On Scalable Coding in the Presence of Decoder Side Information

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.
UC Santa Barbara, CA, US
Email: rose@ece.ucsb.edu

***Abstract*—The problem of scalable coding while exploiting the decoder side information is considered. Prior work considered the two important cases concerning the degraded side information where source X and the side information variables (Y_1, Y_2) form a Markov chain in the order of either $X - Y_1 - Y_2$ or $X - Y_2 - Y_1$. While the encoding schemes for these settings differ considerably, they are both based on the combination of conditional codebook encoding, a standard tool in scalable coding, and random binning, conventionally used in decoder side information problems. In this paper, an encoding scheme is proposed solely on the basis of random binning, which essentially performs scalable and Wyner-Ziv coding simultaneously. Proposed scheme achieves the rate-distortion regions of prior results. A practical advantage of the unifying scheme is the fact that random binning can be realized via practical tools such as nested lattice codes and channel codes. Finally, motivated by the proposed encoding scheme, a network interpretation of scalable coding is considered. An achievable region is derived for this problem setting and the potential benefits of networked scalable coding are shown.**

I. Introduction

This paper considers the problem of scalable coding when each decoder has different side information, correlated with the source. The problem setting is depicted in Figure 1, where decoder 1 receives the base layer at rate R_1 and uses the side information Y_1 to reconstruct the source X with distortion D_1. Decoder 2 receives both the base layer (at rate R_1) and the refinement layer at rate R_2; and with the help of Y_2, reconstructs the source X with distortion D_2.

There are two settings of interest. The first one, analyzed by Steinberg and Merhav in [1], pertains to a particular ordered degradation of side information, *i.e.,* $X - Y_2 - Y_1$ forms a Markov chain in this order. The achievable region for this setting was characterized and conditions for successive refinement were studied. The coding scheme is based on conditional codebook encoding along with Wyner-Ziv binning. We refer to this setting as successive refinement in the Wyner-Ziv setting (SRWZ).

The second case of interest, where the first decoder has better side information, *i.e.,* $X - Y_1 - Y_2$, was studied by Tian and Diggavi in [2]. In the sequel, this setting will be referred as side information scalable coding (SISC)[1]. In [2], the authors derived an achievable region with an encoding scheme that is in sharp contrast to the scheme of [1], in the sense that the main tool of the encoding scheme is nested binning.

These two problem settings can be considered as special cases of scalable coding with side information, with no specific stochastic ordering on the side information. This problem is known to be difficult even without the scalable coding requirement, and the full characterization of achievable regions (without scalable coding) was found only under stochastic ordering of side information, see the works by Heegard and Berger [3] and independently by Kaspi [4].

In this paper, by leveraging the properties of the random binning-conventionally used in distributed coding problems-we propose a coding scheme for both settings of interest. Our contributions are:

- We propose a unified coding scheme based on random binning among three codebooks. We show that the obtained achievable region includes any rate-distortion point achievable by the prior schemes. Moreover, the Markov chain conditions we obtain are less strict than those of [2], suggesting room for possible improvement in the SISC setting. We note again that SRWZ region in [1] is complete for the associated setting, so the proposed scheme achieves all possible points in the SRWZ setting.
- Beyond the classical scalable coding, we consider a *networked* scalable setting motivated by the encoding scheme derived in this paper and the Gray-Wyner network [5]. We derive an achievable region for this setting and show the benefits of such networked considerations.

We note in passing that the proposed encoding scheme is based on random binning only and hence from a practical point of view, it is realizable by Wyner-Ziv-like codes derived from lattices or channel codes [6]–[9]. Our results are based on the realization that scalable coding can be realized by random binning without any loss. See also [10] for a detailed analysis of other implications of this result.

The paper is organized as follows. In Section II, we present the preliminaries and review the prior results. In Section III, we present our encoding scheme and derive the achievable regions associated with different settings. In Section IV, we study the concept of networked scalability. Finally, we discuss future directions in Section V.

[1]We use somewhat vague names SRWZ and SISC to be consistent with the prior work [1], [2].

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

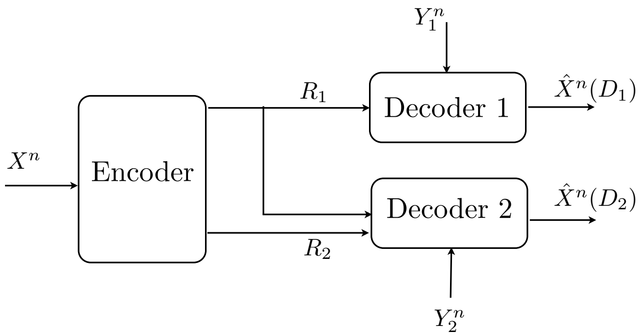

Fig. 1. Problem setup: the classical interpretation of scalable coding with decoder side information. The case where $X - Y_2 - Y_1$ is denoted as SRWZ setting while $X - Y_1 - Y_2$ is called the SISC setting.

II. Preliminaries

A. Notation

Let $\{X_t\}_{t=1}^{\infty}$, $X_t \in \mathcal{X}$, be a discrete memoryless source (DMS) with generic distribution $P_X(x)$. The vector $[X(1), X(2), ..., X(n)]$ is compactly denoted by x^n. Let $\mathcal{Z}$ denote the reproduction alphabet. We employ $H(X)$ to denote the entropy of a discrete random variable X, or differential entropy if X is continuous. For an arbitrary set $\mathcal{A}$, we use $2^{\mathcal{A}}$ to denote the set of all subsets of $\mathcal{A}$, *i.e.,*

$$2^{\mathcal{A}} \triangleq \{\mathcal{S} : \mathcal{S} \subseteq \mathcal{A}\}.$$

Assume a single-letter, bounded, and additive distortion measure $d : \mathcal{X} \times \mathcal{Y} \longrightarrow [0, \infty)$, *i.e.,*

$$d(x^n, y^n) = \frac{1}{n} \sum_{t=1}^{n} d(x_t, y_t) \ . \tag{1}$$

A scalable block code pair (f_1, f_2, g_1, g_2) consists of an encoding function

$$\begin{aligned} f_1 &: \mathcal{X}^n \longrightarrow \mathcal{M}_1 \\ f_2 &: \mathcal{X}^n \longrightarrow \mathcal{M}_2 \end{aligned}$$

and decoders

$$\begin{aligned} g_1 &: \mathcal{M}_1 \times Y_1^n \longrightarrow \mathcal{Z}^n \\ g_2 &: \mathcal{M}_1 \times \mathcal{M}_2 \times Y_2^n \longrightarrow \mathcal{Z}^n \ . \end{aligned}$$

A quadruple (R_1, R_2, D_1, D_2) is called *achievable* if for every $\delta > 0$ and sufficiently large n, there exists a block code (f_1, f_2, g_1, g_2) such that

$$\begin{aligned} \frac{1}{n} \log |\mathcal{M}_1| &\leq R_1 + \delta \\ \frac{1}{n} \log |\mathcal{M}_1||\mathcal{M}_2| &\leq R_1 + R_2 + \delta \\ E\{d(X^n, g_1(f_1(X^n), Y_1^n))\} &\leq D_1 + \delta \\ E\{d(X^n, g_2(f_1(X^n), f_2(X^n), Y_2^n))\} &\leq D_2 + \delta \ . \end{aligned}$$

B. Scalable Coding in the Presence of Side Information

Steinberg and Merhav studied the case of degraded side information in the order of $X - Y_2 - Y_1$, *i.e.,* the SRWZ setting. They showed that the region formalized below is complete. The encoding scheme is intuitive: generate a codebook $\mathcal{C}_1$ with marginal distribution of U_1 and then conditionally generate a codebook $\mathcal{C}_2$ for each codeword u_1^n with the conditional density $P(U_2|U_1^n = u_1^n)$. Next, bin $\mathcal{C}_1$ so that the codeword u_1^n can be decoded with the help of side information Y_1. Next, bin all the conditional codebooks C_2 so that the codewords u_1^n and u_2^n can both be decoded at the decoder with the help of the better side information Y_2. The following theorem presents the achievable region[2] by this scheme.[3]

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

$$\begin{aligned} R_1 &\geq I(X; U_1|Y_1) \\ R_1 + R_2 &\geq I(X; U_2|U_1, Y_2) + I(X; U_1|Y_1) \end{aligned}$$

for a conditional distribution $p(U_1, U_2|X)$ and deterministic decoding functions g_1, g_2 which satisfy

$$D_i \geq \mathbb{E}\{d_i(X, g_i(U_i, Y_i))\} \quad i = 1, 2$$

and the Markov chain $(U_1, U_2) - X - Y_2 - Y_1$.

Tian and Diggavi studied the dual of this problem in the sense that $X - Y_1 - Y_2$ forms a Markov chain. In this setting, they proposed two coding schemes depending on the levels of distortion D_1 and D_2. However, both regions can be unified with the help of an auxiliary random variable, as shown in [2]. This region is presented in the following:

Theorem 2 ([2]). *An achievable region for the SISC setting, $\mathcal{RD}_{SISC}$ is the convex hull of quadruples (R_1, R_2, D_1, D_2) for*

$$\begin{aligned} R_1 &\geq I(X; U_0, U_1|Y_1) \\ R_1 + R_2 &\geq I(X; U_0, U_2|Y_2) + I(X; U_1|Y_1, U_0) \end{aligned}$$

for a conditional distribution $p(U_0, U_1, U_2|X)$ and deterministic decoding functions g_1, g_2 which satisfy

$$D_i \geq \mathbb{E}\{d_i(X, g_i(U_i, Y_i))\} \quad i = 1, 2$$

and the Markov chain $(U_0, U_1, U_2) - X - Y_1 - Y_2$.

III. An Achievable Region for Conventional Scalable Coding with Decoder Side Information

In this section, we propose a unified scheme based on random binning over three codebooks generated independently, as shown in Figure 2. These codebooks, denoted as $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ are generated respectively with the marginal distribution of three auxiliary random variables U_0, U_1 and U_2 at rates r_0, r_1, r_2. The codewords associated with $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ are denoted as u_0^n,

[2]The original achievable region in [1] involves additional terms. However, it was also shown in [1] that the two regions are equivalent.

[3]The cardinality conditions are not essential for our purposes and omitted due to space constraints.

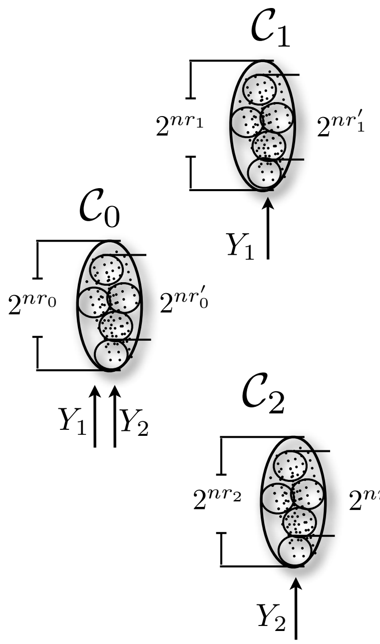

Fig. 2. The overview of the encoding scheme.

u_1^n and u_2^n. After generating the codebooks, we bin the all three codebooks so that there is only one typical tuple of decodable codewords at respective decoders. Let r'_0, r'_1, r'_2 denote the binning rates of the codebooks $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ respectively. The bin indices associated with the codewords u_0^n, u_1^n are transmitted to decoder 1 and the bin index associated with u_2^n is transmitted to decoder 2 as additional (refinement) information. The bit-stream rates are given as:

$$R_1 = r'_0 + r'_1 \tag{2}$$

$$R_2 = r'_2 \tag{3}$$

Although the mechanics of the coding scheme are identical for both settings, the codebook rates and the associated binning rates differ for each setting. The main difference is that for the SRWZ setting, decoder 2 decodes all the received codewords, *i.e.,* the triple u_0^n, u_1^n and u_2^n, while for the SISC case, it decodes only the codewords u_0^n and u_1^n. Hence, in the SRWZ setting, we have to guarantee the joint typicality of the triple (u_0^n, u_1^n, u_2^n), with x^n while in the SISC setting, we only need to guarantee typicality of x^n with the pairs (u_0^n, u_1^n) and (u_0^n, u_2^n). This implies different covering conditions for the two settings. Moreover, in the SRWZ setting, we have to guarantee the joint typicality of (u_0^n, u_1^n) with Y_1^n and also (u_0^n, u_1^n, u_2^n) with Y_2^n for successful joint typicality decoding. The Markov lemma guarantees that if the Markov chains

$$(U_0, U_1) - X - Y_1 \tag{4}$$

$$(U_0, U_1, U_2) - X - Y_2 \tag{5}$$

hold, these typicality conditions are satisfied. Note that (4)-(5) along with $X - Y_2 - Y_1$ imply $(U_0, U_1, U_2) - X - Y_2 - Y_1$. For the SISC case, we need the joint typicality of u_0^n, u_1^n with Y_1^n and also u_0^n, u_2^n with Y_2^n, which imply the Markov chains

$$(U_0, U_1) - X - Y_1 - Y_2 \tag{6}$$

$$(U_0, U_2) - X - Y_2 \tag{7}$$

Note that (6)-(7) are guaranteed to hold if (but not only if) the long Markov chain $(U_0, U_1, U_2) - X - Y_1 - Y_2$ holds. This observation suggests the possibility of obtaining a larger rate region for the SISC setting.

A. SISC Setting

In the SISC setting, decoder 2 cannot decode the three codewords but only two of them, u_0^n and u_2^n, while decoder 1 receives and decodes the tuple u_0^n, u_1^n. For this setting, we do not constrain the distributions to render the triple (u_0^n, u_1^n, u_2^n) to be jointly typical for each typical source-word x^n since this triple is not decoded together in either of the decoders. We have to satisfy the following:

$$\begin{aligned} r_i &\geq I(X; U_i) \quad i \in \{0, 1, 2\} \\ r_0 + r_1 &\geq I(X; U_0, U_1) + I(U_0; U_1) \\ r_0 + r_2 &\geq I(X; U_0, U_2) + I(U_0; U_2)\ . \end{aligned} \tag{8}$$

Note again that there is no sum rate condition since we do not need the joint typicality of the triple (u_0^n, u_1^n, u_2^n). Let us set $r_0 = I(X; U_0)$, $r_1 = I(X, U_0; U_1)$ and $r_2 = I(X, U_0; U_2)$, satisfying (8). Next, we bin the codebooks with rates r'_0, r'_1, and r'_2. The joint decoding conditions yield

$$r_0 - r'_0 + r_1 - r'_1 \leq I(U_0; U_1) + I(U_0, U_1; Y_1) \tag{9}$$

$$r_0 - r'_0 \leq I(Y_1, U_1; U_0) \tag{10}$$

$$r_1 - r'_1 \leq I(Y_1, U_0; U_1) \tag{11}$$

$$r_0 - r'_0 + r_2 - r'_2 \leq I(U_0; U_2) + I(U_0, U_2; Y_2) \tag{12}$$

$$r_0 - r'_0 \leq I(Y_2, U_2; U_0) \tag{13}$$

$$r_2 - r'_2 \leq I(Y_2, U_0; U_2) \tag{14}$$

Now, it can be shown using the Markov chains $(U_0, U_1) - X - Y_1 - Y_2$ and $(U_0, U_2) - X - Y_2$ that one can satisfy (9)-(14) with the rate choices as below:

$$r'_0 = I(X; U_0|Y_1) \tag{15}$$

$$r'_1 = I(X; U_1|U_0, Y_1) \tag{16}$$

$$r'_2 = I(X; U_0, U_2|Y_2) - I(X; U_0|Y_1)\ . \tag{17}$$

Obviously, by (2) and (3), (15)-(17) imply that

$$R_1 = I(X; U_1, U_0|Y_1) \tag{18}$$

$$R_2 = I(X; U_0, U_2|Y_2) - I(X; U_0|Y_1) \tag{19}$$

is an achievable rate pair through our scheme. It then follows through standard rate transfer arguments (see [1], [11], [12] for details) and a comparison with Theorem 2 that we obtain the region $\mathcal{RD}^*_{SISC}$ presented in the following theorem:

Theorem 3. *$\mathcal{RD}^*_{SISC}$ is the convex hull of all rate-distortion quadruples (R_1, R_2, D_1, D_2)*

$$R_1 \geq I(X; U_0, U_1|Y_1)$$

$$R_1 + R_2 \geq I(X; U_0, U_2|Y_2) + I(X; U_1|Y_1, U_0)$$

for a conditional distribution $p(U_0, U_1, U_2|X)$ and deterministic decoding functions g_1, g_2 which satisfy

$$D_i \geq \mathbb{E}\{d_i(X, g_i(U_i, Y_i))\} \quad i = 1, 2$$

and the Markov chains $(U_0, U_1) - X - Y_1 - Y_2$ and $(U_0, U_2) - X - Y_2$.

Corollary 1. $\mathcal{RD}_{SISC} \subseteq \mathcal{RD}^*_{SISC}$

Proof. The rate and distortion expressions of $\mathcal{RD}_{SISC}$ and $\mathcal{RD}^*_{SISC}$ are identical. The Markov chains $(U_0, U_1) - X - Y_1 - Y_2$ and $(U_0, U_2) - X - Y_2$ are implied by $(U_0, U_1, U_2) - X - Y_1 - Y_2$. □

Remark 1. *Note the inclusion may very well be strict, i.e., there can be a point in $\mathcal{RD}^*_{SISC}$ which is not included in $\mathcal{RD}_{SISC}$. This is due to the fact that the long Markov chain $(U_0, U_1, U_2) - X - Y_1 - Y_2$ is a stronger constraint than (6)-(7) and reduces the set of distributions considered in defining the region.*

Remark 2. *It is tempting to question whether the bin index associated with u_1^n, not used in decoder 2 could be used to enlarge the achievable region. Specifically, it might seem that even if we cannot decode the codeword u_1^n, the index of the bin in which u_1^n lies can be used to decrease the number of relevant codewords of $\mathcal{C}_0$. However, the benefit vanishes asymptotically in n, and it can be shown[4] that all codewords in $\mathcal{C}_0$ are jointly typical with at least one codeword in the given bin, irrespective of the rate r_0, hence the bin index is useless.*

B. SRWZ Setting

In the setting where $X - Y_2 - Y_1$, decoder 2 can decode all the received codewords, i.e. the triple u_0^n, u_1^n and u_2^n, while decoder 1 receives and decodes the tuple u_0^n, u_1^n. For this setting, we have to make sure we can find a jointly typical codeword triple u_0^n, u_1^n and u_2^n for each typical source-word x^n.

Clearly, there is no difference between U_0 and U_1 in this coding scheme. Hence, we can set U_0 to constant (*i.e.,* set $r_0 = 0$) without any loss since we can always combine U_0 and U_1 and call this union, the effective U_1. The encoding scheme then simplifies to generating two codebooks associated with U_1 and U_2. We have to satisfy the following due to covering conditions:

$$\begin{aligned} r_i &\geq I(X; U_i) \quad i \in \{1, 2\} \\ r_1 + r_2 &\geq I(X; U_1, U_2) + I(U_1; U_2) \qquad (20) \end{aligned}$$

Let us set $r_1 = I(X; U_1)$ and $r_2 = I(X; U_2|U_1) + I(U_1; U_2)$ to satisfy (20). Next, we bin the codebooks with rates r'_1 and r'_2. The decoding conditions yield

$$\begin{aligned} r_1 - r'_1 &\leq I(U_1; Y_1) \qquad (21) \\ r_1 + r_2 - r'_1 - r'_2 &\leq I(U_1; U_2) + I(U_1, U_2; Y_2) \qquad (22) \\ r_1 - r'_1 &\leq I(Y_2, U_2; U_1) \qquad (23) \\ r_2 - r'_2 &\leq I(Y_2, U_1; U_2)\ . \qquad (24) \end{aligned}$$

[4]The proof is omitted due to the space constraints.

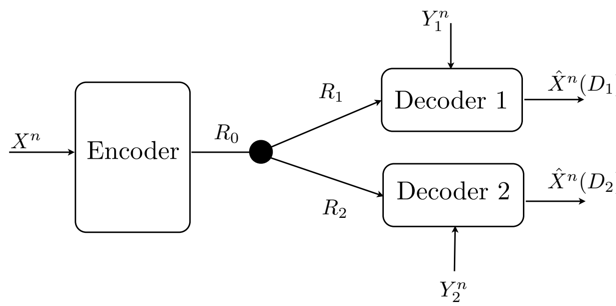

Fig. 3. Problem setup: Network interpretation of scalable coding.

Similar to the SISC case, it can be shown using the Markov chain $(U_1, U_2) - X - Y_2 - Y_1$ that one can satisfy (21)-(24) with the rate choices as below:

$$\begin{aligned} r'_1 &= I(X; U_1|Y_1) \qquad (25) \\ r'_2 &= I(X; U_2|U_1, Y_2) \qquad (26) \end{aligned}$$

Since in this setting, $R_1 = r'_1$ and $R_2 = r'_2$, once again using rate transfer arguments and comparing with Theorem 1 we obtain the region $\mathcal{RD}^*_{SRWZ}$ in the following theorem.

Theorem 4. $\mathcal{RD}^*_{SRWZ} = \mathcal{RD}_{SRWZ}$.

IV. A New Interpretation of Networked Scalable Coding

A major drawback of the conventional scalable coder is that it enforces a rigid hierarchical structure on the bit-stream of different layers. The natural assumption underlying this scalable coding framework is that a user with a better channel will always be able to decode all the base layer bits which were necessary for the base layer reconstruction.

However, it is worthwhile to question this assumption in today's network scenario wherein an intermediate node forwards packets to multiple receivers (see [13] for the potential benefits of such considerations in general networked source coding problems). The SISC problem setting demonstrates the validity of this proposition: the classical scalable compression clearly requires the transmission of redundant information to receivers. Specifically, in our encoding scheme the bin index associated with u_1^n is transmitted to decoder 2 but u_1^n cannot be decoded since the side information Y_2^n is too degraded to identify a unique jointly typical codeword. This fact along with a simple network model given in the seminal paper by Gray and Wyner [5] motivates this problem setup:

Consider the network in Figure 3, where the encoder produces a bit-stream of rate R_0 and an intermediate router transmits the parts of this bit-stream to decoders 1 and 2 at rates R_1 and R_2 respectively. The decoder 1 has better side information but its received rate is smaller than decoder 2, *i.e.,* $X - Y_1 - Y_2$ forms a Markov chain and $R_1 < R_2$. What is the set of achievable rate-distortion quintuples $(R_0, R_1, R_2, D_1, D_2)$?

In the following, we present an achievable region, denoted as $\mathcal{RD}^*_{NSC}$.

Theorem 5. *$\mathcal{RD}^*_{NSC}$ is the convex hull of all rate-distortion quintuples $(R_0, R_1, R_2, D_1, D_2)$*

$$R_0 \geq I(X; U_0, U_2|Y_2) + I(X; U_1|Y_1, U_0)$$
$$R_1 \geq I(X; U_0, U_1|Y_1)$$
$$R_2 \geq I(X; U_0, U_2|Y_2)$$

for a conditional distribution $p(U_0, U_1, U_2|X)$ and deterministic decoding functions g_1, g_2 which satisfy

$$D_i \geq \mathbb{E}\{d_i(X, g_i(U_i, Y_i))\} \quad i = 1, 2$$

and the Markov chains $(U_0, U_1) - X - Y_1$ and $(U_0, U_2) - X - Y_2$.

Proof. The proof is based on the encoding scheme in Theorem 3, with the difference in the rate expressions $R_0 = r'_0 + r'_1 + r'_2$,$R_1 = r'_0 + r'_1$ and $R_0 = r'_0 + r'_2$. Plugging these values in covering and binning conditions given in (8) and (19), the proof is obtained.

□

Remark 3. *It is possible to show that this region is complete for the important quadratic Gaussian case. The proof is based on the optimality of the encoding scheme for SISC setting for the quadratic Gaussian case and completeness of the lossy Gray-Wyner region.*[5]

V. Discussions

The extended version of this paper, which includes the techincal details of the proofs can be found in [12]. Our future work includes the investigation of the converses for the regions in Theorems 3 and 5 and a possible rate-distortion point that shows $\mathcal{RD}_{SISC} \subset \mathcal{RD}^*_{SISC}$ strictly (see Remark 1).

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] 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.

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

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

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

[5] R. Gray and A. Wyner, "Source coding for a simple network," *Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[6] 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.

[7] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," in *Proceedings of the IEEE International Symposium on Information Theory*. IEEE, 2002, p. 91.

[5]Due to space considerations, we omit the proof here.

[8] J. Muramatsu, T. Uyematsu, and T. Wadayama, "Low-density parity-check matrices for coding of correlated sources," *Information Theory, IEEE Trans. on*, vol. 51, no. 10, pp. 3645–3654, 2005.

[9] T. Coleman, A. Lee, M. Médard, and M. Effros, "On some new approaches to practical Slepian-Wolf compression inspired by channel coding," in *Data Compression Conference, 2004. Proceedings. DCC 2004*. IEEE, 2004, pp. 282–291.

[10] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "On source coding in the presence of exploration-exploitation trade-off," *Proc. of IEEE Int. Symp. on Information Theory, 2014*.

[11] E. Tuncel, "The rate transfer argument in two-stage scenarios: When does it matter?" in *IEEE International Symposium on Information Theory, 2009*. IEEE, 2009, pp. 41–45.

[12] E. Akyol, E. Tuncel, K. Rose, and U. Mitra, "On scalable coding in the presence of decoder side information," *working draft available at http://www.arxiv.org*.

[13] K. Viswanatha, E. Akyol, and K. Rose, "On optimum communication cost for joint compression and dispersive information routing," in *Information Theory Workshop (ITW)*. IEEE, 2010, pp. 1–5.