



Rate-Distortion Region of a Gray–Wyner Model with Side Information

Meryem Benammar, Abdellatif Zaidi

► To cite this version:

Meryem Benammar, Abdellatif Zaidi. Rate-Distortion Region of a Gray–Wyner Model with Side Information. Entropy, 2018, Transfer Entropy as a Tool for Hydrodynamic Model Validation, 20 (1), pp.2.1-2.23. 10.3390/e20010002 . hal-01806554

HAL Id: hal-01806554

<https://hal.science/hal-01806554v1>

Submitted on 2 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

Rate-Distortion Region of a Gray-Wyner Model with Side Information

Meryem Benammar ¹  and Abdellatif Zaidi ^{1,†} 

¹ Mathematics and Algorithmic Sciences Lab, Huawei Technologies France, Boulogne-Billancourt, 92100, France ; meryem.benammar@huawei.com

[†] Université Paris-Est, Champs-sur-Marne, 77454, France ; abdellatif.zaidi@huawei.com , abdellatif.zaidi@u-pem.fr.

* Correspondence: meryem.benammar@huawei.com

Academic Editor: name

Version July 27, 2017 submitted to Entropy

Abstract: In this work, we establish a full single-letter characterization of the rate-distortion region of an instance of the Gray-Wyner model with side information at the decoders. Specifically, in this model an encoder observes a pair of memoryless, arbitrarily correlated, sources (S_1^n, S_2^n) and communicates with two receivers over an error-free rate-limited link of capacity R_0 , as well as error-free rate-limited individual links of capacities R_1 to the first receiver and R_2 to the second receiver. Both receivers reproduce the source component S_2^n losslessly; and Receiver 1 also reproduces the source component S_1^n lossily, to within some prescribed fidelity level D_1 . Also, Receiver 1 and Receiver 2 are equipped respectively with memoryless side information sequences Y_1^n and Y_2^n . Important in this setup, the side information sequences are arbitrarily correlated among them, and with the source pair (S_1^n, S_2^n) ; and are not assumed to exhibit any particular ordering. Furthermore, by specializing the main result to two Heegard-Berger models with *successive refinement* and *scalable coding*, we shed light on the roles of the common and private descriptions that the encoder should produce and the role of each of the common and private links. We develop intuitions by analysing the developed single-letter rate-distortion regions of these models, and discuss some insightful binary examples.

Keywords: Rate-distortion, Gray-Wyner, side-information, Heegard-Berger, successive refinement,

1. Introduction

The Gray-Wyner source coding problem was originally formulated, and solved, by Gray and Wyner in [1]. In their original setting, a pair of arbitrarily correlated memoryless sources (S_1^n, S_2^n) is to be encoded and transmitted to two receivers that are connected to the encoder each through a common error-free rate-limited link as well as a private error-free rate-limited link. Because the channels are rate-limited, the encoder produces a compressed bit string W_0 of rate R_0 that it transmits over the common link, and two compressed bit strings, W_1 of rate R_1 and W_2 of rate R_2 , that it transmits respectively over the private link to first receiver and the private link to the second receiver. The first receiver uses the bit strings W_0 and W_1 to reproduce an estimate \hat{S}_1^n of the source component S_1^n to within some prescribed distortion level D_1 , for some distortion measure $d_1(\cdot, \cdot)$. Similarly, the second receiver uses the bit strings W_0 and W_2 to reproduce an estimate \hat{S}_2^n of the source component S_2^n to within some prescribed distortion level D_2 , for some different distortion measure $d_2(\cdot, \cdot)$. In [1], Gray and Wyner characterized the optimal achievable rate triples (R_0, R_1, R_2) and distortion pairs (D_1, D_2) . Figure 1 shows a generalization of Gray-Wyner's original model in which the receivers also observe correlated memoryless side information sequences, Y_1^n at Receiver 1 and Y_2^n at Receiver 2. Some special cases of the Gray-Wyner's model with side information of Figure 1 have been solved (see the "Related

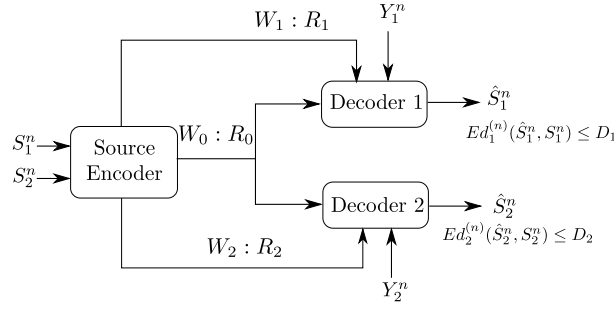


Figure 1. Gray-Wyner network with side information at the receivers.

Work" section below). However, in its most general form, i.e., when the side information sequences are arbitrarily correlated among them and with the sources, this problem has so far eluded single-letter characterization of the optimal rate-distortion region. Indeed, the Gray-Wyner problem with side information subsumes the well known Heegard-Berger problem [2], obtained by setting $R_1 = R_2 = 0$ in Figure 1, which remains, to date, an open problem.

In this paper, we study an instance of the Gray-Wyner's model with side information of Figure 1 in which the reconstruction sets are degraded, meaning, both receivers reproduce the source component S_2^n losslessly and Receiver 1 wants also to reproduce the source component S_1^n lossily, to within some prescribed distortion level D_1 . It is important to note that, while the reconstruction sets are nested, and so degraded, no specific ordering is imposed on the side information sequences, which then can be arbitrarily correlated among them and with the sources (S_1^n, S_2^n) .

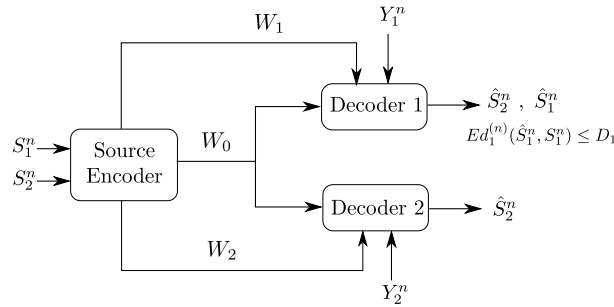


Figure 2. Gray-Wyner model with side information at both receivers and degraded reconstruction sets

As in the Gray-Wyner original coding scheme, the encoder produces a common description of the sources pair (S_1^n, S_2^n) that is intended to be recovered by both receivers, as well as individual or private descriptions of (S_1^n, S_2^n) that are destined to be recovered each by a distinct receiver. Because the side information sequences do *not* exhibit any specific ordering, the choice of the information that each description should carry, and, the links over which each is transmitted to its intended receiver, are challenging questions that we answer in this work.

In order to build the understanding of the role of each of the links and of the descriptions in the optimal coding scheme for the setting of Figure 2, we will investigate as well two important underlying problems which are Heegard-Berger type models with refinement links as shown in Figure 3. In both models, only one of the two refinement individual links has non-zero rate.

In the model of Figure 3a, the receiver that accesses the additional rate-limited link (i.e., Receiver 1) is also required to reproduce a lossy estimate of the source component S_1^n , in addition to the source component S_2^n which is to be reproduced losslessly by both receivers. We will refer to this model as a "Heegard-Berger problem with successive refinement". Reminiscent of successive refinement source coding, this model may be appropriate to model applications in which descriptions of only

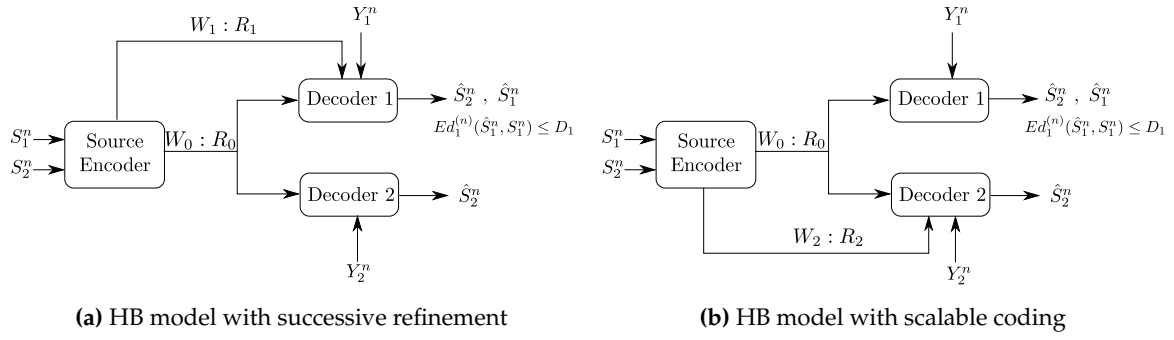


Figure 3. Two classes of Heegard-Berger models

some components (e.g., S_2^n) of the source suffices at the first use of the data; and descriptions of the remaining components (e.g., S_1^n) are needed only at a later stage.

The model of Figure 3b has the individual rate-limited link connected to the receiver that is required to reproduce only the source component S_2^n . We will refer to this model as a “Heegard-Berger problem with scalable coding”, reusing a term that was introduced in [3] for a similar scenario, and in reference to that user 1 may have such a “good quality” side information that only a minimal amount of information from the encoder suffices, and so, in order not to constrain the communication by user 2 with the lower quality side information, an additional rate limited link R_2 is added to balance the decoding capabilities of both users.

1.1. Main Contributions

The main result of this paper is a single-letter characterization of the optimal rate-distortion region of the Gray-Wyner model with side information and degraded reconstruction sets of Figure 2. To this end, we derive a converse proof that is tailored specifically for the model with degraded reconstruction sets that we study here. For the proof of the direct part, we develop a coding scheme that is very similar to one developed in the context of coding for broadcast channels with feedback in [4], but with an appropriate choice of the variables which we specify here. The specification of the main result to the Heegard-Berger models with successive refinement and scalable coding of Figure 3 sheds light on the roles of the common and private descriptions and what they should carry optimally. We develop intuitions by analysing the established single-letter optimal rate-distortion regions of these two models, and illustrate our discussion through some binary examples.

1.2. Related Works

In [4], Shayevitz and Wigger study a two-receiver discrete memoryless broadcast channel with feedback. They develop an efficient coding scheme which treats the feedback signal as a source that has to be conveyed lossily to the receivers in order to refine their messages’ estimates, through a block Markov coding scheme. In doing so, the users’ channel outputs are regarded as side information sequences; and so the scheme clearly connects with the Gray-Wyner model with side information of Figure 1 - as is also clearly explicit in [4]. The Gray-Wyner model with side information for which Shayevitz and Wigger’s develop a (source) coding scheme, as part of their study of the broadcast channel with feedback, assumes general, possibly distinct, distortion measures at the receivers (i.e., not necessarily nested) and side information sequences that are arbitrarily correlated among them and with the source. In this paper we show that when specialized to the model with degraded reconstruction sets of Figure 2 that we study here, Shayevitz and Wigger’s coding scheme for the Gray-Wyner model with side information of [4] yields a rate-distortion region that meets the converse result that we here establish; and so is optimal.

The Gray-Wyner model with side information generalizes another long standing open source coding problem, the famous Heegard-Berger problem [5]. Full single-letter characterization of the optimal rate-distortion function of the Heegard-Berger problem is known only in few specific cases, the most important of which are the cases of i) stochastically degraded side information sequences [5] (see also [6]), ii) Sgarro's result [7] on the corresponding lossless problem, iii) Gaussian sources with quadratic distortion measure [3,8], iv) some instances of conditionally less-noisy side information sequences [9] and v) the recently solved HB model with general side information sequences and degraded reconstruction sets [10], i.e., the model of Figure 2 with $R_1 = R_2 = 0$ — in the lossless case, a few other optimal results were shown, such as for the so-called complementary delivery [11]. A lower bound for general instances of the rate distortion problem with side information at multiple decoders, that is inspired by a linear-programming lower bound for index coding, has been developed recently by Unal and Wagner in [12].

Successive refinement of information was investigated by Equitz *et al.* in [13] wherein the description of the source is successively refined to a collection of receivers which are required to reconstruct the source with increasing quality levels. Extensions of successive refinement to cases in which the receivers observe some side information sequences was first investigated by Steinberg *et al.* in [14] who establish the optimal rate-distortion region under the assumption that the receiver that observes the refinement link, say receiver 1, observes also a *better* side information sequence than the opposite user, i.e. the Markov chain $S \dashv Y_1 \dashv Y_2$ holds. Tian *et al.* give in [8] an equivalent formulation of the result of [14] and extend it to the N-stage successive refinement setting. In [3], Tian *et al.* investigate another setting, coined as "side information scalable coding", in which it is rather the receiver that accesses the refinement link, say receiver 2, which observes the *less good* side information sequence, i.e. $S \dashv Y_1 \dashv Y_2$. Balancing refinement quality and side information asymmetry for such a side-information scalable source coding problem allows authors in [3] to derive the rate-distortion region in the degraded side information case. The previous results on successive refinement in the presence of side information, which were generalized by Timo *et al.* in [15], all assume, however, a specific structure in the side information sequences.

1.3. Outline

An outline of the remainder of this paper is as follows. Section II describes formally the Gray-Wyner model with side information and degraded reconstruction sets of Figure 2 that we study in this paper. Section III contains the main result of this paper, a full single-letter characterization of the rate-distortion region of the model of Figure 2, together with some useful discussions and connections. A formal proof of the direct and converse parts of this result appear in Section VI. In Section IV and Section V, we specialize the result respectively to the Heegard-Berger model with successive refinement of Figure 3a and the Heegard-Berger model with scalable coding of Figure 3b. These sections also contain insightful discussions illustrated by some binary examples.

Notation

Throughout the paper we use the following notations. The term pmf stands for probability mass function. Upper case letters are used to denote random variables, e.g., X ; lower case letters are used to denote realizations of random variables, e.g., x ; and calligraphic letters designate alphabets, i.e., \mathcal{X} . Vectors of length n are denoted by $X^n = (X_1, \dots, X_n)$, and X_i^j is used to denote the sequence (X_i, \dots, X_j) , whereas $X_{<i>} \triangleq (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$. The probability distribution of a random variable X is denoted by $P_X(x) \triangleq \mathbb{P}(X = x)$. Sometimes, for convenience, we write it as P_X . We use the notation $\mathbb{E}_X[\cdot]$ to denote the expectation of random variable X . A probability distribution of a random variable Y given X is denoted by $P_{Y|X}$. The set of probability distributions defined on an alphabet \mathcal{X} is denoted by $\mathcal{P}(\mathcal{X})$. The cardinality of a set \mathcal{X} is denoted by $\|\mathcal{X}\|$. For random variables X , Y and Z , the notation $X \dashv Y \dashv Z$ indicates that X, Y and Z , in this order, form a Markov Chain, i.e., $P_{XYZ}(x, y, z) = P_Y(y)P_{X|Y}(x|y)P_{Z|Y}(z|y)$. The set $\mathcal{T}_{[X]}^{(n)}$ denotes the set of sequences strongly typical

with respect to the probability distribution P_X and the set $\mathcal{T}_{[X|Y^n]}^{(n)}$ denotes the set of sequences x^n jointly typical with y^n with respect to the joint p.m.f. P_{XY} . Throughout this paper, we use $h_2(\alpha)$ to denote the entropy of a Bernoulli (α) source, i.e., $h_2(\alpha) = -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha)$. Also, the indicator function is denoted by $\mathbb{1}(\cdot)$. For real-valued scalars a and b , with $a \leq b$, the notation $[a, b]$ means the set of real numbers that are larger or equal than a and smaller or equal b . For integers $i \leq j$, $[i : j]$ denotes the set of integers comprised between i and j , i.e., $[i : j] = \{i, i + 1, \dots, j\}$. Finally, throughout the paper, logarithms are taken to base 2.

2. Problem Setup and Formal Definitions

Consider the Gray-Wyner source coding model with side information and degraded reconstruction sets shown in Figure 2. Let $(\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{Y}_1 \times \mathcal{Y}_2, P_{\mathcal{S}_1, \mathcal{S}_2, \mathcal{Y}_1, \mathcal{Y}_2})$ be a discrete memoryless vector source with generic variables S_1, S_2, Y_1 and Y_2 . Also, let $\hat{\mathcal{S}}_1$ be a reconstruction alphabet and, d_1 a distortion measure defined as:

$$\begin{aligned} d_1 : \mathcal{S}_1 \times \hat{\mathcal{S}}_1 &\rightarrow \mathbb{R}_+ \\ (s_1, \hat{s}_1) &\rightarrow d_1(s_1, \hat{s}_1) . \end{aligned} \quad (1)$$

Definition 1. An $(n, M_{0,n}, M_{1,n}, M_{2,n}, D_1)$ code for the Gray-Wyner source coding model with side information and degraded reconstruction sets of Figure 2 consists of:

- Three sets of messages $\mathcal{W}_0 \triangleq [1 : M_{0,n}]$, $\mathcal{W}_1 \triangleq [1 : M_{1,n}]$, and $\mathcal{W}_2 \triangleq [1 : M_{2,n}]$.
- Three encoding functions, f_0, f_1 and f_2 defined, for $j \in \{0, 1, 2\}$ as

$$\begin{aligned} f_j : \mathcal{S}_1^n \times \mathcal{S}_2^n &\mapsto \mathcal{W}_j \\ (S_1^n, S_2^n) &\mapsto W_j = f_j(S_1^n, S_2^n) . \end{aligned} \quad (2)$$

- Two decoding functions g_1 and g_2 , one at each user:

$$\begin{aligned} g_1 : \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{Y}_1^n &\mapsto \hat{\mathcal{S}}_2^n \times \hat{\mathcal{S}}_1^n \\ (W_0, W_1, Y_1^n) &\mapsto (\hat{S}_{2,1}^n, \hat{S}_1^n) = g_1(W_0, W_1, Y_1^n) , \end{aligned} \quad (3)$$

and

$$\begin{aligned} g_2 : \mathcal{W}_0 \times \mathcal{W}_2 \times \mathcal{Y}_2^n &\mapsto \hat{\mathcal{S}}_2^n \\ (W_0, W_2, Y_2^n) &\mapsto \hat{S}_{2,2}^n = g_2(W_0, W_2, Y_2^n) . \end{aligned} \quad (4)$$

The expected distortion of this code is given by

$$\mathbb{E} \left(d_1^{(n)}(S_1^n, \hat{S}_1^n) \right) \triangleq \mathbb{E} \frac{1}{n} \sum_{i=1}^n d_1(S_{1,i}, \hat{S}_{1,i}) . \quad (5)$$

The probability of error is defined as

$$P_e^{(n)} \triangleq \mathbb{P}(\hat{S}_{2,1}^n \neq S_2^n \text{ or } \hat{S}_{2,2}^n \neq S_2^n) . \quad (6)$$

□

Definition 2. A rate triple (R_0, R_1, R_2) is said to be D_1 -achievable for the Gray-Wyner source coding model with side information and degraded reconstruction sets of Figure 2 if there exists a sequence of $(n, M_{0,n}, M_{1,n}, M_{2,n}, D_1)$ codes such that:

$$\limsup_{n \rightarrow \infty} P_e^{(n)} = 0 , \quad (7)$$

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left(d_1^{(n)}(S_1^n, \hat{S}_1^n) \right) \leq D_1 , \quad (8)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log_2(M_{j,n}) \leq R_j \text{ for } j \in \{0, 1, 2\} \quad (9)$$

The rate-distortion region \mathcal{RD} of this problem is defined as the union of all rate-distortion quadruples (R_0, R_1, R_2, D_1) such that (R_0, R_1, R_2) is D_1 -achievable, i.e.,

$$\mathcal{RD} \triangleq \cup \{ (R_0, R_1, R_2, D_1) : (R_0, R_1, R_2) \text{ is } D_1\text{-achievable} \} . \quad (10)$$

□

As we already mentioned, we shall also study the special case Heegard-Berger type models shown in Figure 3. The formal definitions for these models are similar to the above, and we omit them here for brevity.

3. Gray-Wyner Model with Side Information and Degraded Reconstruction Sets

In the following, we establish the main result of this work, i.e., the single-letter characterization of the optimal rate-distortion region \mathcal{RD} of the Gray-Wyner model with side information and degraded reconstructions sets shown in Figure 2. We then describe how the result subsumes and generalizes existing rate-distortion regions for this setting under different assumptions.

Theorem 1. *The rate-distortion region \mathcal{RD} of the Gray-Wyner problem with side information and degraded reconstruction set of Figure 2 is given by the sets of all rate-distortion quadruples (R_0, R_1, R_2, D_1) satisfying:*

$$R_0 + R_1 \geq H(S_2|Y_1) + I(U_0U_1; S_1|S_2Y_1) \quad (11a)$$

$$R_0 + R_2 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) \quad (11b)$$

$$R_0 + R_1 + R_2 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) + I(U_1; S_1|U_0S_2Y_1) \quad (11c)$$

for some product pmf $P_{U_0U_1S_1S_2Y_1Y_2}$, such that:

1) the following Markov chain is valid:

$$(Y_1, Y_2) \text{---} (S_1, S_2) \text{---} (U_0, U_1) \quad (12)$$

2) and there exists a function $\phi : \mathcal{Y}_1 \times \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{S}_2 \rightarrow \hat{\mathcal{S}}_1$ such that:

$$\mathbb{E}d_1(S_1, \hat{S}_1) \leq D_1 . \quad (13)$$

Proof: The detailed proof of the direct part and the converse part of this theorem appear in Section VI.

The proof of converse, which is the most challenging part, uses appropriate combinations of bounding techniques for the transmitted rates based on the system model assumptions and Fano's inequality, a series of analytic bounds based on the underlying Markov chains, and most importantly, a proper use of Csiszár-Körner sum identity in order to derive single letter bounds.

As for the proof of achievability, it combines the optimal coding scheme of the Heegard-Berger problem with degraded reconstruction sets [10] and the double-binning based scheme of Shayevitz and Wigger [4, Theorem 2] for the Gray-Wyner problem with side information, and is outlined in the following.

The encoder produces a common description of (S_1^n, S_2^n) that is intended to be recovered by both receivers, and an individual description that is intended to be recovered only by Receiver 1. The common description is chosen as $V_0^n = (U_0^n, S_2^n)$ and is thus designed so as to describe all of S_2^n , which both receivers are required to reproduce losslessly, but also all or part of S_1^n , depending on the desired distortion level D_1 . Since we make no assumptions on the side information sequences, this is meant to account for possibly unbalanced side information pairs (Y_1^n, Y_2^n) , in a manner that is similar to [10]. The message that carries the common description is obtained at the encoder through the technique of

double-binning of Tian and Diggavi in [3], used also by Shayevitz and Wigger [4, Theorem 2] for a Gray-Wyner model with side information. In particular, similar to the coding scheme of [4, Theorem 2], the double-binning is performed in two ways, one that is tailored for Receiver 1 and one that is tailored for Receiver 2.

More specifically, the codebook of the common description is composed of codewords v_0^n that are drawn randomly and independently according to the product law of P_{V_0} ; and is partitioned uniformly into $2^{n\tilde{R}_{0,0}}$ superbins, indexed with $\tilde{w}_{0,0} \in [1 : 2^{n\tilde{R}_{0,0}}]$. The codewords of each superbin of this codebook are partitioned in two distinct ways. In the first partition, they are assigned randomly and independently to $2^{n\tilde{R}_{0,1}}$ subbins indexed with $\tilde{w}_{0,1} \in [1 : 2^{n\tilde{R}_{0,1}}]$, according to a uniform pmf over $[1 : 2^{n\tilde{R}_{0,1}}]$. Similarly, in the second partition, they are assigned randomly and independently to $2^{n\tilde{R}_{0,2}}$ subbins indexed with $\tilde{w}_{0,2} \in [1 : 2^{n\tilde{R}_{0,2}}]$, according to a uniform pmf over $[1 : 2^{n\tilde{R}_{0,2}}]$. The codebook of the private description is composed of codewords u_1^n that are drawn randomly and independently according to the product law of $P_{U_1|V_0}$. This codebook is partitioned similarly uniformly into $2^{n\tilde{R}_{1,0}}$ superbins indexed with $\tilde{w}_{1,0} \in [1 : 2^{n\tilde{R}_{1,0}}]$, each containing $2^{n\tilde{R}_{1,1}}$ subbins indexed with $\tilde{w}_{1,1} \in [1 : 2^{n\tilde{R}_{1,1}}]$ codewords u_1^n .

Upon observing a typical pair $(S_1^n, S_2^n) = (s_1^n, s_2^n)$, the encoder finds a pair of codewords (v_0^n, u_1^n) that is jointly typical with (s_1^n, s_2^n) . Let $\tilde{w}_{0,0}$, $\tilde{w}_{0,1}$ and $\tilde{w}_{0,2}$ denote respectively the indices of the superbin, subbin of the first partition and subbin of the second partition of the codebook of the common description, in which lies the found v_0^n . Similarly, let $\tilde{w}_{1,0}$ and $\tilde{w}_{1,1}$ denote respectively the indices of the superbin and subbin of the codebook of the individual description in which lies the found u_1^n . The encoder sets the common message W_0 as $W_0 = (\tilde{w}_{0,0}, \tilde{w}_{1,0})$ and sends it over the error-free rate-limited common link of capacity R_0 . Also, it sets the individual message W_1 as $W_1 = (\tilde{w}_{0,1}, \tilde{w}_{1,1})$ and sends it the error-free rate-limited link to Receiver 1 of capacity R_1 ; and the individual message W_2 as $W_2 = \tilde{w}_{0,2}$ and sends it the error-free rate-limited link to Receiver 2 of capacity R_2 . For the decoding, Receiver 2 utilizes the second partition of the codebook of the common description; and looks in the subbin of index $\tilde{w}_{0,2}$ of the superbin of index $\tilde{w}_{0,0}$ for a unique v_0^n that is jointly typical with its side information y_2^n . Receiver 1 decodes v_0^n similarly, utilizing the first partition of the codebook of the common description and its side information y_1^n . It also utilizes the codebook of the individual description; and looks in the subbin of index $\tilde{w}_{1,1}$ of the superbin of index $\tilde{w}_{1,0}$ for a unique u_1^n that is jointly typical with the pair (y_1^n, v_0^n) . In the formal proof in Section IV, we argue that with an appropriate choice of the communication rates $\tilde{R}_{0,0}$, $\tilde{R}_{0,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$ and $\tilde{R}_{1,1}$, as well as the sizes of the subbins, this scheme achieves the rate-distortion region of Theorem 1. \square

A few remarks that connect Theorem 1 to known results on related models are in order.

Remark 1. The setting of figure 1 generalizes two important settings which are the Gray-Wyner problem, through the presence of side-information sequences Y_1^n and Y_2^n , and the Heegard-Berger problem, through the presence of private links of rates R_1 and R_2 . As such, the coding scheme for the setting of Figure 2 differs from that of the Gray-Wyner problem and that of the Heegard-Berger problem in many aspects as shown in Figure 4.

First, the presence of side information sequences imposes the use of binning for each of the produced descriptions V_0^n , V_1^n and V_2^n in the Gray-Wyner code construction. However, unlike the binning performed in the Heegard-Berger coding scheme, the binning of the common codeword V_0^n needs to be performed with two different indices, each tailored to a side information sequence at the respective receivers, i.e., double binning. Another different aspect is the role of the private and common links. When in Gray-Wyner's original work, these links carried each a description, i.e., V_0^n on the common link and V_1^n resp. V_2^n on the private links of rates R_1 resp. R_2 , and when in the Heegard-Berger the three descriptions V_0^n , V_1^n and V_2^n are all carried through the common link only, in the optimal coding scheme of the setting of figure 2, the private and common links play different roles. Indeed, the common description V_0^n and the private description V_j^n are transmitted on both the common link and the private link of rates R_0 and R_j , for $j \in \{1, 2\}$, through rate-splitting. As such, these key differences imply an intricate interplay between the side information sequences and the role of the common and private links, which we will emphasize later on in sections IV and V.

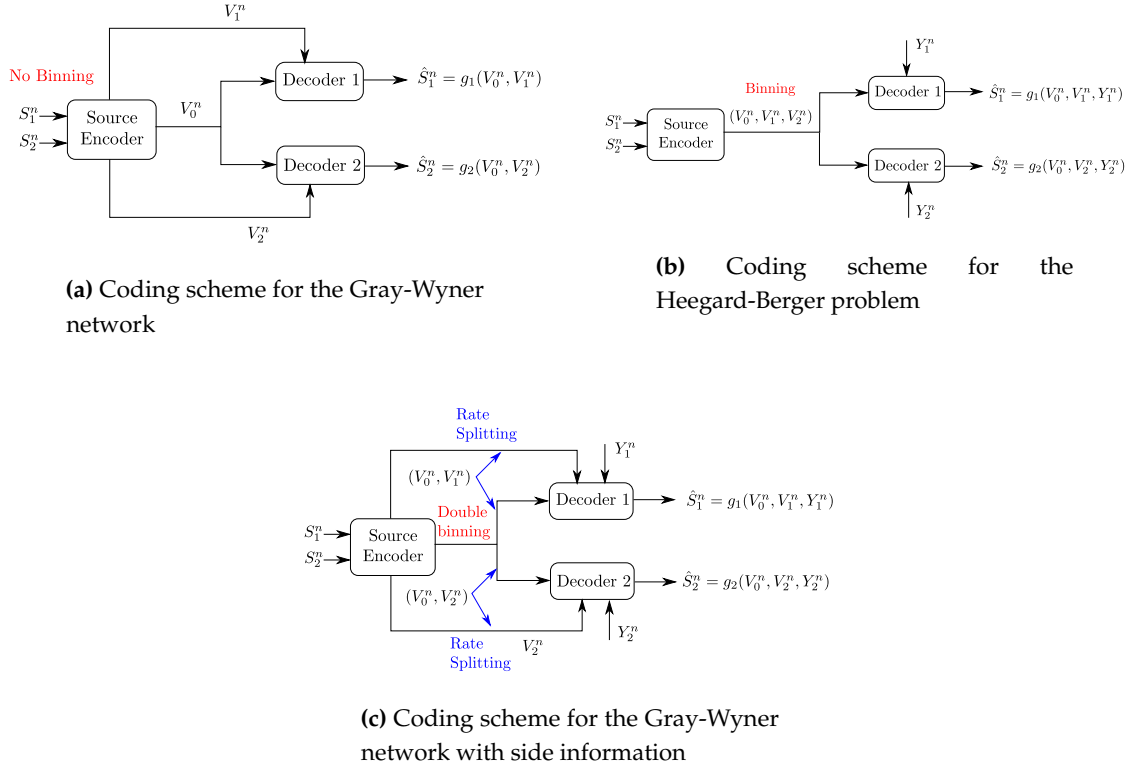


Figure 4. Comparison of coding schemes for the Gray-Wyner network with side information, the Gray-Wyner network and the Heegard-Berger problem.

Remark 2. In the special case in which $R_1 = R_2 = 0$, the Gray-Wyner model with side information and degraded reconstruction sets of Figure 2 reduces to a Heegard-Berger problem with arbitrary side information sequences and degraded reconstruction sets, a model that was studied, and solved, recently in the authors' own recent work [10]. Theorem 1 can then be seen as a generalization of [10, Theorem1] to the case in which the encoder is connected to the receivers also through error-free rate-limited private links of capacity R_1 and R_2 respectively. The most important insight in the Heegard-Berger problem with degraded reconstruction sets is the role that the common description V_0 should play in such a setting. Authors show in [10, Theorem1] that the optimal choice of this description is to contain, intuitively, the common source S_2 intended to both users, and, maybe less intuitive, an additional description U_0 , i.e. $V_0 = (U_0, S_2)$, which is used to piggyback part of the source S_1 in the common codeword though not required by both receivers, in order to balance the asymmetry of the side information sequences. In sections IV and V we show that the utility of this description will depend on both the side information sequences and the rates of the private links. \square

Remark 3. In [16], Timo et al. study the Gray-Wyner source coding model with side information of Figure 1. They establish the rate-region of this model in the specific case in which the side information sequence Y_2^n is a degraded version of Y_1^n , i.e., $(S_1, S_2) \rightarrow Y_1 \rightarrow Y_2$ is a Markov chain, and both receivers reproduce the component S_2^n and Receiver 1 also reproduces the component S_1^n , all in a lossless manner. The result of Theorem 1 generalizes that of [16, Theorem 5] to the case of side information sequences that are arbitrarily correlated among them and with the source pair (S_1, S_2) and lossy reconstruction of S_1 . In [16], Timo et al. also investigate, and solve, a few other special cases of the model, such as those of single source $S_1 = S_2$ [16, Theorem 4] and complementary delivery $(Y_1, Y_2) = (S_2, S_1)$ [16, Theorem 6]. The results of [16, Theorem 4] and [16, Theorem 6] can be recovered from Theorem 1 as special cases of it. Theorem 1 also generalizes [16, Theorem 6] to the case of lossy reproduction of the component S_1^n . \square

4. The Heegard-Berger Problem with Successive Refinement

An important special case of the Gray-Wyner source coding model with side information and degraded reconstruction sets of Figure 2 is the case in which $R_2 = 0$. The resulting model, a Heegard-Berger problem with successive refinement, is shown in Figure 3a.

In this section, we derive the optimal rate distortion region for this setting, and show how it compares to existing results in literature. Besides, we discuss the utility of the common description U_0 depending, not only on the side information sequences structures, but also on the refinement link rate R_1 . We illustrate through a binary example that the utility of U_0 , namely the optimality of the choice of a non-degenerate $U_0 \neq \emptyset$, is governed by the quality of the refinement link rate R_1 and the side information structure.

4.1. Rate-Distortion Region

The following theorem states the optimal rate-distortion region of the Heegard-Berger problem with successive refinement of Figure 3a.

Corollary 1. *The rate-distortion region of the Heegard-Berger problem with successive refinement of Figure 3a is given by the set of rate-distortion triples (R_0, R_1, D_1) satisfying:*

$$R_0 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) \quad (14a)$$

$$R_0 + R_1 \geq H(S_2|Y_1) + I(U_0U_1; S_1|S_2Y_1) \quad (14b)$$

$$R_0 + R_1 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) + I(U_1; S_1|U_0S_2Y_1) \quad (14c)$$

for some product pmf $P_{U_0U_1S_1S_2Y_1Y_2}$, such that:

1) the following Markov chain is valid:

$$(U_0, U_1) \multimap (S_1, S_2) \multimap (Y_1, Y_2) \quad (15)$$

2) and there exists a function $\phi : \mathcal{Y}_1 \times \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{S}_2 \rightarrow \hat{\mathcal{S}}_1$ such that:

$$\mathbb{E}d_1(S_1, \hat{S}_1) \leq D_1. \quad (16)$$

Proof: The proof of Corollary 1 follows from that of Theorem 1 by setting $R_2 = 0$ therein.

Remark 4. Recall the coding scheme of Theorem 1. If $R_2 = 0$, the second partition of the codebook of the common description, which is relevant for Receiver 2, becomes degenerate since, in this case, all the codewords v_0^n of a superbin $\mathcal{B}_{00}(\tilde{w}_{0,0})$ are assigned to a single subbin. Correspondingly, the common message that the encoder sends over the common link carries only the index $\tilde{w}_{0,0}$ of the superbin $\mathcal{B}_{00}(\tilde{w}_{0,0})$ of the codebook of the common description in which lies the typical pair $v_0^n = (s_2^n, u_0^n)$, in addition to the index $\tilde{w}_{1,0}$ of the subbin $\mathcal{B}_{10}(\tilde{w}_{1,0})$ of the codebook of the individual description in which lies the recovered typical u_1^n . The constraint (14a) on the common rate R_0 is in accordance with that Receiver 2 utilizes only the index $\tilde{w}_{0,0}$ in the decoding. Furthermore, note that the constraints (14b) and (14c) on the sum-rate $(R_0 + R_1)$ can be combined as

$$R_0 + R_1 \geq \max \{I(U_0S_2; S_1S_2|Y_1), I(U_0S_2; S_1S_2|Y_2)\} + I(U_1; S_1|U_0S_2Y_1) \quad (17)$$

which resembles the Heegard-Berger result of [2, Theorem 2, p. 733].

Remark 5. As we already mentioned, the result of Corollary 1 holds for side information sequences that are arbitrarily correlated among them and with the sources. In the specific case in which the user who gets the refinement rate-limited link also has the “better-quality” side information, in the sense that $(S_1, S_2) \multimap Y_1 \multimap Y_2$

forms a Markov chain, the rate-distortion region of Corollary 1 reduces to the set of all rate-distortion triples (R_0, R_1, D_1) that satisfy

$$R_0 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) \quad (18a)$$

$$R_0 + R_1 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) + I(U_1; S_1|U_0S_2Y_1) . \quad (18b)$$

for some joint pmf $P_{U_0U_1S_1S_2Y_1Y_2}$ for which (15) and (16) hold. This result can also be obtained from previous works on successive refinement for the Wyner-Ziv source coding problem by Steinberg and Merhav [14, Theorem 1] and Tian and Diggavi [8, Theorem 1]. The results of [14, Theorem 1] and [8, Theorem 1] hold for possibly distinct, i.e., not necessarily nested, distortion measures at the receivers; but they require the aforementioned Markov chain condition which is pivotal for their proofs. Thus, for the considered degraded reconstruction sets setting, Corollary 1 can be seen as generalizing [14, Theorem 1] and [8, Theorem 1] to the case in which the side information sequences are arbitrarily correlated among them and with the sources (S_1, S_2) , i.e., do not exhibit any ordering. \square

Remark 6. In the case in which it is the user who gets only the common rate-limited link that has the “better-quality” side information, in the sense that $(S_1, S_2) \rightarrow Y_2 \rightarrow Y_1$ forms a Markov chain, the rate distortion region of Corollary 1 reduces to the set of all rate-distortion triples (R_0, R_1, D_1) that satisfy

$$R_0 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) \quad (19a)$$

$$R_0 + R_1 \geq H(S_2|Y_1) + I(U_0U_1; S_1|S_2Y_1) \quad (19b)$$

for some joint pmf $P_{U_0U_1S_1S_2Y_1Y_2}$ for which (15) and (16) hold. This result can also be conveyed from [3]. Specifically, in [3] Tian and Diggavi study a therein referred to as “side-information scalable” source coding setup where the side informations are degraded, and the encoder produces two descriptions such that the receiver with the better-quality side information (Receiver 2 if $(S_1, S_2) \rightarrow Y_2 \rightarrow Y_1$ is a Markov chain) uses only the first description to reconstruct its source while the receiver with the low-quality side information (Receiver 1 if $(S_1, S_2) \rightarrow Y_2 \rightarrow Y_1$ is a Markov chain) uses the two descriptions in order to reconstruct its source. They establish inner and outer bounds on the rate-distortion region of the model, which coincide when either one of the decoders requires a lossless reconstruction or when the distortion measures are degraded and deterministic. Similar to the previous remark, Corollary 1 can be seen as generalizing the aforementioned results of [3] to the case in which the side information sequences are arbitrarily correlated among them and with the sources (S_1, S_2) . \square

Remark 7. A crucial remark that is in order for the Heegard-Berger problem with successive refinement of Figure 3a, is that, depending on the rate of the refinement link R_1 , resorting to a common auxiliary variable U_0 might be unnecessary. Indeed, in the case in which S_1 needs to be recovered losslessly at the first receiver, for instance, parts of the rate-region can be achieved without resorting to the common auxiliary variable U_0 , setting $U_0 = \emptyset$, while other parts of the rate region can only be achieved through a non-trivial choice of U_0 .

As such, if $R_1 \geq H(S_1|S_2Y_1)$, then letting $U_0 = \emptyset$ yields the optimal rate region. To see this, note that the rate constraints under lossless construction of S_1 write as:

$$R_0 \geq H(S_1S_2|Y_2) - H(S_1|S_2Y_2U_0) \quad (20a)$$

$$R_0 + R_1 \geq H(S_1S_2|Y_1) \quad (20b)$$

$$R_0 + R_1 \geq H(S_1S_2|Y_2) - H(S_1|S_2Y_2U_0) + H(S_1|U_0S_2Y_1) \quad (20c)$$

which, can be rewritten as follows

$$R_0 \geq H(S_1S_2|Y_2) + \min_{P_{U_0|S_1S_2}} [(H(S_1|S_2Y_1U_0) - R_1)^+ - H(S_1|S_2Y_2U_0)] \quad (21a)$$

$$R_0 + R_1 \geq H(S_1 S_2 | Y_1) \quad (21b)$$

288 where $(x)^+ \triangleq \max\{0, x\}$.

Under the constraint that $R_1 \geq H(S_1 | S_2 Y_1)$, the constraints in (21) reduce to the following

$$R_0 \geq H(S_1 S_2 | Y_2) - \max_{P_{U_0 | S_1 S_2}} H(S_1 | S_2 Y_2 U_0) \quad (22a)$$

$$R_0 + R_1 \geq H(S_1 S_2 | Y_1). \quad (22b)$$

289 Next, by noting that $\max_{P_{U_0 | S_1 S_2}} H(S_1 | S_2 Y_2 U_0) = H(S_1 | S_2 Y_2)$ is achieved by $U_0 = \emptyset$, the claim follows.

290

291 However, when $R_1 < H(S_1 | S_2 Y_1)$, the choice of $U_0 = \emptyset$ might be strictly sub-optimal (as shown in the
292 following binary example). \square

293 4.2. Binary Example

294 Let X_1, X_2, X_3 and X_4 be four independent $\text{Ber}(1/2)$ random variables. Let the sources be
295 $S_1 \triangleq (X_1, X_2, X_3)$ and $S_2 \triangleq X_4$. Now, consider the Heegard-Berger model with successive refinement
296 shown in Figure 5. The first user, which gets both the common and individual links, observes the side
297 information $Y_1 = (X_1, X_4)$ and wants to reproduce the pair (S_1, S_2) losslessly. The second user gets
298 only the common link, has side information $Y_2 = (X_2, X_3)$ and wants to reproduce only the component
299 S_2 , losslessly.

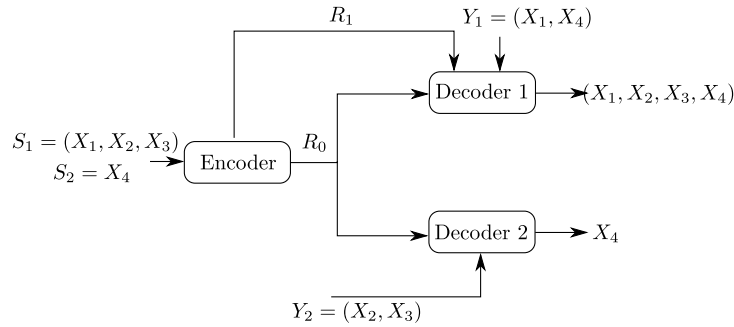


Figure 5. Binary Heegard-Berger example with successive refinement

300 The side information at the decoders do *not* exhibit any degradedness ordering, in the sense that none
301 of the Markov chain conditions of Remark 5 and Remark 6 hold. The following claim provides the
302 rate-region of this binary example.

Claim 1. The rate region of the binary Heegard-Berger example with successive refinement of Figure 5 is given by the set of rate pairs (R_0, R_1) that satisfy

$$R_0 \geq 1 \quad (23a)$$

$$R_0 + R_1 \geq 2. \quad (23b)$$

Proof. The proof of Claim 1 follows easily by computing the rate region

$$R_0 \geq H(S_1 S_2 | Y_2) - H(S_1 | S_2 Y_2 U_0) \quad (24a)$$

$$R_0 + R_1 \geq H(S_1 S_2 | Y_1) \quad (24b)$$

$$R_0 + R_1 \geq H(S_1 S_2 | Y_2) - H(S_1 | S_2 Y_2 U_0) + H(S_1 | U_0 S_2 Y_1) \quad (24c)$$

in the binary setting under study.
First, we note that

$$H(S_1 S_2 | Y_2) = H(X_1 X_4 | X_2 X_3) = 2 \quad (25)$$

$$H(S_1 S_2 | Y_1) = H(X_2 X_3 | X_1 X_4) = 2. \quad (26)$$

which allows then to rewrite the rate region as

$$R_0 \geq 2 - H(X_1 | X_4 U_0) \geq 2 - H(X_1 | X_4) = 1 \quad (27a)$$

$$R_0 + R_1 \geq 2 + \max\{0, H(X_2 X_3 | X_1 X_4 U_0) - H(X_1 | X_2 X_3 X_4 U_0)\} \geq 2 \quad (27b)$$

The proof of the claim follows by noticing that the following inequalities hold with equality for the choices $U_0 = (X_2, X_3)$ or $U_0 = X_2$ or $U_0 = X_3$.
□

The rate region of Claim 1 is depicted in Figure 6. It is insightful to notice that although the second user is only interested in reproducing the component $S_2 = X_4$, the optimal coding scheme that achieves this region sets the common description that is destined to be recovered by both users as one that is composed of not only S_2 but also some part $U_0 = (X_2, X_3)$, or $U_0 = X_2$ or $U_0 = X_3$, of the source component S_1 (though the latter is not required by the second user). A possible intuition is that this choice of U_0 is useful for user 1, who wants to reproduce $S_1 = (X_1, X_2, X_3)$, and its transmission to also the second user does not cost any rate loss since this user has available side information $Y_2 = (X_2, X_3)$.

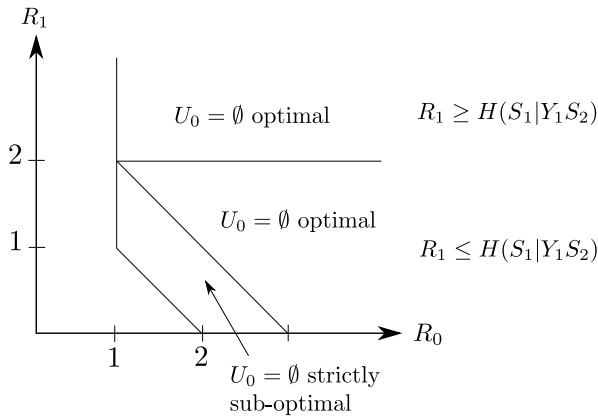


Figure 6. Rate region of the binary example of Figure 5. The choices $U_0 = (X_2, X_3)$ or $U_0 = X_2$ or $U_0 = X_3$ are optimal irrespective of the value of R_1 , while the degenerate choice $U_0 = \emptyset$ is optimal only in some slices of the region.

5. The Heegard-Berger Problem with Scalable Coding

In the following, we consider the model of Figure 3b. As we already mentioned, the reader may find it appropriate for the motivation to think about the side information Y_2^n as being of lower quality than Y_1^n , in which case, the refinement link that is given to the second user is intended to improve its decoding capability. In this section, we describe the optimal coding scheme for this setting, and show that it can be recovered, independently, from the work of Timo *et al.* [15] through a careful choice of the coding sets. Next, we illustrate through a binary example the interplay between the utility of the common description U_0 and the side information sequences, and the refinement rate R_2 .

5.1. Rate-Distortion Region

The following theorem states the rate-distortion region of the Heegard-Berger model with scalable coding of Figure 3b.

Corollary 2. *The rate-distortion region of the Heegard-Berger model with scalable coding of Figure 3b is given by the set of all rate-distortion triples (R_0, R_2, D_1) that satisfy*

$$R_0 \geq H(S_2|Y_1) + I(U_0U_1; S_1|S_2Y_1) \quad (28a)$$

$$R_0 + R_2 \geq H(S_2|Y_2) + I(U_0; S_1|S_2Y_2) + I(U_1; S_1|U_0S_2Y_1) \quad (28b)$$

for some product pmf $P_{U_0U_1S_1S_2Y_1Y_2}$, such that:

1) the following Markov chain is valid:

$$(U_0, U_1) \dashv\vdash (S_1, S_2) \dashv\vdash (Y_1, Y_2) \quad (29)$$

2) and there exists a function $\phi : \mathcal{Y}_1 \times \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{S}_2 \rightarrow \hat{\mathcal{S}}_1$ such that:

$$\mathbb{E}d_1(S_1, \hat{S}_1) \leq D_1. \quad (30)$$

Proof. The proof of Corollary 2 follows from that of Theorem 1 by setting $R_1 = 0$ therein. \square

Remark 8. *In the specific case in which Receiver 2 has a better-quality side information in the sense that $(S_1, S_2) \dashv\vdash Y_2 \dashv\vdash Y_1$ forms a Markov chain, the rate distortion region of Corollary 2 reduces to one that is described by a single rate-constraint, namely*

$$R_0 \geq H(S_2|Y_1) + I(U; S_1|S_2Y_1) \quad (31)$$

for some conditional $P_{U|S_1S_2}$ that satisfies $\mathbb{E}[d_1(S_1, \hat{S}_1)] \leq D_1$. This is in accordance with the observation that, in this case, the transmission to Receiver 1 becomes the bottleneck, as Receiver 2 can recover the source component S_2 losslessly as long as so does Receiver 1.

Remark 9. *Consider the case in which S_1 needs to be recovered losslessly as well at receiver 1. Then, the rate region is given by (??), which can be expressed similarly as follows*

$$R_0 \geq H(S_1S_2|Y_1) \quad (32a)$$

$$R_0 + R_2 \geq H(S_1S_2|Y_2) + \min_{P_{U_0|S_1S_2}} [H(S_1|U_0S_2Y_1) - H(S_1|U_0S_2Y_2)]. \quad (32b)$$

An important comment here is that the optimization problem in $P_{U_0|S_1S_2}$ does not depend on the refinement link R_2 , and the optimal solution to it, i.e., the optimal choice of U_0 , meets the solution to the Heegard-Berger problem without refinement link, $R_2 = 0$, rendering it optimal for all choices of R_2 , which is a main difference with the Heegard-Berger problem with refinement link of Figure 3a in which the solution to the Heegard-Berger problem (with $R_1 = 0$) might not be optimal for all values of R_1 .

Remark 10. In [15, Theorem 1], Timo et al. present an achievable rate-region for the multistage successive-refinement problem with side information. Timo et al. consider distortion measures of the form $\delta_l : \mathcal{X} \times \hat{\mathcal{X}}_l \rightarrow \mathbb{R}_+$, where \mathcal{X} is the source alphabet and $\hat{\mathcal{X}}_l$ is the reconstruction at decoder l , $l \in \{1, \dots, t\}$; and for this reason this result is not applicable as is to the setting of Figure 3b, in the case of two decoders. However, the result of [15, Theorem 1] can be extended to accommodate a distortion measure at the first decoder that is vector-valued; and the direct part of Corollary 2 can then be obtained by applying this extension.

Table 1. Auxiliary random variables associated with the subsets that appear in (36).

	\mathcal{T}_0	\mathcal{T}_1	\mathcal{T}_2
$\mathcal{A}_{\mathcal{T}_j}^-$	\emptyset	\emptyset	U_1
$\mathcal{A}_{\mathcal{T}_j}^\supset$	\emptyset	U_{12}	U_{12}
$\mathcal{A}_{\mathcal{T}_j}^\pm$	$\{U_1, U_2\}$	\emptyset	\emptyset
$\mathcal{A}_{\mathcal{T}_j}^\dagger$	\emptyset	\emptyset	\emptyset
$\mathcal{A}_{\mathcal{T}_j,1}^\dagger$	\emptyset	\emptyset	\emptyset
$\mathcal{A}_{\mathcal{T}_j,2}^\dagger$	\emptyset	\emptyset	\emptyset

Specifically, in the case of two decoders, i.e., $t = 2$, and with $X = (S_1, S_2)$, and two distortion measures $\delta_1 : \mathcal{S}_1 \times \mathcal{S}_2 \times \hat{\mathcal{S}}_{1,1} \times \hat{\mathcal{S}}_{1,2} \rightarrow \{0, 1\} \times \mathbb{R}_+$ and $\delta_2 : \mathcal{S}_1 \times \mathcal{S}_2 \times \hat{\mathcal{S}}_{1,2} \times \hat{\mathcal{S}}_{2,2} \rightarrow \{0, 1\}$ chosen such that

$$\delta_1((s_1, s_2), (\hat{s}_{1,1}, \hat{s}_{2,1})) = (d_H(s_2, \hat{s}_{2,1}), d_1(s_1, \hat{s}_{1,1})) \quad (33)$$

and

$$\delta_2((s_1, s_2), (\hat{s}_{1,2}, \hat{s}_{2,2})) = d_H(s_2, \hat{s}_{2,2}) \quad (34)$$

where $d_H(\cdot, \cdot)$ is the Hamming distance, letting $d_1 = (0, D_1)$ and $d_2 = 0$, a straightforward extension of [15, Theorem 1] to this setting yields a rate-region that is described by the following rate constraints (using the notation of [15, Theorem 1])

$$R_0 \geq \Phi(\mathcal{T}_0, 1) + \Phi(\mathcal{T}_1, 1) \quad (35a)$$

$$R_0 + R_2 \geq \Phi(\mathcal{T}_0, 2) + \Phi(\mathcal{T}_1, 2) + \Phi(\mathcal{T}_2, 2) \quad (35b)$$

where $\mathcal{T}_0 = \{1, 2\}$, $\mathcal{T}_1 = \{1\}$, $\mathcal{T}_2 = \{2\}$, and for $j = 0, 1, 2$ and $l \in 1, 2$ such that $\mathcal{T}_j \cap \{1, \dots, l\} \neq \emptyset$, the function $\Phi(\mathcal{T}_j, l)$, $j = 0, 1, 2$, is defined as

$$\Phi(\mathcal{T}_j, l) = I(S_1 S_2 \mathcal{A}_{\mathcal{T}_j}^\dagger; U_{\mathcal{T}_j} | \mathcal{A}_{\mathcal{T}_j}^\supset) - \min_{l' \in \mathcal{T}_j \cap [1:l]} I(U_{\mathcal{T}_j}; \mathcal{A}_{\mathcal{T}_j, l'}^\dagger Y_{l'} | \mathcal{A}_{\mathcal{T}_j}^\supset) \quad (36)$$

where $\mathcal{A} = \{U_{12}, U_1, U_2\}$ and the sets $\mathcal{A}_{\mathcal{T}_j}^-$, $\mathcal{A}_{\mathcal{T}_j}^\supset$, $\mathcal{A}_{\mathcal{T}_j}^\pm$, $\mathcal{A}_{\mathcal{T}_j}^\dagger$, $\mathcal{A}_{\mathcal{T}_j,1}^\dagger$, $\mathcal{A}_{\mathcal{T}_j,2}^\dagger$, evaluated in this case, are given in Table 1. It is easy to see that the region described by (35) can be written more explicitly in this case as

$$R_0 \geq I(U_{12}; S_1 S_2 | Y_1) \quad (37a)$$

$$R_0 + R_2 \geq \max\{I(U_{12}; S_1 S_2 | Y_1), I(U_{12}; S_1 S_2 | Y_2)\} + I(U_1; S_1 S_2 | Y_1 U_{12}) + I(U_2; S_1 S_2 | Y_2 U_{12}). \quad (37b)$$

Also, setting $U_{12} = (U_0, S_2)$ and $U_2 = S_2$ in (37) one recovers the rate-region of Corollary 2. (Such a connection can also be stated for the result of Corollary 1).

5.2. Binary Example

Consider the setting of ???. Let X_1, X_2, X_3 and X_4 be four independent $\text{Ber}(1/2)$ random variables. Let the sources be $S_1 \triangleq (X_1, X_2, X_3)$ and $S_2 \triangleq X_4$. Now, consider the Heegard-Berger model with scalable coding shown in Figure 7. The first user, which gets both only the common link, observes the side information $Y_1 = (X_1, X_4)$ and wants to reproduce the pair (S_1, S_2) losslessly. The second user gets both the common and private links, has side information $Y_2 = (X_2, X_3)$ and wants to reproduce only the component S_2 , losslessly.

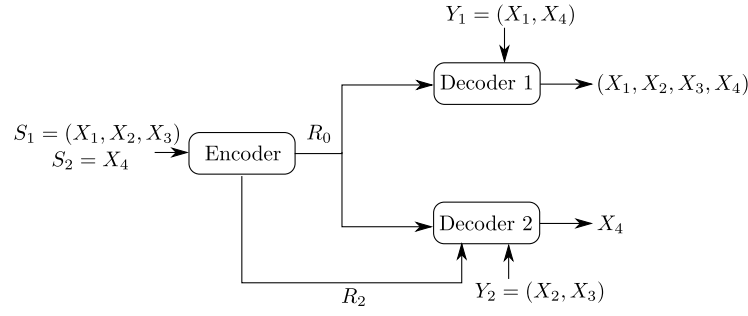


Figure 7. Binary Heegard-Berger example with scalable coding

Claim 2. The rate region of the binary Heegard-Berger example with scalable coding of Figure ?? is given by the set of all rate pairs (R_0, R_2) that satisfy $R_2 \geq 0$ and $R_0 \geq 2$.

Proof. The proof of Claim 2 follows easily by specializing, and computing, the result of Remark 9 for the example at hand. First note that

$$R_0 + R_2 \geq H(S_2 S_1 | Y_2) + \min_{P_{U_0 | S_1 S_2}} [H(S_1 | U_0 S_2 Y_1) - H(S_1 | U_0 S_2 Y_2)] \quad (38a)$$

$$= 2 + \min_{P_{U_0 | S_1 S_2}} [H(X_2 X_3 | X_1 X_4 U_0) - H(X_1 | X_2 X_3 X_4 U_0)] \quad (38b)$$

$$\geq 2 + \min_{P_{U_0 | S_1 S_2}} [-H(X_1 | X_2 U_0)] \quad (38c)$$

$$\geq 1 \quad (38d)$$

where equality in all previous inequalities is satisfied with $U_0 = (X_2, X_3)$ or with $U_0 = X_2$ or $U_0 = X_3$. Note as well that the single rate constraint on R_0 writes as:

$$R_0 \geq H(S_1 S_2 | Y_1) \quad (39a)$$

$$= 2 \quad (39b)$$

which renders the sum-rate constraint redundant and ends the proof of the claim. \square

The optimal rate region of Claim 2 is depicted in Figure 8, as the region delimited by the lines $R_0 = 1$ and $R_2 = 0$. Note that for this example, the source component X_2 , which is the only source component that is required by Receiver 2, needs to be transmitted entirely on the common link so as to be recovered losslessly also by Receiver 1. For this reason, the refinement link is not-constrained and appears to be useless for this example.

There is a sharp difference with the binary Heegard-Berger example with successive refinement of Figure 5 for which the refinement link may sometimes be instrumental to reducing the required rate on the common link. With scalable coding, the refinement link with rate R_0 does not improve the rate transmitted on the common link.

Also, it is insightful to notice that for this example, because of the side information configuration, the choice $U_0 = \emptyset$ in Corollary 2 is strictly suboptimal and results in the smaller region that is described by

$$R_0 \geq 2 \quad (40a)$$

$$R_0 + R_2 \geq 3. \quad (40b)$$

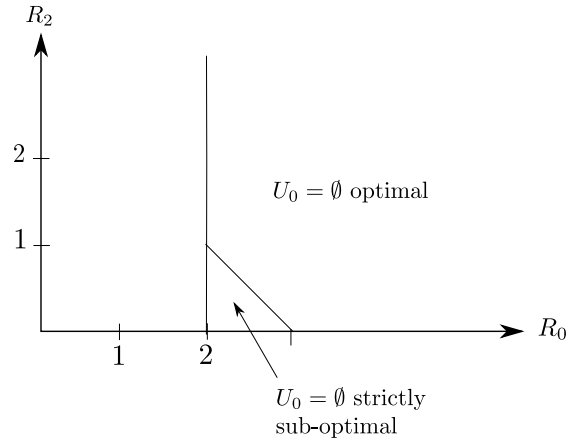


Figure 8. The optimal rate region for the setting of Figure 7 given by $(R_0 \geq 2, R_2 \geq 0)$. The choice of $U_0 = \emptyset$ is optimal only in a slice of the region.

6. Proof of Theorem 1

In the following, we give the proof of the converse part and the direct part of Theorem 1.

The converse part is strongly dependent on the system model we investigate and consists in a series of careful bounding steps resorting to Fano's inequality, Markov chains and Csiszár-Körner sum-identity.

The proof of achievability is two-fold, and consists in proving a general result that holds for a Gray-Wyner setting with side information, and then deriving the optimal choice of the auxiliary codewords involved for the specific setting with degraded reconstruction sets.

6.1. Proof of Converse Part

Assume that a rate triple (R_0, R_1, R_2) is D_1 -achievable. Let then $W_j = f_j(S_1^n, S_2^n)$, where $j \in \{0, 1, 2\}$, be the encoded indices and let $\hat{S}_1^n = g_1(W_0, W_1, Y_1^n)$ be the reconstruction sequence at the first decoder such that $\mathbb{E}d_1^{(n)}(S_1^n, \hat{S}_1^n) \leq D_1$.

Using Fano's inequality, the lossless reconstruction of the source S_2^n at both decoders implies that there exists a sequence $\epsilon_n \xrightarrow{n \rightarrow \infty} 0$ such that:

$$H(S_2^n | W_0 W_1 Y_1^n) \leq n\epsilon_n \quad (41)$$

$$H(S_2^n | W_0 W_2 Y_2^n) \leq n\epsilon_n \quad (42)$$

We start by showing the following sum-rate constraint,

$$R_0 + R_1 + R_2 \geq H(S_2 | Y_2) + I(U_0; S_1 | S_2 Y_2) + I(U_1; S_1 | U_0 S_2 Y_1). \quad (43)$$

We have that

$$\begin{aligned} & n(R_0 + R_1 + R_2) \\ & \geq H(W_0) + H(W_2) + H(W_1) \end{aligned} \quad (44a)$$

$$\geq H(W_0) + H(W_2 | W_0) + H(W_1) \quad (44b)$$

$$= H(W_0 W_2) + H(W_1) \quad (44c)$$

$$\geq H(W_0 W_2 | Y_2^n) + H(W_1 | W_0 S_2^n Y_1^n) \quad (44d)$$

$$\geq I(W_0 W_2; S_1^n S_2^n | Y_2^n) + I(W_1; S_1^n | W_0 S_2^n Y_1^n) \quad (44e)$$

$$= H(S_1^n S_2^n | Y_2^n) - H(S_1^n S_2^n | W_0 W_2 Y_2^n) + H(S_1^n | W_0 S_2^n Y_1^n) - H(S_1^n | W_0 W_1 S_2^n Y_1^n) \quad (44f)$$

$$\stackrel{(a)}{\geq} H(S_1^n S_2^n | Y_2^n) - H(S_1^n | W_0 W_2 S_2^n Y_2^n) + H(S_1^n | W_0 S_2^n Y_1^n) - H(S_1^n | W_0 W_1 S_2^n Y_1^n) - n\epsilon_n \quad (44g)$$

$$\geq H(S_1^n S_2^n | Y_2^n) - H(S_1^n | W_0 S_2^n Y_2^n) + H(S_1^n | W_0 S_2^n Y_1^n) - H(S_1^n | W_0 W_1 S_2^n Y_1^n) - n\epsilon_n \quad (44h)$$

where (a) stems from Fano's inequality (42), which results from the lossless reconstruction of S_2^n at receiver 2.

Let us define then:

$$A \triangleq H(S_1^n | W_0 S_2^n Y_1^n) - H(S_1^n | W_0 S_2^n Y_2^n), \quad (45)$$

$$B \triangleq H(S_1^n | W_0 W_1 S_2^n Y_1^n). \quad (46)$$

In the following, we aim for single-letter bounds on the two quantities A and B .

Since the side information sequences Y_1^n and Y_2^n are not degraded and do not exhibit any structure, together with the sources (S_1^n, S_2^n) , single-letterizing the quantity A can be obtained through some judicious bounding steps that are reported below, in which some important Markov chain are shown to hold and quantities are manipulated appropriately, together with several invocations of Csiszár-Körner sum identity.

Let us start by writing that

$$A \triangleq H(S_1^n | W_0 S_2^n Y_1^n) - H(S_1^n | W_0 S_2^n Y_2^n) \quad (47a)$$

$$= I(S_1^n; Y_2^n | W_0 S_2^n) - I(S_1^n; Y_1^n | W_0 S_2^n) \quad (47b)$$

$$= \sum_{i=1}^n [I(S_1^n; Y_{2,i} | W_0 Y_2^{i-1} S_2^n) - I(S_1^n; Y_{1,i} | W_0 Y_{1,i+1} S_2^n)] \quad (47c)$$

$$\stackrel{(a)}{=} \sum_{i=1}^n [I(S_1^n Y_{1,i+1}^n; Y_{2,i} | W_0 Y_2^{i-1} S_2^n) - I(S_1^n Y_2^{i-1}; Y_{1,i} | W_0 Y_{1,i+1} S_2^n)] \quad (47d)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n [I(S_1^n; Y_{2,i} | W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n) - I(S_1^n; Y_{1,i} | W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n)] \quad (47e)$$

$$\stackrel{(c)}{=} \sum_{i=1}^n [I(S_{1,i}; Y_{2,i} | W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n) - I(S_{1,i}; Y_{1,i} | W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n)] \quad (47f)$$

$$= \sum_{i=1}^n [H(S_{1,i} | Y_{1,i} W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n) - H(S_{1,i} | Y_{2,i} W_0 Y_2^{i-1} Y_{1,i+1}^n S_2^n)] \quad (47g)$$

$$= \sum_{i=1}^n [H(S_{1,i} | Y_{1,i} S_{2,i} U_{0,i}) - H(S_{1,i} | Y_{2,i} S_{2,i} U_{0,i})] \quad (47h)$$

where $U_{0,i} \triangleq (W_0, Y_2^{i-1}, Y_{1,i+1}^n, S_{2,<i>})$ (note that the lossless reconstruction of S_2^n at both receivers is instrumental to the definition of U_0 which plays the role of the common auxiliary variable in the proof of converse), and where (a) follows using the following Csiszár-Körner sum-identity

$$\sum_{i=1}^n I(Y_2^{i-1}; Y_{1,i} | S_1^n W_0 Y_{1,i+1}^n S_2^n) = \sum_{i=1}^n I(Y_{1,i+1}^n; Y_{2,i} | S_1^n W_0 Y_2^{i-1} S_2^n), \quad (48)$$

(b) follows using the Csiszár-Körner sum-identity given by

$$\sum_{i=1}^n I(Y_2^{i-1}; Y_{1,i} | W_0 Y_{1,i+1}^n S_2^n) = \sum_{i=1}^n I(Y_{1,i+1}^n; Y_{2,i} | W_0 Y_2^{i-1} S_2^n), \quad (49)$$

while (c) is the consequence of the following sequence of Markov chains

$$(S_1^{i-1}, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n, Y_{1,i+1}^n, Y_2^{i-1}) \multimap (S_{1,i}, S_{2,i}) \multimap Y_{j,i} \quad (50a)$$

$$\stackrel{(a)}{\Rightarrow} (S_1^{i-1}, S_{1,i+1}^n, S_2^{i-1}, S_{2,i+1}^n, Y_{1,i+1}^n, Y_2^{i-1}, W_0) \multimap (S_{1,i}, S_{2,i}) \multimap Y_{j,i} \quad (50b)$$

$$\Rightarrow (S_1^{i-1}, S_{1,i+1}^n) \ominus (S_2^{i-1}, S_{2,i+1}^n, Y_{1,i+1}^n, Y_2^{i-1}, W_0, S_{1,i}, S_{2,i}) \ominus Y_{j,i} \quad (50c)$$

where (50a) results from that the source sequences $(S_1^n, S_2^n, Y_1^n, Y_2^n)$ are memoryless, while (a) is a consequence of that W_0 is a function of the pair of sequences (S_1^n, S_2^n) .

To upper-bound the term B, note the following

$$B \triangleq H(S_1^n | W_0 W_1 S_2^n Y_1^n) \quad (51a)$$

$$= \sum_{i=1}^n H(S_{1,i} | W_0 W_1 S_2^n Y_1^{i-1}) \quad (51b)$$

$$= \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} W_0 S_{2,<i>} Y_{1,i+1}^n S_1^{i-1} W_1 Y_1^{i-1}) \quad (51c)$$

$$\stackrel{(a)}{=} \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} W_0 S_{2,<i>} Y_{1,i+1}^n S_1^{i-1} Y_2^{i-1} W_1 Y_1^{i-1}) \quad (51d)$$

$$\leq \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} W_0 S_{2,<i>} Y_{1,i+1}^n Y_2^{i-1} W_1 Y_1^{i-1}) \quad (51e)$$

where (a) is a consequence of the following sequence of Markov chains:

$$Y_2^{i-1} \ominus (S_1^{i-1}, S_2^{i-1}, Y_1^{i-1}) \ominus (S_{1,i}, S_{1,i+1}^n, S_{2,i}, S_{2,i+1}^n, Y_{1,i+1}^n) \quad (52a)$$

$$\stackrel{(a)}{\Rightarrow} Y_2^{i-1} \ominus (S_1^{i-1}, S_2^{i-1}, Y_1^{i-1}) \ominus (S_{1,i}, S_{1,i+1}^n, S_{2,i}, S_{2,i+1}^n, Y_{1,i+1}^n, W_0, W_1) \quad (52b)$$

$$\Rightarrow Y_2^{i-1} \ominus (S_1^{i-1}, S_2^{i-1}, Y_1^{i-1}, S_{2,i}, S_{2,i+1}^n, Y_{1,i+1}^n, W_0, W_1) \ominus S_{1,i} . \quad (52c)$$

where (52a) results from that the source sequences $(S_1^n, S_2^n, Y_1^n, Y_2^n)$ are memoryless, while (a) is a consequence of that W_0 and W_1 are each function of the pair of sequences (S_1^n, S_2^n) .

Finally, letting $U_{1,i} \triangleq (W_1, Y_1^{i-1})$ so that the choice of $(U_{0,i}, U_{1,i})$ satisfy the condition: $\hat{S}_{1,i} = g_i(Y_{1,i}, U_{0,i}, U_{1,i}, S_{2,i})$, we write the resulting sum-rate constraint as

$$\begin{aligned} n(R_0 + R_1 + R_2) &\geq nH(S_1 S_2 | Y_2) + \sum_{i=1}^n [H(S_{1,i} | S_{2,i} Y_{1,i} U_{0,i}) - H(S_{1,i} | S_{2,i} Y_{2,i} U_{0,i})] \\ &\quad - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} U_{0,i} U_{1,i}) - n\epsilon_n \end{aligned} \quad (53)$$

Let us now prove that the following bound holds

$$R_0 + R_1 \geq H(S_2 S_1 | Y_1) - H(S_1 | U_0 U_1 Y_1 S_2) . \quad (54)$$

We have

$$n(R_0 + R_1) \geq H(W_0) + H(W_1 | W_0) \quad (55a)$$

$$= H(W_0, W_1) \quad (55b)$$

$$\geq H(W_0 W_1 | Y_1^n) \quad (55c)$$

$$\geq I(W_0 W_1; S_1^n S_2^n | Y_1^n) \quad (55d)$$

$$= H(S_1^n S_2^n | Y_1^n) - H(S_1^n S_2^n | W_0 W_1 Y_1^n) \quad (55e)$$

$$\stackrel{(a)}{\geq} H(S_1^n S_2^n | Y_1^n) - H(S_1^n | W_0 W_1 S_2^n Y_1^n) - n\epsilon_n \quad (55f)$$

$$= nH(S_1 S_2 | Y_1) - B - n\epsilon_n \quad (55g)$$

$$\stackrel{(b)}{\geq} nH(S_1 S_2 | Y_1) - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} U_{0,i} U_{1,i}) - n\epsilon_n . \quad (55h)$$

where (a) is a consequence of Fano's inequality in (41), which results from the lossless reconstruction of S_2^n at receiver 1, and (b) results from the upper bound on B in (51e).

As for the third rate constraint

$$R_0 + R_2 \geq H(S_1 S_2 | Y_2) - H(S_1 | U_0 Y_2 S_2) , \quad (56)$$

we write

$$n(R_0 + R_2) \geq H(W_0 W_2) \quad (57a)$$

$$\geq H(W_0 W_2 | Y_2^n) \quad (57b)$$

$$\geq I(W_0 W_2; S_1^n S_2^n | Y_2^n) \quad (57c)$$

$$= H(S_1^n S_2^n | Y_2^n) - H(S_1^n S_2^n | W_0 W_2 Y_2^n) \quad (57d)$$

$$\stackrel{(a)}{\geq} H(S_1^n S_2^n | Y_2^n) - H(S_1^n | W_0 W_2 S_2^n Y_2^n) - n\epsilon_n \quad (57e)$$

$$\geq H(S_1^n S_2^n | Y_2^n) - H(S_1^n | W_0 S_2^n Y_2^n) - n\epsilon_n \quad (57f)$$

$$= nH(S_1 S_2 | Y_2) - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{2,i} W_0 S_{2,<i> Y_{2,<i> S_{1,i+1}^n) - n\epsilon_n \quad (57g)$$

$$\stackrel{(b)}{=} nH(S_1 S_2 | Y_2) - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{2,i} W_0 S_{2,<i> Y_{2,<i> S_{1,i+1}^n Y_{1,i+1}^n) - n\epsilon_n \quad (57h)$$

$$\geq nH(S_1 S_2 | Y_2) - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{2,i} W_0 S_{2,<i> Y_2^{i-1} Y_{1,i+1}^n) - n\epsilon_n \quad (57i)$$

$$= nH(S_1 S_2 | Y_2) - \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{2,i} U_{0,i}) - n\epsilon_n . \quad (57j)$$

where (a) is a consequence of Fano's inequality in (42) and (b) stems for the following sequence of Markov Chains

$$Y_{1,i+1}^n \ominus (S_{2,i+1}^n, S_{1,i+1}^n, Y_{1,i+1}^n) \ominus (S_{1,i}, S_{1,i}^{i-1}, S_{2,i}, S_{2,i}^{i-1}, Y_{1,i}^{i-1}) \quad (58a)$$

$$\stackrel{(a)}{\Rightarrow} Y_{1,i+1}^n \ominus (S_{2,i+1}^n, S_{1,i+1}^n, Y_{1,i+1}^n) \ominus (S_{1,i}, S_{1,i}^{i-1}, S_{2,i}, S_{2,i}^{i-1}, Y_{1,i}^{i-1}, W_0, W_1) \quad (58b)$$

$$\Rightarrow Y_{1,i+1}^n \ominus (S_{2,i+1}^n, S_{1,i+1}^n, Y_{1,i+1}^n, S_{2,i}, S_{2,i}^{i-1}, Y_{1,i}^{i-1}, W_0, W_1) \ominus S_{1,i} . \quad (58c)$$

where (58a) results from that the source sequences $(S_1^n, S_2^n, Y_1^n, Y_2^n)$ are memoryless, while (a) is a consequence of that W_0 and W_1 are each function of the pair of sequences (S_1^n, S_2^n) .

Let Q be an integer-valued random variable, ranging from 1 to n , uniformly distributed over $[1 : n]$ and independent of all other variables $(S_1, S_2, U_0, U_1, Y_1, Y_2)$. We have

$$\begin{aligned} R_0 + R_1 + R_2 &\geq H(S_1 S_2 | Y_2) + \frac{1}{n} \sum_{i=1}^n [H(S_{1,i} | S_{2,i} Y_{1,i} U_{0,i}) - H(S_{1,i} | S_{2,i} Y_{2,i} U_{0,i})] \\ &\quad - \frac{1}{n} \sum_{i=1}^n H(S_{1,i} | S_{2,i} Y_{1,i} U_{0,i} U_{1,i}) - n\epsilon_n \end{aligned} \quad (59a)$$

$$\begin{aligned} &= H(S_1 S_2 | Y_2) + \sum_{i=1}^n \mathbb{P}(Q = i) [H(S_{1,Q} | S_{2,Q} Y_{1,Q} U_{0,Q}, Q = i) - H(S_{1,Q} | S_{2,Q} Y_{2,Q} U_{0,Q}, Q = i)] \\ &\quad - \sum_{i=1}^n \mathbb{P}(Q = i) H(S_{1,Q} | S_{2,Q} Y_{1,Q} U_{0,Q} U_{1,Q}, Q = i) - n\epsilon_n \end{aligned} \quad (59b)$$

$$\begin{aligned} &= H(S_1 S_2 | Y_2) + H(S_{1,Q} | S_{2,Q} Y_{1,Q} U_{0,Q} Q) - H(S_{1,Q} | S_{2,Q} Y_{2,Q} U_{0,Q} Q) \\ &\quad - H(S_{1,Q} | S_{2,Q} Y_{1,Q} U_{0,Q} U_{1,Q} Q) - n\epsilon_n . \end{aligned} \quad (59c)$$

$$\begin{aligned} &\stackrel{(a)}{=} H(S_1 S_2 | Y_2) + H(S_1 | S_2 Y_1 U_{0,Q} Q) - H(S_1 | S_2 Y_2 U_{0,Q} Q) \\ &\quad - H(S_1 | S_2 Y_1 U_{0,Q} U_{1,Q} Q) - n\epsilon_n . \end{aligned} \quad (59d)$$

where (a) is a consequence of that all sources $(S_1^n, S_2^n, Y_1^n, Y_2^n)$ are memoryless. Let us now define $U_1 \triangleq (Q, U_{1,Q})$ and $U_0 \triangleq (Q, U_{0,Q})$, we obtain

$$R_0 + R_1 + R_2 \geq H(S_1 S_2 | Y_2) + H(S_1 | S_2 Y_1 U_0) - H(S_1 | S_2 Y_2 U_0) - H(S_1 | S_2 Y_1 U_0 U_1). \quad (60)$$

The two other rate constraints can be written in a similar fashion,

$$R_0 + R_1 \geq H(S_2 S_1 | Y_1) - H(S_1 | U_0 U_1 Y_1 S_2) \quad (61a)$$

$$R_0 + R_2 \geq H(S_1 S_2 | Y_2) - H(S_1 | U_0 Y_2 S_2); \quad (61b)$$

and this completes the proof of converse. \square

6.2. Proof of Direct Part

We first show that the rate-distortion region of the proposition that will follow is achievable. The achievability of the rate-distortion region of Theorem 1 follows by choosing then the random variable V_0 of the proposition as $V_0 = (U_0, S_2)$.

Proposition 1. *An inner bound on the rate-distortion region of the Gray-Wyner model with side information and degraded reconstruction sets of Figure 2 is given by the set of all rate-distortion quadruples (R_0, R_1, R_2, D_1) that satisfy*

$$R_0 + R_1 \geq I(V_0 U_1; S_1 S_2 | Y_1) \quad (62a)$$

$$R_0 + R_2 \geq I(V_0; S_1 S_2 | Y_2) \quad (62b)$$

$$R_0 + R_1 + R_2 \geq \max \{I(V_0; S_1 S_2 | Y_1), I(V_0; S_1 S_2 | Y_2)\} + I(U_1; S_1 S_2 | V_0 Y_1) \quad (62c)$$

for some choice of the random variables (V_0, U_1) such that $(V_0, U_1) \oplus (S_1, S_2) \oplus (Y_1, Y_2)$ and there exist functions $g_{1,1}$, $g_{2,1}$, and $g_{2,2}$ such that:

$$\hat{S}_1 = g_{1,1}(V_0, U_1, Y_1) \quad (63a)$$

$$S_2 = g_{2,1}(V_0, U_1, Y_1) \quad (63b)$$

$$S_2 = g_{2,2}(V_0, Y_2), \quad (63c)$$

and

$$\mathbb{E} d_1(S_1; \hat{S}_1) \leq D_1. \quad (64)$$

Proof of Proposition 1: We now describe a coding scheme that achieves the rate-distortion region of Proposition 1. The scheme is very similar to one that is developed by Shayevitz and Wigger [4, Theorem 2] for a Gray-Wyner model with side information. In particular, similar to [4, Theorem 2] it uses a double-binning technique for the common codebook, one that is relevant for Receiver 1 and one that is relevant for Receiver 2. Note, however, that, formally, the result of Proposition 1 cannot be obtained by readily applying [4, Theorem 2] as is; and one needs to extend the result of [4, Theorem 2] in a manner that accounts for that the source component S_2^n is to be recovered losslessly by both decoders. This can be obtained by extending the distortion measure of [4, Theorem 2] to one that is vector-valued, i.e., $d((s_1, s_2), (\hat{s}_1, \hat{s}_2)) = (d_1(s_1, \hat{s}_1), d_H(s_2, \hat{s}_2))$, where $d_H(\cdot, \cdot)$ denotes the Hamming distance. For reasons of completeness, we provide here a proof of Proposition 1.

Our scheme has the following parameters: a conditional joint pmf $P_{V_0 U_1 | S_1 S_2}$ that satisfies (63) and (64), and non-negative communication rates $T_0, T_1, T_{0,0}, T_{0,p}, T_{1,0}, T_{1,1}, \tilde{R}_{0,0}, \tilde{R}_{0,1}, \tilde{R}_{0,2}, \tilde{R}_{1,0}$ and $\tilde{R}_{1,1}$ such that

$$T_0 = T_{0,0} + T_{0,p}, \quad 0 \leq \tilde{R}_{0,0} \leq T_{0,0}, \quad 0 \leq \tilde{R}_{0,1} \leq T_{0,p}, \quad 0 \leq \tilde{R}_{0,2} \leq T_{0,p} \quad (65a)$$

$$T_1 = T_{1,0} + T_{1,1} \quad , \quad 0 \leq \tilde{R}_{1,0} \leq T_{1,0} \quad , \quad 0 \leq \tilde{R}_{1,1} \leq T_{1,1}. \quad (65b)$$

Codebook Generation

- 1) Randomly and independently generate 2^{nT_0} length- n codewords $v_0^n(k_0)$ indexed with the pair of indices $k_0 = (k_{0,0}, k_{0,p})$, where $k_{0,0} \in [1 : 2^{nT_{0,0}}]$ and $k_{0,p} \in [1 : 2^{nT_{0,p}}]$. Each codeword $v_0^n(k_0)$ has i.i.d entries drawn according to $\prod_{i=1}^n P_{V_0}(v_{0,i}(k_0))$. The codewords $\{v_0^n(k_0)\}$ are partitioned into superbins whose indices will be relevant for both receivers; and each superbin is partitioned into two different ways, each into subbins whose indices will be relevant for a distinct receiver (i.e., double-binning). This is obtained by partitioning the indices $\{(k_{0,0}, k_{0,p})\}$ as follows. We partition the $2^{nT_{0,0}}$ indices $\{k_{0,0}\}$ into $2^{n\tilde{R}_{0,0}}$ bins by randomly and independently assigning each index $k_{0,0}$ to an index $\tilde{w}_{0,0}(k_{0,0})$ according to a uniform pmf over $[1 : 2^{n\tilde{R}_{0,0}}]$. We refer to each subset of indices $\{k_{0,0}\}$ with the same index $\tilde{w}_{0,0}$ as a bin $\mathcal{B}_{00}(\tilde{w}_{0,0})$, $\tilde{w}_{0,0} \in [1 : 2^{n\tilde{R}_{0,0}}]$. Also, we make two distinct partitions of the $2^{nT_{0,p}}$ indices $\{k_{0,p}\}$, each relevant for a distinct receiver. In the first partition, which is relevant for Receiver 1, the indices $\{k_{0,p}\}$ are assigned randomly and independently each to an index $\tilde{w}_{0,1}(k_{0,p})$ according to a uniform pmf over $[1 : 2^{n\tilde{R}_{0,1}}]$. We refer to each subset of indices $\{k_{0,p}\}$ with the same index $\tilde{w}_{0,1}$ as a bin $\mathcal{B}_{01}(\tilde{w}_{0,1})$, $\tilde{w}_{0,1} \in [1 : 2^{n\tilde{R}_{0,1}}]$. Similarly, in the second partition, which is relevant for Receiver 2, the indices $\{k_{0,p}\}$ are assigned randomly and independently each to an index $\tilde{w}_{0,2}(k_{0,p})$ according to a uniform pmf over $[1 : 2^{n\tilde{R}_{0,2}}]$; and refer to each subset of indices $\{k_{0,p}\}$ with the same index $\tilde{w}_{0,2}$ as a bin $\mathcal{B}_{02}(\tilde{w}_{0,2})$, $\tilde{w}_{0,2} \in [1 : 2^{n\tilde{R}_{0,2}}]$.
- 2) For each $k_0 \in [1 : 2^{nT_0}]$, randomly and independently generate 2^{nT_1} length- n codewords $u_1^n(k_1, k_0)$ indexed with the pair of indices $k_1 = (k_{1,0}, k_{1,1})$, where $k_{1,0} \in [1 : 2^{nT_{1,0}}]$ and $k_{1,1} \in [1 : 2^{nT_{1,1}}]$. Each codeword $u_1^n(k_1, k_0)$ is with i.i.d elements drawn according to $\prod_{i=1}^n P_{U_1|V_0}(u_{1,i}(k_1, k_0)|v_{0,i}(k_0))$. We partition the $2^{nT_{1,0}}$ indices $\{k_{1,0}\}$ into $2^{n\tilde{R}_{1,0}}$ bins by randomly and independently assigning each index $k_{1,0}$ to an index $\tilde{w}_{1,0}(k_{1,0})$ according to a uniform pmf over $[1 : 2^{n\tilde{R}_{1,0}}]$. We refer to each subset of indices $\{k_{1,0}\}$ with the same index $\tilde{w}_{1,0}$ as a bin $\mathcal{B}_{10}(\tilde{w}_{1,0})$, $\tilde{w}_{1,0} \in [1 : 2^{n\tilde{R}_{1,0}}]$. Similarly, we partition the $2^{nT_{1,1}}$ indices $\{k_{1,1}\}$ into $2^{n\tilde{R}_{1,1}}$ bins by randomly and independently assigning each index $k_{1,1}$ to an index $\tilde{w}_{1,1}(k_{1,1})$ according to a uniform pmf over $[1 : 2^{n\tilde{R}_{1,1}}]$; and refer to each subset of indices $\{k_{1,1}\}$ with the same index $\tilde{w}_{1,1}$ as a bin $\mathcal{B}_{11}(\tilde{w}_{1,1})$, $\tilde{w}_{1,1} \in [1 : 2^{n\tilde{R}_{1,1}}]$.
- 3) Reveal all codebooks and its partitions to the encoder, the codebook of $\{v_0^n(k_0)\}$ and its partitions to both receivers, and the codebook of $\{u_1^n(k_1, k_0)\}$ and its partitions to only Receiver 1.

Encoding

Upon observing the source pair $(S_1^n, S_2^n) = (s_1^n, s_2^n)$, the encoder finds an index $k_0 = (k_{0,0}, k_{0,p})$ such that the codeword $v_0^n(k_0)$ is jointly typical with (s_1^n, s_2^n) , i.e.,

$$(s_1^n, s_2^n, v_0^n(k_0)) \in \mathcal{T}_{[S_1 S_2 V_0]}^{(n)}. \quad (66)$$

By the covering lemma [17, Chapter 3], the encoding in this step is successful as long as n is large and

$$T_0 \geq I(V_0; S_1 S_2). \quad (67)$$

Next, it finds an index $k_1 = (k_{1,0}, k_{1,1})$ such that the codeword $u_1^n(k_1, k_0)$ is jointly typical with the triple $(s_1^n, s_2^n, v_0^n(k_0))$, i.e.,

$$(s_1^n, s_2^n, v_0^n(k_0), u_1^n(k_1, k_0)) \in \mathcal{T}_{[S_1 S_2 V_0 U_1]}^{(n)}. \quad (68)$$

Again, by the covering lemma [17, Chapter 3], the encoding in this step is successful as long as n is large and

$$T_1 \geq I(U_1; S_1 S_2 | V_0). \quad (69)$$

Let $\tilde{w}_{0,0}$, $\tilde{w}_{0,1}$ and $\tilde{w}_{0,2}$ be the bin indices such that $k_{0,0} \in \mathcal{B}_{00}(\tilde{w}_{0,0})$, $k_{0,p} \in \mathcal{B}_{01}(\tilde{w}_{0,1})$ and $k_{0,p} \in \mathcal{B}_{02}(\tilde{w}_{0,2})$. Also, let $\tilde{w}_{1,0}$ and $\tilde{w}_{1,1}$ be the bin indices such that $k_{1,0} \in \mathcal{B}_{10}(\tilde{w}_{1,0})$ and $k_{1,1} \in \mathcal{B}_{11}(\tilde{w}_{1,1})$. The encoder then sends the product message $W_0 = (\tilde{w}_{0,0}, \tilde{w}_{1,0})$ over the error-free rate-limited common link of capacity R_0 . Also, it sends the product message $W_1 = (\tilde{w}_{0,1}, \tilde{w}_{1,1})$ over the error-free rate-limited individual link to Receiver 1 of capacity R_1 , and the message $W_2 = \tilde{w}_{0,2}$ over the error-free rate-limited individual link to Receiver 2 of capacity R_2 .

Decoding

Receiver 1 gets the messages $(W_0, W_1) = (\tilde{w}_{0,0}, \tilde{w}_{1,0}, \tilde{w}_{0,1}, \tilde{w}_{1,1})$. It seeks a codeword $v_0^n(k_0)$ and a codeword $u_1^n(k_1, k_0)$, with the indices $k_0 = (k_{0,0}, k_{0,p})$ and $k_1 = (k_{1,0}, k_{1,1})$ satisfying $k_{0,0} \in \mathcal{B}_{00}(\tilde{w}_{0,0})$, $k_{0,p} \in \mathcal{B}_{01}(\tilde{w}_{0,1})$, $k_{1,0} \in \mathcal{B}_{10}(\tilde{w}_{1,0})$ and $k_{1,1} \in \mathcal{B}_{11}(\tilde{w}_{1,1})$, and such that

$$(v_0^n(k_0), u_1^n(k_1, k_0), y_1^n) \in \mathcal{T}_{[V_0 U_1 Y_1]}^{(n)}. \quad (70)$$

By the multivariate packing lemma [17, Chapter 12], the error in this decoding step at Receiver 1 vanishes exponentially as long as n is large and

$$T_{0,0} - \tilde{R}_{0,0} + T_{0,p} - \tilde{R}_{0,1} \leq I(V_0; Y_1) \quad (71a)$$

$$T_{1,0} - \tilde{R}_{1,0} + T_{1,1} - \tilde{R}_{1,1} \leq I(U_1; Y_1 | V_0). \quad (71b)$$

445

Receiver 1 then sets its reproduced codewords $\hat{s}_{2,1}^n$ and \hat{s}_1^n respectively as

$$\hat{s}_{2,1}^n = g_{2,1}(v_0^n(k_0), u_1^n(k_1, k_0), y_1^n) \quad (72a)$$

$$\hat{s}_1^n = g_1(v_0^n(k_0), u_1^n(k_1, k_0), y_1^n). \quad (72b)$$

Similary, Receiver 2 gets the message $(W_0, W_2) = (\tilde{w}_{0,0}, \tilde{w}_{1,0}, \tilde{w}_{0,2})$. It seeks a codeword $v_0^n(k_0)$, with $k_0 = (k_{0,0}, k_{0,p})$ satisfying $k_{0,0} \in \mathcal{B}_{00}(\tilde{w}_{0,0})$ and $k_{0,p} \in \mathcal{B}_{02}(\tilde{w}_{0,2})$, and such that

$$(v_0^n(k_0), y_1^n) \in \mathcal{T}_{[V_0 Y_2]}^{(n)}. \quad (73)$$

Again, using the multivariate packing lemma [17, Chapter 12], the error in this decoding step at Receiver 2 vanishes exponentially as long as n is large and

$$T_{0,0} - \tilde{R}_{0,0} + T_{0,p} - \tilde{R}_{0,2} \leq I(V_0; Y_2). \quad (74)$$

Receiver 2 then sets its reconstructed codeword $\hat{s}_{2,1}^n$ as

$$\hat{s}_{2,2}^n = g_{2,2}(v_0^n(k_0), y_1^n). \quad (75)$$

Summarizing, combining (67), (69), (71) and (74), the communication rates T_0 , T_1 , $T_{0,0}$, $T_{0,p}$, $T_{1,0}$, $T_{1,1}$, $\tilde{R}_{0,0}$, $\tilde{R}_{0,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$ and $\tilde{R}_{1,1}$ satisfy the following inequalities

$$T_0 \geq I(V_0; S_1 S_2) \quad (76a)$$

$$T_1 \geq I(U_1; S_1 S_2 | V_0) \quad (76b)$$

$$T_{0,0} - \tilde{R}_{0,0} + T_{0,p} - \tilde{R}_{0,1} \leq I(V_0; Y_1) \quad (76c)$$

$$T_{0,0} - \tilde{R}_{0,0} + T_{0,p} - \tilde{R}_{0,2} \leq I(V_0; Y_2) \quad (76d)$$

$$T_{1,0} - \tilde{R}_{1,0} + T_{1,1} - \tilde{R}_{1,1} \leq I(U_1; Y_1 | V_0). \quad (76e)$$

Choosing $\tilde{R}_{0,0}$, $\tilde{R}_{1,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$ and $\tilde{R}_{1,1}$ to also satisfy the rate relations

$$R_0 = \tilde{R}_{0,0} + \tilde{R}_{1,0} \quad (77a)$$

$$R_1 = \tilde{R}_{0,1} + \tilde{R}_{1,1} \quad (77b)$$

$$R_2 = \tilde{R}_{0,2}. \quad (77c)$$

and, finally, using Fourier-Motzkin elimination (FME) to successively project out the nuisance variables $T_{0,0}$, $T_{0,p}$, $T_{1,0}$, $T_{1,1}$, T_0 , T_1 , and then $\tilde{R}_{0,0}$, $\tilde{R}_{0,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$ and $\tilde{R}_{1,1}$ from the set of relations formed by (65), (76) and (77), we get the region of Proposition 1.

This completes the proof of the proposition; and so that of the direct part of Theorem 1.

References

1. Gray, R.; Wyner, A. Source coding for a simple network. *Bell System Technical Journal* **1974**, *53*, 1681–1721.
2. Heegard, C.; Berger, T. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory* **1985**, *31*, 727–734.
3. Tian, C.; Diggavi, S.N. Side-information scalable source coding. *Information Theory, IEEE Transactions on* **2008**, *54*, 5591–5608.
4. Shayevitz, O.; Wigger, M. On the capacity of the discrete memoryless broadcast channel with feedback. *IEEE Transactions on Information Theory* **2013**, *59*, 1329–1345.
5. Heegard, C.; Berger, T. Rate distortion when side information may be absent. *Information Theory, IEEE Transactions on* **1985**, *31*, 727–734.
6. Kaspi, A.H. Rate distortion function when side information may be present at the decoder. *IEEE Transactions on Information Theory* **1994**, *40*, 2031–2034.
7. Sgarro, A. Source coding with side information at several decoders. *Information Theory, IEEE Transactions on* **1977**, *23*, 179–182.
8. Tian, C.; Diggavi, S.N. On multistage successive refinement for Wyner–Ziv source coding with degraded side informations. *Information Theory, IEEE Transactions on* **2007**, *53*, 2946–2960.
9. Timo, R.; Oechtering, T.; Wigger, M. Source Coding Problems With Conditionally Less Noisy Side Information. *Information Theory, IEEE Transactions on* **2014**, *60*, 5516–5532.
10. Benammar, M.; Zaidi, A. Rate-distortion function for a heegard-berger problem with two sources and degraded reconstruction sets. *IEEE Transactions on Information Theory* **2016**, *62*, 5080–5092.
11. Timo, R.; Grant, A.; Kramer, G. Rate-distortion functions for source coding with complementary side information. *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 2934–2938.
12. Unal, S.; Wagner, A. An LP bound for rate distortion with variable side information. *Information Theory Proceedings (ISIT), IEEE International Symposium on*, 2016.
13. Equitz, W.H.; Cover, T.M. Successive refinement of information. *IEEE Transactions on Information Theory* **1991**, *37*, 269–275.
14. Steinberg, Y.; Merhav, N. On successive refinement for the Wyner–Ziv problem. *IEEE Transactions on Information Theory* **2004**, *50*, 1636–1654.
15. Timo, R.; Chan, T.; Grant, A. Rate distortion with side-information at many decoders. *Information Theory, IEEE Transactions on* **2011**, *57*, 5240–5257.
16. Timo, R.; Grant, A.; Chan, T.; Kramer, G. Source coding for a simple network with receiver side information. *2008 IEEE International Symposium on Information Theory*. IEEE, 2008, pp. 2307–2311.
17. Gamal, A.E.; Kim, Y.H. *Network information theory*; Cambridge University Press, 2011.