Source Coding in the Presence of Exploration-Exploitation Tradeoff

Emrah Akyol, Urbashi Mitra
Dep. of Electrical Eng.
USC, CA, US
Email: {eakyol, ubli}@usc.edu

Ertem Tuncel
Dep. of Electrical Eng.
UC Riverside, CA, US
Email: ertem@ee.ucr.edu

Kenneth Rose
Dep. of Electrical-Computer Eng.
UCSB, CA, US
Email: rose@ece.ucsb.edu

***Abstract*—Exploration versus exploitation in a sensor field with a mobile agent is examined in the context of source coding. The encoder is the low complexity data gathering agent. The decoder is a high complexity fusion center. The encoder first sends a coarse description of the random field, then transmits a refined description of a region of interest, *i.e.*, a subset of the correlated sources in the first stage and so on. The main source coding challenge is that the receiver wants to refine a subset of the correlated sources that is unknown to the encoder *a priori*. The conventional approach of scalable coding via conditional codebook encoding (CCE) requires a codebook that is exponential in size with respect to the number of sources and also the number of refinement stages. This paper studies an alternative approach, using random binning (RB), in lieu of CCE. The universality of RB plays a key role, as the encoder does not know *a priori* which sources the decoder wants to refine. It is shown that RB does not introduce any loss and can effectively replace CCE while providing significant storage reduction in terms of the number of codewords stored. Achievable rate regions are derived for the single and the multi-terminal encoding settings.**

I. Introduction

In this paper, we study the fundamental limits of scalable coding (SC) when *exploration-exploitation* (E-E) tradeoff is in effect. E-E tradeoff is a common challenge in data gathering over sensor networks. The problem can be described as follows. Suppose there is a central receiver with high computing power, that is tasked with identifying a target in a random field, or reconstructing a part of a random field. An autonomous, low power agent gathers data sampled over this random field with two general objectives: first to gather data from a space as large as possible (exploration); and to approach the target by traveling in the direction of the gradient of the field (exploitation). After gathering data from already deployed sensors, the encoder coarsely compresses (quantizes at high distortion) the data corresponding to the entire field and transmits them to the central receiver. The receiver then analyses the coarse data and determines the samples that it is interested in to reconstruct at a higher resolution and transmits this information; *i.e.*, the subset of samples and associated rates, over a feedback channel to the agent. The agent compresses again and sends a refinement to the central decoder, which then again analyses the subset and so on. This information exchange occurs K times. How should the encoder compress the data at each stage? The difficulty of the problem lies in the fact that the encoder does not know *a priori* which samples will be requested by the central receiver. Hence, a conventional, "conditional codebook encoding" (CCE) approach of successive coding requires a codebook for each source for each possible combination of codewords from all sources to condition upon. From a storage complexity point of view, this approach is infeasible.

To state the problem more formally, consider a set of correlated memoryless sources $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \dots \boldsymbol{X}_L]$ where $\boldsymbol{X}_j$'s are memoryless sources generating independent, identically distributed random variables according to some known joint probability distribution $p_{X_1,\dots,X_L}(\cdot)$. Assume there are K stages in the communication scenario. At stage k, we want to describe a subset $\mathcal{S}_k \subseteq [1, \dots L]$ of sources, denoted as $\boldsymbol{X}_{\mathcal{S}_k}$ by the additional R_k bits per sample sent at this stage. Given a single letter vector distortion [1] measure $d(\cdot,\cdot)$, we would like reconstruct $\boldsymbol{X}_{\mathcal{S}_k}, k \in [1 : K]$ with distortions $D_1, \dots, D_K$. What is the necessary and sufficient set of $(R_1 \dots R_K)$ to achieve distortions $(D_1, \dots D_K)$ over $(\mathcal{S}_1, \dots \mathcal{S}_K)$?

We note that this problem is equivalent to compressing $\boldsymbol{X}$ with distortion measures which are functions of the subsets of interest at each stage[2], d_k, $k \in [1 : K]$. From this perspective, it is tempting to conclude that conventional SC addresses our problem adequately. However, the challenge - from this perspective- is that distortion measures $d_1 \dots d_K$ are not known prior to compression as assumed in conventional SC. Specifically, while compressing at stage i, the subset that will be of interest in the later stages, and hence distortion measures $d_k(\cdot,\cdot)$ is not known for $k > i$.

In this paper, we first show that a classical result of scalable coding, [2], [3], obtained by the CCE method, can also be obtained by random binning (RB). Building on this basic result, we develop our main results on source coding in E-E tradeoff. In the second part of the paper, we extend our results to distributed coding, by using a single binning scheme that takes successive refinement and distributed coding into account. To the best of our knowledge, this paper is the first

[1]Note that there is a vector distortion measure known to prior to compression of all stages, that operates on a subset which is unknown *a priori*. Hence, given the subset, we assume the distortion measure is known.

[2]The source dependent distortion measure, *i.e.*, $d(x, y)$ which depends not only on $x - y$ but also x, in our case, $d_i(x, y)$ depends on the samples that are desired to reconstructed at higher resolution at stage i, see eg. [1] for details.

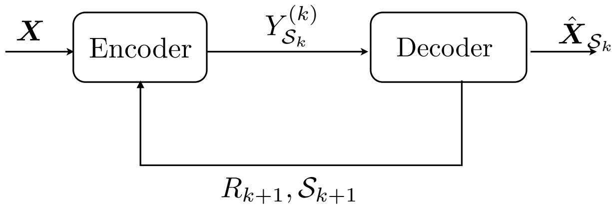

Fig. 1. Problem setting. At stage k encoder sends a refinement of sources in the set $\mathcal{S}_k$ and decoder reconstructs $\boldsymbol{X}_{\mathcal{S}_k}$. The decoder sends the next stage information over a noiseless feedback channel to the encoder.

treatment of "distributed scalable coding" from an information theoretic perspective.

Distributed scalable coding problem is related to another problem, namely scalable coding in the Wyner-Ziv setting, due to the fact that the base layer reconstruction serves as side information for both encoders (note that this reconstruction is not available in either of the encoders) for the refinement layer. Scalable coding in the Wyner-Ziv setting has been studied in [4], [5]. Our approach of using solely RB also offers new insights and results for this problem, see [6] for details.

II. Preliminaries

We use the standard notations, definitions of the additive, single letter distortion measure and the achievable rate regions [7]. We are interested in a set of memoryless correlated sources $\{X_1(t)\}_{t=1}^{\infty}, \ldots, X_L(t)\}_{t=1}$. A vector distortion measure $d(\cdot,\cdot)$ operates on a subset $\mathcal{S}_i$ at stage i. Distortion measure is assumed to be separable in terms of the sources, i.e., $d(X_{\mathcal{S}}, Y_{\mathcal{S}}) = f(\{d_i(X_i, Y_i)\}), \forall i \in \mathcal{S}$. This is a mild assumption that basically states total distortion over a set depends only on the individual distortions of the source and its reconstruction. This paper pertains to two problems in source coding, namely scalable coding and distributed coding. Let us briefly explain the achievable regions for these settings.

A. Scalable Coding

The set of scalably achievable rate distortion quadruples (R_1, R_2, D_1, D_2) is denoted here as $\mathcal{R}_{SC}$. In [2], the achievable region $\mathcal{R}_{SC}$ was characterized:

Theorem 1 ([2], [3]). *$\mathcal{RD}_{SC}$ is the convex hull of quadruples (R_1, R_2, D_1, D_2) for*

$$R_1 \geq I(X; Y_1)$$
$$R_1 + R_2 \geq I(X; Y_1, Y_2)$$

for a conditional distribution $p(Y_1, Y_2|X)$ which satisfy

$$D_1 \geq \mathbb{E}\{d(X, Y_1)\}$$
$$D_2 \geq \mathbb{E}\{d(X, Y_2)\}$$

B. Distributed Coding

In distributed coding setting, two encoders observe two discrete memoryless sources, X_1 and X_2, and describe these sources to central receiver with a distortions D over X_1, X_2.

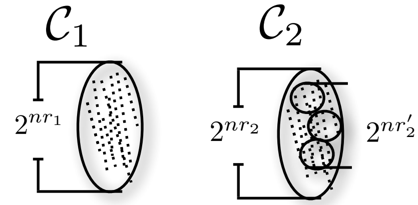

Fig. 2. RB scheme for 2-layer scalable coding

A triple $\mathcal{RD}_{DC} \triangleq (R_1, R_2, D)$ is called *achievable* in distributed coding if with rates R_1, R_2, distortion D is achivable. Exact characterization of $\mathcal{RD}_{DC}$ is unknown, while a well known achievable region due to Berger and Tung, hence denoted here as $\mathcal{RD}_{BT}$, is a result of an intuitive coding scheme called "quantize and bin". Encoders first quantize their observations X_1^n and X_2^n to auxiliary random variables Y_1^n and Y_2^n and utilize binning to losslessly encode Y_1 and Y_2 by Slepian-Wolf encoding. The encoders have to ensure that the received codewords Y_1^n and Y_2^n are jointly typical. One way to guarantee this is to limit the set of Y_1, Y_2 to satisfy the Markov chain $Y_1 - X_1 - X_2 - Y_2$, which renders Y_1^n and Y_2^n jointly typical due to the Markov lemma, see eg. [8].

Theorem 2 ([8], [9]). *$\mathcal{RD}_{BT}$ is the convex hull of quadruples (R_1, R_2, D) for*

$$R_1 \geq I(X_1; Y_1|Y_2)$$
$$R_2 \geq I(X_2; Y_2|Y_1)$$
$$R_1 + R_2 \geq I(X_1, X_2; Y_1, Y_2)$$

for a conditional distribution $p(Y_1, Y_2|X_1, X_2)$ which satisfy

$$D \geq \mathbb{E}\{d([X_1, X_2], g(Y_1, Y_2)\}$$

and the Markov chain $Y_1 - X_1 - X_2 - Y_2$.

III. Main Results

A. Scalable Coding

In this section, we show that using RB in lieu of CCE does not introduce any loss in the conventional successive coding problem. We use $\mathcal{RD}_{RB}$ to denote the achievable region obtained by this new coding scheme. The following theorem formally states our result associated with this setting.

Theorem 3. $\mathcal{RD}_{RB} = \mathcal{RD}_{SC}$

Proof. The encoding scheme is based on RB among two independently generated codebooks are rates r_1 and r_2, denoted as $\mathcal{C}_1$ and $\mathcal{C}_2$, whose codewords are randomly generated according to the marginal probabilities p_{Y_1} and p_{Y_2} respectively. We can find a (y_1^n, y_2^n) tuple jointly typical with any typical x^n if the following are satisfied:

$$\begin{aligned} r_1 &\geq I(X; Y_1) \\ r_2 &\geq I(X; Y_2) \\ r_1 + r_2 &\geq I(X; Y_1) + I(X, Y_1; Y_2) \end{aligned} \tag{1}$$

Let us set $R_1 = r_1 = I(X;Y_1)$ to mimic the rate distortion optimal construction, *i.e.,* there is no binning in $\mathcal{C}_1$. Then, we have to set $r_2 = I(Y_2;X,Y_1)$ to satisfy (1). Since the codebooks are generated independently, the probability of two codewords randomly chosen from Y_1 and Y_2 to be jointly typical under the joint distribution p_{Y_1,Y_2} is approximately $2^{-nI(Y_1;Y_2)}$. To guarantee that the decoder will find only one jointly typical pair given the bin index, the number of codewords in each bin must be approximately $2^{nI(Y_1;Y_2)}$. Then, the number of bins in $\mathcal{C}_2$ (denoted as γ) is

$$\gamma = \frac{2^{nI(Y_2;X,Y_1)}}{2^{nI(Y_1;Y_2)}} = 2^{nI(X;Y_2|Y_1)}.$$

At the refinement layer, only bin index is sent we need a rate of $R_2 = \frac{1}{n}\log(\gamma) = I(X;Y_2|Y_1)$ which is equal to the rate of conventional SC. □

Let us see if this binning scheme provides any storage benefits in itself. Let M_{RB} and M_{CCE} denote the total number of codewords used in RB and CCE methods. Clearly,

$$\begin{aligned} M_{CCE} &= 2^{nI(X;Y_1)} + 2^{nI(X;Y_1)}2^{nI(X;Y_2|Y_1)} \\ &= 2^{nI(X;Y_1)} + 2^{nI(X;Y_1,Y_2)} \end{aligned}$$

while

$$\begin{aligned} M_{RB} &= 2^{nr_1} + 2^{nr_2} \\ &= 2^{nI(X;Y_1)} + 2^{nI(X,Y_1;Y_2)} \end{aligned}$$

Since we can find cases where $I(X,Y_1;Y_2) \gtrless I(X;Y_1,Y_2)$, it is not clear if RB offers any significant storage benefit for this simple setting. As we will show in the next sections, this benefit is significant in the presence of E-E tradeoff.

Remark 1. *This approach can be generalized to L-layer scalable coding, without any performance loss.*

Remark 2. *Implications of this result is not limited to the problem in this paper. Particularly, one of the practical implications is that we can use - for successive refinement purposes- practical tools that realize RB such as lattice codes [10], or Wyner-Ziv codes derived from channel codes [11], [12]. Another implication is on the theoretical study of successive coding in network settings, such as the Wyner-Ziv setting, see eg. [4], [5].*

Remark 3. *Theorem 3 play a key role in developing other results. Particularly, the fact that this result holds for any distortion measure d_2 paves the way to obtain intuitive, high level proofs for the multi-source setting. Note that in multi-source setting analyzed in the next section, our problem can be interpreted as encoding the source scalably with different distortion measures (that depend on the subset) for each layer.*

B. Single Encoder- Multiple Source Setting

Let us consider the setting described earlier, namely E-E source coding. The main challenge is that at stage i, encoder is not aware of the region of interests for later stages, *i.e.*, $\mathcal{S}_k$ for $k > i$. Hence, a conventional CCE approach generates a codebook for each combination of possible $\mathcal{S}_i$'s prior to compression. The achievable region is denoted here as $\mathcal{RD}_{SC}$. The number of codebooks required by this approach is exponential in the number of encoding stages. The proposed random coding scheme however, generates codebooks without any conditioning.

Let us explain how the proposed RB scheme, depicted in Figure 3, works. First, we create K codebooks for each source with $2^{nr_i^{(k)}}$ codewords with the marginal probabilities $p_{Y_i^{(k)}}$ independently, for $i \in \mathcal{I}_L, k \in \mathcal{I}_K$. Let us call these codebooks $\mathcal{C}_i^{(k)}$. Next, we create $2^{n\rho_i^{(k)}}$ non-overlapping bins in $\mathcal{C}_i^{(k)}$ respectively, each containing $2^{n(r_i^{(k)}-\rho_i^{(k)})}$ codewords $\forall i,k$.

At stage k, encoder picks a codeword tuple in the codebooks $Y_{\mathcal{S}_k}^{(k)}$ that is jointly typical with the source and with the codewords that are selected in the previous stages, for each source. Next, the bin indices associated with these codewords are sent to the decoder. The decoder searches for the unique codeword tuple, which is jointly typical in itself and also with the already decoded codewords from the previous stages in each bins.

An interesting property of the proposed scheme is that the bin size is chosen on the fly, depending on the prior subsets[3]. This is done by first enumerating all the codewords in the same order at the encoder and the decoder. At stage k both encoder and decoder can compute the bin size (or the number of bins $\rho_i^{(k)}$) that is needed for successful joint typicality decoding. Encoder then puts the codewords to bins deterministically.[4] Since the decoder has identical list of codewords, given the bin size and bin index, it can identify the bin in which the codeword lies.

Let us call the achievable region obtained by this scheme $\mathcal{RD}_{RB}$.

Theorem 4. *For a given $\mathcal{S}^{(1)}, \ldots \mathcal{S}^{(K)}$, $\mathcal{RD}_{RB}$ is the convex hull of $(R_1, R_2, \ldots R_K, D_1, D_2, \ldots D_K)$*

$$\sum_{i=1}^{k} R_i \geq I(X_{\mathcal{S}^{(1)}} \ldots X_{\mathcal{S}^{(i)}}; Y_{\mathcal{S}^{(1)}}^{(1)} \ldots Y_{\mathcal{S}^{(k)}}^{(k)}) \quad k = 1, 2, \ldots K$$

for a conditional distribution $p(Y_{\mathcal{S}_1}^{(1)} \ldots Y_{\mathcal{S}_K}^{(K)}|X)$ and deterministic decoding functions which satisfy

$$D_k \geq \mathbb{E}\{d(X_{\mathcal{S}_k}, g_k(Y_{\mathcal{S}_1}^1 \ldots Y_{\mathcal{S}_k}^k)\} \quad k = 1, 2, \ldots K$$

Proof. We sketch an intuitive high level proof here, technical details can be found in [14]. Let us recall two results. First one is due to Theorem 1 in [15], encoding correlated sources $\boldsymbol{X}_1 \ldots \boldsymbol{X}_L$ via independent codebook generation and binning

[3]This property yields to use the same codeword for several prior codeword combinations, in other words, we *recycle* the codewords, instead of generating a new one for each possible subset combination as done in CCE scheme.

[4]Binning is deterministic, hence it is slightly different than the original binning scheme used in the seminal paper [13] where bins are also generated randomly. Note however that the codewords are generated randomly, therefore deterministic binning performs essentially the same task as the classical random binning, see [7] for details.

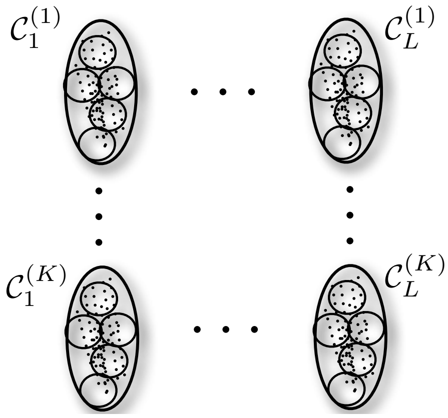

Fig. 3. Overview of the proposed coding scheme for multiple source- single encoder setting. For each stage, we generate L codebooks, one for each source. There are L sources in total, K stages.

does not introduce any loss compared to joint encoding in the classical rate distortion sense. The second result is due to Theorem 3, RB does not introduce any loss in successive coding. Using these two results in conjunction with the observation made earlier- that is the problem at hand is a special case of the SC problem with different distortion measures at each stage, we obtain Theorem 5. □

Let us compare the storage requirements of both schemes for the special case of $L = 2, K = 1$, and for simplicity of computing the size of codebooks, we assume there is no scalable coding.[5] Let again M_{RB} and M_{CCE} denote the total number of codewords used in RB and CCE methods. CCE method generates four codebooks: two of sizes $2^{nI(X_1;Y_1^{(1)})}$, $2^{nI(X_2;Y_2^{(2)})}$ to encode X_1 and X_2 individually and two with size $2^{nI(X_1,X_2;Y_1,Y_2^{(2)})}$ to encode them jointly. Note that we have to generate two of the latter codebooks since we do not know in which sources X_1 and X_2 will be requested.

$$M_{CCE} = 2^{nI(X_1;Y_1^{(1)})} + 2^{nI(X_2;Y_2^{(1)})} + 2\times 2^{nI(X_1X_2;Y_1^{(1)}Y_2^{(1)})}$$

For convenience in comparisons, let us assume $I(X_1X_2;Y_1^{(1)}Y_2^{(1)}) \approx I(X_1;Y_1^{(1)}) + I(X_2;Y_2^{(1)})$, *i.e.*, X_1 and X_2 are nearly independent, therefore $I(Y_1^{(1)};Y_2^{(1)}) = \epsilon \to 0$, hence, $M_{CCE} = 3\times 2^{nI(X_1;Y_1^{(1)})} + 3\times 2^{nI(X_2;Y_2^{(1)})}$

Let us now compute the size of the two codebooks needed in the proposed RB approach. The sum rate inequality in the covering conditions provide an answer as we only need a find a tuple which is jointly typical with x_1^n, x_2^n, then $r_1 \approx r_2 = I(X_1;Y_1^{(1)}) + \epsilon/2$ will satisfy the covering equations. Plugging these values

$$\begin{aligned} M_{RB} &= 2^{nr_1} + 2^{nr_2} \\ &= 2^{n(I(X_1;Y_1^{(1)})+\epsilon/2)} + 2^{n(I(X_2;Y_2^{(1)})+\epsilon/2)} \end{aligned}$$

[5]Note that there is no explicit scalable coding in the example, however the sources X_1 and X_2 can be requested by the decoder successively, hence there is some form on successive coding between X_1 and X_2.

Hence, even without scalable coding and with only two sources, for this simple example, RB requires around $1/3$ storage compared to the conventional CCE approach. More formally, it can be shown that the storage requirements of CCE increases exponentially with the number of sources and the number refinement stages, while RB approach increases only linearly with these quantities.

IV. Extension to Distributed Settings

In this section, we study the setting where there are two encoders. Depending on the set of sources encoders have access to, we address the problem in two different settings.

A. Non-overlapping Sources

In this setting, the sources that different encoders have do not overlap, *i.e.*, say the encoder 1 has $X_{\mathcal{A}}$ and the encoder 2 has $X_{\mathcal{B}}$ and $\mathcal{A}\cap\mathcal{B} = \emptyset$.

Theorem 5. *For a given $\mathcal{A}_1, \dots \mathcal{A}_K$ and $\mathcal{B}_1, \dots \mathcal{B}_K$ regions of interest for each stage, $\mathcal{RD}_{RB}$ is the convex hull of $(R_1^{(1)}, R_2, \dots R_K, D_1, D_2, \dots D_K)$, for $k = 1, 2, \dots K$:*

$$\begin{aligned} \sum_{i=1}^{k} R_1^{(i)} &\geq I(\mathcal{X}_{\mathcal{A}}^{(k)};\mathcal{Y}_{\mathcal{A}}^{(k)}|\mathcal{Y}_{\mathcal{B}}^{(k)}) \\ \sum_{i=1}^{k} R_2^{(i)} &\geq I(\mathcal{X}_{\mathcal{B}}^{(k)};\mathcal{Y}_{\mathcal{B}}^{(k)}|\mathcal{Y}_{\mathcal{A}}^{(k)}) \\ \sum_{i=1}^{k} R_1^{(i)} + R_2^{(i)} &\geq I(\mathcal{X}_{\mathcal{A}}^{(k)},\mathcal{X}_{\mathcal{B}}^{(k)};\mathcal{Y}_{\mathcal{A}}^{(k)},\mathcal{Y}_{\mathcal{B}}^{(k)}) \end{aligned}$$

for a conditional distribution $p(Y_{\mathcal{A}_1}^1 \dots Y_{\mathcal{A}_K}^K|X)$ and deterministic decoding functions which satisfy

$$D_k \geq \mathbb{E}\{d_k([X_{\mathcal{A}_k}, X_{\mathcal{B}_k}], g_k(\mathcal{Y}_{\mathcal{A}}^{(k)},\mathcal{Y}_{\mathcal{B}}^{(k)})\}$$

and the set of Markov chains

$$\mathcal{Y}_{\mathcal{A}}^{(k)} - \mathcal{X}_{\mathcal{A}}^{(k)} - \mathcal{X}_{\mathcal{B}}^{(k)} - \mathcal{Y}_{\mathcal{B}}^{(k)}$$

where $\mathcal{X}_j^{(k)} = X_{j^{(1)}} \dots X_{j^{(k)}}$, $\mathcal{Y}_j^{(k)} = Y_{j^{(1)}}^{(1)} \dots Y_{j^{(k)}}^{(k)}$ for $j \in \{\mathcal{A},\mathcal{B}\}$.

Proof. We provide a sketch of the proof here, the techincal details of the proof can be found in [14]. Essentially, compared to the previous case, encoding stays the same, encoders first generate K codebooks for each source, then bin with the bin size determined the previous subsets. The Markov chains guarantee typicality of the received codewords the at the receiver, due to the Markov lemma, as they do in Berger-Tung region. One difference from single terminal setting is that, at stage k, the reconstructions of the previous layers, $\mathcal{Y}_{\mathcal{B}}^{(k-1)}$ are not available at the first encoder (and vice versa), hence encoders perform scalable coding with decoder side information. This setting corresponds to degraded side information setting of Steinberg and Merhav [4] and we show in a companion paper that RB achieves the optimal performance for this setting. Hence, the region is readily obtained by optimality of RB for scalable coding with side information (see [6]), in conjunction with the application of "quantize and bin" for distributed coding, see [14] for the details. □

Remark 4. *Note that $\mathcal{RD}_{RB}$ is complete for the quadratic Gaussian setting. The proof, omitted due to space constraints, follows from optimality of Berger-Tung region for the quadratic Gaussian setting.*

B. Overlapping Sources

In this setting, the sources that different encoders have overlap: again, let the encoder 1 have $X_{\mathcal{A}}$ and the encoder 2 have $X_{\mathcal{B}}$ and $\mathcal{A} \cap \mathcal{B} = \mathcal{D} \neq \emptyset$. The difference between the non-overlapping case is the Markov chains need to be satisfied, *i.e.*, the set of codewords guaranteed to be typical. Since sources $X_{\mathcal{D}}$ are available in both encoders, we do not need to invoke the Markov lemma to make these codewords associated with these sources typical. Particularly, only the codewords associated with the sets $\mathcal{A} - \mathcal{D}$ and $\mathcal{B} - \mathcal{D}$ have to be made jointly typical via the Markov lemma, the jointly typical pairs in $\mathcal{D}$ can be selected at both encoders as both the source-words and the codebooks associated with $\mathcal{D}$ are available at both encoders. Hence, for sources in $\mathcal{D}$ the problem simplifies to single terminal scalable coding investigated in the previous section. Similar observations were made in [16] for a simpler setting (where there is no scalable coding or E-E tradeoff) to encode sources which have a common part in the Gács-Körner common information [17] sense.

Theorem 6. *For a given $\mathcal{A}_1, \dots \mathcal{A}_K$ and $\mathcal{B}_1, \dots \mathcal{B}_K$ regions of interest for each stage, $\mathcal{RD}_{RB}$ is the convex hull of $(R_1^{(1)}, R_2, \dots R_K, D_1, D_2, \dots D_K)$, for $k = 1, 2, \dots K$:*

$$\sum_{i=1}^{k} R_1^{(i)} \geq I(\mathcal{X}_{\mathcal{A}-\mathcal{D}}^{(k)}; \mathcal{Y}_{\mathcal{A}-\mathcal{D}}^{(k)} | \mathcal{X}_{\mathcal{B}-\mathcal{D}}^{(k)})$$

$$\sum_{i=1}^{k} R_2^{(i)} \geq I(\mathcal{X}_{\mathcal{B}-\mathcal{D}}^{(k)}; \mathcal{Y}_{\mathcal{B}-\mathcal{D}}^{(k)} | \mathcal{X}_{\mathcal{A}-\mathcal{D}}^{(k)})$$

$$\sum_{i=1}^{k} R_1^{(i)} + R_2^{(i)} \geq I(\mathcal{X}_{\mathcal{A}}^{(k)}, \mathcal{X}_{\mathcal{B}}^{(k)}; \mathcal{Y}_{\mathcal{A}}^{(k)}, \mathcal{Y}_{\mathcal{B}}^{(k)})$$

for a conditional distribution $p(Y_{\mathcal{S}_1}^1 \dots Y_{\mathcal{S}_K}^K | X)$ and deterministic decoding functions which satisfy

$$D_k \geq \mathbb{E}\{d_k([X_{\mathcal{A}_k}, X_{\mathcal{B}_k}], g_k(\mathcal{Y}_{\mathcal{A}}^{(k)}, \mathcal{Y}_{\mathcal{B}}^{(k)})\}$$

and the set of Markov chains

$$\mathcal{Y}_{\mathcal{A}-\mathcal{D}}^{(k)} - \mathcal{X}_{\mathcal{A}-\mathcal{D}}^{(k)} - \mathcal{X}_{\mathcal{B}-\mathcal{D}}^{(k)} - \mathcal{Y}_{\mathcal{B}-\mathcal{D}}^{(k)}$$

where $\mathcal{X}_j^{(k)} = X_{j^{(1)}} \dots X_{j^{(k)}}$, $\mathcal{Y}_j^{(k)} = Y_{j^{(1)}}^{(1)} \dots Y_{j^{(k)}}^{(k)}$ for $j \in \{\mathcal{A}, \mathcal{B}\}$.

V. Discussions

In this paper, we studied scalable coding in the presence of E-E tradeoff. The main idea is to perform scalable coding via RB for both single terminal and distributed cases. It is shown that RB does not introduce any loss compared to the CCE approach and provides significant storage advantage when E-E tradeoff is in effect.

Our future work includes the investigation of numerical results for the quadratic Gaussian setting as well as extensions of these ideas to source-channel coding and network settings.

Acknowledgment

This work is supported in part by these grants: NSF CCF-1118075, NSF CCF-0917343, NSF CCF-1117896, NSF CNS-1213128, ONR N00014-09-1-0700, AFOSR FA9550-12-1-0215, and DOT CA-26-7084-00.

References

[1] T. Berger, *Rate Distortion Theory. A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[2] V. Koshelev, "Hierarchical coding of discrete sources," *Problemy peredachi informatsii*, vol. 16, no. 3, pp. 31–49, 1980.

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

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

[5] C. Tian and S. Diggavi, "Side-information scalable source coding," *IEEE Trans. on Information Theory*, vol. 54, no. 12, pp. 5591–5608, 2008.

[6] E. Akyol, E. Tuncel, K. Rose, and U. Mitra, "On scalable coding in the presence of decoder side information," *Proc. of IEEE Int. Symp. on Information Theory, 2014*.

[7] A. El Gamal and Y. Kim, *Network information theory*. Cambridge University Press, 2011.

[8] T. Berger, "Multiterminal source coding," *The Information Theory approach to communications*, vol. 229, pp. 171–231, 1977.

[9] S. Tung, "Multiterminal rate-distortion theory," Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, 1978.

[10] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[11] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *IEEE Trans. on Communications*, vol. 57, no. 2, pp. 376–387, 2009.

[12] Z. Liu, S. Cheng, A. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Trans. on Information Theory*, vol. 52, no. 10, pp. 4358–4379, 2006.

[13] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources (corresp.)," *Information Theory, IEEE Transactions on*, vol. 21, no. 2, pp. 226–228, 1975.

[14] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "On source coding in the presence of exploration-exploitation trade-off," *working draft available at http://www.arxiv.org*.

[15] S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions-part i:(n, k) source-channel erasure codes," *IEEE Trans. on Information Theory*, vol. 50, no. 1, pp. 47–61, 2004.

[16] A. B. Wagner, B. G. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4035–4057, 2011.

[17] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.