The Wyner-Ziv Problem With Multiple Sources*

Michael Gastpar

June 6, 2002

Communication Systems Department
Ecole Polytechnique Fédérale (EPFL)
Lausanne, Switzerland
Email: Michael.Gastpar@epfl.ch

Abstract

We consider the problem of separately compressing multiple sources in a lossy fashion for a decoder that has access to side information. For the case of a single source, this problem has been completely solved by Wyner and Ziv. For the case of two sources, we establish an achievable rate region, an inner bound to the rate region, and a partial converse. The partial converse applies to the case when the sources are conditionally independent given the side information, and it differs significantly from prior art in that it applies also to the symmetric case where all sources are encoded with respect to fidelity criteria. Moreover, we also show that in this special case, there is no difference between the minimum rate needed to encode the sources jointly, and the minimum sum rate needed for separate encoding.

Index Terms: Wyner-Ziv problem, side information, distributed lossy source coding.

1 Introduction

Suppose two correlated discrete memoryless sources have to be compressed separately from each other in a lossy fashion, i.e., with respect to a fidelity criterion. Moreover, suppose that the decoder has access to side information which is also correlated with the two sources. This situation is depicted in Figure 1. Clearly, the two encoders could compress their respective

*This work was presented in part at the Joint Conference on Communications and Coding, Saas Fee, Switzerland, March 2-9, 2002.

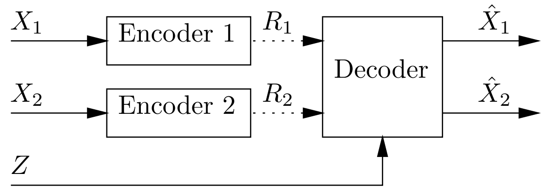

Figure 1: Separate compression of two sources with side information at the receiver.

sources ignoring each other's presence as well as that of the side information. The smallest rates under such an assumption are given by standard single-source rate-distortion theory [2].

As it turns out, however, both the fact that the two sources are dependent and the side information Z permit to lower the necessary rates (and/or the incurred distortions). The first gain (stemming from the fact that the sources are dependent) has been studied and solved by Slepian and Wolf [3] for the case of lossless compression. They considered the system of Figure 1 without the side information Z. Their surprising result is that the total rate needed for separate (lossless) compression of two sources X_1 and X_2 is the same as the rate needed for joint compression of the two sources, which is their joint entropy $H(X_1, X_2)$. When the compression is lossy, the dependence between the sources still permits to lower the rates (see e.g. [4]), but no final results are available to date. The second gain, i.e., the one due to side information, has been determined for the case of lossy compression of a single source by Wyner and Ziv [1]; the result is called the Wyner-Ziv rate-distortion function.

In this paper, we study the Wyner-Ziv problem with two (and more) discrete[1] memoryless dependent sources X_1, X_2 and a side information random variable Z. More precisely, X_1, X_2 and Z are random variables whose respective alphabets are arbitrary finite sets. Their joint distribution

$$p(x_1, x_2, z) \tag{1}$$

is given, and the desired result is a rate region, i.e., the set of rate pairs (R_1, R_2) that permit

[1]The conjecture is that our main results extend without modification to the case of continuous alphabets; however, the proofs in this paper are limited to discrete alphabets, in line with most treatments of related issues. Notable exceptions to this include Wyner's extension [5] of [1] to continuous alphabets, and [6].

the decoder to satisfy the distortion constraints

$$Ed_1(X_1, \hat{X}_1) \leq D_1, \tag{2}$$

$$Ed_2(X_2, \hat{X}_2) \leq D_2, \tag{3}$$

where $\hat{X}_i$ is the estimate that the receiver produces of X_i, using all the received codeword indices together with the side information. The distortion measures $d_1(\cdot)$ and $d_2(\cdot)$ are arbitrary. We denote this rate region by $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$.

In Section 2, we derive an achievable region $\mathcal{R}_a(D_1, D_2) \subseteq \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$. The coding technique that leads to this region is an extension of the code of [3] (and its lossy extension given in [4]), combined with the code construction of Wyner and Ziv [1] (which is actually also based on [3]).

In Section 3, we discuss inner bounds to the rate region $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$. For the problem without side information, in spite of numerous attempts, the rate region of [4] could not be shown to be optimal, i.e., there is no converse to this rate region. Certain partial answers appear in the literature (see e.g. [7, 8, 6]). All those answers apply to cases where one of the two sources is either not to be reconstructed, or encoded perfectly. Consequently, it is not surprising that we cannot show $\mathcal{R}_a(D_1, D_2)$ to be the optimal rate region for the problem of Figure 1, either. Instead, we determine a general inner bound to the rate region, denoted by $\mathcal{R}_i(D_1, D_2) \supseteq \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$. This inner region is generally larger than $\mathcal{R}_a(D_1, D_2)$.

Nevertheless, in Section 4, we show that the two regions do coincide in one situation that is *not* an extension of the converses studied for the case without side information (as reported in [7, 8, 6]): Our converse applies to a *symmetric* case, i.e., *both* sources are encoded, and *both* sources have to be reconstructed with respect to a fidelity criterion. More precisely, our rate region is optimal when the two sources are *conditionally independent* given the side information. Note that this means that the sources can be arbitrarily dependent as long as the side information makes them conditionally independent. To our knowledge, this is the first such result for a distributed lossy compression problem.

In Section 5, we consider a complementary question which is of natural interest to distributed source coding problems: What rate penalty is to be paid for doing distributed compression, rather than joint? Slepian and Wolf proved the surprising result that for the case of lossless distributed compression (Figure 1, without side information), there is *no* *rate penalty* [3]. For lossy compression, this does not hold. This was first observed for the Wyner-Ziv problem with a single source: the rate *can* be lowered in general if the side information is known at the encoder [1]. Consequently, it is suspected that for the case of lossy distributed compression, there is a penalty as well. However, this question could not be settled because as pointed out above, the rate region for lossy distributed compression is not known. In extension of these arguments, it must be suspected that for the problem of Figure 1, there is a rate penalty with respect to joint encoding as shown in Figure 2, at least in the general case. In this paper, we prove that in the special case where X_1 and X_2 are conditionally independent given Z, this penalty vanishes. To our knowledge, this is the first such result for a distributed lossy compression problem.

The results of this paper have natural applications to distributed compression and signal processing. One such extension is described in [9]. At the same time, they are also relevant to establish certain rate regions for multiple-relay channels. As a matter of fact, Figure 1 can be understood as a relay network: The two boxes labeled "encoder" are two relays that observe X_1 and X_2, respectively, and the strategy is for the relays to compress their observations for a final destination that observes Z. Since X_1, X_2 and Z were all produced by the source of the relay network, they are generally correlated. This is further explained in [10].

2 An achievable rate region

An achievable rate region can be obtained by extending the coding scheme introduced by Slepian and Wolf in [3], which is sometimes called "binning." In summary, the code leading to the achievable rate region given by Theorem 2 is a cascade of a suitable vector quantizer with a binning operation for the codeword indices. In particular, the encoder of the source X_1 must apply a binning operation with respect to both the codeword index of source X_2 *and* the side information Z, and source X_2 must do likewise. Given the two bin indices selected by the two encoders, the decoder then uses the side information to undo the binning and retrieve the correct quantization cell indices. But this requires the two bin indices to be jointly typical with the side information. More precisely, the fact that each bin index is jointly typical with its corresponding source sequence must imply that the bin indices and the side information form a jointly typical triplet. The key to such an implication is the

Markov lemma.

Lemma 1 (Markov lemma). *Suppose $Z - X_1 - W_1$. For a fixed $(z^n, x_1^n) \in A_\epsilon^{*(n)}$, draw a W_1^n according to $p(w_1|x_1)$, which implies (with high probability) that $(x_1^n, W_1^n) \in A_\epsilon^{*(n)}$. With high probability, $(z^n, x_1^n, W_1^n) \in A_\epsilon^{*(n)}$.*

Proof. The lemma is quoted without proof in [11, Lemma 14.8.1, p. 436]; with reference to [4, 12]. In [4], the Markov lemma (Lemma 4.1) has a prominent role as well as a proof. □

To understand the lemma, note that from the fact that a sequence z^n is jointly typical with x_1^n, and x_1^n is jointly typical with w_1^n, it does not yet follow that the three sequences form a jointly typical triplet. One sufficient condition to ensure that they *do* form a jointly typical triplet is precisely the stated Markov relationship. This lemma is at the heart of our achievable rate region; the necessary extension is given in the appendix. We did not find a weaker condition that permits to infer the same conclusions.

Once the decoder has retrieved the correct codeword indices, it can use the side information a second time in order to remove a part of the quantization noise. This is possible because the side information is available in unquantized form. The rates that can be achieved by this coding scheme can be expressed as follows:

Theorem 2. *$\mathcal{R}_a(D_1, D_2) \subseteq \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$, where $\mathcal{R}_a(D_1, D_2)$ is the set of all rate pairs (R_1, R_2) such that there exists a pair (W_1, W_2) of discrete random variables with $p(w_1, w_2, x_1, x_2, z) = p(w_1|x_1)p(w_2|x_2)p(x_1, x_2, z)$ for which the following conditions are satisfied*

$$R_1 \geq I(X_1; W_1) - I(Z, W_2; W_1) \tag{4}$$

$$R_2 \geq I(X_2; W_2) - I(Z, W_1; W_2) \tag{5}$$

$$R_1 + R_2 \geq I(X_1; W_1) + I(X_2; W_2) - I(Z, W_2; W_1) - I(Z, W_1; W_2) + I(W_1; W_2|Z), \tag{6}$$

and for which there exist functions $g_1(\cdot)$ and $g_2(\cdot)$ such that

$$Ed_1(X_1, g_1(W_1, W_2, Z)) \leq D_1, \quad \text{and} \tag{7}$$

$$Ed_2(X_2, g_2(W_2, W_1, Z)) \leq D_2. \tag{8}$$

Proof. The proof of this theorem is given in the appendix.

In order to easily compare this achievable rate region to the inner bound derived below, we prove that the rate region of Theorem 2 can be rewritten in a different shape:

Corollary 3 (equivalent statement). *The rate bounds of Theorem 2 can be expressed equivalently as*

$$R_1 \geq I(X_1, X_2; W_1|Z, W_2) \tag{9}$$

$$R_2 \geq I(X_1, X_2; W_2|Z, W_1) \tag{10}$$

$$R_1 + R_2 \geq I(X_1, X_2; W_1, W_2|Z). \tag{11}$$

Proof. The proof of this corollary is given in the appendix.

The rate region $\mathcal{R}_a$ can easily be extended to more than two sources. For brevity and since the proof does not contain new ideas, we omit the explicit statement, pointing out that the key lies in Item 3 of the proof of Theorem 2: In the case of more than two sources, many more error events must be analyzed. While this is straightforward (using appropriate extensions of Lemma 10), the resulting expressions are rather cumbersome.

3 A General Inner Bound

In this section, we present a region $\mathcal{R}_i(D_1, D_2)$ which contains the desired rate-distortion region $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$. The region $\mathcal{R}_i(D_1, D_2)$ follows from standard outer bounding arguments; it is a slight extension of the arguments given in [4, Thm. 6.2].

Theorem 4 (inner region). $\mathcal{R}_i(D_1, D_2) \supseteq \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$*, where* $\mathcal{R}_i(D_1, D_2)$ *is the set of all rate pairs* (R_1, R_2) *such that there exists a pair* (W_1, W_2) *of discrete random variables with* $p(w_1|x_1, x_2, z) = p(w_1|x_1)$ *and* $p(w_2|x_1, x_2, z) = p(w_2|x_2)$ *for which the following conditions are satisfied*

$$R_1 \geq I(X_1X_2; W_1|ZW_2) \tag{12}$$

$$R_2 \geq I(X_1X_2; W_2|ZW_1) \tag{13}$$

$$R_1 + R_2 \geq I(X_1X_2; W_1W_2|Z) \tag{14}$$

and for which there exist functions $g_1(\cdot)$ *and* $g_2(\cdot)$ *such that*

$$Ed_1(X_1, g_1(W_1, W_2, Z)) \leq D_1, \quad \textit{and} \tag{15}$$

$$Ed_2(X_2, g_2(W_2, W_1, Z)) \leq D_2. \tag{16}$$

The region $\mathcal{R}_i(D_1, D_2)$ given in Theorem 4 cannot generally be shown to coincide with $\mathcal{R}_a(D_1, D_2)$, and hence, no final rate-distortion result can be given for the Wyner-Ziv rate-distortion problem with multiple sources. The difference between the two regions is best understood by comparing Corollary 3 with Theorem 4: The mutual information expressions are exactly the same both for $\mathcal{R}_i(D_1, D_2)$ and $\mathcal{R}_a(D_1, D_2)$; the difference occurs only in the degrees of freedom in choosing the auxiliary random variables W_1 and W_2, i.e., the difference lies in the space over which the minimization is carried out.

The achievability of the rate region $\mathcal{R}_a(D_1, D_2)$ relies on the Markov lemma. But this means that achievability can be guaranteed only if

$$p(x_1, x_2, z, w_1, w_2) = p(x_1, x_2, z)p(w_1|x_1)p(w_2|x_2). \tag{17}$$

In the derivation of an inner region, we clearly cannot make such a strong assumption. For $\mathcal{R}_i(D_1, D_2)$, all we can be sure of is

$$p(x_1, x_2, z, w_1) = p(x_1, x_2, z)p(w_1|x_1) \tag{18}$$

$$p(x_1, x_2, z, w_2) = p(x_1, x_2, z)p(w_2|x_2). \tag{19}$$

The difference between $\mathcal{R}_a(D_1, D_2)$ and $\mathcal{R}_i(D_1, D_2)$ stems therefore from the difference between the two sets of auxiliary random variables. Hence, one way to obtain a final result would be to set up the probability space in such a way that these two sets of auxiliary random variables coincide. However, as we show in the next section, there is another way to obtain a final result: If X_1 and X_2 are conditionally independent given Z, the two sets of auxiliary random variables do not coincide, but the mutual information and distortion functionals simplify to the extent that they no longer depend on the additional degrees of freedom in the choice of the auxiliary random variables in $\mathcal{R}_i(D_1, D_2)$. For this reason, the two regions coincide.

4 Partial Converse: X_1 and X_2 are conditionally independent given Z

While the two rate regions derived in this paper, $\mathcal{R}_a(D_1, D_2)$ and $\mathcal{R}_i(D_1, D_2)$, do not coincide in general, we now analyze a special case in which they indeed do coincide, hence establishing a true rate-distortion result. This special case is when X_1 and X_2 are independent given Z, i.e., when

$$p(x_1, x_2, z) = p(x_1|z)p(x_2|z)p(z). \tag{20}$$

The first step in our derivation is to rewrite the achievable region, introducing the simplifications due to the assumption that X_1 and X_2 are conditionally independent given Z.

Corollary 5. *If X_1 and X_2 are conditionally independent given Z, $\mathcal{R}_a(D_1, D_2) \subseteq \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$, where $\mathcal{R}_a(D_1, D_2)$ is the set of all rate pairs (R_1, R_2) such that there exists a pair (W_1, W_2) of discrete random variables with $p(w_1, w_2, x_1, x_2, z) = p(w_1|x_1)p(w_2|x_2)p(x_1|z)p(x_2|z)p(z)$ for which the following conditions are satisfied*

$$R_1 \geq I(X_1; W_1) - I(Z; W_1) \tag{21}$$

$$R_2 \geq I(X_2; W_2) - I(Z; W_2) \tag{22}$$

and for which there exist functions $g_1(\cdot)$ and $g_2(\cdot)$ such that

$$Ed_1(X_1, g_1(W_1, W_2, Z)) \leq D_1, \quad \textit{and} \tag{23}$$

$$Ed_2(X_2, g_2(W_2, W_1, Z)) \leq D_2. \tag{24}$$

Proof. The term $I(W_1; W_2|Z)$ in the sum rate bound is zero. Therefore, the sum rate bound becomes just the sum of the two side bounds, and hence can be omitted: the achievable rate region becomes a square. The side bounds can be simplified by noting that

$$I(Z, W_2; W_1) = I(Z; W_1) + \underbrace{I(W_1; W_2|Z)}_{=0} \tag{25}$$

where the last term is zero. □

The inner bound to the rate region derived in this paper, $\mathcal{R}_i(D_1, D_2)$ can also be simplified in the special case when X_1 and X_2 are conditionally independent given Z. Using this assumption, we can give the following new inner bound to the rate region:

Corollary 6. *If X_1 and X_2 are conditionally independent given Z, $\mathcal{R}'_i(D_1, D_2) \supseteq \mathcal{R}_i(D_1, D_2)$, and hence $\mathcal{R}'_i(D_1, D_2) \supseteq \mathcal{R}_{X_1,X_2|Z}(D_1, D_2)$ where $\mathcal{R}'_i(D_1, D_2)$ is the set of all rate pairs (R_1, R_2) such that there exists a pair (W_1, W_2) of discrete random variables with $p(w_1|x_1, x_2, z) = p(w_1|x_1)$ and $p(w_2|x_1, x_2, z) = p(w_2|x_2)$ for which the following conditions are satisfied*

$$R_1 \geq I(X_1; W_1) - I(Z; W_1) \tag{26}$$

$$R_2 \geq I(X_2; W_2) - I(Z; W_2) \tag{27}$$

and for which there exist functions $g_1(\cdot)$ and $g_2(\cdot)$ such that

$$Ed_1(X_1, g_1(W_1, W_2, Z)) \leq D_1, \quad \textit{and} \tag{28}$$

$$Ed_2(X_2, g_2(W_2, W_1, Z)) \leq D_2. \tag{29}$$

Proof. The goal is to lower bound the results of Theorem 4. Clearly, we can enlarge the region $\mathcal{R}_i(D_1, D_2)$ by removing the sum rate bound. Moreover, Equation (12), which requires

$$R_1 \geq I(X_1, X_2; W_1 | Z, W_2), \tag{30}$$

can be relaxed to yield

$$R_1 \geq I(X_1; W_1 | Z W_2) + I(X_2; W_1 | Z, W_2, X_1) \tag{31}$$

$$\geq I(X_1; W_1 | Z W_2). \tag{32}$$

To lower bound this term, we can write out the following term:

$$I(X_1; W_1, W_2 | Z) = I(X_1; W_1 | Z) + I(X_1; W_2 | Z W_1) \tag{33}$$

$$= I(X_1; W_2 | Z) + I(X_1; W_1 | Z W_2). \tag{34}$$

However, under the assumption that X_1 and X_2 are conditionally independent given Z, it is true that $I(X_1; W_2 | Z) = 0$. This directly implies

$$R_1 \geq I(X_1; W_1 | Z). \tag{35}$$

By the Markov relationship $Z - X_1 - W_1$, this can be rewritten as

$$R_1 \geq I(X_1; W_1) - I(Z; W_1). \tag{36}$$

The same derivation applies to R_2, and hence, an inner region $\mathcal{R}_i' \supseteq \mathcal{R}_i$ is given by

$$R_1 \geq I(X_1; W_1) - I(Z; W_1) \tag{37}$$

$$R_2 \geq I(X_2; W_2) - I(Z; W_2). \tag{38}$$

□

The main result of this section follows by combining Corollaries 5 and 6, and by observing that the additional degrees of freedom in Corollary 6 (more particularly, the additional freedom in choosing the auxiliary random variables W_1 and W_2) do not permit to lower the involved mutual information and distortion functionals. It follows that the rate regions described by Corollaries 5 and 6 are actually the same, and hence, that they correspond to the desired rate-distortion region $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$.

Theorem 7. *If X_1 and X_2 are conditionally independent given Z, then $\mathcal{R}_a(D_1, D_2) = \mathcal{R}_i(D_1, D_2) = \mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$.*

Proof. Both rate regions have the same shape *except* that the minimization in the converse is extended over a larger space (relaxed Markov condition). However, since all of the involved mutual information functionals only depend on the joint marginals of (X_1, W_1), (X_2, W_2), (Z, W_1), and (Z, W_2), the additional degrees of freedom cannot lower their values.

More precisely, suppose the minimization of the outer bound has yielded a certain rate pair (R_1, R_2) with a pair of auxiliary random variables (W_1, W_2) satisfying the conditions of Corollary 6, according to some distribution

$$p(w_1, w_2, x_1, x_2, z) = p(w_1, w_2|x_1, x_2)p(x_1|z)p(x_2|z)p(z). \tag{39}$$

But then, construct the auxiliary random variables (W_1', W_2') such that

$$p(w_1'|x_1) = \sum_{w_2,x_2} p(w_1, w_2|x_1, x_2)p(x_2|x_1), \tag{40}$$

$$p(w_2'|x_2) = \sum_{w_1,x_1} p(w_1, w_2|x_1, x_2)p(x_1|x_2). \tag{41}$$

The joint distribution

$$p(w_1', w_2', x_1, x_2, z) = p(w_1'|x_1)p(w_2'|x_2)p(x_1|z)p(x_2|z)p(z) \tag{42}$$

achieves the same values in the mutual information functionals in Corollaries 5 and 6, but W_1' and W_2' are achievable since they satisfy the conditions of Corollary 5.

For the distortion D_1, the key argument is that X_1 and W_2 are conditionally independent given Z. To make this precise, pick any decoding function $g_1(W_1, W_2, Z)$ that achieves $Ed_1(X_1, g_1(W_1, W_2, Z)) = D$. Now suppose that the decoder knows X_2. Clearly, it can still apply the encoding function if it wishes to do so, which implies that there exists a function $\tilde{g}_1(W_1, X_2, Z)$ satisfying $Ed_1(X_1, \tilde{g}_1(W_1, X_2, Z)) = D$. However, determine

$$x_2^*(z) = \arg\min_{x_2} \sum_{x_1} d_1(x_1, \tilde{g}_1(w_1, x_2, z)p(x_1|z). \tag{43}$$

Hence, choosing the decoding function $g'(w_1, z) = \tilde{g}_1(w_1, x_2^*(z), z)$ implies

$$D = Ed_1(X_1, \tilde{g}_1(W_1, X_2, Z)) \quad \geq \quad Ed_1(X_1, g_1'(W_1, Z)). \tag{44}$$

This also holds if W_1 is replaced by W_1' on the right hand side since both have the same distribution with respect to X_1 and Z. This implies that any distortion D_1 achievable for the pair of auxiliary random variables (W_1, W_2) is also achievable for the pair of auxiliary random variables (W_1', W_2'). The same holds for D_2, which completes the proof. □

By the nature of the arguments leading to Theorem 7, and in particular by the fact that the rate region becomes a square, it is clear that the result carries over to the case of more than two sources.

5 Separate vs Joint Encoding

In this section, we compare separate and joint encoding of the two sources. Separate encoding is illustrated in Figure 1, and joint encoding in Figure 2. Clearly, the minimum rate needed for joint encoding cannot be larger than the minimum sum rate needed for separate encoding. For the case of lossless encoding and no side information, Slepian and Wolf [3] found that the two rates are equal.

Unfortunately, the same does not seem to hold generally when the compression is lossy. Final results for the Slepian-Wolf problem with distortion are still missing, but a simpler problem has been studied in detail: The problem of source coding with side information. The lossy case has been solved by Wyner and Ziv [1]. The comparison is between the case where the side information is known both at the encoder and at the decoder, and the case where it is only known at the decoder. Wyner and Ziv found that there is generally a difference in rate between the two cases. This difference vanishes in certain special cases, including the jointly Gaussian case with mean-squared error. This suggests that in the Slepian-Wolf problem with lossy compression, there is a rate loss between separate and joint encoding of the two sources.

For the Wyner-Ziv problem with multiple sources (as shown in Figure 1), two types of rate losses could be studied. On the one hand, there may be a rate penalty because the side information is not available at either encoder. Scenarios of interest may also include asymmetric cases where some encoders know the side information, while others do not.

On the other hand, there may be a rate penalty because the sources have to be encoded separately rather than jointly. In the present paper, we only study the latter kind of rate losses, i.e., the difference between the minimum sum rate required in the system depicted in Figure 1 and the minimum rate required in Figure 2.

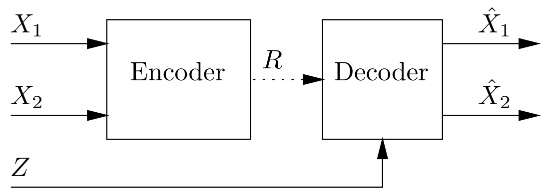

Figure 2: Joint compression of two sources with side information at the receiver.

The system of Figure 2 is almost the original Wyner-Ziv problem. The only difference is that the distortion constraint has a vector form,

$$Ed_1(X_1, \hat{X}_1) \leq D_1 \quad (45)$$

$$Ed_2(X_2, \hat{X}_2) \leq D_2. \quad (46)$$

The solution of Wyner and Ziv in [1] naturally extends to this case. The minimum rate required in the system of Figure 2 can be expressed as

$$R^{WZ}_{(X_1,X_2)|Z}(D_1, D_2) = \min_{p(w|x_1,x_2)} I(X_1, X_2; W|Z), \quad (47)$$

where the minimization is over all auxiliary random variables W for which there exist functions $g_1(\cdot)$ and $g_2(\cdot)$ satisfying

$$Ed_1(X_1, g_1(W, Z)) \leq D_1 \quad (48)$$

$$Ed_2(X_2, g_2(W, Z)) \leq D_2. \quad (49)$$

Hence, in this section, we compare the rate $R^{WZ}_{(X_1,X_2)|Z}(D_1, D_2)$ with the object of study of this paper, i.e., the rate region $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$. In line with the comments made above for the separate lossy compression of multiple sources, it must be expected that this rate loss is nonzero. However, to date, no final statement can be made since the region $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$ is not known in general.

For one special case, however, we determined $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$ in this paper: when X_1 and X_2 are conditionally independent given the side information Z. Here, we can compare $R^{WZ}_{(X_1,X_2)|Z}(D_1, D_2)$ and $\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1, D_2)$, and it turns out that the rate loss is zero. In other words, when the sources are conditionally independent given the side information, the source coding problem of Figure 2 can be solved by *separately* encoding X_1 and X_2, i.e., by using a system as shown in Figure 1. This can be established by an argument similar to the proof of Theorem 7. More precisely, we prove the following statement:

Theorem 8. *If X_1 and X_2 are conditionally independent given the side information Z, then*

$$\min_{(R_1,R_2)\in\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1,D_2)} R_1+R_2 \quad = \quad R^{WZ}_{(X_1,X_2)|Z}(D_1,D_2). \tag{50}$$

Proof. Clearly, the minimum sum rate for separate coding cannot be smaller than the minimum rate for joint coding, i.e.,

$$\min_{(R_1,R_2)\in\mathcal{R}^{WZ}_{X_1,X_2|Z}(D_1,D_2)} R_1+R_2 \quad \geq \quad R^{WZ}_{(X_1,X_2)|Z}(D_1,D_2). \tag{51}$$

We have to prove that when X_1 and X_2 are independent given Z, the inequality also holds in the other direction. Recall from [1] that

$$\begin{aligned} R^{WZ}_{(X_1,X_2)|Z}(D_1,D_2) &= \min_W I(X_1,X_2;W|Z) && (52)\\ &= \min_W I(X_1;W|Z)+I(X_2;W|Z,X_1) && (53)\\ &\geq \min_W I(X_1;W|Z)+I(X_2;W|Z), && (54)\end{aligned}$$

where the last inequality holds because

$$\begin{aligned} I(X_2;X_1,W|Z) &= \underbrace{I(X_2;X_1|Z)}_{=0}+I(X_2;W|Z,X_1) && (55)\\ &= I(X_2;W|Z)+I(X_2;X_1|Z,W). && (56)\end{aligned}$$

Now select two new random variables W_1 and W_2 depending arbitrarily both on X_1 and X_2 in such a way that $I(X_1;W_1,W_2|Z)=I(X_1;W|Z)$ and $I(X_2;W_1,W_2|Z)=I(X_2;W|Z)$. A trivial way to satisfy this is by choosing $W_1=W_2=W$, and hence,

$$\begin{aligned} R^{WZ}_{(X_1,X_2)|Z}(D_1,D_2) &\geq \min_{W_1,W_2} I(X_1;W_1,W_2|Z)+I(X_2;W_1,W_2|Z) && (57)\\ &\geq \min_{W_1,W_2} I(X_1;W_1|Z)+I(X_2;W_2|Z) && (58)\\ &= \min_{W_1} I(X_1;W_1|Z)+\min_{W_2} I(X_2;W_2|Z), && (59)\end{aligned}$$

where Equation (57) follows because the introduction of W_1 and W_2 can only make the minimization space larger, and Equation (58) follows by writing out

$$I(X_1;W_1,W_2|Z) \quad = \quad I(X_1;W_1|Z)+I(X_1;W_2|Z,W_1). \tag{60}$$

The auxiliary random variables W_1 and W_2 can be chosen arbitrarily, and the joint density takes on the form

$$p(w_1, w_2, x_1, x_2, z) \quad = \quad p(w_1, w_2|x_1, x_2)p(x_1|z)p(x_2|z)p(z). \tag{61}$$

Construct alternative auxiliary random variables (W_1', W_2') such that

$$p(w_1'|x_1) \quad = \quad \sum_{w_2, x_2, z} p(w_1, w_2|x_1, x_2)p(x_2|x_1), \tag{62}$$

$$p(w_2'|x_2) \quad = \quad \sum_{w_1, x_1, z} p(w_1, w_2|x_1, x_2)p(x_1|x_2). \tag{63}$$

The joint distribution

$$p(w_1', w_2', x_1, x_2, z) \quad = \quad p(w_1'|x_1)p(w_2'|x_2)p(x_1|z)p(x_2|z)p(z) \tag{64}$$

achieves the same values in the mutual information functionals in Equation (59). It can be shown by precisely the same argument as in the proof of Theorem 7 that (W_1', W_2') also permit to achieve the same distortions D_1 and D_2. However, if the auxiliary random variables satisfy Equation (64), then Corollary 5 guarantees that

$$\min_{W_1} I(X_1; W_1|Z) + \min_{W_2} I(X_2; W_2|Z) \quad \in \quad \mathcal{R}^{WZ}_{X_1, X_2|Z}(D_1, D_2), \tag{65}$$

i.e., these rates can be achieved by separate encoding, which, in combination with Equation (59), implies that

$$\min_{(R_1, R_2) \in \mathcal{R}^{WZ}_{X_1, X_2|Z}(D_1, D_2)} R_1 + R_2 \quad \leq \quad R^{WZ}_{(X_1, X_2)|Z}(D_1, D_2). \tag{66}$$

This completes the proof. □

Theorem 8 establishes therefore that in the special case when the sources are conditionally independent given the side information, there is no rate loss due to separate encoding of the sources, just like in the problem of Slepian and Wolf [3].

By the nature of the arguments leading to Theorem 8, it is clear that the result carries over to the case of more than two sources.

6 Conclusion

Distributed lossy compression as shown in Figure 1 (without the side information) is a long-standing open problem. The best known achievable rate region is the one given in [4], and it does not coincide with any converse bound; no final rate-distortion results can be given. In this paper, we investigated distributed lossy compression with side information, and we established final rate-distortion results for one special case. Our problem can be seen as the extension of the Wyner-Ziv problem [1] to multiple sources, or as the extension of the problem of Slepian and Wolf [3] to the case of lossy compression and the presence of side information. As a matter of fact, the achievable rate region is found by extending and combining the strategies of [3] (or, more precisely, its lossy extension that was presented in [4]), and [1]. An inner bound to the rate region was also determined in extension of the arguments in [4]. In contrast to earlier work, inner bound and achievable rate region could be shown to coincide in a *symmetric* case, i.e., where both sources are compressed with respect to a fidelity criterion, and both sources are to be reconstructed. More precisely, the bounds coincide when the sources are conditionally independent given the side information. In that special case, it was also shown that there is no rate loss between separate and joint compression of the sources. To our knowledge, this is the first rate-distortion result for a truly distributed lossy compression problem. While we provide explicit solutions and proofs for the case of two sources, it is clear that the key arguments can be extended directly to the case of more than two sources.

The results of this paper are of interest to distributed compression and signal processing, such as sensor networks. This is further explained and extended in [9]. Another application of the results of this paper is to relay networks. More precisely, the coding strategies needed for the problem studied in this paper are also useful to exploit the potential furnished by relays. This is further explained in [10].

Acknowledgments

The author would like to acknowledge the excellent manuscript of Toby Berger [4], as well as stimulating discussions with Pier Luigi Dragotti, Gerhard Kramer, and Martin Vetterli.

A Proofs

A.1 Notation

In the proofs, we need to denote *sequences* of random variables. Subsequent symbols of the source X_1 are denoted $X_{1,1}, X_{1,2}, X_{1,3}, \dots$, and we will use the notation

$$X_{1,i}^{j} \stackrel{def}{=} \{X_{1,i}, X_{1,i+1}, \dots, X_{1,j}\}. \tag{67}$$

We also use the shorthand $X_1^n \stackrel{def}{=} X_{1,1}^n$.

Moreover, we use the symbol $A_\epsilon^{*(n)}$ to denote the strongly typical set as defined in [11, p. 359].

A.2 Achievable Rate Region

Lemma 9 (extended Markov lemma). *Suppose that*

$$p(w_1, w_2, x_1, x_2, z) = p(w_1|x_1)p(w_2|x_2)p(x_1, x_2, z). \tag{68}$$

For a fixed $(z^n, x_1^n, x_2^n) \in A_\epsilon^{*(n)}$, draw a W_1^n according to $p(w_1|x_1)$, which implies (with high probability) that $(x_1^n, W_1^n) \in A_\epsilon^{*(n)}$, and draw a W_2^n according to $p(w_2|x_2)$, which implies (with high probability) that $(x_2^n, W_2^n) \in A_\epsilon^{*(n)}$. With high probability, $(z^n, x_1^n, x_2^n, W_1^n, W_2^n) \in A_\epsilon^{*(n)}$.

Proof. We use the Markov lemma twice, first to establish that $((x_2, z), x_1, w_1)$ are jointly typical, and that $((x_1, z), x_2, w_2)$ are jointly typical. Using the Markov relationship $W_1 - (X_1, Z, X_2) - W_2$, the Markov lemma implies that (w_1, x_1, z, x_2, w_2) are also jointly typical. □

Lemma 10. *If $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \sim p(x^n)p(y^n)p(z^n)$, i.e. they have the same marginals as $p(x^n, y^n, z^n)$ but they are independent, then the probability that $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$ is in the strongly typical set $A_\epsilon^{*(n)}$ of $p(x^n, y^n, z^n)$ can be bounded by*

$$Pr((\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \in A_\epsilon^{*(n)}) \leq 2^{-n(I(Z,X;Y)+I(Z,Y;X)-I(X;Y|Z)-4\epsilon)}. \tag{69}$$

Proof. The lemma can be established along the lines of [11, Thm. 13.6.2]. We here give the explicit argument for the case of weak typicality, i.e., extending the proof of [11, Thm.

8.6.1], in particular Equations (8.55)-(8.57) on page 197:

$$
\begin{aligned}
Pr((\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \in A_\epsilon^{(n)}) &= \sum_{(x^n,y^n,z^n)\in A_\epsilon^{(n)}} p(x^n)p(y^n)p(z^n) && (70)\\
&\leq 2^{n(H(X,Y,Z)+\epsilon)}2^{-n(H(X)-\epsilon)}2^{-n(H(Y)-\epsilon)}2^{-n(H(Z)-\epsilon)} && (71)\\
&= 2^{-n(H(X)+H(Y)+H(Z)-H(X,Y,Z)-4\epsilon)}, && (72)
\end{aligned}
$$

where $A_\epsilon^{(n)}$ denotes the typical set as defined in [11, p. 51]. The expression in the exponent can be rewritten as $I(Z,X;Y) + I(Z,Y;X) - I(X;Y|Z)$, which completes the proof. □

Proof of Theorem 2. Fix $p(w_1|x_1)$ and $p(w_2|x_2)$ as well as $g_1(w_1, w_2, z)$ and $g_2(w_2, w_1, z)$. Suppose that $Ed_1(X_1, g_1(W_1, W_2, Z)) \leq D_1$ and $Ed_2(X_2, g_2(W_2, W_1, Z)) \leq D_2$.

Generation of Codebooks: Generate $2^{nR_1'}$ codewords of length n, sampled iid from the marginal distribution $p(w_1)$. Label these as $\mathbf{w_1}(s_1)$, with $s_1 \in \left\{1, 2, \ldots, 2^{nR_1'}\right\}$.

Generate $2^{nR_2'}$ codewords of length n, sampled iid from the marginal distribution $p(w_2)$. Label these as $\mathbf{w_2}(s_2)$, with $s_2 \in \left\{1, 2, \ldots, 2^{nR_2'}\right\}$.

Provide 2^{nR_1} random bins with indices t_1. Randomly assign to every codeword $\mathbf{w_1}(s_1)$ a bin index $t_1 \in \{1, 2, \ldots, 2^{nR_1}\}$. Denote the set of codeword indices s_1 with bin index t_1 as $B_1(t_1)$.

Provide 2^{nR_2} random bins with indices t_2. Randomly assign to every codeword $\mathbf{w_2}(s_2)$ a bin index $t_2 \in \{1, 2, \ldots, 2^{nR_2}\}$. Denote the set of codeword indices s_2 with bin index t_2 as $B_2(t_2)$.

Encoding: Given a source sequence X_1^n, the first encoder looks for a codeword $W_1^n(s_1)$ such that $(X_1^n, W_1^n(s_1)) \in A_\epsilon^{*(n)}$. The first encoder sends the index t_1 of the bin in which s_1 belongs.

Given a source sequence X_2^n, the second encoder looks for a codeword $W_2^n(s_2)$ such that $(X_2^n, W_2^n(s_2)) \in A_\epsilon^{*(n)}$. The second encoder sends the index t_2 of the bin in which s_2 belongs.

Decoding: The decoder looks for a pair $(W_1^n(s_1), W_2^n(s_2))$ such that $s_1 \in B_1(t_1)$, $s_2 \in B_2(t_2)$, and $(W_1^n(s_1), W_2^n(s_2), Z^n) \in A_\epsilon^{*(n)}$. If he finds a unique (s_1, s_2), he calculates $(\hat{X}_1^n, \hat{X}_2^n)$.

Analysis of Error Events:

1. The encoders do not find jointly typical codewords. This is prevented if

$$R_1' > I(X_1; W_1) \tag{73}$$

$$R_2' > I(X_2; W_2) \tag{74}$$

2. The two true codewords and the side information do not form a jointly typical triplet.

   We have assumed that

$$p(w_1, w_2, x_1, x_2, z) = p(w_1|x_1)p(w_2|x_2)p(x_1, x_2, z). \tag{75}$$

   Hence, by Lemma 9 (the extended Markov lemma), the probability of this error event goes to zero, and the decoder will identify the correct pair (s_1, s_2) with high probability.

3. There exists an alternative choice of two codewords that, together with the side information, form a jointly typical triplet.

   Denote the correct codeword indices by s_1 and s_2. First, we consider the case where the first codeword index is in error while the second is retrieved correctly. By [11, Lemma 13.6.2], a particular pair (s_1', s_2) occurs with a probability that can be bounded as

$$Pr\{(w_1(s_1'), w_2(s_2), z) \in A_\epsilon^{*(n)}\} \leq 2^{-n(I(W_1;W_2,Z)-7\epsilon)}. \tag{76}$$

   The decoder may confuse with any one of the codewords $\mathbf{w}_1(s_1')$ that have bin index t_1, i.e., the same bin index as the true codeword $\mathbf{w_1}(s_1)$. Hence, the error event is

$$Pr\{\exists s_1' \in B(t_1), s_1' \neq s_1 : (w_1(s_1'), w_2(s_2), z) \in A_\epsilon^{*(n)}\} \tag{77}$$

$$\leq \sum_{s_1' \neq s_1, s_1' \in B_1(t_1)} Pr\{(w_1(s_1'), w_2(s_2), z) \in A_\epsilon^{*(n)}\} \tag{78}$$

$$\leq 2^{R_1 - R_1'} 2^{-n(I(W_1;W_2,Z)-7\epsilon)}. \tag{79}$$

   The same derivation applies to the case where the first index is correctly retrieved while the second is in error.

   The third case is when both indices are in error. It is at this point that we need the extension of [11, Lemma 13.6.2], which was given above as Lemma 10. Using that lemma, the probability that a particular erroneous pair (s'_1, s'_2) occurs can be bounded by

$$
\begin{aligned}
& Prob\{(W_1^n(s'_1), W_2^n(s'_2), Z^n) \in A_\epsilon^{*(n)}\} && (80)\\
& \le 2^{-n(I(Z,W_2;W_1)+I(Z,W_1;W_2)-I(W_1;W_2|Z)-14\epsilon)}, && (81)
\end{aligned}
$$

and hence, the probability of the error event can be bounded as

$$
\begin{aligned}
& Prob\{\exists(s'_1, s'_2), s'_1 \in B(t_1), s'_1 \neq s_1, s'_2 \in B(t_2), s'_2 \neq s_2 : && (82)\\
& \quad (W_1^n(s'_1), W_2^n(s'_2), Z^n) \in A_\epsilon^{*(n)}\} && (83)\\
& \le \sum_{s'_1 \in B(t_1), s'_1 \neq s_1, s'_2 \in B(t_2), s'_2 \neq s_2} Pr\{(w_1(s'_1), w_2(s_2), z) \in A_\epsilon^{*(n)}\} && (84)\\
& \le 2^{R_1 - R'_1 + R_2 - R'_2} 2^{-n(I(Z,W_2;W_1)+I(Z,W_1;W_2)-I(W_1;W_2|Z)-7\epsilon)}. && (85)
\end{aligned}
$$

In conclusion, this error event will have vanishingly small error probability (as n tends to infinity) when the following rate conditions are satisfied:

$$
\begin{aligned}
R'_1 - R_1 &< I(Z, W_2; W_1) && (86)\\
R'_2 - R_2 &< I(Z, W_1; W_2) && (87)\\
R'_1 - R_1 + R'_2 - R_2 &< I(Z, W_2; W_1) + I(Z, W_1; W_2) - I(W_1; W_2|Z) && (88)
\end{aligned}
$$

4. If the indices (s_1, s_2) are decoded correctly, then $(X_1^n, X_2^n, W_1^n(s_1), W_2^n(s_2), Z^n) \in A_\epsilon^{*(n)}$. Therefore, the empirical joint distribution is close to $p(x_1, x_2, z)p(w_1|x_1)p(w_2|x_2)$, which by assumption has distortion (D_1, D_2).

□

Proof of Corollary 3. To prove this, we start by noting that

$$
I(X_1, X_2; W_1|Z, W_2) = I(X_1; W_1|Z, W_2) + \underbrace{I(X_2; W_1|Z, X_1, W_2)}_{=0}. \quad (89)
$$

Hence, we show equivalently that $I(X_1; W_1|Z, W_2) = I(X_1; W_1) + I(Z; W_1)$, as follows:

$$
\begin{aligned}
I(X_1, Z, W_2; W_1) &= I(Z, W_2; W_1) + I(X_1; W_1|Z, W_2) && (90)\\
&= I(X_1; W_1) + \underbrace{I(Z, W_2; W_1|X_1)}_{=0}, && (91)
\end{aligned}
$$

where the last term is zero due to the Markov condition. For the sum rate bound, we can write out as follows:

$$I(X_1, X_2; W_1, W_2|Z) \tag{92}$$

$$= I(X_1; W_1, W_2|Z) + I(X_2; W_1, W_2|Z, X_1) \tag{93}$$

$$= I(X_1; W_2|Z) + I(X_1; W_1|Z, W_2) + \underbrace{I(X_2; W_1|Z, X_1)}_{=0} + I(X_2; W_2|Z, X_1, W_1). \tag{94}$$

To transform the last term, note

$$\begin{aligned} I(X_1, X_2; W_2|Z, W_1) &= I(X_1; W_2|Z, W_1) + I(X_2; W_2|Z, X_1, W_1) &(95)\\ &= I(X_2; W_2|Z, W_1) + \underbrace{I(X_1; W_2|Z, X_2, W_1)}_{=0}. &(96)\end{aligned}$$

Hence, by plugging in, we obtain

$$I(X_1, X_2; W_1, W_2|Z) \tag{97}$$

$$= I(X_1; W_1|Z, W_2) + I(X_2; W_2|Z, W_1) + I(X_1; W_2|Z) - I(X_1; W_2|Z, W_1). \tag{98}$$

Now, the last term can be written as follows:

$$\begin{aligned} I(X_1, W_1; W_2|Z) &= I(X_1; W_2|Z) + \underbrace{I(W_1; W_2|Z, X_1)}_{=0} &(99)\\ &= I(W_1; W_2|Z) + I(X_1; W_2|Z, W_1), &(100)\end{aligned}$$

which establishes the equivalence of the two descriptions of our achievable rate region. □

A.3 Inner Bound To The Rate Region

Proof of Theorem 4. Our proof goes along the lines of the proof of [4, Thm. 6.2]. For an observed source sequence X_1^n, Encoder 1 must provide the decoder with an index, denoted by T_1. The following is true about T_1:

$$
\begin{aligned}
nR_1 &\geq H(T_1) && (101)\\
&\overset{(a)}{\geq} H(T_1|T_2, Z^n) && (102)\\
&\geq H(T_1|T_2, Z^n) - H(T_1|T_2, Z^n, X_1^n, X_2^n) && (103)\\
&\overset{(b)}{=} I(X_1^n X_2^n; T_1|T_2, Z^n) && (104)\\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; T_1|T_2, Z^n, X_{1,1}^{i-1}, X_{2,1}^{i-1}) && (105)\\
&\overset{(b)}{=} \sum_{i=1}^{n} H(X_{1,i}, X_{2,i}|T_2, Z^n, X_{1,1}^{i-1}, X_{2,1}^{i-1}) - H(X_{1,i}, X_{2,i}|T_1, T_2, Z^n, X_{1,1}^{i-1}, X_{2,1}^{i-1}) &&\\
&\overset{(d)}{=} \sum_{i=1}^{n} H(X_{1,i}, X_{2,i}|W_{2,i}, Z_i) - H(X_{1,i}, X_{2,i}|W_{1,i}, W_{2,i}, Z_i) && (106)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; W_{1,i}|W_{2,i}, Z_i). && (107)
\end{aligned}
$$

For (a), recall that further conditioning cannot increase entropy; (b) is the definition of mutual information; and (c) is the chain rule for mutual information [11, Thm. 2.5.2]. For (d), we define $W_{1,i} = (T_1, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Z_1^{i-1}, Z_{i+1}^n)$, and $W_{2,i} = (T_2, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Z_1^{i-1}, Z_{i+1}^n)$. Note that with this definition, it is indeed true that $W_{1,i}$ is conditionally independent of $(X_{2,i}, Z_i)$ for given $X_{1,i}$, and the corresponding is true for $W_{2,i}$.

For the sum rate bound, we find similarly

$$
\begin{aligned}
n(R_1 + R_2) &\geq H(T_1, T_2) && (108)\\
&\overset{(a)}{\geq} H(T_1, T_2|Z^n) && (109)\\
&= I(T_1, T_2; X_1^n, X_2^n|Z^n) && (110)\\
&\overset{(b)}{=} \sum_{i=1}^{n} H(X_{1,i}, X_{2,i}|Z^n, X_1^{i-1}, X_2^{i-1}) - H(X_{1,i}, X_{2,i}|T_1, T_2, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Z^n) &&\\
&\overset{(c)}{=} \sum_{i=1}^{n} H(X_{1,i}, X_{2,i}|Z_i) - H(X_{1,i}, X_{2,i}|T_1, T_2, X_{1,1}^{i-1}, X_{2,1}^{i-1}, Z^n) && (111)\\
&\overset{(d)}{=} \sum_{i=1}^{n} H(X_{1,i}, X_{2,i}|Z_i) - H(X_{1,i}, X_{2,i}|W_{1,i}, W_{2,i}, Z_i) && (112)\\
&= \sum_{i=1}^{n} I(X_{1,i} X_{2,i}; W_{1,i} W_{2,i}|Z_i) && (113)
\end{aligned}
$$

For (a), recall that further conditioning cannot increase entropy; (b) is the chain rule for mutual information; and (c) holds because $(X_{1,i}, X_{2,i})$ is conditionally independent of

$Z_1^{i-1}, Z_{i+1}^n, X_1^{i-1}, X_2^{i-1}$, given Z_i. For (d), we use again the definitions of W_1 and W_2. Note that the sum rate bound can be proved just like in the case of the single-source Wyner-Ziv problem, see e.g. [11, p. 440].

The result now follows from a standard convexity argument that is no different from the one used in the proof of [4, Thm. 6.2], wherefore the details are omitted. $\square$

References

[1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–11, January 1976.

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

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, 1973.

[4] T. Berger, "Multiterminal source coding," *Lectures presented at CISM Summer School on the Information Theory Approach to Communications*, July 1977.

[5] A. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and Control*, vol. 38, pp. 60–80, 1978.

[6] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 43, pp. 1912–1923, Novermber 1997.

[7] A. H. Kaspi and T. Berger, "Rate-distortion for correlated source with partially separated encoders," *IEEE Transactions on Information Theory*, vol. 28, pp. 828–840, November 1982.

[8] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Transactions on Information Theory*, vol. 35, pp. 228–236, March 1989.

[9] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *submitted to 2002 International Workshop on Multimedia Signal Processing*, May 2002. Full manuscript in preparation.

[10] M. Gastpar, G. Kramer, and P. Gupta, "Antenna-clustering strategies for the multiple-relay channel," *to be submitted to IEEE Trans Info Theory*, 2002.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems.* New York: Academic Press, 1981.