Multiuser Successive Refinement and Multiple Description Coding

Chao Tian, Jun Chen and Suhas N. Diggavi

Abstract

We consider the multiuser successive refinement (MSR) problem, where the users are connected to a central server via links with different noiseless capacities, and each user wishes to reconstruct in a successive-refinement fashion. An achievable region is given for the two-user two-layer case and it provides the complete rate-distortion region for the Gaussian source under the MSE distortion measure. The key observation is that this problem includes multiple description (MD) coding as a subsystem, and the techniques useful in the MD problem can be extended to this case. We show that the coding scheme based on the universality of random binning is sub-optimal, because multiple Gaussian side informations only at the decoders do incur performance loss, in contrast to the case of single side information at the decoder. We further show that unlike the single user case, when there are multiple users, the loss of performance by a multistage coding approach can be unbounded for the Gaussian source. The result suggests that in such a setting, the benefit of using successive refinement is not likely to justify the accompanying performance loss. The MSR problem is also related to the source coding problem where each decoder has its individual side information, while the encoder has the complete set of the side informations. The MSR problem further includes several variations of the MD problem, for which the specialization of the general result is investigated and the implication is discussed.

I. Introduction

Multiuser information theory has attracted much attention recently because of the growth in the complexity and capability of the practical communication networks. In this work, we consider the multiuser successive refinement (MSR) problem formulated by Pradhan and Ramchandran in [1]. In this problem, a server is to provide multimedia data to users connected to the server through channels with different (noiseless) capacities, e.g., a dial-up connection vs. a high-speed cable connection. The server performs the transmission in a broadcasting manner in order to reduce operating cost, and thus the users with bad channels will only receive a (known) subset of the bitstream, while the users with good channels will be able to receive the complete bitstream. Furthermore, to reduce the delay for each user, the server would also like to provide the bitstream in a successive refinement fashion user-wise. The "multiusers" in the MSR problem thus receive degraded message sets, while the "successive refinement" refers to the fact that there are multiple rounds (layers) of such transmissions.

A diagram is given in Fig. 1 for a system with two users and two layers. We will assume the user with good channel connection will remain so for the complete transmission, however the exact rates $R_{11}, R_{12}, R_{21}, R_{22}$ can

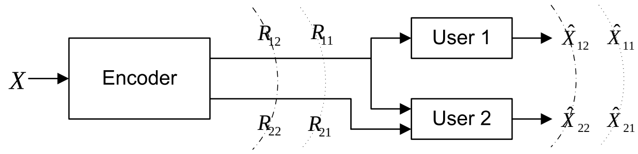

Fig. 1. A system diagram with two users and two layers. The “good” user (user 2) receives the complete message while the “bad” user (user 1) only receive a subset of the message. There are two rounds of transmissions, illustrated by the dotted curve and dash-dot curve, and in each round the users form their reconstructions accordingly.

vary. If the channel capacity is fixed during the transmission, then $\frac{R_{11}}{R_{21}} = \frac{R_{12}}{R_{22}}$, which is a special case of this general setting. We will only consider the two-user two-layer system in this work.

The notion of successive refinement of information in the single user setting was introduced by Koshelev [2] and by Equitz and Cover [3] (see also [4]), and the problem is well researched. The main question is whether the requirement of encoding a source progressively necessitates a higher rate than encoding without the progressive requirement. A source is successively refinable if encoding in multiple stages incurs no rate loss as compared with optimal rate-distortion encoding at the separate distortion levels. The reassuring result by Equitz and Cover is that many familiar sources, such as the Gaussian source under the mean squared error (MSE) distortion measure and discrete sources under Hamming distortion measure, are in fact successively refinable. Lastras and Berger [5] further showed that when the source has a real alphabet and the distortion measure is MSE, even when it is not successively refinable, the rate loss is bounded by a fixed constant.

Naturally, in the multiuser scenario we are interested in answering whether such a progressive coding requirement causes any performance loss, and if so, whether the loss is bounded. In this work, we provide an achievable rate-distortion region for the problem with two users and two layers by embedding a multiple description problem inside it, and show that this region is tight for the Gaussian case. Furthermore, the loss of performance to a single layer coding can indeed be **unbounded**, which suggests unless there is a significant reason calling for a progressive coding, the loss of performance makes it a less attractive approach.

The MSR problem includes the multiple description (MD) problem as a subsystem, and the techniques in the MD literature (notably [6] and [7]), are our main tools in this work. We show that the coding scheme given in [1] based on random binning is sub-optimal, because multiple Gaussian side informations only at the decoders incur performance loss compared to the side information also available at the encoder; this is in contrast to the case of single side information at the decoder, where there is no essential loss [8]. The MSR problem is also related to the problem considered in [9], where each decoder has its individual side information, and the encoder has the complete set of side informations. The MSR problem further includes several variations of the MD problem, such as the MD problem with central refinement (MDCR) [10], as well as the conditional MD problem. We will discuss the MDCR problem in detail and reveal the implication of the general result from MSR.

The distortion-rate (D-R) region and the rate-distortion (R-D) region given for the Gaussian MSR problem can be easily reduced to that for the MD problem. Though the Gaussian MD region has been known for more than

TABLE I

THE CORRESPONDENCE BETWEEN SYSTEM OF FIG. 1 AND THAT IN FIG. 2.

double subscript	(11)	(21)	(12)	(22)
single subscript	1	2	3	4

25 years, as pointed out in [11], the expressions given in the literature are usually not complete (even incorrect if being used without caution) and we hope this confusion can be clarified by this work[1].

It is worth pointing out that the formulation considered in this work is from the source coding point of view, which implies a coding system where source and channel coding are separated. However, it is well known that under the degraded broadcast channel, for which the considered coding approach is arguably the most suitable, a source-channel separation approach is not optimal (see, for example [12]). It is nevertheless useful to consider the current formulation, since firstly joint source-channel coding (JSCC) schemes are often more complex, and secondly the performance using source-channel separation can be used to compare with that of any JSCC schemes to measure the possible performance loss.

The rest of the paper is organized as follows. In Section II the problem is formally defined and some related background is given. An achievable region is given in Section III. In Section IV, we prove that the given achievable region is tight for the Gaussian source, then analyze the performance loss comparing with single layer coding and discuss a special case with fixed channel configuration. Section V discusses the MDCR problem as the a special case of the problem being treated and Section VI concludes the paper.

II. PROBLEM DEFINITION

Let $\mathcal{X}$ be a finite set and let $\mathcal{X}^n$ be the set of all n-vectors with components in $\mathcal{X}$. Denote an arbitrary member of $\mathcal{X}^n$ as $x^n = (x_1, x_2, ..., x_n)$, or alternatively as $\boldsymbol{x}$; $(x_i, x_2, ..., x_j)$ will also be written as $x_{i,\dots,j}$. Upper case is used for random variables and vectors. A discrete memoryless source (DMS) $(\mathcal{X}, P_X)$ is an infinite sequence $\{X_i\}_{i=1}^{\infty}$ of independent copies of a random variable X in $\mathcal{X}$ with a generic distribution P_X with $P_X(x^n) = \prod_{i=1}^{n} P_X(x_i)$. Let $\hat{\mathcal{X}}$ be a finite reconstruction alphabet, and for simplicity we assume that the decoders all use this reconstruction alphabet. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$ be a distortion measure. The single-letter distortion of a vector is defined as

$$d(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i), \quad \forall \boldsymbol{x} \in \mathcal{X}^n, \quad \hat{\boldsymbol{x}} \in \hat{\mathcal{X}}^n. \tag{1}$$

Instead of directly considering the system depicted in Fig. 1, we consider the equivalent system given in Fig. 2. The reformulation is crucial, which makes the rather involved relations between descriptions more explicit. The double subscript in Fig. 1 is simplified to single subscript, whose correspondence is made clear in Table I.

[1]Feng and Effros clarified the Gaussian MD region in terms of R-D characterization in [11], but the interpretation of the degenerate region was not made explicit in their work.

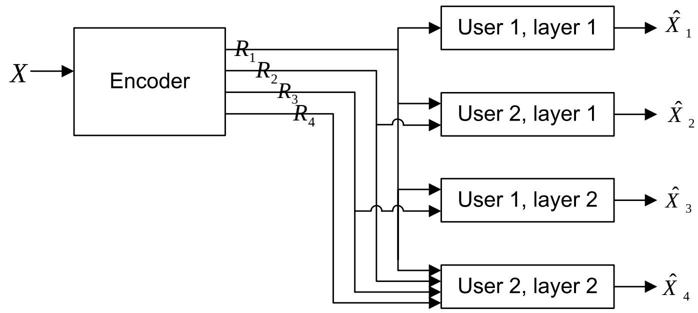

Fig. 2. The equivalent system diagram to the two-user two-layer system depicted in Fig. 1.

***Definition** 1:* An $(n, M_{1...4}, D_{1...4})$ MSR code for source X consists of 4 encoding functions ϕ_i and 4 decoding functions ψ_i, $i = 1, 2, 3, 4$:

$$\phi_i \quad : \quad \mathcal{X}^n \rightarrow I_{M_i},\ i = 1, 2, 3, 4$$

where $I_k = \{1, 2, ..., k\}$ and

$$\psi_1 : I_{M_1} \rightarrow \hat{\mathcal{X}}^n, \quad \psi_2 : I_{M_1} \times I_{M_2} \rightarrow \hat{\mathcal{X}}^n, \quad \psi_3 : I_{M_1} \times I_{M_3} \rightarrow \hat{\mathcal{X}}^n, \quad \psi_4 : I_{M_1} \times I_{M_2} \times I_{M_3} \times I_{M_4} \rightarrow \hat{\mathcal{X}}^n,$$

such that

$$\mathbb{E}d(X^n, \psi_1(\phi_1(X^n))) \leq D_1, \quad \mathbb{E}d(X^n, \psi_2(\phi_1(X^n), \phi_2(X^n))) \leq D_2,$$

$$\mathbb{E}d(X^n, \psi_3(\phi_1(X^n), \phi_3(X^n))) \leq D_3, \quad \mathbb{E}d(X^n, \psi_4(\phi_1(X^n), \phi_2(X^n), \phi_3(X^n), \phi_4(X^n))) \leq D_4,$$

where $\mathbb{E}$ is the expectation operation.

For the rest of the paper, we will often refer to the output of the encoding function ϕ_i as description i.

***Definition** 2:* A rate distortion eight-tuple $(R_{1...4}, D_{1...4})$ is said to be achievable, if for any $\epsilon > 0$ and sufficiently large n, there exist an $(n, M_{1...4}, D_1 + \epsilon, D_2 + \epsilon, D_3 + \epsilon, D_4 + \epsilon)$ MSR code, such that $R_i + \epsilon \geq \frac{1}{n} \log(M_i)$ for $i = 1, 2, 3, 4$.

The MSR rate-distortion region, denoted by $\mathcal{Q}$, is the set of all achievable eight-tuples. The problem of characterizing this region is difficult in general, because the problem at hand can be reduced to the well-known multiple description (MD) problem, which is a long standing open problem. In the multiple description problem, one sends two descriptions over two unreliable channels, either of which can break down; the question is to characterize the achievable rate-distortion quintuple consisting of the two description rates and the distortions of individual description, as well as that resulting from the two descriptions jointly. To reduce the MSR problem to the MD problem, we only need to set $R_1 = R_4 = 0$ and $D_1 = \infty$.

The literature on the MD problem is vast (see [13] for a review) and new results are emerging, but the problem remains open. Inner bound exists [7] and it was shown recently that this bound is in fact quite good for source with real alphabet under MSE distortion measure [14]. Nevertheless, the Gaussian source with MSE distortion measure, for which this inner bound is tight, is the only case that the R-D region is completely characterized. Given these facts, our focus will not be on finding a complete solution for the general MSR problem. Instead, we will extend the coding scheme in [7] to give an achievable region, and then focus on the quadratic Gaussian case, for which the achievable region is indeed tight.

We now briefly outline the coding scheme given by El Gamal and Cover in [7] for the MD problem: given joint distribution $P_{XX_1X_2X_3}$, generate two length-n codebooks using the marginals P_{X_1} and P_{X_2}, respectively. It is well known that if approximately $nI(X;X_1)$ and $nI(X;X_2)$ codewords are generated for the two codebooks, respectively, then with high probability we can find codewords X_1^n, respectively X_2^n, jointly typical with any X^n vector in the individual codebook. However, to guarantee the chosen codewords X_1^n, X_2^n are also jointly typical together with X^n, the codebook sizes have to increase. The resulting increased rate is the expense paid to "match" X_1^n and X_2^n. Then for the matched X_1^n and X_2^n vector, a conditional codebook using $P_{X_3|X_1X_2}$ can be further added. The decoders then use the codewords X_1^n, X_2^n and X_3^n as the reconstructions.

III. An Achievable Region

Several schemes were outlined in [1] for the MSR problem which can achieve several specific operating points. One of them is to treat $\hat{X}_1$ and $\hat{X}_2$, which are the reconstructions in the first layer for user 1 and user 2, respectively, as side informations at the decoder and use the Wyner-Ziv binning approach [8] to generate description 3, and also use binning for description 4. The intuition behind this scheme is that the binning strategy has certain "universal" property that whenever the bin is sufficiently small to decode with some side information, it can also be decoded with better quality side information (see [15]). Since $\hat{X}_2$ is a better quality side information than $\hat{X}_1$, user 2 can also decode the description 3, which is meant for user 1; furthermore, user 2 can use it to improve its estimation.

Though the above observation is important, and perhaps provides the insight for the important result on symmetric N description problem [16], it is not optimal for the current problem. Notice that since the receiver 2 has access to both description 1 and description 2, it can also reconstruct $\hat{X}_1$, in addition to $\hat{X}_2$ (which is its desired reconstruction). As such, a conditional codebook on $\hat{X}_1$ is more suitable, since it is available at both the encoder and the decoder.

It should now be clear that MD coding method can be used in MSR, if we treat $\hat{X}_1$ as the common side information available at both the encoder and decoders when encoding $\hat{X}_2$, $\hat{X}_3$ and $\hat{X}_4$. Conditioned on $\hat{X}_1$, two codebooks of size 2^{nR_2} and 2^{nR_3}, respectively, are generated using $P_{\hat{X}_2|\hat{X}_1}$ and $P_{\hat{X}_3|\hat{X}_1}$; as discussed in the last section, in order to find $\hat{X}_2^n$ and $\hat{X}_3^n$ (conditioned on $\hat{X}_1$) jointly typical with X^n (i.e., matched) in these two codebooks with high probability, the codebook sizes should be chosen accordingly. Given the codeword $\hat{X}_1^n$ and the matched codewords $\hat{X}_2^n$ and $\hat{X}_3^n$, in the last coding stage a codeword in the codebook of size 2^{nR_4} generated by $P_{\hat{X}_4|\hat{X}_1\hat{X}_2\hat{X}_3}$ is chosen which is jointly typical with the source vector X^n and the previously chosen codewords $\hat{X}_1^n, \hat{X}_2^n$ and $\hat{X}_3^n$. Now an achievable region is readily available using standard techniques, though it is not clear if it is optimal.

Here we would like to bring attention to a quite subtle and often-overlooked fact: even when $R_1 = 0$, the reduced system is still not the same as an MD system. Notice there are three rates (R_2, R_3, R_4) to characterize here, instead of the two rates in the MD problem. This problem, which we refer to as the multiple descriptions with central refinement (MDCR) problem [10], will be treated in more depth later. Though not the same, the MDCR system is not unfamiliar: the coding scheme in [7] in fact uses such a structure.

Given the discussion above, we next state an achievable region without detailed proof for the sake of brevity[2]. Define the region $\mathcal{Q}_{ach}$ to be the set of all rate distortion eight-tuples $(R_{1...4}, D_{1...4})$ for which there exist four random variables $\hat{X}_1, \hat{X}_2, \hat{X}_3, \hat{X}_4$ in finite alphabet $\hat{\mathcal{X}}$ such that

$$\mathbb{E}d(X, \hat{X}_i) \leq D_i,\ \ i = 1, 2, 3, 4, \tag{2}$$

and the non-negative rate vector satisfies:

$$R_1 \geq I(X; \hat{X}_1), \quad \sum_{i=1,2} R_i \geq I(X; \hat{X}_1\hat{X}_2), \quad \sum_{i=1,3} R_i \geq I(X; \hat{X}_1\hat{X}_3), \tag{3}$$

$$\sum_{i=1,2,3} R_i \geq I(X; \hat{X}_1\hat{X}_2\hat{X}_3) + I(\hat{X}_2; \hat{X}_3|\hat{X}_1), \quad \sum_{i=1,2,3,4} R_i \geq I(X; \hat{X}_1\hat{X}_2\hat{X}_3\hat{X}_4) + I(\hat{X}_2; \hat{X}_3|\hat{X}_1). \tag{4}$$

The following theorem provides an achievable region.

***Theorem** 1:*

$$\mathcal{Q}_{ach} \subseteq \mathcal{Q}.$$

If $R_1 = R_4 = 0$, it is seen that $\mathcal{Q}_{ach}$ degenerates to the achievable region given by El Gamal and Cover in [7]. The region is characterized by a set of sum-rate bounds, instead of the individual rate for (R_1, R_2, R_3, R_4). It is not immediate clear that the aforementioned coding scheme (with individual rates) can achieve the complete region characterized by the sum-rates in Theorem 1. However, a moment of thought reveals that the structure of this system implies that for any code with rates (r_1, r_2, r_3, r_4), we can freely move the rates r_2 (and r_3) into r_1, and rate r_4 into r_1, r_2, r_3 to construct new codes (see [18] for a thorough explanation on the successive refinement problem). Thus indeed the region given in Theorem 1 is achievable.

Though we consider discrete source so far, the results can be generalized to well-behaved continuous sources. In the next section, we prove a converse for the Gaussian source under MSE measure, and show that the region given in Theorem 1 is tight for the Gaussian source. It is worth clarifying that the converse result for the Gaussian MSR problem is not implied by that of the Gaussian source with common side information at both the encoder and the decoders [17], because the optimal first codebook in MSR is not necessarily a codebook generated with any single letter marginal distribution $P_{\hat{X}_1}$, while the common side information is always an i.i.d. random variable in the setting of [17]; moreover, the MSR problem is further complicated by the included MDCR sub-system.

IV. The Gaussian Source

In this section, we focus on the Gaussian source with MSE distortion measure. All the logarithm are natural log.

[2]See also [17] for a similar result for multiple descriptions when both encoder and decoders have access to common side information.

A. The Distortion-Rate Region for the Gaussian Source

Theorem 2 below gives the distortion-rate (D-R) region for the Gaussian source, and Theorem 2′ gives the rate-distortion (R-D) region. Similarly to the MD problem, the D-R region is simpler than the R-D region due to the degenerate regions (see [11] and [19]). It will be illustrated that the R-D region can be established from the D-R region.

We will follow the approach by Ozarow [6] to prove the D-R region. Wang and Viswanath [20] recently introduced an elegant proof technique for the n-description vector-Gaussian sum-rate problem with central and individual distortion constraints, however we have not found it suitable for the MSR problem, mainly due to the addition of the encoding function ϕ_4.

***Theorem** 2:* For the Gaussian source $X \sim \mathcal{N}(0, \sigma_x^2)$ under MSE distortion measure, the achievable distortion-rate region for rates (R_1, R_2, R_3, R_4) is given by

$$\begin{aligned} d_1 \geq \sigma_x^2 \exp[-2R_1], \qquad & d_2 \geq \sigma_x^2 \exp[-2(R_1+R_2)] \\ d_3 \geq \sigma_x^2 \exp[-2(R_1+R_3)], \qquad & d_4 \geq \tfrac{\sigma_x^2 \exp[-2(R_1+R_2+R_3+R_4)]}{1-(|\sqrt{\Pi}-\sqrt{\Delta}|^+)^2} \end{aligned} \tag{5}$$

where $|x|^+ = \max(x, 0)$ and

$$\begin{aligned} d_1^* \triangleq \sigma_x^2 \exp[-2R_1], \quad \hat{d}_2 \triangleq \min(d_2, d_1^*), \quad \hat{d}_3 \triangleq \min(d_3, d_1^*) \\ \Delta \triangleq \frac{\hat{d}_2 \hat{d}_3}{d_1^{*2}} - \exp[-2(R_2+R_3)], \quad \Pi \triangleq (1-\frac{\hat{d}_2}{d_1^*})(1-\frac{\hat{d}_3}{d_1^*}). \end{aligned} \tag{6}$$

Define the function $R(D) = \frac{1}{2}\log\frac{1}{D}$. The following theorem describes the R-D region for the Gaussian source.

Theorem $2'$: For the Gaussian source $X \sim \mathcal{N}(0, \sigma_x^2)$ under MSE distortion measure, the achievable rate-distortion region for distortions (d_1, d_2, d_3, d_4) is given as follows.

$$R_1 \geq R_1^* \triangleq R(\frac{\min(d_1, \sigma_x^2)}{\sigma_x^2}). \tag{7}$$

For any rates $R_1 \geq R_1^*$ and $R_4 \geq 0$, define $\hat{d}_4 \triangleq d_4 \exp(2R_4)$. The achievable rates (R_2, R_3) are given as:

$$R_2 \geq R(\hat{d}_2/d_1^*), \quad R_3 \geq R(\hat{d}_3/d_1^*), \quad R_2 + R_3 \geq \begin{cases} R(\hat{d}_4/d_1^*) & 0 < \hat{d}_4 < \hat{d}_2 + \hat{d}_2 - d_1^*; \\ 0 & \hat{d}_4 > (\hat{d}_2^{-1} + \hat{d}_3^{-1} - (d_1^*)^{-1})^{-1}; \\ R(\hat{d}_4/d_1^*) + L & otherwise, \end{cases} \tag{8}$$

where

$$L = \frac{1}{2}\log\frac{(d_1^* - \hat{d}_4)^2}{(d_1^* - \hat{d}_4)^2 + (\sqrt{d_1^* - \hat{d}_2}\sqrt{d_1^* - \hat{d}_3} + \sqrt{\hat{d}_2 - \hat{d}_4}\sqrt{\hat{d}_3 - \hat{d}_4})^2}. \tag{9}$$

There is one degenerate case in the D-R region (when $\Pi < \Delta$), and there are two degenerate cases in the R-D region (the first two cases in the last inequality of (8)). They are degenerate in the sense that any eight-tuple in those regions is worse than or equal to (in each component) an eight-tuple on the boundary of the non-degenerate region. This interpretation is made more explicit at the end of the forward proof for Theorem 2.

The region given in Theorem 2 reduces to the Gaussian MD region, when $R_1 = R_4 = 0$. The form of this achievable region is not surprising, given the aforementioned achievable scheme and the Guassian MD region in

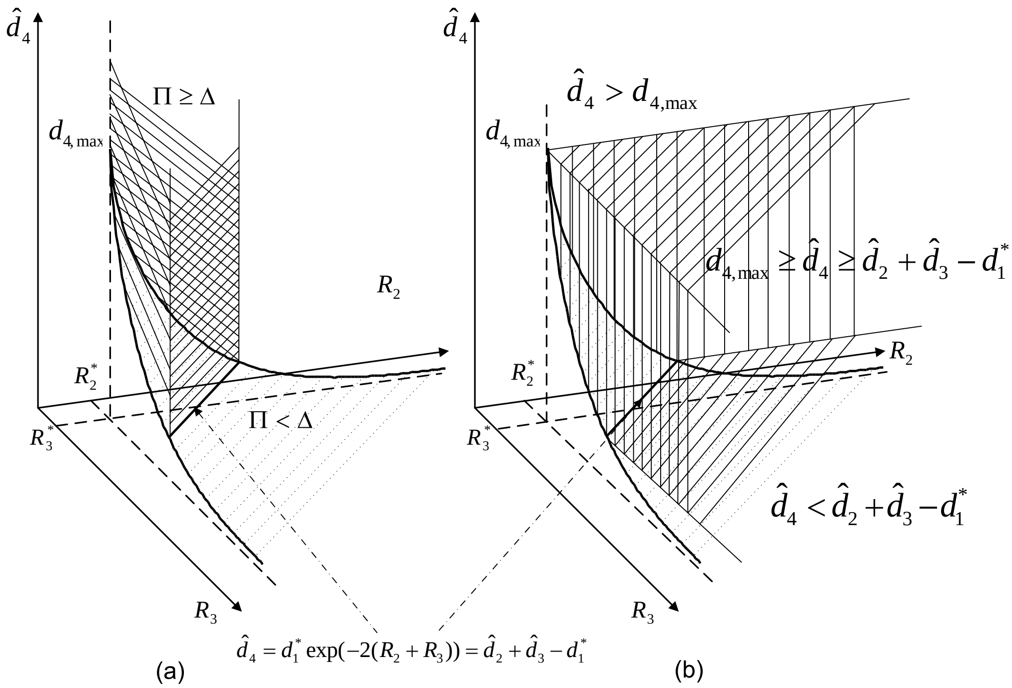

Fig. 3. The equivalence of the R-D and D-R characterization of $(R_2, R_3, \hat{d}_4)$ (a) D-R characterization (b) R-D characterization. $R_2^* = R(\hat{d}_2/d_1^*)$, $R_3^* = R(\hat{d}_3/d_1^*)$, and $d_{4,\max} = (\hat{d}_2^{-1} + \hat{d}_3^{-1} - (d_1^*)^{-1})^{-1}$

[6] (with the additional degenerate case made explicit here). However, the converse is not yet clear due to the involvement of the coding function ϕ_1 and ϕ_4. More precisely, the following two questions regarding ϕ_1 and ϕ_4, respectively, can be asked:

1) Is a Gaussian codebook optimal for encoder ϕ_1?
2) Because the additional information provided by ϕ_4, should the codebooks generated for ϕ_2 and ϕ_3 still have the same structure as the MD codebooks? In other words, should there still be matched codewords for (almost) every typical source sequence in the codebooks for ϕ_2 and ϕ_3 (or the matching can be moved into ϕ_4)?

In the proof, we will show that the answers to both the questions are positive. The main difference from the well-known proof by Ozarow [6] for the Gaussian MD problem is the addition of coding stages ϕ_1 and ϕ_4, which makes the converse more involved, and the entropy power inequality has to be applied twice in the proof.

Before proceeding to the proof of Theorem 2, we illustrate the equivalence of the D-R and the R-D characterizations. The first three inequalities in (5) are clearly equivalent to (7) and the first two inequalities in (8). Thus we only need to establish that with them, the region of triples $(R_2, R_3, \hat{d}_4)$ characterized by the last inequality in (5) and that characterized by the last inequality in (8) are equivalent for any valid and fixed $(R_1, R_4, d_1, d_2, d_3)$. This is illustrated geometrically in Fig. 3. It is seen that the same region above the surface can be described either in two regimes as in the D-R characterization, or in three regimes as in the R-D characterization. Their function forms specifying the regions can be shown to be equivalent with some amount of algebra from the functions given in Theorem 2 and Theorem 2′. Note that the same argument is true for the MD problem, simply by taking $R_1 = R_4 = 0$.

Proof: [Theorem 2]

Converse: The following bounds are straightforward by conventional rate-distortion theory:

$$d_1 \geq \sigma_x^2 \exp(-2R_1), \quad d_2 \geq \sigma_x^2 \exp[-2(R_1+R_2)], \quad d_3 \geq \sigma_x^2 \exp[-2(R_1+R_3)]. \tag{10}$$

We also have

$$\begin{aligned} I(X^n;\hat{X}_4^n) &\overset{(a)}{\leq} I(X^n;\phi_1\phi_2\phi_3\phi_4) \leq H(\phi_1\phi_2\phi_3\phi_4) \\ &= H(\phi_1) + H(\phi_2|\phi_1) + H(\phi_3|\phi_1) - I(\phi_2;\phi_3|\phi_1) + H(\phi_4|\phi_1\phi_2\phi_3) \\ &\overset{(b)}{\leq} n(R_1+R_2+R_3+R_4) - I(\phi_2;\phi_3|\phi_1), \end{aligned}$$

where (a) is because $\hat{X}_4^n$ is determined by the encoding functions $(\phi_1,\phi_2,\phi_3,\phi_4)$, and (b) is because conditioning reduces entropy as $H(\phi_3|\phi_1) \leq H(\phi_3)$ and $H(\phi_2|\phi_1) \leq H(\phi_2)$, and because of the cardinalities of the encoding functions. By converse to source coding theorem

$$d_4 \geq D(\frac{1}{n}I(X^n;\hat{X}_4^n)) \geq \sigma_x^2 \exp[-2(R_1+R_2+R_3+R_4)] \exp[\frac{2}{n}I(\phi_2;\phi_3|\phi_1)]. \tag{11}$$

Because $\hat{X}_2^n$ and $\hat{X}_3^n$ are functions of ϕ_1,ϕ_2 and ϕ_1,ϕ_3, respectively, it is seen that

$$I(\phi_2;\phi_3|\phi_1) \geq I(\hat{X}_2^n;\hat{X}_3^n|\phi_1). \tag{12}$$

As in the proof by Ozarow [6] for the MD problem, let $Y = X + N$, where N is zero mean Gaussian with variance ϵ and independent of X. Because of the following identity

$$\begin{aligned} I(\hat{X}_2^n;\hat{X}_3^nY^n|\phi_1) &= I(\hat{X}_2^n;\hat{X}_3^n|Y^n\phi_1) + I(\hat{X}_2^n;Y^n|\phi_1) \\ &= I(\hat{X}_2^n;\hat{X}_3^n|\phi_1) + I(\hat{X}_2^n;Y^n|\hat{X}_3^n\phi_1) \\ &= I(\hat{X}_2^n;\hat{X}_3^n|\phi_1) + I(\hat{X}_2^n\hat{X}_3^n;Y^n|\phi_1) - I(\hat{X}_3^n;Y^n|\phi_1), \end{aligned}$$

we have

$$\begin{aligned} I(\hat{X}_2^n;\hat{X}_3^n|\phi_1) &\geq I(\hat{X}_2^n;Y^n|\phi_1) + I(\hat{X}_3^n;Y^n|\phi_1) - I(\hat{X}_2^n\hat{X}_3^n;Y^n|\phi_1) \\ &= H(Y^n|\phi_1) - H(Y^n|\hat{X}_2^n\phi_1) + H(Y^n|\phi_1) - H(Y^n|\hat{X}_3^n\phi_1) - H(Y^n|\phi_1) + H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1) \\ &= H(Y^n|\phi_1) - H(Y^n|\hat{X}_2^n\phi_1) - H(Y^n|\hat{X}_3^n\phi_1) + H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1) \\ &= I(Y^n;\hat{X}_2^n\phi_1) + I(Y^n;\hat{X}_3^n\phi_1) - 2H(Y^n) + H(Y^n|\phi_1) + H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1) \\ &\geq I(Y^n;\hat{X}_2^n) + I(Y^n;\hat{X}_3^n) - 2H(Y^n) + H(Y^n|\phi_1) + H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1). \end{aligned} \tag{13}$$

Taking Y as a Gaussian source, the distortion between Y^n and $\hat{X}_2^n$ is upper bounded by $n(d_2+\epsilon)$, by the converse to source coding theorem

$$I(Y^n;\hat{X}_2^n) \geq nR(d_2+\epsilon) = \frac{n}{2}\log\frac{\sigma_x^2+\epsilon}{d_2+\epsilon}. \tag{14}$$

Similarly

$$I(Y^n;\hat{X}_3^n) \geq \frac{n}{2}\log\frac{\sigma_x^2+\epsilon}{d_3+\epsilon}. \tag{15}$$

The following steps are the main difference between our converse proof and Ozarow's, and it is worth noting the complication introduced by the coding function ϕ_1. We apply the conditional entropy power inequality [21] on the term $H(Y^n|\phi_1)$, which gives

$$H(Y^n|\phi_1) \geq \frac{n}{2}\log[\exp(\frac{2}{n}H(X^n|\phi_1)) + 2\pi e\epsilon]. \tag{16}$$

However notice that

$$H(X^n|\phi_1) = H(X^n) - I(X^n;\phi_1) \geq \frac{n}{2}\log(2\pi e\sigma_x^2) - nR_1,$$

which gives

$$H(Y^n|\phi_1) \geq \frac{n}{2}\log[\exp(\frac{2}{n}[\frac{n}{2}\log(2\pi e\sigma_x^2) - nR_1]) + 2\pi e\epsilon] = \frac{n}{2}\log[2\pi e(\sigma_x^2\exp(-2R_1)+\epsilon)]. \tag{17}$$

Apply the entropy power inequality again on the term $H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1)$, we have

$$H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1) \geq \frac{n}{2}\log[\exp(\frac{2}{n}H(X^n|\hat{X}_2^n\hat{X}_3^n\phi_1)) + 2\pi e\epsilon].$$

It follows that

$$\begin{aligned} H(X^n|\hat{X}_2^n\hat{X}_3^n\phi_1) &= H(X^n|\phi_1) - I(X^n;\hat{X}_2^n\hat{X}_3^n|\phi_1) \\ &\overset{(a)}{=} H(X^n|\phi_1) - H(\hat{X}_2^n\hat{X}_3^n|\phi_1) \\ &= H(X^n) - I(X^n;\phi_1) - H(\hat{X}_2^n|\phi_1) - H(\hat{X}_3^n|\phi_1) + I(\hat{X}_2^n;\hat{X}_3^n|\phi_1) \\ &\overset{(b)}{\geq} H(X^n) - I(X^n;\phi_1) - H(\phi_2|\phi_1) - H(\phi_3|\phi_1) + I(\hat{X}_2^n;\hat{X}_3^n|\phi_1) \\ &\geq \frac{n}{2}\log(2\pi e\sigma_x^2) - n(R_1+R_2+R_3) + I(\hat{X}_2^n;\hat{X}_3^n|\phi_1), \end{aligned}$$

where (a) follows from the fact that $\hat{X}_2^n\hat{X}_3^n$ are functions of X^n, and (b) from the fact $\hat{X}_2^n$ and $\hat{X}_3^n$ are functions of (ϕ_1,ϕ_2) and (ϕ_1,ϕ_3), respectively. This leads to

$$\begin{aligned} H(Y^n|\hat{X}_2^n\hat{X}_3^n\phi_1) &\geq \frac{n}{2}\log\left\{\exp\left(\frac{2}{n}[\frac{n}{2}\log(2\pi e\sigma_x^2) - n(R_1+R_2+R_3) + I(\hat{X}_2^n;\hat{X}_3^n|\phi_1)]\right) + 2\pi e\epsilon\right\} \\ &= \frac{n}{2}\log\left\{2\pi e\left(\sigma_x^2\exp[-2(R_1+R_2+R_3)]\exp[\frac{2}{n}I(\hat{X}_2^n;\hat{X}_3^n|\phi_1)] + \epsilon\right)\right\}. \end{aligned} \tag{18}$$

Define

$$t \triangleq \exp[\frac{2}{n}I(\hat{X}_2^n;\hat{X}_3^n|\phi_1)], \quad d_1^* \triangleq \sigma_x^2\exp(-2R_1), \tag{19}$$

and summarize all the bounds in (13), (17) and (18), we have

$$t \geq \frac{d_1^*+\epsilon}{(d_2+\epsilon)(d_2+\epsilon)}[td_1^*\exp[-2(R_2+R_3)]+\epsilon]. \tag{20}$$

Isolating t we have

$$t \geq \frac{\epsilon(d_1^* + \epsilon)}{(d_2 + \epsilon)(d_3 + \epsilon) - (d_1^* + \epsilon)d_1^* \exp[-2(R_1 + R_2)]}, \tag{21}$$

notice that because $d_2 \geq d_1^* \exp(-2R_2)$ and $d_3 \geq d_1^* \exp(-2R_3)$, the denominator is always positive, as long as ϵ is positive. To get the tightest bound, we maximize t over ϵ. Define

$$\Delta = \frac{d_2 d_3}{d_1^{*2}} - \exp[-2(R_2 + R_3)], \quad \Pi = (1 - \frac{d_2}{d_1^*})(1 - \frac{d_3}{d_1^*}). \tag{22}$$

Then we choose the following value of ϵ

$$\epsilon = \begin{cases} \frac{d_1^* \sqrt{\Delta}}{\sqrt{\Pi} - \sqrt{\Delta}} & \Pi \geq \Delta; \\ \infty & \text{otherwise.} \end{cases} \tag{23}$$

After some algebraic calculation, we have for the case $\Pi \geq \Delta$,

$$t \geq \frac{1}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2}, \tag{24}$$

and subsequently using (11), (12) and (24)

$$d_4 \geq \frac{\sigma_x^2 \exp[-2(R_1 + R_2 + R_3 + R_4)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2}. \tag{25}$$

For the case $\Pi < \Delta$, we have $t \geq 1$, which gives the trivial bound of

$$d_4 \geq \sigma_x^2 \exp[-2(R_1 + R_2 + R_3 + R_4)]. \tag{26}$$

Forward: Construct the following random variables

$$U_1 = X + N_1, \quad X' = X - \mathbb{E}(X|U_1), \quad U_2 = X' + N_2, \quad U_3 = X' + N_3, \quad U_4 = X' + N_4, \tag{27}$$

where N_1, N_2, N_3, N_4 are zero mean jointly Gaussian, independent of X, and having the covariance matrix

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \rho\sigma_2\sigma_3 & 0 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^3 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix} \tag{28}$$

It can be seen that X' is essentially the innovation of X given U_1. The decoding functions are

$$\hat{X}_1 = f_1(U_1) = \mathbb{E}(X|U_1) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_1^2} U_1, \qquad \hat{X}_2 = f_2(U_1, U_2) = \hat{X}_1 + \mathbb{E}(X'|U_2) \tag{29}$$

$$\hat{X}_3 = f_3(U_1, U_2) = \hat{X}_1 + \mathbb{E}(X'|U_3), \qquad \hat{X}_4 = f_4(U_1, U_2, U_3, U_4) = \hat{X}_1 + \mathbb{E}(X'|U_2 U_3 U_4). \tag{30}$$

We have that the following rate R_1 is achievable

$$R_1 \geq I(X; U_1) = \frac{1}{2} \log \frac{\sigma_x^2 + \sigma_1^2}{\sigma_1^2}. \tag{31}$$

Choose σ_1 such that the above inequality holds with equality. Then we have

$$d_1 = \mathbb{E}(X'^2) = \frac{\sigma_1^2 \sigma_x^2}{\sigma_1^2 + \sigma_x^2} = \sigma_x^2 \exp(-2R_1). \tag{32}$$

We have also

$$d_2 = \mathbb{E}[X - \hat{X}_2]^2 = \mathbb{E}[X' - \mathbb{E}(X'|U_2)]^2 = \frac{d_1 \sigma_2^2}{d_1 + \sigma_2^2} \tag{33}$$

Notice also

$$I(X; U_2|U_1) = I(X'; X' + N_2|U_1) \overset{(a)}{=} I(X'; X' + N_2) = \frac{1}{2} \log \frac{d_1 + \sigma_2^2}{\sigma_2^2}, \tag{34}$$

where (a) is true because U_1 is independent of X' and N_2. Choose σ_2^2 such that

$$R_2 \geq \frac{1}{2} \log \frac{d_1 + \sigma_2^2}{\sigma_2^2}, \tag{35}$$

which can always be done because the function is continuous. Similarly choose σ_3^2 such that

$$R_3 \geq \frac{1}{2} \log \frac{d_1 + \sigma_3^2}{\sigma_3^2}. \tag{36}$$

The rates satisfying the following bound are achievable

$$\begin{aligned} R_1 + R_2 + R_3 &\geq I(X; U_1) + I(X; U_2 U_3|U_1) + I(U_2; U_3|U_1) \\ &= R_1 + I(X'; U_2 U_3|U_1) + I(X' + N_2; X' + N_3|U_1) \\ &= R_1 + I(X'; U_2 U_3) + I(U_2; U_3) \\ &= R_1 - H(N_2 N_3) + H(U_2) + H(U_3), \end{aligned}$$

which gives

$$R_2 + R_3 \geq \frac{1}{2} \log \frac{(d_1 + \sigma_2^2)(d_1 + \sigma_3^2)}{(1 - \rho^2)\sigma_2^2 \sigma_3^2} = \frac{d_1^2}{d_2 d_3 (1 - \rho^2)}. \tag{37}$$

When $\Pi \geq \Delta$, choose

$$\rho = -\sqrt{1 - \frac{d_1^2 \exp[-2(R_1 + R_2)]}{d_2 d_3}}. \tag{38}$$

Then

$$I(X; U_4|U_1 U_2 U_3) = I(X'; U_4|U_2 U_3) = H(U_4|U_2 U_3) - H(N_4) = \frac{1}{2} \log \frac{\sigma_4^2 + d_4^*}{\sigma_4^2},$$

where

$$d_4^* \triangleq \mathbb{E}(X' - E(X'|U_2 U_3))^2 = \frac{d_1 \sigma_2^2 \sigma_3^2 (1 - \rho^2)}{d_1 \sigma_2^2 \sigma_3^2 (1 - \rho^2) + d_1(\sigma_2^2 + \sigma_3^2) - 2\rho d_1 \sigma_2 \sigma_3}. \tag{39}$$

Choose σ_4^2 such that

$$R_4 = \frac{1}{2} \log \frac{\sigma_4^2 + d_4^*}{\sigma_4^2}. \tag{40}$$

We further have (and after some simplification)

$$\begin{aligned} d_4 &= \mathbb{E}(X' - \mathbb{E}(X'|U_2 U_3 U_4))^2 && (41) \\ &= \frac{\exp(-2R_4) d_1 \sigma_2^2 \sigma_3^2 (1 - \rho^2)}{\sigma_2^2 \sigma_3^2 (1 - \rho^2) + d_1(\sigma_2^2 + \sigma_3^2) - 2\rho d_1 \sigma_2 \sigma_3} && (42) \\ &= \frac{\sigma_x^2 \exp[-2(R_1 + R_2 + R_3 + R_4)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2}. && (43) \end{aligned}$$

Thus if $\Pi \geq \Delta$, i.e., $d_1 + d_1 \exp[-2(R_2+R_3)] \geq d_2 + d_3$, then this achievable region matches the outer bounds.

When $\Pi < \Delta$, i.e., $d_1 + d_1 \exp[-2(R_2+R_3)] < d_2 + d_3$, we can find some $d_2' \leq d_2$ and $d_3' \leq d_3$, where at least one of the inequalities is strict, such that $d_1 + d_1 \exp[-2(R_2+R_3)] = d_2' + d_3'$; with this choice of (d_2', d_3'), apparently $\Pi' = \Delta'$, which further implies that the distortion quadruple $(d_1, d_2', d_3', \sigma_x^2 \exp[-2(R_1+R_2+R_3+R_4)])$ is achievable[3]. Thus the quadruple $(d_1, d_2, d_3, \sigma_x^2 \exp[-2(R_1+R_2+R_3+R_4)])$ is also achievable for the case $\Pi < \Delta$, and for both the cases the achievable region indeed matches the outer bounds. ■

B. Fixed Channel Configuration and the Performance Loss

One case of interest is that the good channel and the bad channel used to transmit are fixed as $R_1/R_2 = R_3/R_4$, and we will consider the performance loss in this case. Suppose $R_2 = \alpha R_1$ and $R_4 = \alpha R_3$.

Consider the cases where only the first layer performance or only the second layer performance is in consideration. For the form case, i.e., the first layer, user 1 has description of rate R_1 while user 2 has joint description of rate $(1+\alpha)R_1$, which results in minimum distortions $\sigma_x^2 \exp(-2R_1)$ and $\sigma_x^2 \exp[-2(1+\alpha)R_1]$. For the latter case, i.e., optimized only for the second layer, user 1 has description of rate $R_1 + R_3$ while user 2 has joint description of rate $(1+\alpha)(R_1+R_3)$, which results in minimum distortions $\sigma_x^2 \exp[-2(R_1+R_3)]$ and $\sigma_x^2 \exp[-2(1+\alpha)(R_1+R_3)]$.

Now for an MSR system to achieve the same minimum second layer distortions as if only the second layer is in consideration. Then by the result from the previous section, we see that $\Pi \leq \Delta$, which gives

$$d_2 + \sigma_x^2 \exp[-2(R_1+R_3)] \geq d_1^*[1 + \exp[-2(R_2+R_3)]],$$

which further gives

$$d_2 \geq d_1^*[1 + \exp(-2(\alpha R_1 + R_3)) - \exp(-2R_3)]. \tag{44}$$

For a single layer system optimized for the first layer, distortion $d_1^* = \sigma_x^2 \exp(-2R_1)$ and $d_2^* = \sigma_x^2 \exp[-2(1+\alpha)R_1]$ are achievable, thus the loss on d_2 can be as large as

$$\frac{d_2}{d_2^*} = \frac{[1 + \exp(-2(\alpha R_1 + R_3)) - \exp(-2R_3)]}{\exp(-2\alpha R_1)} = \exp(2\alpha R_1) + \exp(-2R_3) - \exp[2(\alpha R_1 - R_3)].$$

Thus we see that as $R_1 \to \infty$, **the performance loss compared to a single layer system can be unbounded**. However, the distortion d_1 is not jeopardized by the progressive encoding requirement. In other words (d_1, d_3, d_4) can be matched to an optimal coding system with coding rate $(R_1, R_1+R_3, (1+\alpha)(R_1+R_3))$, with the distortion d_2 being quite large. If d_2 is of little importance, then such a system can be utilized; otherwise, the performance loss needed to improve d_2 is large, and can hardly be compensated by the added functionality.

C. MD Coding vs. Wyner-Ziv Coding

The coding approach proposed in [1] is based on the Wyner-Ziv (WZ) coding, which treats the reconstruction $\hat{X}_1$ and $\hat{X}_2$ as side informations at the decoder. In this section we compare the performance by the WZ-based coding approach with that by the MD-based coding approach.

[3]This degenerate region was not treated by Ozarow in [6], and it sometimes causes certain confusion.

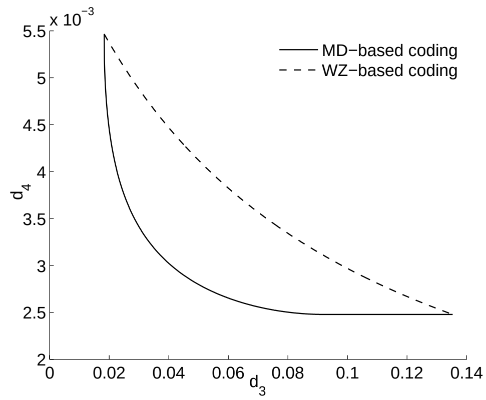

Fig. 4. Comparison of the distortion regions (d_3, d_4) of the two coding approaches at $R_1 = R_3 = 1.0$ nat and $R_2 = R_4 = 0.5$ nat with fixed $d_1 = d_1^*$ and $d_2 = d_2^*$. The range of d_3 is $[\sigma_x^2 \exp[-2(R_1 + R_3)], d_1^*]$.

To compare the two coding scheme, fix $d_1 = d_1^* = \sigma_x^2 \exp(-2R_1)$ and $d_2 = d_2^* = \sigma_x^2 \exp[-2(R_1 + R_2)]$. For the WZ-based approach, since the ϕ_3 and ϕ_4 are successive refinement by definition, the WZ-based coding is in fact the successive Wyner-Ziv problem with degraded side information at the decoder considered by Steinberg and Merhav in [22]. Though $\hat{X}_1$ and $\hat{X}_2$ are not necessarily physically degraded, as pointed out in [22], the achievable region is only dependent on the pairwise distribution between the source and the side information, thus statistical degradedness and physical degradedness have no essential difference. It is known that the Gaussian source and side informations can always be taken as statistically degraded, and thus the general result in [22] can be readily used. The rate-distortion region for the Gaussian source was given explicitly in [23], and can be (modified accordingly and) written as follows. Choose σ_1^2 and σ_2^2 such that

$$d_1^* = \frac{\sigma_x^2(\sigma_1^2 + \sigma_2^2)}{\sigma_x^2 + \sigma_1^2 + \sigma_2^2}, \quad d_2^* = \frac{\sigma_x^2 \sigma_2^2}{\sigma_x^2 + \sigma_2^2} \tag{45}$$

and define $\gamma \triangleq \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, then the achievable distortions using WZ-based coding are given by

$$d_3' \geq \exp(-2R_3) d_3^* = \sigma_x^2 \exp[-2(R_1 + R_3)] \tag{46}$$

$$d_4' \geq \exp[-2(R_3 + R_4)] \frac{\sigma_x^2 \sigma_1^2 \sigma_2^2}{(\sigma_x^2 + \sigma_1^2 + \sigma_2^2)((1-\gamma)^2 \min(d_3', d_1^*) + \gamma \sigma_1^2)}. \tag{47}$$

On the other hand, the MD-based coding approach can achieve

$$d_3 \geq \sigma_x^2 \exp[-2(R_1 + R_3)], \quad d_4 \geq \frac{\sigma_x^2 \exp[-2(R_1 + R_2 + R_3 + R_4)]}{1 - (|\sqrt{\Pi} - \sqrt{\Delta}|^+)^2}, \tag{48}$$

where we have for this special case

$$\Delta = \frac{d_2^* d_3}{d_1^{*2}} - \exp[-2(R_2 + R_3)] = \exp(-2R_2)[\frac{d_3}{d_1^*} - \exp(-2R_3)] \tag{49}$$

$$\Pi = (1 - \frac{d_2^*}{d_1^*})(1 - \frac{d_3}{d_1^*}) = [1 - \exp(-2R_2)](1 - \frac{d_3}{d_1^*}). \tag{50}$$

A set of typical tradeoff curves of (d_3, d_4) for the WZ-based approach and MD-based approach are given in Fig. 4 for fixed rates (R_1, R_2, R_3, R_4), and we can see the gap is non-zero. As such, the WZ-based approach is suboptimal

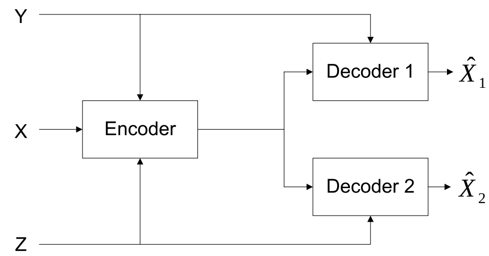

Fig. 5. The system diagram for the EDSI problem.

except two extreme operating points: one point is when $d_3 = d_1^*$ and the other one is $d_3 = \sigma_x^2 \exp[-2(R_1 + R_3)]$. This in fact illustrates the role of side informations at only the decoders or at both the encoder and the decoders are quite different: it is known that for the Gaussian source there is no loss between the cases when a *single* Gaussian side information is available at both the encoder and the decoder, or at the decoder only, however when there are multiple side informations at different decoders, they are no longer equivalent. This observation perhaps was firstly made explicit in [9].

The MSR problem in fact has a more subtle connection with the encoder/decoder side information (EDSI) problem considered in [9] (see also [25]), which is depicted in Fig. 5. One particular interesting case for the EDSI problem is when the source and side informations are physically degraded as $X \leftrightarrow Y \leftrightarrow Z$, which in the Gaussian case means $Y = X + N_1$ and $Z = Y + N_2$ where N_1 and N_2 are independent Gaussian noise. Now if we take $\hat{X}_1$ and $\hat{X}_2$ in MSR as the side informations Z and Y, respectively, and let $R_4 = 0$, then MSR can be considered as a relaxed version of the EDSI problem, because in MSR the codeword $\hat{X}_1$ and $\hat{X}_2$ do not have to be generated by any marginal distribution as specified in the EDSI problem; furthermore, decoder 1 always has $\hat{X}_1$ (corresponding to Z) in addition to $\hat{X}_2$ (corresponding to Y) in MSR. As such, an outer bound for the EDSI problem can be found by using the Gaussian MSR region for this degraded case. As shown in [9], this outer bound is indeed achievable by using a hybrid conditioning/binning scheme [4].

V. A Variation of the MD Problem: the MDCR Problem

As aforementioned, when $R_1 = 0$, the problem being considered reduces to the MD problem with central refinement (MDCR), and the six-tuple of rates and distortions are to be characterized. Again we focus on the Gaussian case; we will continue to use the notations (R_2, R_3, R_4) and descriptions ϕ_2, ϕ_3, ϕ_4 and assuming $R_1 = 0$ and no description ϕ_1 exists.

El Gamal and Cover constructed an MD scheme for general sources based on the MDCR method in [7]. More precisely, the description ϕ_4 of rate R_4 is split and combined into the existing two descriptions ϕ_2 and ϕ_3, and the resulting two descriptions are of rates $R_2' = R_2 + \beta R_4$ and $R_3' = R_3 + (1-\beta)R_4$ for some $0 \leq \beta \leq 1$. It is

[4] The (same) outer bound for this case given in [9] was derived by applying the conditional version of the results of [24], which is indeed closely related to the MD problem.

known that in the Gaussian MD problem, there is no need for the central refinement coding to achieve the complete distortion region; i.e., $R_4 = 0$ is sufficient to achieve the complete distortion region given the two description rates R'_2 and R'_3. A natural question to ask is whether it is possible to construct an optimal Gaussian MD system using an MDCR system with nonzero rate R_4.

The answer to the above question in fact negative, which is implied by Theorem 2. To see this, assume the distortion d_2^* and d_3^* in both an MDCR-based system and an optimal MD system. By Theorem 2, we see that for an MDCR system with non-zero R_4

$$d_4 \geq \frac{\sigma_x^2 \exp[-2(R_2 + R_3 + R_4)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2} \tag{51}$$

where

$$\Delta = \frac{d_2^* d_3^*}{\sigma_x^4} - \exp[-2(R_2 + R_3)], \quad \Pi = (1 - \frac{d_2^*}{\sigma_x^2})(1 - \frac{d_3^*}{\sigma_x^2}). \tag{52}$$

For an optimal MD system, the distortion resulting from the joint description can be

$$d'_4 \geq \frac{\sigma_x^2 \exp[-2(R'_2 + R'_3)]}{1 - (\sqrt{\Pi'} - \sqrt{\Delta'})^2} \tag{53}$$

where

$$\Delta' = \frac{d_2^* d_3^*}{\sigma_x^4} - \exp[-2(R'_2 + R'_3)], \quad \Pi' = (1 - \frac{d_2^*}{\sigma_x^2})(1 - \frac{d_3^*}{\sigma_x^2}). \tag{54}$$

To keep the rates of the two system equal, we have $R'_2 = R_2 + \beta R_4$ and $R'_3 = R_3 + (1-\beta) R_4$ for some $0 \leq \beta \leq 1$. It is now seen that the MDCR approach with non-zero R_4 is suboptimal because $d_4 > d'_4$, due to the fact that $\Delta' > \Delta$ and $\Pi' = \Pi$. This is a stronger result than the known one that it is sufficient for $R_4 = 0$ to achieve optimality: **it is in fact necessary for R_4 to be zero in order to be optimal in the Gaussian MD case**.

This result suggests any system based on the MDCR approach when the refinement rate is not zero is not optimal for the Gaussian source: one such example is the system constructed with dithered lattice quantizers in [26].

A. The high-rate asymptotics for balanced descriptions

We consider balanced MDs in this subsection, and further assume $\sigma_x^2 = 1$. Suppose in an MD system, the rate of two descriptions are at equally high rate of R' each, and the side distortions are both $d'_2 = d'_3$. It can be shown that if the side distortion is of the form $d'_2 = b 2^{-2(1-\eta)R'}$, where $0 \leq \eta < 1$ and $b \geq 1$, the central distortion of an MD system can asymptotically (at low distortion) achieve

$$d'_4 \geq \begin{cases} 2^{-2R'}/2(b + \sqrt{b^2 - 1}) & \eta = 0; \\ 2^{-2R'(1+\eta)}/4b & 0 < \eta < 1. \end{cases} \tag{55}$$

Notice the condition $0 < \eta < 1$ in fact corresponds to the condition that $1 \gg d'_2$ and $d'_2 \gg d'_4$ at high rate. In this case, the central and side distortion product remains bounded by a constant at fixed rate, which is $d'_4 d'_2 \geq \frac{2^{-4R'}}{4}$, independent of the tradeoff between them. This product has been used as the information theoretical bound to measure the efficiency of quantization methods [27], [28]. Below, the performance of the optimal MD system is compared with that of an MDCR-based system in this high-rate and high-refinement-rate case.

For an MDCR-based MD system, R_4 is allocated to the refinement stage, and thus each of the first stage descriptions is of rate $R' - R_4/2$. Keeping the side distortion of this system $d_2 = d_2' = b2^{-2(1-\eta)R'}$ for an easier comparison, consider the case $1 > \eta > 0$, and let $R_4 = 2\eta_1 R'$, where $1 - \eta_1$ is the ratio between R_2 and R_2'. Then it can be shown (through some algebra) that using the MDCR approach, we can achieve

$$d_4 \geq \begin{cases} 2^{-2R'(1+\eta)}/2(b+\sqrt{b^2-1}) & \eta_1 = \eta; \\ 2^{-2R'(1+\eta)}/4b & 0 \leq \eta_1 < \eta. \end{cases} \tag{56}$$

This implies that if the first stage has sufficient excess marginal rate, i.e., $\eta_1 < \eta$, then the performance loss from the optimal MD system by the MDCR approach with non-zero R_4, in terms of the distortion product, is asymptotically zero in the range of $1 \gg d_2$ and $d_2 \gg d_4$. However, as the rate allocated to the refinement stage increases, the excess marginal rate in the first stage decreases. When $\eta_1 = \eta$, the performance loss is a factor of $\frac{2b}{(b+\sqrt{b^2-1})}$. If the first stage is without excess marginal rate, which means $\eta_1 = \eta$ and $b = 1$, then the loss is a factor of 2 comparing to the MD system without taking such an MDCR approach.

This discussion suggests that the MDCR approach is appealing for the high-rate case, if $1 \gg d_2$ and $d_2 \gg d_4$ is the desired operating range. However, the first stage should reserve enough excess marginal rate in order to avoid the performance loss. Taking the MD system in [26] as an example, using certain sub-optimal lattices for ϕ_2 and ϕ_3 is potentially able to achieve (asymptotic) optimal performance, but using two good lattices as ϕ_2 and ϕ_3 will not be, because the excess marginal rate is diminishing as the dimension increases.

VI. Conclusion

We considered the problem of multiuser successive refinement. An achievable region was provided, which was shown to be tight for the Gaussian source under MSE measure. It was shown that different from the single user case, the MSR coding necessitates performance loss, which can be unbounded. The results rely on the recognition that a multiple description system is embedded inside the MSR system. The MSR system also includes a variation of the MD system, namely the MDCR problem. This problem was treated with some depth, which revealed some interesting implications in designing the MD coding system.

References

[1] S. Pradhan and K. Ramchandran, "Multiuser successive refinement," in *Proc. Conference on Information Sciences and Systems*, Mar. 2002.

[2] V. N. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Information Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[4] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. Information Theory*, vol. 40, no. 1, pp. 253–259, Jan. 1994.

[5] L. Lastras and T. Berger, "All sources are nearly successively refinable," *IEEE Trans. Information Theory*, vol. 47, no. 3, pp. 918–926, Mar. 2001.

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

[7] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Information Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[8] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[9] E. Perron, S. Diggavi, and E. Telatar, "On the role of encoder side-information in source coding for multiple decoders," to appear *Proc. 2006 International Symposium on Information Theory*, Jul. 2006.

[10] C. Tian, J. Chen, S. Hemami, and T. Berger, "Multiple descriptions with central refinement," in *Proc. Conference on Information Sciences and Systems*, Mar. 2005.

[11] H. Feng and M. Effros, "On the achievable region for multiple description source codes on Gaussian sources," in *Proc. IEEE Symposium Information Theory*, p.195, Jun.-Jul 2003.

[12] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisted," *IEEE Trans. Information Theory*, vol. 49, no. 5, pp. 1147–1158, May. 2003.

[13] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 47, no. 6, pp. 2199 – 2224, Sep. 2001.

[14] L. A. Lastras and V. Castelli, "Near sufficiency of random coding for two descriptions," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 681–695, Feb. 2006.

[15] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Information Theory*, vol. 23, no. 2, pp. 179–182, Mar. 1977.

[16] S.S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions - part i: (n, k) source-channel erasure codes," *IEEE Trans. Information Theory*, vol. 50, pp. 47–61, Jan. 2004.

[17] S. Diggavi and V. Vaishampayan, "On multiple description source coding with decoder side information," in *Proc. IEEE Information Theory Workshop*, pp. 88-93, Oct. 2004.

[18] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source codes," *IEEE Trans. Information Theory*, vol. 45, no. 6, pp. 1887–1910, Sep. 1999.

[19] J. Chen, C. Tian, S. Hemami, and T. Berger, "Multiple description quantization via Gram-Schmidt orthogonalization," submitted to *IEEE Trans. Information Theory*.

[20] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," submitted to *IEEE Trans. Information Theory*.

[21] T. M. Cover and J. A. Thomas, *Elements of information theory*, New York: Wiley, 1991.

[22] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Information Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[23] C. Tian and S. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side information," in *EPFL Technical Report*, Jan. 2006.

[24] A. Kaspi, "Rate-distortion when side-information may be present at the decoder," *IEEE Trans. Information Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[25] J.K. Wolf, "Source coding for a noiseless broadcast channel," in *Proc. Conference on Information Sciences and Systems*, Mar. 2004.

[26] Y. Frank-Dayan and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Trans. Information Theory*, vol. 48, no. 1, pp. 192–204, Jan. 2002.

[27] C. Tian and S. S. Hemami, "Universal multiple description scalar quantizer: analysis and design," *IEEE Trans. Information Theory*, vol. 50, no. 9, pp. 2089–2102, Sep. 2004.

[28] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Information Theory*, vol. 47, no. 5, pp. 1718–1734, Jul. 2001.