

The Rate-Distortion Function for Product of Two Sources with Side-Information at Decoders

Shun Watanabe *Member, IEEE*

Abstract—This paper investigates a lossy source coding problem in which two decoders can access their side-information respectively. The correlated sources are a product of two component correlated sources, and we exclusively investigate the case such that each component is degraded. We show the rate-distortion function for that case, and give the following observations. When the components are degraded in matched order, the rate distortion function of the product sources is equal to the sum of the component-wise rate distortion functions. On the other hand, the former is strictly smaller than the latter when the component sources are degraded in mismatched order. The converse proof for the mismatched case is motivated by the enhancement technique used for broadcast channels. For binary Hamming and Gaussian examples, we evaluate the rate-distortion functions.

Index Terms—Heegard-Berger Problem, Rate-Distortion, Reversely Degraded, Side-Information

I. INTRODUCTION

The source coding problem for correlated sources has been regarded as an important research area in information theory, and various types of coding problems were studied so far (e.g. [1], [2], [3], [4], [5]). In particular, our focus in this paper is the lossy coding problem posed by Heegard and Berger [6].

In the problem, there is one encoder and multiple decoders (see Fig. 1). In this paper, we only treat the case with two decoders. The encoder sends an encoded version of principal source X . The decoder 1 reproduces the principal source within prescribed distortion level by the help of side-information Y , and the decoder 2 reproduces the principal source within prescribed distortion level by the help of side-information Z .

In this setting, Heegard and Berger showed an upper bound on the rate distortion function. They also showed that the upper bound is tight if the side-information is degraded, i.e., X , Y , and Z form a Markov chain in this order. So far, there is no conclusive result, i.e., an upper bound and a lower bound coincide, without the degraded assumption, and whether Heegard and Berger's upper bound is tight or not for non-degraded cases has been a long-standing open problem¹.

Part of this paper was presented at 2011 IEEE International Symposium on Information Theory, Saint Petersburg, Russia.

The author is with the Department of Information Science and Intelligent Systems, University of Tokushima, 2-1, Minami-josanjima, Tokushima 770-8506, Japan, and with the Institute for System Research, University of Maryland, College Park, MD 20742, USA, e-mail:shun-wata@is.tokushima-u.ac.jp.

Manuscript received ; revised

¹Sgarro's problem [7] can be regarded as a lossless special case of Heegard and Berger's problem, and his result [7, Theorem 1] holds without the degraded assumption.

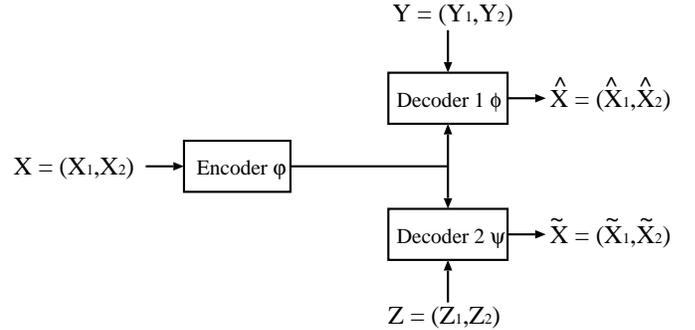


Fig. 1. The coding system investigated in this paper.

In [8], Steinberg and Merhav investigated the successive refinement for the Wyner-Ziv problem, which is a generalization of Heegard and Berger's problem. In [9], Tian and Diggavi investigated the multistage successive refinement for the Wyner-Ziv problem. In these literatures [8], [9], the side-information is assumed to be degraded. In [10], Tian and Diggavi also investigated the side-information scalable source coding, in which the side-information is reversely degraded with respect to the successive refinement. When the refinement layer's rate of the side-information scalable source coding is 0, it is nothing but Heegard and Berger's problem. In such a case, there is no difference between the degraded and the reversely degraded.

In order to provide some insight to Heegard and Berger's problem, we investigate a special case of this problem in this paper. Specifically, we consider the case such that the correlated sources (X, Y, Z) is a cartesian product of two components correlated sources (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) and the components are independent of each other (see Fig. 1). Furthermore, we exclusively consider the case such that each component is degraded, i.e., either

$$\begin{aligned} X_1 &\leftrightarrow Y_1 \leftrightarrow Z_1, \\ X_2 &\leftrightarrow Y_2 \leftrightarrow Z_2 \end{aligned} \quad (1)$$

or

$$\begin{aligned} X_1 &\leftrightarrow Y_1 \leftrightarrow Z_1, \\ X_2 &\leftrightarrow Z_2 \leftrightarrow Y_2 \end{aligned} \quad (2)$$

is satisfied, where $A \leftrightarrow B \leftrightarrow C$ represents that the random variables (A, B, C) form Markov chain in this order.

When (1) is satisfied, the joint sources (X, Y, Z) are degraded. Thus, Heegard and Berger's result suggests that their upper bound is tight. On the otherhand, when (2) is satisfied, the joint sources are not degraded. Thus, whether Heegard and Berger's upper bound is tight or not is unclear so far. In

this paper, we show that the upper bound is tight whenever (2) holds by finding a tight lower bound (a converse), i.e., we characterize the rate-distortion function. To the best of the author's knowledge, this is the first example such that the rate-distortion function is characterized without the degraded assumption².

The problem setting treated in this paper is interesting not only because we can obtain a conclusive result, but it is also interesting by the following reason. Since the component correlated sources in our problem setting are independent of each other, one might think that a combination of the component-wise optimal scheme is optimal in total and the rate-distortion function of our problem setting is just the summation of the component-wise rate distortion functions. However, this is not the case, i.e., the rate distortion function of product sources can be strictly smaller than the summation of the component-wise rate distortion functions even though the components are independent of each other. To explain this fact intuitively, let us consider an example illustrated in Fig. 2. When two components are encoded and decoded separately, 1 bit must be sent for each components, which means 2 bits must be sent to reproduce (X_1, X_2) at both decoders. On the otherhand, if the encoder sends $X_1 \oplus X_2$, then both decoders can reproduce (X_1, X_2) as in the network coding [15]³. Thus, when the components are encoded and decoded jointly, 1 bit suffices for the decoders to reproduce (X_1, X_2) . As we can find from this example, the rate distortion function of product sources is not trivial, and it is interesting to characterize the rate distortion function for our problem setting.

It should be noted that the present work is motivated by the results on product of two broadcast channels by Poltyrev [17] and El Gamal [18]. The broadcast channel [19] is also a long-standing open problem in the network information theory even for two receivers. When there is an ordering between the two receivers (such as degraded, less noisy, and more capable), then conclusive results have been obtained [20], [21], [22], [23], [24]. Poltyrev and El Gamal's conclusive results are few examples without such orderings. The result in this paper can be regarded as a source coding counterpart of Poltyrev and El Gamal's results. However, there is a subtlety of distortions in our problem setting that do not exist in the broadcast channel.

Recently, Weingarten *et al.* solved the capacity region of the MIMO Gaussian broadcast channel [25]. The MIMO Gaussian broadcast channel is not degraded in general. In [25], the authors introduced a technique called enhancement. There

²At the same time as the first version of this paper appeared in the conference, Timo *et al.* also showed the rate-distortion function for some special cases of the lossy complementary delivery problem [11] (see also [12]), which can be also regarded as special cases of non-degraded Heegard and Berger's problem. More specifically, Timo *et al.* solved the lossy complementary delivery problem for the binary symmetric sources with Hamming distortion measures, and general sources with small distortions and Hamming distortion measures. Recently, Timo *et al.* also solved another special case of Heegard and Berger's problem by introducing the conditional less noisy condition, which subsumes the degraded condition [13], [14].

³This example can be also regarded as a special case of the complementary delivery problem. The relationship between the complementary delivery problem and the network coding was pointed out in [16]. A similar example was also investigated in [11, Example 1] (see also [12, Example 1]) as an example of the lossy complementary delivery problem.

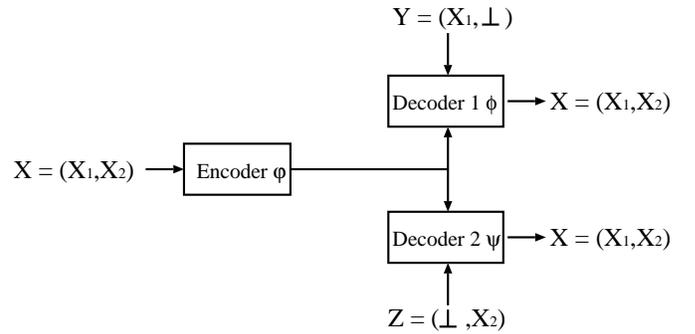


Fig. 2. An intuitive example such that the rate-distortion function (with distortion 0) for the product source is strictly smaller than the summation of the component-wise rate distortion functions. X_1 and X_2 are independent uniform binary random variables, and \perp represent a constant random variable. When two components are encoded and decoded separately, 1 bit must be sent for each components, which means 2 bits must be sent to reproduce (X_1, X_2) at both decoders. On the otherhand, if the encoder sends $X_1 \oplus X_2$, then both decoders can reproduce (X_1, X_2) as in the network coding [15]. Thus, when the components are encoded and decoded jointly, 1 bit suffices for the decoders to reproduce (X_1, X_2) .

are two roles for the enhancement in the converse proof of the MIMO Gaussian broadcast channel. One of them is a reduction of a MIMO non-degraded Gaussian broadcast channel to a MIMO degraded Gaussian broadcast channel. As was pointed out in [5, Section 9.4], Poltyrev's result [17] can be also derived by a straightforward application of the enhancement argument. An application of the enhancement argument to our problem will be also discussed in this paper. Actually, it turns out that a lower bound on the rate-distortion function derived by a straightforward application of the enhancement argument is loose in general.

The rest of the paper is organized as follows. In Section II, we explain the problem setting treated in this paper, and also explain known results. In Section III, we show our main result and its proof. In Section IV, we show the binary Hamming example and the Gaussian example.

II. PRELIMINARIES

In this section, we formally define the problem setup and review Heegard and Berger's results [6].

Let $(X, Y, Z) = ((X_1, X_2), (Y_1, Y_2), (Z_1, Z_2))$ be product of correlated sources, i.e., components (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) are independent of each other. The alphabet of the sources are denoted by $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, and $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2$ respectively, where we assume that these alphabets are finite unless otherwise specified in the Gaussian example. Let (X^n, Y^n, Z^n) be n independent and identically distributed copies of (X, Y, Z) .

Let $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \tilde{\mathcal{X}}_1$, and $\tilde{\mathcal{X}}_2$ be reproduction alphabets, and for $i = 1, 2$ let

$$\hat{d}_i : \mathcal{X}_i \times \hat{\mathcal{X}}_i \rightarrow [0, \infty), \quad (3)$$

$$\tilde{d}_i : \mathcal{X}_i \times \tilde{\mathcal{X}}_i \rightarrow [0, \infty) \quad (4)$$

be distortion measures. Then, let

$$\begin{aligned} \hat{d}_{\text{sum}}(x_1, x_2, \hat{x}_1, \hat{x}_2) &= \hat{d}_1(x_1, \hat{x}_1) + \hat{d}_2(x_2, \hat{x}_2), \\ \tilde{d}_{\text{sum}}(x_1, x_2, \tilde{x}_1, \tilde{x}_2) &= \tilde{d}_1(x_1, \tilde{x}_1) + \tilde{d}_2(x_2, \tilde{x}_2), \end{aligned}$$

be the sum distortion measures.

For blocklength n , the coding system treated in this paper consists of one encoder⁴

$$\varphi : \mathcal{X}^n \rightarrow \{1, \dots, M\}$$

and two decoders

$$\phi : \{1, \dots, M\} \times \mathcal{Y}^n \rightarrow \hat{\mathcal{X}}_1^n \times \hat{\mathcal{X}}_2^n$$

and

$$\psi : \{1, \dots, M\} \times \mathcal{Z}^n \rightarrow \tilde{\mathcal{X}}_1^n \times \tilde{\mathcal{X}}_2^n.$$

For quadruplet $\mathbf{D} = (\hat{D}_1, \hat{D}_2, \tilde{D}_1, \tilde{D}_2)$, rate R is said to be \mathbf{D} -achievable if, for each $\gamma > 0$, there exists a code (φ, ϕ, ψ) with a sufficiently large blocklength n such that

$$\frac{1}{n} \log M \leq R + \gamma$$

and

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[\hat{d}_i(X_{it}, \hat{X}_{it})] \leq \hat{D}_i + \gamma, \quad (5)$$

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[\tilde{d}_i(X_{it}, \tilde{X}_{it})] \leq \tilde{D}_i + \gamma \quad (6)$$

for $i = 1, 2$ are satisfied, where $(\hat{X}_1^n, \hat{X}_2^n) = \phi(\varphi(X^n), Y^n)$ and $(\tilde{X}_1^n, \tilde{X}_2^n) = \psi(\varphi(X^n), Z^n)$. Then, the rate-distortion function is defined as

$$R(\mathbf{D}) := \inf\{R : R \text{ is } \mathbf{D}\text{-achievable}\}.$$

Note that we place the individual distortion constraints in (5) and (6), which are slightly different from those in the original Heegard and Beger's problem [6]. By replacing (5) and (6) with

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[\hat{d}_{\text{sum}}(X_{1t}, X_{2t}, \hat{X}_{1t}, \hat{X}_{2t})] \leq \hat{D} + \gamma,$$

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[\tilde{d}_{\text{sum}}(X_{1t}, X_{2t}, \tilde{X}_{1t}, \tilde{X}_{2t})] \leq \tilde{D} + \gamma$$

respectively, we can define the rate-distortion function $R_{\text{sum}}(\hat{D}, \tilde{D})$ for the sum distortions. Since the sum distortions are special cases of joint distortions, they are special cases of [6].

From the definitions, we obviously have

$$\begin{aligned} & R_{\text{sum}}(\hat{D}, \tilde{D}) \\ & \leq \min\{R(\mathbf{D}) : \hat{D}_1 + \hat{D}_2 \leq \hat{D}, \tilde{D}_1 + \tilde{D}_2 \leq \tilde{D}\}. \end{aligned} \quad (7)$$

When (1) or (2) hold, the opposite inequality can be also proved via the single letter characterization (see Proposition 5 and Theorem 7)⁵.

Remark 1: We can also define the rate-distortion function $R(\hat{D}, \tilde{D})$ for general joint distortions \hat{d} and \tilde{d} . The single letter characterization of $R(\hat{D}, \tilde{D})$ under the condition of (1) can be derived from [6]. However, under the condition of (2),

⁴Since it is obvious from the context, we omit subscript n from the encoder, the decoders, and the message size to simplify the notations.

⁵It is not clear whether the opposite inequality hold or not in general.

the single letter characterization of $R(\hat{D}, \tilde{D})$ is not clear (see Remark 11).

Remark 2: It should be noted that the results in this paper can be easily extended to the weighted sum distortion measures

$$\begin{aligned} \hat{d}_{\text{wsum}}(x_1, x_2, \hat{x}_1, \hat{x}_2) &= \hat{\alpha} \hat{d}_1(x_1, \hat{x}_1) + \hat{\beta} \hat{d}_2(x_2, \hat{x}_2), \\ \tilde{d}_{\text{wsum}}(x_1, x_2, \tilde{x}_1, \tilde{x}_2) &= \tilde{\alpha} \tilde{d}_1(x_1, \tilde{x}_1) + \tilde{\beta} \tilde{d}_2(x_2, \tilde{x}_2), \end{aligned}$$

for some $\hat{\alpha}, \hat{\beta}, \tilde{\alpha}, \tilde{\beta} \geq 0$.

In [6], Heegard and Berger showed an upper bound on the rate-distortion function.

Proposition 3: ([6, Theorem 2]⁶) Let (W, \hat{U}, \tilde{U}) be auxiliary random variables satisfying

- 1) $(W, \hat{U}, \tilde{U}) \leftrightarrow X \leftrightarrow (Y, Z)$.
- 2) There exist functions $\hat{X}'_i(W, \hat{U}, Y)$ and $\tilde{X}'_i(W, \tilde{U}, Z)$ such that $\mathbb{E}[\hat{d}_i(X_i, \hat{X}'_i)] \leq \hat{D}_i$ and $\mathbb{E}[\tilde{d}_i(X_i, \tilde{X}'_i)] \leq \tilde{D}_i$ for $i = 1, 2$.
- 3) $|\mathcal{W}| \leq |\mathcal{X}| + 7$, $|\hat{\mathcal{U}}| \leq |\mathcal{X}| \cdot |\mathcal{W}| + 2$, and $|\tilde{\mathcal{U}}| \leq |\mathcal{X}| \cdot |\mathcal{W}| + 2$, where $\mathcal{W}, \hat{\mathcal{U}}$, and $\tilde{\mathcal{U}}$ are alphabets of W, \hat{U} , and \tilde{U} respectively.

Then, we have

$$\begin{aligned} R(\mathbf{D}) &\leq \max\{I(W; X|Y), I(W; X|Z)\} \\ &\quad + I(\hat{U}; X|Y, W) + I(\tilde{U}; X|Z, W). \end{aligned}$$

Remark 4: In [6], Heegard and Berger also showed an upper bound on the rate-distortion function for more than three decoders. However, Timo *et. al.* pointed out that the statement of [6, Theorem 2] for more than three decoders is invalid, and only the statement for two decoders is valid [26]. In [26], they also showed a corrected upper bound on the rate-distortion function for more than three decoders.

When the component sources are degraded in matched order, i.e.,

$$\begin{aligned} X_1 &\leftrightarrow Y_1 \leftrightarrow Z_1, \\ X_2 &\leftrightarrow Y_2 \leftrightarrow Z_2 \end{aligned} \quad (8)$$

are satisfied, then the joint sources (X, Y, Z) are degraded, i.e.,

$$X \leftrightarrow Y \leftrightarrow Z.$$

For the degraded sources, Heegard and Berger [6] showed that the upper bound in Proposition 3 is tight. In particular for product of two sources, we have the following statement.

Proposition 5: ([6, Theorem 3]) If the components sources are degraded in matched order, i.e., (8) is satisfied, then we have

$$\begin{aligned} R(\mathbf{D}) &= R^*(\mathbf{D}) \\ &:= \min[I(W_1; X_1|Z_1) + I(U_1; X_1|Y_1, W_1) \\ &\quad + I(W_2; X_2|Z_2) + I(U_2; X_2|Y_2, W_2)], \end{aligned}$$

⁶Proposition 3 is a slight modification of [6, Theorem 2] to component distortion functions. The third condition, i.e., the cardinality bound was not stated in [6, Theorem 2], and first shown in [26, Example 2]. Since our problem in this paper places individual distortion constraints on each component of the product source, we have to increase the cardinalities with respect to [26, Example 2].

where the minimization is taken over all auxiliary random variables W_1, W_2, U_1, U_2 satisfying the following:

- 1) $(W_i, U_i) \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ for $i = 1, 2$.
- 2) $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ are independent of each other.
- 3) There exist functions $\hat{X}_i(W_i, U_i, Y_i)$ and $\tilde{X}_i(W_i, Z_i)$ such that $\mathbb{E}[\hat{d}_i(X_i, \hat{X}_i)] \leq \hat{D}_i$ and $\mathbb{E}[\tilde{d}_i(X_i, \tilde{X}_i)] \leq \tilde{D}_i$ for $i = 1, 2$.
- 4) $|\mathcal{W}_i| \leq |\mathcal{X}_i| + 2$ and $|\mathcal{U}_i| \leq (|\mathcal{X}_i| + 1)^2$ for $i = 1, 2$, where \mathcal{W}_i and \mathcal{U}_i are alphabets of W_i and U_i respectively.

Furthermore, we also have

$$R_{\text{sum}}(\hat{D}, \tilde{D}) = \min\{R^*(\mathbf{D}) : \hat{D}_1 + \hat{D}_2 \leq \hat{D}, \tilde{D}_1 + \tilde{D}_2 \leq \tilde{D}\}.$$

Remark 6: Technically, the result in [6, Theorem 3] does not directly imply Proposition 5, because Proposition 5 states the stronger condition on the auxirially random variables, i.e., $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ are independent of each other. We give a proof of Proposition 5 in Appendix A for readers' convenience.

Note that $R^*(\mathbf{D})$ is nothing but the summation of the component-wise rate distortion functions, i.e.,

$$R^*(\mathbf{D}) = R_1^*(\hat{D}_1, \tilde{D}_1) + R_2^*(\hat{D}_2, \tilde{D}_2),$$

where

$$R_i^*(\hat{D}_i, \tilde{D}_i) = \min\{I(W_i; X_i|Z_i) + I(U_i; X_i|Y_i, W_i)\} \quad (9)$$

and the minimization in (9) is taken over all (U_i, W_i) satisfying the conditions 1, 3, and 4 in Proposition 5. This fact implies that the optimal scheme for the degraded product sources is to combine the component-wise optimal scheme.

When sources (X, Y, Z) are not necessarily degraded, whether the upper bound in Proposition 3 is tight or not has been an open problem for a long time. In the next section, we will show that the upper bound is tight if the component sources satisfy (2).

III. MAIN RESULTS

A. Statement of Results

In this section, we consider the case in which the component sources are degraded in mismatched order, i.e.,

$$\begin{aligned} X_1 &\leftrightarrow Y_1 \leftrightarrow Z_1, \\ X_2 &\leftrightarrow Z_2 \leftrightarrow Y_2 \end{aligned} \quad (10)$$

are satisfied. In this case, the joint sources (X, Y, Z) are not degraded, and the rate-distortion function $R(\mathbf{D})$ has not been clarified by any literatures. The following is our main result, which will be proved in Section III-C.

Theorem 7: Suppose that (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) are independent of each other and (10) is satisfied. Then, we have

$$\begin{aligned} R(\mathbf{D}) &= R^\dagger(\mathbf{D}) \\ &:= \min\{\max\{I(W_1; X_1|Y_1) + I(W_2; X_2|Y_2), \\ &\quad I(W_1; X_1|Z_1) + I(W_2; X_2|Z_2)\} \\ &\quad + I(U_1; X_1|Y_1, W_1) + I(U_2; X_2|Z_2, W_2)\}, \end{aligned}$$

where the minimization is taken over all auxiliary random variables W_1, W_2, U_1, U_2 satisfying the following:

- 1) $(W_i, U_i) \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ for $i = 1, 2$.
- 2) $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ are independent of each other.
- 3) There exist functions $\hat{X}_1(W_1, U_1, Y_1)$, $\hat{X}_2(W_2, Y_2)$, $\tilde{X}_1(W_1, Z_1)$, and $\tilde{X}_2(W_2, U_2, Z_2)$ such that

$$\mathbb{E}[\hat{d}_i(X_i, \hat{X}_i)] \leq \hat{D}_i$$

and

$$\mathbb{E}[\tilde{d}_i(X_i, \tilde{X}_i)] \leq \tilde{D}_i$$

for $i = 1, 2$.

- 4) $|\mathcal{W}_i| \leq |\mathcal{X}_i| + 3$ and $|\mathcal{U}_i| \leq |\mathcal{X}_i| \cdot (|\mathcal{X}_i| + 3) + 1$ for $i = 1, 2$, where \mathcal{W}_i and \mathcal{U}_i are alphabets of W_i and U_i respectively.

Furthermore, we also have

$$R_{\text{sum}}(\hat{D}, \tilde{D}) = \min\{R^\dagger(\mathbf{D}) : \hat{D}_1 + \hat{D}_2 \leq \hat{D}, \tilde{D}_1 + \tilde{D}_2 \leq \tilde{D}\}.$$

When the distortion levels are all 0, we have the following corollary, which can be also derived as a straightforward consequence of Sgarro's result [7, Theorem 1].

Corollary 8: When the distortion measures are the Hamming distortion measure and $(\hat{D}_1, \hat{D}_2, \tilde{D}_1, \tilde{D}_2) = \mathbf{0} = (0, 0, 0, 0)$, we have

$$\begin{aligned} R(\mathbf{0}) &= \max\{H(X_1|Y_1) + H(X_2|Y_2), \\ &\quad H(X_1|Z_1) + H(X_2|Z_2)\} \\ &= \max\{H(X_1, X_2|Y_1, Y_2), H(X_1, X_2|Z_1, Z_2)\}. \end{aligned}$$

Remark 9: It should be noted that

$$\begin{aligned} &\max\{H(X_1|Y_1) + H(X_2|Y_2), H(X_1|Z_1) + H(X_2|Z_2)\} \\ &\leq \max\{H(X_1|Y_1), H(X_1|Z_1)\} \\ &\quad + \max\{H(X_2|Y_2), H(X_2|Z_2)\} \end{aligned} \quad (11)$$

$$= H(X_1|Z_1) + H(X_2|Y_2) \quad (12)$$

hold, and the equality in the inequality (11) does not necessarily hold in general, where the equality in (12) follows from (10). Note that (11) is the rate that is needed when we apply Sgarro's coding scheme to each component. This fact implies that the combination of the component-wise optimal scheme is not necessarily optimal even though the components are independent of each other. This phenomenon also appears for lossy cases, which will be exemplified in Section IV.

B. Comparison to Scalable Source Coding

In [10], Tian and Diggavi proposed a coding scheme that is different from [6]. Although joint encoding and decoding is required to achieve the rate-distortion function given in Theorem 7, we can construct a code that achieve the rate-distortion function from component-wise coding scheme of [10] in a similar manner as the example of Fig. 2.

When we apply the coding scheme of [10] to the first component source (X_1, Y_1, Z_1) , the source X_1^n is quantized into the common description W_1^n and the private description

U_1^n . Then, we apply the bin coding to the common description W_1^n at rate

$$I(W_1; X_1|Y_1) = I(W_1; X_1) - I(W_1; Y_1), \quad (13)$$

where the rate $I(W_1; X_1)$ corresponds to the quantization rate and the rate $I(W_1; Y_1)$ corresponds to the reduction of the rate by the bin coding. Note that the equality in (13) requires the Markov condition $(W_1, U_1) \leftrightarrow X_1 \leftrightarrow (Y_1, Z_1)$. Furthermore, we apply the bin coding to W_1^n at extra rate

$$I(W_1; Y_1|Z_1) = I(W_1; Y_1) - I(W_1; Z_1).$$

By using the first bin index I_1 , the first decoder (with Y_1^n) can reconstruct the common description W_1^n . By using both the first bin index I_1 and the extra bin index I_2 , the second decoder (with Z_1^n) can reconstruct W_1^n . After that the private description U_1^n is transmitted to the first decoder at rate

$$I(U_1; X_1|Y_1, W_1).$$

Similarly, when we apply the coding scheme of [10] to the second component source (X_2, Y_2, Z_2) , the source X_2^n is quantized into the common description W_2^n and the private description U_2^n . Then, we apply the bin coding to the common description W_2^n at rates

$$I(W_2; X_2|Z_2) = I(W_2; X_2) - I(W_2; Z_2)$$

and

$$I(W_2; Z_2|Y_2) = I(W_2; Z_2) - I(W_2; Y_2)$$

respectively so that the first decoder (with Y_2^n) can reconstruct W_2^n from both the first bin index J_1 and the second bin index J_2 and the second decoder (with Z_2^n) can reconstruct W_2^n from J_1 . The private description U_2^n is also transmitted to the second decoder at rate

$$I(U_2; X_2|Z_2, W_2).$$

By using the above two component-wise coding scheme, we can construct a joint encoding and decoding scheme as follows. First, the encoder sends $(I_1, J_1, I_2 \oplus J_2)$. This requires the rate

$$I(W_1; X_1|Y_1) + I(W_2; X_2|Z_2) + \max[I(W_1; Y_1|Z_1), I(W_2; Z_2|Y_2)].$$

Note that the first (second) decoder can obtain J_2 (I_2) by first reconstructing W_1^n (W_2^n) and then subtracting I_2 (J_2) from $I_2 \oplus J_2$. The encoder also sends the private descriptions U_1^n and U_2^n at rates $I(U_1; X_1|Y_1, W_1)$ and $I(U_2; X_2|Z_2, W_2)$ respectively. Consequently, the total rate coincides with the rate-distortion function given in Theorem 7.

If we use a straightforward combination of the component-wise coding scheme, I_2 and J_2 will be transmitted separately instead of $I_2 \oplus J_2$, and the rate loss from the joint coding scheme is

$$\min[I(W_1; Y_1|Z_1), I(W_2; Z_2|Y_2)]. \quad (14)$$

C. Proof of Theorem 7

1) *Direct Part:* The direct part is a straightforward consequence of Proposition 3.

For any auxiliary random variables (W_1, W_2, U_1, U_2) satisfying the conditions in Theorem 7, let

$$\begin{aligned} W &= (W_1, W_2), \\ \hat{U} &= U_1, \\ \tilde{U} &= U_2, \\ \hat{X}'_1(W, \hat{U}, Y) &= \hat{X}_1(W_1, U_1, Y_1), \\ \hat{X}'_2(W, \hat{U}, Y) &= \hat{X}_2(W_2, Y_2), \\ \tilde{X}'_1(W, \tilde{U}, Z) &= \tilde{X}_1(W_1, Z_1), \\ \tilde{X}'_2(W, \tilde{U}, Z) &= \tilde{X}_2(W_2, U_2, Z_2). \end{aligned}$$

Then, Proposition 3 implies Theorem 7. The direct part for $R_{\text{sum}}(\hat{D}, \tilde{D})$ follows from (7). ■

2) *Converse Part:* As we have mentioned in Section II, Heegard and Berger showed the converse coding theorem for degraded case. In the course of the proof, they essentially showed the following lemma, which can be shown only for the degraded case. Although our purpose is to show the converse coding theorem for the non-degraded case, we need the following lemma in our converse proof of Theorem 7. A proof of Lemma 10 is given in Appendix B.

Lemma 10: Let

$$(A, B, C) = ((A_1, A_2), (B_1, B_2), (C_1, C_2))$$

be product of correlated sources such that (A_1, B_1, C_1) and (A_2, B_2, C_2) are independent of each other and

$$A_i \leftrightarrow B_i \leftrightarrow C_i \quad (15)$$

for both $i = 1$ and $i = 2$. Let (A^n, B^n, C^n) be n independent identically distributed copies of (A, B, C) . Then, for any (possibly stochastic) function $T_n = f_n(A^n)$, we have

$$\begin{aligned} H(T_n) &\geq \sum_{t=1}^n [I(T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t}|C_{1t}) \\ &\quad + I(B_{1t}^+, B_2^n; A_{1t}|B_{1t}, T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n) \\ &\quad + I(T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+; A_{2t}|C_{2t}) \\ &\quad + I(B_{2t}^+, A_{2t}|B_{2t}, T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+)], \end{aligned}$$

where we use the notations $B_{1t}^- = (B_{11}, \dots, B_{1(t-1)})$, $B_{1t}^+ = (B_{1(t+1)}, \dots, B_{1n})$, and etc.

We now prove the converse part for $R(\mathbf{D})$. Suppose that the rate R is \mathbf{D} -achievable. Then, for any $\gamma > 0$ there exists a code (φ, ϕ, ψ) such that

$$\frac{1}{n} H(S_n) \leq \frac{1}{n} \log M \leq R + \gamma \quad (16)$$

and

$$\frac{1}{n} \sum_{t=1}^n \hat{D}_{it} \leq \hat{D}_i + \gamma, \quad (17)$$

$$\frac{1}{n} \sum_{t=1}^n \tilde{D}_{it} \leq \tilde{D}_i + \gamma \quad (18)$$

for $i = 1, 2$ are satisfied, where we set $S_n = \varphi(X^n)$, $\hat{D}_{it} = \mathbb{E}[\hat{d}_i(X_{it}, \hat{X}_{it}^n)]$ and $\tilde{D}_{it} = \mathbb{E}[\tilde{d}_i(X_{it}, \tilde{X}_{it}^n)]$ for $(\hat{X}_1^n, \hat{X}_2^n) = \phi(\varphi(X^n), Y^n)$ and $(\tilde{X}_1^n, \tilde{X}_2^n) = \psi(\varphi(X^n), Z^n)$.

The key idea of the proof is to derive two lower bounds on $H(S_n)$ by using Lemma 10 as follows. First, let $T_n = S_n$, $(A_1, B_1, C_1) = (X_1, Y_1, Z_1)$ and $(A_2, B_2, C_2) = (X_2, Z_2, Y_2)$. Then, since (A, B, C) satisfies (15), we can use Lemma 10, and we have

$$\begin{aligned}
 & \frac{1}{n}H(S_n) \\
 & \geq \frac{1}{n} \sum_{t=1}^n [I(S_n, Y_{1t}^-, Z_{1t}^-, Z_{1t}^+, Z_2^n; X_{1t}|Z_{1t}) \\
 & \quad + I(Y_{1t}^+; X_{1t}|Y_{1t}, S_n, Y_{1t}^-, Z_{1t}^-, Z_{1t}^+, Z_2^n) \\
 & \quad + I(S_n, Y_1^n, Z_1^n, Z_{2t}^-, Z_{2t}^+; X_{2t}|Z_{2t})] \\
 & = \frac{1}{n} \sum_{t=1}^n [I(S_n, Y_{1t}^-, Y_2^n, Z_{1t}^-, Z_{1t}^+, Z_2^n; X_{1t}|Z_{1t}) \\
 & \quad + I(Y_{1t}^+; X_{1t}|Y_{1t}, S_n, Y_{1t}^-, Y_2^n, Z_{1t}^-, Z_{1t}^+, Z_2^n) \\
 & \quad + I(S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_1^n, Z_{2t}^-, Z_{2t}^+; X_{2t}|Z_{2t})] \quad (19) \\
 & = \frac{1}{n} \sum_{t=1}^n [I(W_{1t}; X_{1t}|Z_{1t}) + I(U_{1t}; X_{1t}|Y_{1t}, W_{1t}) \\
 & \quad + I(W_{2t}, U_{2t}; X_{2t}|Z_{2t})] \\
 & = I(W_{1T}; X_{1T}|Z_{1T}, T) + I(U_{1T}; X_{1T}|Y_{1T}, W_{1T}, T) \\
 & \quad + I(W_{2T}, U_{2T}; X_{2T}|Z_{2T}, T) \\
 & = I(W_{1T}, T; X_{1T}|Z_{1T}) + I(U_{1T}; X_{1T}|Y_{1T}, W_{1T}, T) \\
 & \quad + I(W_{2T}, T, U_{2T}; X_{2T}|Z_{2T}) \quad (20)
 \end{aligned}$$

where we used the fact that Y_2 is degraded version of Z_2 in (19), i.e.,

$$\begin{aligned}
 Y_2^n & \leftrightarrow (S_n, Y_{1t}^-, Z_1^n, Z_2^n) \leftrightarrow X_{1t}, \\
 Y_2^n & \leftrightarrow (S_n, Y_{1t}^-, Y_{1t}, Z_{1t}^-, Z_{1t}^+, Z_2^n) \leftrightarrow X_{1t}, \\
 Y_2^n & \leftrightarrow (S_n, Y_1^n, Z_{1t}^-, Z_{1t}^+, Z_2^n) \leftrightarrow X_{1t},
 \end{aligned}$$

and we set

$$\begin{aligned}
 W_{1t} & = (S_n, Y_{1t}^-, Y_2^n, Z_{1t}^-, Z_{1t}^+, Z_2^n), \\
 U_{1t} & = Y_{1t}^+, \\
 W_{2t} & = (S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_1^n, Z_{2t}^-), \\
 U_{2t} & = Z_{2t}^+,
 \end{aligned}$$

and T is the uniform random numbers on $\{1, \dots, n\}$ that are independent of the other random variables. Note that $W_{1t}, U_{1t}, W_{2t}, U_{2t}$ satisfy $(W_{it}, U_{it}) \leftrightarrow X_{it} \leftrightarrow (Y_{it}, Z_{it})$ for $i = 1, 2$.

Similarly, let $T_n = S_n$, $(A_1, B_1, C_1) = (X_2, Z_2, Y_2)$ and $(A_2, B_2, C_2) = (X_1, Y_1, Y_1)$. Then, since (A, B, C) satisfies

(15), we can use Lemma 10, and we have

$$\begin{aligned}
 & \frac{1}{n}H(S_n) \\
 & \geq \frac{1}{n} \sum_{t=1}^n [I(S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_{2t}^-; X_{2t}|Y_{2t}) \\
 & \quad + I(Z_{2t}^+; X_{2t}|Z_{2t}, S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_{2t}^-) \\
 & \quad + I(S_n, Y_{1t}^-, Y_{1t}^+, Y_2^n, Z_2^n; X_{1t}|Y_{1t})] \\
 & = \frac{1}{n} \sum_{t=1}^n [I(S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_1^n, Z_{2t}^-; X_{2t}|Y_{2t}) \\
 & \quad + I(Z_{2t}^+; X_{2t}|Z_{2t}, S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_1^n, Z_{2t}^-) \\
 & \quad + I(S_n, Y_{1t}^-, Y_{1t}^+, Y_2^n, Z_{1t}^-, Z_{1t}^+, Z_2^n; X_{1t}|Y_{1t})] \quad (21) \\
 & = \frac{1}{n} \sum_{t=1}^n [I(W_{2t}; X_{2t}|Y_{2t}) + I(U_{2t}; X_{2t}|Z_{2t}, W_{2t}) \\
 & \quad + I(W_{1t}, U_{1t}; X_{1t}|Y_{1t})], \\
 & = I(W_{2T}; X_{2T}|Y_{2T}, T) \\
 & \quad + I(U_{2T}; X_{2T}|Z_{2T}, W_{2T}, T) \\
 & \quad + I(W_{1T}, U_{1T}; X_{1T}|Y_{1T}, T) \\
 & = I(W_{2T}, T; X_{2T}|Y_{2T}) \\
 & \quad + I(U_{2T}; X_{2T}|Z_{2T}, W_{2T}, T) \\
 & \quad + I(W_{1T}, T, U_{1T}; X_{1T}|Y_{1T}), \quad (22)
 \end{aligned}$$

where we used the fact that Z_1 is degraded version of Y_1 in (21), i.e.,

$$\begin{aligned}
 Z_1^n & \leftrightarrow (S_n, Y_1^n, Y_2^n, Z_{2t}^-) \leftrightarrow X_{2t}, \\
 Z_1^n & \leftrightarrow (S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_2^n) \leftrightarrow X_{2t}, \\
 Z_1^n & \leftrightarrow (S_n, Y_1^n, Y_{2t}^-, Y_{2t}^+, Z_{2t}^-, Z_{2t}) \leftrightarrow X_{2t}.
 \end{aligned}$$

Since (W_{1t}, U_{1t}, Y_{1t}) and (W_{2t}, Y_{2t}) include (S_n, Y_1^n, Y_2^n) respectively, there exist functions $\hat{X}_{1t}(W_{1t}, U_{1t}, Y_{1t})$ and $\hat{X}_{2t}(W_{2t}, Y_{2t})$ satisfying

$$\begin{aligned}
 \mathbb{E}[\hat{d}_1(X_1, \hat{X}_{1t})] & = \hat{D}_{1t}, \\
 \mathbb{E}[\hat{d}_2(X_2, \hat{X}_{2t})] & = \hat{D}_{2t}
 \end{aligned}$$

respectively. Similarly, since (W_{1t}, Z_{1t}) and (W_{2t}, U_{2t}, Z_{2t}) include (S_n, Z_1^n, Z_2^n) respectively, there exist functions $\tilde{X}_{1t}(W_{1t}, Z_{1t})$ and $\tilde{X}_{2t}(W_{2t}, U_{2t}, Z_{2t})$ satisfying

$$\begin{aligned}
 \mathbb{E}[\tilde{d}_1(X_1, \tilde{X}_{1t})] & = \tilde{D}_{1t}, \\
 \mathbb{E}[\tilde{d}_2(X_2, \tilde{X}_{2t})] & = \tilde{D}_{2t}
 \end{aligned}$$

respectively. Thus, there exist functions

$$(\hat{X}_1(W_{1T}, T, U_{1T}, Y_{1T}), \hat{X}_2(W_{2T}, T, Y_{2T}))$$

for the first decoder and functions

$$(\tilde{X}_1(W_{1T}, T, Z_{1T}), \tilde{X}_2(W_{2T}, T, U_{2T}, Z_{2T}))$$

for the second decoder satisfying

$$\mathbb{E}[\hat{d}(X_i, \hat{X}_i)] \leq \hat{D}_i + \gamma \quad (23)$$

$$\mathbb{E}[\tilde{d}(X_i, \tilde{X}_i)] \leq \tilde{D}_i + \gamma \quad (24)$$

for $i = 1, 2$. Thus, by combining (16), (20), and (22), and by taking $W_1 = (W_{1T}, T)$, $U_1 = U_{1T}$, $W_2 = (W_{2T}, T)$, and

$U_2 = U_{2T}$, we have that there exist W_1, W_2, U_1, U_2 satisfying (23) and (24) and

$$\begin{aligned} R(\mathbf{D}) &\geq I(W_1; X_1|Z_1) + I(U_1; X_1|Y_1, W_1) \\ &\quad + I(W_2, U_2; X_2|Z_2) - \gamma, \\ R(\mathbf{D}) &\geq I(W_1, U_1; X_1|Y_1) \\ &\quad + I(W_2; X_2|Y_2) + I(U_2; X_2|Z_2, W_2) - \gamma. \end{aligned}$$

Although the auxirially random variables $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ chosen above are not necessarily independent of each other, they do not appear together in any one term. Thus we can take $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ to be independent of each other.

By applying the cardinality bound on the auxiliary random variables, which will be proved in Appendix C, we have

$$R(\mathbf{D}) \geq R^\dagger(\mathbf{D} + \gamma \mathbf{1}) - \gamma,$$

where $\mathbf{1} = (1, 1, 1, 1)$. Since $\gamma > 0$ is arbitrary, by the continuity of $R^\dagger(\mathbf{D})$ with respect to \mathbf{D} , we have the converse part for $R(\mathbf{D})^7$. The converse part for $R_{\text{sum}}(\hat{D}, \tilde{D})$ can be proved almost in the same manner. ■

Remark 11: In the above converse proof, we derived the independence between $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ by using the fact that they do not appear together in any term one term. Thus, we cannot derive the independence between them if we employ general joint distortion measures \hat{d} and \tilde{d} . Without this independence, we cannot prove the matching direct part from Proposition 3 because

$$\begin{aligned} &I(W_1, W_2; X_1, X_2|Y_1, Y_2) \\ &= I(W_1; X_1|Y_1) + I(W_2; X_2|Y_2), \\ &I(W_1, W_2; X_1, X_2|Z_1, Z_2) \\ &= I(W_1; X_1|Z_1) + I(W_2; X_2|Z_2) \end{aligned}$$

do not hold in general. For this reason, the single letter characterization of $R(\hat{D}, \tilde{D})$ is not clear.

Remark 12: In the above converse argument, we reduced the proof to the degraded case by setting $(A_1, B_1, C_1) = (X_1, Y_1, Z_1)$ and $(A_2, B_2, C_2) = (X_2, Z_2, Z_2)$, or by setting $(A_1, B_1, C_1) = (X_2, Z_2, Y_2)$ and $(A_2, B_2, C_2) = (X_1, Y_1, Y_1)$. This reduction argument is motivated by the enhancement technique introduced by Weingarten *et. al.* [25], in which the converse proof of the MIMO (not necessarily degraded) broadcast channel was reduced to that of the MIMO degraded broadcast channel. This kind of argument was implicitly used in [18]. As is pointed out in [5, Section 9.4], the result in [17] can be obtained by a straightforward application of the enhancement argument.

It should be noted that the following straightforward application of the enhancement argument gives only loose converse in our problem. Suppose that $((X_1, X_2), (Y_1, Y_2), (Z_1, Z_2))$ satisfies the Markov conditions in (10), and let

$$R(\mathbf{D}|((X_1, X_2), (Y_1, Y_2), (Z_1, Z_2)))$$

⁷Since the cardinalities of the auxiliary random variables are bounded, $R^\dagger(\mathbf{D})$ can be described as a finite dimensional optimization problem and the continuity of $R^\dagger(\mathbf{D})$ with respect to \mathbf{D} follows from the continuity of the mutual information with respect to the test channel.

be the rate-distortion function for this source. Let

$$R(\mathbf{D}|((X_1, X_2), (Y_1, Z_2), (Z_1, Z_2)))$$

and

$$R(\mathbf{D}|((X_1, X_2), (Y_1, Y_2), (Y_1, Z_2)))$$

be the rate-distortion functions for the enhanced sources respectively. Then, we have

$$\begin{aligned} &R(\mathbf{D}|((X_1, X_2), (Y_1, Y_2), (Z_1, Z_2))) \\ &\geq \max\{R(\mathbf{D}|((X_1, X_2), (Y_1, Z_2), (Z_1, Z_2))), \\ &\quad R(\mathbf{D}|((X_1, X_2), (Y_1, Y_2), (Y_1, Z_2)))\}. \end{aligned} \quad (25)$$

As will be exemplified in Section IV-B, this lower bound is loose in general.

IV. EXAMPLES

To illustrate our main result, we consider a binary example and a Gaussian example.

A. Binary Example

In this section, we evaluate the rate distortion function for the binary Hamming example. We first review some known result of the binary Hamming version of the rate-distortion function where the side-information may be absent [6]. This result will be used to investigate the rate-distortion function for product of two binary sources.

Let X be the uniform binary source, and let Y be the output of the binary symmetric channel with crossover probability $p < \frac{1}{2}$, where the input is X . Let Z be a constant, and let d be the Hamming distortion measure. The rate-distortion function of this situation is given by

$$R_b(\hat{D}, \tilde{D}) = \min[I(W; X) + I(U; X|W, Y)], \quad (26)$$

where the minimization is taken over all auxiliary random variables W and U satisfying the following:

- 1) $(W, U) \leftrightarrow X \leftrightarrow Y$.
- 2) There exist functions $\hat{X}(W, U, Y)$ and $\tilde{X}(W)$ such that $\mathbb{E}[d(X, \hat{X})] \leq \hat{D}$ and $\mathbb{E}[d(X, \tilde{X})] \leq \tilde{D}$.
- 3) $|\mathcal{W}| \leq |\mathcal{X}| + 2$ and $|\mathcal{U}| \leq (|\mathcal{X}| + 2)^2$.

An explicit form of $R_b(\hat{D}, \tilde{D})$ was first studied in [6], and a loose upper bound was obtained. After that, Kerpez [27] and Fleming and Effros [28] also studied this problem. Finally, Tian and Diggavi [9] derived an explicit form of $R_b(\hat{D}, \tilde{D})$.

For $0 \leq q \leq 1$, let

$$G_p(q) = h(p * q) - h(q),$$

where $h(\cdot)$ is the binary entropy function and $p * q = p(1-q) + (1-p)q$ is the binary convolution. It was shown in [27] that the rate distortion region can be partitioned into four subregions, three of which are degenerate.

- Region I: $0 \leq \tilde{D} < \frac{1}{2}$ and $0 \leq \hat{D} < \min\{\tilde{D}, p\}$. In this region, $R_b(\hat{D}, \tilde{D})$ is a function of both \hat{D} and \tilde{D} , and it is the only non-degenerate case.

- Region II: $\tilde{D} \geq \frac{1}{2}$ and $0 \leq \hat{D} < p$. In this region, the common description W is not needed, and $R_b(\hat{D}, \tilde{D})$ reduces to the Wyner-Ziv rate-distortion function, i.e.,

$$\begin{aligned} R_b(\hat{D}, \tilde{D}) &= R_p^{WZ}(\hat{D}) \\ &= \min_{(\beta, \theta): 0 \leq \theta \leq 1, 0 \leq \beta \leq p, \theta\beta + (1-\theta)p = \hat{D}} [\theta G(\beta)]. \end{aligned}$$

- Region III: $0 \leq \tilde{D} < \frac{1}{2}$ and $\hat{D} \geq \min\{\tilde{D}, p\}$. In this region, the refinement description U is not needed, and $R_b(\hat{D}, \tilde{D})$ reduces to the ordinary rate-distortion function, i.e.,

$$R_b(\hat{D}, \tilde{D}) = 1 - h(\tilde{D}).$$

- Region IV: $\tilde{D} \geq \frac{1}{2}$ and $\hat{D} \geq p$. In this region, clearly both descriptions W and U can be constant, and $R_b(\hat{D}, \tilde{D}) = 0$.

To describe the rate-distortion function for region I, we need to introduce some notations. For parameters $(D, \alpha, \beta, \theta, \tau)$ satisfying

$$0 \leq D \leq \frac{1}{2}, \quad 0 \leq \alpha, \beta \leq p, \quad 0 \leq \tau \leq \theta \leq 1,$$

we define

$$\begin{aligned} B_p(D, \alpha, \beta, \theta, \tau) &= (\theta - \tau)G_p(\alpha) + \tau G_p(\beta) \\ &\quad + (1 - \theta)G_p(\gamma(D, \alpha, \beta, \theta, \tau)), \end{aligned}$$

where

$$\gamma(D, \alpha, \beta, \theta, \tau) = \begin{cases} \frac{D - (\theta - \tau)(1 - \alpha) - \tau\beta}{1 - \theta} & \theta \neq 1 \\ \frac{1}{2} & \theta = 1 \end{cases}.$$

We also define

$$\begin{aligned} \mathcal{Q}_p(\hat{D}, \tilde{D}) &= \{(\tilde{D}, \alpha, \beta, \theta, \tau) : \\ &(1 - \theta)p \leq \tilde{D} - (\theta - \tau)(1 - \alpha) - \tau\beta \leq (1 - \theta)(1 - p), \\ &0 \leq \tau \leq \theta \leq 1, \\ &0 \leq \alpha, \beta \leq p, \\ &(\theta - \tau)\alpha + \tau\beta + (1 - \theta)p \leq \hat{D}, \\ &\tilde{D} \leq \tilde{D}\}. \end{aligned}$$

For region I, Tian and Diggavi [9] showed⁸

$$\begin{aligned} R_b(\hat{D}, \tilde{D}) &= \min_{(\tilde{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\hat{D}, \tilde{D})} [1 - h(\tilde{D} * p) + B_p(\tilde{D}, \alpha, \beta, \theta, \tau)]. \end{aligned} \quad (27)$$

The righthand side of (26) can be rewritten as

$$\min[I(W; Y) + I(U, W; X|Y)], \quad (28)$$

⁸Tian and Diggavi also showed that the restriction to the equalities $\tilde{D} = \tilde{D}$ and $(\theta - \tau)\alpha + \tau\beta + (1 - \theta)p = \hat{D}$ in the definition of $\mathcal{Q}_p(\hat{D}, \tilde{D})$ does not increase the rate-distortion function. However, in the case of the product of two sources, it is not clear whether such a restriction does not increase the rate-distortion function.

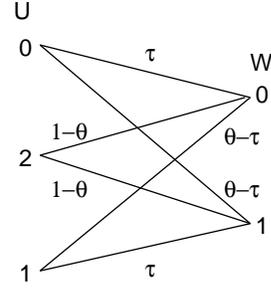


Fig. 3. The test channel between U and W .

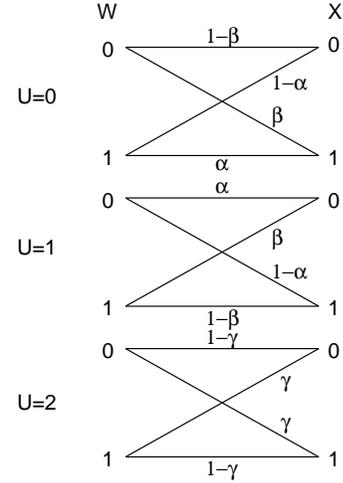


Fig. 4. The test channel between W and X .

and $R_b(\hat{D}, \tilde{D})$ is achieved by reverse test channels described in Figs. 3 and 4. Note that

$$\begin{aligned} &\min_{(\tilde{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\hat{D}, \tilde{D})} [1 - h(\tilde{D} * p) + B_p(\tilde{D}, \alpha, \beta, \theta, \tau)] \\ &= R_p^{WZ}(\hat{D}) \end{aligned} \quad (29)$$

for $\tilde{D} = \frac{1}{2}$ and

$$\begin{aligned} &\min_{(\tilde{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\hat{D}, \tilde{D})} [1 - h(\tilde{D} * p) + B_p(\tilde{D}, \alpha, \beta, \theta, \tau)] \\ &= 1 - h(\tilde{D} * p) + G_p(\tilde{D}) \\ &= 1 - h(\tilde{D}) \end{aligned} \quad (30)$$

for $\hat{D} = \min\{\tilde{D}, p\}$ and $\tilde{D} \leq \frac{1}{2}$, which will be proved in Appendix D. Thus, we can also write

$$\begin{aligned} R_b(\hat{D}, \tilde{D}) &= \min_{(\tilde{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\hat{D}, \tilde{D})} [1 - h(\tilde{D} * p) + B_p(\tilde{D}, \alpha, \beta, \theta, \tau)] \end{aligned}$$

for any (\hat{D}, \tilde{D}) .

Now, we consider the rate-distortion function for product of two binary sources. Let X_1 and X_2 be the independent uniform binary sources. Let Y_1 be the output of the binary symmetric channel with crossover probability $p_1 < \frac{1}{2}$, where the input is X_1 . Let Z_2 be the outputs of the binary symmetric channel with crossover probability $p_2 < \frac{1}{2}$, where the input is X_2 . Then, let Y_2 and Z_1 be constant. Obviously, this pair

of correlated sources satisfy the Markov conditions in (10). In this case, we have the following.

Theorem 13: For any $\mathbf{D} = (\hat{D}_1, \tilde{D}_1, \hat{D}_2, \tilde{D}_2)$, we have

$$\begin{aligned} R(\mathbf{D}) &= \min[\max\{1 - h(\tilde{D}_1 * p_1), 1 - h(\tilde{D}_2 * p_2)\} \\ &\quad + B_{p_1}(\tilde{D}_1, \alpha_1, \beta_1, \theta_1, \tau_1) \\ &\quad + B_{p_2}(\tilde{D}_2, \alpha_2, \beta_2, \theta_2, \tau_2)], \end{aligned} \quad (31)$$

where the minimizations are taken over

$$(\tilde{D}_1, \alpha_1, \beta_1, \theta_1, \tau_1) \in \mathcal{Q}_{p_1}(\hat{D}_1, \tilde{D}_1)$$

and

$$(\tilde{D}_2, \alpha_2, \beta_2, \theta_2, \tau_2) \in \mathcal{Q}_{p_2}(\hat{D}_2, \tilde{D}_2)$$

respectively.

Proof: See Appendix E. \blacksquare

In the following, for a symmetric case, we compare the rate-distortion function, the upper bound derived by the component-wise scheme, and the lower bound derived by the straightforward enhancement. Let $p_1 = p_2 = p < \frac{1}{2}$. Let $d_c(p)$ the critical distortion [2], i.e., the distortion satisfying

$$\frac{G_p(d_c(p))}{d_c(p) - p} = G'_p(d_c(p)).$$

Let $\tilde{D}_1 = \hat{D}_2 = \bar{D}$, where $d_c(p) < \bar{D} < \frac{1}{2}$. Let $\hat{D}_1 = \tilde{D}_2 = D$. From (31), it is clear that the summation of the component-wise rate-distortion functions is

$$2 \min_{(\tilde{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(D, \bar{D})} [1 - h(\tilde{D} * p) + B_p(\tilde{D}, \alpha, \beta, \theta, \tau)],$$

which is strictly larger than the joint rate-distortion function obtained from Theorem 13.

Suppose that $D \geq \bar{D}$. In this case, in a similar manner as (30), we can show that the joint rate-distortion function is

$$R(\mathbf{D}) = 1 - h(\bar{D} * p) + 2G_p(\bar{D}).$$

On the other hand, from Proposition 5, the rate-distortion function of the source satisfying (8) is the summation of the component-wise rate-distortion functions. Thus, the lower bound in (25) is given by

$$1 - h(\bar{D} * p) + G_p(\bar{D}) + R_p^{WZ}(\bar{D}).$$

Since $R_p^{WZ}(\bar{D}) < G_p(\bar{D})$ for $d_c(p) < \bar{D} < \frac{1}{2}$, the lower bound in (25) is loose.

Suppose that $D \leq d_c(p)$. In this case, in a similar manner as [9, Corollary 2], we can show that the joint rate-distortion function is

$$1 - h(\bar{D} * p) + 2G_p(D). \quad (32)$$

The lower bound in (25) coincide with (32) in this case, and thus tight.

B. Gaussian Example

In this section, we evaluate the rate distortion function for the Gaussian example. We consider jointly Gaussian sources (X_i, Y_i, Z_i) given by $Y_i = X_i + N_{i,y}$ and $Z_i = X_i + N_{i,z}$, where $N_{i,y}$ and $N_{i,z}$ are Gaussian noises with variances Σ_{i,N_y} and Σ_{i,N_z} such that $\Sigma_{1,N_y} < \Sigma_{1,N_z}$ and $\Sigma_{2,N_z} < \Sigma_{2,N_y}$ respectively. The conditional variance of X_i given Y_i is denoted by $\Sigma_{i,x|y}$ etc.. To avoid tedious degenerate cases, we assume that $\hat{D}_i < \Sigma_{i,x|y}$ and $\tilde{D}_i < \Sigma_{i,x|z}$ for $i = 1, 2$.

In the above setting, the rate-distortion function is given by the following theorem. The theorem can be proved by first showing that Gaussian auxiliary random variables suffice, and then by elementary calculation.

Theorem 14: We have

$$\begin{aligned} R(\mathbf{D}) &= \max \left[\frac{1}{2} \log \frac{\Sigma_{1,x|y}}{(\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1})^{-1}} \right. \\ &\quad \left. + \frac{1}{2} \log \frac{\Sigma_{2,x|y}}{\hat{D}_2}, \right. \\ &\quad \left. \frac{1}{2} \log \frac{\Sigma_{1,x|z}}{\tilde{D}_1} + \frac{1}{2} \log \frac{\Sigma_{2,x|z}}{(\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1})^{-1}} \right] \\ &\quad + \frac{1}{2} \log \frac{(\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1})^{-1}}{(B_1^* + \Sigma_{1,N_y}^{-1})^{-1}} \\ &\quad + \frac{1}{2} \log \frac{(\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1})^{-1}}{(B_2^* + \Sigma_{2,N_z}^{-1})^{-1}}, \end{aligned}$$

where

$$\begin{aligned} B_1^* &= \max[\hat{D}_1^{-1} - \Sigma_{1,N_z}^{-1}, \hat{D}_1^{-1} - \Sigma_{1,N_y}^{-1}], \\ B_2^* &= \max[\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1}, \hat{D}_2^{-1} - \Sigma_{2,N_z}^{-1}]. \end{aligned}$$

Note that the component-wise rate-distortion functions are given by

$$\begin{aligned} R_1^*(\hat{D}_1, \tilde{D}_1) &= \frac{1}{2} \log \frac{\Sigma_{1,x|z}}{\tilde{D}_1} \\ &\quad + \frac{1}{2} \log \frac{(\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1})^{-1}}{(B_1^* + \Sigma_{1,N_y}^{-1})^{-1}}, \\ R_2^*(\hat{D}_2, \tilde{D}_2) &= \frac{1}{2} \log \frac{\Sigma_{2,x|y}}{\hat{D}_2} \\ &\quad + \frac{1}{2} \log \frac{(\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1})^{-1}}{(B_2^* + \Sigma_{2,N_z}^{-1})^{-1}}. \end{aligned}$$

By noting $\Sigma_{1,N_y} < \Sigma_{1,N_z}$ and $\Sigma_{2,N_y} > \Sigma_{2,N_z}$, we have

$$\begin{aligned} \frac{\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1}}{\Sigma_{1,x|y}^{-1}} &= \frac{\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1}}{\Sigma_{1,x|z}^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1}} \\ &< \frac{\tilde{D}_1^{-1}}{\Sigma_{1,x|z}^{-1}} \end{aligned}$$

and

$$\begin{aligned} \frac{\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1}}{\Sigma_{2,x|z}^{-1}} &= \frac{\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1}}{\Sigma_{2,x|y}^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1}} \\ &< \frac{\hat{D}_2^{-1}}{\Sigma_{2,x|y}^{-1}}. \end{aligned}$$

Thus, we have

$$R(\mathbf{D}) < R_1^*(\hat{D}_1, \tilde{D}_1) + R_2^*(\hat{D}_2, \tilde{D}_2),$$

which implies that the combination of the component-wise optimal scheme is suboptimal for Gaussian product sources.

Next, we consider the lower bound in (25). Let

$$\begin{aligned} D_{1,\min} &= \min\{\hat{D}_1, \tilde{D}_1\}, \\ D_{2,\min} &= \min\{\hat{D}_2, \tilde{D}_2\}. \end{aligned}$$

Then, using the same notations as in Remark 12, we have

$$\begin{aligned} R(\mathbf{D} | ((X_1, X_2), (Y_1, Z_2), (Z_1, Z_2))) \\ &= \frac{1}{2} \log \frac{\Sigma_{1,x|z}}{\hat{D}_1} + \frac{1}{2} \log \frac{(\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1})^{-1}}{(B_1^* + \Sigma_{1,N_y}^{-1})^{-1}} \\ &\quad + \frac{1}{2} \log \frac{\Sigma_{2,x|z}}{D_{2,\min}} \end{aligned}$$

and

$$\begin{aligned} R(\mathbf{D} | ((X_1, X_2), (Y_1, Y_2), (Y_1, Z_2))) \\ &= \frac{1}{2} \log \frac{\Sigma_{1,x|y}}{D_{1,\min}} \\ &\quad + \frac{1}{2} \log \frac{\Sigma_{2,x|y}}{\hat{D}_2} + \frac{1}{2} \log \frac{(\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1})^{-1}}{(B_2^* + \Sigma_{2,N_z}^{-1})^{-1}}. \end{aligned}$$

Thus, if

$$\hat{D}_1 \leq (\tilde{D}_1^{-1} - \Sigma_{1,N_z}^{-1} + \Sigma_{1,N_y}^{-1})^{-1}, \quad (33)$$

$$\tilde{D}_2 \leq (\hat{D}_2^{-1} - \Sigma_{2,N_y}^{-1} + \Sigma_{2,N_z}^{-1})^{-1}, \quad (34)$$

then we have

$$\begin{aligned} B_1^* + \Sigma_{1,N_y}^{-1} &= D_{1,\min}^{-1} = \hat{D}_1^{-1}, \\ B_2^* + \Sigma_{2,N_z}^{-1} &= D_{2,\min}^{-1} = \tilde{D}_2^{-1}, \end{aligned}$$

and the lower bound in (25) is tight. However, if (33) or (34) are not satisfied, then the lower bound in (25) is not necessarily tight.

In the following, for a symmetric case, we compare the rate-distortion function, the upper bound derived by the component-wise scheme, and the lower bound derived by the straightforward enhancement, i.e., the lower bound in (25). We set $\Sigma_{X_1} = \Sigma_{X_2} = \Sigma_X$, $\Sigma_{1,N_y} = \Sigma_{2,N_z} = \Sigma_N$, $\Sigma_{1,N_z} = \Sigma_{2,N_y} = \Sigma_{\bar{N}}$, $\hat{D}_1 = \hat{D}_2 = D$, and $\tilde{D}_1 = \tilde{D}_2 = \bar{D}$, where $\Sigma_N < \Sigma_{\bar{N}}$. In this case, we have

$$\begin{aligned} \Sigma_{1,x|y} &= \Sigma_{2,x|z} =: \Sigma_{x|s}, \\ \Sigma_{1,x|z} &= \Sigma_{2,x|y} =: \bar{\Sigma}_{x|s}, \end{aligned}$$

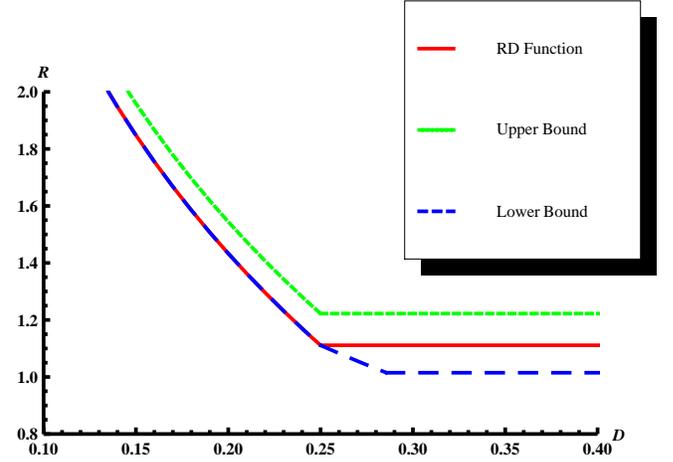


Fig. 5. The red solid curve is the rate-distortion function. The green dotted curve is the upper bound derived by the component-wise scheme. The blue dashed curve is the lower bound derived by the straightforward application of the enhancement.

and

$$\begin{aligned} R(\mathbf{D}) &= \frac{1}{2} \log \frac{\Sigma_{x|s}}{(\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1} + \Sigma_N^{-1})^{-1}} + \frac{1}{2} \log \frac{\bar{\Sigma}_{x|s}}{\bar{D}} \\ &\quad + \log \frac{(\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1} + \Sigma_N^{-1})^{-1}}{(B^* + \Sigma_N^{-1})^{-1}}, \end{aligned} \quad (35)$$

where

$$B^* = \max[\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1}, D^{-1} - \Sigma_N^{-1}].$$

We also have

$$\begin{aligned} R_1^*(D, \bar{D}) + R_2^*(\bar{D}, D) \\ &= \log \frac{\bar{\Sigma}_{x|s}}{\bar{D}} + \log \frac{(\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1} + \Sigma_N^{-1})^{-1}}{(B^* + \Sigma_N^{-1})^{-1}}. \end{aligned} \quad (36)$$

The lower bound in (25) is given by

$$\begin{aligned} \frac{1}{2} \log \frac{\bar{\Sigma}_{x|s}}{\bar{D}} + \frac{1}{2} \log \frac{(\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1} + \Sigma_N^{-1})^{-1}}{(B^* + \Sigma_N^{-1})^{-1}} \\ + \frac{1}{2} \log \frac{\Sigma_{x|s}}{D_{\min}}, \end{aligned} \quad (37)$$

where $D_{\min} := \min\{D, \bar{D}\}$.

The distortion such that (33) and (34) hold with equality is given by

$$D^* := (\bar{D}^{-1} - \Sigma_{\bar{N}}^{-1} + \Sigma_N^{-1})^{-1}.$$

For fixed \bar{D} , the rate-distortion function, the upper bound, and the lower bound are functions of D . From (35) and (36), we can find that the rate-distortion function and the upper bound are constant for $D \geq D^*$. On the other hand, from (37), we can find that the lower bound is constant for $D \geq \bar{D}$. For $\Sigma_X = 1$, $\Sigma_N = 1$, $\Sigma_{\bar{N}} = 2$, and $\bar{D} = \frac{2}{7}$, we plot the rate-distortion function, the upper bound, and the lower bound in Fig. 5. In this case, note that $D^* = \frac{1}{4}$. We can find that the upper bound is loose for every D , and that the lower bound is loose for $D > D^*$.

V. CONCLUSION

In this paper, we investigated the lossy coding problem for a product of two sources with two decoders, and characterized the rate-distortion function.

It is important to extend our result to the case in which there exists correlation between component sources. One of such examples is vector Gaussian sources. As was mentioned in Remark 12, the converse proof in this paper is motivated by the enhancement argument introduced by Weingarten *et al.* [25]. However, as we have exemplified in Section IV, the bound derived by the straightforward application of the enhancement argument is loose in general. Thus, some ingenious way of enhancement might be needed to solve the vector Gaussian Heegard and Berger problem. This topic will be investigated in elsewhere.

ACKNOWLEDGMENT

The author would like to thank Prof. Tomohiko Uyematsu and Prof. Hiroshige Yamamoto for valuable comments. He also would like to thank Prof. Yasutada Oohama for valuable discussions. He also appreciate the comments from anonymous reviewers of ISIT2011, especially for pointing out the result in Section III-B. The comments from anonymous reviewers of this journal are also appreciated, which substantially improved the clarity of this paper.

APPENDIX

A. Proof of Proposition 5

Since the direct part directly follows from Proposition 3, we only prove the converse part. We proved the converse part for $R(\mathbf{D})$. Suppose that R is \mathbf{D} -achievable. Then, for any $\gamma > 0$, there exists a code (φ, ϕ, ψ) satisfying (16)–(18), where we use the same notation as in Section III-C2. We will lower bound $H(S_n)$ by using Lemma 10. Let $T_n = S_n$, $(A_1, B_1, C_1) = (X_1, Y_1, Z_1)$ and $(A_2, B_2, C_2) = (X_2, Y_2, Z_2)$. Then, from Lemma 10, we have

$$\begin{aligned}
 & \frac{1}{n}H(S_n) \\
 & \geq \frac{1}{n} \sum_{t=1}^n [I(S_n, Y_{1t}^-, Z_{1t}^-, Z_{1t}^+, Z_2^n; X_{1t}|Z_{1t}) \\
 & \quad + I(Y_{1t}^+, Y_2^n; X_{1t}|Y_{1t}, S_n, Y_{1t}^-, Z_{1t}^-, Z_{1t}^+, Z_2^n) \\
 & \quad + I(S_n, Y_1^n, Y_{2t}^-, Z_1^n, Z_{2t}^-, Z_{2t}^+, X_{2t}|Z_{2t}) \\
 & \quad + I(Y_{2t}^+; X_{2t}|Y_{2t}, S_n, Y_1^n, Y_{2t}^-, Z_1^n, Z_{2t}^-, Z_{2t}^+)] \\
 & = \frac{1}{n} \sum_{t=1}^n [I(W_{1t}; X_{1t}|Z_{1t}) + I(U_{1t}; X_{1t}|Y_{1t}, W_{1t}) \\
 & \quad + I(W_{2t}; X_{2t}|Z_{2t}) + I(U_{2t}; X_{2t}|Y_{2t}, W_{2t})] \\
 & = I(W_{1T}; X_{1T}|Z_{1T}, T) + I(U_{1T}; X_{1T}|Y_{1T}, W_{1T}, T) \\
 & \quad + I(W_{2T}; X_{2T}|Z_{2T}, T) + I(U_{2T}; X_{2T}|Y_{2T}, W_{2T}, T) \\
 & = I(W_{1T}, T; X_{1T}|Z_{1T}) + I(U_{1T}; X_{1T}|Y_{1T}, W_{1T}, T) \\
 & \quad + I(W_{2T}, T; X_{2T}|Z_{1T}) + I(U_{2T}; X_{2T}|Y_{2T}, W_{2T}, T),
 \end{aligned}$$

where we set

$$\begin{aligned}
 W_{1t} &= (S_n, Y_{1t}^-, Z_{1t}^-, Z_{1t}^+, Z_2^n), \\
 U_{1t} &= (Y_{1t}^+, Y_2^n), \\
 W_{2t} &= (S_n, Y_1^n, Y_{2t}^-, Z_1^n, Z_{2t}^-, Z_{2t}^+), \\
 U_{2t} &= Y_{2t}^+,
 \end{aligned}$$

and T is the uniform random number on $\{1, \dots, n\}$ that are independent of the other random variables. Note that $W_{1t}, U_{1t}, W_{2t}, U_{2t}$ satisfy $(W_{it}, U_{it}) \leftrightarrow X_{it} \leftrightarrow (Y_{it}, Z_{it})$ for $i = 1, 2$.

In a similar reason as in Section III-C2, there exist functions $\hat{X}_i(W_{iT}, T, U_{iT}, Y_{iT})$ and $\tilde{X}_i(W_{iT}, T, Z_{iT})$ satisfying (23) and (24) for $i = 1, 2$. Thus, by taking $W_1 = (W_{1T}, T)$, $U_1 = U_{1T}$, $W_2 = (W_{2T}, T)$, and $U_2 = U_{2T}$, we have that there exist W_1, W_2, U_1, U_2 satisfying (23) and (24) and

$$\begin{aligned}
 R(\mathbf{D}) &\geq I(W_1; X_1|Z_1) + I(U_1; X_1|Y_1, W_1) \\
 &\quad + I(W_2; X_2|Z_2) + I(U_2; X_2|Y_2, W_2).
 \end{aligned}$$

Although the auxirially random variables $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ chosen above are not necessarily independent of each other, they never appear in any term simultaneously. Thus we can take $(W_1, U_1, X_1, Y_1, Z_1)$ and $(W_2, U_2, X_2, Y_2, Z_2)$ to be independent of each other. By using the support lemma [4], we have the statement on the cardinalities of the auxiliary alphabets. Since $\gamma > 0$ is arbitrary, by the continuity of $R^*(\mathbf{D})$ with respect to \mathbf{D} , we have the converse part for $R(\mathbf{D})$. The converse part for $R_{\text{sum}}(\hat{D}, \tilde{D})$ can be proved almost in the same manner. ■

B. Proof of Lemma 10

The lemma is proved in a similar manner as Heegard and Berger's converse argument. Our strategy is to regard (A^n, B^n, C^n) as correlated sources of block length $2n$. Then, we use Heegard and Berger's converse argument to the independently but not identical distributed sources of length $2n$.

First, by chain rules, we have

$$\begin{aligned}
 & H(T_n) \\
 & \geq I(T_n; A^n|C^n) \\
 & = I(T_n, B^n; A^n|C^n) - I(B^n; A^n|T_n, C^n) \\
 & = \sum_{t=1}^n [I(T_n, B_1^n, B_2^n; A_{1t}|A_{1t}^-, C_1^n, C_2^n) \\
 & \quad - I(B_{1t}; A_1^n, A_2^n|T_n, B_{1t}^-, C_1^n, C_2^n) \\
 & \quad + I(T_n, B_1^n, B_2^n; A_{2t}|A_{1t}^-, A_{2t}^-, C_1^n, C_2^n) \\
 & \quad - I(B_{2t}; A_1^n, A_2^n|T_n, B_1^-, B_{2t}^-, C_1^n, C_2^n)].
 \end{aligned}$$

Since (A_{1t}, C_{1t}) and $(A_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n)$ are independent, we have

$$\begin{aligned}
 & I(T_n, B_1^n, B_2^n; A_{1t}|A_{1t}^-, C_1^n, C_2^n) \\
 & = I(T_n, A_{1t}^-, B_1^n, B_2^n, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t}|C_{1t}) \\
 & \geq I(T_n, B_1^n, B_2^n, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t}|C_{1t}).
 \end{aligned}$$

Similarly, since (A_{2t}, C_{2t}) and $(A_1^n, A_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+)$ are independent, we have

$$\begin{aligned} & I(T_n, B_1^n, B_2^n; A_{2t} | A_1^n, A_{2t}^-, C_1^n, C_2^n) \\ & \geq I(T_n, B_1^n, B_2^n, C_1^-, C_{2t}^+, C_{2t}^+; A_{2t} | C_{2t}). \end{aligned}$$

Furthermore, since the Markov chains

$$B_{1t} \leftrightarrow (A_{1t}, C_{1t}) \leftrightarrow (T_n, A_{1t}^-, A_{1t}^+, A_2^-, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n)$$

and

$$B_{2t} \leftrightarrow (A_{2t}, C_{2t}) \leftrightarrow (T_n, A_1^n, A_{2t}^-, A_{2t}^+, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+)$$

hold, we have

$$\begin{aligned} & I(B_{1t}; A_1^n, A_2^n | T_n, B_{1t}^-, C_1^n, C_2^n) \\ & = I(B_{1t}; A_{1t} | T_n, B_{1t}^-, C_1^n, C_2^n) \end{aligned}$$

and

$$\begin{aligned} & I(B_{2t}; A_1^n, A_2^n | T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n) \\ & = I(B_{2t}; A_{2t} | T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n). \end{aligned}$$

Thus, we have

$$\begin{aligned} & H(T_n) \\ & \geq \sum_{t=1}^n [I(T_n, B_1^n, B_2^n, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t} | C_{1t}) \\ & \quad - I(B_{1t}; A_{1t} | T_n, B_{1t}^-, C_1^n, C_2^n) \\ & \quad + I(T_n, B_1^n, B_2^n, C_1^-, C_{2t}^+, C_{2t}^+; A_{2t} | C_{2t}) \\ & \quad - I(B_{2t}; A_{2t} | T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n)]. \end{aligned} \quad (38)$$

By chain rules, we have

$$\begin{aligned} & I(T_n, B_1^n, B_2^n, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t} | C_{1t}) \\ & \quad - I(B_{1t}; A_{1t} | T_n, B_{1t}^-, C_1^n, C_2^n) \\ & = I(T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t} | C_{1t}) \\ & \quad + I(B_{1t}; A_{1t} | T_n, B_{1t}^-, C_1^n, C_2^n) \\ & \quad + I(B_{1t}^+, B_2^n; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_1^n, C_2^n) \\ & \quad - I(B_{1t}; A_{1t} | T_n, B_{1t}^-, C_1^n, C_2^n) \\ & = I(T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t} | C_{1t}) \\ & \quad + I(B_{1t}^+, B_2^n; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_1^n, C_2^n). \end{aligned} \quad (39)$$

Similarly, we have

$$\begin{aligned} & I(T_n, B_1^n, B_2^n, C_1^n, C_{2t}^-, C_{2t}^+; A_{2t} | C_{2t}) \\ & \quad - I(B_{2t}; A_{2t} | T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n) \\ & = I(T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+; A_{2t} | C_{2t}) \\ & \quad + I(B_{2t}^+, A_{2t} | B_{2t}, T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n). \end{aligned} \quad (40)$$

From (8), we have

$$C_{1t} \leftrightarrow B_{1t} \leftrightarrow (T_n, A_{1t}, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n).$$

Thus, we have

$$\begin{aligned} & I(B_{1t}^+; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_1^n, C_2^n) \\ & = I(B_{1t}^+, C_{1t}; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n) \\ & \geq I(B_{1t}^+; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n). \end{aligned} \quad (41)$$

Similarly, from (8), we have

$$\begin{aligned} & I(B_{2t}^+; A_{2t} | B_{2t}, T_n, B_1^n, B_{2t}^-, C_1^n, C_2^n) \\ & \geq I(B_{2t}^+; A_{2t} | B_{2t}, T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+). \end{aligned} \quad (42)$$

Finally, by substituting (39)–(42) into (38), we have

$$\begin{aligned} H(T_n) & \geq \sum_{t=1}^n [I(T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n; A_{1t} | C_{1t}) \\ & \quad + I(B_{1t}^+, B_2^n; A_{1t} | B_{1t}, T_n, B_{1t}^-, C_{1t}^-, C_{1t}^+, C_2^n) \\ & \quad + I(T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+; A_{2t} | C_{2t}) \\ & \quad + I(B_{2t}^+; A_{2t} | B_{2t}, T_n, B_1^n, B_{2t}^-, C_1^n, C_{2t}^-, C_{2t}^+)]. \end{aligned}$$

■

C. Proof of Cardinality Bounds

We prove the cardinality bounds by using the support lemma [29], [5]. We prove by two steps. In the first step, we reduce the cardinality of W_1 and W_2 . We consider $|\mathcal{X}_1| + 3$ continuous functions of $P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)$ as follows:

$$f_{0, x_1}(P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)) := \sum_{u_1} P_{U_1 X_1 | W_1}(u_1, x_1 | w_1),$$

for $x_1 = 1, \dots, |\mathcal{X}_1| - 1$ and

$$\begin{aligned} f_{I, Y}(P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)) & := H(X_1 | Y_1) \\ & \quad - H(X_1 | Y_1, W_1 = w_1) + I(U_1; X_1 | Y_1, W_1 = w_1), \\ f_{I, Z}(P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)) & := H(X_1 | Z_1) \\ & \quad - H(X_1 | Z_1, W_1 = w_1) + I(U_1; X_1 | Y_1, W_1 = w_1), \\ f_{\hat{d}}(P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)) & := \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(w_1, U_1, Y_1)) \right], \\ f_{\tilde{d}}(P_{U_1 X_1 | W_1}(\cdot, \cdot | w_1)) & := \mathbb{E} \left[\tilde{d}_1(X_1, \tilde{X}_1(w_1, Z_1)) \right]. \end{aligned}$$

By using the support lemma to these functions, there exists random variable W'_1 with cardinality $|\mathcal{W}'_1| \leq |\mathcal{X}_1| + 3$ and the corresponding random variable U'_1 , i.e., $P_{U'_1 W'_1 X_1}(u_1, w_1, x_1) = P_{W'_1}(w_1) P_{U_1 X_1 | W_1}(u_1, x_1 | w_1)$, such that the marginal P_{X_1} is preserved and

$$\begin{aligned} & I(W'_1; X_1 | Y_1) + I(U'_1; X_1 | Y_1, W'_1) \\ & = I(W_1; X_1 | Y_1) + I(U_1; X_1 | Y_1, W_1) \\ & I(W'_1; X_1 | Z_1) + I(U'_1; X_1 | Y_1, W'_1) \\ & = I(W_1; X_1 | Z_1) + I(U_1; X_1 | Y_1, W_1), \\ & \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(W'_1, U'_1, Y_1)) \right] \\ & = \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(W_1, U_1, Y_1)) \right], \\ & \mathbb{E} \left[\tilde{d}_1(X_1, \tilde{X}_1(W'_1, Z_1)) \right] \\ & = \mathbb{E} \left[\tilde{d}_1(X_1, \tilde{X}_1(W_1, Z_1)) \right]. \end{aligned}$$

Similarly, there exists W'_2 with cardinality $|\mathcal{W}'_2| \leq |\mathcal{X}_2| + 3$ and the corresponding random variable U'_2 such that P_{X_2} is

preserved and

$$\begin{aligned}
 & I(W'_2; X_2|Y_2) + I(U'_2; X_2|Z_2, W'_2) \\
 &= I(W_2; X_2|Y_2) + I(U_2; X_2|Z_2, W_2) \\
 & I(W'_2; X_2|Z_2) + I(U'_2; X_2|Z_2, W'_2) \\
 &= I(W_2; X_2|Z_2) + I(U_2; X_2|Z_2, W_2), \\
 & \mathbb{E} \left[\hat{d}_2(X_2, \hat{X}_2(W'_2, Y_2)) \right] \\
 &= \mathbb{E} \left[\hat{d}_2(X_2, \hat{X}_2(W_2, Y_2)) \right], \\
 & \mathbb{E} \left[\tilde{d}_2(X_2, \tilde{X}_2(W'_2, U'_2, Z_2)) \right] \\
 &= \mathbb{E} \left[\tilde{d}_2(X_2, \tilde{X}_2(W_2, U_2, Z_2)) \right].
 \end{aligned}$$

In the next step, we reduce the cardinality of U'_1 and U'_2 . We consider $|\mathcal{W}'_1| \cdot |\mathcal{X}_1| + 1$ continuous functions of $P_{W'_1 X_1 | U'_1}(\cdot, \cdot | u_1)$ as follows:

$$\begin{aligned}
 & g_{0, w_1, x_1}(P_{W'_1 X_1 | U'_1}(\cdot, \cdot | u_1)) = P_{W'_1 X_1 | U'_1}(w_1, x_1 | u_1) \\
 & \text{for } (w_1, x_1) = 1, \dots, |\mathcal{W}'_1| \cdot |\mathcal{X}_1| - 1 \text{ and} \\
 & g_I(P_{W'_1 X_1 | U'_1}(\cdot, \cdot | u_1)) \\
 &= H(X_1|Y_1, W'_1) - H(X_1|Y_1, W'_1, U_1 = u_1), \\
 & g_{\hat{d}}(P_{W'_1 X_1 | U'_1}(\cdot, \cdot | u_1)) \\
 &= \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(W'_1, u_1, Y_1)) \right].
 \end{aligned}$$

By using the support lemma to these functions, there exists U''_1 with cardinality $|\mathcal{U}''_1| \leq |\mathcal{W}'_1| \cdot |\mathcal{X}_1| + 1 = |\mathcal{X}_1|(|\mathcal{X}_1| + 3) + 1$ such that the marginal $P_{W'_1 X_1}$ is preserved,

$$I(U''_1; X_1|Y_1, W'_1) = I(U'_1; X_1|Y_1, W'_1),$$

and

$$\begin{aligned}
 & \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(W'_1, U''_1, Y_1)) \right] \\
 &= \mathbb{E} \left[\hat{d}_1(X_1, \hat{X}_1(W'_1, U'_1, Y_1)) \right].
 \end{aligned}$$

Similarly, there exists U''_2 with cardinality $|\mathcal{U}''_2| \leq |\mathcal{X}_2|(|\mathcal{X}_2| + 3) + 1$ such that $P_{W'_2 X_2}$ is preserved,

$$I(U''_2; X_2|Z_2, W'_2) = I(U'_2; X_2|Z_2, W'_2),$$

and

$$\begin{aligned}
 & \mathbb{E} \left[\tilde{d}_2(X_2, \tilde{X}_2(W'_2, U''_2, Z_2)) \right] \\
 &= \mathbb{E} \left[\tilde{d}_2(X_2, \tilde{X}_2(W'_2, U'_2, Z_2)) \right].
 \end{aligned}$$

By relabeling $(W'_1, U''_1, W'_2, U''_2)$ as (W_1, U_1, W_2, U_2) , we have the cardinality bounds. ■

D. Proof of (29) and (30)

a) *Proof of (29):* By noting that $G_p(\cdot)$ is a non-negative and convex function, for any $(\check{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\check{D}, \check{D})$ we have

$$\begin{aligned}
 B_p(\check{D}, \alpha, \beta, \theta, \tau) &= (\theta - \tau)G_p(\alpha) + \tau G_p(\beta) \\
 &\quad + (1 - \theta)G_p(\gamma(\check{D}, \alpha, \beta, \theta, \tau)) \\
 &\geq (\theta - \tau)G_p(\alpha) + \tau G_p(\beta) \\
 &\geq \theta G_p\left(\frac{\theta - \tau}{\theta}\alpha + \frac{\tau}{\theta}\beta\right) \\
 &\geq R_p^{WZ}(\hat{D}).
 \end{aligned}$$

Thus, the lefthand side of (29) is larger than or equal to the righthand side. On the other hand, for $\check{D} = \frac{1}{2}$, by setting $\check{D} = \check{D}$, $\alpha = \beta$ and $\tau = \frac{\theta}{2}$, and optimizing (β, θ) , we can show that the lefthand side achieves the righthand side in (29).

b) *Proof of (30):* By noting that $G_p(\cdot)$ is a convex function, for any $(\check{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\check{D}, \check{D})$ we have

$$\begin{aligned}
 B_p(\check{D}, \alpha, \beta, \theta, \tau) &= (\theta - \tau)G_p(\alpha) + \tau G_p(\beta) \\
 &\quad + (1 - \theta)G_p(\gamma(\check{D}, \alpha, \beta, \theta, \tau)) \\
 &= (\theta - \tau)G_p(1 - \alpha) + \tau G_p(\beta) \\
 &\quad + (1 - \theta)G_p(\gamma(\check{D}, \alpha, \beta, \theta, \tau)) \\
 &\geq G_p(\check{D}).
 \end{aligned}$$

Since $1 - h(p * \check{D})$ and $G_p(\check{D})$ are monotone decreasing for $\check{D} \leq \frac{1}{2}$, the lefthand side of (30) is larger than or equal to the righthand side. On the other hand, when $\hat{D} = p$, by setting $\check{D} = \check{D}$, $\theta = \tau = \alpha = \beta = 0$, we can show that the lefthand side coincides with the righthand side in (30). When $\hat{D} = \check{D}$, by setting $\check{D} = \check{D}$, $\theta = \tau = 1$, $\alpha = 0$, and $\beta = \check{D}$, we can show that the lefthand side coincides with the righthand side in (30).

E. Proof of Theorem 13

First, note that (28) can be written as

$$1 - H(Y|W) + H(Y|U, W) - H(X|U, W),$$

where we used the relations

$$I(W; Y) = 1 - H(Y|W), \quad (43)$$

$$I(U, W; X|Y) = H(Y|U, W) - H(X|U, W). \quad (44)$$

To prove (27), Tian and Diggavi essentially showed the following in [9, Appendix 5].

Lemma 15: Let (U, W) be auxiliary random variables satisfying the conditions 1 and 2 right after (26). Then, we have

$$1 - H(Y|W) \geq 1 - h(\check{D} * p),$$

$$H(Y|U, W) - H(X|U, W) \geq B_p(\check{D}, \alpha, \beta, \theta, \tau)$$

for some $(\check{D}, \alpha, \beta, \theta, \tau) \in \mathcal{Q}_p(\hat{D}, \hat{D})$.

By noting that Z_1 and Y_2 are constant and by using chain rules, for a fixed auxiliary random variable (U_1, W_1, U_2, W_2) , we can rewrite the rate condition of Theorem 7 as

$$\begin{aligned}
 & \max\{I(U_1, W_1; X_1|Y_1) + I(W_2; Z_2) + I(U_2, W_2; X_2|Z_2), \\
 & I(W_1; Y_1) + I(U_1, W_1; X_1|Y_1) + I(U_2, W_2; X_2|Z_2)\}.
 \end{aligned}$$

Then, by using Lemma 15 and the relations in (43) and (44), we have

$$\begin{aligned}
 & I(U_1, W_1; X_1|Y_1) + I(W_2; Z_2) + I(U_2, W_2; X_2|Z_2) \\
 & \geq 1 - h(\check{D}_2 * p_2) + B_{p_1}(\check{D}_1, \alpha_1, \beta_1, \theta_1, \tau_1) \\
 & \quad + B_{p_2}(\check{D}_2, \alpha_2, \beta_2, \theta_2, \tau_2)
 \end{aligned}$$

and

$$\begin{aligned}
 & I(W_1; Y_1) + I(U_1, W_1; X_1|Y_1) + I(U_2, W_2; X_2|Z_2) \\
 & \geq 1 - h(\check{D}_1 * p_1) + B_{p_1}(\check{D}_1, \alpha_1, \beta_1, \theta_1, \tau_1) \\
 & \quad + B_{p_2}(\check{D}_2, \alpha_2, \beta_2, \theta_2, \tau_2)
 \end{aligned}$$

for some

$$(\tilde{D}_1, \alpha_1, \beta_1, \theta_1, \tau_1) \in \mathcal{Q}_{p_1}(\hat{D}_1, \tilde{D}_1)$$

and

$$(\tilde{D}_2, \alpha_2, \beta_2, \theta_2, \tau_2) \in \mathcal{Q}_{p_2}(\tilde{D}_2, \hat{D}_2).$$

Thus, the lefthand side is larger than or equal to the righthand side in (31). We can prove the other direction of inequality by using the reverse test channels described in Figs. 3 and 4.

REFERENCES

- [1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.
- [3] T. Berger, *Multiterminal Source Coding*, ser. CISM Courses and Lectures: The Information Theory Approach to Communication. Springer, 1978, vol. 229.
- [4] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [5] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, 2011.
- [6] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 727–734, November 1985.
- [7] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Inform. Theory*, vol. 23, no. 2, pp. 179–182, March 1977.
- [8] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1636–1654, August 2004.
- [9] C. Tian and S. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side informations," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2946–2960, August 2007.
- [10] —, "Side-information scalable source coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5591–508, December 2008.
- [11] R. Timo, A. Grant, and G. Kramer, "Rate-distortion functions for source coding with complementary side information," in *Proceedings of IEEE International Symposium on Information Theory*, Saint Petersburg, Russia, August 2011, pp. 2988–2992.
- [12] —, "Lossy broadcasting with complementary side information," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 104–131, January 2013.
- [13] R. Timo, T. J. Oechtering, and M. Wigger, "Source coding with conditionally less noisy side information," in *Proceedings of IEEE Information Theory Workshop*, Lausanne, Switzerland, September 2012, pp. 55–59.
- [14] R. Timo, A. Grant, and G. Kramer, "Source coding problems with conditionally less noisy side information," arXiv:1212.2396.
- [15] R. Ahlswede, N. Cai, S.-Y. R., and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.
- [16] A. Kimura, T. Uyematsu, S. Kuzuoka, and S. Watanabe, "Universal source coding over generalized complementary delivery networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 1360–1373, March 2009.
- [17] G. S. Poltyrev, "Capacity for parallel broadcast channels with degraded components," *Probl. Inf. Transm.*, vol. 13, no. 2, pp. 23–35, 1977.
- [18] A. El Gamal, "Capacity of the product and sum of two unmatched broadcast channels," *Probl. Inf. Transm.*, vol. 16, no. 1, pp. 3–23, 1977.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.
- [20] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Inf. Transm.*, vol. 10, no. 3, pp. 3–14, 1974.
- [21] R. Ahlswede and J. Körner, "Source coding with side information and a converse for the degraded broadcast channel," *IEEE Trans. Inform. Theory*, vol. 21, no. 6, pp. 629–637, November 1975.
- [22] J. Körner and K. Marton, "Comparison of two noisy channels," *Keszthely Colloquium on Information Theory*, pp. 411–423, 1975.
- [23] A. ElGamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 166–169, March 1979.
- [24] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4207–4214, September 2010.
- [25] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.
- [26] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5240–5257, August 2011.
- [27] K. J. Kerpez, "The rate-distortion function of a binary symmetric source when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 33, no. 3, pp. 448–452, May 1987.
- [28] M. Fleming and M. Effros, "On rate-distortion with mixed types of side information," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1698–1705, April 2006.
- [29] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

Shun Watanabe (M'09) received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 2005, 2007, and 2009 respectively. Since April 2009, he has been an Assistant Professor in the Department of Information Science and Intelligent Systems of the University of Tokushima. Since April 2013, he also has been a visiting Assistant Professor in the Institute for Systems Research of the University of Maryland, College Park. His current research interests are in the areas of information theory, quantum information theory, and quantum cryptography.