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1)-I(\mathbf{X}_1;\mathbf{X}_2)\},
\end{aligned}$$

where (a) follows from the data processing inequality and (b) follows from the assumption that $I(\mathbf{U}_1;\mathbf{X})=R(D_1)$, the assumption that $E[d(\mathbf{U}_1,\mathbf{X})]=E[d(\mathbf{X}+\mathbf{N}_1,\mathbf{X})]=D_1$, and Shannon's converse to the Rate-Distortion Theorem.

For the first brace on the right hand side of (c), we obtain the following bound,

$$\begin{aligned}
I(\mathbf{X}+\mathbf{N}_1;\mathbf{X})-I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1) &\overset{(a)}{\leq} I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}\mid\mathbf{X}_1)\\
&\overset{(b)}{\leq} \frac{n}{2}\log_2(2-D_1),
\end{aligned}$$

where, as usual, (a) is obtained by applying the chain rule in two different ways to $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X},\mathbf{X}_1)$ and invoking the non-negativity of $I(\mathbf{X}+\mathbf{N}_1;\mathbf{X}_1\mid\mathbf{X})$, and (b) is Inequality (15).

The second brace can be upper bounded with a calculation formally identical to the one leading to Inequality (24), giving:

$$I(\mathbf{X}_1;\mathbf{X}+\mathbf{N}_1)-I(\mathbf{X}_1;\mathbf{X}_2)\leq\frac{n}{2}\log_2(2-D_2).$$

In summary, regardless of the relative values of D_1 and D_2,

$$I(\mathbf{U}_1;\mathbf{X}+\mathbf{N}_2)-I(\mathbf{X}_1;\mathbf{X}_2)\leq\frac{n}{2}\log_2\left((2-D_1)(2-D_2)\right).$$

This remark, in conjunction with (22) and (21) completes the proof of the theorem.

V. Conclusions

We have shown that, although not tight, the El Gamal and Cover achievable region for two descriptions is universally nearly tight. Mathematically, our contribution is a new single-letter outer bound for the achievable rates given prescribed distortion levels. The outer bound has been deliberately crafted to match a particular inner bound that can be found through the El Gamal and Cover coding theorem and a forward test channel first introduced by Frank-Dayan and Zamir. We succeed in demonstrating that the gaps between the outer and inner bounds on the individual rates and the sum rate are never greater than 0.5 and 1.5 bits/sample, respectively. Thus there is a sense in which the two descriptions problem is nearly solved.

An intriguing possibility is that known achievable regions for other unsolved multi-terminal information theory problems may in fact possess the type of competitive optimality that the EGC theorem exhibits. We expect to address these issues in the future and also expect that the general methodology described in this work will have wide applicability in these investigations.

VI. Appendix

A. Proof of Lemma 1

Since subscripts are currently being employed to differentiate among vectors of random variables, we shall use the unconventional notation that upper-scripts denote elements within a vector. For example, $\mathbf{G}_3^i$ denotes the ith element of the random vector $\mathbf{G}_3$, $\mathbf{G}_3^{-i}$ denotes all such elements but the ith one, and $\mathbf{G}_3^{1,\cdots,i}$ denotes elements 1 through i.

We first extend to vector-valued random variables a sequence of inequalities whose version for scalars appears in Zamir [11]. We write

$$\begin{aligned}
& I(\mathbf{A};\mathbf{A}+\mathbf{G}_1,\cdots,\mathbf{A}+\mathbf{G}_k|\mathbf{B}) \\
&= I(\mathbf{A}-\mathbf{B};\mathbf{A}-\mathbf{B}+\mathbf{G}_1,\cdots,\mathbf{A}-\mathbf{B}+\mathbf{G}_k|\mathbf{B}) \\
&\overset{(a)}{\leq} I(\mathbf{A}-\mathbf{B};\mathbf{A}-\mathbf{B}+\mathbf{G}_1,\cdots,\mathbf{A}-\mathbf{B}+\mathbf{G}_k) \\
&\overset{(b)}{=} I(\mathbf{W};\mathbf{W}+\mathbf{G}_1,\cdots,\mathbf{W}+\mathbf{G}_k) \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(\mathbf{W};\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i|\mathbf{W}^{1,\cdots,i-1}+\mathbf{G}_1^{1,\cdots,i-1},\cdots,\mathbf{W}^{1,\cdots,i-1}+\mathbf{G}_k^{1,\cdots,i-1}) \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} I(\mathbf{W};\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i) \\
&\overset{(e)}{=} \sum_{i=1}^{n} I(\mathbf{W}^i;\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i)+I(\mathbf{W}^{-i};\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i|\mathbf{W}^i) \\
&\overset{(f)}{=} \sum_{i=1}^{n} I(\mathbf{W}^i;\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i),
\end{aligned} \tag{25}$$

where (a) follows from the chain $\mathbf{B}\to(\mathbf{A}-\mathbf{B})\to(\mathbf{A}-\mathbf{B}+\mathbf{G}_1,\cdots,\mathbf{A}-\mathbf{B}+\mathbf{G}_k)$, in (b) we have defined $\mathbf{W}\overset{\triangle}{=}\mathbf{A}-\mathbf{B}$ (c) follows from the chain rule for mutual information, (d) follows from the fact that for every $i\in\{1,\cdots,n\}$,

$$(\mathbf{W}^{1,\cdots,i-1}+\mathbf{G}_1^{1,\cdots,i-1},\cdots,\mathbf{W}^{1,\cdots,i-1}+\mathbf{G}_k^{1,\cdots,i-1})\to\mathbf{W}\to(\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i)$$

Step (e) follows from the chain rule for mutual information and (f) follows from $\mathbf{W}^{-i}\to\mathbf{W}^i\to(\mathbf{W}^i+\mathbf{G}_1^i,\cdots,\mathbf{W}^i+\mathbf{G}_k^i)$. We now argue that each of the n mutual information expressions in (25) can be written as a difference of finite-valued differential entropies. Define $r_i\overset{\triangle}{=}E(\mathbf{W}^i-E\mathbf{W}^i)^2$. Note that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} r_i &= \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{W}^i-E\mathbf{W}^i)^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{W}^i)^2
\end{aligned}$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{A}^i - \mathbf{B}^i)^2\right] \leq \Delta < +\infty, \tag{26}$$

by the lemma assumption, and thus in particular $r_i < +\infty$ for all $i \in \{1, \cdots, n\}$. Because conditioning never increases differential entropy,

$$\begin{aligned} h(\mathbf{W}^i + \mathbf{G}_1^i, \cdots, \mathbf{W}^i + \mathbf{G}_k^i) &\geq h(\mathbf{W}^i + \mathbf{G}_1^i, \cdots, \mathbf{W}^i + \mathbf{G}_k^i | \mathbf{W}^i) \\ &= h(\mathbf{G}_1^i, \cdots, \mathbf{G}_k^i) \\ &> -\infty, \end{aligned} \tag{27}$$

where the last inequality follows from the assumption of positivity of the variances $\sigma_1^2, \cdots, \sigma_k^2$. For every $i \in \{1, \cdots, n\}$, the covariance matrix of the random k-vector in the left of Inequality (27) is

$$\Sigma_i = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_k^2) + r_i 11^T. \tag{28}$$

In the above, the symbol 11^T represents a square matrix with every entry equal to one. The size of this matrix is chosen so as to be appropriate for the arithmetic operation in context. [3]

It is a basic information theoretic fact that the entropy of a vector that possesses a given covariance matrix is never greater than the entropy of a Gaussian vector with the same auto-covariance matrix. Therefore

$$h(\mathbf{W}^i + \mathbf{G}_1^i, \cdots, \mathbf{W}^i + \mathbf{G}_k^i) \leq \frac{1}{2}\log_2\left[(2\pi e)^k \det \Sigma_i\right] < +\infty. \tag{29}$$

It then follows that expansion of the mutual information as a difference of differential entropies is well defined for every $i \in \{1, \cdots, n\}$, yielding the upper bound

$$I(\mathbf{W}^i; \mathbf{W}^i + \mathbf{G}_1^i, \cdots, \mathbf{W}^i + \mathbf{G}_k^i) \leq \frac{1}{2}\log_2\left[(2\pi e)^k \det \Sigma_i\right] - \frac{1}{2}\log_2\left[(2\pi e)^k \Pi_{j=1}^k \sigma_j^2\right]. \tag{30}$$

To evaluate the upper bound, we shall use the following lemma:

Lemma 3: For Σ_i as defined in Equation (28), with positive $\sigma_1, \cdots, \sigma_k$ and nonnegative r_i,

$$\det \Sigma_i = \left(\Pi_{j=1}^k \sigma_j^2\right)\left(1 + \sum_{j=1}^{k}\frac{r_i}{\sigma_j^2}\right).$$

The finishing arguments then are as follows:

$$\begin{aligned} n^{-1} I(\mathbf{W}; \mathbf{W} + \mathbf{G}_1, \cdots, \mathbf{W} + \mathbf{G}_k) &\leq \frac{1}{2}\sum_{i=1}^{n}\frac{1}{n}\log_2\left(1 + \sum_{j=1}^{k}\frac{r_i}{\sigma_j^2}\right) \\ &\overset{(a)}{\leq} \frac{1}{2}\log_2\left(1 + \sum_{j=1}^{k}\frac{\frac{1}{n}\sum_{i=1}^{n} r_i}{\sigma_j^2}\right) \\ &\overset{(b)}{\leq} \frac{1}{2}\log_2\left(1 + \sum_{j=1}^{k}\frac{\Delta}{\sigma_j^2}\right), \end{aligned}$$

where step (a) follows from Jensen's inequality and step (b) follows from Inequality (26). This concludes the proof of the lemma. □

[3]More specifically, the symbol "1" in the notation refers to a column vector with all entries equal to one; "1^T" is the transpose of this vector and "11^T" is the result of the matrix multiplication of these vectors.

B. Proof of Lemma 3

Since the subindex in r_i is irrelevant to this lemma, in the subsequent discussion we drop it and simply refer to r. Since the case when $r = 0$ is trivial, we assume that $r > 0$. We claim that for every $1 \leq j < k$,

$$\begin{aligned} & \det\left(\mathrm{diag}(\sigma_1^2, \cdots, \sigma_k^2) + r11^T\right) \\ = & \left(\Pi_{l=1}^{j}\left(\sigma_l^2 + \gamma_l\right)\right) \det\left(\mathrm{diag}(\sigma_{j+1}^2, \cdots, \sigma_k^2) + \gamma_{j+1}11^T\right), \end{aligned} \tag{31}$$

where for $1 \leq m \leq k+1$, we have defined

$$\gamma_m \triangleq \left(\frac{1}{r} + \sum_{l=1}^{m-1} \frac{1}{\sigma_l^2}\right)^{-1}.$$

Substituting $j = k - 1$ in Equation (31) we obtain

$$\begin{aligned} \det\left(\mathrm{diag}(\sigma_1^2, \cdots, \sigma_k^2) + r11^T\right) &= \Pi_{l=1}^{k}\left(\sigma_l^2 + \gamma_l\right) \\ &= r\gamma_1^{-1}\Pi_{l=1}^{k}\left(\sigma_l^2 + \gamma_l\right). \end{aligned} \tag{32}$$

Note that for all $l \in \{1, \cdots, k\}$,

$$\gamma_l^{-1}(\sigma_l^2 + \gamma_l) = \sigma_l^2 \gamma_{l+1}^{-1}.$$

and therefore the expression in Equation (32) is equal to

$$r\left(\Pi_{l=1}^{k}\sigma_l^2\right)\gamma_{k+1}^{-1} = \left(\Pi_{l=1}^{k}\sigma_l^2\right)\left(1 + \sum_{i=1}^{k}\frac{r}{\sigma_l^2}\right),$$

which proves the lemma. It remains to argue that Equation (31) is correct. This can be seen as follows: for any m positive real numbers $c_1, \cdots, c_m$ and positive d,

$$\begin{aligned} & \det\left(\mathrm{diag}(c_1, \cdots, c_m) + d11^T\right) \\ = & \det\left(\left[\begin{array}{c|c} c_1 + d & d1^T \\ \hline d1 & \mathrm{diag}\left(c_2, \cdots, c_m\right) + d11^T \end{array}\right]\right) \\ \overset{(a)}{=} & (c_1 + d)\det\left(\mathrm{diag}(c_2, \cdots, c_m) + \left(\frac{1}{d} + \frac{1}{c_1}\right)^{-1} 11^T\right), \end{aligned} \tag{33}$$

where step (a) follows from the relation

$$\det\left(\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\right) = \det(A)\det(D - CA^{-1}B),$$

which can be found in [25], Chapter 10.1, Equation 5. The proof now proceeds by induction. Equation (31) with $j = 1$ can be seen to be true by making the substitutions

$$\begin{aligned} m &\to k \\ c_l &\to \sigma_l^2 \quad 1 \leq l \leq m \\ d &\to \gamma_1 \end{aligned}$$

in (33). Finally, assuming (31) is correct for some $j \leq k-2$, and making the substitutions

$$\begin{aligned} m &\rightarrow k-j \\ c_l &\rightarrow \sigma_{j+l}^2 \quad 1 \leq l \leq m \\ d &\rightarrow \gamma_{j+1} \end{aligned}$$

in Equation (33), we obtain that the statement in (31) for j implies the same statement for $j+1$. The lemma is now proved.

□

C. Proof of Lemma 2

Note that

$$\begin{aligned} I(B,C,D;A) &\stackrel{(a)}{=} I(C,B;A) + I(D;A \mid C,B) \\ &\stackrel{(b)}{=} I(D;A) + I(C,B;A \mid D) \\ &\stackrel{(c)}{=} I(D;A), \end{aligned}$$

where both (a) and (b) are consequences of two different applications of the chain rule to $I(B,C,D;A)$ and (c) follows from the assumption that $(CB) \rightarrow D \rightarrow A$. As a consequence,

$$\begin{aligned} I(D;A|CB) &= I(D;A) - I(CB;A) \\ &\stackrel{(a)}{=} I(D;A) - I(B;A) - I(C;A|B) \\ &\stackrel{(b)}{\leq} I(D;A) - I(B;A) \\ &\stackrel{(c)}{=} I(D;A|B), \end{aligned}$$

where (a) follows from the chain rule, (b) follows from the non-negativity of mutual information and (c) follows from the chain rule applied twice to $I(BD;A)$ and the fact that $B \rightarrow D \rightarrow A$.

VII. Acknowledgments

The authors greatly appreciate Toby Berger's suggestion that "nearly-tight" outer bounds for unsolved multi-terminal information theory problems were both important and feasible. The authors also thank Hanying Feng and Ram Zamir for numerous exchanges in which their contributions to the two descriptions problem were discussed.

References

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

[2] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Conv. Rec.*, vol. 7, pp. 142–163, 1959.

[3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, 1976.

[5] J. Kieffer, "A survey of the theory of source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 5, pp. 1473–1490, 1993.

[6] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, 1980.

[7] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 721–726, 1985.

[8] A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, 1982.

[9] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 502–521, 1987.

[10] H. Feng and M. Effros, “On the rate loss of multiple description source codes,” *IEEE Trans. Inform. Theory*, vol. IT-51, pp. 671–683, February 2005.

[11] R. Zamir, “The rate loss in the Wyner-Ziv problem,” *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 2073–2084, 1996.

[12] L. Lastras and T. Berger, “All sources are nearly successively refinable,” *IEEE Trans. Inform. Theory*, vol. IT-47, pp. 918–926, 2001.

[13] H. Feng and M. Effros, “Improved bounds for the rate loss of multiresolution source codes,” *IEEE Trans. Inform. Theory*, vol. IT-49, no. 4, pp. 809–820, 2003.

[14] L. Lastras-Montaño and V. Castelli, “Near tightness of the El Gamal and Cover region for two descriptions,” in *Data Compression Conference, Snowbird, Utah*, 2005, pp. 3–12.

[15] R. Zamir, “Gaussian codes and shannon bounds for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-45, no. 7, pp. 2629–2636, 1999.

[16] Y. Frank-Dayan and R. Zamir, “Dithered lattice-based quantizers for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-48, no. 1, pp. 192–204, 2002.

[17] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, 1997.

[18] H. Feng and M. Effros, “On the achievable region for multiple description source codes on Gaussian sources,” in *International Symposium on Information Theory, Yokohama, Japan*, 2003, p. 195.

[19] Z. Zhang and T. Berger, “Multiple description source coding with no excess marginal rate,” *IEEE Trans. Inform. Theory*, vol. IT-41, no. 2, pp. 349–357, Mar. 1995.

[20] J. Ziv, “On universal quantization,” *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344–347, 1985.

[21] R. Zamir and M. Feder, “Information rates of pre/post-filtered dithered quantizers,” *IEEE Trans. Inform. Theory*, vol. IT-42, no. 5, pp. 1340–1353, 1996.

[22] H. Feng, “Rate loss of network source codes,” Ph.D. dissertation, California Institute of Technology, 2002.

[23] J. Chen, C. Tian, T. Berger, and S. Hemami, “Multiple description quantization via gram-schmidt orthogonalization,” *Submitted to the IEEE Transactions on Information Theory*.

[24] A. Kashyap, L. A. L.-M. no, C. H. Xia, and Z. Liu, “Distributed source coding in dense sensor networks,” in *Proceedings of the Data Compression Conference, Snowbird, Utah*, 2005, pp. 13–22.

[25] S. R. Searle, *Matrix Algebra Useful For Statistics*, ser. Wiley Series in probability and mathematical statistics. John Wiley & Sons, 1982.