

# Common Reconstructions in the Successive Refinement Problem with Receiver Side Information

Badri N. Vellambi, *Senior Member, IEEE* and Roy Timo, *Member, IEEE*

## Abstract

We study a variant of the successive refinement problem with receiver side information where the receivers require identical reconstructions. We present general inner and outer bounds for the rate region for this variant and present a single-letter characterization of the admissible rate region for several classes of the joint distribution of the source and the side information. The characterization indicates that the side information can be fully used to reduce the communication rates via binning; however, the reconstruction functions can depend only on the Gács-Körner common randomness shared by the two receivers. Unlike existing (inner and outer) bounds to the rate region of the general successive refinement problem, the characterization of the admissible rate region derived for several settings of the variant studied requires only one auxiliary random variable. Using the derived characterization, we establish that the admissible rate region is not continuous in the underlying source source distribution even though the problem formulation does not involve zero-error or functional reconstruction constraints.

## I. INTRODUCTION

**T**HIS paper considers a *common-receiver reconstructions (CRR)* variant of the *successive refinement* problem with receiver side information where the source reconstructions at the receivers are required to be identical (almost always). An encoder is required to compress the output of a discrete memoryless source (DMS) into two messages:

- a *common* message that is reliably delivered to both receivers, and
- a *private* message that is reliably delivered to one receiver.

Each receiver has some side information jointly correlated with the source (and the other receiver's side information), and is required to output source reconstruction that meets a certain fidelity requirement. The CRR condition requires that these reconstructions be identical to one another.

The CRR problem described above can be viewed as an abstraction of a communication scenario that could arise when conveying data (e.g., meteorological or geological survey data, or an MRI scan) over a network for storage in separate data clusters storing (past) records of the data. The records, which serve as side information, could be an earlier survey data or a previous scan, depending on the specific application. The framework considered here is the source coding problem that arises when data is to be communicated over a degraded broadcast channel to two receivers that have prior side information, and the three terminals (the transmitter and the two receivers) use a separate source-channel coding paradigm [1].

The problem of characterizing the *achievable rate-distortion region* of the general successive refinement problem with receiver side information is open [2]–[4]. The version of the successive refinement problem where the private message is absent, known as the Heegard-Berger problem, is also open [4]–[6]. However, complete characterization exists for specific settings of both successive refinement and Heegard-Berger problems. For example, the rate region of the successive refinement problem is known when the side information of the receiver that receives one message is a degraded version of side information of the other receiver [2]. Similarly, the Heegard-Berger problem has been solved when the side information is degraded [5], mismatched degraded [7], or conditionally less noisy [8]. Additionally, the HB problem has also been solved under list decoding constraints (closely related to logarithmic-loss distortion functions) [9], degraded message sets [10], and many vector Gaussian formulations [11], [12].

B. N. Vellambi is with the New Jersey Institute of Technology, Newark, NJ 07102, USA, badri.n.vellambi@ieee.org.

R. Timo was with the Institute for Communications Engineering at the Technical University of Munich. He is now with Ericsson Research, Stockholm, roy.timo@ericsson.com.

Some of the material in this paper was presented at the 2013 IEEE Information Theory Workshop in Seville, Spain, and the 2014 IEEE International Symposium on Information Theory in Honolulu, USA, 2014. This work was supported by the Australian Research Council's Discovery Project DP120102123.

The common reconstruction variant of the Wyner-Ziv problem was first motivated and solved by Steinberg [13]. Common reconstructions in other problems were then considered in [14]. Benammar and Zaidi recently considered the HB problem under a three-way common reconstructions condition with degraded message sets [10], [15]. In our previous work [16], we characterized the rate region for several cases of the HB problem with the CRR requirement. In this work, we present single-letter inner and outer bounds for the rate region of the successive refinement problem with receiver side information and the CRR requirement (termed as the SR-CRR problem). For several specific cases of the underlying joint distribution between the source and the side information random variables (including those in our previous work [17]), we prove that the inner and outer bounds match, and therefore yield a characterization of the rate region.

The characterization indicates that while the receiver side information can be fully utilized for reducing the communication rate by means of binning, only the Gács-Körner common randomness between the random variables (i.e., both auxiliary and side information) available to the two receivers can be used for generating the reconstructions. This feature is also seen in our characterization for the HB problem with the CRR requirement in [16]. This single-letter characterization for the rate region of the SR-CRR problem derived in this work is *unique* in the sense that it is the first rate region formulation where the Gács-Körner common randomness explicitly appears in the single-letter constraint corresponding to the receiver source reconstructions. Unlike the best-known bounds for the successive refinement problem, the characterization of the SR-CRR rate region (when the source satisfies a certain support condition) requires only one auxiliary random variable that is decoded by both receivers. Thus, the CRR requirement obviates the need for a second auxiliary random variable to absorb the private message.

The paper is organized as follows. Section II-A introduces some basic notation; Section II-B reviews the concept of Gács-Körner common randomness; and Section II-C formally defines the successive refinement problem with the common receiver reconstruction constraint. The characterization of the paper's main contributions are summarized in Section III, including the single-letter characterization of the rate region, and the proof of the discontinuity of the characterization with the source distribution. The reader will be directed to the respective appendices for the proofs of the results contained in Section III. Finally, Section IV concludes this work.

## II. NOTATION, GÁCS-KÖRNER COMMON RANDOMNESS, AND PROBLEM SETUP

### A. Notation

Let  $\mathbb{N} = \{1, 2, \dots\}$  denote the natural numbers. Uppercase letters (e.g.,  $S, U, V$ ) represent random variables (RVs), and the script versions (e.g.,  $\mathcal{S}, \mathcal{U}, \mathcal{V}$ ) denote the corresponding alphabets. All alphabets in this work are assumed to be finite. Realizations of RVs are given by lowercase letters (e.g.,  $s, u, v$ ). For RVs  $A, B, C$ , we denote  $A - B - C$  if they form a Markov chain. Given RVs  $A$  and  $B$ , we let  $A \equiv B$  if and only if

$$H(A|B) = H(B|A) = 0. \quad (1)$$

Given  $k$  jointly correlated random variables  $A_1, \dots, A_k$ , their support set is defined by

$$\mathcal{S}(A_1, \dots, A_k) = \{(a_1, \dots, a_k) \in \mathcal{A}_1 \times \dots \times \mathcal{A}_k : p_{A_1 \dots A_k}(a_1, \dots, a_k) > 0\}.$$

For a set  $T$ , and  $a, b \in T$ , we let

$$\mathbb{1}\{a = b\} := \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases}, \quad (2)$$

and let  $\bar{\mathbb{1}}\{a = b\} := 1 - \mathbb{1}\{a = b\}$ . Vectors are indicated by superscripts, and their components by subscripts; for example,  $x^n := (x_1, \dots, x_n)$  and  $x^{n \setminus i} := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ . We will use  $\mathfrak{m}$  to denote the least positive probability mass over the support of random variables; for example, given a joint probability mass function (henceforth, pmf)  $p_{A_1 \dots A_k}$ ,

$$\mathfrak{m}_{A_1} = \min \{p_{A_1}(a) : a \in \mathcal{S}(A_1)\}, \quad (3)$$

$$\mathfrak{m}_{A_1 \dots A_k} = \min \{p_{A_1 \dots A_k}(a_1, \dots, a_k) : (a_1, \dots, a_k) \in \mathcal{S}(A_1, \dots, A_k)\}, \quad (4)$$

$$\mathfrak{m}_{A_1|A_2} = \min \{p_{A_1|A_2}(a_1|a_2) : (a_1, a_2) \in \mathcal{S}(A_1, A_2)\}. \quad (5)$$

The probability of an event  $E$  is denoted by  $\mathbb{P}(E)$ , and  $\mathbb{E}$  denotes the expectation operator. Lastly, for  $\varepsilon > 0$ , the set of  $\varepsilon$ -letter-typical sequences of length  $n$  according to pmf  $p_X$  is denoted by  $T_\varepsilon^n[p_X]$  [18].

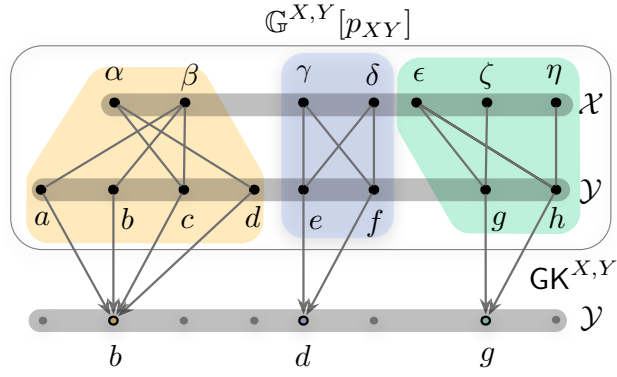


Fig. 1. Illustration of  $\mathbb{G}^{X,Y}[p_{XY}]$  and  $\text{GK}^{X,Y}$ .

### B. Gács-Körner Common Randomness

Given two jointly correlated random variables  $X$  and  $Y$ , the Gács and Körner's *common randomness* between  $X$  and  $Y$  [19] is the random variable  $Z$  with the largest entropy such that  $H(Z|X) = H(Z|Y) = 0$ . This notion of common randomness will play a key role in this paper. To define this notion of common randomness we introduce the following terminology. Given  $(X, Y) \sim p_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$ , let  $\mathbb{G}^{X,Y}[p_{XY}]$  denote the bipartite graph with left nodes  $\mathcal{X}$ , right nodes  $\mathcal{Y}$ , and edges between  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$  if and only if  $p_{XY}(x, y) > 0$ . Now define an equivalence relation on  $\mathcal{Y}$  by  $y_1 \rightleftharpoons y_2$  if and only if they are in the same connected component of  $\mathbb{G}^{X,Y}[p_{XY}]$ . Finally, let

$$\text{GK}^{X,Y} : \mathcal{Y} \rightarrow \mathcal{Y} \quad (6)$$

be any mapping satisfying

$$\text{GK}^{X,Y}(y_1) = \text{GK}^{X,Y}(y_2) \quad \text{iff} \quad y_1 \rightleftharpoons y_2. \quad (7)$$

Of course, there are multiple choices for the *Gács-Körner* mapping in (6). However, all such choices are equivalent in the sense that if  $\text{GK}_1^{X,Y}$  and  $\text{GK}_2^{X,Y}$  satisfy (7) then

$$\text{GK}_1^{X,Y}(Y) \equiv \text{GK}_2^{X,Y}(Y). \quad (8)$$

As an illustration, let  $\mathcal{X} = \{a, b, c, d, e, f, g, h\}$  and let  $\mathcal{Y} = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta\}$ . Consider the following pmf  $p_{XY}$ .

$$p_{XY} = \frac{1}{20} \begin{bmatrix} \begin{array}{cccc|cccc} a & b & c & d & e & f & g & h \\ \hline 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{array} \end{bmatrix} \begin{array}{l} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \zeta \\ \eta \end{array} \quad (9)$$

Figure 1 illustrates the bipartite graph representation of  $p_{XY}$ , and depicts one possible choice for the Gács-Körner mapping  $\text{GK}^{X,Y}$  satisfying the requirement in (7). Notice that that  $\mathbb{G}^{X,Y}[p_{XY}]$  contains three connected components, and hence  $\text{GK}^{X,Y}$  is a ternary RV taking values in  $\{b, d, g\}$  if the chosen mapping is the one illustrated in the figure. Note that for this pmf, there are  $4 \times 2 \times 2 = 16$  equivalent choices for the mapping.

From the definition above, two properties of the Gács and Körner's common randomness are evident.

- The Gács and Körner's common randomness between two random variables is *symmetric*, i.e.,

$$\text{GK}^{X,Y}(Y) \equiv \text{GK}^{Y,X}(X). \quad (10)$$

Note however that the above two Gács and Körner's common randomness variables take values over different alphabets even though each is a function of the other, i.e.,  $\text{GK}^{X,Y}(Y)$  and  $\text{GK}^{Y,X}(X)$  are random variables over  $\mathcal{Y}$  and  $\mathcal{X}$ , respectively.

- Since the Gács and Körner's common randomness depends on the pmf  $p_{XY}$  only through the bipartite graph, the Gács and Körner's common randomness between RVs computed using two pmfs  $p_{XY}$  and  $q_{XY}$  are identical if: (a)  $\mathbb{G}^{X,Y}[p_{XY}]$  and  $\mathbb{G}^{X,Y}[q_{XY}]$  are graph-isomorphic; and (b) the probabilities of the components of  $p_{XY}$  and that of  $q_{XY}$  are permutations of one another. To illustrate this, consider  $p_{XY}$  of (9) and the following pmf  $q_{XY}$ .

$$q_{XY} = \frac{1}{20} \begin{array}{c} \begin{array}{cccc|cc} a & b & c & d & e & f & g & h \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{l} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \zeta \\ \eta \end{array} \quad (11)$$

The Gács and Körner's common randomness depends between  $X$  and  $Y$  computed using either  $p_{XY}$  or  $q_{XY}$  yields a ternary random variable with the pmf  $[\frac{7}{20} \ \frac{3}{10} \ \frac{7}{20}]$ .

In the remainder of this work, for a given pmf  $p_{XY}$  over  $\mathcal{X} \times \mathcal{Y}$ , we will assume an arbitrary but fixed choice for the Gács and Körner's common randomness mapping that satisfies (7) without explicitly specifying this mapping. On account of notational ease, we will also drop the argument, and abbreviate it by

$$\text{GK}^{X,Y} := \text{GK}^{X,Y}(Y). \quad (12)$$

It is to be assumed that the argument is always the second random variable in the superscript, and consequently, the Gács and Körner's common randomness  $\text{GK}^{X,Y}$  is a random variable over the alphabet  $\mathcal{Y}$  of the second variable. This is, quite simply, a only a *notational bias*, since Gács and Körner's common randomness is indeed symmetric.

Before we proceed to formally state the problem investigated and the main results, we present two results that pertain solely to Gács-Körner common randomness that we will need in the main section of this work. Both results present decompositions of the Gács-Körner common randomness between two random variables when additional information about the support of the joint pmf of the two variables is known.

*Lemma 1:* Suppose that the support set of  $(A_1, A_2, U, V) \sim q_{A_1 A_2 U V}$  satisfies

$$\text{S}(A_1, A_2, U, V) = \text{S}(A_1, A_2) \times \text{S}(U, V). \quad (13)$$

Then,

$$\text{GK}^{A_1 U, A_2 V} \equiv (\text{GK}^{A_1, A_2}, \text{GK}^{U, V}). \quad (14)$$

*Proof:* The proof can be found in Appendix A. ■

*Lemma 2:* If  $(X, Y, Z) \sim q_{XYZ}$  satisfies

$$\text{S}(X, Y, Z) = \text{S}(X, Y) \times \text{S}(Z), \quad (15)$$

then

$$\text{GK}^{X, YZ} \equiv \text{GK}^{X, Y}. \quad (16)$$

*Proof:* Define a constant random variable  $W$  over a singleton alphabet, say  $\{w\}$  and let  $q_W$  denote the degenerate pmf of  $W$ . Define  $q_{XYZ} = q_W q_{XYZ}$ . Then, one can see that

$$\text{S}(W, X, Y, Z) = \text{S}(X, Y, Z) \times \text{S}(W) = \text{S}(X, Y) \times \text{S}(Z) \times \text{S}(W) \quad (17)$$

$$= \text{S}(X, Y) \times \text{S}(Z, W). \quad (18)$$

Then, an application of Lemma 1 yields

$$\text{GK}^{X, YZ} \equiv (\text{GK}^{X, Y}, \text{GK}(Z, W)) \stackrel{(a)}{\equiv} \text{GK}^{X, Y}, \quad (19)$$

where (a) follows since  $W$  is a constant RV. ■

### C. Problem Setup — Successive Refinement with the CRR Constraint

Let pmf  $p_{SUV}$  on  $\mathcal{S} \times \mathcal{U} \times \mathcal{V}$ , a reconstruction alphabet  $\hat{\mathcal{S}}$ , and a bounded distortion function  $d : \mathcal{S} \times \hat{\mathcal{S}} \rightarrow [0, \bar{D}]$  be given. We assume that  $\bar{D} \in (0, \infty)$ . A DMS  $(S, U, V) \sim p_{SUV}$  emits an i.i.d. sequence

$$(S^n, U^n, V^n) = (S_1, U_1, V_1), (S_2, U_2, V_2), \dots, (S_n, U_n, V_n). \quad (20)$$

As illustrated in Fig. 2, the engineering problem is to encode the *source*  $S^n$  into a common message  $M_{uv}$  communicated to both receivers, and a private message  $M_v$  communicated only the receiver having the side information  $V^n$  so that the following three conditions are satisfied:

- 1) A receiver with access to the *side information*  $U^n$  and the common message  $M_{uv}$  can output an estimate  $\hat{S}^n$  of  $S^n$  to within a prescribed average (per-letter) distortion  $D$ .
- 2) A receiver with access to the *side information*  $V^n$  and both (common and private) messages  $M_{uv}$  and  $M_v$  can output an estimate  $\tilde{S}^n$  of  $S^n$  to within a prescribed average (per-letter) distortion  $D$ .
- 3) The estimates  $\hat{S}^n$  and  $\tilde{S}^n$  (both defined on the set  $\hat{\mathcal{S}}$ ) are identical to one another almost always.

The aim then is to characterize the rates of the common and private messages that need to be communicated to achieve the above requirements. The following definition formally defines the problem.

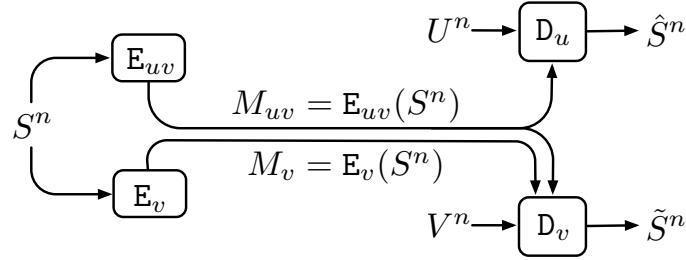


Fig. 2. Successive refinement with receiver side information and common receiver reconstructions (CRR).

**Definition 1:** Fix  $D \geq 0$ . We say that a rate pair  $(r_{uv}, r_v) \in [0, \infty)^2$  is  $D$ -admissible if for each  $\varepsilon > 0$  there exist a sufficiently large blocklength  $n \in \mathbb{N}$ , and:

(a) two encoders

$$E_{uv} : \mathcal{S}^n \rightarrow \{1, 2, \dots, \lceil 2^{n(r_{uv} + \varepsilon)} \rceil\}, \quad (21a)$$

$$E_v : \mathcal{S}^n \rightarrow \{1, 2, \dots, \lceil 2^{n(r_v + \varepsilon)} \rceil\}; \quad (21b)$$

(b) and two decoders

$$D_u : \{1, 2, \dots, \lceil 2^{n(r_{uv} + \varepsilon)} \rceil\} \times \mathcal{U}^n \rightarrow \hat{\mathcal{S}}^n, \quad (21c)$$

$$D_v : \{1, 2, \dots, \lceil 2^{n(r_{uv} + \varepsilon)} \rceil\} \times \{1, \dots, \lceil 2^{n(r_v + \varepsilon)} \rceil\} \times \mathcal{V}^n \rightarrow \hat{\mathcal{S}}^n, \quad (21d)$$

such that the reconstructions

$$\hat{S}^n := D_u(E_{uv}(S^n), U^n), \quad (22a)$$

$$\tilde{S}^n := D_v(E_{uv}(S^n), E_v(S^n), V^n), \quad (22b)$$

satisfy:

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [d(S_i, \hat{S}_i)] \leq D + \varepsilon, \quad (23a)$$

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [d(S_i, \tilde{S}_i)] \leq D + \varepsilon, \quad (23b)$$

$$\mathbb{P}[\hat{S}^n \neq \tilde{S}^n] \leq \varepsilon. \quad (23c)$$

*Definition 2:* The  $D$ -admissible rate region  $\mathcal{R}(D)$  of the successive refinement problem with CRR is the set of all  $D$ -admissible rate pairs.

The main problem of interest in this paper is to characterize the  $D$ -admissible rate region  $\mathcal{R}(D)$ . Define  $\underline{D}$  by

$$\underline{D} := \min_{\phi: \mathcal{S} \rightarrow \hat{\mathcal{S}}} \mathbb{E} d(S, \phi(S)). \quad (24)$$

On one hand, if  $0 \leq D < \underline{D}$ , then  $\mathcal{R}(D) = \emptyset$ . On the other, if  $D > \overline{D}$ , then  $\mathcal{R}(D) = [0, \infty)^2$ . For the interesting interval of non-trivial values of  $D \in [\underline{D}, \overline{D}]$ , the rate region  $\mathcal{R}(D)$  is a closed and convex subset of  $[0, \infty)^2$ . This nontrivial interval of values of  $D$  will be the subject of our investigation for the remainder of the paper.

### III. MAIN RESULTS

In this section, we will present inner (achievability) and outer (converse) bounds for the  $D$ -admissible rate region  $\mathcal{R}(D)$ , and we show that these bounds are tight in a variety of nontrivial settings. We will characterize these bounds through the following three rate regions defined over three corresponding spaces of auxiliary random variable pmfs.

#### A. Three single-letter rate regions and their properties

*Definition 3:* For  $k \in \mathbb{N}$ , let  $\mathcal{P}_{D,k}^*$  denote the set of all pmfs  $q_{ASUV}$  defined on  $\mathcal{A} \times \mathcal{S} \times \mathcal{U} \times \mathcal{V}$  such that  $(A, S, U, V) \sim q_{ASUV}$  satisfies the following conditions:

- (i)  $q_{SUV} = p_{SUV}$ ;
- (ii)  $|\mathcal{A}| \leq k$ ;
- (iii)  $A - S - (U, V)$ ; and
- (iv) there exists a function  $f: \mathcal{A} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  for which

$$\mathbb{E} [d(S, f(\text{GK}^{AU, AV}))] \leq D. \quad (25)$$

*Definition 4:* Let  $\mathcal{R}_k^*(D)$  denote the set of all rate pairs  $(r_{uv}, r_v) \in [0, \infty)^2$  satisfying

$$r_{uv} \geq I(S; A|U) \quad (26a)$$

$$r_v + r_{uv} \geq I(S; A|V) \quad (26b)$$

for some  $(A, S, U, V) \sim q_{ASUV} \in \mathcal{P}_{D,k}^*$ .

*Definition 5:* For  $k \in \mathbb{N}$ , let  $\mathcal{P}_{D,k}^\dagger$  denote the set of all pmfs  $q_{ABCSUV}$  defined on  $\mathcal{A} \times \mathcal{B} \times \mathcal{C} \times \mathcal{S} \times \mathcal{U} \times \mathcal{V}$  such that  $(A, B, C, S, U, V) \sim q_{ABCSUV}$  satisfies the following constraints:

- (i)  $q_{SUV} = p_{SUV}$ ;
- (ii)  $\max\{|\mathcal{A}|, |\mathcal{B}|, |\mathcal{C}|\} \leq k$ ;
- (iii)  $(A, B, C) - S - (U, V)$ ; and
- (iv) there exists a function  $f: \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  for which

$$\mathbb{E} [d(S, f(\text{GK}^{ABU, ACV}))] \leq D. \quad (27)$$

*Definition 6:* Let  $\mathcal{R}_k^\dagger(D)$  denote the set of all rate pairs  $(r_{uv}, r_v) \in [0, \infty)^2$  satisfying

$$r_{uv} \geq I(S; A, B|U), \quad (28a)$$

$$r_v + r_{uv} \geq I(S; B|A, C, U, V) + I(S; A, C|V) \quad (28b)$$

for some  $(A, B, C, S, U, V) \sim q_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$ .

*Definition 7:* For  $k \in \mathbb{N}$ , let

$$\mathcal{P}_{D,k}^\ddagger = \left\{ q_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger : B - (A, S) - C \right\}. \quad (29)$$

*Definition 8:* Let  $\mathcal{R}_k^\ddagger(D)$  denote the set of all rate pairs  $(r_{uv}, r_v) \in [0, \infty)^2$  satisfying

$$r_{uv} \geq I(S; A, B|U) \quad (30a)$$

$$r_v + r_{uv} \geq \max\{I(S; A|U), I(S; A|V)\} + I(S; B|A, U) + I(S; C|A, V) \quad (30b)$$

for some  $(A, B, C, S, U, V) \sim q_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$ .

We can establish the following preliminary inclusions between the three rate regions defined above.

*Lemma 3:* For any  $k \in \mathbb{N}$ ,  $\mathcal{R}_k^*(D) \subseteq \mathcal{R}_k^\dagger(D)$ .

*Proof:* By simply choosing  $B = C = A$  with  $|A| \leq k$ , we cover all rate pairs that line in  $\mathcal{R}_k^*(D)$ , and hence,  $\mathcal{R}_k^*(D)$  is a subset of  $\mathcal{R}_k^\dagger(D)$ . ■

*Lemma 4:* For any  $k \in \mathbb{N}$ ,  $\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}_k^\dagger(D)$ .

*Proof:* First, note that  $\mathcal{P}_{D,k}^\dagger \subseteq \mathcal{P}_{D,k}^\dagger$ . So we are done if we show that the RHS of (28b) is numerically smaller than that of (30b) for any pmf in  $\mathcal{P}_{D,k}^\dagger$ . To do that, pick  $p_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$  and consider the following argument.

$$\begin{aligned} I(S; B|A, C, U, V) + I(S; A, C|V) &= H(B|A, C, U, V) - H(B|A, C, S, U, V) + I(S; A, C|V) \\ &\stackrel{(a)}{=} H(B|A, C, U, V) - H(B|A, C, S) + I(S; A, C|V) \\ &\stackrel{(b)}{=} H(B|A, C, U, V) - H(B|A, S) + I(S; A, C|V) \\ &\stackrel{(c)}{\leq} H(B|A, U) - H(B|A, S) + I(S; A, C|V) \\ &\stackrel{(d)}{=} I(S; B|A, U) + I(S; A, C|U), \end{aligned} \tag{31}$$

where (a) follows from the chain  $(A, B, C) - S - (U, V)$ ; (b) follows from the chain  $B - (A, S) - C$ ; (c) follows by dropping variables in the conditioning; and finally (d) follows by reintroducing  $U$  in the second term of (31) without affecting the numerically affecting the terms. Finally, the claim follows by noting that (32) is bounded below by the RHS of (30b) thereby completing the proof of this claim. ■

While the above two inclusions hold true for all DMSs  $p_{SUV}$ , we can establish stronger results if we know something more about  $p_{SUV}$ . In specific, if we know that the pmf  $p_{SUV}$  satisfies the full-support condition of (33), then the following *reverse* inclusion also holds albeit with some alphabet size readjustment. In other words, when the full-support condition is met, any rate pair that meets (28a) and (28b) (with auxiliary RVs  $A, B$ , and  $C$ ) also meets (26a) and (26b) for a different auxiliary RV  $A$  with an appropriately larger alphabet.

*Lemma 5:* If the support of  $(S, U, V) \sim p_{SUV}$  satisfies

$$\mathcal{S}(S, U, V) = \mathcal{S}(S) \times \mathcal{S}(U, V), \tag{33}$$

then

$$\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}_{k^2}^*(D), \quad \forall k \in \mathbb{N}. \tag{34}$$

*Proof:* Lemma 5 is proved in Appendix B. ■

Observe that each of the three rate regions defined above (Definitions 4, 6, and 8) can potentially be enlarged by merely increasing  $k$ . In other words, we are guaranteed to have

$$\mathcal{R}_1^\lambda(D) \subseteq \mathcal{R}_2^\lambda(D) \subseteq \mathcal{R}_3^\lambda(D) \subseteq \cdots, \quad \lambda \in \{\star, \dagger, \dagger\}. \tag{35}$$

Since we do not impose restrictions on how we encode the source, we can allow the alphabet sizes of the auxiliary RVs to be finite but arbitrarily large. Hence, it makes sense to introduce the following notation for the limiting rate regions allowing the alphabets of auxiliary random variables to be any finite set.

*Definition 9:*

$$\mathcal{R}^\lambda(D) := \bigcup_{k \in \mathbb{N}} \mathcal{R}_k^\lambda(D), \quad \lambda \in \{\star, \dagger, \dagger\}. \tag{36}$$

However, from a computational point of view, it is preferable that the sequence of sets in (35) not grow indefinitely with  $k$ . The following result ensures that this, indeed, does not happen. It quantifies the bounds on the alphabet size of the auxiliary random variables beyond which there is no strict enlargement of these regions.

*Lemma 6:* For all integers  $k \in \mathbb{N}$ , we have the following.

$$\mathcal{R}_k^*(D) \subseteq \mathcal{R}^*(D) = \mathcal{R}_{|S|+2}^*(D), \tag{37a}$$

$$\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}^\dagger(D) = \mathcal{R}_{|S|(|S|+6)|\hat{S}|^{|U|+4}}^\dagger(D), \tag{37b}$$

$$\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}^\dagger(D) = \mathcal{R}_{|S|(|S|+4)|\hat{S}|^{|U|+1}}^\dagger(D). \tag{37c}$$

*Proof:* The proof for claim for  $\mathcal{R}^\dagger(D)$  is presented in detail in Appendix C. The proof for (37a) and (37c) is almost identical to that of (37b), and the difference are highlighted in Remarks 5 and 6 in Appendix C. ■

We conclude this section with two properties of the above three regions.

*Lemma 7:* The regions  $\mathcal{R}^*(D)$ ,  $\mathcal{R}^\dagger(D)$ , and  $\mathcal{R}^\ddagger(D)$  are convex.

*Proof:* A proof of the claim for  $\mathcal{R}^*(D)$  can be found in Appendix D. The proofs of the convexity of the other two regions are identical, and are omitted. ■

*Lemma 8:* Let for each  $i \in \mathbb{N}$ ,  $D_i \in [\underline{D}, \overline{D}]$  and  $(r_{uv}^{(i)}, r_v^{(i)}) \in \mathcal{R}_{D_i, k}^*$  be given. Suppose that  $\lim_{i \rightarrow \infty} D_i = D$ , and  $\lim_{i \rightarrow \infty} (r_{uv}^{(i)}, r_v^{(i)}) = (r_{uv}, r_v)$ . Then,  $(r_{uv}, r_v) \in \mathcal{R}_{D, k}^*$ .

*Proof:* A proof of the claim for  $\mathcal{R}^*(D)$  can be found in Appendix E. The proofs for the other two regions are identical, and are omitted. ■

By simply choosing  $D_i = D \in [\underline{D}, \overline{D}]$ ,  $i \in \mathbb{N}$ , we obtain the following result.

*Remark 1:* For any  $D \in [\underline{D}, \overline{D}]$ , the regions  $\mathcal{R}^*(D)$ ,  $\mathcal{R}^\dagger(D)$ , and  $\mathcal{R}^\ddagger(D)$  are topologically closed.

### B. A single-letter characterization for $\mathcal{R}(D)$

In this section, we present our main results on the single-letter characterization of the  $D$ -admissible rate region. The first two present inner and outer bounds sandwiching the  $D$ -admissible rate region using the three limiting rate regions given in Definitions 4, 6, and 8.

*Theorem 1:* For any  $D \in [\underline{D}, \overline{D}]$ , the regions  $\mathcal{R}^*(D)$  and  $\mathcal{R}^\dagger(D)$  are inner bounds to the  $D$ -admissible rate region  $\mathcal{R}(D)$  of the successive refinement problem with the CRR constraint, i.e.,

$$\mathcal{R}^*(D) \stackrel{(a)}{\subseteq} \mathcal{R}^\dagger(D) \stackrel{(b)}{\subseteq} \mathcal{R}(D). \quad (38)$$

*Proof:* The inclusion (a) follows from Lemmas 3 and 6 above, and a proof of the inclusion (b) can be found in Appendix F. ■

*Theorem 2:* For any  $D \in [\underline{D}, \overline{D}]$ , the rate region  $\mathcal{R}^\ddagger(D)$  is an outer bound to the  $D$ -admissible rate region  $\mathcal{R}(D)$  of the successive refinement problem with the CRR constraint, i.e.,

$$\mathcal{R}(D) \stackrel{(b)}{\subseteq} \mathcal{R}^\ddagger(D).$$

*Proof:* The proof of the inclusion in (a) can be found in Appendix G. ■

In the absence of the CRR constraint, Steinberg and Merhav's original solution to the *physically-degraded side information* version of the successive refinement problem required three auxiliary random variables (later simplified to two by Tian and Diggavi [20]) and two reconstruction functions. Benammar and Zaidi's solution to their formulation of the successive refinement problem with a common source reconstruction required two auxiliary random variables and a reconstruction function [10], [15]. The following result, which is the main result in this work, establishes a single-letter characterization of the  $D$ -admissible rate region  $\mathcal{R}(D)$  for several cases of side information. Unlike other characterizations, the rate region  $\mathcal{R}(D)$  is completely described by a single auxiliary random variable, and a single reconstruction function whose argument is not the side information and the auxiliary random variable, but the Gács-Körner common randomness shared by the two receivers.

*Theorem 3:* If the DMS  $(S, U, V) \sim p_{SUV}$  falls into one of the following cases,

- A.  $S(S, U, V) = S(S) \times S(U, V)$ ,
- B.  $S - V - U$ ,
- C.  $S - U - V$  and  $S(S, U) = S(S) \times S(U)$ ,
- D.  $H(S|U) = 0$ ,
- E.  $S - U - V$  and  $S(U, V) = S(U) \times S(V)$ , or
- F.  $\min\{H(U|S), H(V|S)\} = 0$ ,

then for any  $D \in [\underline{D}, \overline{D}]$ , the inner bound  $\mathcal{R}^*(D)$  and the outer bound  $\mathcal{R}^\ddagger(D)$  match, and

$$\mathcal{R}(D) = \mathcal{R}^*(D) = \mathcal{R}^\ddagger(D).$$

*Proof:* Theorem 3 is proved in Appendix I. ■

*Remark 2:*  $\mathcal{R}(D) = \mathcal{R}^*(D)$  when  $H(S|V) = 0$  or  $H(U|V) = 0$ , but these are subsumed by Case B above.



Since in each of the cases in Theorem 3, the characterization is given by precisely one auxiliary random variable, it is only natural to wonder as to when a *quantize-and-bin* strategy is optimal. In this strategy, the auxiliary random variable  $A$  that the encoder encodes the source into is simply the reconstruction that the receivers require. The encoder upon identifying a suitable sequence of reconstruction symbols simply uses a binning strategy to reduce the rate of communication prior to forwarding the bin index to the receivers. Thus in this strategy, all three terminals (i.e., the transmitter included) are aware of the common reconstruction. To analyze cases in which the quantize-and-bin approach is optimal, we define the corresponding rate region.

*Definition 10:* The *quantize-and-bin* rate region  $\mathcal{R}_{\text{QB}}^*(D)$  is the union of all pairs  $(r_{uv}, r_v) \in [0, \infty)^2$  such that

$$r_{uv} \geq I(S; \hat{S}|U), \quad (39a)$$

$$r_v + r_{uv} \geq I(S; \hat{S}|V), \quad (39b)$$

where the union is taken over all reconstructions  $\hat{S}$  on  $\hat{\mathcal{S}}$  with  $\hat{S} - S - (U, V)$  and  $\mathbb{E}[d(S, \hat{S})] \leq D$ .

Clearly, by setting  $A = \hat{S}$  in the proof of Theorem 1 detailed in Appendix F, we infer that  $\mathcal{R}_{\text{QB}}^*(D) \subseteq \mathcal{R}^*(D)$ . Consequently, we can see that the quantize-and-bin region is always achievable. The following result establishes three conditions under which the quantize-and-bin strategy is not merely achievable, but optimal, i.e.,  $\mathcal{R}_{\text{QB}}^*(D) = \mathcal{R}^*(D)$ .

*Theorem 4:* If the DMS  $(S, U, V) \sim p_{SUV}$  falls into one of the following cases,

- A.  $\mathcal{S}(S, U, V) = \mathcal{S}(S) \times \mathcal{S}(U, V)$  and  $H(\text{GK}^{U,V}(V)) = 0$ ,
- B.  $S - U - V$  and  $\mathcal{S}(U, V) = \mathcal{S}(U) \times \mathcal{S}(V)$ , or
- C.  $\min\{H(U|S), H(V|S)\} = 0$ ,

then, the quantize-and-bin strategy is optimal, i.e.,

$$\mathcal{R}(D) = \mathcal{R}^*(D) = \mathcal{R}^\dagger(D) = \mathcal{R}_{\text{QB}}^*(D). \quad (40)$$

*Proof:* The proof can be found in Appendix J. ■

### C. A Binary Example

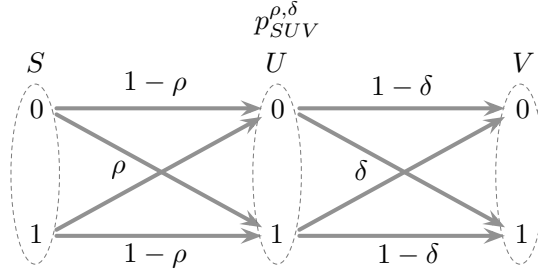


Fig. 3. A family of DMSs based on the binary symmetric channel.

In this section, we will present a binary example with  $\mathcal{S} = \hat{\mathcal{S}} = \{0, 1\}$  and with the reconstruction distortion measure  $d$  being the binary Hamming distortion measure

$$d(s, \hat{s}) = \begin{cases} 1 & \text{if } \hat{s} \neq s \\ 0 & \text{otherwise} \end{cases}. \quad (41)$$

As illustrated in Figure 3, let  $\{(S, U, V) \sim p_{SUV}^{\rho, \delta}\}_{\rho, \delta \in [0, 1]}$  denote a family of DMSs such that (a)  $S$  is an equiprobable binary source; (b)  $S - U - V$  form a Markov chain; and (c) the channels  $p_{U|S}$ , and  $p_{V|U}$  are binary symmetric channels with crossover probabilities  $\rho$  and  $\delta$ , respectively. Note that for any  $0 < \rho, \delta < 1$ , the pmf  $p_{SUV}^{\rho, \delta}$  satisfies the conditions of both Case A and Case B of Theorem 4 above. Hence, we see that the quantize-and-bin strategy is optimal, and the optimal tradeoff between communication rates on the common and private links can be obtained without having to time-share between various operating points (corresponding to different average distortions). The following result presents an explicit characterization of the  $D$ -admissible rate region  $\mathcal{R}(D)[p_{SUV}^{\rho, \delta}]$  for this class of sources.

*Lemma 9:* If<sup>1</sup>  $0 < \rho, \delta < 1$  and  $D \in [0, \frac{1}{2}]$ , then

$$\mathcal{R}(D)[p_{SUV}^{\rho, \delta}] = \left\{ (r_{uv}, r_v) \in [0, \infty)^2 : \begin{array}{l} r_{uv} \geq h(\rho * D) - h(D) \\ r_v + r_{uv} \geq h((\rho * \delta) * D) - h(D) \end{array} \right\}, \quad (42)$$

where  $x * y := x(1 - y) + y(1 - x)$  denotes the *binary convolution* operation, and  $h$  is the *binary entropy function*. Otherwise if  $D \geq 1/2$ , then  $\mathcal{R}(D)[p_{SUV}^{\rho, \delta}] = [0, \infty)^2$ .

*Proof:* If the distortion  $D \geq 1/2$ , then we can trivially meet the distortion requirement by setting  $\hat{S}^n = \tilde{S}^n = 0^n$ . Consider the non-trivial range of distortions  $0 \leq D < 1/2$ . For  $\delta \in (0, 1)$ , the corresponding joint pmf  $p_{SUV}^{\rho, \delta}$  satisfies  $S - U - V$  and  $S(U, V) = S(U) \times S(V)$ . Therefore, Case B of Theorem 4 applies, and we have

$$\mathcal{R}(D)[p_{SUV}^{\rho, \delta}] = \mathcal{R}_{\text{QB}}^*(D)[p_{SUV}^{\rho, \delta}]. \quad (43)$$

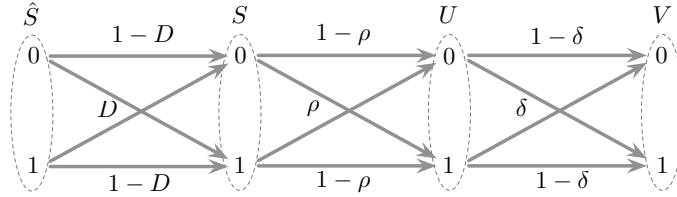


Fig. 4. Illustration of the quantization  $\hat{S}$  used in the proof of Lemma 9.

We first show that the right hand side of (42) is an inner bound for  $\mathcal{R}_{\text{QB}}^*(D)[p_{SUV}^{\rho, \delta}]$ . Choose  $\hat{S}$  as illustrated in Figure 4 with  $\mathbb{P}[\hat{S} = 1] = \mathbb{P}[\hat{S} = 0] = 1/2$  and  $\hat{S} - S - U - V$ . Then,  $\mathbb{E}[d(S, \hat{S})] = D$  and

$$I(S; \hat{S}|U) = h(D * \rho) - h(D), \quad (44a)$$

$$I(S; \hat{S}|V) = h((\rho * \delta) * D) - h(D). \quad (44b)$$

Thus, from the above, we conclude that

$$\left\{ (r_{uv}, r_v) \in [0, \infty)^2 \mid \begin{array}{l} r_{uv} \geq h(\rho * D) - h(D) \\ r_v + r_{uv} \geq h((\rho * \delta) * D) - h(D) \end{array} \right\} \subseteq \mathcal{R}_{\text{QB}}^*(D)[p_{SUV}^{\rho, \delta}] \stackrel{(43)}{=} \mathcal{R}(D)[p_{SUV}^{\rho, \delta}]. \quad (45)$$

To establish that choosing  $\hat{S}$  according to Fig. 4 suffices to cover the entire rate region, we need to show that the LHS of the above equation is an outer bound for  $\mathcal{R}_{\text{QB}}^*(D)[p_{SUV}^{\rho, \delta}]$ . To that end, we proceed as follows.

$$\begin{aligned} \mathcal{R}(D)[p_{SUV}^{\rho, \delta}] &\stackrel{(a)}{=} \bigcup_{\substack{\hat{S} - S - (U, V) \\ \mathbb{E}[d(S, \hat{S})] \leq D}} \left\{ (r_{uv}, r_v) \in [0, \infty)^2 \mid \begin{array}{l} r_{uv} \geq I(S; \hat{S}|U) \\ r_v + r_{uv} \geq I(S; \hat{S}|V) \end{array} \right\} \\ &\stackrel{(b)}{\subseteq} \left\{ (r_{uv}, r_v) \in [0, \infty)^2 \mid \begin{array}{l} r_{uv} \geq \min_{\substack{\hat{S} - S - U \\ \mathbb{E}[d(S, \hat{S})] \leq D}} I(S; \hat{S}|U) \\ r_v + r_{uv} \geq \min_{\substack{\hat{S} - S - V \\ \mathbb{E}[d(S, \hat{S})] \leq D}} I(S; \hat{S}|V) \end{array} \right\} \\ &\stackrel{(c)}{\subseteq} \left\{ (r_{uv}, r_v) \in [0, \infty)^2 \mid \begin{array}{l} r_{uv} \geq h(\rho * D) - h(D) \\ r_v + r_{uv} \geq h((\rho * \delta) * D) - h(D) \end{array} \right\}, \end{aligned} \quad (46)$$

where (a) follows from Case B of Theorem 4 of Sec. III-A and the definition of  $\mathcal{R}_{\text{QB}}^*(D)$ ; (b) creates an outer bound by relaxing the optimizing problem; and (c) follows from using Steinberg's *common-reconstruction function for the binary symmetric source* (see (14) of [13]) to obtain the solutions to the two minimizations in (46). The channel  $p_{S|\hat{S}}$  that *simultaneously* minimizes both optimization problems in (46) is precisely the choice in Fig. 4. ■

<sup>1</sup>The case where  $\delta = 0$  or 1 will be treated in the next section.

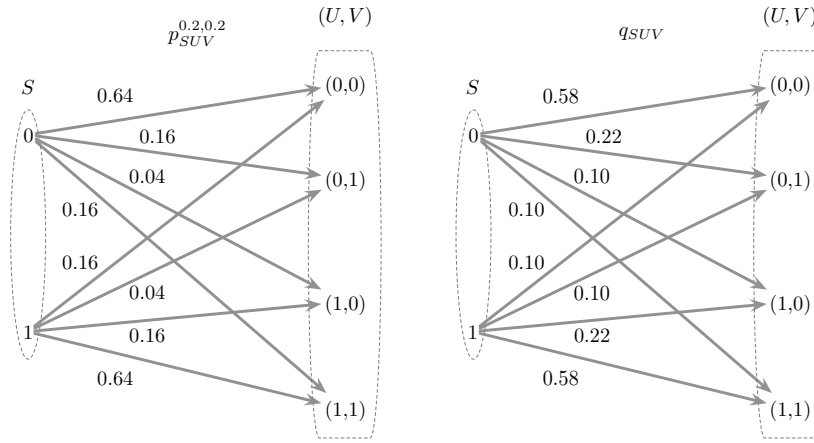


Fig. 5. Two pmfs that have the same  $D$ -admissible rate regions.

Notice carefully that the proof of the above result uses one important aspect of Steinberg's characterization of the quantize-and-bin rate region for the point-to-point rate-distortion problem.

- When the source and the receiver side-information are related by a binary symmetric channel, the optimal *reverse test channel*  $p_{S|\hat{S}}$  to be a binary symmetric channel with crossover probability  $D$  independent of the crossover probability of the channel relating the source and the receiver side information (see (14) of [13]).

Consequently, the following general observation holds.

*Remark 3:* The  $D$ -admissible rate region for any source  $q_{SUV}$  that meets the conditions of Case A or Case B of Theorem 4 and for which  $q_{U|S}$  and  $q_{V|S}$  are binary symmetric channels with crossover probabilities  $\rho, \rho * \delta$ , respectively, is given by (42).

Consider the two pmfs given in Fig. 5 in support of the remark. A simple computation will yield that when  $(S, U, V) \sim q_{SUV}$  or  $(S, U, V) \sim p_{SUV}^{0.2,0.2}$ : (a)  $\mathcal{S}(S, U, V) = \mathcal{S} \times \mathcal{U} \times \mathcal{V}$ ; (b)  $q_{U|S}$  is a binary symmetric channel with crossover probability 0.2; and (c)  $q_{V|S}$  is a binary symmetric channel with crossover probability 0.32. Remark 3 assures that the quantize-and-bin strategy is optimal for both these sources, and that  $\mathcal{R}(D)[p_{SUV}^{0.2,0.2}] = \mathcal{R}(D)[q_{SUV}]$ .

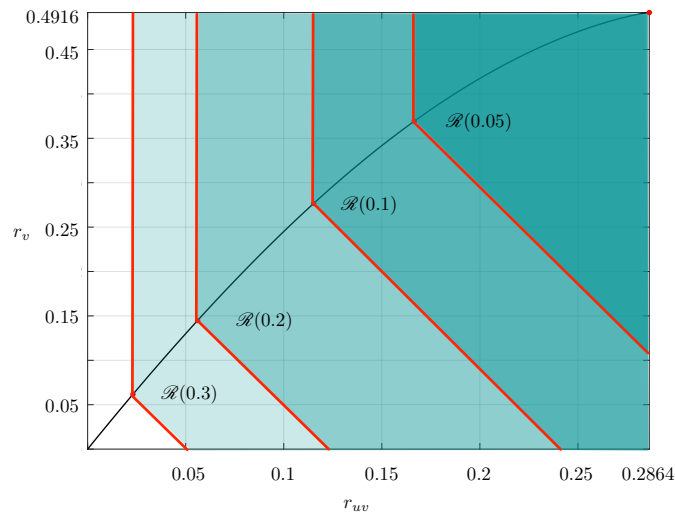


Fig. 6. The  $D$ -admissible rate region for  $p_{SUV}^{\rho,\delta}$  for  $\rho = 0.05$  and  $\rho = 0.2$ .

Figure 6 presents a simulation of the  $D$ -admissible rate region for  $p_{SUV}^{\rho,\delta}$  for  $\rho = 0.05$  and  $\rho = 0.2$  as an illustration of the result in Lemma 9. The figure presents the rate region for four values of  $D$ , namely  $D = 0.05, 0.1, 0.2$

and 0.3. The rate region for each of these values for  $D$  is bounded by two lines – one with slope  $-1$  corresponding to the message sum rate  $r_{uv} + r_v$ , and one with infinite slope corresponding to the common message rate  $r_{uv}$ . When  $D = 0.5$ , no communication is required, and the rate region is the entire non-negative quadrant. As  $D$  is made smaller, the  $D$ -admissible rate region shrinks, and the minimum required communication rate for the common message increases. The admissible rate region shrinks until eventually  $D = 0$ , at which point, the corresponding admissible rate region is given by

$$r_{uv} \geq H(S|U) = h(\rho) = h(0.05) = 0.2864, \quad (48a)$$

$$r_{uv} + r_v \geq H(S|V) = h(\rho * \delta) = h(0.23) = 0.7780. \quad (48b)$$

This region is entirely outside this figure except for the vertex, which is located at the top-right corner of the figure.

#### D. On the Discontinuity of $\mathcal{R}(D)$

In source coding problems, the continuity of rate regions with the underlying source distributions allows for small changes in source distributions to translate to small changes in boundary of the rate region. Continuity is therefore essential in practice to allow the communications system engineer to estimate the source distribution and use the estimate to choose a suitable system operating point. When a single-letter characterization of a source coding rate region is known, it is possible to establish its continuity w.r.t the underlying rate region using the continuity of Shannon's information measures on finite alphabets [21, Chap. 2.3]. For example, [22, Lem. 7.2] considers the continuity of the standard rate-distortion function, and [23] and [24] study the semicontinuity of various source networks. However, the rate regions of certain source-coding problems are known to be discontinuous in the source distribution especially when they involve zero-error or functional reconstruction constraints [22, Ch. 11], [23], [25].

Despite the absence of any such reconstruction constraints, it turns out that the  $D$ -admissible rate region studied here is *discontinuous* in the pmf  $p_{SUV}$ . The discontinuity arises rather due to the fact that we require the two reconstructions that are generated at two different locations in the network to agree (albeit with vanishing block error probability). Intuitively, in each of the cases where a single-letter characterization of the  $D$ -admissible rate region is known, the discontinuity can be attributed to the Gács-Körner common randomness in the argument of the single-letter reconstruction function; the Gács-Körner common randomness, and more precisely, its entropy can easily be seen to be discontinuous in the pmf  $p_{SUV}$ .

We illustrate this phenomenon by a simple example. Recall the  $D$ -admissible rate region  $\mathcal{R}(D)[p_{SUV}^{\rho, \delta}]$  of the binary example given above. We now establish the discontinuity of this problem by showing that

$$\lim_{\delta \downarrow 0} \mathcal{R}(D)[p_{SUV}^{\rho, \delta}] \neq \mathcal{R}(D)[p_{SUV}^{\rho, 0}], \quad 0 < \rho < D < 1/2. \quad (49)$$

Suppose that  $(S, U, V) \sim p_{SUV}^{\rho, 0}$  and  $1 > D > \rho > 0$ . Since  $\delta = 0$ , we have  $U = V$ , and consequently, neither Case A nor Case B of Theorem 4 is no longer applicable in identifying the  $D$ -admissible rate region. However, since  $D > \rho$  and  $U = V$ , we can obviously achieve the distortion  $D$  at  $(r_{uv}, r_v) = (0, 0)$  by simply choosing  $\hat{S}^n = U^n = V^n = \tilde{S}^n$ . This yields an average distortion of  $\rho < D$ , since

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[d(S_i, \hat{S}_i)] = \rho. \quad (50)$$

Thus,

$$\mathcal{R}(D)[p_{SUV}^{\rho, 0}] = [0, \infty)^2. \quad (51)$$

Lemma 9 determines  $\mathcal{R}(D)[p_{SUV}^{\rho, \delta}]$  for  $\delta \in (0, 1)$ . The mapping  $\delta \mapsto h((\rho * \delta) * D)$  is continuous on  $[0, 1]$ , so

$$\lim_{\delta \downarrow 0} \mathcal{R}(D)[p_{SUV}^{\rho, \delta}] = \left[ \lim_{\delta \downarrow 0} h((\rho * \delta) * D) - h(D), \infty \right) \times [0, \infty) = [h(\rho * D) - h(D), \infty) \times [0, \infty). \quad (52)$$

From (51) and (52), we see that the  $D$ -admissible rate region of  $p_{SUV}^{\rho, \delta}$  does not approach that of  $p_{SUV}^{\rho, 0}$  as  $\delta \rightarrow 0$ .

*Remark 4:* The above argument can also be used to show the discontinuity at  $\delta = 1$ , i.e.,

$$\lim_{\delta \uparrow 1} \mathcal{R}(D)[p_{SUV}^{\rho, \delta}] \neq \mathcal{R}(D)[p_{SUV}^{\rho, 1}]. \quad (53)$$

#### IV. CONCLUSIONS

In this work, we look at a variant of the two-receiver successive refinement problem with the common receiver reconstructions requirement. We present general inner and outer bound for this variant. The outer bound is unique in the sense it is the first information-theoretic single-letter characterization where the source reconstruction at the receivers is explicitly achieved via a function of the Gács-Körner common randomness between the random variables (both auxiliary and side information) available to the two receivers. Using these bounds, we derive a single-letter characterization of the admissible rate region and the optimal coding strategy for several settings of the joint distribution between the source and the receiver side information variables. Using this characterization, we then establish the discontinuity of the admissible rate region with respect to the underlying source distribution even though the problem formulation does not involve zero-error or functional reconstruction constraints.

#### APPENDIX A

##### PROOF OF LEMMA 1

Suppose that we have  $(a_1, u) \rightleftharpoons (a'_1, u')$  in  $\mathbb{G}^{A_1 U, A_2 V}[p_{A_1 U A_2 V}]$ . Then, there must exist a positive integer  $k \in \mathbb{N}$  and sets  $\{(a_2^{(j)}, v^{(j)}) : 1 \leq j \leq k\}$  and  $\{(a_1^{(j)}, u^{(j)}) : 1 \leq j < k\}$  such that

$$p_{A_1 U A_2 V}(a_1, u, a_2^{(1)}, v^{(1)}) > 0, \quad (54a)$$

$$p_{A_1 U A_2 V}(a_1^{(j)}, u^{(j)}, a_2^{(j)}, v^{(j)}) > 0, \quad 1 \leq j < k, \quad (54b)$$

$$p_{A_1 U A_2 V}(a_1^{(j)}, u^{(j)}, a_2^{(j+1)}, v^{(j+1)}) > 0, \quad 1 \leq j < k, \quad (54c)$$

$$p_{A_1 U A_2 V}(a'_1, u', a_2^{(k)}, v^{(k)}) > 0. \quad (54d)$$

Since  $S(A_1, A_2, U, V) = S(A_1, A_2) \times S(U, V)$ , it follows that the above equations hold if and only if

$$p_{A_1 A_2}(a_1, a_2^{(1)}) > 0 \text{ and } p_{UV}(u, v^{(1)}) > 0, \quad (55a)$$

$$p_{A_1 A_2}(a_1^{(j)}, a_2^{(j)}) > 0 \text{ and } p_{UV}(u^{(j)}, v^{(j)}) > 0, \quad 1 \leq j < k, \quad (55b)$$

$$p_{A_1 A_2}(a_1^{(j)}, a_2^{(j+1)}) > 0 \text{ and } p_{UV}(u^{(j)}, v^{(j+1)}) > 0, \quad 1 \leq j < k, \quad (55c)$$

$$p_{A_1 A_2}(a'_1, a_2^{(k)}) > 0 \text{ and } p_{UV}(u', v^{(k)}) > 0. \quad (55d)$$

Thus, it follows that

$$(a_1, u) \rightleftharpoons (a'_1, u') \text{ in } \mathbb{G}^{A_1 V, A_2 U}[p_{A_1 V A_2 U}] \Leftrightarrow \begin{cases} a_1 \rightleftharpoons a'_1 \text{ in } \mathbb{G}^{A_1, A_2}[p_{A_1 A_2}] \\ u \rightleftharpoons u' \text{ in } \mathbb{G}^{U, V}[p_{UV}] \end{cases}. \quad (56)$$

Similarly, it can be shown that

$$(a_2, v) \rightleftharpoons (a'_2, v') \text{ in } \mathbb{G}^{A_1 V, A_2 U}[p_{A_1 V A_2 U}] \Leftrightarrow \begin{cases} a_2 \rightleftharpoons a'_2 \text{ in } \mathbb{G}^{A_1, A_2}[p_{A_1 A_2}] \\ v \rightleftharpoons v' \text{ in } \mathbb{G}^{U, V}[p_{UV}] \end{cases}. \quad (57)$$

Consequently, from (7), it follows that  $\text{GK}^{A_1 U, A_2 V} \equiv (\text{GK}^{A_1, A_2}, \text{GK}^{U, V})$ . ■

#### APPENDIX B

##### PROOF OF LEMMA 5

Let  $(r_{uv}, r_v) \in \mathcal{R}_{D,k}^\dagger$ . Then, there must exist pmf  $p_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$  and function  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that

$$\mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D, \quad (58)$$

and

$$r_{uv} \geq I(S; A, B|U), \quad (59a)$$

$$r_v + r_{uv} \geq I(S; B|A, C, U, V) + I(S; A, C|V). \quad (59b)$$

For this choice of  $p_{ABCSUV}$ , let us define

$$\eta^* := \min_{\substack{(u,v) \in \mathcal{S}(U,V) \\ s \in \mathcal{S}(S)}} \frac{p_{UV|S}(u,v|s)}{p_{UV}(u,v)}. \quad (60)$$

Since (33) holds, we are guaranteed that  $\eta^* > 0$ . Then, for any  $(a, b, c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}$  and  $(u, v) \in \mathcal{U} \times \mathcal{V}$ ,

$$\begin{aligned} p_{ABCUV}(a, b, c, u, v) &= \sum_{s \in \mathcal{S}(S)} p_{ABCS}(a, b, c, s) p_{UV|S}(u, v|s) \\ &\geq \eta^* \sum_{s \in \mathcal{S}(S)} p_{ABCS}(a, b, c, s) p_{UV}(u, v) \\ &= \eta^* p_{ABC}(a, b, c) p_{UV}(u, v). \end{aligned} \quad (61)$$

$$\begin{aligned} p_{ABCUV}(a, b, c, u, v) &= \sum_{s \in \mathcal{S}(S)} p_{ABC|S}(a, b, c, s) p_{UV|S}(u, v|s) \\ &\leq \frac{1}{m_S} \sum_{s \in \mathcal{S}(S)} p_{ABCS}(a, b, c, s) p_{UVS}(u, v, s) \\ &\leq \frac{p_{UV}(u, v)}{m_S} \sum_{s \in \mathcal{S}(S)} p_{ABCS}(a, b, c, s) \\ &= \frac{p_{ABC}(a, b, c) p_{UV}(u, v)}{m_S}. \end{aligned} \quad (62)$$

Hence, we also have  $\mathcal{S}(A, B, C, U, V) = \mathcal{S}(A, B, C) \times \mathcal{S}(U, V)$ . An application of Lemma 1 of Sec. II-B with  $A_1 = (A, C)$  and  $A_2 = (A, B)$  then yields the following conclusion.

$$\text{GK}^{ABU, ACV} \equiv (\text{GK}^{AB, AC}, \text{GK}^{U, V}). \quad (63)$$

Now, define random variable  $\tilde{A} := \text{GK}^{AB, AC}$  over  $\tilde{\mathcal{A}} \subseteq \mathcal{A} \times \mathcal{C}$ , and let  $\tilde{p}_{\tilde{A}SUV}$  denote the pmf of  $(\tilde{A}, S, U, V)$ . Then, by construction,  $\tilde{A} - S - (U, V)$ . Note that both  $\tilde{A}$  and  $\text{GK}^{U, V}$  are functions of  $\text{GK}^{\tilde{A}V, \tilde{A}U}$ , i.e.,

$$H(\tilde{A}, \text{GK}^{U, V} | \text{GK}^{\tilde{A}V, \tilde{A}U}) = 0. \quad (64)$$

Combining the above with (63), yields  $H(\text{GK}^{ABU, ACV} | \text{GK}^{\tilde{A}U, \tilde{A}V}) = 0$ . Let  $\tilde{f} : \tilde{\mathcal{A}} \times \mathcal{V} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  be such that

$$\tilde{f}(\text{GK}^{\tilde{A}U, \tilde{A}V}) = \text{GK}^{ABU, ACV}. \quad (65)$$

Then, from (58),  $\mathbb{E}[d(S, f(\tilde{f}(\text{GK}^{\tilde{A}V, \tilde{A}U})))] \leq D$ . Hence, since  $\tilde{\mathcal{A}} \subseteq \mathcal{A} \times \mathcal{C}$ , we definitely have  $\tilde{p}_{\tilde{A}SUV} \in \mathcal{P}_{D, k^2}^*$ . Lastly, note that

$$r_{uv} \stackrel{(59a)}{\geq} I(S; A, B|U) \geq I(S; \text{GK}^{AB, AC}|U) = I(S; \tilde{A}|U), \quad (66)$$

$$r_v + r_{uv} \stackrel{(59b)}{\geq} I(S; A, C|V) \geq I(S; \text{GK}^{AB, AC}|V) = I(S; \tilde{A}|V). \quad (67)$$

Hence, it follows that  $(r_{uv}, r_v) \in \mathcal{R}_{D, k^2}^*$ . ■

## APPENDIX C PROOF OF LEMMA 6

We will begin by limiting the size for the auxiliary RV  $A$ , and then present bounds for the alphabet sizes for  $B$  and  $C$ , respectively. To bound the size of  $\mathcal{A}$ , we need to preserve: (a)  $\{p_S(s) : s \in \mathcal{S}\}$ ; (b) six information functionals  $H(S|A, U)$ ,  $H(S|A, V)$ ,  $H(S|A, B, U)$ ,  $H(S|A, C, V)$ ,  $H(S|A, C, U, V)$ , and  $H(S|A, B, C, U, V)$ ; and lastly, (c) the reconstruction constraint (25).

We begin by fixing  $\tilde{s} \in \mathcal{S}$ . Define the following  $|\mathcal{S}| + 6$  continuous, real-valued functions on  $\mathcal{P}(\mathcal{B} \times \mathcal{C} \times \mathcal{S})$ , the

set of all pmfs on  $\mathcal{B} \times \mathcal{C} \times \mathcal{S}$ . Let  $\pi \in \mathcal{P}(\mathcal{B} \times \mathcal{C} \times \mathcal{S})$ .

$$\psi_s(\pi) = \sum_{(b,c) \in \mathcal{B} \times \mathcal{C}} \pi(b, c, s), \quad s \in \mathcal{S} \setminus \{\check{s}\}, \quad (68a)$$

$$\psi_1(\pi) = - \sum_{s,u} \left( \sum_{b,c} \pi(b, c, s) p_{U|S}(u|s) \right) \log_2 \left( \frac{\sum_{b,c} \pi(b, c, s) p_{U|S}(u|s)}{\sum_{b,c,s'} \pi(b, c, s') p_{U|S}(u|s')} \right), \quad (68b)$$

$$\psi_2(\pi) = - \sum_{s,v} \left( \sum_{b,c} \pi(b, c, s) p_{V|S}(v|s) \right) \log_2 \left( \frac{\sum_{b,c} \pi(b, c, s) p_{V|S}(v|s)}{\sum_{b,c,s'} \pi(b, c, s') p_{V|S}(v|s')} \right), \quad (68c)$$

$$\psi_3(\pi) = - \sum_{b,s,u} \left( \sum_c \pi(b, c, s) p_{U|S}(u|s) \right) \log_2 \left( \frac{\sum_c \pi(b, c, s) p_{U|S}(u|s)}{\sum_{c,s'} \pi(b, c, s') p_{U|S}(u|s')} \right), \quad (68d)$$

$$\psi_4(\pi) = - \sum_{c,s,v} \left( \sum_b \pi(b, c, s) p_{V|S}(v|s) \right) \log_2 \left( \frac{\sum_b \pi(b, c, s) p_{V|S}(v|s)}{\sum_{b,s'} \pi(b, c, s') p_{V|S}(v|s')} \right), \quad (68e)$$

$$\psi_5(\pi) = - \sum_{b,s,u,v} \left( \sum_c \pi(b, c, s) p_{UV|S}(u, v|s) \right) \log_2 \left( \frac{\sum_c \pi(b, c, s) p_{UV|S}(u, v|s)}{\sum_{c,s'} \pi(b, c, s') p_{UV|S}(u, v|s')} \right), \quad (68f)$$

$$\psi_6(\pi) = - \sum_{b,c,s,u,v} \pi(b, c, s) p_{UV|S}(u, v|s) \log_2 \left( \frac{\pi(b, c, s) p_{UV|S}(u, v|s)}{\sum_{s'} \pi(b, c, s') p_{UV|S}(u, v|s')} \right). \quad (68g)$$

Note that preserving condition (25) is not straightforward because of the presence of Gács-Körner common randomness function. Consequently, this condition has to be split into two parts, which have to be combined together non-trivially. However, this approach requires the application of the Support Lemma [26, p. 631] infinitely many number of times, along with a suitable limiting argument. To preserve (25), define for each  $m \in \mathbb{N}$ , a continuous function  $\psi_{7,m} : \mathcal{P}(\mathcal{B} \times \mathcal{C} \times \mathcal{S}) \rightarrow \mathbb{R}$  by

$$\psi_{7,m}(\pi) = \min_{\substack{g: \mathcal{B} \times \mathcal{U} \rightarrow \hat{\mathcal{S}} \\ f: \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}}} \left[ \sum_{b,c,s,u,v} \pi(b, c, s) p_{UV|S}(u, v|s) \left( m \bar{D} \mathbb{I}\{f(c, v) = g(b, u)\} + d(s, f(c, v)) \right) \right]. \quad (69)$$

Note that  $\psi_{7,m}$  links together the distortion requirement with the probability that the reconstructions are different. Pick any pmf  $\mathbf{p}_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$ . Then, by definition, it follows that there exist functions  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  and  $g : \mathcal{A} \times \mathcal{B} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$

$$\mathbb{P}[f(A, C, V) \neq g(A, B, U)] = 0, \quad (70a)$$

$$\mathbb{E}[d(S, f(A, C, V))] \leq D. \quad (70b)$$

Consequently,

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_{7,m}(\mathbf{p}_{BCS|A=a}) \leq D, \quad \forall m \in \mathbb{N}. \quad (71)$$

Combining the above with (68a)-(68g), we see that

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_s(\mathbf{p}_{BCS|A=a}) = \mathbf{p}_S(s) = p_S(s), \quad s \in \mathcal{S} \setminus \{\check{s}\}, \quad (72a)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_1(\mathbf{p}_{BCS|A=a}) = H(S|A, U)[\mathbf{p}], \quad (72b)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_2(\mathbf{p}_{BCS|A=a}) = H(S|A, V)[\mathbf{p}], \quad (72c)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_3(\mathbf{p}_{BCS|A=a}) = H(S|A, B, U)[\mathbf{p}], \quad (72d)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_4(\mathbf{p}_{BCS|A=a}) = H(S|A, C, V)[\mathbf{p}], \quad (72e)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_5(\mathbf{p}_{BCS|A=a}) = H(S|A, B, U, V)[\mathbf{p}], \quad (72f)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_6(\mathbf{p}_{BCS|A=a}) = H(S|A, B, C, U, V)[\mathbf{p}], \quad (72g)$$

$$\sum_{a \in \mathcal{A}} p_A(a) \psi_{7,m}(\mathbf{p}_{BCS|A=a}) \leq D, \quad \forall m \in \mathbb{N}. \quad (72h)$$

For each  $m \in \mathbb{N}$ , apply the Support Lemma with  $|\mathcal{S}| + 6$  functions  $(\{\psi_s\}_{s \in \mathcal{S} \setminus \{\check{s}\}}, \psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_{7,m})$  to identify a pmf  $q_{A_m BCSUV}$  with  $|\mathcal{A}_m| \leq |\mathcal{S}| + 6$  such that  $(A_m, B, C) - S - (U, V)$  and

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_s(q_{BCS|A_m=a}) = \mathbf{p}_S(s) = p_S(s), \quad s \in \mathcal{S} \setminus \{\check{s}\}, \quad (73a)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_1(q_{BCS|A_m=a}) = H(S|A, U)[\mathbf{p}], \quad (73b)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_2(q_{BCS|A_m=a}) = H(S|A, V)[\mathbf{p}], \quad (73c)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_3(q_{BCS|A_m=a}) = H(S|A, B, U)[\mathbf{p}], \quad (73d)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_4(q_{BCS|A_m=a}) = H(S|A, C, V)[\mathbf{p}], \quad (73e)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_5(q_{BCS|A_m=a}) = H(S|A, B, U, V)[\mathbf{p}], \quad (73f)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_6(q_{BCS|A_m=a}) = H(S|A, B, C, U, V)[\mathbf{p}], \quad (73g)$$

$$\sum_{a \in \mathcal{A}_m} q_{A_m}(a) \psi_{7,m}(q_{BCS|A_m=a}) \leq D. \quad (73h)$$

After possibly renaming of elements, we may assume that the alphabet of each of the auxiliary RVs  $A_m$  is the same, say  $\mathcal{A}^*$ . Note that the optimal reconstruction functions  $f_m, g_m$  (see (69)) for the choice  $q_{A_m BCSUV}$  meeting (73h) satisfy

$$\mathbb{P}[f_m(A_m, C, V) \neq g_m(A_m, B, U)] \leq \frac{D}{Dm} \leq \frac{1}{m}, \quad (74a)$$

$$\mathbb{E}[d(S, f_m(A_m, C, V))] \leq D. \quad (74b)$$

Since the number of functions from the set  $\mathcal{A}^* \times \mathcal{B} \times \mathcal{U}$  (or  $\mathcal{A}^* \times \mathcal{C} \times \mathcal{V}$ ) to  $\hat{\mathcal{S}}$  is finite, the sequence  $\{(f_m, g_m)\}_{m \in \mathbb{N}}$  must contain infinitely many repetitions of at least one pair of reconstruction functions. Therefore, let

$$\omega_0 = \min \{i \in \mathbb{N} : |\{j \in \mathbb{N} : (f_j, g_j) = (f_i, g_i)\}| = \infty\}.$$

Let  $\{q_{A_{m_j} BCSUV}\}_{j \in \mathbb{N}}$  be the subsequence of  $\{q_{A_m BCSUV}\}_{m \in \mathbb{N}}$  with  $(f_{m_j}, g_{m_j}) = (f_{\omega_0}, g_{\omega_0})$ . By the Bolzano-Weierstrass theorem [27], we can find a subsequence  $\{q_{A_{m_{j_l}} BCSUV}\}_{l \in \mathbb{N}} \subseteq \{q_{A_{m_j} BCSUV}\}_{j \in \mathbb{N}}$  that converges to, say,  $q_{A^* BCSUV}^*$ . Since  $\{\psi_s\}_{s \in \mathcal{S} \setminus \{\check{s}\}}, \psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6$  and  $\{\psi_{7,m}\}_{m \in \mathbb{N}}$  are continuous in their arguments, by taking appropriate limits of (73a)-(73e), we see that

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_s(q_{BCS|A^*=a}) \stackrel{(73a)}{=} \mathbf{p}_S(s), \quad s \in \mathcal{S} \setminus \{\check{s}\}, \quad (75a)$$



$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_1(q_{BCS|A^*=a}) \stackrel{(73b)}{=} H(S|A, U)[p], \quad (75b)$$

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_2(q_{BCS|A^*=a}) \stackrel{(73c)}{=} H(S|A, V)[p], \quad (75c)$$

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_3(q_{BCS|A^*=a}) \stackrel{(73d)}{=} H(S|A, B, U)[p], \quad (75d)$$

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_4(q_{BCS|A^*=a}) \stackrel{(73e)}{=} H(S|A, C, V)[p], \quad (75e)$$

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_5(q_{BCS|A^*=a}) \stackrel{(73f)}{=} H(S|A, B, U, V)[p], \quad (75f)$$

$$\sum_{a \in \mathcal{A}^*} q_{A^*}(a) \psi_6(q_{BCS|A^*=a}) \stackrel{(73g)}{=} H(S|A, B, C, U, V)[p]. \quad (75g)$$

Let  $\mathcal{S} := \{(a, b, c, u, v) : f_{\omega_0}(a, c, v) \neq g_{\omega_0}(a, b, u)\}$ . Then, we have

$$\begin{aligned} \mathbb{P}[f_{\omega_0}(A^*, C, V) \neq g_{\omega_0}(A, B, U)] &= \mathbb{P}[(A^*, B, C, U, V) \in \mathcal{S}] \\ &= \lim_{l \rightarrow \infty} \mathbb{P}[(A_{m_{j_l}}, B, C, U, V) \in \mathcal{S}] \\ &= \lim_{l \rightarrow \infty} \mathbb{P}[f_{m_{j_l}}(A_{m_{j_l}}, C, V) \neq g_{m_{j_l}}(A_{m_{j_l}}, B, U)] \stackrel{(74a)}{=} 0. \end{aligned} \quad (76)$$

Note that the above equality holds because  $(f_{\omega_0}, g_{\omega_0}) = (f_{m_{j_l}}, g_{m_{j_l}})$  for all  $l \in \mathbb{N}$ . Similarly,

$$\mathbb{E}[d(S, f_{\omega_0}(A^*, C, V))] = \lim_{j \rightarrow \infty} \mathbb{E}[d(S, f_{\omega_0}(A_{m_j}, C, V))] \leq D. \quad (77)$$

Thus, we may, without loss of generality, restrict the size of the alphabet of  $A$  to be  $|\mathcal{S}| + 6$ .

Next, to restrict the size of the alphabets of  $B$  and  $C$ , we proceed by first picking  $(A, B, C, S, U, V) \in \mathcal{P}_{D,k}^\dagger$  with  $|\mathcal{A}| \leq |\mathcal{S}| + 6$ . By definition, it follows that there exist functions  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  and  $g : \mathcal{A} \times \mathcal{B} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$

$$\mathbb{P}[f(A, C, V) \neq g(A, B, U)] = 0, \quad (78a)$$

$$\mathbb{E}[d(S, f(A, C, V))] \leq D. \quad (78b)$$

Unfortunately, we cannot limit the sizes of  $\mathcal{B}$  and  $\mathcal{C}$  by invoking Carathéodory's Theorem because of the inability to preserve  $I(S; AB|U)$ ,  $(S; AC|V)$  and the constraint  $\mathbb{P}[f(A, C, V) \neq g(A, B, U)] = 0$  simultaneously.

However, without loss of generality we can assume that  $|\mathcal{S}(\mathcal{B})| \leq |\hat{\mathcal{S}}|^{|\mathcal{U}|}$ . To see why that is the case, let  $\{u_1, \dots, u_k\}$  be an enumeration of  $\mathcal{U}$ . Define auxiliary random variable  $\bar{B}$  by

$$\bar{B} := (g(A, B, u_1), \dots, g(A, B, u_k)) \in \hat{\mathcal{S}}^{|\mathcal{U}|}. \quad (79)$$

By construction,  $\bar{B}$  is a function of  $A$  and  $B$ , and hence, we are guaranteed that  $(A, \bar{B}, C) - S - (U, V)$ , and

$$H(g(A, B, U) | A, \bar{B}, U) = 0. \quad (80)$$

Combining the above with (78a) and (78b), we conclude the existence of functions  $\bar{g} : \mathcal{A} \times \bar{\mathcal{B}} \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$  such that

$$\mathbb{P}[f(A, C, V) \neq \bar{g}(A, \bar{B}, U)] = 0, \quad (81)$$

$$\mathbb{E}[d(S, f(A, C, V))] \leq D. \quad (82)$$

Further, since  $H(\bar{B} | A, B) = 0$ , we are guaranteed that

$$I(S; A, B|U) \geq I(S; A, \bar{B}|U), \quad (83a)$$

$$I(S; B|A, C, U, V) \geq I(S; \bar{B}|A, C, U, V). \quad (83b)$$

It then follows from Definitions 5 and 6 of Sec. III-A that considering random variables  $A$  and  $B$  with  $|\mathcal{A}| > |\mathcal{S}| + 6$  and  $|\mathcal{B}| > |\hat{\mathcal{S}}|^{|\mathcal{U}|}$  does not enlarge the region, i.e., we can identify different  $A^*$  and  $\bar{B}$  using the above argument that operate at the same rate pair.

We are now only left with bounding the size of  $\mathcal{C}$ , for which, we can repeat now repeat steps similar to that of  $A$ . This time, we preserve: (a) the distribution  $p_{ABS}$ ; (b) three information functionals  $H(S|A, C, V)$ ,  $H(S|A, C, U, V)$  and  $H(S|A, B, C, U, V)$ ; and (c) the reconstruction constraint. Proceeding similarly as in the case of for the random variable  $A$ , we conclude that  $|\mathcal{C}| \leq |\mathcal{S}|(|\mathcal{S}| + 6)|\hat{\mathcal{S}}|^{|\mathcal{U}|} + 3$  suffices. Thus, it follows that

$$\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|}+3}^\dagger(D) = \mathcal{R}^\dagger(D), \quad k \in \mathbb{N}. \quad (84)$$

■

*Remark 5:* Since  $\mathcal{R}_k^*(D)$  has only one auxiliary random variable, the proof of (37a) of Lemma 6 follows closely the portion of the above proof corresponding to the reduction of the size of  $A$  alone. For the purposes of Lemma 6, we only need to preserve two information functionals, namely  $H(S|A, U)$ , and  $H(S|A, V)$ . Hence, the proof will only use  $\{\psi_s : s \in \mathcal{S} \setminus \{\tilde{s}\}\}$ ,  $\psi_1$ ,  $\psi_2$ , and  $\psi_{7,m}$  to conclude that  $|\mathcal{A}| \leq |\mathcal{S}| + 2$  suffices, and hence

$$\mathcal{R}_k^*(D) \subseteq \mathcal{R}_{|\mathcal{S}|+2}^*(D) = \mathcal{R}^*(D), \quad k \in \mathbb{N}. \quad (85)$$

*Remark 6:* For the proof of (37c) of Lemma 6, we only need to preserve four information functionals, namely  $H(S|A, U)$ ,  $H(S|A, V)$ ,  $H(S|A, B, U)$  and  $H(S|A, C, V)$ . Hence, the proof will only use  $\{\psi_s : s \in \mathcal{S} \setminus \{\tilde{s}\}\}$ ,  $\psi_1$ ,  $\psi_2$ ,  $\psi_3$ ,  $\psi_4$  and  $\psi_{7,m}$  to conclude that  $|\mathcal{A}| \leq |\mathcal{S}| + 4$  suffices. The bound for  $\mathcal{B}$  is the same as in the above proof. The final argument for bounding  $\mathcal{C}$  requires the preservation of (a) the distribution  $p_{ABS}$ ; (b) the information functionals  $H(S|A, C, V)$ ; and (c) the reconstruction constraint.

$$\mathcal{R}_k^\dagger(D) \subseteq \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+4)|\hat{\mathcal{S}}|^{|\mathcal{U}|}+1}^\dagger(D) = \mathcal{R}^\dagger(D), \quad k \in \mathbb{N}. \quad (86)$$

#### APPENDIX D PROOF OF LEMMA 7

Since  $\mathcal{R}^*(D) \subseteq \mathbb{R}^2$ , we only need to show that the line segment between any two points in  $\mathcal{R}^*(D)$  lies completely within  $\mathcal{R}^*(D)$ . To do so, pick  $(r'_u, r'_{uv}), (r''_u, r''_{uv}) \in \mathcal{R}^*(D)$  and  $\lambda \in [0, 1]$ . Then, by definition, we can find pmfs  $q_{A'SUV}, q_{A''SUV} \in \mathcal{P}_{D, |\mathcal{S}|+2}^*$  such that

$$r'_{uv} \geq I(S; A'|U), \quad (87a)$$

$$r'_v + r'_{uv} \geq I(S; A'|V), \quad (87b)$$

$$r''_{uv} \geq I(S; A''|U), \quad (87c)$$

$$r''_v + r''_{uv} \geq I(S; A''|V). \quad (87d)$$

Further, there must also exist functions  $f' : \mathcal{A}' \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$ ,  $g' : \mathcal{A}' \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$ ,  $f'' : \mathcal{A}'' \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$ , and  $g'' : \mathcal{A}'' \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that

$$\mathbb{E}(S, f'(A', U)) = \sum_{a', s, u, v} q_{A'SUV}(a', s, u, v) d(s, f'(a', u)) \leq D, \quad (88a)$$

$$\mathbb{E}(S, f''(A'', U)) = \sum_{a'', s, u, v} q_{A''SUV}(a'', s, u, v) d(s, f''(a'', u)) \leq D, \quad (88b)$$

and

$$\mathbb{P}[f'(A', U) \neq g'(A', V)] = 0, \quad (89a)$$

$$\mathbb{P}[f''(A'', U) \neq g''(A'', V)] = 0, \quad (89b)$$

Without loss of generality, we may assume that the alphabets of  $A'$  and  $A''$  are disjoint, i.e.,  $\mathcal{A}' \cap \mathcal{A}'' = \emptyset$ . Let us define  $\tilde{\mathcal{A}} \triangleq \mathcal{A}' \cup \mathcal{A}''$ . Now, define a joint pmf  $q_{T\tilde{\mathcal{A}}SUV}^*$  over  $\{0, 1\} \times \tilde{\mathcal{A}} \times \mathcal{S} \times \mathcal{U} \times \mathcal{V}$  as follows:

1. Let  $T \in \{0, 1\}$  be an RV such that  $\mathbb{P}[T = 0] = \lambda$ .
2. Let

$$q_{\tilde{\mathcal{A}}SUV|T=\alpha}^* = \begin{cases} q_{A'SUV}, & \alpha = 0 \\ q_{A''SUV}, & \alpha = 1 \end{cases}. \quad (90)$$

Then by definition  $T$  is a function of  $\tilde{A}$  (since  $T = 0$  if and only if  $\tilde{A} \in \mathcal{A}'$ ) and  $T$  is independent of  $(S, U, V)$ . Let  $q_{\tilde{A}, S, U, V}^*$  be the marginal of  $(\tilde{A}, S, U, V)$  obtained from  $q_{T, \tilde{A}, S, U, V}^*$ . Then, the following hold:

(i)  $q_{\tilde{A}, S, U, V}^* \in \mathcal{P}_{D, 2|S|+4}^*$ . This follows by defining  $\tilde{f} : \tilde{\mathcal{A}} \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$  and  $\tilde{g} : \tilde{\mathcal{A}} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  by

$$\tilde{f}(\tilde{A}, U) = \begin{cases} f'(A', U), & \tilde{A} \in \mathcal{A}' \\ f''(A'', U), & \tilde{A} \in \mathcal{A}'' \end{cases}, \quad (91a)$$

$$\tilde{g}(\tilde{A}, V) = \begin{cases} g'(A', V), & \tilde{A} \in \mathcal{A}' \\ g''(A'', V), & \tilde{A} \in \mathcal{A}'' \end{cases}, \quad (91b)$$

and verifying that

$$\begin{aligned} \mathbb{P}[\tilde{f}(\tilde{A}, U) \neq \tilde{g}(\tilde{A}, V)] &= \sum_{t \in \{0,1\}} q_T^*(t) \mathbb{P}[\tilde{f}(\tilde{A}, U) \neq \tilde{g}(\tilde{A}, V) | T = t] \\ &\stackrel{(90), (91)}{=} \lambda \mathbb{P}[f'(A', U) \neq g'(A', V)] + (1 - \lambda) \mathbb{P}[f''(A'', U) \neq g''(A'', V)] \stackrel{(89)}{=} 0. \end{aligned} \quad (92)$$

$$\mathbb{E}(S, \tilde{f}(\tilde{A}, U)) = \sum_{\tilde{a}, s, u, v} q_{\tilde{A} S U V}^*(\tilde{a}, s, u, v) d(s, \tilde{f}(\tilde{a}, u)) \quad (93)$$

$$= \sum_{\tilde{a}, s, u, v} q_T^*(t) q_{\tilde{A} S U V | T}^*(\tilde{a}, s, u, v | t) d(s, \tilde{f}(\tilde{a}, u)) \quad (94)$$

$$\stackrel{(90), (91)}{=} \lambda \mathbb{E}(S, f'(A', U)) + (1 - \lambda) \mathbb{E}(S, f''(A'', U)) \stackrel{(88)}{\leq} D. \quad (95)$$

(ii) Further, we also have that

$$\begin{aligned} I(S; \tilde{A} | U) &\stackrel{(a)}{=} I(S; \tilde{A}, T | U) \stackrel{(b)}{=} I(S; \tilde{A} | U, T) = \sum_{t \in \{0,1\}} \mathbb{P}[T = t] I(S; \tilde{A} | U, T = t) \\ &= \lambda I(S; A' | U) + (1 - \lambda) I(S; A'' | U), \end{aligned} \quad (96)$$

where (a) follows since  $T$  is a function of  $\tilde{A}$ , and (b) follows by the independence of  $T$  and  $(A, U)$ . Similarly, we can also show that  $I(S; \tilde{A} | V) = \lambda I(S; A' | V) + (1 - \lambda) I(S; A'' | V)$ .

It then follows that

$$\lambda r'_{uv} + (1 - \lambda) r''_{uv} \geq \lambda I(S; A' | U) + (1 - \lambda) I(S; A'' | U) = I(S; \tilde{A} | U), \quad (97a)$$

$$\lambda(r_v + r'_{uv}) + (1 - \lambda)(r_v'' + r''_{uv}) \geq \lambda I(S; A' | V) + (1 - \lambda) I(S; A'' | V) = I(S; \tilde{A} | V). \quad (97b)$$

Hence, it follows that any point that is a linear combination of  $(r'_u, r'_{uv}), (r'_u, r''_{uv}) \in \mathcal{R}^*(D)$  is a point in  $\mathcal{R}_{2|S|+6}^*(D)$ , which by Lemma 6, is identical to  $\mathcal{R}^*(D)$ . Hence, the claim of convexity follows.

## APPENDIX E PROOF OF LEMMA 8

From Definition 4, we can find  $q_{A_i S U V} \in \mathcal{P}_{D_i, k}^*$ , and functions  $f_i : \mathcal{A}_i \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  and  $g_i : \mathcal{A}_i \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$  such that

$$r_{uv}^{(i)} \geq I(S; A_i | U), \quad (98a)$$

$$r_v^{(i)} + r_{uv}^{(i)} \geq I(S; A_i | V), \quad (98b)$$

and

$$\mathbb{P}[g_i(A_i, U) \neq f_i(A_i, V)] = 0, \quad (99a)$$

$$\mathbb{E}[d(S, f_i(A_i, V))] \leq D_i. \quad (99b)$$

Perhaps after a round of renaming, we may assume that the alphabets of  $\mathcal{A}_i$ s are identical, i.e.,  $\mathcal{A}_i = \mathcal{A}$ . Since there are only a finite number of functions from  $\mathcal{A} \times \mathcal{V}$  or  $\mathcal{A} \times \mathcal{U}$  to  $\hat{\mathcal{S}}$ , the sequence  $(f_1, g_1), (f_2, g_2), \dots$  must contain infinitely many copies of some pair of functions. Let

$$\omega = \min \{j \in \mathbb{N} : |\{i \in \mathbb{N} : (f_i, g_i) = (f_j, g_j)\}| = \infty\}. \quad (100)$$

Let  $\{i_j\}_{j \in \mathbb{N}}$  be a subsequence of indices such that  $(f_{i_j}, g_{i_j}) = (f_\omega, g_\omega)$  for all  $j \in \mathbb{N}$ . Consider the sequence of pmfs  $\{q_{A_{i_j}SUV}\}_{j \in \mathbb{N}}$ . Since  $\mathcal{A} \times \mathcal{S} \times \mathcal{U} \times \mathcal{V}$  is a finite set, by Bolzano-Weierstrass theorem [27], a subsequence of pmfs must be convergent. Let  $\{i_{j_k}\}_{k \in \mathbb{N}}$  be one such subsequence, and let  $q_{A_{i_{j_k}}SUV} \rightarrow q_{\hat{A}SUV}$ . By the continuity of the information functional [21], we see that

$$I(\hat{A}; U, V|S) = \lim_{k \rightarrow \infty} I(A_{i_{j_k}}; U, V|S) = 0. \quad (101)$$

Now, let  $\mathcal{F} = \{(a, u, v) \in \mathcal{A} \times \mathcal{U} \times \mathcal{V} : f_\omega(a, v) \neq g_\omega(a, u)\}$ . Then,

$$\begin{aligned} \mathbb{P}[f_\omega(\hat{A}, V) \neq g_\omega(\hat{A}, U)] &= \mathbb{P}[(\hat{A}, U, V) \in \mathcal{F}] \\ &= \sum_{(a, u, v) \in \mathcal{F}} q_{\hat{A}UV}(a, u, v) \\ &= \sum_{(a, u, v) \in \mathcal{F}} \left( \lim_{k \rightarrow \infty} q_{A_{i_{j_k}}UV}(a, u, v) \right) \\ &= \lim_{k \rightarrow \infty} \sum_{(a, u, v) \in \mathcal{F}} q_{A_{i_{j_k}}UV}(a, u, v) \\ &= \lim_{k \rightarrow \infty} \mathbb{P}[f_\omega(A_{i_{j_k}}, V) \neq g_\omega(A_{i_{j_k}}, U)] \stackrel{(99)}{=} 0. \end{aligned} \quad (102)$$

Further, we see that

$$\begin{aligned} \mathbb{E}[d(S, f_\omega(\hat{A}, V))] &= \sum_{a, s, v} q_{\hat{A}SV}(a, s, v) d(s, f_\omega(a, v)) \\ &= \sum_{a, s, v} \left( \lim_{k \rightarrow \infty} q_{A_{i_{j_k}}UV}(a, u, v) d(s, f_\omega(a, v)) \right) \\ &= \lim_{k \rightarrow \infty} \sum_{a, s, v} q_{A_{i_{j_k}}UV}(a, u, v) d(s, f_\omega(a, v)) \\ &= \lim_{k \rightarrow \infty} \mathbb{E}[d(S, f_\omega(A_{i_{j_k}}, V))] \stackrel{(99)}{\leq} \lim_{k \rightarrow \infty} D_i = D. \end{aligned} \quad (103)$$

Note that in the above two arguments, we have used the fact that  $(f_{i_{j_k}}, g_{i_{j_k}}) = (f_\omega, g_\omega)$  for  $k \in \mathbb{N}$ . Combining (101), (102), and (103), we see that  $q_{\hat{A}SUV} \in \mathcal{P}_{D,k}^*$ . Lastly, using the continuity of the information functional [21], we see that

$$r_{uv} = \lim_{k \rightarrow \infty} r_{uv}^{(i_{j_k})} \geq \lim_{k \rightarrow \infty} I(S; A_{i_{j_k}}|U) = I(S; \hat{A}|U), \quad (104a)$$

$$r_{uv} + r_v = \lim_{k \rightarrow \infty} r_{uv}^{(i_{j_k})} + r_v^{(i_{j_k})} \geq \lim_{k \rightarrow \infty} I(S; A_{i_{j_k}}|V) = I(S; \hat{A}|U). \quad (104b)$$

Hence, it follows that  $(r_{uv}, r_v) \in \mathcal{R}_k^*(D)$ . ■

## APPENDIX F

### PROOF OF THEOREM 1 (INNER BOUND)

Pick  $p_{ABCSUV} \in \mathcal{P}_{D,k}^\dagger$  and  $\varepsilon \in (0, \frac{1}{6})$ . Let  $f : \mathcal{A} \times \mathcal{B} \times \mathcal{U} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  and  $g : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  be such that

$$f_U(A, B, U) = f_V(A, C, V) = \text{GK}^{ABU, ACV}(A, C, V).$$

Let us denote

$$R_a = I(S; A|U) + 2\varepsilon, \quad (105a)$$

$$R'_a = \max\{I(S; A|V) - I(S; A|U), 0\}, \quad (105b)$$

$$R''_a = \max\{\min\{I(A; U), I(A; V)\} - \varepsilon, 0\}, \quad (105c)$$

$$R_b = I(S; B|A, U) + 2\varepsilon, \quad (105d)$$

$$R'_b = \max\{I(U; B|A) - \varepsilon, 0\}, \quad (105e)$$

$$R_c = I(S; C|A, V) + 2\varepsilon, \quad (105f)$$

$$R'_c = \max\{I(V; C|A) - \varepsilon, 0\}. \quad (105g)$$

We build a codebook using the marginals  $p_A$ ,  $p_{B|A}$  and  $p_{C|A}$  obtained from the chosen joint pmf. The codebooks for the three auxiliary RVs are constructed as follows using the structure illustrated in Fig. 7.

- For each triple  $(i, i', i'') \in \llbracket 1, 2^{nR_a} \rrbracket \times \llbracket 1, 2^{nR'_a} \rrbracket \times \llbracket 1, 2^{nR''_a} \rrbracket$ , generate a random codeword  $A^n(i, i', i'') \sim p_A^n$  independent of all other codewords. Note that by the choice of rates, the total rate of the  $A$ -codebook is

$$R_a + R'_a + R''_a \stackrel{(105)}{>} I(S; A). \quad (106)$$

- For each triple  $(i, i', i'') \in \llbracket 1, 2^{nR_a} \rrbracket \times \llbracket 1, 2^{nR'_a} \rrbracket \times \llbracket 1, 2^{nR''_a} \rrbracket$ , and pair  $(\ell, \ell') \in \llbracket 1, 2^{nR_b} \rrbracket \times \llbracket 1, 2^{nR'_b} \rrbracket$ , generate a random codeword  $B^n(i, i', i'', \ell, \ell') \sim \prod_{k=1}^n p_{B|A_k}(i, i', i'')$  independent of all other codewords. Note that by the choice of rates, the total rate of the  $B$ -codebook is

$$R_b + R'_b \stackrel{(105)}{>} I(S; B|A). \quad (107)$$

- Similarly, for each triple  $(i, i', i'') \in \llbracket 1, 2^{nR_a} \rrbracket \times \llbracket 1, 2^{nR'_a} \rrbracket \times \llbracket 1, 2^{nR''_a} \rrbracket$ , and pair  $(l, l') \in \llbracket 1, 2^{nR_c} \rrbracket \times \llbracket 1, 2^{nR'_c} \rrbracket$ , generate a random codeword  $C^n(i, i', i'', l, l')$  randomly using  $\prod_{k=1}^n p_{C|A_k}(i, i', i'')$  independent of all other codewords. Note that by the choice of rates, the total rate of the  $C$ -codebook is

$$R_c + R'_c \stackrel{(105)}{>} I(S; C|A). \quad (108)$$

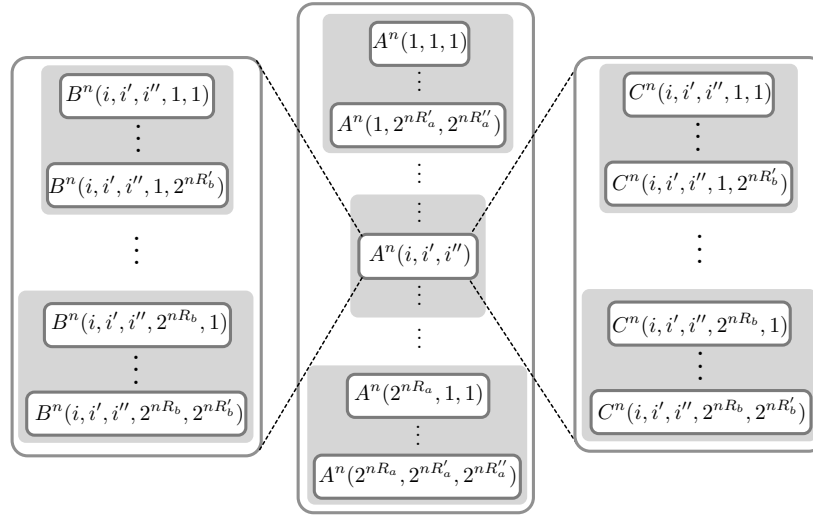


Fig. 7. A random-coding achievable scheme.

Upon receiving a realization  $s^n$  of  $S^n$ , the encoder does the following:

1. It searches for a triple  $(i, i', i'')$  such that  $(A^n(i, i', i''), s^n) \in T_\varepsilon^n[p_{AS}]$ .
2. It then searches for a pair  $(\ell, \ell')$  such that  $(A^n(i, i', i''), B^n(i, i', i'', \ell, \ell'), s^n) \in T_\varepsilon^n[p_{ABS}]$ , and a pair  $(l, l')$  such that  $(A^n(i, i', i''), C^n(i, i', i'', l, l'), s^n) \in T_\varepsilon^n[p_{ACS}]$ . Using (106)-(108) and invoking the lossy source coding theorem [26, p. 57], and the Covering Lemma [26, p. 62], we see that

$$\mathbb{P}[\exists (i, i', i'', \ell, \ell') : (A^n(i, i', i''), B^n(i, i', i'', \ell, \ell'), C^n(i, i', i'', l, l'), s^n) \in T_\varepsilon^n[p_{ABCS}]] \xrightarrow{n \rightarrow \infty} 0. \quad (109)$$

3. The encoder conveys  $(i, \ell)$  to both receivers, and  $(i', l)$  to the receiver with side information  $V$ . Note that this strategy corresponds to the following rates

$$r_{uv} = R_a + R_b = I(S; A, B|U) + 4\varepsilon, \quad (110a)$$

$$\begin{aligned} r_{uv} + r_v &= R_a + R'_a + R_b + R_c \\ &= \max\{I(S; A|U), I(S; A|V)\} + I(S; B|A, U) + I(S; C|A, V) + 6\varepsilon. \end{aligned} \quad (110b)$$

Further, for any  $(i, i', i'', \ell, \ell', l, l')$ , by the Markov Lemma [26, p. 296], we are guaranteed that

$$\mathbb{P} \left[ \begin{array}{c} (A^n(i, i', i''), B^n(i, i', i''), C^n(i, i', i''), S^n, U^n, V^n) \notin T_{2\varepsilon}^n[p_{ABCSUV}] \\ \mid \\ (A^n(i, i', i''), B^n(i, i', i''), C^n(i, i', i''), S^n) \in T_{\varepsilon}^n[p_{ABCS}] \end{array} \right] \xrightarrow{n \rightarrow \infty} 0. \quad (111)$$

Thus for sufficiently large  $n$ , the probability with which we will find a tuple  $(i, i', i'', \ell, \ell', l, l')$  such that the corresponding codewords and the source and side information realizations is  $\varepsilon$ -typical can be made arbitrarily small. Moreover, as a consequence of the Packing Lemma [26, p. 46], we are also guaranteed that

$$\mathbb{P} \left[ \begin{array}{c} \exists (\tilde{i}', \tilde{i}'') \neq (i', i'') \text{ s.t. } (A^n(i, \tilde{i}', \tilde{i}''), U^n) \in T_{2\varepsilon}^n[p_{AU}] \\ \mid \\ (A^n(i, i', i''), U^n) \in T_{2\varepsilon}^n[p_{AU}] \end{array} \right] \xrightarrow{n \rightarrow \infty} 0, \quad (112a)$$

$$\mathbb{P} \left[ \begin{array}{c} \exists (\tilde{i}', \tilde{i}'') \neq (i', i'') \text{ s.t. } (A^n(i, \tilde{i}', \tilde{i}''), V^n) \in T_{2\varepsilon}^n[p_{AV}] \\ \mid \\ (A^n(i, i', i''), V^n) \in T_{2\varepsilon}^n[p_{AV}] \end{array} \right] \xrightarrow{n \rightarrow \infty} 0, \quad (112b)$$

$$\mathbb{P} \left[ \begin{array}{c} \exists \tilde{\ell}' \neq \ell \text{ s.t. } (A^n(i, i', i''), B^n(i, i', i''), \tilde{\ell}', U^n) \in T_{2\varepsilon}^n[p_{ABU}] \\ \mid \\ (A^n(i, i', i''), B^n(i, i', i''), \ell, U^n) \in T_{2\varepsilon}^n[p_{ABU}] \end{array} \right] \xrightarrow{n \rightarrow \infty} 0, \quad (112c)$$

$$\mathbb{P} \left[ \begin{array}{c} \exists \tilde{\ell}' \neq l \text{ s.t. } (A^n(i, i', i''), C^n(i, i', i''), \tilde{\ell}', V^n) \in T_{2\varepsilon}^n[p_{ACV}] \\ \mid \\ (A^n(i, i', i''), C^n(i, i', i''), l, V^n) \in T_{2\varepsilon}^n[p_{ACV}] \end{array} \right] \xrightarrow{n \rightarrow \infty} 0. \quad (112d)$$

From (109), (111), (112), we can choose  $n$  large enough such that when averaging over all realizations of the random codebooks, the probability with which all of the following events occur is at least  $\varepsilon$ .

- (a) the encoder will be able to identify indices  $(i, i', i'', l, l', \ell, \ell')$  such that the corresponding codewords and the realization of  $S^n$  are jointly  $\varepsilon$ -letter typical;
- (b) the identified codewords and the realizations of  $(S^n, U^n, V^n)$  are jointly  $2\varepsilon$ -letter typical;
- (c) the receiver with side information  $V$  will identify the indices  $(i'', l')$  determined by the encoder; and
- (d) the receiver with side information  $U$  will identify the indices  $i', i'', \ell'$  determined by the encoder.

Then, there must exist a realization of the three codebooks  $\{a^n(i, i', i'')\}$ ,  $\{b^n(i, i', i'', \ell, \ell')\}$ , and  $\{c^n(i, i', i'', l, l')\}$  such that the above four events occur simultaneously with a probability of at least  $1 - \varepsilon$ . For this realization of the codebooks, with probability of at least  $1 - \varepsilon$ , the realizations of the source and side informations  $(s^n, u^n, v^n)$ , and the selected codewords will be jointly  $2\varepsilon$ -letter typical, i.e.,

$$(a^n(i, i', i''), b^n(i, i', i'', \ell, \ell'), c^n(i, i', i'', l, l'), s^n, u^n, v^n) \in T_{2\varepsilon}^n[p_{ABCSUV}]. \quad (113)$$

Note that letter typicality ensures that the support of the empirical distribution  $\tilde{q}_{ABCSUV}$  induced by the tuple  $(a^n(i, i', i''), b^n(i, i', i'', \ell, \ell'), c^n(i, i', i'', l, l'), s^n, u^n, v^n)$  matches that of  $q_{ABCSUV}$ . Thus, it follows that whenever (113) holds, we have for each  $k = 1, \dots, n$ ,

$$\hat{s}_k = f_U(a_k(i, i', i''), b_k(i, i', i'', \ell, \ell'), u_k) = f_V(a_k(i, i', i''), c_k(i, i', i'', l, l'), v_k) = \tilde{s}_k. \quad (114)$$

Lastly, since (113) holds with a probability of at least  $1 - \varepsilon$ , it is also true that the reconstructions at either receiver will offer an expected distortion of no more than  $D(1 + 2\varepsilon)$  with a probability of at least  $1 - \varepsilon$ . The proof is complete by noting that  $\varepsilon$  can be chosen to be arbitrarily small. ■

## APPENDIX G

### PROOF OF THEOREM 2 (OUTER BOUND)

Let  $(r_{uv}, r_v)^T \in \mathcal{R}(D)$ . Let  $0 < \iota < 1$ . Choose  $0 < \eta < \log 2$  such that

$$\max \left\{ 2|\mathcal{S}|(e^\eta - 1), \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2|\mathcal{S}|(e^\eta - 1) \log |\mathcal{S}| \right\} < \iota. \quad (115)$$

Let  $\varepsilon = \min \left\{ \iota, \left( \frac{1}{4} \mathbf{m}_{SUV} \mathbf{m}_S (1 - e^{-\eta}) \right)^2 \right\}$ . By Definition 1, there exist functions  $\mathbf{E}_{uv}$ ,  $\mathbf{E}_v$ ,  $\mathbf{D}_u$ ,  $\mathbf{D}_v$  satisfying (23). Let  $M_{uv} := \mathbf{E}_{uv}(S^n)$ ,  $M_v := \mathbf{E}_v(S^n)$ ,  $\hat{S}^n := \mathbf{D}_u(M_{uv}, U^n)$ , and  $\tilde{S}^n := \mathbf{D}_v(M_{uv}, M_v, V^n)$ . Then,

$$\begin{aligned}
 r_{uv} + \varepsilon &\geq \frac{1}{n} H(M_{uv}) \\
 &\geq \frac{1}{n} H(M_{uv} | U^n) \\
 &\geq \frac{1}{n} I(S^n; M_{uv} | U^n) \\
 &= \frac{1}{n} \sum_{i=1}^n I(S_i; M_{uv} | U^n, S^{i-1}) \\
 &= \frac{1}{n} \sum_{i=1}^n I(S_i; M_{uv}, U^{n \setminus i}, S^{i-1} | U_i) \\
 &\stackrel{(a)}{=} \frac{1}{n} \sum_{i=1}^n I(S_i; A_i, B_i | U_i) \\
 &\stackrel{(b)}{=} I(S_Q; A_Q, B_Q | U_Q, Q) \\
 &\stackrel{(c)}{=} I(S_Q; A_Q, B_Q, Q | U_Q) \\
 &\stackrel{(d)}{=} I(S; \bar{A}, \bar{B} | U),
 \end{aligned} \tag{116}$$

where in (a), we let  $A_i := (M_{uv}, U^{i-1})$  and  $B_i := (U_{i+1}^n, S^{i-1})$ ; in (b), we introduce the time-sharing random variable  $Q$  that is uniform over  $\{1, \dots, n\}$ ; in (c) we use the fact that  $Q$  is independent of  $(S_Q, U_Q)$ ; and in (d), we denote  $\bar{A} := (A_Q, Q)$ , and  $\bar{B} = (B_Q, Q)$ . Similarly,

$$\begin{aligned}
 r_{uv} + r_v + 2\varepsilon &\geq \frac{1}{n} H(M_{uv}, M_v) \\
 &\geq \frac{1}{n} I(S^n; M_{uv}, M_v | V^n) \\
 &= \frac{1}{n} \left[ I(S^n; M_{uv}, M_v, U^n | V^n) - I(S^n; U^n | M_{uv}, M_v, V^n) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; M_{uv}, M_v, U^n | V^n, S^{i-1}) - I(S^n; U_i | M_{uv}, M_v, V^n, U^{i-1}) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; M_{uv}, M_v, U^n, V^{n \setminus i}, S^{i-1} | V_i) - I(S^n; U_i | M_{uv}, M_v, V^n, U^{i-1}) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; A_i, B_i, M_v, V^{n \setminus i}, U_i | V_i) - H(U_i | A_i, M_v, V^n) + H(U_i | A_i, M_v, V^n, S^n) \right] \\
 &\stackrel{(a)}{=} \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; A_i, B_i, C_i, U_i | V_i) - H(U_i | A_i, C_i, V_i) + H(U_i | A_i, C_i, V_i, S_i) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; A_i, B_i, C_i, U_i | V_i) - I(S_i; U_i | A_i, C_i, V_i) \right] \\
 &= \frac{1}{n} \sum_{i=1}^n \left[ I(S_i; A_i, C_i | V_i) + I(S_i; B_i | A_i, C_i, U_i, V_i) \right] \\
 &\stackrel{(b)}{=} I(S_Q; A_Q, C_Q | V_Q, Q) + I(S_Q; B_Q | A_Q, C_Q, U_Q, V_Q, Q) \\
 &\stackrel{(c)}{=} I(S; \bar{A}, \bar{C} | V) + I(S; \bar{B} | \bar{A}, \bar{C}, U, V),
 \end{aligned} \tag{117}$$

where in (a), we have denoted  $C_i := (M_v, V^{n \setminus i})$  and used the chain

$$U_i - (M_{uv}, M_v, V^n, U^{i-1}, S_i) - (M_{uv}, M_v, V^n, U^{i-1}, S^n); \quad (118)$$

in (b) we make use of the uniform time-sharing random variable  $Q$ ; and in (c), we use the independence of  $Q$  and  $(S_Q, U_Q, V_Q)$  and define  $\bar{A} := (A_Q, Q)$ ,  $\bar{B} = (B_Q, Q)$ , and  $\bar{C} = (C_Q, Q)$ . Note that the following holds.

$$I(U, V; \bar{A}, \bar{B}, \bar{C} | S) = I(U_Q, V_Q; A_Q, B_Q, C_Q | S_Q, Q) = \sum_{i=1}^n \frac{1}{n} I(U_i, V_i; A_i, B_i, C_i | S_i) \stackrel{(118)}{=} 0. \quad (119)$$

Now, note that  $H(M_{uv}, M_v, V^n | A_i, C_i, V_i) = 0$  and  $H(M_{uv}, U^n | A_i, B_i, U_i) = 0$ . Hence,  $\hat{S}_i$  and  $\tilde{S}_i$  are functions of  $(A_i, B_i, U_i)$  and  $(A_i, C_i, V_i)$ , resp. Let  $\tilde{S}_i := g_i(A_i, B_i, U_i)$  and  $\hat{S}_i := f_i(A_i, C_i, V_i)$ . Let  $\bar{f} : \bar{\mathcal{A}} \times \bar{\mathcal{C}} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  and  $\bar{g} : \bar{\mathcal{A}} \times \bar{\mathcal{B}} \times \mathcal{U} \rightarrow \hat{\mathcal{S}}$  be defined as follows. For each  $i \in \{1, \dots, n\}$  and  $(a, b, c, u, v) \in \mathcal{A}_i \times \mathcal{B}_i \times \mathcal{C}_i \times \mathcal{U} \times \mathcal{V}$ ,

$$\bar{f}((a, i), (c, i), v) := f_i(a, c, v), \quad (120a)$$

$$\bar{g}((a, i), (b, i), u) := g_i(a, b, u). \quad (120b)$$

In other words,

$$\tilde{S}_Q = \bar{f}((A_Q, Q), (C_Q, Q), V_Q), \quad (121a)$$

$$\hat{S}_Q = \bar{g}((A_Q, Q), (B_Q, Q), U_Q). \quad (121b)$$

Using the above notation, we can then verify that

$$\begin{aligned} \mathbb{P}[\bar{f}(\bar{A}, \bar{C}, V) \neq \bar{g}(\bar{A}, \bar{B}, U)] &= \sum_{i=1}^n \frac{1}{n} \mathbb{P}[\bar{f}(\bar{A}, \bar{C}, V) \neq \bar{g}(\bar{A}, \bar{B}, U) | Q = i] \\ &= \sum_{i=1}^n \frac{\mathbb{P}[\tilde{S}_i \neq \hat{S}_i]}{n} \leq \varepsilon \leq \iota, \end{aligned} \quad (122)$$

$$\begin{aligned} \mathbb{E}[d(S, \bar{f}(\bar{A}, \bar{C}, V))] &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[d(S_i, f_i(A_i, C_i, V_i) | Q = i)] \\ &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[d(S_i, \tilde{S}_i)] \stackrel{(c)}{\leq} D + \varepsilon \leq D + \iota, \end{aligned} \quad (123)$$

$$\begin{aligned} \mathbb{E}[d(S, \bar{g}(\bar{A}, \bar{B}, U))] &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[d(S_i, g_i(A_i, B_i, U_i) | Q = i)] \\ &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[d(S_i, \hat{S}_i)] \stackrel{(c)}{\leq} D + \varepsilon \leq D + \iota. \end{aligned} \quad (124)$$

Now, we have to establish the existence of auxiliary RVs such that the RHS of (122) is, in fact, zero. To do so, we make use of the two *pruning theorems* in Appendix H. The first step is to only allow realizations of auxiliary random variables for which the reconstructions agree most of the time, and prune out the rest. To this end, define

$$\mathcal{E} := \left\{ (\bar{a}, \bar{b}, \bar{c}) \in \mathcal{S}(\bar{A}, \bar{B}, \bar{C}) : \mathbb{P}[\bar{f}(\bar{A}, \bar{C}, V) \neq \bar{g}(\bar{A}, \bar{B}, U) \mid (\bar{A}, \bar{B}, \bar{C}) = (\bar{a}, \bar{b}, \bar{c})] \leq \sqrt{\varepsilon} \right\}. \quad (125)$$

By a simple application of Markov's inequality, one can argue that

$$\mathbb{P}[(\bar{A}, \bar{B}, \bar{U}) \in \mathcal{E}] \geq 1 - \sqrt{\varepsilon}. \quad (126)$$

Define auxiliary RVs  $\hat{A} \in \bar{\mathcal{A}}$ ,  $\hat{B} \in \bar{\mathcal{B}}$  and  $\hat{C} \in \bar{\mathcal{C}}$  with  $(\hat{A}, \hat{B}, \hat{C}) - \bar{S} - (\bar{U}, \bar{V})$  by

$$p_{\hat{A}\hat{B}\hat{C}S}(a, b, c, s) = \begin{cases} \frac{p_{ABCS}(a, b, c, s)}{\mathbb{P}[(A, B, C) \in \mathcal{E} | S = s]}, & (a, b, c) \in \mathcal{E} \text{ and } s \in \mathcal{S}(S) \\ 0, & \text{otherwise} \end{cases}. \quad (127)$$

The above pmf  $p_{\hat{A}\hat{B}\hat{C}S}$  is precisely the pmf  $p_{ABCS}$  obtained by an application of Pruning Method A defined in



(143) of Appendix H applied with  $\delta = \sqrt{\varepsilon} < m_S(1 - e^{-\eta})$ . Hence, the properties of Theorem 5 of Appendix H can be applied to  $p_{\hat{A}\hat{B}\hat{C}S}$ . First, by an application of Property (d) of Theorem 5 of Appendix H, we see that

$$|I(S; AB|U) - I(S; \hat{A}\hat{B}|U)| = |H(S|ABU) - H(S|\hat{A}\hat{B}U)| \leq \iota. \quad (128)$$

$$|I(S; AC|V) - I(S; \hat{A}\hat{C}|V)| = |H(S|ACV) - H(S|\hat{A}\hat{C}V)| \leq \iota. \quad (129)$$

Similarly, using Property (e) of Theorem 5 of Appendix H, we see that

$$|I(S; B|ACUV) - I(S; \hat{B}|\hat{A}\hat{C}UV)| = \left( \begin{array}{c} |H(S|ACUV) - H(S|\hat{A}\hat{C}UV)| \\ + |H(S|ABCUV) - H(S|\hat{A}\hat{B}\hat{C}UV)| \end{array} \right) \leq 2\iota. \quad (130)$$

From (123), (124) and Property (a) of Theorem 5 of Appendix H, we see that

$$\mathbb{E}[d(S, \bar{f}(\hat{A}, \hat{C}, V))] \leq \mathbb{E}[d(S, \bar{f}(\bar{A}, \bar{C}, V))] + \|p_{ACSV} - p_{\hat{A}\hat{C}SV}\| \bar{D} \quad (131a)$$

$$\leq D + \iota + \bar{D}\iota \quad (131b)$$

$$\mathbb{E}[d(S, \bar{g}(\hat{A}, \hat{B}, U))] \leq \mathbb{E}[d(S, \bar{g}(\bar{A}, \bar{B}, U))] + \|p_{ABSU} - p_{\hat{A}\hat{B}SU}\| \bar{D} \quad (131c)$$

$$\leq D + \iota + \bar{D}\iota. \quad (131d)$$

Since  $S(\hat{A}, \hat{B}, \hat{C}) = \mathcal{E}$ , for any  $(\bar{a}, \bar{b}, \bar{c}) \in \mathcal{E}$ , we can invoke Property (b) of Theorem 5 of Appendix H with  $\mathcal{F} := \{(u, v) : \bar{f}(\bar{a}, \bar{c}, v) \neq \bar{g}(\bar{a}, \bar{b}, u)\}$  to infer that

$$\mathbb{P}[\bar{f}(\hat{A}, \hat{C}, V) \neq \bar{g}(\hat{A}, \hat{B}, U) \mid (\hat{A}, \hat{B}, \hat{C}) = (\bar{a}, \bar{b}, \bar{c})] \leq e^\eta \sqrt{\varepsilon} = 2\sqrt{\varepsilon}, \quad (132)$$

Now, let's proceed by pruning  $p_{\hat{A}\hat{B}\hat{C}S}$  further by using Pruning Method B defined in (159) of Appendix H with  $\delta = \frac{2\sqrt{\varepsilon}}{m_{UVS}}$ . Let us define

$$\mathcal{E}' = \left\{ (a, b, c, s) : p_{S|\hat{A}\hat{B}\hat{C}}(s|\bar{a}, \bar{b}, \bar{c}) > \frac{2\sqrt{\varepsilon}}{m_{UVS}} \right\}, \quad (133)$$

and auxiliary RVs  $A \in \bar{A}$ ,  $B \in \bar{B}$ , and  $C \in \bar{C}$  with  $(A, B, C) - S - (U, V)$  by

$$p_{ABCS}(a, b, c, s) = \begin{cases} \frac{p_{\hat{A}\hat{B}\hat{C}S}(a, b, c, s)}{\mathbb{P}[(\hat{A}, \hat{B}, \hat{C}, S) \in \mathcal{E}' | S = s]}, & (a, b, c, s) \in \mathcal{E}' \\ 0, & \text{otherwise} \end{cases}. \quad (134)$$

Since  $p_{ABCS}$  is obtained from  $p_{\hat{A}\hat{B}\hat{C}S}$  by Pruning Method B, the properties of Theorem 6 of Appendix H can be applied to  $p_{\hat{A}\hat{B}\hat{C}S}$ . Combining (132) with Property (b) of Theorem 6 of Appendix H, we see that

$$\mathbb{P}[\bar{f}(A, C, V) \neq \bar{g}(A, B, U) \mid (A, B, C) = (\bar{a}, \bar{b}, \bar{c})] = 0, \quad (\bar{a}, \bar{b}, \bar{c}) \in \mathcal{E}. \quad (135)$$

Since by construction,  $S(A, B, C) \subseteq S(\hat{A}, \hat{B}, \hat{C}) \subseteq \mathcal{E}$ , it follows that

$$\mathbb{P}[\bar{f}(A, C, V) \neq \bar{g}(A, B, U)] = 0. \quad (136)$$

Invoking Property (a) of Theorem 6 of Appendix H, it follows that

$$\mathbb{E}[d(S, \bar{f}(A, C, V))] \leq \mathbb{E}[d(S, \bar{f}(\hat{A}, \hat{C}, V))] + \bar{D}\|p_{ACSV} - p_{\hat{A}\hat{C}SV}\| \quad (137a)$$

$$\leq D + \iota + 2\bar{D}\iota \quad (137b)$$

$$\mathbb{E}[d(S, \bar{g}(A, B, U))] \leq \mathbb{E}[d(S, \bar{g}(\hat{A}, \hat{B}, U))] + \bar{D}\|p_{ABSU} - p_{\hat{A}\hat{B}SU}\| \quad (137c)$$

$$\leq D + \iota + 2\bar{D}\iota. \quad (137d)$$

By an application of Property (e) of Theorem 6 of Appendix H, we see that

$$|I(S; AB|U) - I(S; \hat{A}\hat{B}|U)| = |H(S|ABU) - H(S|\hat{A}\hat{B}U)| \leq \iota, \quad (138a)$$

$$|I(S; AC|V) - I(S; \hat{A}\hat{C}|V)| = |H(S|ACV) - H(S|\hat{A}\hat{C}V)| \leq \iota. \quad (138b)$$

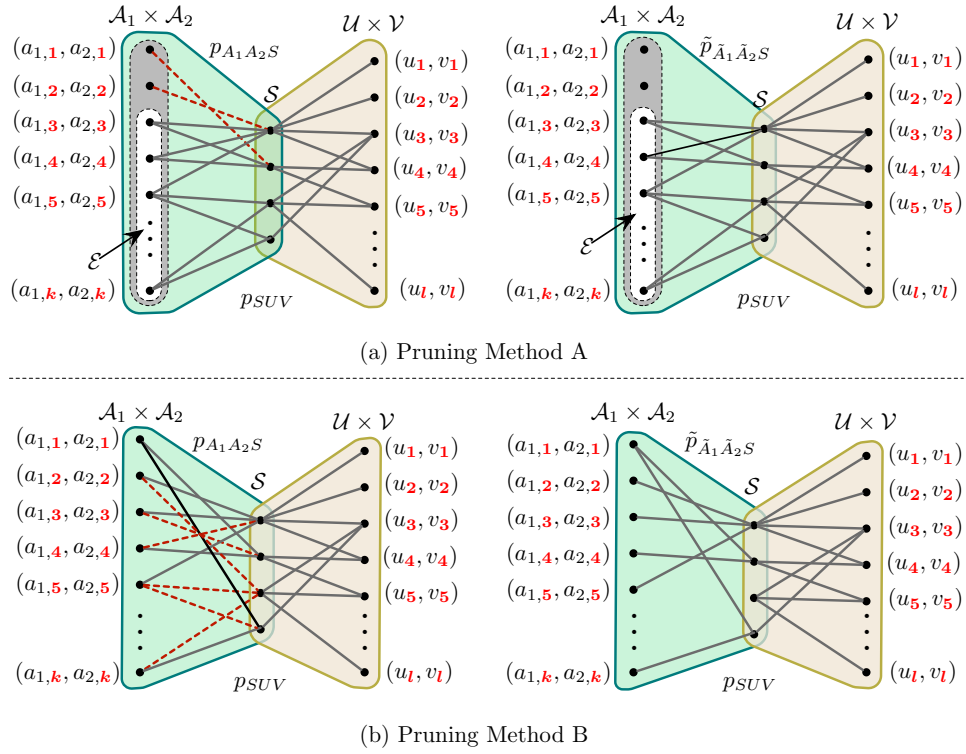


Fig. 8. Illustration of the two kinds of pruning.

Similarly, an application of Property (f) of Theorem 6 of Appendix H yields

$$|I(S; B|ACUV) - I(S; \hat{B}|\hat{A}\hat{C}UV)| = \left( \begin{array}{l} |H(S|ACUV) - H(S|\hat{A}\hat{C}UV)| \\ + |H(S|ABCUV) - H(S|\hat{A}\hat{B}\hat{C}UV)| \end{array} \right) \leq 2\iota. \quad (139)$$

Combining (128)-(131) with (137)-(139), we have auxiliary RVs  $(A, B, C) - S - (U, V)$  such that

$$r_{uv} + 2\iota \geq I(S; AB|U), \quad (140a)$$

$$r_v + r_{uv} + 2\iota \geq I(S; AC|V) + I(S; B|ACUV), \quad (140b)$$

$$\mathbb{P}[\bar{f}(A, C, V) \neq \bar{g}(A, B, U)] = 0, \quad (140c)$$

$$\mathbb{E}[d(S, \bar{f}(A, C, V))] \leq D + \iota + 2\bar{D}\iota, \quad (140d)$$

$$\mathbb{E}[d(S, \bar{g}(A, B, U))] \leq D + \iota + 2\bar{D}\iota. \quad (140e)$$

Thus, from (37b) of Lemma 6 of Sec. III-A, it follows that  $(r_{uv} + \iota, r_v + \iota) \in \mathcal{R}^\dagger(D + \iota + 2\bar{D}\iota)$ . Finally, by constructing an appropriate sequence of infinitesimals and invoking Lemma 8 of Sec. III-A, we conclude that  $(r_{uv}, r_v) \in \mathcal{R}^\dagger(D)$ . ■

## APPENDIX H PRUNING THEOREMS

We now present two *pruning theorems* that concern any five random variables  $(A_1, A_2, S, B_1, B_2)$  where  $(A_1, A_2) - S - (B_1, B_2)$  forms a Markov chain. The theorems will be applied to the CRR successive refinement problem, where  $S$  will be identified as the source,  $A_1, A_2$  will be associated with auxiliary random variables, and  $B_1$  and  $B_2$  are the side information random variables available at each of the receivers. The pruning theorems will help us to understand how small alterations to the marginal pmf of  $(A_1, A_2, S)$  will change

- the joint pmf of  $(A_1, A_2, S, B_1, B_2)$  with respect to variational distance, and
- information functionals such as  $I(S; A_1|B_1)$ ,  $I(S; A_2|B_2)$  and  $I(S; A_1 A_2|B_1 B_2)$ .

Figure 8 illustrates the two kinds of pruning. In the Pruning Method A, we have  $(A_1, A_2, S) \sim p_{A_1 A_2 S}$  as shown in Figure 8(a). We select an appropriate threshold  $0 < \delta < 1$ , and consider any subset  $\mathcal{E} \subseteq \mathcal{A}_1 \times \mathcal{A}_2$  with  $\mathbb{P}[(A_1, A_2) \in \mathcal{E}] \geq 1 - \delta$ . We then take the joint pmf  $p_{A_1 A_2 S}$  and construct a new joint pmf  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}$  whose support set satisfies  $S(\tilde{A}_1, \tilde{A}_2, S) \subseteq \mathcal{E} \times \mathcal{S}$ . In other words, the edges belonging to  $\mathcal{E}^c \times \mathcal{S}$  in the bipartite graph of  $p_{A_1 A_2 S}$  indicated by red dashed lines in Figure 8(a) are removed (and the remaining edges are scaled appropriately) to define  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}$ .

In the Pruning Method B that is illustrated in Figure 8(b), we first select an appropriate threshold  $0 < \delta < 1$ , and consider any subset of edges  $\mathcal{E} \subseteq \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{S}$  with

$$p_{S|A_1, A_2}(s|a_1, a_2) > \delta, \quad \forall (a_1, a_2, s) \in \mathcal{E} \quad (141)$$

Edges that are not in  $\mathcal{E}$  (shown by red dashed lines in Fig. 8(b)) are removed, and the probability mass of the rest of the edges are scaled appropriately to construct a new  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}$  such that  $S(\tilde{A}_1, \tilde{A}_2, S) \subseteq \mathcal{E}$ . The precise details of the two kinds of pruning, and the required results pertaining to them are elaborated below.

#### A. Pruning Method A

Suppose that we have a random tuple  $(A_1, A_2, S, B_1, B_2)$  over  $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2$  distributed according to  $p_{A_1 A_2 S B_1 B_2}$  such that  $(A_1, A_2) - S - (B_1, B_2)$  and  $S(S) = \mathcal{S}$ . Pick any  $\delta, \eta > 0$  satisfying

$$\delta \leq m_S(1 - e^{-\eta}), \quad (142)$$

and consider any subset  $\mathcal{E} \subseteq \mathcal{A}_1 \times \mathcal{A}_2$  satisfying  $\mathbb{P}[(A_1, A_2) \in \mathcal{E}] \geq 1 - \delta$ . Let

$$\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) := \begin{cases} \frac{p_{A_1 A_2 S}(a_1, a_2, s)}{\mathbb{P}[(A_1, A_2) \in \mathcal{E} | S = s]} & \text{if } (a_1, a_2) \in \mathcal{E} \\ 0 & \text{if } (a_1, a_2) \in \mathcal{E}^c \end{cases}, \quad (143)$$

and  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} := \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S} p_{B_1 B_2 | S}$ . Define  $\Xi : [0, \frac{m_{SUV}}{|\mathcal{S}|}] \rightarrow \mathbb{R}$  by

$$\Xi(x) := \frac{|\mathcal{S}|x}{m_{SUV} - |\mathcal{S}|x} \log \frac{|\mathcal{S}|^2(m_{SUV} - |\mathcal{S}|x)}{x}. \quad (144)$$

*Theorem 5:*  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}$  defines a valid joint pmf with  $(\tilde{A}_1, \tilde{A}_2) - S - (B_1, B_2)$  for which the following holds:

(a)  $\|\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} - p_{A_1 A_2 S B_1 B_2}\|_1 \leq \frac{2\delta}{m_S - \delta} \leq 2(e^\eta - 1)$ .

(b) For any event  $\mathcal{F} \subseteq \mathcal{B}_1 \times \mathcal{B}_2$  and  $(a_1, a_2) \in \mathcal{E}$ ,

$$\mathbb{P}[(B_1, B_2) \in \mathcal{F} | (\tilde{A}_1, \tilde{A}_2) = (a_1, a_2)] \leq e^\eta \mathbb{P}[(B_1, B_2) \in \mathcal{F} | (A_1, A_2) = (a_1, a_2)]. \quad (145)$$

(c)  $D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \leq \eta$ .

(d)  $|H(S | \tilde{A}_1 B_1) - H(S | A_1 B_1)| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2(e^\eta - 1) \log |\mathcal{S}|$ .

(e)  $|H(S | \tilde{A}_1 \tilde{A}_2 B_1 B_2) - H(S | A_1 A_2 B_1 B_2)| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2(e^\eta - 1) \log |\mathcal{S}|$ .

*Proof:* For brevity, we will omit the subscripts on the joint distributions  $p_{A_1 A_2 S B_1 B_2}$  and  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}$  (and their marginals) throughout the proof; for example,  $p(a_1, s) = p_{A_1 S}(a_1, s)$  and  $\tilde{p}(a_1, s) = \tilde{p}_{A_1 S}(a_1, s)$ .

*Assertion (a):* For any  $(a_1, a_2, s, b_1, b_2)$  with  $(a_1, a_2) \in \mathcal{E}$  and  $p(a_1, a_2, s, b_1, b_2) > 0$ , we have

$$1 \leq \frac{\tilde{p}(a_1, a_2, s, b_1, b_2)}{p(a_1, a_2, s, b_1, b_2)} = \frac{\tilde{p}(a_1, a_2, s)}{p(a_1, a_2, s)} = \frac{p(s)}{\mathbb{P}[(A_1, A_2) \in \mathcal{E}, S = s]} \stackrel{(a)}{\leq} \frac{m_S}{m_S - \delta}, \quad (146)$$

which (a) follows because

$$1 \geq \frac{\mathbb{P}[(A_1, A_2) \in \mathcal{E}, S = s]}{p(s)} \geq \frac{p(s) - \mathbb{P}[(A_1, A_2) \in \mathcal{E}]}{p(s)} \geq 1 - \frac{\mathbb{P}[(A_1, A_2) \in \mathcal{E}]}{p(s)} \geq \frac{m_S - \delta}{m_S}. \quad (147)$$

Thus, from (146) we now have

$$\sum_{(a_1, a_2, s, b_1, b_2) \in \mathcal{E} \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2} \left| p(a_1, a_2, s, b_1, b_2) - \tilde{p}(a_1, a_2, s, b_1, b_2) \right|$$

$$\begin{aligned}
&\leq \sum_{(a_1, a_2, s, b_1, b_2) \in \mathcal{E} \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2} \left( \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} - 1 \right) p(a_1, a_2, s, b_1, b_2) \\
&= \left( \frac{\delta}{\mathbf{m}_S - \delta} \right) \mathbb{P}[(A_1, A_2) \in \mathcal{E}].
\end{aligned} \tag{148}$$

Next, we can bound the contribution for  $\mathcal{E}^c \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2$  by

$$\sum_{(a_1, a_2, s, b_1, b_2) \in \mathcal{E}^c \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2} |p(a_1, a_2, s, b_1, b_2) - \tilde{p}(a_1, a_2, s, b_1, b_2)| = \mathbb{P}[(A_1, A_2) \in \mathcal{E}^c] \leq \delta. \tag{149}$$

Combining (148) and (149) yields Assertion (b), since  $\delta + \frac{\delta}{\mathbf{m}_S - \delta} < \frac{2\delta}{\mathbf{m}_S - \delta} = 2(e^\eta - 1)$ .

*Assertion (b):* It is sufficient to prove the assertion for any singleton  $\mathcal{F} = \{(b_1, b_2)\}$  with  $(b_1, b_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ . Let  $(a_1, a_2) \in \mathcal{E}$ . Without loss of generality, we may assume that  $p(b_1, b_2|a_1, a_2) > 0$ , since if not, the claim is trivially true. Thus,

$$\begin{aligned}
\frac{\tilde{p}(b_1, b_2|a_1, a_2)}{p(b_1, b_2|a_1, a_2)} &= \frac{\sum_{s \in \mathcal{S}(S)} \tilde{p}(a_1, a_2, s) p(b_1, b_2|s)}{\tilde{p}(a_1, a_1) p(b_1, b_2|a_1, a_2)} \stackrel{(146)}{\leq} \left( \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} \right) \frac{\sum_{s \in \mathcal{S}(S)} p(a_1, a_2, s) p(b_1, b_2|s)}{\tilde{p}(a_1, a_2) p(b_1, b_2|a_1, a_2)} \\
&= \left( \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} \right) \frac{p(a_1, a_2, b_1, b_2)}{\tilde{p}(a_1, a_2) p(b_1, b_2|a_1, a_2)} \\
&= \left( \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} \right) \frac{p(a_1, a_2)}{\tilde{p}(a_1, a_2)} \stackrel{(a)}{\leq} \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} \stackrel{(142)}{\leq} e^\eta,
\end{aligned} \tag{150}$$

where (a) follows by rearranging (146), and summing over elements of  $\mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2$ .

*Assertion (c):*

$$\begin{aligned}
D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) &= D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S} \| p_{A_1 A_2 S}) \\
&= \sum_{(a_1, a_2, s) \in \mathcal{S}(\tilde{A}_1 \tilde{A}_2 S)} \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) \log \frac{\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s)}{p_{A_1 A_2 S}(a_1, a_2, s)} \\
&\stackrel{(146)}{\leq} \sum_{(a_1, a_2, s) \in \mathcal{S}(\tilde{A}_1 \tilde{A}_2 S)} \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) \log \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} \leq \eta.
\end{aligned} \tag{151}$$

*Assertion (d):*

$$\begin{aligned}
D_{KL}(\tilde{p}_{S|\tilde{A}_1 B_1} \| p_{S|A_1 B_1}) &:= \sum_{a_1 b_1} \tilde{p}_{\tilde{A}_1 B_1}(a_1, b_1) D_{KL}(\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) \| p_{S|A_1 B_1}(\cdot|a_1, b_1)) \\
&\leq D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \stackrel{(151)}{\leq} \log \frac{\mathbf{m}_S}{\delta - \mathbf{m}_S} = \eta.
\end{aligned} \tag{152}$$

An application of Pinsker's followed by Jensen's inequality yields the following.

$$\sum_{a_1 b_1} \tilde{p}_{\tilde{A}_1 B_1}(a_1, b_1) \|\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) - p_{S|A_1 B_1}(\cdot|a_1, b_1)\|_1 \leq \sqrt{2\eta}. \tag{153}$$

Let  $\mathcal{D} := \{(a_1, b_1) : \|\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) - p_{S|A_1 B_1}(\cdot|a_1, b_1)\|_1 \leq \sqrt[4]{2\eta}\}$ . Then, by Markov's inequality,

$$\mathbb{P}[(\tilde{A}_1, B_1) \in \mathcal{D}] \geq 1 - \sqrt[4]{2\eta}. \tag{154}$$

Further for every  $(a_1, b_1) \in \mathcal{D}$ , Lemma 2.5 of [22] guarantees that

$$\Delta_{a_1, b_1} := \left| H(S | (\tilde{A}_1, B_1) = (a_1, b_1)) - H(S | (A_1, B_1) = (a_1, b_1)) \right| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|}{\sqrt[4]{2\eta}}. \tag{155}$$

Finally,

$$\left| H(S | \tilde{A}_1 B_1) - H(S | A_1 B_1) \right| \leq \sum_{a_1, b_1} \tilde{p}_{\tilde{A}_1, B_1}(a_1, b_1) \Delta_{a_1, b_1} + \|\tilde{p}_{\tilde{A}_1, B_1} - p_{A_1, B_1}\|_1 \cdot \log |\mathcal{S}|$$

$$\begin{aligned}
&\leq \sum_{(a_1, b_1) \in \mathcal{D}} \tilde{p}_{\tilde{A}_1, B_1}(a_1, b_1) \Delta_{a_1, b_1} + \mathbb{P}[(\tilde{A}_1, B_1) \notin \mathcal{D}] \log |\mathcal{S}| + \frac{2\delta \log |\mathcal{S}|}{\mathbf{m}_S - \delta} \\
&\leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2(e^\eta - 1) \log |\mathcal{S}|.
\end{aligned} \tag{156}$$

*Assertion (e):* The proof of (e) is identical to that of (d), with the only difference being that the commencement of the argument is as follows.

$$D_{KL}(\tilde{p}_{S|\tilde{A}_1 \tilde{A}_2 B_1 B_2} \| p_{S|A_1 A_2 B_1 B_2}) \leq D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \stackrel{(151)}{\leq} \log \frac{\mathbf{m}_S}{\delta - \mathbf{m}_S} = \eta. \tag{157}$$

The remaining steps are identical to those in (d), with the exception that all four variables (either  $(\tilde{A}_1, \tilde{A}_2, B_1, B_2)$  or  $(A_1, A_2, B_1, B_2)$ ) appear in the conditioning. ■

### B. Pruning Method B

Let pmf  $p_{A_1 A_2 S B_1 B_2}$  over  $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{S} \times \mathcal{B}_1 \times \mathcal{B}_2$  be given such that  $(A_1, A_2) - S - (B_1, B_2)$  and  $\mathbf{S}(S) = \mathcal{S}$ . Let  $\delta \leq \mathbf{m}_S(1 - e^{-\eta})$  for some  $0 < \eta < 1$ , and let

$$\mathcal{E} := \{(a_1, a_2, s) \in \mathbf{S}(A_1, A_2, S) : p_{S|A_1 A_2}(s|a_1, a_2) > \delta\}, \tag{158a}$$

$$\mathcal{E}_s := \{(a_1, a_2) : (a_1, a_2, s) \in \mathcal{E}\}, \quad s \in \mathcal{S}, \tag{158b}$$

$$\mathcal{E}_{a_1, a_2} := \{s : (a_1, a_2, s) \in \mathcal{E}\}, \quad s \in \mathcal{S}. \tag{158c}$$

Define pmf  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}$  by

$$\tilde{p}_{A_1 A_2 S}(a_1, a_2, s) = \begin{cases} \frac{p_{A_1 A_2 S}(a_1, a_2, s)}{\mathbb{P}[(A_1, A_2) \in \mathcal{E}_s | S = s]}, & (a_1, a_2, s) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}, \tag{159}$$

and extend it to pmf  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}$  using the Markov chain  $(\tilde{A}_1, \tilde{A}_2) - S - (B_1, B_2)$  with  $\tilde{p}_{S B_1 B_2} := p_{S B_1 B_2}$ . Let  $\Xi(\cdot)$  be as defined in (144). The following properties hold for the pmf  $\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}$  defined above.

*Theorem 6:* (a)  $\|\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} - p_{A_1 A_2 S B_1 B_2}\|_1 \leq \frac{2\delta|\mathcal{S}|}{\mathbf{m}_S - \delta} = 2|\mathcal{S}|(e^\eta - 1)$ .

(b) Given  $\mathcal{F} \subseteq \mathcal{B}_1 \times \mathcal{B}_2$  and  $(a_1, a_2) \in \mathbf{S}(A_1, A_2)$  with  $\mathbb{P}[(B_1, B_2) \in \mathcal{F} | (A_1, A_2) = (a_1, a_2)] \leq \delta \mathbf{m}_{B_1 B_2 | S}$ ,

$$\mathbb{P}[(B_1, B_2) \in \mathcal{F} | (\tilde{A}_1, \tilde{A}_2) = (a_1, a_2)] \leq \delta \mathbf{m}_{B_1 B_2 | S} = 0. \tag{160}$$

(c)  $D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \leq \log \frac{\mathbf{m}_S}{\mathbf{m}_S - \delta} = \eta$ .

(d)  $|H(S|\tilde{A}_1 B_1) - H(S|A_1 B_1)| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2|\mathcal{S}|(e^\eta - 1) \log |\mathcal{S}|$ .

(e)  $|H(S|\tilde{A}_1 \tilde{A}_2 B_1 B_2) - H(S|A_1 A_2 B_1 B_2)| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2|\mathcal{S}|(e^\eta - 1) \log |\mathcal{S}|$ .

*Proof:* The proof follows on the same steps as Theorem 5 and the difference is in the evaluation of the probabilities of the normalization term in (159).

*Assertion (a):* Since for  $(a_1, a_2, s) \in \mathcal{E}$ , we have  $p(s|a_1, a_2) \leq \delta$ , we can argue that

$$\mathbb{P}[(A_1, A_2, S) \in \mathcal{E}^c] = \sum_{(a_1, a_2, s) \in \mathcal{E}^c} p_{A_1 A_2 S}(a_1, a_2, s) \leq \delta \sum_{a_1, a_2, s} p_{A_1 A_2}(a_1, a_2) \leq \delta |\mathcal{S}|. \tag{161}$$

Further, for any  $s \in \mathbf{S}(S)$ ,

$$\begin{aligned}
\mathbb{P}[(A_1, A_2) \in \mathcal{E}_s | S = s] &= \sum_{(a_1, a_2) \in \mathcal{E}_s} \frac{p_{A_1 A_2 S}(a_1, a_2, s)}{p_S(s)} = 1 - \sum_{(a_1, a_2) \notin \mathcal{E}_s} \frac{p_{A_1 A_2 S}(a_1, a_2, s)}{p_S(s)} \\
&\geq 1 - \delta \sum_{a_1, a_2} \frac{p_{A_1 A_2}(a_1, a_2)}{p_S(s)} \geq 1 - \frac{\delta}{\mathbf{m}_S}.
\end{aligned} \tag{162}$$

To prove (a), first note that for any  $(a_1, a_2, b_1, b_2)$  and  $s \in \mathcal{S}$ ,

$$p_{A_1 A_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2) = 0 \Rightarrow \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2) = 0, \quad (163)$$

and if  $(a_1, a_2, s) \in \mathcal{E}$  and  $p_{A_1 A_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2) > 0$ , then

$$1 \leq \frac{\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2)}{p_{A_1 A_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2)} = \frac{1}{\mathbb{P}[(A_1, A_2) \in \mathcal{E}_s | S = s]} \stackrel{(162)}{\leq} \frac{\mathfrak{m}_S}{\mathfrak{m}_S - \delta}. \quad (164)$$

Thus,

$$\sum_{\substack{b_1, b_2 \\ (a_1, a_2, s) \in \mathcal{E}}} |p_{A_1 A_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2) - \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2)| \leq \frac{\delta \mathbb{P}[(A_1, A_2, S) \in \mathcal{E}]}{\mathfrak{m}_S - \delta}. \quad (165)$$

Next, we can bound the contribution for  $\mathcal{E}^c \times \mathcal{B}_1 \times \mathcal{B}_2$  by

$$\sum_{\substack{b_1, b_2 \\ (a_1, a_2, s) \notin \mathcal{E}}} |p_{A_1 A_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2) - \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2}(a_1, a_2, s, b_1, b_2)| = \mathbb{P}[(A_1, A_2, S) \notin \mathcal{E}] \stackrel{(161)}{\leq} \delta |\mathcal{S}|. \quad (166)$$

Combining the above two equations together with the fact that  $\delta |\mathcal{S}| + \frac{\delta}{\mathfrak{m}_S - \delta} < \frac{2\delta |\mathcal{S}|}{\mathfrak{m}_S - \delta}$  establishes (a).

*Assertion (b):* We may assume that  $\mathcal{F} \neq \emptyset$ , or else the claim is trivial. Now, let  $(a_1, a_2) \in \mathcal{S}(A_1, A_2)$ . Then,  $(a_1, a_2) \in \mathcal{S}(\tilde{A}_1, \tilde{A}_2)$  because

$$\begin{aligned} \tilde{p}_{\tilde{A}_1, \tilde{A}_2}(a_1, a_2) \cdot \mathbb{P}[(A_1, A_2) \in \mathcal{E}_s | S = s] &= \sum_{s \in \mathcal{E}_{a_1, a_2}} p_{A_1 A_2 S}(a_1, a_2, s) \\ &= p_{A_1, A_2}(a_1, a_2) - \sum_{s \notin \mathcal{E}_{a_1, a_2}} p_{A_1 A_2 S}(a_1, a_2, s) \\ &\geq p_{A_1, A_2}(a_1, a_2) - \sum_{s \notin \mathcal{E}_{a_1, a_2}} \delta p_{A_1 A_2}(a_1, a_2) \\ &= (1 - \delta |\mathcal{S}|) p_{A_1, A_2}(a_1, a_2) > 0. \end{aligned} \quad (167)$$

Now, by hypothesis,

$$\mathbb{P}[(B_1, B_2) \in \mathcal{F} | (A_1, A_2) = (a_1, a_2)] = \sum_{(s', b_1, b_2) \in \mathcal{S}(S) \times \mathcal{F}} p_{S|A_1 A_2}(s' | a_1, a_2) p_{B_1 B_2|S}(b_1, b_2 | s') \leq \delta m_{B_1 B_2|S}. \quad (168)$$

Thus, for any  $(b_1, b_2) \in \mathcal{F}$  and  $s \in \mathcal{S}$  such that  $p_{B_1 B_2|S}(b_1, b_2 | s) > 0$ , it must be true that

$$p_{S|A_1 A_2}(s | a_1, a_2) \leq \sum_{(s', b_1, b_2) \in \mathcal{S}(S) \times \mathcal{F}} \frac{p_{S|A_1 A_2}(s' | a_1, a_2) p_{B_1 B_2|S}(b_1, b_2 | s')}{p_{B_1 B_2|S}(b_1, b_2 | s)} \stackrel{(168)}{\leq} \frac{\delta m_{B_1 B_2|S}}{p_{B_1 B_2|S}(b_1, b_2 | s)} \leq \delta, \quad (169)$$

Hence,  $(a_1, a_2, s) \notin \mathcal{E}$ . Thus, for any  $(b_1, b_2) \in \mathcal{F}$  and  $s \in \mathcal{S}$  such that  $p_{B_1 B_2|S}(b_1, b_2 | s) > 0$ ,  $\tilde{p}_{S|\tilde{A}_1 \tilde{A}_2}(s | a_1, a_2) = 0$ , and therefore, (b) follows.

*Assertion (c):* To prove (c), we proceed as follows.

$$\begin{aligned} D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) &= D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S} \| p_{A_1 A_2 S}) \\ &= \sum_{a_1, a_2, s} \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) \log \frac{\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s)}{p_{A_1 A_2 S}(a_1, a_2, s)} \\ &= \sum_{s \in \mathcal{S}(S), (a_1, a_2) \in \mathcal{E}_s} \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) \log \frac{1}{\mathbb{P}[(A_1, A_2) \in \mathcal{E}_s | S = s]} \\ &\stackrel{(162)}{\leq} \sum_{\substack{s \in \mathcal{S}(S) \\ (a_1, a_2) \in \mathcal{E}_s}} \tilde{p}_{\tilde{A}_1 \tilde{A}_2 S}(a_1, a_2, s) \log \frac{\mathfrak{m}_S}{\mathfrak{m}_S - \delta} = \eta. \end{aligned} \quad (170)$$

*Assertion (d):* Consider the following.

$$\sum_{a_1 b_1} \tilde{p}_{\tilde{A}_1 B_1}(a_1, b_1) D_{KL}(\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) \| p_{S|A_1 B_1}(\cdot|a_1, b_1)) \leq D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \stackrel{(170)}{\leq} \eta. \quad (171)$$

An application of Pinsker's followed by Jensen's inequality yields the following.

$$\sum_{a_1 b_1} \tilde{p}_{\tilde{A}_1 B_1}(a_1, b_1) \|\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) - p_{S|A_1 B_1}(\cdot|a_1, b_1)\|_1 \leq \sqrt{2\eta}. \quad (172)$$

Let  $\mathcal{D} := \{(a_1, b_1) : \|\tilde{p}_{S|\tilde{A}_1 B_1}(\cdot|a_1, b_1) - p_{S|A_1 B_1}(\cdot|a_1, b_1)\|_1 \leq \sqrt[4]{2\eta}\}$ . Then, by Markov's inequality,

$$\mathbb{P}[(\tilde{A}_1, B_1) \in \mathcal{D}] \geq 1 - \sqrt[4]{2\eta}. \quad (173)$$

Further for every  $(a_1, b_1) \in \mathcal{D}$ , by Lemma 2.5 of [22], we have

$$\Delta_{a_1, b_1} := \left| H(S|(\tilde{A}_1, B_1) = (a_1, b_1)) - H(S|(A_1, B_1) = (a_1, b_1)) \right| \leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|}{\sqrt[4]{2\eta}}. \quad (174)$$

Finally,

$$\begin{aligned} |H(S|\tilde{A}_1 B_1) - H(S|A_1 B_1)| &\leq \sum_{a_1, b_1} \tilde{p}_{\tilde{A}_1, B_1}(a_1, b_1) \Delta_{a_1, b_1} + \|\tilde{p}_{\tilde{A}_1, B_1} - p_{A_1, B_1}\|_1 \log |\mathcal{S}| \\ &\leq \sum_{(a_1, b_1) \in \mathcal{D}} \tilde{p}_{\tilde{A}_1, B_1}(a_1, b_1) \Delta_{a_1, b_1} + \mathbb{P}[(\tilde{A}_1, B_1) \notin \mathcal{D}] \log |\mathcal{S}| + \frac{2\delta |\mathcal{S}| \log |\mathcal{S}|}{\mathbf{m}_S - \delta} \\ &\leq \sqrt[4]{2\eta} \log \frac{|\mathcal{S}|^2}{\sqrt[4]{2\eta}} + 2|\mathcal{S}|(e^\eta - 1) \log |\mathcal{S}|. \end{aligned} \quad (175)$$

*Assertion (e):* The proof of (e) is identical to that of (d), with the only difference being that the commencement of the argument is as follows.

$$\begin{aligned} \sum_{a_1 a_2, b_1, b_2} \tilde{p}_{\tilde{A}_1 \tilde{A}_2, B_1, B_2}(a_1 a_2, b_1, b_2) D_{KL}(\tilde{p}_{S|\tilde{A}_1 \tilde{A}_2 B_1 B_2}(\cdot|a_1 a_2, b_1, b_2) \| p_{S|A_1 A_2 B_1 B_2}(\cdot|a_1 a_2, b_1, b_2)) \\ \leq D_{KL}(\tilde{p}_{\tilde{A}_1 \tilde{A}_2 S B_1 B_2} \| p_{A_1 A_2 S B_1 B_2}) \stackrel{(170)}{\leq} \eta. \end{aligned} \quad (176)$$

The remaining steps are identical to those in (d), with the exception that all four variables, i.e., either  $(\tilde{A}_1, \tilde{A}_2, B_1, B_2)$  or  $(A_1, A_2, B_1, B_2)$ , appear in the conditioning. ■

## APPENDIX I PROOF OF THEOREM 3

In each case, the overall approach is to show that a rate pair  $(r_{uv}, r_v)$  that is included in the outer bound rate region  $\mathcal{R}^\dagger(D)$  is also included in  $\mathcal{R}^D$ . To establish that, in each case, we pick a rate pair  $(r_{uv}, r_v) \in \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D)$  and establish explicitly that the rate pair is also an element of  $\mathcal{R}_k^*(D)$  for some  $k$ . The proof is then complete by invoking the achievability of  $\mathcal{R}_k^*(D)$  proved in Theorem 1 of Sec. III-B.

*Case A:* The proof follows from the following series of arguments:

$$\mathcal{R}^*(D) \stackrel{(a)}{\subseteq} \mathcal{R}(D) \stackrel{(b)}{\subseteq} \mathcal{R}^\dagger(D) \stackrel{(c)}{=} \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D) \stackrel{(d)}{\subseteq} \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D), \quad (177)$$

where (a) and (b) follow from Theorems 1 and 2 of Sec. III-B; (c) from Lemma 6 of Sec. III-A; and (d) from Lemma 5 of Sec. III-A, where the full-support condition specific to this case is incorporated.

*Case B:* Since  $S - V - U$ , from the outer bound in Theorem 2 can be rewritten as follows.

$$\mathcal{R}(D) \subseteq \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D) := \bigcup_{q \in \mathcal{P}_{D, |\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger} \left\{ (r_{uv}, r_v) : \begin{array}{l} r_{uv} \geq I(S; A, B|U) \\ r_v + r_{uv} \geq I(S; A, B, C|V) \end{array} \right\}. \quad (178)$$

Note that the sum rate constraint incorporates the chain  $S - V - U$  specific to this case. Now, let us fix a pmf  $q_{ABCSUV} \in \mathcal{P}_{D, |S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4}}^\dagger$ . Let  $f_U : \mathcal{A} \times \mathcal{B} \times \mathcal{U} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  and  $f_V : \mathcal{A} \times \mathcal{C} \times \mathcal{U} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  be such that

$$f_U(A, B, U) = f_V(A, C, V) = \mathbf{GK}^{ABU, ACV}. \quad (179)$$

Now,  $\mathbb{P}[f_U(A, B, U) \neq f_V(A, C, V)] = 0$  can be expanded using  $C - (A, B, V) - U$  in the following manner.

$$\sum_{a,b,v \in \mathbf{S}(A,B,V)} q_{ABV}(a,b,v) \sum_{u,c} q_{C|ABV}(c|a,b,v) q_{U|V}(u|v) \mathbb{1}\{f_U(a,b,u) = f_V(a,c,v)\} = 0. \quad (180)$$

Hence, for any  $(a,b,v) \in \mathbf{S}(A,B,V)$ ,

$$\sum_{u,c} q_{C|ABV}(c|a,b,v) q_{U|V}(u|v) \mathbb{1}\{f_U(a,b,u) = f_V(a,c,v)\} = 0. \quad (181)$$

Then, by Lemma 10 of Appendix K, it follows that  $(a,b,v) \in \mathbf{S}(A,B,V)$ , there exists  $\eta(a,b,v)$  such that

$$\mathbb{P}[f_U(A, B, U) = \eta(a,b,v) \mid (A, B, V) = (a,b,v)] = 1, \quad (182a)$$

$$\mathbb{P}[f_V(A, C, V) = \eta(a,b,v) \mid (A, B, V) = (a,b,v)] = 1. \quad (182b)$$

Hence, we are guaranteed of the existence of a function  $g_V : \mathcal{A} \times \mathcal{B} \times \mathcal{V} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  such that

$$g_V(A, B, V) = f_U(A, B, U) = f_V(A, C, V). \quad (183)$$

Thus,  $\tilde{A} := g_V(A, B, V)$  is a function of  $(A, B, U)$  as well as a function of  $(A, B, V)$ . Then, notice that

$$r_{uv} \stackrel{(178)}{\geq} I(S; A, B|U) \geq I(S; \tilde{A}|U), \quad (184a)$$

$$r_v + r_{uv} \stackrel{(178)}{\geq} I(S; A, B, C|V) \geq I(S; \tilde{A}|V), \quad (184b)$$

and further  $\tilde{A} - S - (U, V)$ . Lastly, we see that the alphabet of  $\tilde{A} \subseteq \mathcal{A} \times \mathcal{C} \times \mathcal{V}$ , and hence

$$|\tilde{A}| \leq (|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4})^2 |\mathcal{U}|. \quad (185)$$

From (183)-(185), we conclude that  $q_{\tilde{A}SUV} \in \mathcal{P}_{D, (|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4})^2 |\mathcal{U}|}^*$  and  $(r_{uv}, r_v) \in \mathcal{R}_{(|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4})^2 |\mathcal{U}|}^*(D)$ .

Since  $q_{ABCSUV}$  is any arbitrary pmf in  $\mathcal{P}_{D, |S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4}}^\dagger$ , it follows that

$$\mathcal{R}_{|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4}}^\dagger(D) \subseteq \mathcal{R}_{(|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4})^2 |\mathcal{U}|}^*(D). \quad (186)$$

*Case C:* Let  $(r_{uv}, r_v) \in \mathcal{R}_{|S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4}}^\dagger(D)$ . Then, there must exist  $q_{ABCSUV} \in \mathcal{P}_{D, |S|(|S|+6)|\hat{S}|^{|\mathcal{U}|+4}}^\dagger$ , and a function  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{S}$  such that  $\mathbb{E}[d(S, f(\mathbf{GK}^{ABU, ACV}))] \leq D$ , and

$$r_{uv} \geq I(S; A, B|U), \quad (187a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) + I(S; B|A, C, U). \quad (187b)$$

Since  $(A, B, C) - S - (U, V)$ , we have for any  $(a,b,c,u,v) \in \mathbf{S}(A,B,C,U,V)$ ,

$$\begin{aligned} q_{ABCUV}(a,b,c,u,v) &= \sum_{s \in \mathbf{S}(S)} q_{ABCSUV}(a,b,c,s,u,v) \\ &= \sum_{s \in \mathbf{S}(S)} q_{ABCS}(a,b,c,s) q_{UV|S}(u,v|s) \\ &\leq \frac{1}{\mathbf{m}_S} \sum_{s \in \mathbf{S}(S)} q_{ABCS}(a,b,c,s) q_{SUV}(s,u,v) \\ &\leq \frac{1}{\mathbf{m}_S} q_{UV}(u,v) \sum_{s \in \mathbf{S}(S)} q_{ABCS}(a,b,c,s) \end{aligned}$$



$$\leq \frac{1}{\mathfrak{m}_S} q_{ABC}(a, b, c) q_{UV}(u, v). \quad (188)$$

Further, since  $S(S, U) = S(S) \times S(U)$ , it follows that for any  $(a, b, c, u, v) \in S(A, B, C, U, V)$ ,

$$\begin{aligned} q_{ABCUV}(a, b, c, u, v) &= \sum_{s \in S} q_{ABCS}(a, b, c, s) q_{U|S}(u|s) q_{V|U}(v|u) \\ &\geq \mathfrak{m}_{U|S} \sum_{s \in S} q_{ABCS}(a, b, c, s) q_{V|U}(v|u) \\ &= \mathfrak{m}_{U|S} q_{ABC}(a, b, c) q_{V|U}(v|u) \\ &\geq \mathfrak{m}_{U|S} q_{ABC}(a, b, c) q_{UV}(u, v). \end{aligned} \quad (189)$$

Hence,

$$S(A, B, C, U, V) = S(A, B, C) \times S(U, V). \quad (190)$$

Then, from Lemma 1 of Sec. II-B, it follows that

$$\text{GK}^{ABU, ACV} \equiv (\text{GK}^{AB, AC}, \text{GK}^{U, V}). \quad (191)$$

Now, define  $\tilde{A} := \text{GK}^{AB, AC}$ , and let  $p_{\tilde{A}SUV}$  denote the pmf of  $(\tilde{A}, S, U, V)$ . Then, by construction the Markov chain  $\tilde{A} - S - (U, V)$  holds. Further,

$$H(\tilde{A}, \text{GK}^{U, V} \mid \text{GK}^{\tilde{A}V, \tilde{A}U}) = 0. \quad (192)$$

Then, by (191), there must exist a function  $\tilde{f} : \tilde{\mathcal{A}} \times \mathcal{V} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  such that

$$\tilde{f}(\text{GK}^{\tilde{A}U, \tilde{A}V}) \equiv \text{GK}^{ABU, ACV}. \quad (193)$$

Then, from (187), we see that

$$\mathbb{E}[d(S, f(\tilde{f}(\text{GK}^{\tilde{A}V, \tilde{A}U})))] \leq D. \quad (194)$$

Since  $\tilde{\mathcal{A}} \subseteq \mathcal{A} \times \mathcal{C}$ , we are guaranteed that  $|\tilde{\mathcal{A}}| \leq (|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2$ , and hence,  $p_{\tilde{A}SUV} \in \mathcal{P}_{D, (|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*$ . Further,

$$r_{uv} \geq I(S; A, B|U) \geq I(S; \text{GK}^{AC, AB}|U) = I(S; \tilde{A}|U), \quad (195a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) \geq I(S; \text{GK}^{AC, AB}|V) = I(S; \tilde{A}|V). \quad (195b)$$

Consequently,  $(r_{uv}, r_v) \in \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D)$ . Since the rate pair was chosen arbitrarily, it follows that

$$\mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D) \subseteq \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D). \quad (196)$$

*Case D:* In this case, since  $H(S|U) = 0$ , the receiver with side information  $U$  does not require the encoder to communicate any message. Hence, we see that if  $(r_{uv}, r_v) \in \mathcal{R}(D)$ , then any  $(r'_{uv}, r'_v) \in \mathcal{R}(D)$  provided  $r'_{uv} + r'_v \geq r_{uv} + r_v$ , i.e., only the sum-rate constraint is relevant.

Let  $(r_{uv}, r_v) \in \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D)$ . Then, there must exist a joint pmf  $q_{ABCSUV} \in \mathcal{P}_{D, |\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger$  and a function  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that  $\mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D$ , and

$$r_{uv} \geq I(S; A, B|U) = 0, \quad (197a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) + I(S; B|A, C, U, V) = I(S; A, C|V). \quad (197b)$$

Let functions  $f_U, g_V$  be defined such that  $f_U(A, B, U) = f_V(A, C, V) = f(\text{GK}^{ABU, ACV})$ . Now, since  $H(S|U) = 0$ , it follows that  $B - (A, C, U) - V$ . Then, we have

$$\sum_{a, c, u} q_{ACU}(a, c, u) \sum_{b, v} q_{B|ACU}(b|a, c, u) p_{V|U}(v|u) \mathbb{1}\{f_U(a, b, u) = f_V(a, c, v)\} = 0. \quad (198)$$

Then, by Lemma 10 of Appendix K, for each  $(a, c, u) \in \mathcal{S}(A, C, U)$ , there must exist an  $\eta(a, c, u)$  such that

$$\mathbb{P}[f_U(A, B, U) = \eta(a, c, u) \mid (A, C, U) = (a, c, u)] = 1, \quad (199a)$$

$$\mathbb{P}[f_V(A, C, V) = \eta(a, c, u) \mid (A, C, U) = (a, c, u)] = 1. \quad (199b)$$

Hence, there must exist a function  $\tilde{f}_U : \mathcal{A} \times \mathcal{C} \times \mathcal{U} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  such that

$$\mathbb{P}[\tilde{f}_U(A, C, U) \neq f_V(A, C, V)] = 0. \quad (200)$$

Further, we also have

$$\mathbb{E}[d(S, \tilde{f}_U(A, C, U))] = \mathbb{E}[d(S, f_V(A, C, V))] = \mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D \quad (201)$$

Now, set  $\tilde{A} = (A, C)$ , and let  $q_{\tilde{A}, S, U, V}$  be the joint pmf of  $(\tilde{A}, S, U, V)$ . From (200), (201), and from the fact that  $|\tilde{A}| \leq |\mathcal{A}||\mathcal{C}|$ , it follows that  $q_{\tilde{A}, S, U, V} \in \mathcal{P}_{D, (|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*$ . Further, from (197a) and (197b), it follows that

$$r_{uv} \geq I(S; A, B|U) = I(S; \tilde{A}|U) = 0, \quad (202a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) = I(S; \tilde{A}|V). \quad (202b)$$

Hence,  $(r_{uv}, r_v) \in \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D)$ . Since the rate pair was chosen arbitrarily, it follows that

$$\mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D) \subseteq \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D). \quad (203)$$

*Case E:* Pick  $(r_{uv}, r_v) \in \mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D)$ . Then, there must exist  $q_{ABCSUV} \in \mathcal{P}_{D, |\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger$  and a function  $f : \mathcal{A} \times \mathcal{C} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that  $\mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D$ , and

$$r_{uv} \geq I(S; A, B|U), \quad (204a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) + I(S; B|A, C, U), \quad (204b)$$

where in the sum rate we have incorporated side information degradedness. Since  $(A, B, C) - S - U - V$ , and  $\mathcal{S}(U, V) = \mathcal{S}(U) \times \mathcal{S}(V)$ , it follows that

$$\mathcal{S}(A, B, C, U, V) = \mathcal{S}(A, B, C, U) \times \mathcal{S}(V). \quad (205)$$

Then, an invocation of Lemma 2 of Sec. II-B yields

$$\text{GK}^{ABU, ACV} \equiv \text{GK}^{ABU, AC}. \quad (206)$$

Hence, there must exist a function  $\tilde{f} : \mathcal{A} \times \mathcal{C} \rightarrow \mathcal{A} \times \mathcal{C} \times \mathcal{V}$  such that  $\tilde{f}(\text{GK}^{ABU, AC}) = \text{GK}^{ABU, ACV}$ . Using this function, let us now define

$$\tilde{A} := \tilde{f}(\text{GK}^{ABU, AC}). \quad (207)$$

Then,  $\hat{\mathcal{S}} = f(\tilde{A})$ , and

$$\mathbb{E}[d(S, \hat{\mathcal{S}})] = \mathbb{E}[d(S, f(\tilde{A}))] = \mathbb{E}[d(S, f(\tilde{f}(\text{GK}^{ABU, AC})))] = \mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D. \quad (208)$$

Since  $\tilde{A}$  is both a function of  $(A, B, U)$  and  $(A, C)$ , we see that  $|\tilde{A}| \leq |\mathcal{A}||\mathcal{C}| = (|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2$ , and

$$r_{uv} \geq I(S; A, B|U) = I(S; A, B, \tilde{A}|U) \geq I(S; \tilde{A}|U), \quad (209a)$$

$$r_v + r_{uv} \geq I(S; A, C, \tilde{A}|V) + I(S; B|A, C, U) \geq I(S; \tilde{A}|V). \quad (209b)$$

Further,  $\tilde{A} - S - U - V$ . Then, from (208) and (209), we conclude that  $q_{\tilde{A}, S, U, V} \in \mathcal{P}_{D, (|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*$  and  $(r_{uv}, r_v) \in \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D)$ . Since  $(r_{uv}, r_v)$  was chosen arbitrarily, it follows that

$$\mathcal{R}_{|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4}}^\dagger(D) \subseteq \mathcal{R}_{(|\mathcal{S}|(|\mathcal{S}|+6)|\hat{\mathcal{S}}|^{|\mathcal{U}|+4})^2}^*(D). \quad (210)$$

*Case F:* Pick  $(r_{uv}, r_v) \in \mathcal{R}_{|S|(|S|+6)|\hat{S}|^{|U|+4}}^\dagger(D)$ . Then, there must exist pmf  $q_{ABCSUV} \in \mathcal{P}_{D, |S|(|S|+6)|\hat{S}|^{|U|+4}}^\dagger$  and a reconstruction function  $f : \mathcal{A} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that  $\mathbb{E}[d(S, f(\text{GK}^{ABU, ACV}))] \leq D$ , and

$$r_{uv} \geq I(S; A, B|U), \quad (211a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) + I(S; B|A, C, U, V). \quad (211b)$$

Suppose that  $H(V|S) = 0$ . Since  $H(A|\text{GK}^{ABU, ACV}) = 0$ , we have

$$\begin{aligned} I(\text{GK}^{ABU, ACV}; U, V|S) &\leq I(\text{GK}^{ABU, ACV}, A, C; U, V|S) \\ &= I(A, C; U, V|S) + I(\text{GK}^{ABU, ACV}; U, V|A, C, S) \\ &\stackrel{(a)}{=} I(\text{GK}^{ABU, ACV}; U|A, C, S, V) \stackrel{(b)}{=} 0, \end{aligned} \quad (212)$$

where in (a), we use  $(A, C) - S - (U, V)$  and that  $V$  is a function of  $S$ ; and in (b), we use the fact that  $\text{GK}^{ABU, ACV}$  is a function of  $(A, V)$ . Hence,

$$\text{GK}^{ABU, ACV} - S - (U, V). \quad (213)$$

Similarly, when  $H(U|S) = 0$ , we arrive at the same conclusion by reversing the roles of  $U$  and  $V$ . By defining  $\tilde{A} := \text{GK}^{ABU, ACV}$ , we see that  $\tilde{A} - S - (U, V)$  with  $|\mathcal{A}| \leq (|S|(|S|+3)|\hat{S}|^{|U|}+3)^3$ , and that  $\mathbb{E}[d(S, f(\tilde{A}))] \leq D$ . Further,

$$r_{uv} \geq I(S; A, B|U) \geq I(S; \tilde{A}|U) \quad (214a)$$

$$r_v + r_{uv} \geq I(S; A, C|V) + I(S; B|A, C, U, V) \geq I(S; \tilde{A}|V) \quad (214b)$$

Hence,  $(r_{uv}, r_v) \in \mathcal{R}_{(|S|(|S|+6)|\hat{S}|^{|U|+4})^3}^*(D)$ . Since the rate pair was chosen arbitrarily, it follows that

$$\mathcal{R}_{|S|(|S|+6)|\hat{S}|^{|U|+4}}^\dagger(D) \subseteq \mathcal{R}_{(|S|(|S|+6)|\hat{S}|^{|U|+4})^2}^*(D). \quad (215)$$

## APPENDIX J PROOF OF THEOREM 4

Clearly we have  $\mathcal{R}_{\text{QB}}^*(D) \subseteq \mathcal{R}^*(D)$ . Since Theorem 3 applies to Cases A, B and C, we need only show that  $\mathcal{R}_{\text{QB}}^*(D) \supseteq \mathcal{R}^*(D)$  in each case. Let  $(r_{uv}, r_v) \in \mathcal{R}^*(D)$ . Then, there must exist  $q_{ASUV} \in \mathcal{P}_{D, |S|+2}^\dagger$ , and a function  $f : \mathcal{A} \times \mathcal{V} \rightarrow \hat{\mathcal{S}}$  such that  $\mathbb{E}[d(S, f(\text{GK}^{AU, AV}))] \leq D$ , and

$$r_{uv} \geq I(S; A|U), \quad (216a)$$

$$r_v + r_{uv} \geq I(S; A|V). \quad (216b)$$

The rest of the proof for each case is as follows.

*Case A:* Since  $S(S, U, V) = S(S) \times S(U, V)$ , it follows from that Lemma 1 of Sec. II-B that

$$\text{GK}^{AU, AV} = (A, \text{GK}^{U, V}). \quad (217)$$

Since  $H(\text{GK}^{U, V}(V)) = 0$ , we conclude from the above equation that  $\text{GK}^{AU, AV}(A, V) \equiv A$ .

Consequently, there must exist a function  $\tilde{f} : \mathcal{A} \rightarrow \hat{\mathcal{S}}$  such that reconstruction  $\tilde{f}(A) = f(\text{GK}^{AU, AV})$ . Define random variable  $\hat{S} = \tilde{f}(A) = f(\text{GK}^{AU, AV})$ . Then,  $(U, V) - S - A - \hat{S}$  and

$$r_{uv} \geq I(S; A|U) \geq I(S; \hat{S}|U), \quad (218a)$$

$$r_v + r_{uv} \geq I(S; A|V) \geq I(S; \hat{S}|V). \quad (218b)$$

Further,

$$D \geq \mathbb{E}[d(S, f(\text{GK}^{AU, AV}))] = \mathbb{E}[d(S, \hat{S})], \quad (219)$$

Hence, it follows that  $(r_{uv}, r_v) \in \mathcal{R}_{\text{QB}}^*(D)$ .

*Case B:* In this setting, the Markov chain  $A - S - U - V$  and the support condition  $S(U, V) = S(U) \times S(V)$  together imply  $S(A, U, V) = S(A, U) \times S(V)$ . Thus, from Lemma 2 of Sec. II-B, we see that

$$\text{GK}^{AU, AV} \equiv \text{GK}^{AU, A} \equiv A. \quad (220)$$

The rest of the proof then follows by setting  $\hat{S} = f(A) = f(\text{GK}^{AU, AV})$  and is identical to that of Case A above.

*Case C:* Repeating the steps of Case F of Theorem 3, we see in this case that  $\text{GK}^{AU, AV} = S - (U, V)$ . The proof is then complete by choosing  $\hat{S} = f(\text{GK}^{AU, AV})$ , and verifying that  $\hat{S} - S - (U, V)$ , (218) and (219) hold.

## APPENDIX K

### A LEMMA ON FUNCTIONS OF INDEPENDENT RANDOM VARIABLES

*Lemma 10:* Let  $X \sim p$  and  $Y \sim q$  be independent random variables. Suppose that we are given a finite set  $\mathcal{Z}$ , and functions  $f : \mathcal{X} \rightarrow \mathcal{Z}$  and  $g : \mathcal{Y} \rightarrow \mathcal{Z}$  satisfying

$$\mathbb{P}[f(X) \neq g(Y)] \leq \delta < \frac{1}{25}. \quad (221)$$

There exists an  $z^* \in \mathcal{Z}$  such that

$$\mathbb{P}[f(X) = z^*] \geq \frac{1 + \sqrt{1 - \delta - \sqrt{\delta}}}{2}, \quad (222a)$$

$$\mathbb{P}[g(Y) = z^*] \geq \frac{1 + \sqrt{1 - \delta - \sqrt{\delta}}}{2}. \quad (222b)$$

*Proof:* Let for  $z \in \mathcal{Z}$ ,  $p_z := \mathbb{P}[f(X) = z]$  and  $q_z := \mathbb{P}[g(Y) = z]$ . Then for any  $z \in \mathcal{Z}$ ,

$$|p_z - q_z| = \sqrt{(p_z - q_z)^2} \leq \sqrt{p_z(1 - q_z) + (1 - p_z)q_z} \leq \sqrt{\mathbb{P}[f(X) \neq g(Y)]} \leq \sqrt{\delta}. \quad (223)$$

Then, the following holds

$$\begin{aligned} \sum_z p_z^2 &\geq \sum_z p_z(p_z + q_z - q_z) \geq \sum_z p_z q_z - \sum_z p_z |q_z - p_z| \\ &\geq \mathbb{P}[f(X) = g(Y)] - \sqrt{\delta} \geq 1 - \delta - \sqrt{\delta} > \frac{19}{25}. \end{aligned} \quad (224)$$

Now, let  $M_p = \max_z p_z$ ,  $M_q = \max_z q_z$ ,  $z_p = \arg \max_z p_z$  and  $z_q = \arg \max_z q_z$ . Then,  $M_p > \frac{19}{25}$ , since

$$M_p = \sum_z M_p p_z \geq \sum_z p_z^2 \geq 1 - \delta - \sqrt{\delta} > \frac{19}{25}. \quad (225)$$

Also,  $z_p$  and  $z_q$ , have to be identical, because  $q_z$  can strictly exceed  $\frac{1}{2}$  for only one  $z \in \mathcal{Z}$ , and

$$q_{z_p} \geq p_{z_p} - |p_{z_p} - q_{z_p}| \stackrel{(223)}{\geq} M_p - \sqrt{\delta} > \frac{19}{25} - \frac{1}{5} > \frac{1}{2}. \quad (226)$$

Since the problem is symmetric, we see that  $M_q = q_{z_q} > 1 - \delta - \sqrt{\delta}$  as well. At this point we are done if the RHS of (222a) and (222b) were  $1 - \delta - \sqrt{\delta}$ . To improve this estimate, consider for  $\gamma \in (0, 1)$ , the following optimization problem and its solution.

$$\max_{\substack{\forall z \in \mathcal{Z}, r_z \in [0, \gamma] \\ \sum_z r_z = 1}} \sum_z r_z^2 = \left\lfloor \frac{1}{\gamma} \right\rfloor \gamma^2 + \left(1 - \left\lfloor \frac{1}{\gamma} \right\rfloor \gamma\right)^2 \leq \gamma^2 + (1 - \gamma)^2. \quad (227)$$

Since  $M_p$  is the maximum positive value taken by  $\{p_z : z \in \mathcal{Z}\}$ , it follows that

$$1 - \delta - \sqrt{\delta} \leq \sum_z p_z^2 \stackrel{(223)}{\leq} M_p^2 + (1 - M_p)^2, \quad (228)$$

which necessitates that

$$\mathbb{P}[f(X) = z_p] = M_p \in \left[0, \frac{1 - \sqrt{1 - \delta - \sqrt{\delta}}}{2}\right] \cup \left[\frac{1 + \sqrt{1 - \delta - \sqrt{\delta}}}{2}, 1\right]. \quad (229)$$

Selecting the choice meeting (225) eliminates the first interval. Finally, reversing the roles of  $p$  and  $q$  along with the fact that  $z_p = z_q$  concludes the proof. ■

## REFERENCES

- [1] B. N. Vellambi and R. Timo, "Lossy broadcasting with common transmitter-receiver reconstructions," 2014 Australian Communications Theory Workshop (AusCTW 2014), Sydney, Australia, 2014, pp. 138-143.
- [2] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1636-1654, Aug. 2004.
- [3] C. Tian and S. N. Diggavi, "side information scalable source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5591-5608, Dec. 2008.
- [4] R. Timo, T. Chan and A. Grant, "Rate distortion with side information at many decoders," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5240 - 5257, Aug. 2011.
- [5] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727-734, Nov. 1985.
- [6] A. Kaspi, "Rate-distortion function when side information may be present at the decoder," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 2031-2034, Nov. 1994.
- [7] S. Watanabe, "The rate-distortion function for product of two sources with side information at decoders," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5678-5691, Sep. 2013.
- [8] R. Timo, T. J. Oechtering and M. Wigger, "Source coding problems with conditionally less noisy side information," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5516-5532, Sep. 2014.
- [9] Y.-K. Chia, "On multiterminal source coding with list decoding constraints," *IEEE International Symposium on Information Theory (ISIT 2014)*, Honolulu, USA, June 2014.
- [10] M. Benammar and A. Zaidi, "Rate-distortion function for a Heegard-Berger problem with two sources and degraded reconstruction sets," *ArXiv:1508.06434*, Aug. 2015.
- [11] S. Unal and A. Wager, "Vector-Gaussian rate-distortion function with variable side information," *proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, USA, June 2014.
- [12] S. Unal and A. Wager, "Vector-Gaussian multi-decoder rate-distortion: Trace Constraints," *proc. 50th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, USA, Mar. 2016.
- [13] Y. Steinberg, "Coding and common reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4995-5010, Nov. 2009.
- [14] B. Ahmadi, R. Tandon, O. Simeone, and H. V. Poor, "Heegard-Berger and cascade source coding problems with common reconstruction constraints," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1458-1474, Mar. 2013.
- [15] M. Benammar and A. Zaidi, "Rate-Distortion of a Heegard-Berger Problem with Common Reconstruction Constraint," *International Zurich Seminar on Communications (IZS)*, pp. 150 - 155, Zurich, Mar. 2016.
- [16] B. N. Vellambi and R. Timo, "The Heegard-Berger problem with common receiver reconstructions," in *2013 IEEE Information Theory Workshop (ITW 2013)*, Sep. 2013, pp. 1-5.
- [17] B. N. Vellambi and R. Timo, "Successive refinement with common receiver reconstructions," 2014 IEEE International Symposium on Information Theory, Honolulu, HI, 2014, pp. 2664-2668.
- [18] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265-444, 2007.
- [19] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149-162, 1973.
- [20] C. Tian and S. N. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side informations," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2946-2960, Aug. 2007.
- [21] R. Yeung, "Information theory and network coding," Springer 2008.
- [22] I. Csiszár and J. Körner, "Information theory: Coding theorems for discrete memoryless systems," 2nd edition, Cambridge University Press, 2011.
- [23] W. Gu, M. Effros and M. Bakshi, "A continuity theory for lossless source coding over networks," *Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, Sep. 2008.
- [24] J. Chen and A. Wagner, "A semicontinuity theorem and its application to network source coding," *IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July 2008.
- [25] T. S. Han and K. Kobayashi, "A dichotomy of functions  $f(x, y)$  of correlated sources  $(x, y)$  from the viewpoint of the achievable rate region," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 69-76, Jan. 1987.
- [26] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, 1st edition, Cambridge University Press, Jan. 2012.
- [27] T. M. Apostol, *Mathematical Analysis*, 2nd edition, Addison-Wesley, 1974.