Multiple Description Coding With Many Channels

Raman Venkataramani, *Member, IEEE*, Gerhard Kramer, *Member, IEEE*, and Vivek K Goyal, *Senior Member, IEEE*

***Abstract*—An achievable region for the L-channel multiple description coding problem is presented. This region generalizes two-channel results of El Gamal and Cover and of Zhang and Berger. It further generalizes three-channel results of Gray and Wyner and of Zhang and Berger. A source that is successively refinable on chains is shown to be successively refinable on trees. A new outer bound on the rate-distortion (RD) region for memoryless Gaussian sources with mean squared error distortion is also derived. The achievable region meets this outer bound for certain symmetric cases.**

***Index Terms*—Multiple descriptions, rate-distortion (RD) theory, source coding.**

I. Introduction

Multiple description (MD) coding arose in connection with communicating speech over the telephone network. The idea was to split the information from a call into two parts that are sent on two separate links or paths. Normally both parts are received and are combined to achieve the usual voice quality. However, an outage of one link or the other can now be accommodated by reducing the voice quality. This idea of *channel splitting* inspired the following question: Given an information source and a number of channels that can fail, what are the concurrent limitations on the data rates and transmission qualities?

This question was formalized by A. D. Wyner in 1979 and became known as the *multiple description problem*. The two-channel problem is as follows. An encoder is given a sequence $X^N = X^{(1)}, X^{(2)}, \ldots, X^{(N)}$ and maps it into two descriptions J_1 and J_2 having B_1 and B_2 bits, respectively. J_1 and J_2 are sent over the respective first and second channel, and each description either arrives error free at the receiver or is lost. The receiver uses one of three decoders. The *central decoder* is used if both J_1 and J_2 are received and it produces an estimate X_0^N of X^N. One of two *side decoders* is used if one of J_1 or J_2 is received and it produces the estimate X_1^N or X_2^N. The rates of the descriptions are denoted $R_i = B_i/N$ bits per source symbol, $i = 1, 2$, and the distortions attained by these reproductions are denoted D_i, $i = 0, 1, 2$. Of course, if neither description is received the receiver can only guess at what the source sequence is.

Manuscript received September 1, 2002; revised April 4, 2003.

R. Venkataramani was with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA. He is now with Harvard University, Cambridge, MA 02140 USA (e-mail: ramanv@ieee.org).

G. Kramer is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: gkr@bell-labs.com).

V. K Goyal was with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA. He is now with the University of California, Berkeley, CA 94720 USA (e-mail: v.goyal@ieee.org).

Communicated by R. Zamir, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2003.815767

A. Two Examples

The two most commonly studied MD problems are the binary-symmetric source (BSS) with Hamming distortion and the Gaussian source with squared-error distortion. Consider first the BSS, for which the $X^{(n)}$ are independent and identically distributed (i.i.d.) binary random variables taking on the values 0 and 1 with probability $1/2$. The average Hamming distortion between a source sequence x^N and its reproduction x_i^N is

$$d_H^N\left(x^N, x_i^N\right) = \frac{1}{N}\sum_{n=1}^{N} d_H\left(x^{(n)}, x_i^{(n)}\right)$$

where $d_H(x, x_i) = 0$ if $x = x_i$ and $d_H(x, x_i) = 1$ if $x \neq x_i$. Suppose that we require

$$D_0 = \mathrm{E}[d_H^N(X^N, X_0^N)] \to 0, \qquad \text{as } N \to \infty$$

i.e., the central decoder should reproduce X^N with vanishingly small error probability for large N. The channel splitting approach is to transmit the even-numbered bits across the first channel and the odd-numbered bits across the second channel. Thus, the rates are $R_1 = R_2 = 0.5$ bits per symbol and the average distortions are $D_0 = 0$, $D_1 = \mathrm{E}[d_H^N(X^N, X_1^N)] = 0.25$, and $D_2 = \mathrm{E}[d_H^N(X^N, X_2^N)] = 0.25$. The last two distortions are achieved by simply guessing at those bits which one does not know. However, one can do better than channel splitting—more sophisticated codes can achieve

$$D_1 = D_2 = (\sqrt{2} - 1)/2 \approx 0.207.$$

Next, the Gaussian source has the $X^{(n)}$ as i.i.d. Gaussian random variables with zero mean and unit variance. The mean squared error distortion between x^N and x_i^N is

$$d_G^N\left(x^N, x_i^N\right) = \frac{1}{N}\sum_{n=1}^{N}\left(x^{(n)} - x_i^{(n)}\right)^2.$$

For example, to attain a central distortion of D_0, $0 < D_0 \leq 1$, one can encode X^N at a rate as close as desired to Shannon's rate-distortion (RD) function $R(D_0) = -\log_2(D_0)/2$ bits per symbol [1, Sec. 28], [2] for sufficiently large N. However, if small side distortions D_1 and D_2 are desired then $R_1 + R_2 > R(D_0)$ might be required.

B. Historical Summary

Perhaps the earliest information-theoretic treatment of an MD problem can be found in [3], [4]. The first general result was El Gamal and Cover's achievable region for two channels [5]. Ozarow proved this region to be optimal for the Gaussian source and mean squared error distortion [6]. This result was extended to create high-rate bounds for other memoryless sources in [7] and for stationary Gaussian processes in [8]. Various bounds for

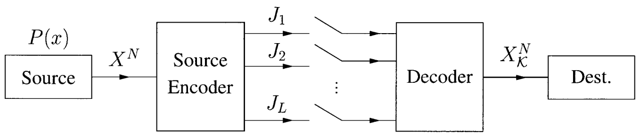

Fig. 1. The L-channel MD model.

the BSS with Hamming distortion were developed in [9]–[11]. An achievable region for the BSS with many channels was derived in [12]. Some specialized but final results for the BSS can be found in [13]–[15]. The important two-channel problem with "no excess rate" was solved in [16]. Our main results first appeared in [17].

A special case of the MD problem is where one considers only the distortions D_0 and D_1; this is known as the *successive-refinement* problem. This problem was solved for two channels by Gray [3]. Further results can be found in [18]–[25]. Another special case of the MD problem is the *symmetric* case where all the rates are equal and the distortion depends on the number of descriptions received but not on which particular descriptions are received. An achievable rate region for this problem has been recently determined by Pradhan, Puri, and Ramchandran [26], [27].

Motivated largely by the analogy between uses of a channel and sending a packet on a data network that loses packets but makes no errors within received packets, the construction of practical MD codes has been an active research area. Widespread interest followed the publication of [28], though techniques had been developed at Bell Labs in the late 1970s and early 1980s. For more on the history of MD coding and a survey of practical techniques and applications, see [29].

This paper is organized as follows. Section II presents our main results, which are an achievable rate region for the L-channel MD problem (Theorem 1) and an outer bound for the Gaussian source with mean squared error distortion (Theorem 2). Several prior results are shown to be subsumed by our results in Section III. In Section IV, we introduce and solve a successive-refinement problem for tree structures. In Section V, we show that the inner bound given by Theorem 1 meets the outer bound given by Theorem 2 for a restricted class of MD problems. Finally, Section VI concludes the paper.

II. Problem and Main Results

The L-channel MD problem is depicted in Fig. 1. A source emits a sequence $X^N = X^{(1)}, X^{(2)}, \ldots, X^{(N)}$ of i.i.d. random variables taking on values in the finite alphabet $\mathcal{X}$. Let $\mathcal{L} = \{1, \ldots, L\}$. There are L encoding functions $f_l(\cdot)$, $l \in \mathcal{L}$, that map X^N to the *descriptions* $J_l = f_l(X^N)$, where J_l is a sequence having B_l bits or $\log_e(2)\, B_l$ nats. The rate of description J_l is thus $R_l = \log_e(2)\, B_l/N$ nats per symbol.

Description J_l is transmitted over channel l and is either received error-free or lost completely. The receiver thus encounters one of 2^L configurations depending on which J_l are received. Excepting the trivial case, we can represent the receiver as a collection of $2^L - 1$ decoding functions $g_{\mathcal{K}}(\cdot)$, $\mathcal{K} \subseteq \mathcal{L}$, $\mathcal{K} \neq \emptyset$. The decoder whose inputs are $\{J_k: k \in \mathcal{K}\}$ reproduces X^N as

$$g_{\mathcal{K}}(J_k: k \in \mathcal{K}) = X_{\mathcal{K}}^N = X_{\mathcal{K}}^{(1)}, \ldots, X_{\mathcal{K}}^{(N)}.$$

The letters $X_{\mathcal{K}}^{(n)}$, $n = 1, \ldots, N$, take on values in the finite reproduction alphabet $\mathcal{X}_{\mathcal{K}}$.

The receiver usually cannot reconstruct X^N perfectly. We consider the case where its distortion is measured in a *per letter* fashion via

$$D_{\mathcal{K}} = \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N} d_{\mathcal{K}}\left(X^{(n)}, X_{\mathcal{K}}^{(n)}\right)\right] \tag{1}$$

where $d_{\mathcal{K}}(\cdot, \cdot)$ is a letter distortion function mapping pairs $(x, x_{\mathcal{K}}) \in \mathcal{X} \times \mathcal{X}_{\mathcal{K}}$ into the nonnegative reals, and where $\mathrm{E}[\cdot]$ denotes expectation. Observe that $d_{\mathcal{K}}(\cdot, \cdot)$ can depend on $\mathcal{K}$. We assume that $d_{\mathcal{K}}(\cdot, \cdot)$ is upper-bounded by $d_{\max}$ for our proofs. However, when considering the Gaussian source with mean squared error distortion we make the usual step of dropping the restrictions that $d_{\mathcal{K}}(\cdot, \cdot)$ is bounded and that the alphabets $\mathcal{X}_{\mathcal{K}}$ are finite (see, e.g., [5]).

We say that the rates $\{R_l: l \in \mathcal{L}\}$ and distortions $\{D_{\mathcal{K}}: \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset\}$ are *achievable* if there are encoding functions $f_l(\cdot)$ having rates R_l, $l \in \mathcal{L}$, and decoding functions $g_{\mathcal{K}}(\cdot)$ giving distortions $D_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset$. Our problem is to find the closure of the region of achievable rates and distortions for all N, called the *RD region*. This region has $[L + (2^L - 1)]$ dimensions.

A. An Achievable Region for Discrete Sources

Our first result is an inner bound to the RD region. Let $2^{\mathcal{L}}$ be the power set of $\mathcal{L}$, i.e., the collection of all subsets of $\mathcal{L}$. The set difference between collections of sets $\mathcal{C}$ and $\mathcal{D}$ is denoted as $\mathcal{C} - \mathcal{D} = \{\mathcal{M} \in \mathcal{C}: \mathcal{M} \notin \mathcal{D}\}$. We write $R_{\mathcal{K}}$ as a shorthand for $\sum_{k \in \mathcal{K}} R_k$ and $X_{(\mathcal{C})}$ for $\{X_{\mathcal{N}}: \mathcal{N} \in \mathcal{C}\}$. We interpret $X_{(\{\emptyset\} - \{\emptyset\})} = X_{(\emptyset)}$ as a constant and $|\mathcal{K}|$ as the cardinality of $\mathcal{K}$. Finally, for random variables X, Y, and Z we use the common notation $H(X)$, $H(X|Y)$, $I(X; Y)$, and $I(X; Y|Z)$ for entropies and mutual informations [30, Ch. 2].

Theorem 1: Let $X_{(2^{\mathcal{L}})}$ be any set of 2^L random variables jointly distributed with X, where $X_{\emptyset}$ takes on values in some finite alphabet $\mathcal{X}_{\emptyset}$ and each $X_{\mathcal{K}}$ takes on values in the reproduction alphabet $\mathcal{X}_{\mathcal{K}}$, $\mathcal{K} \neq \emptyset$. Then the RD region contains the rates and distortions satisfying

$$D_{\mathcal{K}} \geq \mathrm{E}[d_{\mathcal{K}}(X, X_{\mathcal{K}})] \tag{2}$$

$$R_{\mathcal{K}} \geq (|\mathcal{K}| - 1) I(X; X_{\emptyset}) - H\left(X_{(2^{\mathcal{K}})} \middle| X\right) + \sum_{\mathcal{M} \subseteq \mathcal{K}} H\left(X_{\mathcal{M}} \middle| X_{(2^{\mathcal{M}} - \{\mathcal{M}\})}\right) \tag{3}$$

for every $\mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}$.

Proof: See Appendix A. □

Remark 1: The usual time-sharing arguments can be applied to Theorem 1, e.g., the convex hull of the region defined by (2) and (3) is also in the RD region.

Remark 2: One can generalize Theorem 1 by replacing the $X_{\mathcal{K}}$ in (3) with artificial reproduction random variables $\hat{X}_{\mathcal{K}}$ that take on values in some finite alphabets $\hat{\mathcal{X}}_{\mathcal{K}}$. The $X_{\mathcal{K}}$ in (2) are recovered by using $X_{\mathcal{K}} = \psi_{\mathcal{K}}(\hat{X}_{\mathcal{K}})$ for some functions $\psi_{\mathcal{K}}(\cdot)$. Theorem 1 is, thus, the special case of this result if one chooses the $\psi_{\mathcal{K}}(\cdot)$ to be identity mappings. However, we do not know if this generalization improves Theorem 1.

Remark 3: For $L = 1$, Theorem 1 with $X_\emptyset$ a constant gives the set of (R_1, D_1) corresponding to Shannon's RD function.

Remark 4: For $L = 2$, Theorem 1 generalizes a result of El Gamal and Cover [5, Theorem 1] and of Zhang and Berger [15, Theorem 1]. We prove these claims in Section III-A.

Remark 5: For $L = 3$, Theorem 1 generalizes a result of Zhang and Berger [15, Theorem 1*] and of Gray and Wyner [4, direct half of Theorem 8]. We prove these claims in Section III-B.

Remark 6: Theorem 1 holds more generally for well-behaved continuous sources and distortion functions if the entropies $H(\cdot)$ are replaced by differential entropies $h(\cdot)$. We will simply assume that Theorem 1 is valid and apply it to the Gaussian source with mean squared error distortion. A formal proof justifying this step might proceed along the lines of [31] that generalizes the results of [32] from discrete to nondiscrete sources. We expect that such a proof will lean heavily on the proof in Appendix A, just as the proof in [31] is based on the proof in [32].

B. An Achievable Region for the Gaussian Source

Consider the Gaussian source with mean squared error distortion. We apply Theorem 1 and Remark 6 with the following choice of $P(x, x_{(2^{\mathcal{L}})})$ (see also [5, Sec. IV] and [6, Sec. III]). Let $W_1, \ldots, W_L$ be zero-mean, Gaussian random variables independent of X and having covariance matrix $\boldsymbol{Q}$, i.e., the i, j entry of $\boldsymbol{Q}$ is $\mathrm{E}[W_i W_j]$. Let $U_l = X + W_l$ and $X_{\mathcal{K}} = \mathrm{E}[X|U_{(\mathcal{K})}]$. We use results from least squares estimation [33, p. 237] to obtain

$$X_{\mathcal{K}} = \mathrm{E}\left[X\,\underline{U}_{\mathcal{K}}^T\right] \cdot \mathrm{E}\left[\underline{U}_{\mathcal{K}}\underline{U}_{\mathcal{K}}^T\right]^{-1} \underline{U}_{\mathcal{K}} = \underline{e}^T(\boldsymbol{J} + \boldsymbol{Q}_{\mathcal{K}})^{-1}\underline{U}_{\mathcal{K}} \tag{4}$$

where $\underline{U}_{\mathcal{K}}$ is the column vector of the U_k, $k \in \mathcal{K}$, T denotes transposition, $\underline{e}$ is an all-ones column vector, $\boldsymbol{J}$ is an all-ones matrix, and $\boldsymbol{Q}_{\mathcal{K}}$ is the covariance matrix for the W_k, $k \in \mathcal{K}$. We also set $X_\emptyset = 0$.

Suppose we insert (4) into (3). We then run into the difficulty that $X_{\mathcal{K}}$ is a function of the X_k, $k \in \mathcal{K}$, so that

$$h(X_{\mathcal{K}}|X_{(2^{\mathcal{K}} - \{\mathcal{K}\})}) = -\infty, \qquad \text{for } |\mathcal{K}| > 1.$$

One can remedy this by canceling such entropies on the right-hand side of (3) before evaluating them. The result is that (3) becomes

$$R_{\mathcal{K}} \geq -h\left(X_{(\mathcal{K})}\middle| X\right) + \sum_{k\in\mathcal{K}} h(X_k). \tag{5}$$

Evaluating both (2) and (5) using (4), we have

$$D_{\mathcal{K}} \geq \frac{\det \boldsymbol{Q}_{\mathcal{K}}}{\det(\boldsymbol{J} + \boldsymbol{Q}_{\mathcal{K}})} \tag{6}$$

$$R_{\mathcal{K}} \geq -\frac{1}{2}\log\left(\frac{\det \boldsymbol{Q}_{\mathcal{K}}}{\prod_{k\in\mathcal{K}}(1 + \mathrm{E}[W_k^2])}\right) \tag{7}$$

where $\det \boldsymbol{Q}$ denotes the determinant of $\boldsymbol{Q}$. We use these bounds in Section V to determine certain boundary points of the RD region. Of course, distributions other than (4) might yield larger achievable regions than (6) and (7).

Finally, we give an alternate bound on the rates achieved by the above distribution by using (6) in (7)

$$R_{\mathcal{K}} \geq -\frac{1}{2}\log\left(D_{\mathcal{K}}\,\frac{\det(\boldsymbol{J} + \boldsymbol{Q}_{\mathcal{K}})}{\prod_{k\in\mathcal{K}}(1 + \mathrm{E}[W_k^2])}\right). \tag{8}$$

Observe that Shannon's RD theorem would permit $R_{\mathcal{K}} \geq -\log(D_{\mathcal{K}})/2$. Thus, the fraction to the right of $D_{\mathcal{K}}$ in (8) (which is no larger than unity by Hadamard's inequality [34, Theorem 7.8.1]) limits how close one gets to Shannon's RD function by using Theorem 1 with (4).

C. An Outer Bound for the Gaussian Source

We give an outer bound on the RD region for the Gaussian source that generalizes a result of Ozarow [6]. We call a collection of disjoint sets $\{\mathcal{K}_m\}_{m=1}^M = \{\mathcal{K}_1, \ldots, \mathcal{K}_M\}$ a *partition* of the set $\mathcal{K}$ if $\bigcup_{m=1}^M \mathcal{K}_m = \mathcal{K}$.

Theorem 2: The achievable rates R_l, $l \in \mathcal{L}$, and distortions $D_{\mathcal{K}}$, $\mathcal{K} \in 2^{\mathcal{L}}$, satisfy

$$e^{-2R_{\mathcal{K}}} \leq \min_{\{\mathcal{K}_m\}_{m=1}^M} \inf_{\lambda\geq 0}\left(D_{\mathcal{K}}\,\frac{\prod_{m=1}^M (D_{\mathcal{K}_m} + \lambda)}{(D_{\mathcal{K}} + \lambda)(1+\lambda)^{M-1}}\right) \tag{9}$$

where the minimization is over all partitions of $\mathcal{K}$.

Proof: See Appendix B, where we derive a somewhat stronger bound than (9). □

Remark 7: For $L = 1$, Theorem 2 gives Shannon's RD bound.

Remark 8: For $L = 2$, Theorem 2 gives a result of Ozarow [6]. To see this, insert [6, eqn. (11)] into [6, eq. (7)] to obtain

$$D_{\{1,2\}} \geq \sup_{\lambda\geq 0} e^{-2(R_1+R_2)}\,\frac{\lambda(1+\lambda)}{\lambda^2 + \lambda(1 + \Delta - \Pi) + \Delta} \tag{10}$$

where

$$\Delta = D_{\{1\}}D_{\{2\}} - e^{-2(R_1+R_2)}$$

and

$$\Pi = \left(1 - D_{\{1\}}\right)\left(1 - D_{\{2\}}\right).$$

We can rewrite (10) as

$$e^{-2(R_1+R_2)} \leq \inf_{\lambda\geq 0}\left(D_{\{1,2\}}\,\frac{(D_{\{1\}} + \lambda)(D_{\{2\}} + \lambda)}{(D_{\{1,2\}} + \lambda)(1+\lambda)}\right) \tag{11}$$

which is the same as (9) with $\mathcal{K} = \{1, 2\}$, $\mathcal{K}_1 = \{1\}$, and $\mathcal{K}_2 = \{2\}$.

Remark 9: For $L \geq 3$, Theorem 2 is used in Section V to determine certain boundary points of the RD region.

III. Existing Results as Special Cases

A. Existing Two-Channel Regions

Consider Theorem 1 with $L = 2$ and $X_\emptyset$ a constant. The result is the region $\mathcal{R}_{\mathrm{EGC}}$ of El Gamal and Cover [5, Theorem 1]

$$D_\mathcal{K} \geq \mathrm{E}[d_\mathcal{K}(X, X_\mathcal{K})], \quad \mathcal{K} \in \{\{1\}, \{2\}, \{1, 2\}\} \quad (12)$$

$$R_\mathcal{K} \geq I(X; X_\mathcal{K}), \quad \mathcal{K} \in \{\{1\}, \{2\}\} \quad (13)$$

$$R_1 + R_2 \geq I(X; X_1 X_2 X_{12}) + I(X_1; X_2). \quad (14)$$

Thus, $\mathcal{R}_{\mathrm{EGC}}$ is included in Theorem 1. As we have done in (14), we will often simplify notation by dropping the braces when no confusion arises, e.g., we write X_1 and X_{12} for the respective $X_{\{1\}}$ and $X_{\{1,2\}}$.

Remark 10: $\mathcal{R}_{\mathrm{EGC}}$ is the RD region for the Gaussian source and mean squared error distortion [6].

Remark 11: $\mathcal{R}_{\mathrm{EGC}}$ is the RD region for any source and distortion measure if there is no excess sum-rate, i.e., $R_1 + R_2 = R(D_{12})$ where $R(\cdot)$ is Shannon's RD function [16].

Consider next Zhang and Berger's [15, Theorem 1]. This theorem states that an inner bound $\mathcal{R}_{\mathrm{ZB}}$ to the RD region is the set of $(D_1, D_2, D_{12}, R_1, R_2)$ satisfying

$$D_\mathcal{K} \geq \mathrm{E}[d_\mathcal{K}(X, X_\mathcal{K})], \quad \mathcal{K} \in \{\{1\}, \{2\}, \{1, 2\}\} \quad (15)$$

$$R_\mathcal{K} \geq I\left(X; \hat{X}_0 \hat{X}_\mathcal{K}\right), \quad \mathcal{K} \in \{\{1\}, \{2\}\} \quad (16)$$

$$R_1 + R_2 \geq 2I\left(X; \hat{X}_0\right) + I\left(\hat{X}_1; \hat{X}_2 \middle| \hat{X}_0\right) + I\left(X; \hat{X}_1 \hat{X}_2 \middle| \hat{X}_0\right) \quad (17)$$

where $X_1 = \phi_1(\hat{X}_0, \hat{X}_1)$, $X_2 = \phi_2(\hat{X}_0, \hat{X}_2)$, and $X_{12} = \phi_{12}(\hat{X}_0, \hat{X}_1, \hat{X}_2)$. The functions $\phi_1(\cdot)$, $\phi_2(\cdot)$, and $\phi_{12}(\cdot)$ are mappings to the respective reproduction alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_{12}$. We can bound the informations in (16) as

$$I\left(X; \hat{X}_0 \hat{X}_\mathcal{K}\right) = I\left(X; \hat{X}_0 \hat{X}_\mathcal{K} X_\mathcal{K}\right) \geq I\left(X; \hat{X}_0 X_\mathcal{K}\right) \quad (18)$$

where the equality follows because $X_\mathcal{K}$ is a function of $\hat{X}_0$ and $\hat{X}_\mathcal{K}$. Similarly, we bound the right-hand side of (17) as

$$2I\left(X; \hat{X}_0\right) + I\left(\hat{X}_1; \hat{X}_2 \middle| \hat{X}_0\right) + I\left(X; \hat{X}_1 \hat{X}_2 \middle| \hat{X}_0\right) \geq 2I\left(X; \hat{X}_0\right) + I\left(X_1; X_2 \middle| \hat{X}_0\right) + I\left(X; X_1 X_2 X_{12} \middle| \hat{X}_0\right). \quad (19)$$

In other words, one cannot shrink $\mathcal{R}_{\mathrm{ZB}}$ by replacing the $\hat{X}_\mathcal{K}$ with $X_\mathcal{K}$ and adding X_{12} to the last term in (17). But the resulting region is just Theorem 1 with $L = 2$ and $X_\emptyset = \hat{X}_0$. Thus, the region obtained with Theorem 1 contains all points in $\mathcal{R}_{\mathrm{ZB}}$.

Remark 12: $\mathcal{R}_{\mathrm{ZB}}$ is larger than $\mathcal{R}_{\mathrm{EGC}}$ for the BSS and Hamming distortion [15, Sec. 3]. However, we do not know if Theorem 1 improves on $\mathcal{R}_{\mathrm{ZB}}$.

B. Existing Three-Channel Regions

Zhang and Berger's $L = 2$ region $\mathcal{R}_{\mathrm{ZB}}$ is based on a region $\mathcal{R}^*_{\mathrm{ZB}}$ for $L = 3$ [15, Theorem 1*]. This $L = 3$ region is concerned only with the distortions D_3, D_{13}, D_{23}, D_{123},

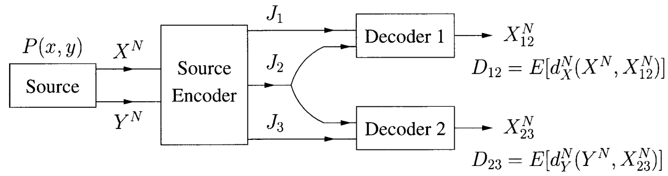

Fig. 2. Gray and Wyner's three-channel problem.

and states that an inner bound to the RD region is the set of $(D_3, D_{13}, D_{23}, D_{123}, R_1, R_2, R_3)$ satisfying

$$D_\mathcal{K} \geq \mathrm{E}[d_\mathcal{K}(X, X_\mathcal{K})], \quad \mathcal{K} \in \{3, 13, 23, 123\} \quad (20)$$

$$R_3 \geq I(X; X_3) \quad (21)$$

$$R_\mathcal{K} \geq I\left(X; \hat{X}_\mathcal{K} | X_3\right), \quad \mathcal{K} \in \{1, 2\} \quad (22)$$

$$R_1 + R_2 \geq I\left(\hat{X}_1; \hat{X}_2 \middle| X_3\right) + I\left(X; \hat{X}_1 \hat{X}_2 \middle| X_3\right) \quad (23)$$

where X_3 takes on values in $\mathcal{X}_3$, $X_{13} = \phi_{13}(X_3, \hat{X}_{13})$, $X_{23} = \phi_{23}(X_3, \hat{X}_{23})$, and $X_{123} = \phi_{123}(X_3, \hat{X}_{13}, \hat{X}_{23})$. The functions $\phi_\mathcal{K}(\cdot)$ are again mappings to the reproduction alphabets $\mathcal{X}_\mathcal{K}$. We can perform steps similar to (18) and (19) by replacing the $\hat{X}_\mathcal{K}$ with $X_\mathcal{K}$, and adding X_{123} to (23). The resulting region is at least as large as $\mathcal{R}^*_{\mathrm{ZB}}$ and is achievable with Theorem 1 by using $L = 3$ and $X_\emptyset = X_1 = X_2 = X_{12} = 0$. Thus, the region obtained with Theorem 1 contains all points in $\mathcal{R}^*_{\mathrm{ZB}}$.

We next consider perhaps the earliest information-theoretic treatment of an MD problem. Gray and Wyner in [4] studied an $L = 3$ channel problem where the source puts out pairs (X, Y) and where only the distortions D_{12} and D_{23} are of interest. Distortion D_{12} is independent of Y and distortion D_{13} is independent of X, as depicted in Fig. 2. We apply Theorem 1 with $L = 3$, $X_2 = W$ with $\mathcal{X}_2$ arbitrary, and

$$X_\emptyset = X_1 = X_3 = X_{13} = X_{123} = 0.$$

Furthermore, we choose $P(x, y, w, x_{12}, x_{13})$ to factor as $P(x, y, w)P(x_{12}|x, w)P(x_{13}|y, w)$. The resulting bounds are

$$D_\mathcal{K} \geq \mathrm{E}[d_\mathcal{K}((X, Y), X_\mathcal{K})], \quad \mathcal{K} \in \{13, 23\} \quad (24)$$

$$R_2 \geq I(XY; W) \quad (25)$$

$$R_1 + R_2 \geq I(XY; W) + I(X; X_{12}|W) \quad (26)$$

$$R_2 + R_3 \geq I(XY; W) + I(Y; X_{23}|W) \quad (27)$$

$$R_1 + R_2 + R_3 \geq I(XY; W) + I(X; X_{12}|W) + I(Y; X_{23}|W). \quad (28)$$

We can satisfy (25)–(28) by choosing the rates as

$$R_1 \geq I(X; X_{12}|W) \quad (29)$$

$$R_2 \geq I(XY; W) \quad (30)$$

$$R_3 \geq I(Y; X_{23}|W). \quad (31)$$

A final simplification arises because the distortion functions have the special form

$$d_{12}((X, Y), X_{12}) = d^*_{12}(X, X_{12})$$

and

$$d_{23}((X, Y), X_{23}) = d^*_{23}(Y, X_{23}).$$

This allows one to replace $I(X; X_{12}|W)$ and $I(Y; X_{23}|W)$ by conditional RD functions, as was done in [4, p. 1702]. In summary, the region of Theorem 1 contains all points in Gray and Wyner's achievable region. This region is, in fact, optimal for their problem [4, Theorem 8].

C. Successive Refinement

Suppose the alphabets $\mathcal{X}_{\mathcal{K}}$ are the same for all $\mathcal{K}$, and that the functions $d_{\mathcal{K}}(\cdot, \cdot)$ are the same for all $\mathcal{K}$. Consider a source sequence encoded at rate R_1 to produce a distortion of D_1 per symbol, and such that an additional R_2 bits of rate produces a lower distortion D_{12}. If $D_1 = D(R_1)$ and $D_2 = D(R_1 + R_2)$ are simultaneously achievable, where $D(\cdot)$ is the distortion-rate function, the source is said to be *successively refinable* [21]–[23]. In general, we say that a source is *successively refinable on a chain* or *in L stages* if the RD region is given by

$$D_{12\cdots K} \geq D(R_1 + \cdots + R_K), \qquad K = 1, 2, \cdots, L. \quad (32)$$

The L-stage successive refinement problem is a special case of the MD problem where only the distortions $D_1, D_{12}, \ldots, D_{12\cdots L}$ are considered. Suppose we choose all $X_{\mathcal{K}}$ to be constant except for $X_1, X_{12}, \ldots, X_{12\cdots L}$. Applying Theorem 1, we have

$$D_{12\cdots K} \geq \mathrm{E}\left[d_{12\cdots K}(X, X_{12\cdots K})\right] \quad (33)$$

$$\sum_{k=1}^{K} R_k \geq I(X; X_1 X_{12} \cdots X_{12\cdots K}) \quad (34)$$

where $K = 1, 2, \ldots, L$. In fact, the bounds (33) and (34) define the RD region. One can show this by invoking the same arguments used to prove the converse of Shannon's RD theorem. We have thus recovered a general result of Rimoldi [23, Theorem 3].

It is known [21] that the source X is successively refinable on a chain if and only if there exists a Markov chain

$$X_1 \to X_{12} \to \cdots \to X_{12\cdots L} \to X$$

such that $I(X; X_{12\cdots K}) = R(D_{12\cdots K})$. One can show that the Gaussian source is successively refinable on a chain and its RD region is given by (32) where $D(R) = e^{-2R}$.

IV. Successive Refinement on Trees

We now discuss a generalization of the successive refinement problem that we call *successive refinement on trees*. We begin with a definition.

Definition 1: A finite collection $\boldsymbol{C}$ of sets of positive integers is said to be a *tree structure* if a) for any nonempty $\mathcal{K} \in \boldsymbol{C}$, there is a unique $p(\mathcal{K}) \in \boldsymbol{C}$ called the *parent node* of $\mathcal{K}$, such that $p(\mathcal{K}) \subset \mathcal{K}$ and $|p(\mathcal{K})| = |\mathcal{K}| - 1$, and b) for distinct $\mathcal{K}_1, \mathcal{K}_2 \in \boldsymbol{C} \backslash \{\emptyset\}$, $\mathcal{V}(\mathcal{K}_1) \neq \mathcal{V}(\mathcal{K}_2)$ where $\mathcal{V}(\mathcal{K}) = \mathcal{K} - p(\mathcal{K})$.

Clearly, $\emptyset \in \boldsymbol{C}$ for any tree structure. For instance, Fig. 3 depicts the tree structure

$$\boldsymbol{C} = \{\emptyset, \{1\}, \{2\}, \{1, 3\}, \{1, 4\}, \{2, 5\}\}. \quad (35)$$

The quantity $\mathcal{V}(\mathcal{K}) = \mathcal{K} - p(\mathcal{K})$ is labeled on the branch from $p(\mathcal{K})$ to $\mathcal{K}$. Without loss of generality, we assume that $\bigcup_{\mathcal{K} \in \boldsymbol{C}} \mathcal{K} = \{1, \ldots, L\}$. We will say that a source X is *successively refinable on trees* if for any tree structure $\boldsymbol{C}$, the RD region is given by

$$D_{\mathcal{K}} \geq D(R_{\mathcal{K}}), \qquad \mathcal{K} \in \boldsymbol{C} \quad (36)$$

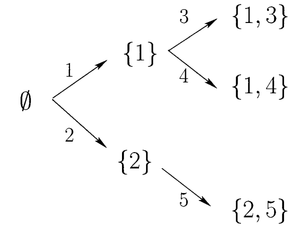

Fig. 3. Example of a tree structure.

where $D(\cdot)$ is the distortion-rate function. We prove the following result.

Theorem 3: X is a successively refinable on a chain if and only if it is successively refinable on trees.

Proof: Clearly, X is successively refinable on a chain if it is successively refinable on trees. We prove the converse by transforming the refinement problem on a tree into one on a chain by appropriately decomposing the descriptions and their rates. Observe that (36) is an outer bound on the RD region by Shannon's RD theorem. It thus suffices to show that, for given R_l, we can achieve distortions $D(R_{\mathcal{K}})$. Without loss of generality, we assume that each R_l is nonzero.

We write $\mathcal{K} \preceq \mathcal{K}'$ (or $\mathcal{K}$ precedes $\mathcal{K}'$) if $R_{\mathcal{K}} \leq R_{\mathcal{K}'}$. Let $\mathcal{K}_0 \preceq \mathcal{K}_1 \preceq \cdots \preceq \mathcal{K}_M$ be an ordering of the elements of $\boldsymbol{C}$. Clearly, $\mathcal{K}_0 = \emptyset$. Since X is successively refinable on a chain and $\{R_{\mathcal{K}_m}\colon m = 1, \ldots, M\}$ is a nondecreasing sequence, there exists a Markov chain

$$\tilde{X}_1 \to \tilde{X}_{12} \to \cdots \to \tilde{X}_{12\cdots M} \to X$$

such that $I(X; \tilde{X}_{12\cdots m}) = R_{\mathcal{K}_m}$ and

$$\tilde{D}_{12\cdots m} := \mathrm{E}\left[d\left(X, \tilde{X}_{12\cdots m}\right)\right] = D\left(R_{\mathcal{K}_m}\right). \quad (37)$$

Consider an encoding scheme that successively refines the source sequence X^N in M stages. The incremental descriptions are denoted $\tilde{J}_m$, $m = 1, 2, \ldots, M$, and their rates are $\tilde{R}_m = R_{\mathcal{K}_m} - R_{\mathcal{K}_{m-1}} \geq 0$, i.e., $H(\tilde{J}_m) = N\tilde{R}_m$. Let $\tilde{X}_{1\cdots m}^N$ be the output of the decoder whose inputs are $\tilde{J}_1, \ldots, \tilde{J}_m$. We define

$$\underline{J}_l = \left\{\tilde{J}_m\colon m = m_1 + 1, \ldots, m_2\right\} \quad (38)$$

$$R_l = R_{\mathcal{K}_{m_2}} - R_{\mathcal{K}_{m_1}} \quad (39)$$

where m_1 and m_2 are unique indexes such that $\mathcal{V}(\mathcal{K}_{m_2}) = \{l\}$ and $\mathcal{K}_{m_1} = p(\mathcal{K}_{m_2})$. We have

$$H(\underline{J}_l) = \sum_{m=m_1+1}^{m_2} N\tilde{R}_m$$

$$= N\left(R_{\mathcal{K}_{m_2}} - R_{\mathcal{K}_{m_1}}\right) = NR_l \quad (40)$$

$$\{\underline{J}_k\colon k \in \mathcal{K}_m\} = \left\{\tilde{J}_k\colon k = 1, 2, \ldots, m\right\}. \quad (41)$$

Using (41), we see that $X_{\mathcal{K}_m}^N := \tilde{X}_{1\cdots m}^N$ is a function of $\{\underline{J}_k\colon k \in \mathcal{K}_m\}$. Thus, we have constructed an encoding scheme producing $\underline{J}_l$ with rate R_l such that

$$D_{\mathcal{K}_m} = \tilde{D}_{12\cdots m} = D\left(R_{\mathcal{K}_m}\right). \quad (42)$$

□

V. Boundary Points for a Symmetric Gaussian MD Problem

Consider again the Gaussian source and mean squared error distortion. Suppose that we have symmetric rates and distortions in the sense that $R_l = R_1$ for all l and $D_{\mathcal{K}} = D_1$ for all $\mathcal{K}$, $\mathcal{K} \neq \mathcal{L}$. This means that we have only three parameters to consider: R_1, D_1, and $D_{\mathcal{L}}$. We determine the best central distortion given R_1 and D_1.

Theorem 4: Consider the MD problem with a Gaussian source, mean squared error distortion, $R_l = R_1$ for all l, $D_l = D_1$ for all l, $R_1 \geq -\log(D_1)/2$, and $D_{\mathcal{K}} \geq D_1$ for all $\mathcal{K}$ with $1 < |\mathcal{K}| < L$. The best central distortion for this problem is

$$D_{\mathcal{L}} = \sup_{\lambda \geq 0} \left(\frac{e^{-2LR_1} \lambda (1+\lambda)^{L-1}}{(D_1+\lambda)^L - e^{-2LR_1}(1+\lambda)^{L-1}} \right). \quad (43)$$

Proof: Suppose first that $D_1 = 1$, which means that there is no side distortion constraint and we can achieve $D_{\mathcal{L}} = \exp(-2LR_1)$. This is obviously the best distortion, and (43) gives the same result with $\lambda \to \infty$. So suppose from here on that $D_1 < 1$. Theorem 2 implies that the right-hand side of (43) is a lower bound on $D_{\mathcal{L}}$. We prove achievability using the distribution used in Section II-B with

$$\boldsymbol{Q} = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix} \quad (44)$$

where we require ρ to satisfy $-1/(L-1) \leq \rho \leq 1$ to ensure that $\boldsymbol{Q}$ is positive semidefinite. We will, in fact, need to consider only nonpositive ρ. One can show that

$$\det \boldsymbol{Q}_{\mathcal{K}} = \sigma^{2|\mathcal{K}|}(1-\rho)^{|\mathcal{K}|-1}[1 + (|\mathcal{K}|-1)\rho] \quad (45)$$

$$\det(\boldsymbol{J} + \boldsymbol{Q}_{\mathcal{K}}) = [\sigma^2(1-\rho)]^{|\mathcal{K}|-1} \cdot \left\{ |\mathcal{K}| + \sigma^2[1 + (|\mathcal{K}|-1)\rho] \right\}. \quad (46)$$

Inserting these identities into (6) and (7), we obtain the achievable region

$$D_{\mathcal{K}} \geq \frac{\sigma^2[1 + (|\mathcal{K}|-1)\rho]}{|\mathcal{K}| + \sigma^2[1 + (|\mathcal{K}|-1)\rho]} \quad (47)$$

$$e^{-2R_{\mathcal{K}}} \leq \left[\frac{\sigma^2}{1+\sigma^2} \right]^{|\mathcal{K}|} (1-\rho)^{|\mathcal{K}|-1}[1 + (|\mathcal{K}|-1)\rho]. \quad (48)$$

Consider first the bounds (47). We choose $\sigma^2 = D_1/(1-D_1)$ so that $D_1 = \sigma^2/(1+\sigma^2)$. Then (47) holds with equality for each $\mathcal{K}$ with $|\mathcal{K}| = 1$. For $\mathcal{K}$ with $1 < |\mathcal{K}| < L$, the bounds are automatically satisfied because the right-hand side of (47) is at most D_1. For $\mathcal{K} = \mathcal{L}$, we make (47) an equality to obtain the smallest $D_{\mathcal{L}}$ given ρ. We also introduce the variable $\mu := -D_1\rho/[1 - D_1(1-\rho)]$ which means that

$$\rho = \frac{-\mu(1-D_1)}{(1+\mu)D_1}. \quad (49)$$

We insert (49) into (47) with $\mathcal{K} = \mathcal{L}$ and choose the smallest $D_{\mathcal{L}}$ to obtain

$$D_{\mathcal{L}} = \frac{(D_1+\mu) - \mu L(1-D_1)}{(D_1+\mu) + L(1-D_1)} \quad (50)$$

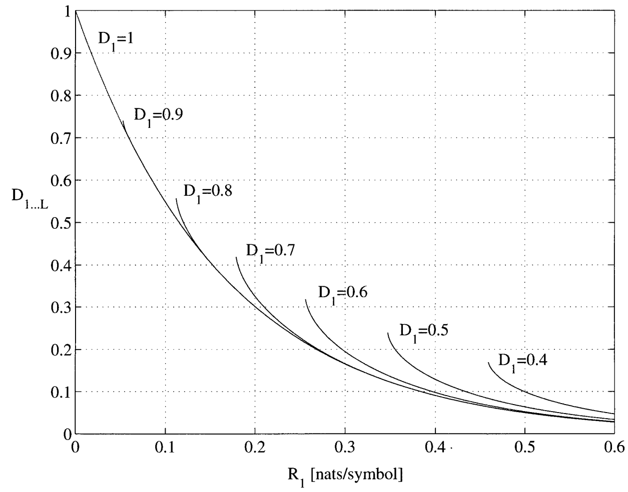

Fig. 4. Best central distortion for a symmetric MD problem with three channels.

$$\Rightarrow \; D_{\mathcal{L}} + \mu = \frac{(D_1+\mu)(1+\mu)}{(D_1+\mu) + L(1-D_1)}. \quad (51)$$

Consider next the bounds (48). For $|\mathcal{K}| = 1$ we get Shannon's RD bound. For $\mathcal{K}$ with $1 < |\mathcal{K}| \leq L$ one can check that (48) gets progressively more restrictive with $|\mathcal{K}|$ if ρ is nonpositive. We thus consider only the $\mathcal{K} = \mathcal{L}$ bound and insert (49)–(51) into (48) to obtain

$$e^{-2LR_1} \leq D_{\mathcal{L}} \frac{(D_1+\mu)^L}{(D_{\mathcal{L}}+\mu)(1+\mu)^{L-1}} \quad (52)$$

which has the same form as (9) with the partition $\mathcal{K}_m = \{m\}$ for $m = 1, 2, \ldots, L$.

Note that we can choose $\mu = 0$ (which gives $D_{\mathcal{L}} = 1/[LD_1^{-1} - (L-1)]$) because R_1 satisfies Shannon's RD bound. As we increase μ, we find that both $D_{\mathcal{L}}$ and the right-hand side of (52) decrease. We must, therefore, stop increasing μ when either (52) holds with equality or when $\rho = -1/(L-1)$. However, the latter condition implies $D_{\mathcal{L}} = 0$ which means that we can achieve equality in (52). The result is the same as (9) with $R_{\mathcal{L}} = LR_1$. Finally, we rewrite (9) to obtain (43). □

As an example, consider $L = 3$. Fig. 4 shows how the best central distortion behaves with R_1 for various D_1. For instance, for $D_1 = 1$, the best central distortion is simply the RD function $D_{\mathcal{L}} = \exp(-2LR_1)$. Consider next $D_1 = 0.5$ for which we require $R_1 \geq -\log(0.5)/2 \approx 0.347$. Now the best $D_{\mathcal{L}}$ is larger than $\exp(-2LR_1)$. In fact, as we decrease D_1, the best $D_{\mathcal{L}}$ moves further away from the $\exp(-2LR_1)$ curve.

Remark 13: Theorem 4 includes both the *high-distortion* and *low-distortion* cases [6].

Remark 14: When the single-description distortion is minimized $(D_1 = e^{-2R_1})$, the best central distortion is

$$D_{\mathcal{L}} = \frac{1}{Le^{2R_1} - (L-1)}.$$

This is calculated from Theorem 4 by first noting that the supremum in (43) is obtained by taking $\lambda \to 0^+$ and then evaluating this limit with L'Hôpital's rule. With the technique in [26], one designs for a minimum number of received descriptions n satisfying $1 \le n \le L$, and achieves distortions

$$D_{\mathcal{K}} = \frac{n}{|\mathcal{K}|e^{2nR_1} - (|\mathcal{K}| - n)}, \qquad n \le |\mathcal{K}| \le L. \tag{53}$$

Nothing is guaranteed for $D_{\mathcal{K}}$ if $|\mathcal{K}| < n$. Hence, we have shown the tightness of the main theorem of [26] for the case of $n = 1$. Whether (53) with $n > 1$ gives points on the boundary of the achievable region remains an open question.

VI. Concluding Remarks

We have presented a general achievable region (inner bound) for the L-channel multiple description problem. This region generalizes a result of El Gamal and Cover for the two-description problem ($L = 2$) and results of Zhang and Berger. We have given an outer bound to the RD region for the Gaussian source with mean squared error distortion. We showed that the outer bound meets our achievable region when only the central decoder and the L single-description decoders (with equal distortions) are of interest. However, the RD region for the Gaussian case with all decoders present is still unknown.

Appendix A
Proof of Theorem 1

Consider first the case $X_\emptyset = 0$; we will later consider general $X_\emptyset$. Suppose we are given distortions $D_{\mathcal{K}}$ for which there exists a $P(x_{(2^{\mathcal{L}})}|x)$ so that (2) is satisfied for all $\mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}$. Let $n_{x^n}(u)$ be the number of occurrences of the letter $u \in \mathcal{X}$ in x^n. Let $T_\epsilon(X)$ denote the set of ϵ-typical sequences with respect to P_X, i.e., $T_\epsilon(X)$ is the set of x^N for which

$$|n_{x^N}(u)/N - P_X(u)| \le \epsilon/|\mathcal{X}| \tag{54}$$

for all $u \in \mathcal{X}$. We encode and decode as follows.

1. *Codebook*: Let j_l, $l = 1, \ldots, L$, denote indexes between 1 and e^{NR_l}. The codebooks are formed by assigning a codeword $x_{\mathcal{K}}^N(j_{(\mathcal{K})})$ to each $\mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}$ and $j_{(\mathcal{K})}$. The components $x_{\mathcal{K}}^{(n)}(j_{(\mathcal{K})})$, $n = 1, \ldots N$, of the codewords are chosen independently at random via the distribution

$$P_{X_{\mathcal{K}}|X_{(2^{\mathcal{K}}-\{\mathcal{K}\})}}\left(\cdot\left|x^{(n)}_{(2^{\mathcal{K}}-\{\mathcal{K}\})}(j_{(\mathcal{K})})\right.\right). \tag{55}$$

Note that to generate the $x_{\mathcal{K}}^N$ codebook, we first need the $x_{\mathcal{M}}^N$ codebooks for all $\mathcal{M} \in 2^{\mathcal{K}} - \{\mathcal{K}\}$.

2. *Encoding*: Given the source sequence x^N, the encoder tries to find a $j_{(\mathcal{L})}$ for which x^N is jointly ϵ-typical with the codewords corresponding to $j_{(\mathcal{L})}$, i.e., $j_{(\mathcal{L})}$ must satisfy

$$\left(x^N, x^N_{(2^{\mathcal{L}})}(j_{(\mathcal{L})})\right) \in T_\epsilon\left(X, X_{(2^{\mathcal{L}})}\right). \tag{56}$$

If such a $j_{(\mathcal{L})}$ does not exist, we set $j_l = 1$ for $l = 1, \ldots, L$. The j_l are transmitted over the L channels.

3. *Decoding*: Decoder $D_{\mathcal{K}}$ takes $j_{(\mathcal{K})}$ as its input and produces the output $x_{\mathcal{K}}^N(j_{(\mathcal{K})})$.

As usual, it suffices to consider only $x^N \in T_\epsilon(X)$. Suppose further that there are $j_{(\mathcal{L})}$ so that (56) is satisfied. One can check that

$$(x^N, y^N) \in T_\epsilon(X, Y) \;\Rightarrow\; x^N \in T_\epsilon(X),\; y^N \in T_\epsilon(Y). \tag{57}$$

Thus, we have

$$\begin{aligned}
\sum_{n=1}^{N} \frac{d_{\mathcal{K}}(x^{(n)}, x_{\mathcal{K}}^{(n)})}{N}
&= \sum_{u\in\mathcal{X},\, v\in\mathcal{X}_{\mathcal{K}}} \frac{n_{x^N x^N_{\mathcal{K}}}(u, v)}{N} d_{\mathcal{K}}(u, v) \\
&\le \sum_{u\in\mathcal{X},\, v\in\mathcal{X}_{\mathcal{K}}} \left[P_{XX_{\mathcal{K}}}(u, v) + \frac{\epsilon}{|\mathcal{X}|\,|\mathcal{X}_{\mathcal{K}}|}\right] d_{\mathcal{K}}(u, v) \\
&\le D_{\mathcal{K}} + \epsilon\, d_{\max}
\end{aligned} \tag{58}$$

for all $\mathcal{K} \subseteq \mathcal{L}$, where the first inequality follows by (54), (56), and (57). We can, therefore, bound the average distortion over the ensemble of all codes as

$$\sum_{n=1}^{N} \frac{\mathrm{E}\left[d_{\mathcal{K}}(x^{(n)}, X_{\mathcal{K}}^{(n)})\right]}{N} \le [D_{\mathcal{K}} + \epsilon\, d_{\max}] + P_e d_{\max} \tag{59}$$

where P_e is the probability that (56) does not occur for any $j_{(\mathcal{L})}$ and a fixed $x^N \in T_\epsilon(X)$.

We proceed to find rates $R_{(\mathcal{L})}$ for which P_e approaches zero as N increases. Let $F(j_{(\mathcal{L})})$ denote the event that $j_{(\mathcal{L})}$ gives a good choice of codewords, i.e.,

$$F\left(j_{(\mathcal{L})}\right) = \left\{\left(x^N, X^N_{(2^{\mathcal{L}})}\left(j_{(\mathcal{L})}\right)\right) \in T_\epsilon\left(X, X_{(2^{\mathcal{L}})}\right)\right\}. \tag{60}$$

We can write $P_e = \Pr[\bigcap_{j_{(\mathcal{L})}} F^c(j_{(\mathcal{L})})]$, where A^c denotes the complement of the set A in the sample space. Alternatively, we can write P_e in terms of the random variables

$$\chi\left(j_{(\mathcal{L})}\right) = \begin{cases} 1, & \text{if } F\left(j_{(\mathcal{L})}\right) \text{ occurs} \\ 0, & \text{else.} \end{cases} \tag{61}$$

We have $P_e = \Pr[K = 0]$ where $K = \sum_{j_{(\mathcal{L})}} \chi(j_{(\mathcal{L})})$. Since $K = 0$ implies $|K - \mathrm{E}[K]| \ge \mathrm{E}[K]/2$, we can bound

$$P_e \le \Pr[|K - \mathrm{E}[K]| \ge \mathrm{E}[K]/2] \le \frac{4\,\mathrm{Var}[K]}{(\mathrm{E}[K])^2} \tag{62}$$

where $\mathrm{Var}[K]$ is the variance of K and where the last step follows by the Chebyshev inequality [35, p. 13]. We use the usual typicality arguments (see, e.g., [36, Ch. 12]) to bound $\mathrm{E}[K]$ as

$$\begin{aligned}
\mathrm{E}[K] &= \sum_{j_{(\mathcal{L})}} \mathrm{E}\left[\chi(j_{(\mathcal{L})})\right] = \sum_{j_{(\mathcal{L})}} \Pr\left[F(j_{(\mathcal{L})})\right] \\
&\ge \sum_{j_{(\mathcal{L})}} e^{-N[\gamma(\mathcal{L})+c\epsilon]} \\
&= e^{NR_{\mathcal{L}}} e^{-N[\gamma(\mathcal{L})+c\epsilon]}
\end{aligned} \tag{63}$$

where c is a positive integer independent of ϵ and

$$\gamma(\mathcal{L}) = -H\left(X_{(2^{\mathcal{L}})}\middle| X\right) + \sum_{\mathcal{M}\subseteq\mathcal{L}} H\left(X_{\mathcal{M}}\middle| X_{(2^{\mathcal{M}}-\{\mathcal{M}\})}\right). \tag{64}$$

We have chosen the natural logarithm to compute the entropies.

Consider next $\mathrm{Var}[K] = \mathrm{E}[K^2] - (\mathrm{E}[K])^2$. We have

$$\begin{aligned}
\mathrm{E}[K^2] &= \sum_{j_{(\mathcal{L})}} \sum_{k_{(\mathcal{L})}} \mathrm{E}\left[\chi\left(j_{(\mathcal{L})}\right)\, \chi\left(k_{(\mathcal{L})}\right)\right] \\
&= \sum_{j_{(\mathcal{L})}} \sum_{k_{(\mathcal{L})}} \Pr\left[F\left(j_{(\mathcal{L})}\right) \cap F\left(k_{(\mathcal{L})}\right)\right].
\end{aligned} \tag{65}$$

The probability in (65) depends on whether $j_l = k_l$ for some l, so define $\mathcal{Q}$ as the set of subscripts l for which $j_l = k_l$. This, of course, means that $X_{(2^{\mathcal{Q}})}(j_{(\mathcal{Q})}) = X_{(2^{\mathcal{Q}})}(k_{(\mathcal{Q})})$. Also, because of the conditional independence in (55), all the other choices of codewords in $X_{(2^{\mathcal{L}})}(j_{(\mathcal{L})})$ and $X_{(2^{\mathcal{L}})}(k_{(\mathcal{L})})$ will be independent given $X_{(2^{\mathcal{Q}})}(j_{(\mathcal{Q})})$. Again, using typicality arguments, we can upper-bound the probability in (65) by

$$e^{-N[\gamma(\mathcal{Q})-c\epsilon]}\left[e^{-N[\gamma(\mathcal{L})-\gamma(\mathcal{Q})-c\epsilon]}\right]^2 \tag{66}$$

where we have chosen the c in (63) to be large enough so that (66) is valid. For $\mathcal{Q} = \emptyset$, however, we replace (66) with $\Pr[F(j_{(\mathcal{L})})]\Pr[F(k_{(\mathcal{L})})]$. Next, the number of ways in which $j_{(\mathcal{L})}$ and $k_{(\mathcal{L})}$ can be chosen so that they overlap in $\mathcal{Q}$ is

$$\prod_{l\in\mathcal{Q}} e^{NR_l} \prod_{l\notin\mathcal{Q}} e^{NR_l}\left(e^{NR_l}-1\right) \le e^{NR_{\mathcal{L}}}e^{N(R_{\mathcal{L}}-R_{\mathcal{Q}})}. \tag{67}$$

Inserting the bounds of (66) and (67) into (65), and using (63), we have

$$\mathrm{Var}[K^2] \le \sum_{\mathcal{Q}\in 2^{\mathcal{L}}-\{\emptyset\}} e^{-2N[\gamma(\mathcal{L})-R_{\mathcal{L}}]+N[\gamma(\mathcal{Q})-R_{\mathcal{Q}}]+3Nc\epsilon}. \tag{68}$$

Inserting (68) into (62), we obtain

$$P_e \le \sum_{\mathcal{Q}\in 2^{\mathcal{L}}-\{\emptyset\}} 4\, e^{N[\gamma(\mathcal{Q})-R_{\mathcal{Q}}+5c\epsilon]}. \tag{69}$$

Recall that we can choose ϵ to be any positive number. Thus, as long as $R_{\mathcal{Q}} > \gamma(\mathcal{Q})$ for all nonempty $\mathcal{Q}$, there is a code that satisfies the distortion requirements with probability as close to one as desired. This proves Theorem 1 with $X_\emptyset = 0$.

We use an approach similar to [15] to include $X_\emptyset \ne 0$. Suppose we add to each description $NR_\emptyset$ extra nats that represent a sequence $X_\emptyset^N$. Equivalently, we have an $(L+1)$-channel problem for which the $(L+1)$th channel carries the information representing $X_\emptyset^N$. We use the achievable region derived above and set $X_{\mathcal{K}} = 0$ for $L+1 \notin \mathcal{K}$. After simplifying, we have the following 2^L rate bounds:

$$R_{L+1} > I(X;\, X_{L+1}) \tag{70}$$

$$R_{\mathcal{K}} + R_{L+1} > -H\left(X_{(2^{\mathcal{L}})}X_{L+1}|X\right) + H(X_{L+1}) + \sum_{\mathcal{M}\subseteq\mathcal{L}} H\left(X_{\mathcal{M}}|X_{L+1}X_{(2^{\mathcal{M}}-\{\mathcal{M}\})}\right) \tag{71}$$

where $\mathcal{K}$ is any subset of $2^{\mathcal{L}} - \{\emptyset\}$ (note that $X_\emptyset = 0$ in (71)). However, by including R_{L+1} in channels $1, \ldots, L$ we are actually transmitting at rates $R'_l = R_l + R_{L+1}$. The bounds (71) can thus be rewritten as

$$R'_{\mathcal{K}} > (|\mathcal{K}|-1)R_{L+1} - H\big(X_{(2^{\mathcal{L}})}X_{L+1}\big|\, X\big) + H(X_{L+1}) + \sum_{\mathcal{M}\subseteq\mathcal{L}} H\big(X_{\mathcal{M}}\,\big|X_{L+1}X_{(2^{\mathcal{M}}-\{\mathcal{M}\})}\big). \tag{72}$$

We now choose R_{L+1} to be as small as possible, i.e., $R_{L+1} = I(X;\, X_{L+1}) + \delta$ for small positive δ. We further replace X_{L+1} by $X_\emptyset$ and $R'_{\mathcal{K}}$ by $R_{\mathcal{K}}$ to obtain the bounds (3). This proves Theorem 1.

APPENDIX B
PROOF OF THEOREM 2

Our approach follows closely that of [6, Sec. 2]. We have

$$-\frac{N}{2}\log(D_{\mathcal{K}}) \overset{\text{(a)}}{\le} I\left(X^N;\, X_{\mathcal{K}}^N\right) \overset{\text{(b)}}{=} H\left(X_{\mathcal{K}}^N\right) \overset{\text{(c)}}{\le} H\left(J_{(\mathcal{K})}\right) \tag{73}$$

where (a) follows by Shannon's RD theorem, (b) because $X_{\mathcal{K}}^N$ is a function of X^N, and (c) because $X_{\mathcal{K}}^N$ is a function of $J_{(\mathcal{K})}$. We define

$$I\left(J_{(\mathcal{K})}\right) := \left[\sum_{k\in\mathcal{K}} H(J_k)\right] - H\left(J_{(\mathcal{K})}\right). \tag{74}$$

Note that $H(J_k)$ has at most NR_k nats of information so we can bound

$$H\left(J_{(\mathcal{K})}\right) \le \left[\sum_{k\in\mathcal{K}} NR_k\right] - I\left(J_{(\mathcal{K})}\right). \tag{75}$$

The bounds (73) and (75) give

$$D_{\mathcal{K}} \ge e^{-2R_{\mathcal{K}}}\exp\left(\frac{2}{N}\, I\left(J_{(\mathcal{K})}\right)\right) \tag{76}$$

which is the analog of [6, eq. (7)].

We continue to follow [6] and define $Y^N = X^N + Z^N$ where the $Z^{(n)}, n = 1, \ldots, N$, are zero mean, i.i.d., Gaussian random variables with variance λ. Let $\{\mathcal{K}_m\}_{m=1}^M$ be a partition of $\mathcal{K}$ and consider the quantity

$$I' := \left[\sum_{m=1}^{M} H\left(J_{(\mathcal{K}_m)}\big|\, Y^N\right)\right] - H\left(J_{(\mathcal{K})}\big|\, Y^N\right). \tag{77}$$

We have $I' \ge 0$ because $H(A|C) + H(B|C) \ge H(AB|C)$. We next expand

$$\begin{aligned} I\left(J_{(\mathcal{K})}\right) &= \left[\sum_{m=1}^{M}\sum_{k\in\mathcal{K}_m} H(J_k)\right] - H\left(J_{(\mathcal{K})}\right) \\ &= \left[\sum_{m=1}^{M} I\left(J_{(\mathcal{K}_m)}\right) + H\left(J_{(\mathcal{K}_m)}\right)\right] - H\left(J_{(\mathcal{K})}\right). \end{aligned} \tag{78}$$

Subtracting I' from the right-hand side of (78), we have

$$I\left(J_{(\mathcal{K})}\right) \ge \left[\sum_{m=1}^{M} I\left(J_{(\mathcal{K}_m)}\right) + I\left(J_{(\mathcal{K}_m)};\, Y^N\right)\right] - I\left(J_{(\mathcal{K})};\, Y^N\right). \tag{79}$$

Note that this bound is useless if $M = 1$. Thus, one should consider only nontrivial partitions. For those $\mathcal{K}$ having only one element one can simply use the bound $I(J_{(\mathcal{K})}) \ge 0$.

Consider now the terms $I(J_{(\mathcal{K}_m)};\, Y^N)$ in (79). Observe that if one can get distortion $D_{\mathcal{K}_m}$ for X^N using $J_{(\mathcal{K}_m)}$ then one can also get distortion $D_{\mathcal{K}_m} + \lambda$ for Y^N. Thus, we can use Shannon's RD therem to bound

$$\begin{aligned} I\left(J_{(\mathcal{K}_m)};\, Y^N\right) &\ge \frac{N}{2}\log\left(\frac{\mathrm{Var}[Y]}{D_{\mathcal{K}_m}+\lambda}\right) \\ &= \frac{N}{2}\log\left(\frac{1+\lambda}{D_{\mathcal{K}_m}+\lambda}\right). \end{aligned} \tag{80}$$

This is basically the same as [6, eq. (9)]. Similarly, we use the same steps as in [6] to arrive at

$$I\left(J_{(\mathcal{K})}; Y^N\right) \leq \frac{N}{2}\log\left(\frac{1+\lambda}{e^{-2R_{\mathcal{K}}}e^{2I(J_{(\mathcal{K})})/N}+\lambda}\right). \quad (81)$$

This is the analog of [6, eq. (10)].

We next define $t_{\mathcal{K}} = \exp[2I(J_{(\mathcal{K})})/N]$ and combine (79)–(81) to obtain

$$t_{\mathcal{K}} \geq \left[\prod_{m=1}^{M} t_{\mathcal{K}_m}\right]\cdot\frac{(1+\lambda)^M}{\prod_{m=1}^{M}(D_{\mathcal{K}_m}+\lambda)}\cdot\frac{\left(e^{-2R_{\mathcal{K}}}t_{\mathcal{K}}+\lambda\right)}{1+\lambda}.$$

Solving for $t_{\mathcal{K}}$ we get the analog of [6, eq. (11)] as

$$t_{\mathcal{K}} \geq \frac{\lambda(1+\lambda)^{M-1}\prod_{m=1}^{M}t_{\mathcal{K}_m}}{\left[\prod_{m=1}^{M}(D_{\mathcal{K}_m}+\lambda)\right]-(1+\lambda)^{M-1}e^{-2R_{\mathcal{K}}}\prod_{m=1}^{M}t_{\mathcal{K}_m}}. \quad (82)$$

We can choose λ to maximize the right-hand side of (82). The $t_{\mathcal{K}_m}$ can be computed by recursively applying (82) starting with $t_{\{l\}} = 1$ for $l = 1, \ldots, L$. Substituting (82) into (76) we have

$$e^{-2R_{\mathcal{K}}} \leq D_{\mathcal{K}}\frac{\prod_{m=1}^{M}(D_{\mathcal{K}_m}+\lambda)}{(D_{\mathcal{K}}+\lambda)(1+\lambda)^{M-1}\prod_{m=1}^{M}t_{\mathcal{K}_m}}. \quad (83)$$

This bound strengthens (9) because $t_{\mathcal{K}_m} \geq 1$. However, we use only the weakened form (9) for our computations.

Acknowledgment

The authors wish to thank the Associate Editor and the three reviewers for their useful comments.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948. Reprinted in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, Eds. Piscataway, NJ: IEEE Press, 1993, pp. 5–83.

[2] ——, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Int. Conv. Rec.*, pp. 142–163, Mar. 1959. Reprinted in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, Eds. Piscataway, NJ: IEEE Press, 1993, pp. 325–350.

[3] R. M. Gray, “Source coding a symmetric source over a simple network,” in *Proc. 1973 Hawaii Int. Conf. System Science*, Honolulu, HI, Jan. 1973, pp. 354–355.

[4] R. M. Gray and A. D. Wyner, “Source-coding for a simple network,” *Bell Syst. Tech. J.*, vol. 53, pp. 1681–1721, Nov. 1974.

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

[6] L. Ozarow, “On a source-coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[7] R. Zamir, “Gaussian codes and Shannon bounds for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 2629–2635, Nov. 1999.

[8] P. L. Dragotti, S. D. Servetto, and M. Vetterli, “Optimal filter banks for multiple description coding: Analysis and synthesis,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 2036–2052, July 2002.

[9] H. S. Witsenhausen, “On source networks with minimal breakdown degradation,” *Bell Syst. Tech. J.*, vol. 59, pp. 1083–1087, July–Aug. 1980.

[10] J. K. Wolf, A. D. Wyner, and J. Ziv, “Source-coding for multiple descriptions,” *Bell Syst. Tech. J.*, vol. 59, pp. 1417–1426, Oct. 1980.

[11] H. S. Witsenhausen and A. D. Wyner, “On team guessing with independent information,” *Math. Oper. Res.*, vol. 6, pp. 293–304, May 1981.

[12] H. S. Witsenhausen, “An achievable region for the breakdown degradation problem with multiple channels,” Bell Lab. Tech. Memo., TM-81-11217-3, Jan. 1981.

[13] T. Berger and Z. Zhang, “Minimum breakdown degradation in binary source coding,” *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 807–814, Nov. 1983.

[14] R. Ahlswede, “On multiple descriptions and team guessing,” *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 543–549, July 1986.

[15] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 502–521, July 1987.

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

[17] R. Venkataramani, G. Kramer, and V. K. Goyal, “Bounds on the achievable region for certain multiple description coding problems,” in *Proc. 2001 IEEE Int. Symp. Information Theory*, Washington, DC, June 24–29, 2001, p. 148.

[18] V. Koshelev, “Multilevel source coding and data-transmission theorem,” in *Proc. VII All-Union Conf. Theory of Coding and Data Transm.*, vol. 1, Vilnius, U.S.S.R., 1978, pp. 85–92.

[19] ——, “Hierarchical coding of discrete sources,” *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[20] ——, “An evaluation of the average distortion for discrete scheme of sequential approximation,” *Probl. Pered. Inform.*, vol. 17, no. 3, pp. 20–33, 1981.

[21] W. H. R. Equitz and T. M. Cover, “Successive refinement of information,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[22] W. H. R. Equitz and T. M. Cover, “Addendum to ‘Successive refinement of information’,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1465–1466, July 1993.

[23] B. Rimoldi, “Successive refinement of information: Characterization of the achievable rates,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[24] F.-W. Fu and R. W. Yeung, “On the rate-distortion region for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 2012–2021, July 2002.

[25] R. Venkataramani, G. Kramer, and V. K Goyal, “Successive refinement on trees: A special case of a new MD coding region,” in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2001, pp. 293–301.

[26] S. S. Pradhan, R. Puri, and K. Ramchandran, “(n, k) source-channel erasure codes: Can parity bits also refine quality?,” in *Proc. Conf. Information Sciences and Systems*. Baltimore, MD: The Johns Hopkins Univ., 2001.

[27] R. Puri, S. S. Pradhan, and K. Ramchandran, “n-channel symmetric multiple descriptions: New rate regions,” in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June 30–July 5 2002, p. 93.

[28] V. A. Vaishampayan, “Design of multiple description scalar quantizers,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[29] V. K Goyal, “Multiple description coding: Compression meets the network,” *IEEE Signal Processing Mag.*, vol. 18, pp. 74–93, Sept. 2001.

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

[31] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder—II: General sources,” *Inform. Contr.*, vol. 38, pp. 60–80, July 1978.

[32] A. D. 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, Jan. 1976.

[33] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The John Hopkins Univ. Press, 1996.

[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[35] R. G. Gallager, *Discrete Stochastic Processes*. Boston, MA: Kluwer, 1996.

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