Multiuser Successive Refinement and Multiple Description Coding

Chao Tian, Member, IEEE, Jun Chen, Member, IEEE, Suhas N. Diggavi, Member, IEEE

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 the multiple description (MD) problem (with two descriptions) as a subsystem, and the techniques useful in the MD problem can be extended to this case. It is shown 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. It is further shown 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.

Index Terms—Multiple description coding, rate distortion, source coding, successive refinement.

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 “multuisers” 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 \( R_{11} \) (the first subscript specifies which user and the second subscript specifies which round of transmission), \( R_{12}, R_{21}, R_{22} \) can vary. If the transmission rate 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; this special case is important in practice as we expect that the channels between the transmitter and the receivers remain the same over the two rounds of transmission. We will only consider the two-user two-layer system in this work.

This work was presented in part at 2006 IEEE Information Theory Workshop in Oct. 2006, Chengdu, China.

![Fig. 1. A system diagram with two users and two layers.](image-url)

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; i.e., when the necessary encoding rate does not increase comparing to a single stage coding. 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 the source is not successively refinable, the rate loss is bounded by a universal constant.

In the multiuser setting we are interested in understanding whether the progressive coding requirement necessitates 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 (two-description) 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 system design.

The MSR problem includes the problem of multiple descriptions (MD) 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.
consists of 4 encoding functions single subscript, whose correspondence is made clear in Table I.

Let \( \hat{x} \) be a finite reconstruction alphabet, and for simplicity we \( \hat{x} \) be a finite set and let \( \phi_i \) be the set of all \( \hat{x} \) with a generic distribution \( \hat{x} \) with components in \( \mathbb{R}^n \). Denote an arbitrary member of \( \hat{x} \) as \( x^n = (x_1, x_2, ..., x_n) \), or alternatively as \( x_i = (x_i, x_2, ..., x_j) \) will also be written as \( x_{i1}, x_{i2}, ..., x_{in} \). Upper case is used for random variables and vectors. A discrete memoryless source (DMS) 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 \( \hat{X} \) be a finite reconstruction alphabet, and for simplicity we assume that the decoders all use this reconstruction alphabet. Let \( d : \mathcal{X} \times \hat{X} \rightarrow [0, \infty) \) be a distortion measure. The single-letter distortion of a vector is defined as

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

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. 2 is simplified to single subscript, whose correspondence is made clear in Table I.

Definition 1: An \((n, M_1, M_2, D_1, \delta)\) MSR code for source \( X \) consists of 4 encoding functions \( \phi_i \) and 4 decoding functions \( \psi_i \).

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.

The distortion-rate (D-R) region and the rate-distortion (R-D) region given for the Gaussian MSR problem can be easily reduced to those for the MD problem. Though the Gaussian MD region has been known for more than 25 years, as pointed out in [11] (discussed more extensively in [12]), the expressions given in the literature are to those for the MD problem. Though the Gaussian MD region has been known for more than 25 years, as pointed out in [11] (discussed more extensively in [12]), the expressions given in the literature are

\[
E \{d^n(x^n, \hat{x}^n)\} = \sum_{i=1}^{n} d(x_i, \hat{x}_i).
\]

\[
\begin{align*}
\phi_i & : \mathcal{X}^n \rightarrow I_M, \quad i = 1, 2, 3, 4 \\
\psi_1 : I_M \rightarrow \hat{X}^n, & \quad \psi_2 : I_M \times I_M \rightarrow \hat{X}^n, \\
\psi_3 : I_M \times I_M \times I_M \rightarrow \hat{X}^n, & \quad \psi_4 : I_M \times I_M \times I_M \times I_M \rightarrow \hat{X}^n
\end{align*}
\]

where \( I_k = \{1, 2, 3, 4\} \) and

\[
\begin{align*}
E \{d^n(x^n, \hat{x}^n)\} & \leq D_1, \\
E \{d^n(x^n, \hat{x}^n)\} & \leq D_2, \\
E \{d^n(x^n, \hat{x}^n)\} & \leq D_3, \\
E \{d^n(x^n, \hat{x}^n)\} & \leq D_4
\end{align*}
\]

where \( E \) is the expectation operation.

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 MD 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 [16] for a review) and new results are emerging, but the problem remains open. Inner bound exists [7] and it was shown that this bound is in fact quite good for source with real alphabet under MSE distortion measure [17], but it is not tight in general [18]. The Gaussian source with MSE distortion measure (and the recent extension to the vector Gaussian problem [19]), 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...
not the same as the MD problem. Notice there are three rates $(R_2, R_3, R_4)$ to characterize here, instead of the two rates in the MD problem; it is nevertheless a special case of the general three description 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. Define the region $Q_{ach}$ to be the set of all rate distortion eight-tuples $(R_1,4, D_{14...4})$ for which there exist four random variables $X_1, X_2, X_3, X_4$ in finite alphabet $\mathcal{X}$ such that
\[
\mathbb{E}d(X, \hat{X}_i) \leq D_i, \quad i = 1, 2, 3, 4, \tag{4}
\]
and the non-negative rate vector satisfies:
\[
\begin{align*}
R_1 & \geq I(X; \hat{X}_1) \tag{5} \\
\sum_{i=1,2} R_i & \geq I(X; \hat{X}_1 \hat{X}_2), \tag{6} \\
\sum_{i=1,3} R_i & \geq I(X; \hat{X}_1 \hat{X}_3), \tag{7} \\
\sum_{i=1,2,3} R_i & \geq I(X; \hat{X}_1 \hat{X}_2 \hat{X}_3) + I(\hat{X}_2; \hat{X}_3 | \hat{X}_1), \tag{8} \\
\sum_{i=1,2,3,4} R_i & \geq I(X; \hat{X}_1 \hat{X}_2 \hat{X}_3 \hat{X}_4) + I(\hat{X}_2; \hat{X}_3 | \hat{X}_1). \tag{9}
\end{align*}
\]

The following theorem provides an achievable region.

**Theorem 1:**

$$Q_{ach} \subseteq Q.$$ If $R_1 = R_4 = 0$, $Q_{ach}$ 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 immediately 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 $r_4$ into $r_1$, and $r_4$ into $r_1, r_2, r_3$ to construct new codes; see [23] (also [4]) for a thorough explanation regarding a similar property in the successive refinement problem. Thus indeed the region given in Theorem 1 is achievable.

It is not clear whether the region $Q_{ach}$ is convex. Interestingly, a generalization of this region, denoted as $Q_{ach}$, is indeed convex, and we now provide this generalized region. Let the region $Q_{ach}'$ be the set of all rate distortion eight-tuples $(R_1,4, D_{14...4})$ for which there exist four random variables $U_1, U_2, U_3, U_4$ in finite alphabets $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \mathcal{U}_4$ such that there exists deterministic functions
\[
\begin{align*}
g_1 : U_1 & \rightarrow \hat{X}, & g_2 : U_1 \times U_2 & \rightarrow \hat{X}, \tag{10} \\
g_3 : U_1 \times U_3 & \rightarrow \hat{X}, & g_4 : U_1 \times U_2 \times U_3 \times U_4 & \rightarrow \hat{X}. \tag{11}
\end{align*}
\]
satisfying
\[
\begin{align*}
\mathbb{E}d(X, g_1(U_1)) & \leq D_1, & \mathbb{E}d(X, g_2(U_1, U_2)) & \leq D_2, \tag{12} \\
\mathbb{E}d(X, g_3(U_1, U_3)) & \leq D_3, & \mathbb{E}d(X, g_4(U_1, U_2, U_3, U_4)) & \leq D_4. \tag{13}
\end{align*}
\]
and the non-negative rate vector satisfies:
\[ R_1 \geq I(X; U_1), \]  
\[ \sum_{i=1,2} R_i \geq I(X; U_1 U_2), \]  
\[ \sum_{i=1,3} R_i \geq I(X; U_1 U_3), \]  
\[ \sum_{i=1,2,3} R_i \geq I(X; U_1 U_2 U_3) + I(U_2; U_3|U_1), \]  
\[ \sum_{i=1,2,3,4} R_i \geq I(X; U_1 U_2 U_3 U_4) + I(U_2; U_3|U_1). \]  

**Theorem 1':**
\[ \mathcal{Q}_{ach} \subseteq \mathcal{Q}. \]

The proof of this theorem follows the exact same line as that of Theorem 1. Moreover, it can be shown straightforwardly that \( \mathcal{Q}_{ach} \subseteq \mathcal{Q}_{ach}' \), but it is not clear whether the inclusion in the other direction is also true. Though the region \( \mathcal{Q}_{ach}' \) is more general, it is also more complex due to the involvement of several decoding functions. In fact, as suggested in [18], the achievable region given by El Gamal and Cover [7] originally had a form with several decoding functions, which was later largely abandoned in favor of the region defined without such functions. The region \( \mathcal{Q}_{ach}' \) is more suitable when we consider the Gaussian source, and we thus include both forms here for completeness.

It is not difficult to show that \( \mathcal{Q}_{ach}' \) is convex. Let \( P(U_1^0, U_2^0, U_3^0, U_4^0|X) \) and \( P(U_1^0, U_2^0, U_3^0, U_4^0|X) \) be two conditional distributions which provide rate vectors in \( \mathcal{Q}_{ach}' \). Let \( Q \) be a Bernoulli random variable with \( \Pr(Q = 0) = \lambda \) and \( \Pr(Q = 1) = 1 - \lambda \), which is independent of everything, then it is easily seen
\[ \lambda I(X; U_1^0) + (1 - \lambda) I(X; U_1^1) = \lambda I(X; U_1^0|Q = 0) + (1 - \lambda) I(X; U_1^1|Q = 1) \]
\[ = I(X; U_1^1, Q). \]  

Moreover
\[ \lambda I(U_2^0; U_3^0|U_1^2) + (1 - \lambda) I(U_2^0; U_3^0|U_1^3) = I(U_2^0; U_3^0|U_1^2, Q). \]  

Similar relations can be derived for the other mutual information quantities. Now define \( U_1 = (U_1^0, Q) \), and the decoding functions can be defined accordingly; it follows that this convex combination of the rate vectors is indeed in \( \mathcal{Q}_{ach}' \).

Though we have considered discrete sources so far, the results can be generalized to Gaussian sources using the techniques in [24][25]. In the next section, we prove a converse for the Gaussian source under the MSE distortion 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 between the encoder and the decoders [22], because the optimal first codebook in MSR is not necessarily a codebook generated with any single letter marginal distribution \( P_{X_1} \), while the common side information is always an i.i.d. random variable in the setting of [22]; moreover, the MSR problem is further complicated by the included MDCR sub-system.

### IV. THE GAUSSIAN SOURCE

In this section, we shall focus on the Gaussian source with MSE distortion measure, and establish the distortion-rate as well as the rate-distortion region.

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. Similar to the MD problem, the D-R region is simpler than the R-D region due to less number of degenerate regions (see [11] and [26]). It will be illustrated that the R-D region can be established from the D-R region. We shall first present the theorems, and then follow the approach by Ozarow [6] to establish the D-R region.

**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
\[ d_1 \geq \sigma_x^2 \exp[-2R_1], \]
\[ d_2 \geq \sigma_x^2 \exp[-2(R_1 + R_2)], \]
\[ d_3 \geq \sigma_x^2 \exp[-2(R_1 + R_3)], \]
\[ d_4 \geq \frac{\sigma_x^2 \exp[-2(R_1 + R_2 + R_3 + R_4)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2} \]  

where \(|x|^+ = \max(x, 0)\) and
\[ d_i = \frac{\sigma_x^2 \exp[-2R_i]}{1 - d_i^*} \]
\[ \Pi = (1 - \frac{d_2}{d_1})(1 - \frac{d_3}{d_1}), \]
\[ \Delta = \frac{d_2d_3}{d_1^2} - \exp[-2(R_2 + R_3)]. \]

Define the function \( R(D) = \frac{1}{2} \log \frac{d}{d_1} \). 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 distortion-rate region for distortions \((d_1, d_2, d_3, d_4)\) is given as follows:
\[ R_1 \geq R_1^* \triangleq R(\min(d_1, \frac{\sigma_x^2}{\sigma_x^2})). \]  

For any rates \( R_1 \geq R_1^* \) and \( R_4 \geq 0 \), define \( d_1^* \triangleq \frac{\sigma_x^2}{\sigma_x^2} \exp(2R_4) \). The achievable rates \((R_2, R_3)\) are given as:
\[ R_2 \geq R(d_2^*/d_1^*), \]
\[ R_3 \geq R(d_3^*/d_1^*), \]
\[ R_2 + R_3 \geq \begin{cases} R(d_4/d_1^*) & 0 < d_4 < d_2 + d_3 - d_1^*; \\ R(d_4/d_1^*) + L & (d_2 - 1)d_3 - (d_1^* - 1)^{-1}; \end{cases} \]
\[ \text{otherwise,} \]
\[ L = \frac{\sqrt{\Pi + \sqrt{\Delta}}}{(d_1^* - d_2^*)^2} + \frac{d_2^* \exp(-2(R_2 + R_3))}{(d_1^* - d_2^*)^2}. \]  

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 \( [29] \)). 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 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 Gaussian MD region in [6] (with the additional degenerate case made explicit here). However, the converse is not yet clear due to the involvement of the coding functions \( \phi_1 \) and \( \phi_4 \). More precisely, the following two questions regarding \( \phi_1 \) and \( \phi_4 \), respectively, can be asked:

1. Is a Gaussian codebook optimal for encoder \( \phi_1 \)?
2. With the additional information provided by \( \phi_1 \), should the codebooks generated for \( \phi_2 \) and \( \phi_3 \) 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 \( \phi_2 \) and \( \phi_3 \) (or the matching can be moved into \( \phi_4 \)?)

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 additional coding stages \( \phi_1 \) and \( \phi_4 \), 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 inequalities in (22)–(24) are clearly equivalent to (27) and the two inequalities in (28). Thus we only need to establish that with them, the region of triples \((R_2, R_3, d_4)\) characterized by (22)–(24) are equivalent for any valid and fixed power inequality has to be applied twice in the proof.

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. The functions given in Theorem 2 and Theorem 2’ specifying the regions can be shown to be equivalent with some amount of algebra (see [11] and [12] for a brief discussion on this algebraic computation in the MD problem). 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:

\[
\begin{align*}
d_1 &\geq \sigma_2^2 \exp(-2R_1), \\
d_2 &\geq \sigma_3^2 \exp(-2[R_1 + R_2]), \\
d_3 &\geq \sigma_4^2 \exp(-2[R_1 + R_3]).
\end{align*}
\]

We also have

\[
I(X^n; \hat{X}_1^n) \leq I(X^n; \phi_1 \phi_2 \phi_3 \phi_4) \leq H(\phi_1 \phi_2 \phi_3 \phi_4) = H(\phi_1) + H(\phi_2 | \phi_1) + H(\phi_3 | \phi_1) - I(\phi_2; \phi_3 | \phi_1) + H(\phi_4 | \phi_1),
\]

where (a) is because \( \hat{X}_1^n \) is determined by the encoding functions \( \phi_1, \phi_2, \phi_3, \phi_4 \), and (b) is because conditioning reduces entropy as \( H(\phi_3 | \phi_1) \leq H(\phi_3) \), \( H(\phi_2 | \phi_1) \leq H(\phi_2) \) and \( H(\phi_4 | \phi_1) \leq H(\phi_4) \), and because of the cardinalities of the encoding functions. By converse to the source coding theorem

\[
d_4 \geq D(1/n I(X^n; \hat{X}_1^n)) \geq \sigma_4^2 \exp(-2[R_1 + R_2 + R_3 + R_4]) \exp[-2/n I(\phi_2; \phi_3 | \phi_1)].
\]

Because \( \hat{X}_2^n \) and \( \hat{X}_3^n \) are functions of \( \phi_1, \phi_2 \) and \( \phi_1, \phi_3 \), respectively, it is seen that

\[
I(\phi_2; \phi_3 | \phi_1) \geq I(\hat{X}_2^n; \hat{X}_3^n | \phi_1).
\]

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

\[
I(\hat{X}_2^n; \hat{X}_3^n | \phi_1) = I(\hat{X}_2^n; \hat{X}_3^n | \hat{X}_1^n) + I(\hat{X}_2^n; \hat{X}_3^n | \hat{X}_2^n) - I(\hat{X}_3^n; \hat{X}_1^n | \hat{X}_2^n).
\]
we have

\[ I(\hat{X}_2^n; \widehat{X}_3^n | \phi_1) \geq I(\hat{X}_2^n; Y^n | \phi_1) + I(\hat{X}_3^n; Y^n | \phi_1) - I(\hat{X}_2^n, \hat{X}_3^n; Y^n | \phi_1) \]

\[ = h(Y^n | \phi_1) - h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) + h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ = h(Y^n | \phi_1) - h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) + h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ = I(Y^n; \hat{X}_2^n | \phi_1) + I(Y^n; \hat{X}_3^n | \phi_1) - 2h(Y^n) + h(Y^n | \phi_1) + h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ \geq I(Y^n; \hat{X}_2^n | \phi_1) + I(Y^n; \hat{X}_3^n | \phi_1) - 2h(Y^n) + h(Y^n | \phi_1) + h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1). \]  

(34)

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

\[ I(Y^n; \hat{X}_2^n) \geq nR_Y(d_2 + \epsilon) = \frac{n}{2} \log \frac{d_2^2 + \epsilon}{d_2^2 + \epsilon}, \]  

(35)

where \( R_Y(D) \) is the rate distortion function for source \( Y \). Similarly

\[ I(Y^n; \hat{X}_3^n) \geq \frac{n}{2} \log \frac{\sigma^2}{d_2^2 + \epsilon}. \]  

(36)

The following steps in our converse proof are different from Ozarow's, and it is worth noting the complication introduced by the coding function \( \phi_1 \). We apply the conditional entropy power inequality [27] on the term \( h(Y^n | \phi_1) \), which gives

\[ h(Y^n | \phi_1) \geq \frac{n}{2} \log \left[ \exp \left( \frac{2}{n} h(X^n | \phi_1) \right) + 2\pi e \right]. \]  

(37)

However notice that

\[ h(X^n | \phi_1) = h(X^n) - I(X^n; \phi_1) \geq \frac{n}{2} \log (2\pi e \sigma^2) - nR_1, \]

which gives

\[ h(Y^n | \phi_1) \geq \frac{n}{2} \log \left[ \exp \left( \frac{2}{n} \log (2\pi e \sigma^2) - nR_1 \right) + 2\pi e \right] \]

\[ = \frac{n}{2} \log \left[ 2\pi e \sigma^2 \exp (-2R_1 + \epsilon) \right]. \]  

(38)

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

\[ h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \geq \frac{n}{2} \log \left[ \exp \left( \frac{2}{n} h(X^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \right) + 2\pi e \right]. \]

It follows that

\[ h(X^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) = h(X^n | \phi_1) - I(X^n; \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ = h(X^n) - I(X^n; \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ = h(X^n) - I(X^n; \phi_1) - H(\hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ - H(\hat{X}_2^n | \phi_1) + I(\hat{X}_2^n; \hat{X}_3^n | \phi_1) \]

\[ \geq \frac{n}{2} \log (2\pi e \sigma^2) - n(R_1 + R_2 + R_3) + I(\hat{X}_2^n; \hat{X}_3^n | \phi_1). \]

where (a) follows from the fact that \( \hat{X}_2^n \) and \( \hat{X}_3^n \) are functions of \( X^n \), and (b) from the fact \( \hat{X}_2^n \) and \( \hat{X}_3^n \) are functions of \( (\phi_1, \phi_2) \) and \( (\phi_1, \phi_3) \), respectively. This leads to

\[ h(Y^n | \hat{X}_2^n, \hat{X}_3^n | \phi_1) \]

\[ \geq \frac{n}{2} \log \left( 2\pi e \right) \]

\[ + \exp \left( \frac{2}{n} \log (2\pi e \sigma^2) - n \sum_{i=1}^{3} R_i + I(\hat{X}_2^n; \hat{X}_3^n | \phi_1) \right) \]

\[ = \frac{n}{2} \log \left( 2\pi e \right) \left( \sigma^2 \exp \left( -2 \sum_{i=1}^{3} R_i \right) \exp \left( \frac{2}{n} I(\hat{X}_2^n; \hat{X}_3^n | \phi_1) + \epsilon \right) \right). \]  

(39)

Define

\[ t \triangleq \exp \left( \frac{2}{n} I(\hat{X}_2^n; \hat{X}_3^n | \phi_1) \right), \]  

(40)

and summarize all the bounds in (34), (35), (36), (38) and (39), and we thus have

\[ t \geq \frac{d_2^2 + \epsilon}{(d_2 + \epsilon)(d_3 + \epsilon)} \left[ d_4 I(d_1^2 + \epsilon) \right] \]  

(41)

Isolating \( t \) we have

\[ t \geq \frac{d_2^2 + \epsilon}{(d_2 + \epsilon)(d_3 + \epsilon) - (d_2^2 + \epsilon)(d_1^2 + \epsilon) \exp [-2(R_2 + R_3)]}. \]  

(42)

notice that because \( d_2 \geq d_1^2 \exp (-2R_2) \) and \( d_3 \geq d_1^2 \exp (-2R_3) \) from (31), the denominator is always positive, as long as \( \epsilon \) is positive. To get the tightest bound, we maximize the lower bound on \( t \) over \( \epsilon \). When \( d_2 \leq d_1^2 \) and \( d_3 \leq d_1^2 \), define

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

(43)

Then we choose the following value of \( \epsilon \)

\[ \epsilon = \min \left\{ \frac{d_2^2 + \epsilon}{(d_2 + \epsilon)(d_3 + \epsilon)}, \quad \Pi^* \geq \Delta^*; \right. \]

\[ \left. \text{otherwise.} \right\} \]  

(44)

After some algebraic calculation, we have for the case \( \Pi^* \geq \Delta^* \),

\[ t \geq \frac{1}{1 - (\Pi^* - \Delta^*)^2}, \]  

(45)

and subsequently using (32), (33), (40) and (45)

\[ d_4 \geq \frac{\sigma^2 \exp [-2(R_1 + R_2 + R_3 + R_4)]}{1 - (\Pi^* - \Delta^*)^2}. \]  

(46)

For the case \( \Pi^* < \Delta^* \), we have \( t \geq 1 \), which gives the trivial bound of

\[ d_4 \geq \sigma^2 \exp [-2(R_1 + R_2 + R_3 + R_4)]. \]  

(47)

This is not yet the bound given in Theorem 2 since the definitions of \( \Pi \) and \( \Delta \) are not the same as those of \( \Pi^* \) and \( \Delta^* \), which are only defined when \( d_2 \leq d_1^2 \) and \( d_3 \leq d_1^2 \). To close this gap, note that if \( d_2 \geq d_1^2 \) (or \( d_3 \geq d_1^2 \)), we may trivially write the lower bound \( \Pi^* \), which coincides with (35) for this case, due to the fact \( \Pi = 0 \) by the definition of \( d_2 \) (or \( d_3 \)). Thus the lower bound in Theorem 2 is established.

**Forward:** Now we shall use the general achievable region given in Theorem 1 to derive an inner bound for the Gaussian source, and show it coincides with the outer bound.

Construct the following random variables

\[ U_1 = X + N_1, \quad U' = X - E(X U_1), \quad U_2 = X' + N_2, \]

\[ U_3 = X' + N_3, \quad U_4 = X' + N_4. \]  

(48)

(49)
where $N_1, N_2, N_3, N_4$ are zero mean jointly Gaussian, independent of $X$, and having the covariance matrix

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

(50)

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

$$
\begin{align*}
\hat{X}_1 &= f_1(U_1) = \mathbb{E}(X|U_1) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} U_1, \\
\hat{X}_2 &= f_2(U_1, U_2) = \hat{X}_1 + \mathbb{E}(X'|U_2), \\
\hat{X}_3 &= f_3(U_1, U_3) = \hat{X}_1 + \mathbb{E}(X'|U_3), \\
\hat{X}_4 &= f_4(U_1, U_2, U_3, U_4) = \hat{X}_1 + \mathbb{E}(X'|U_2 U_3 U_4).
\end{align*}
$$

(51) - (54)

We have that the following rate $R_3$ is achievable

$$R_3 \geq I(X; U_1) = \frac{1}{2} \log \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2}.
$$

(55)

Choose $\sigma_1$ such that the above inequality holds with equality. Then we have

$$d_1 = \mathbb{E}(X'^2) = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \sigma_2^2 \exp(-2R_1) = d_1^*.
$$

(56)

We have also

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

(57)

$$d_3 = \mathbb{E}[(X - \hat{X}_3)^2] = \mathbb{E}[X' - \mathbb{E}(X'|U_3)^2] = \frac{d_1 \sigma_2^2}{d_1 + \sigma_2^2}.
$$

(58)

Note that for optimal distortion quadruples, $d_2 \leq d_1^*$ and $d_3 \leq d_1^*$. It should be chosen, i.e., $d_2 = d_2^*$ and $d_3 = d_3^*$.

Notice also

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

(59)

where (a) is true because $U_1$ is independent of $X'$ and $N_2$. Choose $\sigma_2^2$ such that

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

(60)

which can always be done because the function is continuous. Similarly choose $\sigma_3^2$ such that

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

(61)

From (60), the sum rate satisfying the following bound is achievable

$$R_1 + R_2 + R_3 \geq I(X; U_1) + I(X'; U_2 U_3 | U_1) + I(U_2; U_3 | U_1) = R_1 + I(X'; U_2 U_3 | U_1) + I(X' + N_2; X' + N_2 | U_1) = R_1 + I(X'; U_2 U_3) + I(U_2; U_3) = R_1 - h(N_2 N_3) + h(U_2) + h(U_3),
$$

which gives

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

(62)

When $\Pi \geq \Delta$, we may choose

$$\rho = -\sqrt{1 - \frac{d_1^2 \exp(-2(R_2 + R_3))}{d_2 d_3}}
$$

(63)

Then

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

where

$$d_4^* \overset{\Delta}{=} \mathbb{E}(X' - \mathbb{E}(X'|U_2 U_3))^2 = \frac{d_1 \sigma_2^2 \sigma_3^2 (1 - \rho^2)}{d_1^2 \sigma_3^2 \sigma_4^2 (1 - \rho^2) + d_1 (\sigma_2^2 + \sigma_3^2) - 2\rho d_1 \sigma_2 \sigma_3},
$$

(64)

Choose $\sigma_4^2$ such that

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

(65)

We further have (and after some simplification)

$$d_4 = \mathbb{E}(X' - \mathbb{E}(X'|U_2 U_3) \overset{(a)}{=} \frac{\exp(-2R_4) d_1 \sigma_2^2 \sigma_3^2 (1 - \rho^2)}{d_1^2 \sigma_3^2 \sigma_4^2 (1 - \rho^2) + d_1 \sigma_2^2 + \sigma_3^2 - 2\rho d_1 \sigma_2 \sigma_3},
$$

(66)

$$d_4 = \mathbb{E}(X' - \mathbb{E}(X'|U_2 U_3)^2 \overset{(b)}{=} \frac{\exp(-2(R_1 + R_2 + R_3))}{d_1^2 \sigma_3^2 \sigma_4^2 (1 - \rho^2) + d_1 \sigma_2^2 + \sigma_3^2 - 2\rho d_1 \sigma_2 \sigma_3},
$$

(67)

$$1 - \sqrt{\frac{\Delta}{\Pi - \Delta}}
$$

(68)

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 $d_1 + d_1 \exp[-2(R_1 + R_2 + R_3)] = d_2^* + d_3^*$. To see this, let us first consider choosing $d_1^* \in [d_1, 2^{-2\Pi m}, m = 2, 3]$. By a continuity argument, $d_2^* + d_3^*$ such chosen can take any real value in $[d_1 \exp[-2R_2 + d_3^* \exp(-2R_2)]$, $d_2 + d_3^*$. It is obvious that $d_1^* = d_1 + d_1 \exp[-2(R_2 + R_3)]$ for $d_2 + d_3^*$, and it follows there always exists at least a pair of $(d_2^*, d_3^*)$ such that $d_2^* + d_3^* = d_2 + d_3$, $d_1 + d_1 \exp[-2(R_2 + R_3)]$, i.e., $\Pi = \Delta$. Thus we can conclude the distortion quadruple $(d_1^*, d_2^*, d_3^*, \exp(-2(R_1 + R_2 + R_3)))$ is achievable by (63), which implies that the quadruple $(d_1, d_2, d_3, \exp(-2(R_1 + R_2 + R_3)))$ is achievable. Therefore for both the case $\Pi > \Delta$ and $\Pi < \Delta$, the achievable region 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 former 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 $\sigma_2^2 \exp[-2R_1]$ and $\sigma_2^2 \exp[-2(1 + \alpha)R_1]$. 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 $\sigma_2^2 \exp[-2(R_1 + R_3)]$ and $\sigma_2^2 \exp[-2(1 + \alpha)(R_1 + R_3)]$.

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

$$d_2 + \sigma_2^2 \exp[-2(R_1 + R_3)] \geq d_1^*[1 + \exp[-2(R_2 + R_3)]],
$$

3This degenerate region was not treated by Ozarow in [6], and it sometimes causes certain confusion.
which further gives
\[
d_2 \geq d_1^0 [1 + \exp(-2(2\alpha R_1 + R_3) - \exp(-2R_3)].
\]

For a single layer system optimized for the first layer, distortion \( d_1^0 = \sigma_1^2 \exp(-2R_1) \) and \( d_2^0 = \sigma_2^2 \exp[-2(1 + \alpha)R_1] \) are achievable, thus the loss on \( d_2 \) can be as large as
\[
\frac{d_2}{d_2^0} = \frac{[1 + \exp(-2(2\alpha R_1 + R_3))] - \exp(-2R_3)]}{\exp(-2\alpha R_1)}
= \exp(2\alpha R_1) + \exp(-2R_3) - \exp[2(\alpha R_1 - R_3)].
\]

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 \( (R_1, R_2, R_3, (1 + \alpha)(R_1 + R_3)) \), 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.

To compare the two coding schemes, fix \( d_1 = d_1^0 = \sigma_1^2 \exp(-2R_1) \) and \( d_2 = d_2^0 = \sigma_2^2 \exp[-2(R_1 + R_2)] \). For the WZ-based scheme, since the \( \phi_3 \) and \( \phi_4 \) are successive refinement by definition, the WZ-based coding is in fact the successive Wyner-Ziv based approach, since the side informations at the decoder. In this section we compare the operating points: one point is when \( d_3 = d_3^0 \) and the other one is \( d_3 = \sigma_3^2 \exp[-2(R_1 + R_3)] \), which correspond to description \( \phi_3 \) is either completely useless for decoder 3, or maximally useful for decoder 3, respectively. 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]. Though the Wyner-Ziv coding based approach is sub-optimal in this problem, it does have certain advantage, particularly when the second round of transmission is not encoded together with the first round descriptions, but is made possible when certain network resource becomes available after an initial transmission.

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 except two extreme operating points: one point is when \( d_3 = d_3^0 \) and the other one is \( d_3 = \sigma_3^2 \exp[-2(R_1 + R_3)] \), which is depicted in Fig. 5. The connection is through \( I \setminus \mathcal{A}_2 \), rather than only \( I \setminus \mathcal{A}_2 \).

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 \( d_1 = d_1^0 \) and \( d_2 = d_2^0 \). The range of \( d_3 \) is \([\sigma_3^2 \exp[-2(R_1 + R_3)], d_3^0]\), as well as
\[
\Delta = \frac{d_2^2 d_3^3 - \exp[-2(2R_2 + R_3)]}{d_1^2 d_1^4} - \exp(-2R_3). \tag{76}
\]

As 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 except two extreme operating points: one point is when \( d_3 = d_3^0 \) and the other one is \( d_3 = \sigma_3^2 \exp[-2(R_1 + R_3)] \), which correspond to description \( \phi_3 \) is either completely useless for decoder 3, or maximally useful for decoder 3, respectively. 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]. Though the Wyner-Ziv coding based approach is sub-optimal in this problem, it does have certain advantage, particularly when the second round of transmission is not encoded together with the first round descriptions, but is made possible when certain network resource becomes available after an initial transmission.

The MSR problem in fact has a more subtle connection with the encoder/decoder side information (EDSI) problem considered in [9] (see also [29]), which is depicted in Fig. 5. The connection is through one particular special case for the EDSI problem, when the source and side informations are physically degraded as \( X \leftrightarrow Y \leftarrow Z \), and in the Gaussian case we may write without loss of generality \( Y = X + N_1 \) 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, but we can indeed choose them to have such structure with single-letter distribution \( P_{ZX} \); furthermore, in the MSR problem, decoder one always has \( \hat{X}_1 \) (corresponding to \( Z \)), rather than only \( \hat{X}_2 \) (corresponding to the better side information \( Y \)). As such, if in the MSR system we set \( R_1 = I(X; Z) \) and \( R_2 = I(X; Y|Z) \), then \( D_1 = \mathbb{E}[X - \hat{X}(X|Z)^2 \) and \( D_2 = \mathbb{E}[X - \hat{X}(X|Y)^2 \), an outer bound for the EDSI problem can be found; it is an outer
bound since we can use $P_{XZY}$ as $P_{XU_1U_2}$ in the minimization for the MSR problem, but the chosen $(R_1, R_2, D_1, D_2)$ also allows for other choices of random variables. As shown in [9], this outer bound is indeed achievable by using a hybrid conditioning/binning scheme.

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 $\phi_2, \phi_3, \phi_4$ and assuming $R_1 = 0$ and no description $\phi_1$ exists.

El Gamal and Cover constructed an MD scheme for general sources based on the MDCR method in [7]. More precisely, the description $\phi_4$ of rate $R_4$ is split and combined into the existing two descriptions $\phi_2$ and $\phi_3$, and the resulting two descriptions are of rates $R_3 = R_2 + \beta R_4$ and $R_4' = R_4 + (1 - \beta) R_4$ for some $0 \leq \beta \leq 1$. It is 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 $R_4'$. 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 4. To see this, assume the distortion $d_2$ and $d_3$ in both an MDCR-based system and an optimal MD system, such that $d_2 < d_1'$ and $d_3 < d_1'$. By Theorem 4, we see that for an MDCR system with non-zero $R_4$

$$d_4 \geq \frac{\sigma_2^2 \exp[-2(R_2 + R_4 + R_3)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2}$$

(77)

where

$$\Delta = \frac{d_2 d_3}{\sigma_2^2} - \exp[-2(R_2 + R_3)], \quad \Pi = (1 - \frac{d_2}{\sigma_2})(1 - \frac{d_3}{\sigma_2}).$$

(78)

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

$$d_4' \geq \frac{\sigma_2^2 \exp[-2(R_2 + R_3)]}{1 - (\sqrt{\Pi'} - \sqrt{\Delta'})^2}$$

(79)

where

$$\Delta' = \frac{d_2 d_3}{\sigma_2^2} - \exp[-2(R_2' + R_3')], \quad \Pi' = (1 - \frac{d_2}{\sigma_2})(1 - \frac{d_3}{\sigma_2}).$$

(80)

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 $0 < \beta < 1$; further assume $d_2$ and $d_3$ are chosen such that $\Pi' > \Delta'$, i.e., the MD system does not operate in the degenerate region. 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 in general.

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 [31].

A. The high-rate asymptotics for balanced descriptions

We consider balanced MDs in this subsection, and further assume $\sigma_2^2 = 1$. 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 $0 < \eta < 1$ and $b \geq 1$, the central distortion of an MD system can asymptotically (at low distortion) achieve

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

(81)

Notice the condition $0 < \eta < 1$ in fact corresponds to the condition that $1 > d_2'$. This result suggests any system based on the MDCR approach is appealing for an optimal rate $R' = R_4/2$. Keeping the side distortion of this system $d_2' = d_3' = b 2^{-2(1 - \eta) R'}$ for an easier comparison, consider the case $1 > \eta > 0$, and let $R_4 = 2\eta R'$, where $1 - \eta$ 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

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

(82)

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 $1 > d_2$ and $d_2' > d_2$. However, as the rate allocated to the refinement stage increases, the excess marginal rate in the first stage decreases. When $\eta = \eta_1$, the performance loss is a factor of $2 \sqrt{\eta_1}$. If the first stage is without excess marginal rate, which means $\eta_1 = 1$ 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 $1 > d_2$ and $d_2' > d_2$ is the desired operating range. However, the first stage should reserve sufficient excess marginal rate in order to avoid the performance loss. Taking the MD system in [31] as an example, using certain sub-optimal lattices for $\phi_2$ and $\phi_3$ is potentially able to achieve (asymptotic) optimal performance, but using two good lattices as $\phi_2$ and $\phi_3$ 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 is provided, which is shown to be tight for the
Gaussian source under MSE measure. It is 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 is treated with some depth, which reveals some interesting implications in designing the MD coding system.

For the general problem with an arbitrary $K > 2$ rounds of transmission, or $K > 2$ users, an achievable region can be derived using the technique developed in [34] and [21]. However, even for the Gaussian case, the problem is highly intractable, and a complete characterization appears difficult. Given the results in the current work, we expect the loss of performance for the general case to be more severe than the $K = 2$ case.

REFERENCES

[1] S. Pradhan and K. Ramchandran, “Multiuser successive refinement,” in 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. Peron, S. Diggavi, and E. Telatar, “On the role of encoder side-information in source coding for multiple decoders,” in Proc. 2006 International Symposium on Information Theory, Seattle, WA, Jul. 2006, pp. 331–335.

[10] C. Tian, J. Chen, S. Hemami, and T. Berger, “Multiple descriptions with central refinement,” in 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, Jun-Jul 2003, p. 195.

[12] H. Feng and M. Effros, “Nearly separability of optimal multiple description source codes,” IEEE Trans. Information Theory, submitted for publication.

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

[14] Y. Steinberg and N. Merhav, “On hierarchical joint source-channel coding with degraded side information,” IEEE Trans. Information Theory, vol. 52, no. 3, pp. 886–903, Mar. 2006.

[15] C. Tian and S. Diggavi, “On multistage successive refinement for wyner-ziv source coding with degraded side information,” IEEE Trans. Information Theory, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.

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

[17] L. A. Lastras and V. Castelli, “Near sufficiency of random coding for two descriptions,” IEEE Trans. Information Theory, vol. 52, no. 2, pp. 618–695, Feb. 2006.

[18] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” IEEE Trans. Information Theory, vol. 33, no. 4, pp. 502–521, Jul. 1987.

[19] H. Wang and P. Viswanath, “Vector Gaussian multiple description with individual and central receivers,” IEEE Trans. Information Theory, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

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

[21] 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.

[22] S. Diggavi and V. Vaishampayan, “On multiple description source coding with decoder side information,” in Proc. IEEE Information Theory Workshop, San Antonio, TX, Oct. 2004.

[23] 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.

[24] R. G. Gallager, Information theory and reliable communication, New York: John Wiley, 1968.

[25] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder II: general sources,” Inform. cont., vol. 38, pp. 60–80, 1978.

[26] J. Chen, C. Tian, S. Hemami, and T. Berger, “Multiple description quantization via Gram-Schmidt orthogonalization,” IEEE Trans. Information Theory, vol. 52, no. 12, pp. 5197–5217, Dec. 2006.

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

[28] 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.

[29] J. K. Wolf, “Source coding for a noiseless broadcast channel,” in Conference on Information Sciences and Systems, Mar. 2004.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple description coding with many channels,” IEEE Trans. Information Theory, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.