On multiple description source coding with decoder side information

Suhas N. Diggavi
School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland, CH 1015.
suhas.diggavi@epfl.ch

Vinay A. Vaishampayan
AT&T Shannon Laboratories, 180, Park Ave, Bldg 103, Florham Park, New Jersey, NJ 07932, USA.
vinay@research.att.com

Abstract — ***We formulate a multi-terminal source coding problem, where we are required to construct a multiple-description code for a source sequence when side information about dependent random processes is available at the decoder only, or at both the decoder and the encoder. We describe an achievable rate-distortion region for these problems in two cases: where there is common side-information at the decoders and when they are different. In the quadratic Gaussian case, and when there is common side information among the decoders, we show that the rate region when both the encoder and decoder have access to the side information coincides with that of decoder–only side information. This is analogous to the single-description (Wyner-Ziv) case, and an explicit characterization of the rate-distortion region is provided for this case.***

I. Introduction

This problem is motivated by the applications of distributed source coding to sensor networks [8], and in video compression [9]. The basic idea behind these applications involves compression of correlated sources. In the lossless case, the Slepian-Wolf theorem [10] yields the surprising result that one can compress correlated sources in a distributed manner as efficiently as if they were jointly compressed (or "co-located"). The rate-distortion problem with decoder-only side information was solved by Wyner and Ziv [13]. The rate-distortion region for distributed source coding remains an open question [3].

The question we address in this paper is robust compression of a source when correlated side information is available at the decoder, a generalization of the multiple description problem to include side information. In multiple description source coding, the goal is to produce a set of descriptions (each with its own rate constraints) of a source sequence such that reception of any subset of the descriptions guarantees a certain performance (rate-distortion) between the source and the decoded sequence. The information-theoretic issues of the problem were investigated by several researchers (see for example [4, 7, 14, 1] and references therein). The rate-distortion region for the multiple description problem in the two-channel Gaussian case with squared-error distortions is known [4], [7], though the general case still remains open.

The main question examined in this paper is the achievable rate distortion region for multiple description source coding when side information about a correlated random process is known at the decoder. In a sense, this formulation marries the multiple description source coding problem with the Wyner-Ziv question. For brevity we examine the two-description problem with the understanding that the techniques used can be easily generalized to more than two descriptions. At the completion of our work, we became aware of an independent recent extension of the Wyner-Ziv result to successive refinement of information [11]. However, the successive refinement problem is a special case of multiple description source coding.

In this paper we establish an achievable rate-distortion region for the general case when the decoders have different side information (see Figure 1). However, we mainly focus on the special case where there is common side information (see Figure 2), and demonstrate that for jointly Gaussian source and side information along with the squared-error distortion measure, the rate-region is tight. In the process of doing so, we also establish the two-description rate-distortion region when *both* the encoder and decoder have access to the side information. In [13] it was shown that in the single description Gaussian case, the decoder–only side-information rate-distortion function coincided with that when both encoder and decoder were informed of the side-information. Our result establishes that this is also true in the Gaussian two-description problem with common decoder side-information. However, the rate-region for the general problem posed is open.

The paper is organized as follows. We formally state the problem in Section II. In Sections III we state the main achievability results. In IV, we focus on the special case of jointly Gaussian source and side-information, and establish a converse result. The appendix contains some details of the proof of the main results of this paper.

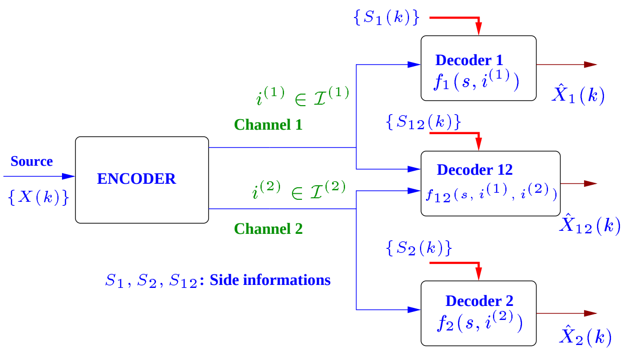

Figure 1: Multiple description source coding with side information for each multiple description decoder.

II. Problem Statement

We first establish the notation for the simpler case where there is common side information, depicted in Figure 2. We assume that $\{X(k)\}, k = 1, \dots, n$ is a sequence of i.i.d. discrete random variables belonging to a finite set $\mathcal{X}$ drawn according to a probability mass function $p(x)$. In a similar manner, we define the side information as a sequence $\{S(k)\}$ of i.i.d. discrete random variables belonging to the finite set $\mathcal{S}$. The source $\{X(k)\}$ and the side information $\{S(k)\}$ are obtained from independent drawings of dependent random variables

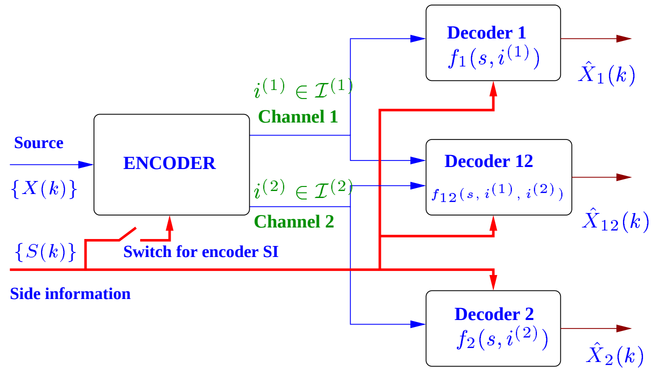

Figure 2: Multiple description with common side information. This is a special case of the scenario in Figure 1, and is the main focus of this paper. The switch determines whether the encoder has access to side information.

X, S from the joint probability distribution $Q(x, s) = \mathbb{P}(X = x, S = s), x \in \mathcal{X}, s \in \mathcal{S}$.

Let $\mathcal{I}_M \triangleq \{0, 1, \ldots, M-1\}$. For an arbitrary finite set $\mathcal{U}$, we also define $\mathcal{U}^n$ as the set of n-vectors with elements in $\mathcal{U}$. Let us define the reproduction alphabets $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \hat{\mathcal{X}}_{12}$, which represent the reproductions for decoder 1, 2 and 12 respectively. We define the following reconstruction maps,

$$\begin{aligned} f_1 &: \mathcal{S}^n \times \mathcal{I}_{M_1} \to \hat{\mathcal{X}}_1^n, \;\; f_2 : \mathcal{S}^n \times \mathcal{I}_{M_2} \to \hat{\mathcal{X}}_2^n \\ f_{12} &: \mathcal{S}^n \times \mathcal{I}_{M_1} \times \mathcal{I}_{M_2} \to \hat{\mathcal{X}}_{12}^n \end{aligned} \tag{1}$$

where the encoder maps are given by

$$g_1 : \mathcal{X}^n \to \mathcal{I}_{M_1}, \;\; g_2 : \mathcal{X}^n \to \mathcal{I}_{M_2} \tag{2}$$

Given the three reproduction alphabets, decoder maps and distortion measures

$$d_m : \mathcal{X} \times \hat{\mathcal{X}}_m \to \mathbb{R}, \quad m = 1, 2, 12, \tag{3}$$

we define the distortion between sequences as the average per-letter distortion,

$$d_m(x^n, \hat{x}_m^n) = \frac{1}{n} \sum_{k=1}^{n} d_m(x(k), \hat{x}_m(k)), \quad m = 1, 2, 12, \tag{4}$$

where the sequences $x^n \in \mathcal{X}^n, \hat{x}_m^n \in \hat{\mathcal{X}}_m^n, m = 1, 2, 12$. Note that we have abused notation a little by reusing d_m both for the per-letter distortion in (3) as well as in sequences in (4). The context should make the notation clear.

A tuple $\{R_1, R_2, D_1, D_2, D_{12}\}$ is said to be achievable if given $\epsilon > 0$, there exists a sequence of encoders g_1, g_2, indexed by n, such that for $M_1 = e^{nR_1}, M_2 = e^{nR_2}$, $g_1(x^n) \in \mathcal{I}_{M_1}, g_2(x^n) \in \mathcal{I}_{M_2}$ and for sufficiently large n,

$$\mathbb{E}[d_m(X^n, \hat{X}^n)] \leq D_m + \epsilon, \quad m = 1, 2, 12. \tag{5}$$

The *rate distortion region* $R(\mathbf{D})$, for distortion $\mathbf{D} = \{D_1, D_2, D_{12}\}$ is the closure of the set of achievable rates $\{R_1, R_2\}$ and distortions $\mathbf{D}$. Our problem is to identify the largest possible subset of the rate-distortion region.

For the closely related problem where $\{S(k)\}$ is available at *both* the encoder and decoder (see Figure 2), the encoder mappings

$$g_1^{(e)} : \mathcal{S}^n \times \mathcal{X}^n \to \mathcal{I}_{M_1}, \;\; g_2^{(e)} : \mathcal{S}^n \times \mathcal{X}^n \to \mathcal{I}_{M_2}, \tag{6}$$

are used. The rate distortion region for this case is similarly defined as in the decoder-only side information case.

Finally, when we have different side informations at the decoders (as depicted in Figure 1), the reconstruction functions defined in (1) change slightly to account for that as,

$$\begin{aligned} f_1 &: \mathcal{S}_1^n \times \mathcal{I}_{M_1} \to \hat{\mathcal{X}}_1^n, \;\; f_2 : \mathcal{S}_2^n \times \mathcal{I}_{M_2} \to \hat{\mathcal{X}}_2^n \\ f_{12} &: \mathcal{S}_{12}^n \times \mathcal{I}_{M_1} \times \mathcal{I}_{M_2} \to \hat{\mathcal{X}}_{12}^n \end{aligned} \tag{7}$$

and the encoder maps remain as in (2).

Also we use the common notation $H(X), H(X|S), I(X;S)$, and $I(X;S|Y)$ for entropies and mutual informations [3].

III. Main Results

We first state the results for the common side information case depicted in Figure 2, and later generalize the results to the case shown in Figure 1. We digress briefly to give an achievable rate region for the case with encoder and decoder side information (Figure 2 with switch closed) by generalizing the results in [4, 14].

Theorem III.1 *Let $(X(1), S(1)), (X(2), S(2)) \ldots$ be a sequence of i.i.d. finite alphabet random variables drawn according to probability mass function $Q(x, s)$. Let the encoder and decoder, defined as in (6) and (1), have access to the side information $\{S(k)\}$. Then any tuple $(R_1, R_2, D_1, D_2, D_{12})$ is achievable if there exist random variables $(\hat{X}_0, \hat{X}_1, \hat{X}_2, \hat{X}_{12})$ jointly distributed with the source random variable X and the side-information random variable S such that,*

$$\begin{aligned} R_1 &> I(X; \hat{X}_0, \hat{X}_1 | S) \\ R_2 &> I(X; \hat{X}_0, \hat{X}_2 | S) \\ R_1 + R_2 &> 2I(X; \hat{X}_0 | S) + I(X; \hat{X}_{12}, \hat{X}_1, \hat{X}_2 | \hat{X}_0, S) \\ &+ I(\hat{X}_1; \hat{X}_2 | \hat{X}_0, S) \end{aligned} \tag{8}$$

for some probability mass function $p(x, s, x_0, x_1, x_2, x_{12}) = Q(x, s) p(x_0, x_1, x_2, x_{12} | x, s)$ and

$$\begin{aligned} D_1 &\geq \mathbb{E}[d_1(X, \hat{X}_1)] \\ D_2 &\geq \mathbb{E}[d_2(X, \hat{X}_2)] \\ D_{12} &\geq \mathbb{E}[d_{12}(X, \hat{X}_{12})] \end{aligned} \tag{9}$$

Proof idea: The proof combines well-known ideas from conditional rate-distortion theory [2, 5] with the proof technique introduced in [4, 14]. Using these, as one would expect, we get the conditional form of the theorems proved in [4, 14] as stated above. ■

The following is one of the main results of this paper.

Theorem III.2 *Let $(X(1), S(1)), (X(2), S(2)) \ldots$ be a sequence of i.i.d. finite alphabet random variables drawn according to probability mass function $Q(x, s)$. Let the encoder and decoder, defined as in (2) and (1), where only the decoder has access to the side information $\{S(k)\}$. Then any tuple $(R_1, R_2, D_1, D_2, D_{12})$ is achievable if there exist random variables (W_0, W_1, W_2, W_{12}) with probability mass* *function* $p(x,s,w_0,w_1,w_2,w_{12}) = Q(x,s)p(w_0,w_1,w_2,w_{12}|x)$, *such that*

$$\begin{aligned} R_1 &> I(X;W_0,W_1|S) \\ R_2 &> I(X;W_0,W_2|S) \\ R_1+R_2 &> 2I(X;W_0|S)+I(X;W_{12},W_1,W_2|W_0,S) \\ &+ I(W_1;W_2|W_0,S) \end{aligned} \tag{10}$$

and there exist reconstruction functions f_1, f_2, f_{12} *which satisfy*

$$\begin{aligned} D_1 &\geq \mathbb{E}[d_1(X,f_1(S,W_1,W_0))] \\ D_2 &\geq \mathbb{E}[d_2(X,f_2(S,W_2,W_0))] \\ D_{12} &\geq \mathbb{E}[d_{12}(X,f_{12}(S,W_{12},W_1,W_2,W_0))]. \end{aligned} \tag{11}$$

■

Proof idea: The basic idea of the proof is combining the multiple description source coding technique *without* side information [4] along with a binning technique similar to that introduced in [10, 13]. A mutual information identity (see (32) in Appendix) enables us to write this in a form which makes the analogy between (8) and (10) more clear. More details are given in Appendix A. ■

Remarks:(i) The form of the probability mass function is equivalent to the Markov chain condition $S \leftrightarrow X \leftrightarrow (W_0,W_1,W_2,W_{12})$. (ii) The usual time-sharing arguments can be applied to convexify the regions given in Theorems III.1 and III.2. Hence the convex hull of the region defined in (8) and (9) are also achievable. (iii)Although we prove the Theorems III.1 and III.2 for discrete random variables, an argument similar to that in [12] can be used to extend this result to well-behaved continuous random variables and distortion measures. In particular, Gaussian sources and Euclidean squared distance distortion measures fall into this category. Therefore, we apply the result in Theorems III.1 and III.2 to this case. (iv) We can easily see that the rate-distortion region when *both* the encoder and the decoder have access to the side information $\{S(k)\}$ forms an inner bound to the rate-distortion region when only the decoder has access to $\{S(k)\}$. (v) The result in [13] showed that the rate-distortion function with decoder only side-information could be written as a minimization of $I(X;W|S)$, where $W \leftrightarrow X \leftrightarrow S$ is an auxilary random variable. In [5] the conditional rate distortion function is shown to be a minimization over $I(X;\hat{X}|S)$, where $\hat{X}$ is the reproduction. Though the minimizations are over different domains, the expressions have a striking analogy as observed in [13]. It is unclear if the analogy between (8) and (10) are fundamental since we do not prove that they are the optimal rate-regions. However, for the Gaussian two-description case, it turns out to be so as shown in Section IV.

The coding technique used to show the achievability of Theorem III.2 can be extended to the case where there are different side information as depicted in Figure 1. This can be done by realizing that the binning strategy used to prove Theorem III.2 is *universal* in that it can work even though the side-information is different [6]. In order to present the simplest results, we examine the case when the side-information between the central and side decoders is "degraded", *i.e.,* $X \leftrightarrow S_{12} \leftrightarrow S_t,\ t=1,2$. This might be reasonable since the central decoder has access to the side information of the decoders 1 & 2, and perhaps more. However, for the achievability idea, this is not a restriction. We can easily generalize the following achievability result when this is not the case, using the private/common information technique of [6].

Theorem III.3 *Let* $(X(1),S_1(1),S_2(1),S_{12}(1)),\ldots$ *be a sequence of i.i.d. finite alphabet random variables drawn according to probability mass function* $Q(x,s_1,s_2,s_{12})$. *Let the encoder and decoder, defined as in (2) and (7), where* only *the decoder has access to the side informations* $\{S_1(k),S_2(k),S_{12}(k)\}$. *Moreover, let the side-informations be degraded,* i.e., $p(x,s_t,s_{12}) = p(x)p(s_{12}|x)p(s_t|s_{12}), t=1,2$. *Then any tuple* (R_1,R_2,D_1,D_2,D_{12}) *is achievable if there exist random variables* (W_0,W_1,W_2,W_{12}) *with probability mass function* $p(x,s_1,s_2,s_{12},w_0,w_1,w_2,w_{12}) = Q(x,s_1,s_2,s_{12})p(w_0,w_1,w_2,w_{12}|x)$, *such that*

$$\begin{aligned} R_1 &> \max\{I(X;W_0|S_1),I(X;W_0|S_2))\} \\ &\quad +I(X;W_1|S_1,W_0) \\ R_2 &> \max\{I(X;W_0|S_1),I(X;W_0|S_2)\} \\ &\quad +I(X;W_2|S_2,W_0), \\ R_1+R_2 &> 2\max\{I(X;W_0|S_1),I(X;W_0|S_2\} \\ &+ I(X;W_{12}|W_1,W_2,W_0,S_{12})+I(X;W_1,W_2|W_0) \\ &+ I(W_1;W_2|W_0)-I(W_1;S_1|W_0)-I(W_2;S_2|W_0) \end{aligned} \tag{12}$$

and there exist reconstruction functions f_1, f_2, f_{12} *which satisfy*

$$\begin{aligned} D_1 &\geq \mathbb{E}[d_1(X,f_1(S_1,W_1,W_0))] \\ D_2 &\geq \mathbb{E}[d_2(X,f_2(S_2,W_2,W_0))] \\ D_{12} &\geq \mathbb{E}[d_{12}(X,f_{12}(S_{12},W_{12},W_1,W_2,W_0))]. \end{aligned} \tag{13}$$

■

IV. THE GAUSSIAN CASE

In this section we focus on the Gaussian case where the side information is common, *i.e.* $S_1(k) = S_2(k) = S_{12}(k) \stackrel{def}{=} S(k)$. Let $\{X(k),S(k)\}$ be a sequence of i.i.d. jointly Gaussian distributed random variables then with no loss of generality we can write,

$$S(k) = \alpha\left[X(k)+U(k)\right], \tag{14}$$

where $\alpha > 0$ and $\{X(k)\}, \{U(k)\}$ are independent Gaussian random variates with $\mathbb{E}[X]=0=\mathbb{E}[U]$, $\mathbb{E}[X^2]=\sigma_X^2$, $\mathbb{E}[U^2]=\sigma_U^2$. Note that the model assumes that $\{X(k)\}, \{U(k)\}$ are i.i.d. sequences.

First, we apply this model to the result in Theorem III.1 (see also Remark (iii) above). In this case, we assume that $\hat{X}_0 = 0$, and $\hat{X}_{12} = \psi_{12}(\hat{X}_1,\hat{X}_2,S)$, for some deterministic function ψ_{12}, *i.e.,* the central reconstruction $\hat{X}_{12}$ is completely determined by the side reconstructions $\hat{X}_1,\hat{X}_2$ and side information S. Note that the best MMSE estimate of X given S is,

$$\hat{X}_{\mathrm{MMSE}|S=s} = \mathbb{E}[X|S=s] = \frac{1}{\alpha}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2}\right)s, \tag{15}$$

and the MMSE estimation error (or the "innovation") is,

$$\begin{aligned} \mathcal{F} &= X-\hat{X}_{\mathrm{MMSE}|S=s} = X-\frac{1}{\alpha}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_U^2}\right)s \\ \sigma_{\mathcal{F}}^2 &\stackrel{def}{=} \mathbb{E}[|\mathcal{F}|^2] = \frac{\sigma_X^2\sigma_U^2}{\sigma_X^2+\sigma_U^2}. \end{aligned} \tag{16}$$

Therefore, consider the following joint distribution for $X,\hat{X}_1,\hat{X}_2,\hat{X}_{12}$,

$$\begin{aligned} \hat{X}_t &= \underbrace{\beta_t\mathcal{F}+N_t}_{V_t}+\hat{X}_{\mathrm{MMSE}|S=s},\ t=1,2 \\ \hat{X}_{12} &= \mathbb{E}[\mathcal{F}|V_1,V_2]+\hat{X}_{\mathrm{MMSE}|S=s}, \end{aligned} \tag{17}$$

where N_1, N_2 are independent of X, S and are zero-mean Gaussian random variables, with $|\rho| < 1$, and

$$\begin{aligned}\beta_1 &= 1 - \frac{D_1}{\sigma_{\mathcal{F}}^2}, \ \ \beta_2 = 1 - \frac{D_2}{\sigma_{\mathcal{F}}^2}, \ \ \mathbb{E}[N_1^2] = \beta_1 D_1 \quad (18)\\ \mathbb{E}[N_2^2] &= \beta_2 D_2, \ \ \mathbb{E}[N_1 N_2] = \rho\sqrt{\beta_1 D_1 \beta_2 D_2}.\end{aligned}$$

Using this in Theorem III.1 we get the following achievability results: $\mathbb{E}(X - \hat{X}_1)^2 = D_1$, $\mathbb{E}(X - \hat{X}_2)^2 = D_2$, and $D_{12} = \mathbb{E}(X - \hat{X}_{12})^2$ is given by

$$\sigma_{\mathcal{F}}^2 \frac{\tilde{D}_1 \tilde{D}_2 (1-\rho^2)}{\tilde{D}_1 + \tilde{D}_2 - \tilde{D}_1\tilde{D}_2(1+\rho^2) - 2\rho\sqrt{\tilde{D}_1\tilde{D}_2(1-\tilde{D}_1)(1-\tilde{D}_2)}}$$

where $\tilde{D}_1 = \frac{D_1}{\sigma_{\mathcal{F}}^2}, \tilde{D}_2 = \frac{D_2}{\sigma_{\mathcal{F}}^2}$. The rates in Theorem III.1 yield,

$$\begin{aligned}R_1 &> \frac{1}{2}\log\left(\frac{\sigma_{\mathcal{F}}^2}{D_1}\right), \ \ R_2 > \frac{1}{2}\log\left(\frac{\sigma_{\mathcal{F}}^2}{D_2}\right), \quad (19)\\ R_1 + R_2 &> \frac{1}{2}\log\left(\frac{1}{\tilde{D}_1\tilde{D}_2(1-\rho^2)}\right)\end{aligned}$$

This implies that, for given $\epsilon > 0$, the following tuple $(R_1, R_2, D_1, D_2, D_{12})$ is achievable:

$$\begin{aligned}D_1 &> \sigma_{\mathcal{F}}^2 e^{-2R_1}, \ D_2 > \sigma_{\mathcal{F}}^2 e^{-2R_2}, \quad (20)\\ D_{12} &> \frac{\sigma_{\mathcal{F}}^2 e^{-2(R_1+R_2)}}{1 - (\sqrt{\tilde{\Pi}} - \sqrt{\tilde{\Delta}})^2}\end{aligned}$$

where

$$\begin{aligned}\tilde{\Pi} &= \left(1 - \frac{D_1}{\sigma_{\mathcal{F}}^2}\right)\left(1 - \frac{D_2}{\sigma_{\mathcal{F}}^2}\right), \quad (21)\\ \tilde{\Delta} &= \left(\frac{D_1}{\sigma_{\mathcal{F}}^2}\right)\left(\frac{D_2}{\sigma_{\mathcal{F}}^2}\right) - e^{-2(R_1+R_2)}\end{aligned}$$

In this case we can show along similar lines to Ozarow's proof for the two-description multiple description problem for Gaussian sources (see [7], Theorem 1) that the region defined in (20) is tight. Therefore we can state the following result

Theorem IV.1 *Let* $(X(1), S(1)), (X(2), S(2))\ldots$ *be a sequence of i.i.d jointly Gaussian random variables. Both the encoder and decoder have access to the side information* $\{S(k)\}$. *If the distortion measures are* $d_m(x, \hat{x}_m) = ||x - \hat{x}_m||^2, m = 1, 2, 12$ *then the set of all achievable tuples* $(R_1, R_2, D_1, D_2, D_{12})$ *are given in (20).*

■

Now we connect this result with our problem of interest, *i.e.,* when only the decoder has access to the side information. In the theorem III.2, let us use $W_0 = 0$, $W_{12} = \psi(W_1, W_2)$, for some deterministic function ψ. Let us define the following reconstruction functions,

$$\hat{X}_t = W_t + \frac{1}{\alpha}\left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2}\right)\left(\frac{D_t}{\sigma_{\mathcal{F}}^2}\right)S, \ \ t = 1, 2 \quad (22)$$

where the auxillary random variables W_1, W_2 are defined as

$$W_t = \left(1 - \frac{D_t}{\sigma_{\mathcal{F}}^2}\right)X + N_t, \ \ t = 1, 2 \quad (23)$$

with N_1, N_2 being zero-mean Gaussian random variables which are the same as defined in (18). We finally define the reconstruction $\hat{X}_{12}$ as,

$$\begin{aligned}\hat{X}_{12} &= \underbrace{c_1 W_1 + c_2 W_2}_{W_{12}} + \quad (24)\\ &\quad + \frac{1}{\alpha}\left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2}\right)(1 - c_1\beta_1 - c_2\beta_2)\,S,\end{aligned}$$

where β_1, β_2 are defined in (18), $[c_1, c_2] = [\beta_1, \beta_2]\mathbf{E}^{-1}$, with $\mathbf{E}$ is equal to

$$\begin{bmatrix} \beta_1 & \beta_1\beta_2 + \rho\sqrt{\beta_1\tilde{D}_1\beta_2\tilde{D}_2} \\ \beta_1\beta_2 + \rho\sqrt{\beta_1\tilde{D}_1\beta_2\tilde{D}_2} & \beta_2 \end{bmatrix}, \quad (25)$$

and as before $\tilde{D}_1 = \frac{D_1}{\sigma_{\mathcal{F}}^2}, \tilde{D}_2 = \frac{D_2}{\sigma_{\mathcal{F}}^2}$. Note that these definitions satisfy the Markov requirement in theorem III.2, *i.e.,* $S \leftrightarrow X \leftrightarrow (W_1, W_2, W_{12})$.

Moreover, due to (22) and (24), $I(X; W_1|\hat{X}_1, S)$, $I(X; W_2|\hat{X}_2, S)$, $I(W_1; W_2|\hat{X}_2, S)$, $I(\hat{X}_2; W_1|\hat{X}_1, S)$ and $I(X; W_1, W_2, W_{12}|\hat{X}_1, \hat{X}_2, \hat{X}_{12}, S)$ all are zero. This along with Theorem III.2 implies that, we can achieve the same rate-distortion tuple as (20). Now from Remark (iv), it is clear that the inner bound to the rate-distortion region of the decoder-only side information is the rate-distortion region when we have *both* encoder and decoder side information. Hence the inner bound in the Gaussian case is given by the region specified in Theorem IV.1. Therefore we can show that the rate region in Theorem III.2 is the same as Theorem IV.1 and is stated below for completeness.

Theorem IV.2 *Let* $(X(1), S(1)), (X(2), S(2))\ldots$ *be a sequence of i.i.d jointly Gaussian random variables. Let only the decoder have access to the side information* $\{S(k)\}$. *If the distortion measures are* $d_m(x, \hat{x}_m) = ||x - \hat{x}_m||^2, m = 1, 2, 12$ *then the set of all achievable tuples* $(R_1, R_2, D_1, D_2, D_{12})$ *are given by (20).*

■

The result in Theorem IV.2 is composed of two components. The first component is that even with side information (both at the encoder and decoder) the rate-distortion region two-description Gaussian multiple description problem can be solved. This is quite natural given that the problem was solved for the case where there was no side-information by El-Gamal and Cover [4] and Ozarow [7]. The second component arises by manipulating the reconstruction functions used in Theorem IV.1, *i.e.,* see (17). This manipulation is shown diagramatically in Figures 3, 4 which are equivalent representations. This property is an extension of the equivalence observed in Figures 3 and 4 of [12].

V. Discussion

Our interest originally arose in investigating robust compression of correlated time-series such as video. Many single description compression systems for video as well as speech need to adapt to the characteristics of the signal and the adaptation information (such as motion vectors) is conveyed to the decoder. In a multiple description framework, it has never been clear how such adaptation information should be handled, or what the associated rate penalty should be. Recently there has been a connection between encoding of such signals with the Wyner-Ziv problem [9]. This allowed design of single description compression schemes where explicit motion information did not have to be sent

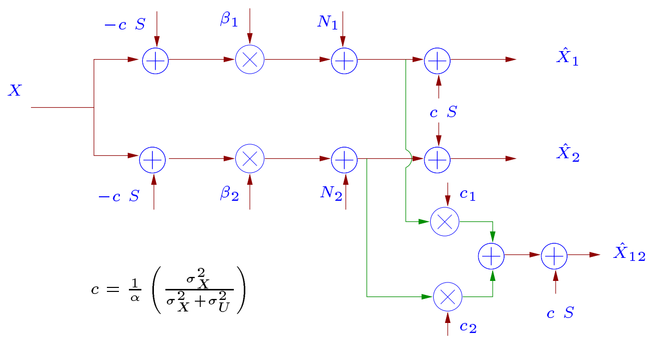

Figure 3: Reconstruction functions for Gaussian source with encoder and decoder side information.

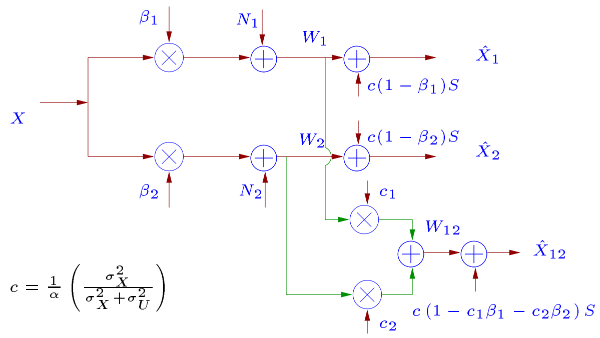

Figure 4: Equivalent to Figure 3 where the reconstructions have been redrawn. This is used to construct the auxillary random variables W_1, W_2, W_{12} required in Theorem III.2 giving the construction stated in (22),(23) and (24).

and motion was inherently captured by using receiver side information. This connection motivated us to look at multiple-description video from this context and therefore the problem formulation studied in this paper. Another motivation to study this problem is in the context of robust distributed compression applications to sensor networks.

A. Proof of Theorem III.2

Due to lack of space we give only the proof outline for Theorem III.2. This can be easily generalized to give the achievability proof for Theorem III.3.

We denote the set of *strongly-typical* sequences by $T_\epsilon^{(n)}$ [3]. The proof as in [13], hinges on the so-called Markov lemma [2].

Let W_0, W_1, W_2, W_{12} satisfy the conditions of Theorem III.2. For brevity of notation, we will denote $\lfloor e^{nA} \rfloor$ by e^{nA} and also doubly use the notation to mean the set $\{1, \ldots, e^{nA}\}$ where the context should make the meaning clear. Also we sometimes drop the indices on the codewords where it is understood that the appropriate indices (i_0, i_1, i_2, i_{12}) can be inserted.

Codebook generation: Choose e^{nA_0} codewords W_0^n independently according to a uniform distribution over the set $T_\epsilon^{(n)}(W_0)$. For each W_0^n draw a collection of n-vectors W_1^n according to the uniform distribution over the set $T_\epsilon^{(n)}(W_1|W_0^n)$. Similarly we draw e^{nA_2} vectors W_2^n from $T_\epsilon^{(n)}(W_2|W_0^n)$. Finally, choose $e^{nA_{12}}$ codewords $\{W_{12}^n\}$ independently according to a uniform distribution over the set $T_\epsilon^{(n)}(W_{12}|W_0^n, W_1^n, W_2^n)$, for every jointly typical $\{W_0^n, W_1^n, W_2^n\}$.

Now we bin all the above codewords. That is, randomly assign $i_0 \in e^{nA_0}$ into one of e^{nB_0} bins using a uniform distribution over the bins. Denote by $\mathcal{B}_0(b_0)$ the indices assigned to bin b_0. Similarly, for each $i_0 \in e^{nA_0}$ we throw $i_t \in e^{nA_t}, t = 1, 2$ into one of $e^{nB_t}, t = 1, 2$ bins using a uniform distribution over the bins (*i.e.*, we conditionally bin W_1^n, W_2^n). We denote by $\mathcal{B}_t(b_t), t = 1, 2$ the indices assigned to bin b_t. Finally for each (i_0, i_1, i_2) we randomly assign $i_{12} \in e^{nA_{12}}$ into one of $e^{nB_{12}}$ bins again using a uniform distribution over the bins. We denote by $\mathcal{B}_{12}(b_{12})$ the indices assigned to bin b_{12}. Note that $|\mathcal{B}_t(b_t)| \doteq e^{n(A_t - B_t)}, t = 0, 1, 2, 12$.

Encoding: Given x^n find if possible (i_0, i_1, i_2, i_{12}) such that,

$$\{x^n, W_0^n(i_0), W_1^n(i_0, i_1), W_2^n(i_0, i_2), W_{12}^n(i_0, i_1, i_2, i_{12})\} \in T_\epsilon^{(n)} \tag{26}$$

If this is not possible, set $(b_0, b_1, b_2, b_{12}) = (0, 0, 0, 0)$. If there exists (i_0, i_1, i_2, i_{12}), such that (26) is satisfied, find (b_0, b_1, b_2, b_{12}) which contain the chosen indices. Split b_{12} into two parts $(b_{12}(1), b_{12}(2))$ (like in [4]), such that the pair together specifies b_{12}. Let $b_{12}(t) \in e^{nB_{12}(t)}, t = 1, 2$, and $B_{12}(1) + B_{12}(2) = B_{12}$. Send $\{b_0, b_1, b_{12}(1)\}$ over channel 1 and $\{b_0, b_2, b_{12}(2)\}$ over channel 2 (see figure 2). Hence the transmission rates are

$$R_1 = B_0 + B_1 + B_{12}(1), \quad R_2 = B_0 + B_2 + B_{12}(2). \tag{27}$$

Decoder: The decoders have access to the side information s^n. The decoder 1 uses (b_0, b_1) to first find $W_0^n(i_0) \in \mathcal{B}_0(b_0)$ such that $\{W_0^n(i_0), s^n\} \in T_\epsilon^{(n)}$. If such a *unique* i_0 does not exist, then the decoder declares an error, *i.e.*, it sets $\hat{X}_1^n$ to an arbitrary sequence in $\mathcal{X}_1^n$. If such a unique i_0 exists, then the decoder attempts to find $W_1^n(i_0, i_1) \in \mathcal{B}_1(b_1)$, such that $\{W_1^n, W_0^n, s^n\} \in T_\epsilon^{(n)}$. Again if such a unique i_1 is not found the decoder declares an error. Finally if the decoder does not declare an error it calculates $\hat{X}_1^n = f_1(W_0^n, W_1^n, s^n)$. Decoder 2 operates similarly and attempts to find $\hat{X}_2^n = f_2(W_0^n, W_2^n; s^n)$. Finally, Decoder 12, has access to $\{b_0, b_1, b_2, b_{12}\}$, and therefore attempts to find W_0^n, W_1^n, W_2^n like in decoders 1 and 2, and declares an error if that fails. If such a unique $\{i_0, i_1, i_2\}$ has been found, it then attempts to find a unique i_{12} such that $W_{12}^n \in \mathcal{B}_{12}(b_{12})$ and $\{s^n, W_0^n, W_1^n, W_2^n, W_{12}^n\} \in T_\epsilon^{(n)}$. If such a unique i_{12} is not found it declares an error. If a unique i_{12} is found, then it forms $\hat{X}_{12}^n = f_{12}(W_0^n, W_1^n, W_2^n, W_{12}^n, s^n)$.

Error analysis: As usual, we can see that the average distortion is $D_m + \epsilon' + P_e d_{\max}, m = 1, 2, 12$, where $\epsilon' > 0$ can be made arbitrarily small, d_{max} is the maximal distortion of $d_m(x, \hat{x}_m)$, over all m, and P_e is the probability that we have either an encoder or a decoder error. Now we bound the set sizes for P_e to go to zero. First we start with the encoder error events, and the set sizes can be bounded by the same technique as used in [4, 14] since the encoder works without side information. This yields,

$$\begin{aligned} A_0 &> I(X; W_0), \quad A_t > I(X; W_t|W_0), \ t = 1, 2, \\ A_1 + A_2 &> I(X; W_1, W_2|W_0) + I(W_1; W_2|W_0) \\ A_{12} &> I(X; W_{12}|W_0, W_1, W_2). \end{aligned} \tag{28}$$

Now for the decoder error events. Notice that the codebook generation and encoding ensures that the conditions of the Markov lemma [3] is satisfied for $S \leftrightarrow X \leftrightarrow (W_0, W_1, W_2, W_{12})$. Therefore, it is easy to see that the for correct index (i_0, i_1, i_2, i_{12}), which satisfies (26), $\{s^n, W_0, W_1, W_2, W_{12}\} \in T_\epsilon^{(n)}$ with probability going to 1 asymptotically. Therefore, the only events left to analyse is when there exists spurious indices $(i_0^{'}, i_1^{'}, i_2^{'}, i_{12}^{'})$ which satisfy $\{s^n, W_0, W_1, W_2, W_{12}\} \in T_\epsilon^{(n)}$. For decoders 1,2 by using an argument similar to [13], it can be shown that this error vanishes asymptotically[1] if,

$$B_0 > A_0 - I(S;W_0), \;\; B_t > A_t - I(S;W_t|W_0), \;\; t = 1,2. \quad (29)$$

For the central decoder the only two additional events to analyze are,

$$F_3^{'} : \exists (i_1^{'}, i_2^{'}) \neq (i_1, i_2) \text{ s.t. } W_t^n(i_0, i_t^{'}) \in \mathcal{B}_t(b_t), t = 1,2, \quad (30)$$
$$\text{and } \{s^n, W_0^n(i_0), W_1^n(i_0, i_1^{'}), W_2^n(i_0, i_2^{'})\} \in T_\epsilon^{(n)}.$$
$$F_4^{'} : \exists i_{12}^{'} \neq i_{12} \text{ s.t. } W_{12}^n(i_0, i_1, i_2, i_{12}^{'}) \in \mathcal{B}_t(b_{12}) \text{ and}$$
$$\{s^n, W_0^n(i_0), W_1^n(i_0, i_1), W_2^n(i_0, i_2), W_{12}^n(i_0, i_1, i_2, i_{12}^{'})\} \in T_\epsilon^{(n)}.$$

Using typicality arguments, we can show the following,

$$\mathbb{P}(F_3^{'}) \to 0, \text{ if } B_t > A_t - I(S;W_t|W_0), t = 1,2 \text{ and} \quad (31)$$
$$B_1 + B_2 > A_1 + A_2 - \{I(S;W_1,W_2|W_0) + I(W_1,W_2|W_0)\},$$
$$\mathbb{P}(F_4^{'}) \to 0, \text{ if } B_{12} > A_{12} - I(S;W_{12}|W_0,W_1,W_2).$$

Since for any random variables Z, W_1, W_2, W_0,

$$I(W_1;W_2|W_0) + I(Z;W_1,W_2|W_0) - I(W_1;Z|W_0) \quad (32)$$
$$- I(W_2;Z|W_0) = I(W_1;W_2|Z,W_0),$$

therefore,

$$I(W_1;Z|W_0) + I(W_2;Z|W_0) \leq I(W_1;W_2|W_0) + I(Z;W_1,W_2|W_0) \quad (33)$$

From (29) we get,

$$B_1 + B_2 > A_1 + A_2 - \{I(S;W_1|W_0) + I(S;W_2|W_0)\}, \quad (34)$$

and from (31) we get

$$B_1 + B_2 > A_1 + A_2 - \{I(Y;W_1,W_2|W_0) + I(W_1,W_2|W_0)\}. \quad (35)$$

But due to (33), it is clear that (35) is dominated by (34). Moreover since the other bounds arising from $\mathbb{P}(F_3^{'}) \to 0$ are contained in (29) we can eliminate the rate bounds arising from $\mathbb{P}(F_3^{'}) \to 0$. Therefore, we get the following set of rate bounds.

$$\begin{aligned} A_0 &> I(X;W_0) \quad (36)\\ A_t &> I(X;W_t|W_0), \;\; t = 1,2\\ A_1 + A_2 &> I(X;W_1,W_2|W_0) + I(W_1;W_2|W_0)\\ A_{12} &> I(X;W_{12}|W_0,W_1,W_2)\\ B_0 &> A_0 - I(S;W_0)\\ B_t &> A_t - I(S;W_t|W_0), \;\; t = 1,2\\ B_{12} &> A_{12} - I(S;W_{12}|W_0,W_1,W_2) \end{aligned}$$

Using (36) together with (27) we can show that the following rate region is achievable ($t = 1,2$).

$$R_t > I(X;W_t|W_0) - I(S;W_t|W_0) + I(X;W_0) - I(S;W_0), \quad (37)$$
$$R_1 + R_2 > 2[I(X;W_0) - I(S;W_0)] + I(X;W_1,W_2|W_0)$$
$$+ I(W_1;W_2|W_0) - \{I(S;W_1|W_0) + I(S;W_2|W_0)\}$$
$$+ I(X;W_{12}|W_0,W_1,W_2) - I(S;W_{12}|W_0,W_1,W_2)$$

[1] These events get modified in the proof of Theorem III.3, but the universality of the binning makes it simultaneously good for S_1, S_2, S_{12} so long as appropriate rate bounds are satisfied.

Now since $S \leftrightarrow X \leftrightarrow (W_0, W_1, W_2, W_{12})$, just as in [13] we can show that,

$$\begin{aligned} &I(X;W_0) - I(S;W_0) = I(X;W_0|S) \quad (38)\\ &I(X;W_{12}|W_0,W_1,W_2) - I(S;W_{12}|W_0,W_1,W_2)\\ &= I(X;W_{12}|W_0,W_1,W_2,S)\\ &I(X;W_t|W_0) - I(S;W_t|W_0) = I(X;W_t|W_0,S), \;\; t = 1,2\\ &I(X;W_1,W_2|W_0) - I(S;W_1,W_2|W_0) = I(X;W_1,W_2|W_0,S) \end{aligned}$$

Moreover, as in (32),

$$\begin{aligned} I(S;W_1,W_2|W_0) &+ I(W_1;W_2|W_0) - I(S;W_1|W_0) \quad (39)\\ &- I(S;W_2|W_0) = I(W_1;W_2|S,W_0) \end{aligned}$$

and thus combining (39) and the last relationship in (38) we get

$$I(X;W_1,W_2|W_0) + I(W_1;W_2|W_0) - I(S;W_1|W_0) \quad (40)$$
$$- I(S;W_2|W_0) = I(W_1;W_2|S,W_0) + I(X;W_1,W_2|W_0,S)$$

Inserting (38) and (40) into (37) we get the result in Theorem III.2.

REFERENCES

[1] R. Ahlswede. The rate distortion region for multiple descriptions without excess rate . *IEEE Transactions on Information Theory*, 31(6):721–726, November 1985.

[2] T. Berger. Multiterminal source coding. In G. Longo, editor, *The information theory approach to communications*, pages 172–231. Springer-Verlag, 1977.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., New York, 1991.

[4] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inform. Th.*, IT-28:851–857, November 1982.

[5] R M. Gray. Conditional rate-distortion theory. Technical Report # 6502-2, Information Systems Laboratory,, Stanford University, October 1972.

[6] Chris Heegard and Toby Berger. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory*, 31(6):727–734, November 1985.

[7] L. Ozarow. On a source coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59:1909–1921, December 1980.

[8] S S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed compression in a dense microsensor network. *IEEE Signal Processing Magazine*, 2:51–60, March 2002.

[9] R. Puri and K. Ramchandran. PRISM: an uplink-friendly multimedia coding paradigm. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, pages 856–859, 2003.

[10] David Slepian and Jack K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

[11] Yossef Steinberg and Neri Merhav. On successive refinement for the Wyner-Ziv problem. In *Proc. IEEE International Symposium on Information Theory (ISIT), Chicago, Illinois*, page 362, July 2004.

[12] A D. Wyner. The rate distortion function for source coding with side information at the decoder – II: General sources. *Information and Control*, 38:60–80, July 1978.

[13] A D. Wyner and J. Ziv. The rate distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, January 1976.

[14] Z. Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Trans. Inform. Th.*, IT-33:502–521, July 1987.