On Scalable Source Coding for Multiple Decoders with Side Information

Chao Tian
School of Computer and Communication Sciences
Laboratory for Information and Communication
Systems (LICOS), EPFL, Lausanne, Switzerland.
Email: chao.tian@epfl.ch

Suhas N. Diggavi
School of Computer and Communication Sciences
Laboratory for Information and Communication
Systems (LICOS), EPFL, Lausanne, Switzerland.
Email: suhas.diggavi@epfl.ch

***Abstract*— The problem of side-information scalable (SI-scalable) source coding and the problem of successive refinement in the Wyner-Ziv (SR-WZ) setting are considered in this paper. Both problems can be understood as special cases of the general problem of scalable lossy source coding for multiple decoders with access to side informations, but the encoder does not. In the former problem, the quality of the side informations deteriorates from the early stage to the later stage, while in the latter problem, the quality of the side informations improves with the stages. The SR-WZ problem was considered by Steinberg and Merhav (T-IT, 2004) where it was solved for the special case of two stages. We provide a generalization of their characterization to multiple stages. We also provide achievable rate and outer bounds for the SI-scalable coding problem. Furthermore, the notion of generalized successively refinability with multiple side informations is introduced, which captures whether progressive encoding to satisfy the distortion constraints for different side information is as good as encoding without progressive requirement. For the quadratic Gaussian case, by combining the results of SI-scalable and SR-WZ coding, we provide a complete answer to the general scalable source coding problem.**

I. Introduction

For the scalable source coding with side information at the decoder(s), we consider encoding a source X into a coarse description of rate R_1 and a refinement description of rate R_2, such that the coarse description can be decoded with side information Y_1 to achieve distortion D_1, and the coarse and refinement descriptions together can be decoded with side information Y_2 to achieve distortion D_2. In order to make it explicit, consider the following two scenarios:

- A server is to broadcast multimedia data stream to multiple users, who have different side informations and maybe also different (viewing) quality requirements. A straightforward coding strategy is to form a description and broadcast it to all the users, who can decode only after receiving the complete description to satisfy the requirement of every user. Therefore, for the users who have strong side information, this strategy causes unnecessary delay. It is then natural to ask whether the encoding can be performed in an opportunistic way. More specifically, we want to construct a bitstream in a progressive manner, such that the users with strong side information can truncate the bitstream and decode, and the users with weak side information will have to receive further data to decode. We will call this kind of coding *side-information scalable* (SI-scalable) source coding [15].
- The encoder constructs the description and transmit it to the decoder, who at the same time is receiving from other routes about the source as side information. From the encoder point of view, the side information at the decoder is improving as it transmits. If a scalable description is constructed by the encoder, the decoder can decode while receiving the description and reconstruct with minimal delay. This is the problem of successive refinement in the Wyner-Ziv setting (SR-WZ), which was recently considered by Steinberg and Merhav [13] (see also [9]).

In the seminal work [17], Wyner and Ziv characterized the rate distortion function for the problem of source coding with side information at the decoder. Therefore, both the above questions are related to a scalable version of the Wyner-Ziv question, where a source is to be encoded in a scalable manner to satisfy (different) distortion requirement at each individual stage. The successive refinement problem was studied by Koshelev [8], and Equitz and Cover [4], whose interest was to determine whether such a progressive encoding incurs any rate loss compared with non-progressive encoding. Rimoldi [10] provided a complete characterization of the rate-distortion region for this problem. Another relevant work is [6] (see also [7]), where the problem of source coding when side information may be absent at the decoder was considered; the result was extended to the multistage case when the side informations are degraded. This is quite similar to the problem being considered here, however without the scalable coding requirement. It can be seen that in this problem, as well as the problems being considered, multiple side informations (and multiple distortion requirements) exist, which implies the encoder in effect is uncertain about which side information can occur. It is thus expected that this uncertainty introduces more difficulty and may also make a higher coding rate necessary.

It is worth pointing out that for the lossless case, both problems are quite straight-forward, and are well understood. The key difference from lossy case is that the quality of the side informations can be naturally determined by the value of $H(X|Y)$. By the seminal work of Slepian and

Wolf [12], $H(X|Y)$ is the minimum rate of encoding X losslessly with side information Y at the decoder, thus in a sense a larger $H(X|Y)$ corresponds to weaker side information. If $H(X|Y_1) \geq H(X|Y_2)$, then the SR-WZ problem degenerates into encoding a single description, and a rate of $H(X|Y_1)$ suffices as shown by Sgarro [11]; on the other hand, if $H(X|Y_1) < H(X|Y_2)$, then the rate $(R_1, R_2) = (H(X|Y_1), H(X|Y_2) - H(X|Y_1))$ is also achievable, as pointed out by Feder and Shulman [5]. Extending this observation and a coding scheme in [1], Draper [2] proposed a universal incremental Slepian-Wolf coding scheme when the distribution is unknown, which inspired Eckford and Yu [3] to design rateless Slepian-Wolf LDPC code.

For the lossless case, there is no loss of optimality by using a scalable coding approach. An immediate question is to ask whether the same is true for the lossy case in terms of rate distortion, which we will show to be not so in general. In order to make the notion of side-information quality precise, we use the Markov condtions: $X \leftrightarrow Y_1 \leftrightarrow Y_2$, for SI-scalable coding and respectively $X \leftrightarrow Y_1 \leftrightarrow Y_2$, for the SR-WZ problem. The SR-WZ problem was recently solved by Steinberg and Merhav [13]. However, because of the reversal of the order in side-information quality, the two problems are quite different.

In this paper, we will provide an achievable region and outer bounds for the SI-scalable coding problem, then relate this result to that for SR-WZ coding. For the quadratic Gaussian case with Gaussian side informations, these bounds completely characterize the rate-distortion region for SI-scalable coding. As a byproduct, the solution for the Gaussian source to SI-scalable coding and SR-WZ coding provides the complete solution for the general scalable coding problem; however, for general sources and distortion measures, the general SI-scalable problem remains open. But, for the SR-WZ problem, we provide a generalization of the characterization given by Steinberg and Merhav for arbitrary number of stages, a question left open in [13].

This paper is organized as follows. In Section II, we formally state the problem and give the notation used in the paper. Section III gives an achievable rate region for the side-information scalable problem as well as some outer bounds. A comparison of side-information scalable source coding and the SR-WZ considered in [13] is made in Section IV. The complete characterization for the special case of the Gaussian quadratic source degraded and reversely degraded side-information cases is given in Section V, which motivates the introduction of generalized successive refinability. We end with a short discussion in Section VI.

II. Notation and Preliminaries

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 $\boldsymbol{x}$ when the dimension n is clear from the context. Upper case is used for random variables and vectors. A discrete memoryless source (DMS) $(\mathcal{X}, P_X)$ is an infinite sequence $\{X\}_{i=1}^{\infty}$ of independent copies of a random variable X in $\mathcal{X}$ with a generic distribution P_X such that the joint distribution is $P_X(x^n) = \prod_{i=1}^n P_X(x_i)$. Let $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, P_{XY_1Y_2})$ be similarly defined.

Let $\hat{\mathcal{X}}$ be a finite reconstruction alphabet, and for simplicity we assume both decoders use this same reconstruction alphabet. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be a distortion measure. The per-letter distortion of a vector is defined as

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

***Definition** 1:* An (n, M_1, M_2, D_1, D_2) rate distortion (RD) scalable code for source X with side information (Y_1, Y_2) consists of 2 encoding functions ϕ_i, $i = 1, 2$, and 2 decoding functions ψ_i, $i = 1, 2$

$$\phi_1 : \mathcal{X}^n \to I_{M_1}, \quad \phi_2 : \mathcal{X}^n \to I_{M_2} \tag{2}$$

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

$$\psi_1 : I_{M_1} \times \mathcal{Y}_1^n \to \hat{\mathcal{X}}^n, \quad \psi_2 : I_{M_1} \times I_{M_2} \times \mathcal{Y}_2^n \to \hat{\mathcal{X}}^n, \tag{3}$$

such that

$$\mathbb{E} d(X^n, \psi_1(\phi_1(X^n), Y_1^n)) \leq D_1, \tag{4}$$

$$\mathbb{E} d(X^n, \psi_2(\phi_1(X^n), \phi_2(X^n), Y_2^n)) \leq D_2 \tag{5}$$

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

***Definition** 2:* A rate pair (R_1, R_2) is said to be (D_1, D_2)-achievable for scalable encoding with side information (Y_1, Y_2), if for any $\epsilon > 0$ and sufficiently large n, there exist an $(n, M_1, M_2, D_1 + \epsilon, D_2 + \epsilon)$ RD scalable code with side informations, such that $R_1 + \epsilon \leq \frac{1}{n} \log(M_1)$ and $R_2 + \epsilon \leq \frac{1}{n} \log(M_2)$.

Denote the collection of all the (D_1, D_2)-achievable rate pair (R_1, R_2) for scalable encoding with side informations as $\mathcal{R}(D_1, D_2)$, and it is the region to be characterized.

The rate-distortion function with degraded side-informations was established in [6] for the non-scalable coding problem. In light of the discussion in Section I, it gives a lower bound on the sum-rate for any RD scalable code with degraded side informations. More precisely, in order to achieve distortion D_1 with side information Y_1, and achieve distortion D_2 together with side information Y_2, when $X \leftrightarrow Y_1 \leftrightarrow Y_2$, the rate-distortion function is:

$$R_{HB}(D_1, D_2) = \min_{p(D_1, D_2)} [I(X; W_2|Y_2) + I(X; W_1|W_2, Y_1)]$$

where $p(D_1, D_2)$ is the set of all random variable $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$ jointly distributed with the generic random variables (X, Y_1, Y_2), such that the following conditions are satisfied[1]: $(W_1, W_2) \leftrightarrow X \leftrightarrow Y_1 \leftrightarrow Y_2$ is a Markov string; and $\hat{X}_1 = f_1(W_1, Y_1)$ and $\hat{X}_2 = f_2(W_2, Y_2)$ satisfy the distortion constraints.

[1]This form is slightly different than the one in [6] where f_1 was defined as $f_1(W_1, W_2, Y)$, but it is straightforward to verify their equivalence.

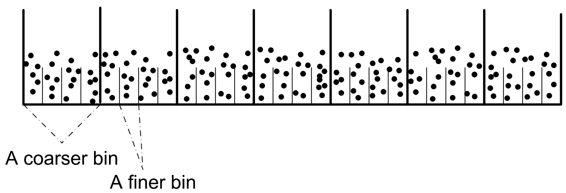

Fig. 1. An illustration of the nested binning structure: each dot stands for a codeword.

III. An Achievable Region and Outer Bounds for Lossy SI-Scalable Coding

Define the region $\mathcal{R}_{ach}(D_1, D_2)$ to be the set of all rate pairs (R_1, R_2) for which there exist random variables (W_1, W_2, V) in finite alphabets $\mathcal{W}_1, \mathcal{W}_2, \mathcal{V}$ such that the following condition are satisfied.

1) $(W_1, W_2, V) \leftrightarrow X \leftrightarrow Y_1 \leftrightarrow Y_2$.
2) There exist deterministic maps $f_j : \mathcal{W}_j \times \mathcal{Y}_j \rightarrow \hat{\mathcal{X}}$ such that

$$\mathbb{E}d(X, f_j(W_j, Y_j)) \leq D_j, \quad j = 1, 2. \tag{6}$$

3) The non-negative rate pairs satisfy:

$$\begin{aligned} R_1 &\geq I(X; V, W_1 | Y_1) \\ R_1 + R_2 &\geq I(X; V, W_2 | Y_2) + I(X; W_1 | Y_1, V). \end{aligned}$$

The following theorem asserts that $\mathcal{R}_{ach}(D_1, D_2)$ is an achievable region.

Theorem *1:* For any discrete memoryless stochastically source with reversely degraded side information, i.e., $X \leftrightarrow Y_1 \leftrightarrow Y_2$,

$$\mathcal{R}(D_1, D_2) \supseteq \mathcal{R}_{ach}(D_1, D_2). \tag{7}$$

To further illustrate the structure of the inner bound, we write

$$I(X; V, W_1 | Y_1) = I(X; V | Y_1) + I(X; W_1 | Y_1, V)$$

and

$$\begin{aligned} I(X; V, W_2 | Y_2) + I(X; W_1 | Y_1, V) = I(X; V | Y_1) \\ + I(X; W_1 | Y_1, V) + I(V; Y_1 | Y_2) + I(X; W_2 | Y_2, V). \end{aligned} \tag{8}$$

The basic idea of the coding strategy can be understood as follows. The first stage encodes $\boldsymbol{V}$ using (perhaps a "coarse") binning, such that the first stage is able to decode it with side information $\boldsymbol{Y_1}$. If necessary, a Wyner-Ziv successive refinement coding (with side information $\boldsymbol{Y_1}$) is added conditioned on the codeword $\boldsymbol{V}$ for the first stage using $\boldsymbol{W_1}$. The second stage enumerates the binning of $\boldsymbol{V}$ upto a level such that $\boldsymbol{V}$ is decodable by the second stage using the weaker side information $\boldsymbol{Y_2}$; if necessary, a Wyner-Ziv successive refinement coding (with side information $\boldsymbol{Y_2}$) is added conditioned on the codeword $\boldsymbol{V}$ for the second stage using $\boldsymbol{W_2}$. The rate $I(V; Y_1 | Y_2) = I(V; Y_1) - I(V; Y_2)$ is used to enumerate the finer bin index within the coarser bin.

Therefore, the key idea to prove this achievable region is simple, which is to use nested binning as illustrated in Fig. 1. Each of the coarser bin contains the same number of finer bins; each finer bin holds certain number of codewords (typical sequences of $\boldsymbol{V}$). They are constructed in such a way that given the specific coarser bin index, the first stage decoder can decode in it with the strong side information; at the second stage, additional bitstream is received by the decoder, which further specifies one of the finer bin in the coarser bin, such that the second stage decoder can decode in this finer bin using the weaker side information. If we assign each codeword, which is a typical sequence of $\boldsymbol{V}$, to a finer bin independently, then its coarser bin index is also independent of that of the other codewords.

An obvious outer bound is given by the intersection of the Wyner-Ziv rate distortion function (denoted by $R_{X|Y_1}(D_1)$) and the rate-distortion function for the problem considered by Heegard and Berger [6] with degraded side information $X \leftrightarrow Y_1 \leftrightarrow Y_2$

$$\begin{aligned} \mathcal{R}_{\cap}(D_1, D_2) = \{(R_1, R_2) : R_1 \geq R_{X|Y_1}(D_1), \\ R_1 + R_2 \geq R_{HB}(D_1, D_2)\}. \end{aligned} \tag{9}$$

A tighter outer bound is also given in [15]. Let us define the region $\mathcal{R}_{out}(D_1, D_2)$ to be the set of all rate pairs (R_1, R_2) for which there exist random variables (W_1, W_2) in finite alphabets $\mathcal{W}_1, \mathcal{W}_2$ such that the following condition are satisfied.

1) $(W_1, W_2) \leftrightarrow X \leftrightarrow Y_1 \leftrightarrow Y_2$.
2) There exist deterministic maps $f_j : \mathcal{W}_j \times \mathcal{Y}_j \rightarrow \hat{\mathcal{X}}$ such that

$$\mathbb{E}d(X, f_j(W_j, Y_j)) \leq D_j, \quad j = 1, 2. \tag{10}$$

3) $|\mathcal{W}_1| \leq |\mathcal{X}| + 2$ and $|\mathcal{W}_2| \leq (|\mathcal{X}| + 1)^2$.
4) The non-negative rate vectors satisfies:

$$\begin{aligned} R_1 &\geq I(X; W_1 | Y_1) \\ R_1 + R_2 &\geq I(X; W_2 | Y_2) + I(X; W_1 | Y_1, W_2). \end{aligned}$$

Theorem *2:* For any discrete memoryless stochastically source with reversely degraded side information, i.e., $X \leftrightarrow Y_1 \leftrightarrow Y_2$,

$$\mathcal{R}_{\cap}(D_1, D_2) \supseteq \mathcal{R}_{out}(D_1, D_2) \supseteq \mathcal{R}(D_1, D_2). \tag{11}$$

The first inclusion of $\mathcal{R}_{\cap}(D_1, D_2) \supseteq \mathcal{R}_{out}(D_1, D_2)$ is obvious, since $\mathcal{R}_{out}(D_1, D_2)$ takes the same form as $R_{X|Y_1}(D_1)$ and $R_{HB}(D_1, D_2)$ when the rates R_1 and $R_1 + R_2$ are considered individually. The proof for the latter inclusion can be found in [15].

In comparing $\mathcal{R}_{ach}(D_1, D_2)$ and $\mathcal{R}_{out}(D_1, D_2)$, it is seen that the main difference is the additional auxiliary random variable V. Roughly speaking, the random variable V reflects the change of coding strategy for different distortion regions. The two coding strategies below are natural for the following two scenarios, which are unified by the achievable region $\mathcal{R}_{ach}(D_1, D_2)$:

- $D_1 \ll D_2$. In this case, the first stage codebook using W_1 is enough to achieve distortion D_2 at the second stage. Thus if W_1 could be decoded at the second stage, too much information has been given. Therefore the first-stage of the SI-scalable coding is constructed by first designing a SR-WZ code for *identical* side-information Y_1. This means that V and W_1 represent the first and second stage of a SR-WZ code with side-information Y_1 giving the coarser and finer descriptions. Furthermore, V is binned using the nested bins. Then in the second stage of the SI-scalable code, by specifying the finer bin index for V, the codeword V can be decoded with the weaker side information. In this case, $W_2 = V$.

- $D_1 \gg D_2$. In this case, the first stage codebook using W_1 is not sufficient to achieve the distortion D_2 at the second stage. Thus we let $W_1 = V$ and bin V using the nested bins. The second stage of the SI-scalable specifies the finer bin index, so that the decoder can find V. To obtain the lower distortion, we design a SR-WZ scheme as above with common side-information Y_2, where V and W_2 are the coarser and finer descriptions, respectively.

IV. Comparison of SI-scalable Coding and SR-WZ Coding

For the SR-WZ problem, it was shown in [13] that $\mathcal{R}(D_1, D_2)$ is equal to the following region $\mathcal{R}^*(D_1, D_2)$, which is defined as the set of all rate pairs (R_1, R_2) for which there exists random variables (W_1, W_2) in finite alphabets $\mathcal{W}_1, \mathcal{W}_2$ such that the following condition are satisfied[2].

1) $(W_1, W_2) \leftrightarrow X \leftrightarrow Y_2 \leftrightarrow Y_1$.
2) There exist deterministic maps $f_j : \mathcal{W}_j \times \mathcal{Y}_j \to \hat{\mathcal{X}}$ such that

$$\mathbf{E}d(X, f_j(W_j, Y_j)) \le D_j, j = 1, 2.$$

3) The non-negative rate pair satisfies:

$$\begin{aligned} R_1 &\ge I(X; W_1|Y_1) \\ R_1 + R_2 &\ge I(X; W_1|Y_1) + I(X; W_2|Y_2, W_1). \end{aligned}$$

Firstly notice that the outer bound $\mathcal{R}_\cap(D_1, D_2)$ is also a valid outer bound for $\mathcal{R}^*(D_1, D_2)$ for the SR-WZ problem (after switching to the Markov string $X \leftrightarrow Y_2 \leftrightarrow Y_1$).

Comparing $\mathcal{R}^*(D_1, D_2)$ of SR-WZ coding with the outer bound $\mathcal{R}_{out}(D_1, D_2)$ of SI-scalable coding, we see the apparent similarity. In both cases, the bounds for the sum-rate suggest the coding strategy along the order of weaker side information to stronger side information. In other words, use the weak side information Y_w (the subscript w for "weak") and encode a description W_w with a rate of $I(X; W_w|Y_w)$, and then use the strong side information Y_s and encode a description W_s with a rate of $I(X; W_s|Y_s, W_w)$. This strategy was shown to be optimal for the sum-rate in [6], and naturally is expected to be optimal in both the problems in question if only sum-rate is considered. Since this coding order is exactly what is required in the SR-WZ problem, it is expected that this bound is achievable. However, for the SI-scalable coding problem, the coding order is in fact reversed, i.e., instead of along the weak to strong direction, it is along the strong to weak direction, thus this bound might not be achievable in general.

Another noteworthy point is that for the SR-WZ problem, the function $f_2(W_2, Y_2)$ in the definition of $\mathcal{R}^*$ can be replaced by a function $f'_2(W_1, W_2, Y_2)$ without essential difference. However, such a freedom is not available for the SI-scalable coding for $\mathcal{R}_{ach}(D_1, D_2)$ or $\mathcal{R}_{out}(D_1, D_2)$.

In [14], we provide a complete characterization of this problem for arbitrary number of stages, a question left open by Steinberg and Merhav. The main result can be summarized as follows.

***Theorem** 3:* For any discrete memoryless stochastically degraded source $X \leftrightarrow Y_N \leftrightarrow Y_{N-1} \leftrightarrow ... \leftrightarrow Y_1$

$$\mathcal{R}(\boldsymbol{D}) = \mathcal{R}^*(\boldsymbol{D}). \tag{12}$$

In the above, we have defined the region $\mathcal{R}(D)$ to be the natural generalization of the region given in Definition 2 for the two-stage case. Also, we have defined the region $\mathcal{R}^*(\boldsymbol{D})$ to be the set of all rate vectors $\boldsymbol{R} = (R_1, R_2, ..., R_N)$ for which there exist N random variables $(W_1, W_2, ..., W_N)$ in finite alphabets $\mathcal{W}_1, \mathcal{W}_2, ..., \mathcal{W}_N$ such that the following conditions are satisfied.

1) $(W_1, W_2, ..., W_N) \leftrightarrow X \leftrightarrow Y_N \leftrightarrow Y_{N-1} \leftrightarrow ... \leftrightarrow Y_1$.
2) There exist deterministic maps $f_j : \mathcal{W}_j \times \mathcal{Y}_j \to \hat{\mathcal{X}}$ for $1 \le j \le N$ such that

$$\mathbb{E}d(X, f_j(W_j, Y_j)) \le D_j, . \tag{13}$$

3) The non-negative rate vector satisfies for $1 \le j \le N$:

$$\sum_{i=1}^{j} R_i \ge \sum_{i=1}^{j} I(X; W_i|W_1, W_2, ..., W_{i-1}, Y_i). \tag{14}$$

V. The Gaussian Source and the Generalized Successive Refinability

The notion of successive refinability with side informations at the decoder was defined in [13], as requiring that progressive coding should incur no rate loss compared to a single stage Wyner-Ziv coding. In this section we discuss the important special case of the quadratic Gaussian source with Gaussian side informations, which is not successively refinable with different side informations in general. However, for both the cases, the outer bound $\mathcal{R}_\cap(D_1, D_2)$ is achievable, which motivates the notion of generalized successive refinability. To distinguish these two kind of successive refinabilities, we will refer the one defined in [13] as *strict* successive refinability.

A. The Gaussian Source

Let $X \sim N(0, \sigma_x^2)$ be a Gaussian random variable. The side information for the SI-scalable coding are given by

$$Y_1 = X + N_1, \quad Y_2 = X + N_1 + N_2 \tag{15}$$

where N_1, N_2 are mutually independent zero-mean Gaussian random variables (also independent of X) with variance σ_1^2 and σ_2^2, respectively. The side informations for SR-WZ coding is given by

$$Y_1 = X + N_1 + N_2, \quad Y_2 = X + N_2. \tag{16}$$

[2]This is in a different form than the rate distortion region proved in [13], however their achievability was shown in that work to be equivalent.

It can be shown that for both SI-scalable and SR-WZ coding, the outer bound $\mathcal{R}_\cap(D_1, D_2)$ is achievable (see [15] for details). Thus, encoding progressively entails no rate loss for the Gaussian case, in the sense that $(R_1, R_2) = (R_{X|Y_1}(D_1), R_{HB}(D_1, D_2) - R_{X|Y_1}(D_1))$ is achievable for both problems.

Consider the special case that $D_1 = D_2 = D^*$. The SR-WZ problem degenerates in this case, because $R_{X|Y_1}(D^*) = R_{HB}(D^*, D^*)$. On the other hand for SI-scalable coding, we have $R_{HB}(D^*, D^*) = R_{X|Y_2}(D^*) \geq R_{X|Y_1}(D^*)$. Thus $(R_{X|Y_1}(D_1), R_{X|Y_1}(D^*) - R_{X|Y_1}(D^*))$ is an achievable rate pair. This implies that under such distortion requirement, the opportunistic coding of assuming the best side information does not entail rate loss to the Wyner-Ziv bound, i.e., it is strictly successively refinenable.

The degraded side information cases, either $X \leftrightarrow Y_1 \leftrightarrow Y_2$ or $X \leftrightarrow Y_2 \leftrightarrow Y_1$, are specially interesting for the quadratic Gaussian case, because for this particular source the rate-distortion region for arbitrary side informations is always equivalent to that when the side information is degraded. To see this, observe that the rate-distortion region is only dependent on the pair-wise marginals of P_{XY_1} and P_{XY_2}. Thus, we can assume without loss of generality that $Y_1 = \alpha_1(X + N_a)$ and $Y_1 = \alpha_2(X + N_b)$, where N_a and N_b are independent Gaussian random variables with variances σ_a^2 and σ_b^2, also independent of X, and α_1 and α_2 are fixed constants. Since scaling does not effect the rate region, α_1 and α_2 can be removed. Without loss of generality, assume $\sigma_a^2 \leq \sigma_b^2$. Let $Y_1' = X + N_a$ and $Y_2' = X + N_a + N_b'$, where N_b' is a Gaussian random variable with variance $\sigma_b^2 - \sigma_a^2$ and is independent of everything else. The pairwise marginals (X, Y_1') and (X, Y_2') remain the same (up to a scaling factor), thus the rate region remains the same. Therefore for the quadratic Gaussian case, the side information can always be taken as degraded.

B. Generalized Successive Refinability

In the above Gaussian quadratic example the total sum rate for the two stages is generally larger than the Wyner-Ziv rate for either of the stages and therefore it is not *strictly succesively refinable* in the sense defined in [13]. Since the sum-rate matches the Heegard-Berger rate, we define the notion of *generalized successive refinability* in the context of lossy source coding with multiple side-informations.

***Definition** 3:* A source X is said to be *generalized successively refinable* with degraded side informations if

$$(R_{X|Y_1}(D_1), R_{HB}(D_1, D_2) - R_{X|Y_1}(D_1)) \in \mathcal{R}(D_1, D_2).$$

The definition is limited to the degraded side information case, since $R_{HB}(D_1, D_2)$ is known for only this case. This notion only considers whether in order to achieve distortion (D_1, D_2) for side information (Y_1, Y_2), a progressive encoder is as good as any arbitrary encoder, but ignores the factor whether $R_{HB}(D_1, D_2) = R_{X|Y_2}$ is achievable.

Therefore, in light this definition, though the Gaussian source with multiple side-informations is not strictly successively refinable, it is nevertheless successively refinable in the generalized sense. This delineation also offers a necessary condition, for strictly successive refinability which might be simpler to test than the condition given in [13] for the SR-WZ case.

***Theorem** 4:* A necessary condition for a discrete memoryless stochastically degraded source $X \leftrightarrow Y_2 \leftrightarrow Y_1$ to be strictly successively refinable with distortion pairs (D_1, D_2), is that $R_{HB}(D_1, D_2) = R_{X|Y_2}(D_2)$.

VI. Conclusion

The problem of side-information scalable source coding and the problem of successive refinement in the Wyner-Ziv setting were considered in this paper. Part of the motivation comes from the fact that in the lossless case a progressive encoding does not cause any rate loss. In order to make explicit the quality of the side informations in the lossy case, Markovian condition is imposed. New achievability result on the SI-scalable coding problem was provided and then related to existing result for the SR-WZ coding. Complete characterization for the quadratic Gaussian case was explicitly determined.

References

[1] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Academic Press, New York, 1981.

[2] S. C. Draper, "Universal incremental Slepian-Wolf coding," in *Proc. 43rd Annual Allerton Conference on communication, control and computing*, Sep. 2002.

[3] A. Eckford and W. Yu, "Rateless Slepian-Wolf codes," in *Proc. Asilomar conference on signals, systems and computers*, 2005.

[4] 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.

[5] M. Feder and N. Shulman, "Source broadcasting with unknown amount of receiver side information," in *Proc. IEEE Information Theory Workshop*, Oct. 2002, pp. 127–130.

[6] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Information Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.

[7] 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.

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

[9] S. Pradhan and K. Ramchandran, "Multiuser successive refinement", Conference on Information Sciences and Systems (CISS), March 2002.

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

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

[12] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information source," *IEEE Trans. Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[13] 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.

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

[15] C. Tian and S. Diggavi, "Side information scalable source coding," in *EPFL Technical Report*, Jan. 2006.

[16] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 21, pp. 294–300, May 1975.

[17] 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.