

# An Achievable Rate-Distortion Region for Multiple Descriptions Source Coding Based on Coset Codes

Farhad Shirani and S. Sandeep Pradhan\*  
Dept. of Electrical Engineering and Computer Science  
Univ. of Michigan, Ann Arbor, MI.

## Abstract

We consider the problem of multiple descriptions (MD) source coding and propose new coding strategies involving both unstructured and structured coding layers. Previously, the most general achievable rate-distortion (RD) region for the  $l$ -descriptions problem was the Combinatorial Message Sharing with Binning (CMSB) region. The CMSB scheme utilizes unstructured quantizers and unstructured binning. In the first part of the paper, we show that this strategy can be improved upon using more general unstructured quantizers and a more general unstructured binning method. In the second part, structured coding strategies are considered. First, structured coding strategies are developed by considering specific MD examples involving three or more descriptions. We show that application of structured quantizers results in strict RD improvements when there are more than two descriptions. Furthermore, we show that structured binning also yields improvements. These improvements are in addition to the ones derived in the first part of the paper. This suggests that structured coding is essential when coding over more than two descriptions. Using the ideas developed through these examples we provide a new unified coding strategy by considering several structured coding layers. Finally, we characterize its performance in the form of an inner bound to the optimal rate-distortion region using computable single-letter information quantities. The new RD region strictly contains all of the previous known achievable regions.

## 1 Introduction

The Multiple-Descriptions (MD) source coding problem arises naturally in a number of applications such as transmission of video, audio and speech over packet networks and fading channels [1] [2]. The multiple-descriptions (MD) source coding setup describes a communications setting consisting of one encoder and several decoders. The encoder receives a discrete memoryless source and wishes to compress it into several *descriptions*. Each decoder receives a specific subset of these descriptions through noiseless links, and produces a reconstruction of the source vector with respect to its own distortion criterion. The parameters of interest are the rates required for transmitting the description and the resulting distortions at the decoders. The objective is to design communications schemes which result in the optimal asymptotic trade-off between these two groups of parameters. The problem has been studied extensively [3] [4] [5] [6] [7] [10], however, the optimal asymptotically achievable rate-distortion (RD) is not known even for the most elementary case when only two descriptions are considered. The two-descriptions setup is depicted in Figure 1. Evidently, for the individual decoders (which receive only one description) to perform optimally the encoder must transmit the two-descriptions according to the optimal Point-to-Point (PtP) source coding schemes. This may require the two-descriptions to be similar to each other. On the other hand, if the descriptions are similar, one of them would be redundant at the central decoder (which receive two descriptions). In fact, this

---

\*This work was supported by NSF grant CCF-1111061 and CCF-1422284. This work was presented in part at IEEE International Symposium on Information Theory (ISIT), July 2014.

decoder requires the two-descriptions to be different from one another in order to yield a better reconstruction. The main challenge in the MD problem is to strike a balance between these two situations. The best known achievable region for the this communications setting is due to Zhang and Berger [6]. In the Zhang-Berger (ZB) strategy, the encoder in the first step sends a common and coarsely quantized version of the source on both descriptions, then in the next step, the encoder sends individual refinements for each decoder on the corresponding descriptions. The ZB coding strategy was generalized in [7] for the case where there are more than two descriptions. In this strategy, first, a common coarsely quantized version of the source is sent to all the decoders, then in the next step, several refinement layers are transmitted. For the symmetric  $l$ -descriptions problem, a coding scheme based on random binning was considered in [8] which outperforms the VKG scheme. This involves generation of independent codebooks followed by random binning. Although the MD problem has a centralized encoder, the strategy involving random binning was proved to be useful. This was further improved upon by a new coding scheme in [9] based on certain parity-check codes. However all the three schemes do not fully exploit the common-information among every subset of individual descriptions. For example in the three-descriptions problem, there can be common-information between the first and second descriptions which is not common with the third description. A new coding scheme called Combinatorial Message Sharing with Binning (CMSB) was considered in [10, 12] which provided a unified achievable RD region for the general  $l$ -descriptions problem. This scheme provided a grand unification of the schemes based on conditional codebooks and the schemes based on random binning, which in turn results in the largest achievable RD region for the problem and subsumes all previous coding schemes. The name is due to the combinatorial number of common-component codebooks present. It can be noted that CMSB scheme is based on a construction of random codes where the codewords are mutually independent, and where the codebooks do not have any algebraic structure.

In this paper, we provide a new coding strategy for the general  $l$ -descriptions problem which strictly subsumes CMSB strategy which is the best known in the literature till now. The coding strategy is based on the common-information perspective. Taking a cue from the two-descriptions ZB strategy, we propose that for the general  $l$ -descriptions problem the encoder constructs a common constituent codebook for each subset of the  $2^l - 1$  decoders. So, for each subset of the decoders there is one common component in the overall coding scheme. This implies that the number of constituent codebooks grows double-exponentially in  $l$ . However, we prove that only an asymptotically exponential number of the codebooks are necessary in terms of contributing to the rate-distortion region, and the rest are redundant. This significantly simplifies the coding strategy. As an example, for the  $l = 3$  case, there are  $2^{2^3-1} = 128$  possible common code components, but only 17 of the corresponding codebooks are non-redundant. It turns out that one can identify all of the non-redundant codebooks by associating them with the Sperner families of sets [13]. As a result, we call the new scheme the Sperner Set Coding (SSC) scheme. The CMSB scheme utilizes 14 codebooks for the 3-descriptions problem. We prove analytically that the addition of the 3 new codebooks in the SSC scheme results in an improved achievable RD region. In other words, we show analytically that the CMSB scheme is not complete. Additionally, we propose a generalized binning approach which improves upon the CMSB scheme and further enhances the SSC scheme. We characterize the asymptotic performance of this coding scheme using computable single-letter information quantities. This forms the first part of the paper. Similar to the coding scheme of CMSB, the SSC scheme uses random unstructured codes.

It has been observed in several other multi-terminal communications settings such as the Broadcast Channel (BC) [16], Interference Channel (IC) [17], variations of the MAC channel [18] [19] and the Distributed Source Coding (DSC) problem [20], that the application of algebraic structured codes results in improvements over random unstructured codes in the asymptotic performance limits. Based on the inherent dualities between the multi-terminal communication problems and the corresponding coding schemes, these observations suggest that one may get such gains in performance even in the MD problem.

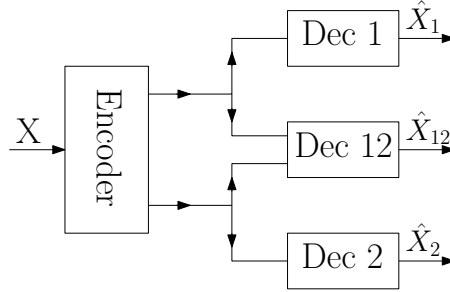


Figure 1: The Two-Descriptions Setup

In the second part of the paper we show that SSC coding scheme which is based on unstructured codes as mentioned above is not complete. We provide several specific examples of 3- and 4-description problems and example-specific coding schemes based on random linear codes that perform strictly better than the above SSC coding scheme. Subsequently, we supplement the above SSC scheme with new coding layers which have algebraic structure. We restrict our attention to the algebraic structure associated with finite fields. We present a unified coding scheme which works for arbitrary sources and distortion measures. We characterize the asymptotic performance of this coding scheme using computable single-letter information quantities. We interpret the SSC coding as capturing the common information components among  $2^l - 1$  decoders using univariate functions, and the algebraic coding supplement as capturing common information among  $2^l - 1$  decoders using bivariate and multivariate functions.

The rest of the paper is organized as follows. Section 2 explains the notation used in the paper. Section 3 provides an overview of the ideas developed in previous works and provides the groundwork for the next sections. In Section 4, we present a new unstructured coding strategy which improves upon the CMSB scheme. We show that there are two different types of gains compared to the previous scheme: the first is due to the addition of several common-component codebook layers, the second is due to a more generalized binning method. In Section 5, we identify examples where improvements due to structured coding materialize in the MD setup. In this section, we investigate three different examples. In two of the examples the achievable RD region is improved via using linear quantizers, and in the other example the gains are due to linear binning. In Section 6 we generalize the ideas in the previous section and provide an achievable RD region for the general  $l$ -descriptions problem. Since the characterization of RD region is involved and complicated we provide the final RD region through several steps, adding new coding layers in each step. Section 7 concludes the paper.

## 2 Definitions and Notation

In this section we introduce the notation used in the paper. We restrict ourselves to finite alphabet random variables. We denote random variables by capital letters such as  $X, U$  and their corresponding alphabets (finite) by sans-serif typeface  $\mathsf{X}, \mathsf{U}$ , respectively. Numbers are denoted by small letters such as  $l, k$ . Sets of numbers are also denoted by the sans-serif typeface such as  $\mathsf{M}, \mathsf{N}$ . Specifically, we denote the set of natural numbers by  $\mathbb{N}$ , and the field of size  $q$  by  $\mathbb{F}_q$ . The set of numbers  $\{1, 2, \dots, m\}$  is also denoted by  $[1, m]$ .  $\alpha_{\mathsf{M}}$  is used to express the vector  $(\alpha_1, \alpha_2, \dots, \alpha_m)$  where  $\mathsf{M} = \{1, 2, \dots, m\}$ . A collection whose elements are sets is called a family of sets and is denoted by the calligraphic typeface  $\mathcal{M}$ . For a given family of sets  $\mathcal{M}$  we define a set  $\widetilde{\mathcal{M}} = \bigcup_{\mathsf{M} \in \mathcal{M}} \mathsf{M}$  as the set of numbers which are the elements of the sets in  $\mathcal{M}$ . The family of sets containing all subsets of  $\mathsf{M}$  is denoted by  $2^{\mathsf{M}}$ . A collection whose elements are families of sets is denoted by

the bold typeface  $\mathbf{M}$ . The collection of families of sets  $\{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_m\}$  is also represented by  $\mathcal{A}_{\mathbf{M}}$ . Random variables are indexed by families of sets as in  $U_{\mathcal{M}}$ . For the purposes of brevity we will write  $U_{M_1, M_2, \dots, M_n}$  instead of  $U_{\mathcal{M}}$  where  $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$  wherever the notation doesn't cause ambiguity.  $U_{\mathcal{M}}^n$  denotes a vector of length  $n$  of random variables, each distributed according to the distribution  $P_{U_{\mathcal{M}}}$ . For  $\epsilon > 0$  and  $n \in \mathbb{N}$ , we denote the set of  $n$ -length vectors which are  $\epsilon$ -typical with respect to  $P_{U_{\mathcal{M}}}$  by  $A_{\epsilon}^n(U_{\mathcal{M}})$ . We use the definition of frequency typicality as given in [14] in this paper.

We denote a set of random variables as follows  $U_{\mathbf{M}} = \{U_{\mathcal{M}} | \mathcal{M} \in \mathbf{M}\}$ . For two collections of families  $\mathbf{M}_1$  and  $\mathbf{M}_2$ , we write  $[U, V]_{(\mathbf{M}_1, \mathbf{M}_2)}$  to denote the unordered collection of random variables  $\{U_{\mathbf{M}_1}, V_{\mathbf{M}_2}\}$ . Let  $\mathbf{N}_i \subset \mathbf{M}_i, i = 1, 2$ , and define  $\bar{\mathbf{N}} = (\mathbf{N}_1, \mathbf{N}_2)$ . We express this as  $\bar{\mathbf{N}} \subset (\mathbf{M}_1, \mathbf{M}_2)$ . Unions, intersections and complements are defined for  $(\mathbf{M}_1, \mathbf{M}_2)$  in the same manner. A family of sets is called a Sperner family of sets if none of its elements is a subset of another element. In other words a family of sets  $\mathcal{S}$  is a Sperner family if  $\nexists \mathbf{N}, \mathbf{N}' \in \mathcal{S}, \mathbf{N} \subsetneq \mathbf{N}'$ . For any given set  $\mathbf{M}$ , the three families  $\emptyset, \{\emptyset\}$  and  $\{\mathbf{M}\}$  are all Sperner families. For a set  $\mathbf{M}$ , we define the collection of families of sets  $\mathbf{S}_{\mathbf{M}}$  as the set of all Sperner families whose elements are subsets of  $\mathbf{M}$  except for the three trivial Sperner families mentioned above. So we have  $\mathbf{S}_{\mathbf{M}} = \{\mathcal{S} | \nexists \mathbf{N}, \mathbf{N}' \in \mathcal{S}, \mathbf{N} \subsetneq \mathbf{N}'\} \setminus \{\emptyset, \{\emptyset\}, \{\mathbf{M}\}\}$ .

For the general  $l$ -descriptions problem, we define the set  $\mathbf{L} \triangleq [1, l]$ , and this set represents the set of all descriptions. Each decoder receives a subset of these descriptions. Let  $l_i \in \mathbf{L}, i \in [1, n]$  for some  $n$ . We denote the decoder which receives descriptions  $l_1, l_2, \dots, l_n$  by the set  $\mathbf{N} = \{l_1, l_2, \dots, l_n\}$ . Define the family of sets  $\mathcal{L} \triangleq 2^{\mathbf{L}} - \{\emptyset\}$ . This family of sets corresponds to the set of all possible decoders. We further explain the notation through an example. Consider the three-descriptions problem. In this case we have  $l = 3$ , the set of descriptions are  $\mathbf{L} = \{1, 2, 3\}$ . There are seven possible decoders. The set of all decoders is  $\mathcal{L} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$ . Consider the two families of sets  $\mathcal{M}_1 = \{\{1, 2\}, \{1, 3\}\}$  and  $\mathcal{M}_2 = \{\{1\}, \{3\}, \{1, 2\}\}$ . In this case,  $\bar{\mathcal{M}}_i = \{1, 2, 3\}, i \in \{1, 2\}$ . Define the set  $\mathbf{M} = \{\mathcal{M}_1, \mathcal{M}_2\}$ . The set of random variables  $\{U_{\mathcal{M}_1}, U_{\mathcal{M}_2}\}$  is denoted by  $U_{\mathbf{M}} = U_{\mathcal{M}_1, \mathcal{M}_2}$ . Here  $\mathcal{M}_1$  is a Sperner family, but  $\mathcal{M}_2$  is not a Sperner family since  $\{1\}, \{1, 2\} \in \mathcal{M}_2$  and  $\{1\} \subsetneq \{1, 2\}$ , furthermore  $\mathcal{M}_1 \in \mathbf{S}_{\mathbf{L}}$  but  $\mathcal{M}_2 \notin \mathbf{S}_{\mathbf{L}}$ . The second part of the paper is involves application of linear codes and their cosets. The following gives a formal definition for such codes,

**Definition 1.** Let  $q$  be a prime number. A  $(k, n)$  linear code  $C$  is characterized by its generator matrix  $G_{k \times n}$  defined on  $\mathbb{F}_q$ .  $C$  is defined as follows:  $C \triangleq \{\mathbf{u}G | \mathbf{u} \in \mathbb{F}_q^k\}$ . A coset code  $C'$  is a shifted version of a linear code and is characterized by a generator matrix  $G_{k \times n}$  and a dither  $\mathbf{b}^n$  defined on  $\mathbb{F}_q$ .  $C'$  is defined as follows:  $C' \triangleq \{\mathbf{u}G + \mathbf{b} | \mathbf{u} \in \mathbb{F}_q^k\}$ .

We will make frequent use of nested linear codes. A pair of nested linear codes is defined as follows,

**Definition 2.** For natural numbers  $k_i < k_o < n$ , let  $G_{k_i \times n}$ , and  $\Delta G_{(k_o - k_i) \times n}$  be matrices on  $\mathbb{F}_q$ . Define  $C_i, C_o$  as the linear codes generated by  $G, [G | \Delta G]$ , respectively.  $(C_i, C_o)$  is called a pair of nested linear codes with the inner code  $C_i$  and the outer code  $C_o$ . Nested coset codes are defined as shifted versions of nested linear codes.

## 3 Preliminaries

### 3.1 Problem Statement

The general  $l$ -descriptions problem is described in this section. The setup is characterized by a discrete memoryless source with probability distribution  $P_X(x), x \in X$ , where  $X$  is a finite set, and the distortion functions  $d_N : X \times \hat{X}_N \rightarrow \mathbb{R}^+, N \in \mathcal{L}$ , where  $\hat{X}_N$  is the reconstruction alphabet. We assume that the

distortion functions are bounded, and that the distortion for the  $n$ -length sequence  $(x^n, \hat{x}^n)$  is given by the average distortion of the components  $(x_i, \hat{x}_i)$ . The discrete, memoryless source  $X$  is fed into an encoder. The encoder upon receiving a block of length  $n$  of source symbols produces  $l$  different indices called *descriptions* of the source. These descriptions are sent to the decoders. Each decoder receives a specific subset of the descriptions. Decoder  $N$ ,  $N \in \mathcal{L}$  receives description  $i$  for all  $i \in N$ . Based on the descriptions it has received, the decoder produces a reconstruction of the source vector.

**Definition 3.** An  $(n, \Theta_1, \Theta_2, \dots, \Theta_l)$  multiple-descriptions code consist of an encoder and  $|\mathcal{L}|$  decoders:

$$e_i : X^n \rightarrow [1, \Theta_i], i \in \mathcal{L},$$

$$f_N : \prod_{i \in N} [1, \Theta_i] \rightarrow \hat{X}_N^n, N \in \mathcal{L}.$$

The achievable rate-distortion (RD) region is defined as follows,

**Definition 4.** The RD vector  $(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}}$  is said to be achievable if for all  $\epsilon > 0$  and sufficiently large  $n$ , there exists an  $(n, \Theta_1, \Theta_2, \dots, \Theta_l)$  multiple-descriptions code such that the following constraints are satisfied:

1.  $\frac{\log \Theta_i}{n} \leq R_i + \epsilon, \forall i \in \mathcal{L}$ ,
2.  $E_{X^n} [d_N(f_N((e_i(X^n))_{i \in N}), X^n)] \leq D_N + \epsilon, \forall N \in \mathcal{L}$ .

The achievable RD region for the  $l$ -descriptions problem is the set of all achievable RD vectors.

**Remark 1.** Although the reconstruction alphabet can be different from the source alphabet, throughout this paper we assume that the two alphabets are the same for the ease of notation. The results hold for the general case.

## 3.2 Prior Works

In this section we present a brief description of some of the previous known schemes, and state the corresponding inner bounds developed for the achievable RD region. One of the early strategies for coding over two descriptions was the El Gamal - Cover (EGC) strategy [5]. Similar to all the other strategies explained in this section, the EGC scheme relies on random, unstructured codebook generation. The following theorem describes the corresponding inner bound to the achievable RD region which results from the EGC scheme. Note that this is an alternative way to characterize the inner bound described in [5].

**Definition 5.** For a joint distribution  $P$  on random variables  $(U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}, X, Q)$  and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_N : \mathcal{U}_N \rightarrow \mathcal{X}, N \in \mathcal{L}\}$ , the set  $\mathcal{RD}_{EGC}(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds:

$$R_1 \geq I(U_{\{1\}}; X|Q), \quad R_2 \geq I(U_{\{2\}}; X|Q), \quad (1)$$

$$R_1 + R_2 \geq I(U_{\{1\}}, U_{\{2\}}; X|Q) + I(U_{\{1\}}; U_{\{2\}}|Q) + I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}, Q), \quad (2)$$

$$D_N \geq E(d_N(g_N(U_N, Q), X)), N \in \mathcal{L}. \quad (3)$$

**Theorem 1 (EGC).** The RD vector  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$  is achievable for the two descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) \in \mathcal{RD}_{EGC}(P, g_{\mathcal{L}})$ .

In the EGC scheme, two codebooks  $C_{\{1\}}$  and  $C_{\{2\}}$  are generated independently based on the marginals  $P_{U_{\{1\}}|Q}$  and  $P_{U_{\{2\}}|Q}$ . The two codebooks should be large enough so that the encoder can find a pair of jointly typical codewords in the two codebooks. If the codebooks were generated jointly based on the joint distribution  $P_{U_{\{1\}}, U_{\{2\}}|Q}$ ,  $R_1 + R_2 \geq I(U_{\{1\}}, U_{\{2\}}; X|Q) + I(U_{\{1,2\}}; X|U_{\{1\}}, U_{\{2\}}, Q)$  would ensure the existence of such jointly typical codewords, however in the EGC scheme, since the codebooks are generated independently, a *rate-penalty* is inflicted on the encoder. The term  $I(U_{\{1\}}; U_{\{2\}}|Q)$  in (2) is a manifestation of this *rate-penalty*. Towards reducing the *rate-penalty* a new coding strategy was introduced. The resulting achievable RD region is called the Zhang-Berger [ZB] region. The region is given in the following theorem:

**Definition 6.** For a joint distribution  $P$  on random variables  $(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}, X)$  and set of reconstruction functions  $g_{\mathcal{L}} = \{g_{\mathbf{N}} : \mathbf{U}_{\mathbf{N}} \rightarrow X, \mathbf{N} \in \mathcal{L}\}$ , the set  $\mathcal{RD}_{\text{ZB}}(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds:

$$\begin{aligned} R_1 &\geq I(U_{\{1,2\}}, U_{\{1\}}; X), & R_2 &\geq I(U_{\{1,2\}}, U_{\{2\}}; X), \\ R_1 + R_2 &\geq I(U_{\{1,2\}}; X) + I(U_{\{1,2\}}, U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}; X) + I(U_{\{1\}}; U_{\{2\}}|U_{\{1,2\}}), \\ D_{\mathbf{N}} &\geq E(d_{\mathbf{N}}(g_{\mathbf{N}}(U_{\mathbf{N}}), X)), \mathbf{N} \in \mathcal{L}. \end{aligned}$$

**Theorem 2 (ZB).** The RD vector  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$  is achievable for the two descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) \in \mathcal{RD}_{\text{ZB}}(P, g_{\mathcal{L}})$ .

The closure of the union of all the achievable vectors is called the ZB rate-distortion region and is denoted by  $\mathcal{RD}_{\text{ZB}}$ :

$$\mathcal{RD}_{\text{ZB}} = \text{cl} \left( \bigcup_{P, g_{\mathcal{L}}} \mathcal{RD}_{\text{ZB}}(P, g_{\mathcal{L}}) \right).$$

The scheme differs from the EGC strategy in the introduction of the random variable  $U_{\{1,2\}}$ . The random variable  $U_{\{1,2\}}$  is called the *common-component* between the two descriptions. In the EGC scheme, in order to send  $U_{\{1\}} = (\tilde{U}_{\{1\}}, U_{\{1,2\}})$  and  $U_{\{2\}} = (\tilde{U}_{\{2\}}, U_{\{1,2\}})$ , one has to pay the following *rate-penalty*:

$$I(U_{\{1\}}; U_{\{2\}}) = H(U_{\{1,2\}}) + I(\tilde{U}_{\{1\}}; \tilde{U}_{\{2\}}|U_{\{1,2\}}).$$

But in the ZB scheme the *rate-penalty* is reduced to:

$$I(U_{\{1,2\}}; X) + I(U_{\{1\}}; U_{\{2\}}|U_{\{1,2\}}) = I(U_{\{1,2\}}; X) + I(\tilde{U}_{\{1\}}; \tilde{U}_{\{2\}}|U_{\{1,2\}}).$$

The following definition provides a characterization of the common-component between two random variables,

**Definition 7.** Let  $X_{\{1\}}$  and  $X_{\{2\}}$  be two random variables.  $W$  is called a *common-component* between  $X_{\{1\}}$  and  $X_{\{2\}}$ , if there exist functions  $h_i : X_{\{i\}} \rightarrow \mathcal{W}, i = 1, 2$  such that  $W = h_1(X_{\{1\}}) = h_2(X_{\{2\}})$  with probability one, and the entropy of  $W$  is positive.

It was shown in [6] that in a certain two-descriptions setup, the addition of  $U_{\{1,2\}}$  enlarges the RD region. We call such a random variable non-redundant. The following definition gives a formal description of a non-redundant random variable:

**Definition 8.** In a given achievable RD region for the  $l$ -descriptions setup, characterized by a collection of auxiliary random variables, an auxiliary random variable  $U$  is called *non-redundant* if the RD region strictly reduces when  $U$  is set as constant.

**Example 1.** We provide an overview of the example in [6] where the ZB rate-distortion region is strictly better than EGC rate-distortion region, since it is used extensively in the following sections. Consider the two-descriptions setting. Here  $X$  is a binary symmetric source (BSS), and the side decoders intend to reconstruct  $X$  with Hamming distortion. The central decoder needs a lossless reconstruction of the source. In [6], it is shown that the rate distortion vector  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) = (0.629, 0.629, 0.11, 0.11, 0)$  is achievable using the ZB scheme but not the EGC scheme.

Typically, in a given RD region, a codebook is associated with each random variable. We call a codebook non-redundant if it is associated with a non-redundant random variable. In ZB coding scheme, the codebook corresponding to  $U_{\{1,2\}}$  is non-redundant.

The idea of constructing a codebook carrying the common-component between the two random variables is the foundation of most of the schemes proposed for the general  $l$ -descriptions problem. One can even interpret the main difference between these schemes to be the way the common-component between different random variables are exploited.

As explained in the introduction, the best known achievable RD region for the  $l$ -descriptions problem is the CMS with binning (CMSB) strategy. In this strategy a combinatorial number of *common-component* random variables are considered. We explain the coding scheme for the three-descriptions case. The codebook structure is shown in Figure 2. There are two layers of codebooks, a layer of Maximum-Distance Separable (MDS) codes and a layer of Source Channel Erasure Codes (SCEC's). The codebook  $C_M$  is decoded at decoder  $N$  if  $\exists N' \in \mathcal{M}, N' \subset N$ . The codebooks are binned independently, and the bin numbers for the MDS code  $C_M$  are carried by description  $i$  if  $i \in \bigcup_{N \in \mathcal{M}} N$ . Whereas the bin number for each SCEC is carried by only one description  $i$  where  $i \in \bigcap_{N \in \mathcal{M}} N$ . Let  $\mathcal{RD}_{CMSB}$  denote the resulting RD region achievable using CMSB strategy (see [10, 11]).

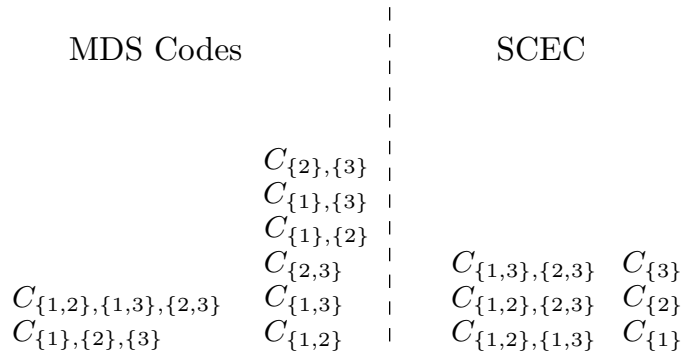


Figure 2: The structure of CMSB codebooks in the three-descriptions problem

## 4 Improvements Using Unstructured Codes

Our objective is to provide a new achievable RD region for the  $l$ -descriptions problem, which improves upon the RD region given by the CMSB strategy. This is based on a new coding scheme involving both unstructured and structured codes. The achievable RD region and the corresponding coding scheme is presented pedagogically in two steps. In the first step, presented in this section, we provide an RD region achievable using unstructured codes. This region is strictly better than the CMSB region. In other words this is an improvement upon the CMSB region using only unstructured codes. In the second step, presented

in the next two sections, this is enhanced with a structured coding layer which improves the performance even further. In other words we show that the codebooks associated with the structured coding layer are non-redundant.

## 4.1 Main Results

We describe the key ideas for the case  $l = 3$ . There are 7 distinct decoders, one associated with every non-empty subset of  $L = \{1, 2, 3\}$ . That is, we identify the set of decoders with  $\mathcal{L} = 2^L \setminus \emptyset$ . The new achievable RD region that we provide improves upon the CMSB rate-distortion region on two factors. The first comes by adding extra codebooks, and the second comes by a more general binning method. Using the common-component perspective, we associate with every non-empty subset  $\mathcal{M}$  of these 7 decoders an auxiliary random variable and a corresponding codebook. That is, we identify the collection of auxiliary variables (and their codebooks) with  $2^{\mathcal{L}} \setminus \emptyset$ . Each codebook is binned multiple times. If a description is received by at least one decoder in  $\mathcal{M}$ , then a bin index of the codebook associated with  $\mathcal{M}$  is sent on that description.

Although it appears that the strategy involves the generation of a doubly-exponential number of codebooks (in  $l$ ), we show that most of these codebooks are redundant, leaving only an asymptotically exponential number of non-redundant codebooks. While the remaining codebooks are generally non-redundant, only a small number of them are such in most of the examples we consider in this paper.

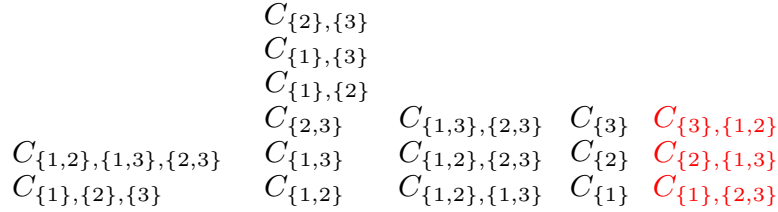


Figure 3: The SSC codebooks present in the three-descriptions problem

It turns out that a codebook is non-redundant if and only if it is associated with a family of sets in  $\mathbf{S}_L$ . So, instead of 63 codebooks, we have just 17. Since the indices of the codebooks are associated with the Sperner families of sets, we call the scheme the Sperner Set Coding (SSC) scheme. A schematic of the codebook collection is shown in Figure 3. We start from the left and from the top. The first two codebooks can be identified as (3, 2) and (3, 1) MDS codes. The next six codebooks can be identified as three (2, 1) MDS codes, and three (2, 2) MDS codes associated with decoders which get two descriptions. The next three can be identified as (3, 2) source-channel erasure codes (SCEC). The next three can be identified as (3, 1) SCEC's (similar to the codebooks used in the EGC rate region). All these 14 codebooks are considered in deriving the CMSB rate region. The final set of codebooks are new. They can be identified as three (2, 1) MDS codes associated with decoders that receive disjoint subsets of descriptions. The following theorem characterizes the achievable RD region for the SSC scheme:

**Definition 9.** For a joint distribution  $P$  on random variables  $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_L$  and  $X$  and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_{\mathbf{N}} : \mathbf{U}_{\mathbf{N}} \rightarrow \mathbf{X}, \mathbf{N} \in \mathcal{L}\}$ , the set  $\mathcal{RD}_{SSC}(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers  $(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i \in \widetilde{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_L}$ :

$$H(U_{\mathbf{M}}|X) \geq \sum_{\mathcal{M} \in \mathbf{M}} (H(U_{\mathcal{M}}) - r_{\mathcal{M}}), \forall \mathbf{M} \subset \mathbf{S}_L, \quad (4)$$



$$H(U_{\mathbf{M}_N} | U_{\mathbf{L} \cup \widetilde{\mathbf{M}}_N}) \leq \sum_{\mathcal{M} \in \mathbf{M}_N \setminus (\mathbf{L} \cup \widetilde{\mathbf{M}}_N)} (H(U_{\mathcal{M}}) + \sum_{i \in \widetilde{\mathcal{M}}} \rho_{\mathcal{M},i} - r_{\mathcal{M}}), \forall \mathbf{L} \subset \mathbf{M}_N, \forall \mathbf{N} \in \mathcal{L}, \quad (5)$$

$$r_{\mathcal{M}} \leq H(U_{\mathcal{M}}), \forall \mathcal{M} \in \mathbf{S}_{\mathbf{L}},$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathbf{N}} = E\{d_{\mathbf{N}}(g_{\mathbf{N}}(U_{\mathbf{N}}), X)\}, \quad (6)$$

where  $\mathbf{M}_N$  is the set of all codebooks decoded at decoder  $\mathbf{N}$ , that is  $\mathbf{M}_N \triangleq \{\mathcal{M} \in \mathbf{S}_{\mathbf{L}} | \exists \mathbf{N}' \subset \mathbf{N}, \mathbf{N}' \in \mathcal{M}\}$ , and  $\widetilde{\mathbf{M}}_N$  denotes the set of all codebooks decoded at decoders  $\mathbf{N}_p \subsetneq \mathbf{N}$  which receive subsets of descriptions received by  $\mathbf{N}$ , that is  $\widetilde{\mathbf{M}}_N \triangleq \bigcup_{\mathbf{N}_p \subsetneq \mathbf{N}} \mathbf{M}_{\mathbf{N}_p}$ .

**Theorem 3.** *The RD vector  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}}$  is achievable for the  $l$ -descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}} \in \mathcal{RD}_{SSC}(P, g_{\mathcal{L}})$ .*

The closure of the union of all such achievable vectors is called the SSC achievable rate-distortion region and is denoted by  $\mathcal{RD}_{SSC}$ ,

$$\mathcal{RD}_{SSC} = cl \left( \bigcup_{P, g_{\mathcal{L}}} \mathcal{RD}_{SSC}(P, g_{\mathcal{L}}) \right).$$

In order to clarify the notation we explain the random variables decoded at each decoder in the three-descriptions problem. When  $l = 3$ , we know  $\mathbf{S}_{\mathbf{L}}$  has 17 elements. In the formulas,  $\mathbf{M}_N$  corresponds to the set of random variables decoded at decoder  $\mathbf{N}$ , whereas  $\widetilde{\mathbf{M}}_N$  corresponds to the set of random variables which are decodable if we have access to strict subsets of the descriptions received by  $\mathbf{N}$ . Here are the random variables decoded at decoders  $\{1\}$  and  $\{2, 3\}$ :

$$\begin{aligned} \text{decoder } \{1\}: & U_{\{1,\{2\},\{3\}\}}, U_{\{1,\{2\}\}}, U_{\{1,\{3\}\}}, U_{\{1,\{2,3\}\}}, U_{\{1\}} \\ \text{decoder } \{2, 3\}: & U_{\{1,\{2\},\{3\}\}}, U_{\{1,2\},\{1,3\},\{2,3\}\}}, U_{\{1,\{2\}\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, \\ & U_{\{1,\{2,3\}\}}, U_{\{2,\{1,3\}\}}, U_{\{3,\{1,2\}\}}, U_{\{1,2\},\{2,3\}\}}, U_{\{1,3\},\{2,3\}\}}, U_{\{2\}}, U_{\{3\}}, U_{\{2,3\}} \end{aligned}$$

So as an example  $\mathbf{M}_{\{1\}} = \left\{ \left\{ \{1\}, \{2\}, \{3\} \right\}, \left\{ \{1\}, \{2\} \right\}, \left\{ \{1\}, \{3\} \right\}, \left\{ \{1\}, \{2, 3\} \right\}, \left\{ \{1\} \right\} \right\}$  which are all the codebooks decoded at decoder  $\{1\}$ . Also  $\widetilde{\mathbf{M}}_{\{2,3\}} = \left\{ \left\{ \{1\}, \{2\}, \{3\} \right\}, \left\{ \{1\}, \{2\} \right\}, \left\{ \{1\}, \{3\} \right\}, \left\{ \{2\}, \{3\} \right\}, \left\{ \{2\}, \{1, 3\} \right\}, \left\{ \{3\}, \{1, 2\} \right\}, \left\{ \{2\} \right\}, \left\{ \{3\} \right\} \right\}$ , and these are all the codebooks which are decoded at decoders  $\{2\}$  and  $\{3\}$ .

**Lemma 1.** *The SSC rate-distortion region is convex.*

*Proof.* See Section A.1 in the appendix. □

**Remark 2.** *For every decoder  $\mathbf{N} \in \mathcal{L}$ , we have defined the reconstruction as a function of the random variable  $U_{\mathbf{N}}$ . However, decoder  $\mathbf{N}$  decodes all random variables  $U_{\mathcal{M}}$  where  $\mathcal{M} \in \mathbf{M}_N$ . The following lemma shows that the RD region does not improve if the reconstruction function is defined as a function of  $U_{\mathbf{M}_N}$  instead.*

**Lemma 2.** *The RD region in Theorem 3 does not change if the reconstruction function at decoder  $\mathbf{N}$  is defined as a function of  $U_{\mathbf{M}_N}$ .*

*Proof.* See Section A.2 in the appendix. □

**Remark 3.** *In the scheme proposed in Theorem 3 there are  $|\mathbf{S}_L|$  codebooks. We know that the size of  $\mathbf{S}_L$  is the number of Sperner families on  $L$  minus three. The number of Sperner families is called the Dedekind numbers [24]. There has been a large body of work in determining the values of Dedekind numbers for different  $l$ . It is known that these numbers grow exponentially in  $l$ . As an example the number of codebooks necessary for  $l = 2, 3$  and  $4$  are  $3, 17$  and  $165$ . However in all of the examples in this paper it turns out that many of the codebooks become redundant and only a small subset are used in the scheme.*

*Proof.* Before proceeding to a more detailed description of the coding strategy we provide a brief outline. For each family of sets  $\mathcal{M} \in \mathbf{S}_L$  the encoder generates a codebook  $C_{\mathcal{M}}$  based on the marginal  $P_{U_{\mathcal{M}}}$  independently of the other codebooks. Intuitively, this codebook is the common-component among all the decoders  $N$  such that  $N \in \mathcal{M}$ , and it is decoded in all decoders  $N' \supset N$ . Codebook  $C_{\mathcal{M}}$  is binned independently and uniformly for each description  $i$  if  $i \in \widetilde{\mathcal{M}}$ . The description will carry the corresponding bin number for the codewords in each of the corresponding codebooks. Each decoder reconstructs its corresponding codewords by finding a unique set of jointly typical codevectors in the bins it has received. The existence of the jointly typical set of codewords is ensured at the encoder by the way of satisfaction of (4), whereas at the decoder unique reconstruction is warranted by (5).

**Codebook Generation:** Fix blocklength  $n$  and positive reals  $(\rho_{\mathcal{M},i}, r_{\mathcal{M}})_{i \in \widetilde{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_L}$ . For every  $\mathcal{M} \in \mathbf{S}_L$ , generate a codebook  $C_{\mathcal{M}}$  based on the marginal  $P_{U_{\mathcal{M}}}$  with size  $2^{nr_{\mathcal{M}}}$ . For the  $i$ th description, if  $i \in \widetilde{\mathcal{M}}$ , bin the codebook  $C_{\mathcal{M}}$  randomly and uniformly into  $2^{n\rho_{\mathcal{M},i}}$  bins (i.e. randomly and uniformly assign an index  $[1, 2^{n\rho_{\mathcal{M},i}}]$  to each codeword in  $C_{\mathcal{M}}$ , and the index is called the *bin-index*).

**Encoding:** Upon receiving the source vector  $X^n$ , the encoder finds a jointly-typical set of codewords  $u_{\mathcal{M}}^n, \mathcal{M} \in \mathbf{S}_L$ . Each description carries the *bin-indices* of all the codewords corresponding to its own binning function.

**Decoding:** Having received the bin-indices from descriptions  $i \in N$ , decoder  $N$  tries to reconstruct  $C_{\mathcal{M}}$  if  $\mathcal{M} \in \mathbf{M}_N$ . In other words the decoder finds a unique vector  $(u_{\mathcal{M}}^n)_{N \in \mathcal{M}}$  of jointly typical sequences in the corresponding bins. If the vector does not exist or is not unique, the decoder declares error.

**Covering Bounds:** Since codebooks are generated randomly and independently, to find a set of vectors  $U_{\mathcal{M}}^n$  that is jointly typical with the source vector  $X^n$ , the mutual covering bounds (4) are necessary based on the mutual covering lemma [21].

**Packing Bounds:** For decoder  $N$ , description  $i$  is received if  $i \in N$ . Since binning is done independently and uniformly, to find a unique set of jointly typical sequences  $(u_{\mathcal{M}}^n)_{N \in \mathcal{M}}$ , the mutual packing bounds (5) are required by the mutual packing lemma [21]. □

**Remark 4.** *There are two main differences between the new scheme and the previous CMSB scheme. First there are additional codebooks present. As an example in Figure 3, the three codebooks in the right column are not present in the CMSB scheme. Second, description  $i$  bins all of the codebooks  $\mathcal{M}$  such that  $i \in \widetilde{\mathcal{M}}$ . We will show in the next sections that these additional codebooks contribute to an enlargement of the achievable RD region. In other words we prove that all of the additional codebooks are non-redundant. Also we show that the new binning strategy improves the achievable RD region.*

## 4.2 Improvements Due to additional codebooks

Consider the general  $l$ -descriptions problem. In this section we prove that a codebook  $C_M$  is non-redundant if  $M \in \mathbf{S}_L$ .

**Remark 5.** *It is straightforward to see that addition of a codebook  $C_M$  where  $M \notin \mathbf{S}_L$  is not going to result in a larger achievable RD region. To see this consider the three descriptions problem and assume we add the codebook  $C_{\{1\},\{1,2\}}$ . By our definition this new codebook is decoded if we either receive description 1 or both descriptions 1 and 2. In this case the codebook is decoded in exactly those decoders where  $C_{\{1\}}$  is decoded. This means that merging these two codebooks does not change the packing bounds whereas it may relax the covering bounds. So such a codebook would be redundant. This is the reason why we consider only those codebooks which are associated with Sperner families.*

**Remark 6.** *There are three Sperner families for which we do not construct codebooks:  $\{\phi, \{\phi\}, \{L\}\}$ . It is clear that  $U_\phi$  and  $U_{\{\phi\}}$  are not necessary since they are not decoded at any decoder. Furthermore one can use the proof provided in [25] to show that  $U_L$  is also redundant.*

The next lemma proves that the random variables considered in Theorem 3 are non-redundant.

**Lemma 3.** *The random variable  $U_M$  is non-redundant for every  $M \in \mathbf{S}_L$ .*

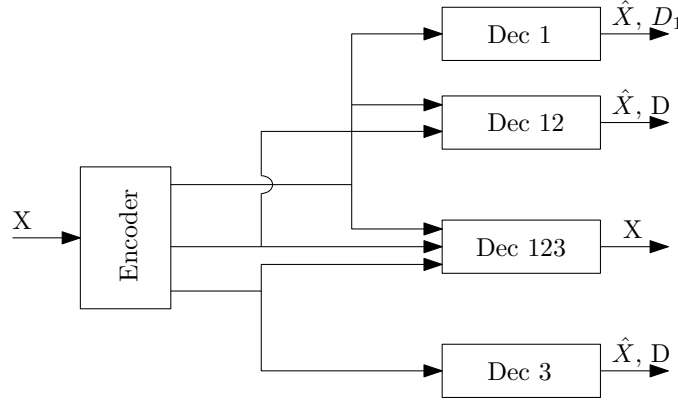


Figure 4: Three Descriptions Setup Showing  $C_{\{1,2\},\{3\}}$  is not redundant.

*Proof.* We provide the proof for the  $l = 3$  case and give an outline of how the proof is generalized for  $l > 3$ . The codebooks  $C_{\{1\}}, C_{\{2\}}, C_{\{3\}}, C_{\{1,2\}}, C_{\{1,3\}}, C_{\{2,3\}}, C_{\{1\},\{2\}}, C_{\{1\},\{3\}}, C_{\{2\},\{3\}}, C_{\{1,2\},\{1,3\}}, C_{\{1,2\},\{2,3\}}, C_{\{1,3\},\{2,3\}}, C_{\{1\},\{2\},\{3\}}, C_{\{1,2\},\{1,3\},\{2,3\}}$  are all present in the CMSB scheme and it was shown that they are non-redundant. The new codebooks are  $C_{\{1,2\},\{3\}}, C_{\{1,3\},\{2\}}$  and  $C_{\{23\},\{1\}}$ . We prove that  $C_{\{1,2\},\{3\}}$  is non-redundant using the following example, the two other codebooks are non-redundant by symmetry.

We build on Example 1 to construct a three-descriptions example as shown in Figure 4. As explained in the previous section, it is known that  $U_{\{1\},\{2\}}$  is non-redundant. Let  $R_i = 0.629, i \in \{1, 2\}$ , and  $D_{\{1,2\}} = 0$ . Let

$$D^* = \min_D \{D | (0.629, 0.629, D, D, 0) \in \mathcal{RD}_{ZB}\}. \quad (7)$$

Let  $\mathbf{P}$  be the set of probability distributions  $P_{U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, X}$ , such that  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}}) = (0.629, 0.629, D^*, D^*, 0)$  belongs to  $\mathcal{RD}_{ZB}(P_{U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, X}, g_{\mathcal{L}})$  for some  $g_{\mathcal{L}}$  as given in Theorem 2. Define the joint distribution  $P_{U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, X}^*$  as follows:

$$P_{U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, X}^* \triangleq \operatorname{arginf}_{P_{U_{\{1\},\{2\}}, U_{\{1\}}, U_{\{2\}}, X} \in \mathbf{P}} I(U_{\{1\},\{2\}}; X).$$

Let  $P_{U_{\{1,2\}}, X}^*$  be the marginal distribution of  $U_{\{1,2\}}$  and  $X$ . Define a random variable  $W$  that is correlated with  $X$  such that  $P_{W,X} = P_{U_{\{1,2\}}, X}^*$ . Let  $N_\delta$  be a binary random variable independent of  $X$  and  $W$  with  $P(N_\delta = 1) = \delta, \delta \in (0, 0.5)$ . Define  $\hat{W} = U_{\{1,2\}} \wedge N_\delta$  where  $\wedge$  denotes the logical AND function. Let  $P_{\hat{W}, X}, P_{X|\hat{W}}$  be the induced joint and conditional distributions, respectively.

**Example 2.** We proceed by explaining the new example. The source  $X$  is a BSS, decoders  $\{1, 2\}$  and  $\{3\}$  want to reconstruct the source with respect to Hamming distortion and the central decoder wants to reconstruct the source losslessly. Decoder  $\{1\}$  wants to reconstruct the source with respect to the distortion function given by:

$$d_{\{1\}}(x, \hat{x}) = -\log(P_{X|\hat{W}}(x|\hat{x}))$$

**Lemma 4.** *The following RD vector does not belong to  $\mathcal{RD}_{CMSB}$ , where  $U_{\{1,2\},\{3\}}$  is constant. The vector belongs to  $\mathcal{RD}_{SSC}$  given in Theorem 3 which is achievable using the SSC scheme:*

$$(R_1, R_2, R_3, D_{\{1\}}, D_{\{1,2\}}, D_{\{3\}}, D_{\{1,2,3\}}) = (I(X; \hat{W}), 0.629 - I(X; \hat{W}), 0.629, D', D^*, D^*, 0),$$

where  $D' = E(d_{\{1\}}(X, \hat{W}))$ .

*Proof.* We provide the intuition behind the proof first. In the coding scheme in Theorem 3, the only codebooks capable of carrying the common-component between decoders  $\{1, 2\}$  and  $\{3\}$  are  $C_{\{1,3\}}, C_{\{2,3\}}, C_{\{1,2,3\}}$  and  $C_{\{1,2\},\{3\}}$ . We have set the distortion constraint at decoder  $\{1\}$  such that this common message can't be carried exclusively on either of the descriptions 1 and 2, but rather both descriptions are necessary for the reconstruction of the common codebook. So the codebook  $C_{\{1,2\},\{3\}}$  can't be empty. The proof is provided in Section A.3 in the appendix.  $\square$

So far we have shown that the additional codebooks are non-redundant when  $l = 3$ . The argument can be extended to the case when  $l > 3$ , an outline of the general argument is provided in appendix A.4.  $\square$

### 4.3 Improvements Due to Binning

The second factor contributing to the gains in the SSC rate-distortion region is the binning method. In the SSC scheme all descriptions  $i \in \widetilde{\mathcal{M}}$  carry independent bin indices of codebook  $C_{\mathcal{M}}$ . This is different from the CMSB strategy where each codebook is binned by a specific subset of the descriptions based on whether the codebook is a SCEC or an MDS codebook. We prove through a three-descriptions example that the RD region enlarges due to binning in the SSC scheme, even with the three additional codebooks. We show in the following example that the bin indices of  $C_{\{1,2\},\{1,3\}}$  should be carried by all descriptions.

**Example 3.** The example is generated by modifying Example 2 and is illustrated in Figure 5. The source  $X$  is BSS.  $d_{\{1\}}(X, \hat{W})$  is defined as in Example 2. Decoders  $\{1, 2\}$  and  $\{1, 3\}$  want to reconstruct the source with Hamming distortion and decoder  $\{1, 2, 3\}$  wants to reconstruct the source losslessly.

**Lemma 5.** *In order to achieve  $(R_1, R_2, R_3, D_{\{1\}}, D_{\{1,2\}}, D_{\{1,3\}}, D_{\{1,2,3\}}) = (I(\hat{W}; X), R - I(\hat{W}; X), R - I(\hat{W}; X), D', D, D, 0)$  we must have  $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} > 0$ .*

*Proof.* See Section A.5 in the appendix.  $\square$

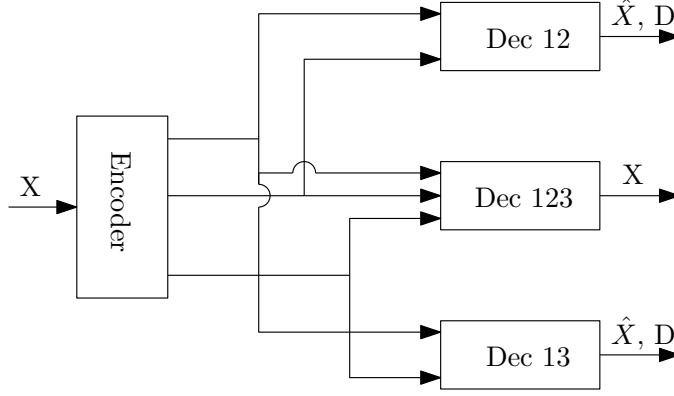


Figure 5: Example Showing Improvements Due to Binning

## 5 Linear Coding Examples

Before providing a unified RD region which uses both unstructured and structured codes (step 2), in this section, for pedagogical reasons, we look at three examples of  $l$ -descriptions problems and provide example-specific coding schemes based on linear codes that perform strictly better than the SSC scheme which is based on unstructured codes. This shows that the SSC region is not complete and a structured coding layer is necessary. These coding schemes are unified and presented in the next section.

### 5.1 Gains Due to Linear Quantizers

We create a three-descriptions setting where reconstructions of bivariate functions are necessary.

**Example 4.** Consider the three-descriptions example in Figure 6. Here  $X$  and  $Z$  are independent BSS. Decoder  $\{1\}$ ,  $\{2\}$  and  $\{3\}$  wish to reconstruct  $X$ ,  $Z$  and  $X+Z$ , respectively, with Hamming distortion. Decoders  $\{1, 2\}$ ,  $\{1, 3\}$ , and  $\{2, 3\}$  wish to reconstruct the pair  $(X, Z)$  with distortion function

$$d_{XZ}((\hat{X}, \hat{Z}), (X, Z)) = d_H(\hat{X}, X) + d_H(\hat{Z}, Z).$$

We are interested in achieving the following RD vector:

$$R_i = 1 - h_b(\delta), i \in \{1, 2, 3\}, D_{\{1\}} = D_{\{2\}} = \delta, D_{\{3\}} = \delta * \delta, D_{\{1,2\}} = D_{\{1,3\}} = D_{\{2,3\}} = 2\delta. \quad (8)$$

First we argue that in this example, description 3 should carry a bivariate function of descriptions 1 and 2. Decoders  $\{1\}$  and  $\{2\}$  operate at the optimal PtP rate-distortion function. So the corresponding descriptions have to allocate all of their rates to satisfy their individual decoder's distortion criteria. Since the distortion constraint at decoder  $\{1\}$  only relates to  $X$ , this description only carries a quantization of  $X$ , and by the same argument description 2 carries a quantization of  $Z$ . Then description 3 has to carry the sum of these two quantizations so that the joint decoders' distortion constraints are all satisfied. Since structured codes are efficient for transmitting bivariate summations of random variables, we expect that using structured codes would give gains in this example as opposed to unstructured codes. First, we prove that the RD vector is achievable using linear codes.

**Lemma 6.** *The RD vector in (8) is achievable.*

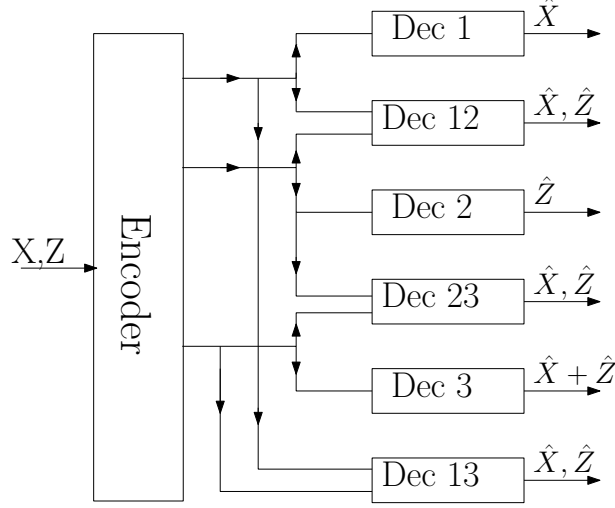


Figure 6: Three-Descriptions Example with a Vector Binary Source

*Proof. Encoding:* Construct a sequence of random linear codes  $C^n$  of rate  $1 - h_b(\delta) + \epsilon_n$ , where  $\epsilon_n$  is going to 0. It is well known that such a sequence of linear codes can be used to quantize a BSS to Hamming distortion  $\delta$ . Define the following:

$$\hat{X}^n = \operatorname{argmin}_{c^n \in C^n} d_H(x^n, c^n)$$

$$\hat{Z}^n = \operatorname{argmin}_{c^n \in C^n} d_H(z^n, c^n)$$

Since  $\hat{X}^n$  and  $\hat{Z}^n$  are codewords and the codebook is linear,  $\hat{X}^n + \hat{Z}^n$  is also a codeword. Description 1 carries the index of  $\hat{X}^n$ , description 2 carries the index of  $\hat{Z}^n$  and description 3 carries the index of  $\hat{X}^n + \hat{Z}^n$ .

**Decoding:** Decoders {1} and {2}, receive  $\hat{X}^n$  and  $\hat{Z}^n$ , respectively, so they satisfy their distortion constraints. Decoder {3} reconstructs  $\hat{X}^n + \hat{Z}^n$ . Lemma 7 shows that the distortion criteria at this decoder is satisfied.

**Lemma 7.** *In the above setting, we have  $\frac{1}{n} E(d_H(\hat{X}^n + \hat{Z}^n, X^n + Z^n)) \rightarrow \delta * \delta$ .*

*Proof.* See Section B.1 in the appendix. □

Decoder {1, 2} receives  $\hat{X}^n$  and  $\hat{Z}^n$ , so it satisfies its distortion requirements. Also decoders {1, 3} and {2, 3} can recover  $\hat{X}^n$  and  $\hat{Z}^n$  by adding  $\hat{X}^n + \hat{Z}^n$  to  $\hat{X}^n$  and  $\hat{Z}^n$ , respectively. This shows that the RD vector in (8) is achievable using linear codes. □

Next we show that the SSC scheme cannot achieve this RD vector.

**Lemma 8.** *The RD vector in (8) does not belong to  $\mathcal{RD}_{SSC}$ , i.e., it is not achievable using the SSC scheme.*

*Proof.* See Section B.2 in the appendix. □

## 5.2 Gains Due to Linear Binning

In the SSC scheme, there are two stages in the codebook generation phase. In the first stage unstructured codebooks are generated randomly and independently, and in the second stage these codebooks are binned randomly in an unstructured fashion for each description. In the previous example it was shown that in the first stage, it is beneficial to generate codebooks with a linear structure. However in that example there was no need for binning. In the next example, we show that the binning operation needs to be carried out in a structured manner as well. This is analogous to the gains observed in the distributed source coding problem [15] where the bin structure needs to be linear. Consider the four-descriptions example in Figure 7.

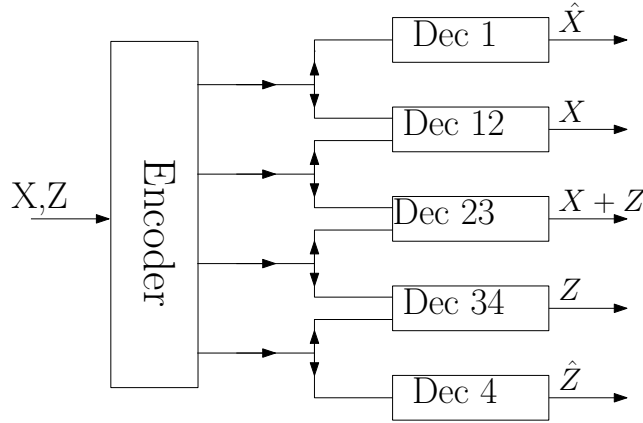


Figure 7: An Example Showing the Gains Due to Linear Binning

**Example 5.**  $X$  and  $Z$  are BSS's.  $X$  and  $Z$  are not independent, and they are related to each other through a binary symmetric channel with bias  $p \in (0, \frac{1}{2})$ . In other words  $X = Z + N_p$  where  $N_p \sim Be(p)$  is independent of  $X$  and  $Z$ . Decoders {1} and {4} wish to decode  $X$  and  $Z$ , respectively, with Hamming distortion. Decoders {1, 2}, {3, 4} and {2, 3} require a lossless reconstruction of  $X$ ,  $Z$  and  $X + Z$ , respectively. We are interested in achieving the following RD vector:

$$R_1 = R_4 = 1 - h_b(p), R_2 = R_3 = h_b(p), D_{\{1\}} = D_{\{4\}} = p \quad (9)$$

We show that the RD vector in (9) is achievable using structured codebooks and linear binning in the next lemma.

**Lemma 9.** *The RD vector in (9) is achievable.*

*Proof. Codebook Generation:* Take an arbitrary sequence of positive numbers  $\epsilon_n$ , where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . For any large  $n \in \mathbb{N}$ , fix  $r_{i,n} = 1 - h_b(p) - \epsilon_n$  and  $r_{o,n} = 1 - h_b(p) + \epsilon_n$ . Construct a family of nested coset codes  $(C_i^n, C_o^n)$  where  $C_i^n \subset C_o^n$  such that the rate of the outer code is  $r_{o,n}$  and the rate of the inner code is  $r_{i,n}$ . Choose  $C_i^n$  such that it is a good channel code for a BSC( $p$ ), and choose  $C_o^n$  such that it is a good source code for quantizing a BSS to Hamming distortion  $p$ . The existence of such nested coset codes is well-known from random coding arguments [27]. Next we bin the space  $\mathbb{F}_2^n$  into shifted versions (cosets) of  $C_i^n$ . Let  $\mathcal{P}_i$  be the Voronoi region of the codeword  $0^n$  in  $C_i^n$ . Any vector  $x^n \in \mathbb{F}_2^n$  can be written in the form  $x^n = v^n + c_i^n, v^n \in \mathcal{P}_i, c_i^n \in C_i^n$ . Define the  $i$ th bin as  $v^n + C_i^n$ . This operation bins the space into  $|\mathcal{P}_i| = 2^{n(h_b(p) + \epsilon_n)}$  bins. The bin number associated with an arbitrary vector  $x^n$  determines exactly the quantization noise resulting from quantizing the vector using  $C_i^n$  with the minimum Hamming distortion criterion. We denote the bin number of  $x^n$  as  $B_i(x^n)$ . A similar binning operation can be performed using  $C_o^n$ . Denote the bin number of  $x^n$  obtained using shifted versions of  $C_o^n$  by  $B_o(x^n)$ .

**Encoding:** The encoder quantizes  $x^n$  and  $z^n$  using  $C_o^n$  to  $Q_o(x^n)$ , and  $Q_o(z^n)$ , respectively. It also finds the bin number of the two source sequences  $B_i(x^n)$  and  $B_i(z^n)$ .  $Q_o(x^n)$  is transmitted on the first description,  $B_i(x^n)$  is transmitted on the second description,  $B_i(z^n)$  is transmitted on the third description, and  $Q_o(z^n)$  is transmitted on the fourth description.

**Decoding:** Since the outer codes are good source codes, the distortion constraints at decoders {1} and {4} are satisfied.

We argue that the Voronoi region of  $0^n$  in  $C_o^n$  is a subset of the one for  $C_i^n$ . This is true since  $C_i^n \subset C_o^n$ . Hence, having  $B_i(x^n)$ , decoders {1, 2} and {3, 4} can calculate  $B_o(x^n)$ . As mentioned above the bin number determines the quantization noise, so the decoders can reconstruct the source losslessly using the bin number and the quantization vector. Decoder {2, 3} receives  $B_i(x^n)$  and  $B_i(z^n)$ . We have  $x^n = Q_i(x^n) + B_i(x^n)$  and  $z^n = Q_i(z^n) + B_i(z^n)$ , so  $B_i(x^n) + B_i(z^n) = x^n + z^n + Q_i(x^n) + Q_i(z^n)$ . Since  $C_i^n$  is linear,  $Q_i(x^n) + Q_i(z^n)$  is a codeword, and  $x^n + z^n$  can be thought of as the noise vector for a BSC( $p$ ). We constructed  $C_i^n$  such that it is a good channel code for BSC( $p$ ), so the decoder can recover  $Q_i(x^n) + Q_i(z^n)$  from  $x^n + z^n + Q_i(x^n) + Q_i(z^n)$ . Then by subtracting the two vectors it can get  $x^n + z^n$ .  $\square$

Although we have used linear codes for quantization as well as binning, the linearity of the binning codebook  $C_i^n$  is critical in this example. In fact, it can be similarly shown that one can achieve the RD vector in (9) with  $C_o^n$  chosen to be a union of random cosets of  $C_i^n$ . This is in contrast with the previous example where the quantizing codebook was required to be linear.

**Lemma 10.** *The RD vector in (9) is not achievable using the SSC scheme.*

*Proof.* See Section B.3 in the appendix.  $\square$

### 5.3 Correlated Quantizations of a Source

It can be noted that in the case of SSC scheme, the unstructured quantizers are generated randomly and independently. As observed in these two examples, in order to efficiently reconstruct the bivariate summation, it is beneficial to use the same linear code for quantizing the source. However, in the two examples the source was a vector with two components which were separately quantized using identical linear codes, and the analysis of the coding scheme required only standard PtP covering and packing bounds for linear codes. In the more general case, evaluation of the performance of identical, and more generally, correlated linear codes for MD quantization, requires new covering and packing bounds. This is illustrated through the following scalar source example which is depicted in Figure 8. The setup is constructed based on the no-excess rate example described in [6] for the two-descriptions problem. In the two-descriptions example, the source  $X$  is BSS, and the distortion functions at all decoders is Hamming distortion. For the special case, called no-excess rate regime, when  $R_1 = R_2 = \frac{1-h(D_0)}{2}$ , it was shown that the EGC region is tight. Here  $D_0$  is the distortion  $D_{\{1,2\}}$  at decoder {1, 2}, and the minimum side distortion  $D_{\{1\}} = D_{\{2\}}$  achievable was shown to be  $\frac{1}{2}(1 - (1 - 2D_0)(2 - \sqrt{2}))$ . The three-descriptions example is given as follows.

**Example 6.** The source  $X$  is BSS, the distortion functions at decoders {1}, {2}, {1, 2}, {1, 3} and {2, 3} are Hamming distortions, and the distortion function at decoder {3} is the following general distortion function,

$$d_{\{3\}}(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ \alpha & \text{if } x = 0, \hat{x} = 1 \\ \beta & \text{if } x = 1, \hat{x} = 0 \end{cases}$$



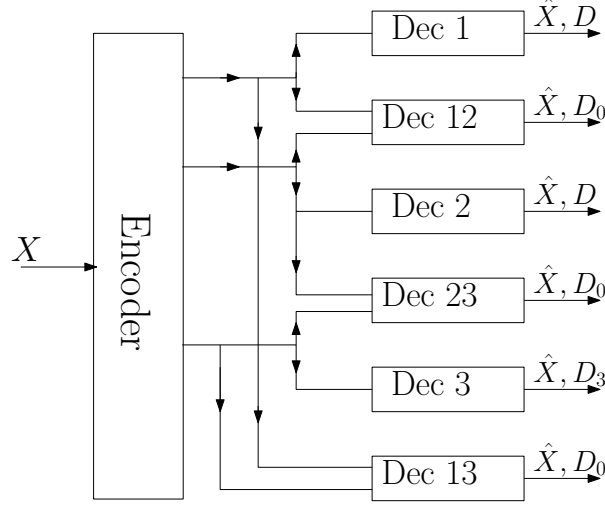


Figure 8: Scalar Source Example with Correlated Quantization

where  $\alpha$  and  $\beta$  are positive real numbers. We are interested in achieving the RD vectors with the following projections:

$$\begin{aligned} R_1 = R_2 &= \frac{1 - h_b(D_0)}{2}, D_{\{1\}} = D_{\{2\}} = \frac{1}{2}(1 - (1 - 2D_0)(2 - \sqrt{2})), \\ D_{\{1,2\}} = D_{\{1,3\}} = D_{\{2,3\}} &= D_0, \end{aligned} \quad (10)$$

Our objective is to evaluate the optimal  $(R_3, D_{\{3\}})$  trade-off. The following lemma provides the RD vectors achievable using linear codes.

**Lemma 11.** *The RD vector in (10) is achievable using linear codes, as long as the following constraints are satisfied:*

$$R_3 \geq \frac{1}{2} + h_b(\sqrt{2} - 1) - h_b\left(\frac{\sqrt{2}}{2}\right) - \frac{h_b(D_0)}{2} \quad (11)$$

$$D_{\{3\}} \geq \alpha(\sqrt{2} - 1)D_0 + \beta\left(\left(\frac{3 - 2\sqrt{2}}{2}\right)(1 - D_0) + \frac{D_0}{2}\right) \quad (12)$$

$$h_b(D_0) + 2h_b\left(\frac{\sqrt{2}}{2}\right) + h_b(2(\sqrt{2} - 1)D_0) + h_b(2(\sqrt{2} - 1)(1 - D_0)) \geq 1. \quad (13)$$

*Proof.* Consider the following definition.

**Definition 10.** *Let  $\mathbb{F}_q$  be a field. Consider 3 random variables  $X$ ,  $U$  and  $V$ , where  $X$  is defined on an arbitrary finite set  $\mathcal{X}$ , and  $U$  and  $V$  are defined on  $\mathbb{F}_q$ . Fix a PMF  $P_{X,U,V}$  on  $\mathcal{X} \times \mathbb{F}_q \times \mathbb{F}_q$ . A sequence of code pairs  $(C_1, C_2)$ , where  $C_j \subset \mathbb{F}_q^n$  for  $j = 1, 2$ , is called  $P_{XUV}$ -covering if  $\forall \epsilon > 0$ ,*

$$P(\{x^n | \exists (u^n, v^n) \in A_\epsilon^n(U, V | x^n) \cap C_1 \times C_2\}) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

First, we derive new covering and packing bounds for joint quantization of a general source  $X$  (i.e. not necessarily binary), using two pairs of nested coset codes. Let  $(C_i, C_o)$  and  $(C_i, C'_o)$  be two pairs of nested coset codes with generator matrices  $G_1$  and  $G_2$  shown in Figure 9 which share the inner code  $C_i$ . If  $r_i = 0$ , the two codebooks are generated independently. On the other hand, if  $r_o = r'_o = r_i$ , the two codebooks are the same, so this construction generalizes the previous constructions.

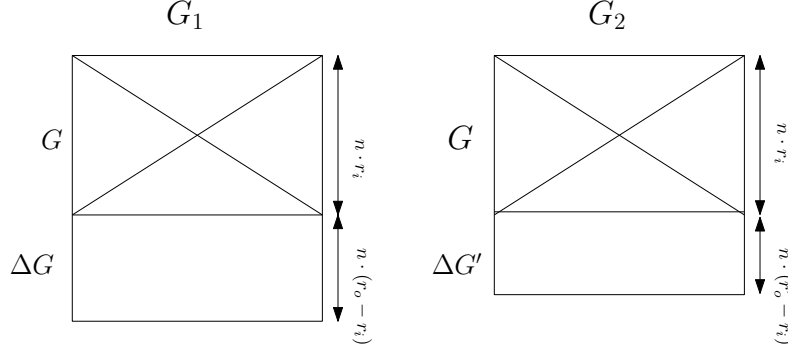


Figure 9: Codebook Construction for Lemma 12

**Lemma 12** (Covering Lemma). *For any  $P_{XUV}$  on  $X \times \mathbb{F}_q \times \mathbb{F}_q$  and rates  $r_o, r'_o$  and  $r_i$  satisfying (14)-(17), there exists a sequence of two pairs of nested coset codes  $(C_o, C_i)$  and  $(C'_o, C_i)$  which are  $P_{XUV}$ -covering.*

$$r_o \geq \log q - H(U|X) \quad (14)$$

$$r'_o \geq \log q - H(V|X) \quad (15)$$

$$r_o + r'_o \geq 2 \log q - H(U, V|X) \quad (16)$$

$$r_o + r'_o - r_i \geq \log q - H(\alpha U \oplus_q \beta V|X), \forall \alpha, \beta \in \mathbb{F}_q \setminus \{0\}, \quad (17)$$

*Proof.* See Section B.4 in the appendix. □

**Remark 7.** *The only difference between the new mutual covering bounds and the ones for independent codebook generation is the presence of the constraint (17). If  $r_i = 0$ , (17) is redundant, so we recover the mutual covering bounds for independent codebook generation as expected. If  $r_i \neq 0$ , (17) is non-redundant. There is an intuitive explanation for this additional bound. Define  $C_3 = \alpha C_1 \oplus_q \beta C_2$ .  $C_3$  is a coset code with generator matrix  $G_3 = [G^t \ G'^t \ \Delta G^t]^t$ , and the size of this codebook is  $2^{n(r_o + r'_o - r_i)}$ . Suppose there are codevectors  $\mathbf{c}_u \in C_1$  and  $\mathbf{c}_v \in C_2$  jointly typical with  $\mathbf{x}$  with respect to  $P_{UVX}$ , then  $\alpha \mathbf{c}_1 \oplus_q \beta \mathbf{c}_2 \in C_3$  is jointly typical with  $\mathbf{x}$  with respect to  $P_{\alpha U \oplus_q \beta V, X}$ . This implies that  $C_3$  should have size at least  $2^{n(\log q - H(\alpha U \oplus_q \beta V|X))}$  by the converse source coding theorem.*

**Definition 11.** *Let  $\mathbb{F}_q, U, V$  and  $X$  be as in Definition 10. A sequence of code pairs  $(C_1, C_2)$  and bin functions  $B_i : C_i \rightarrow [1, 2^{n\epsilon_i}]$ ,  $i \in \{1, 2\}$  is called  $P_{XUV}$ -packing if for all  $\epsilon > 0$ ,*

$$P \left( \left\{ x^n \left| \begin{array}{l} \exists (c_u^n, c_v^n) \neq (c_u'^n, c_v'^n), \\ (c_u^n, c_v^n) \in A_\epsilon^n(U, V|X) \cap C_1 \times C_2, \\ (c_u'^n, c_v'^n) \in A_\epsilon^n(U, V) \cap C_1 \times C_2, \\ B_1(c_u^n) = B_1(c_u'^n), \quad B_2(c_v^n) = B_2(c_v'^n) \end{array} \right. \right\} \rightarrow 0 \text{ as } n \rightarrow \infty. \right.$$

**Lemma 13** (Packing Lemma). *For any  $P_{XUV}$  on  $X \times \mathbb{F}_q \times \mathbb{F}_q$ , there exists a sequence of two pairs of nested coset codes  $(C_o, C_i)$  and  $(C'_o, C'_i)$  and bin function  $B_i, i \in \{1, 2\}$  which are  $P_{XUV}$ -packing, if  $r_o, r'_o, \rho_1$  and  $\rho_2$  satisfy*

$$r_o - \rho_1 \leq \log q - H(U|V), \quad (18)$$

$$r'_o - \rho_2 \leq \log q - H(V|U), \quad (19)$$

$$(r_o - \rho_1) + (r'_o - \rho_2) \leq 2 \log q - H(U, V). \quad (20)$$

*Proof.* See Section B.5 in the appendix. □

We proceed with explaining the achievability scheme. Define the joint distribution in Table 1 on random variables  $V_{\{1\}}, V_{\{2\}}$  and  $X$ .

$\begin{matrix} V_{\{1\}}, V_{\{2\}} \\ X \end{matrix}$	00	01	10	11
0	$\frac{1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2}D_0$	$\frac{\sqrt{2}-1}{2}D_0$	$\frac{3-2\sqrt{2}}{2}D_0$
1	$\frac{1}{2}D_0$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{\sqrt{2}-1}{2}(1 - D_0)$	$\frac{3-2\sqrt{2}}{2}(1 - D_0)$

Table 1: Joint distribution on  $X, V_{\{1\}}$  and  $V_{\{2\}}$ .

**Codebook Generation:** Set  $r = r_o = r'_o = r_i = 1 - \frac{H(V_{\{1\}}, V_{\{2\}}|X)}{2} + \epsilon$ , and  $\rho_1 = \rho_2 = H(V_{\{1\}}) - \frac{H(V_{\{1\}}, V_{\{2\}}|X)}{2} + \epsilon$  and  $\rho_3 = H(V_{\{1\}} \oplus V_{\{2\}}) - \frac{H(V_{\{1\}}, V_{\{2\}}|X)}{2} + \epsilon$ . Construct a family of coset codes  $C$  with rate  $r$ . Also, construct three binning functions  $B_i : C^n \rightarrow [1, 2^{n\rho_i}]$ ,  $i \in \{1, 2, 3\}$ .

**Encoding:** Upon receiving source sequence  $x^n$ , the encoder finds  $\mathbf{c}_1^n$  and  $\mathbf{c}_2^n$  in the codebook, such that they are jointly typical with  $x^n$  with respect to  $P_{V_{\{1\}}, V_{\{2\}}, X}$ . Such a pair of codewords exists as long as the covering bounds in Lemma 12 are satisfied. In the case at hand it can be readily checked that  $r_o, r'_o$  and  $r_i$  satisfy the bounds. Description 1 carries the bin index of  $\mathbf{c}_1^n$  using  $B_1$ , description 2 carries the bin index of  $\mathbf{c}_2^n$  using  $B_2$  and description 3 carries the bin index of  $\mathbf{c}_1^n + \mathbf{c}_2^n$  using  $B_3$ .

**Decoding:** Decoder {1} receives the bin index carried by description 1, and reconstructs  $\mathbf{c}_1^n$  as long as there is a unique codeword in the bin which is typical with respect to  $P_{V_{\{1\}}}$ . The following packing bound ensures correct decoding with arbitrarily small error:

$$H(V_{\{1\}}) \leq 1 - \rho_1 + r_i.$$

By the same arguments decoder {2} reconstructs  $\mathbf{c}_2^n$  correctly. Decoder {3} reconstructs  $\mathbf{c}_1^n + \mathbf{c}_2^n$  with arbitrarily small error since the following packing bound is satisfied:

$$H(V_{\{1\}} + V_{\{2\}}) \leq 1 - \rho_3 + r_i.$$

We conclude that all the decoders which receive two descriptions would have access to  $\mathbf{c}_1^n$  and  $\mathbf{c}_2^n$ . Decoders {1}, {2} and {3} announce their decoded codewords as their reconstruction of the source. The reconstruction function at the decoders which receive two descriptions is given as follows:

$$\hat{x}_i = \begin{cases} 0 & c_{1i} = c_{2i} = 0 \\ 1 & \text{Otherwise} \end{cases}$$

This implies that the RD vector stated in the lemma is achieved from strong typicality.  $\square$

The following lemma shows that some of the RD vectors in Lemma 11 are not achievable using the SSC scheme.

**Lemma 14.** *The RD vector in (10) is not achievable using the SSC scheme for the following values of  $\alpha$  and  $\beta$  and when the equality holds in (13):*

$$\alpha = \log_2 \frac{1 - 2(\sqrt{2} - 1)D_0}{2(2 - \sqrt{2})D_0}, \quad \beta = -\log_2 \frac{1 - 2(\sqrt{2} - 1)(1 - D_0)}{2(2 - \sqrt{2})(1 - D_0)}$$

For example,  $D_0 = 0.035$ ,  $\alpha = 4.566$  and  $\beta = 2.495$  satisfy the above constraints, where we have rounded the parameters up to the third decimal place.

*Proof.* See Section B.6 in the appendix.  $\square$

## 6 Achievable RD Region using Structured Codes

In this section, we provide a new achievable RD region for the general  $l$ -descriptions problem by enhancing the SSC coding scheme with a structured coding layer. We present this region in four stages. In the first stage, we prove that the SSC region can also be achieved using structured codes. In particular, we use independent nested coset codes for each auxiliary random variable, and exploit the pairwise independence of the codewords to show the achievability of the SSC region. In the subsequent stages, we add coding layers that facilitates the reconstruction of multi-variate functions of the auxiliary random variables. The improvements due to these additional layers comes from exploiting the algebraic structure of the codebooks. In the second stage, we only allow the reconstruction of a bivariate summation of codewords. In the third stage we extend this to a multi-variate summation of the codewords. In the fourth stage, we consider the general case involving the reconstruction of an arbitrary number of multi-variate summations at the decoders.

### 6.1 Stage 1: Achievability of the SSC Region Using Nested Coset Codes

**Definition 12.** *For a joint distribution  $P$  on random variables  $U_{\mathbf{M}}$ ,  $\mathbf{M} \in \mathbf{S}_{\mathbf{L}}$  and  $X$ , and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_{\mathbf{N}} : \mathbf{U}_{\mathbf{N}} \rightarrow \mathbf{X}, \mathbf{N} \in \mathcal{L}\}$ , the set  $\mathcal{RD}_1(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers  $(\rho_{\mathbf{M},i}, r_{o,\mathbf{M}})_{i \in \tilde{\mathbf{M}}, \mathbf{M} \in \mathbf{S}_{\mathbf{L}}}$ :*

$$H(U_{\mathbf{M}}|X) \geq \sum_{\mathbf{M} \in \mathbf{M}} (\log q - r_{o,\mathbf{M}}), \quad \forall \mathbf{M} \in \mathbf{S}_{\mathbf{L}} \quad (21)$$

$$H(U_{\mathbf{M}_{\mathbf{N}}}|U_{\mathbf{L}\mathbf{U}\tilde{\mathbf{M}}_{\mathbf{N}}}) \leq \sum_{\mathbf{M} \in \mathbf{M}_{\mathbf{N}} \setminus \mathbf{L}\mathbf{U}\tilde{\mathbf{M}}_{\mathbf{N}}} (\log q + \sum_{j \in [1:L]} \rho_{\mathbf{M},j} - r_{o,\mathbf{M}}), \quad \forall \mathbf{L} \in \mathbf{M}_{\mathbf{N}}, \forall \mathbf{N} \in \mathcal{L} \quad (22)$$

$$R_i = \sum_{\mathbf{M}} \rho_{\mathbf{M},i}, \quad D_{\mathbf{N}} = E\{d_{\mathbf{N}}(h_{\mathbf{N}}(U_{\mathbf{N}}, X))\}.$$

where  $r_{o,\mathbf{M}} \leq \log q, \forall \mathbf{M} \in \mathbf{S}_{\mathbf{L}}$ .

**Theorem 4.** *The RD vector  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}}$  is achievable for the  $l$ -descriptions problem using nested coset codes, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}} \in \mathcal{RD}_1(P, g_{\mathcal{L}})$ .*

*Proof.* The encoding and decoding steps are exactly the same as the ones in the proof of Theorem 3. The only difference is in the codebook generation phase. In this phase, for every  $\mathcal{M} \in \mathbf{S}_L$ , we generate a coset code  $C_{\mathcal{M}}$  with rate  $r_{\mathcal{M}}$ , generator matrix  $G_{\mathcal{M}}$ , and dither  $b_{\mathcal{M}}$ .  $G_{\mathcal{M}}$  and  $b_{\mathcal{M}}$  are generated randomly and uniformly for every  $\mathcal{M}$ . The bounds in (21) are the mutual covering bounds for independently generated coset codes. These bounds ensure encoding can be carried out without error. The bounds in (22) are the mutual packing bounds in each decoder. They ensure errorless decoding. □

**Lemma 15.** *The RD region in Theorem 4 is equal to the SSC RD region.*

*Proof.* See Section C.1 in the appendix. □

## 6.2 Stage 2: Reconstruction of a summation of two codebooks

In the first stage we constructed one codebook for each subset of the decoders. However, only the codebooks corresponding to the Sperner families of sets are shown to be non-redundant. We interpret this using the notion of common-information as defined by Gacs, Körner, Witsenhausen [30] [31]. Let  $K(A_1; A_2)$  denote the common information between any two random variables  $A_1$  and  $A_2$ . The common information among  $m$  random variables  $A_1, A_2, \dots, A_m$  is a vector of length  $(2^m - m - 1)$  of information that is common among every subset of  $m$  random variables of size at least two. When  $m = 3$ , the common information is given by

$$[K(A_1; A_2; A_3), K(A_1; A_2), K(A_1; A_3), K(A_2; A_3)].$$

This was referred to as univariate common information in [17], as each of these components are characterized using univariate functions. We interpret the scheme in the first stage (SSC scheme) as capturing the common-information components among the random variables associated with  $2^l - 1$  decoders using univariate functions.

For  $m = 3$ , this notion of common information was generalized using bivariate functions to the following seven-dimensional vector in [17]:

$$[K(A_1; A_2; A_3), K(A_1; A_2), K(A_1; A_3), K(A_2; A_3), K(A_1; A_2, A_3), K(A_2; A_1, A_3), K(A_3; A_1, A_2)].$$

There are seven degrees of freedom in having information common among 3 random variables. The latter three are called bivariate common information components as they are characterized using bivariate functions of random variables. In this sense, the addition of the structured coding layers in the next stages can be thought of as capturing the common-information among  $2^l - 1$  decoders using bivariate and, more generally, multivariate functions.

We extend the notion of bivariate common information to  $m > 3$  random variables as follows. To characterize a bivariate common information component, we consider three subsets of  $\mathbf{N}_1, \mathbf{N}_2$  and  $\mathbf{N}_3$  of  $\{1, 2, 3, \dots, m\}$ . Define  $K(\tilde{A}_{\mathbf{N}_1}; \tilde{A}_{\mathbf{N}_2}, \tilde{A}_{\mathbf{N}_3})$  as a bivariate common information component among  $A_1, A_2, \dots, A_m$ , where  $\tilde{A}_{\mathbf{N}_i}$  is the information that is common among  $A_{\mathbf{N}_i}$ . For example for  $m = 4$ , let  $\mathbf{N}_1 = \{4\}$ ,  $\mathbf{N}_2 = \{1, 2\}$  and  $\mathbf{N}_3 = \{3\}$ . This characterizes the information in  $A_4$  that can be computed by a conference via a bivariate function of (i) the information common between  $A_1$  and  $A_2$ , and (ii) the information in  $A_3$ . This concept can be extended to define multivariate common information among  $m$  random variables.

We return to our discussion on the achievable RD region for the MD problem, where  $m = 2^l - 1$ . In the second stage, we aim to capture the bivariate common information among random variables associated with  $2^l - 1$  decoders. In particular, we reconstruct a summation of two codebooks. From the above arguments,

instead of one codebook for each subset of decoders as in the first stage, in this stage we need to construct one codebook for every triple of subsets of the decoders. For a given triple of sets of decoders, the third set of decoders reconstruct a bivariate summation of a random variable corresponding to the first subset and a random variable corresponding to the second subset of decoders. This is explained in more detail next. We add two new codebooks to the SSC scheme. The underlying random variables for these two codebooks are denoted by  $V_{\mathcal{A}_1}$  and  $V_{\mathcal{A}_2}$  where  $\mathcal{A}_i \in \mathbf{S}_L, i \in \{1, 2\}, \mathcal{A}_1 \neq \mathcal{A}_2$ . We construct two pairs of nested coset codes for these two random variables. The two nested coset codes have the same inner code. The codebook corresponding to  $V_{\mathcal{A}_i}$  is decoded at decoder N if  $\mathcal{A}_i \in \mathbf{M}_N$ , furthermore, the sum of the two codebooks is decoded at decoder N if  $\mathcal{A}_3 \in \mathbf{M}_N \setminus \{\mathcal{A}_1, \mathcal{A}_2\}$ , where  $\mathcal{A}_3$  is an element of  $\mathbf{S}_L$ . For example, let us choose  $\mathcal{A}_i = \{\{i\}, i \in \{1, 2, 3\}\}$ . In this case the first codebook is decoded whenever description 1 is received, the second codebook is decoded if description 2 is received, and the sum is decoded whenever description 3 is received. This corresponds to the coding schemes we presented for example 4, where  $V_{\mathcal{A}_1} = X + N_\delta$  and  $V_{\mathcal{A}_2} = Z + N'_\delta$ . The following theorem describes the achievable RD region using this scheme.

**Definition 13.** For any three distinct families  $\mathcal{A}_i \in \mathbf{S}_L, i = 1, 2, 3$ , and for a joint distribution  $P$  on random variables  $U_M, M \in \mathbf{S}_L, V_{\mathcal{A}_j}, j \in \{1, 2\}$ , and  $X$ , where the underlying alphabet for all auxiliary random variables is the field  $\mathbb{F}_q$ , and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_N : \mathbf{U}_N \rightarrow X, N \in \mathcal{L}\}$ , the set  $\mathcal{RD}_2(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers  $(\rho_{M,i}, r_{o,M})_{i \in \overline{M}, M \in \mathbf{S}_L}$  and  $\rho_{o,\mathcal{A}_j,i}, r'_{o,\mathcal{A}_j}, i \in \overline{\mathcal{A}_j}, j \in \{1, 2, 3\}$  and  $r_i$ :

$$H(U_M V_E | X) \geq \sum_{M \in \mathbf{M}} (\log q - r_{o,M}) + \sum_{E \in \mathbf{E}} (\log q - r'_{o,E}), \forall \mathbf{M} \subset \mathbf{S}_L, \mathbf{E} \subset \mathbf{A} \quad (23)$$

$$H(U_M, W_{\mathcal{A}_3, \alpha, \beta} | X) \geq \sum_{M \in \mathbf{M}} (\log q - r_{o,M}) + \log q - r'_{o,\mathcal{A}_3}, \forall \mathbf{M} \subset \mathbf{S}_L, \forall \alpha, \beta \in \mathbb{F}_q \setminus \{0\} \quad (24)$$

$$H([U, V, W]_{\overline{\mathbf{M}_N}} | [U, V, W]_{\widehat{\mathbf{M}_N} \cup \overline{\mathbf{L}}}) \leq \sum_{M \in \mathbf{M}_N \setminus \overline{\mathbf{M}_N} \cup \overline{\mathbf{L}}} (\log q + \sum_{j \in \overline{M}} \rho_{M,j} - r_{o,M}) + \sum_{\substack{M \in \mathbf{M}_N \setminus \overline{\mathbf{M}_N} \cup \overline{\mathbf{L}} \\ \cap \{\mathcal{A}_i, i \in [1,3]\}}} (\log q + \sum_{j \in \overline{M}} \rho_{o,M,j} - r'_{o,M}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}_N} \quad (25)$$

$$R_i = \sum_M \rho_{M,i}, \quad D_N = E\{d_N(h_N(U_N, X))\}. \quad (26)$$

where (a)  $\mathbf{A} \triangleq \{\mathcal{A}_1, \mathcal{A}_2\}$ , (b)  $\overline{\mathbf{M}_N} \triangleq (\mathbf{M}_N, \{\mathcal{A}_j, j \in \{1, 2\} | \mathcal{A}_j \in \mathbf{M}_N\}, \{\{\mathcal{A}_3, 1, 1\} | \mathcal{A}_3 \in \mathbf{M}_N\})$ , (c)  $\widehat{\mathbf{M}_N} \triangleq \bigcup_{N' \subset N} \overline{\mathbf{M}_{N'}}$ , (d)  $r'_{o,\mathcal{A}_3} \triangleq r'_{o,\mathcal{A}_1} + r'_{o,\mathcal{A}_2} - r_i$ , (e)  $r_{o,M} \leq \log q$ , and (f)  $W_{\mathcal{A}_3, \alpha, \beta} \triangleq \alpha V_{\mathcal{A}_1} + \beta V_{\mathcal{A}_2}$ <sup>1</sup>.

**Theorem 5.** The RD vector  $(R_i, D_N)_{i \in L, N \in \mathcal{L}}$  is achievable for the  $l$ -descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_i, D_N)_{i \in L, N \in \mathcal{L}} \in \mathcal{RD}_2(P, g_{\mathcal{L}})$ .

Before providing the proof we explain the bounds in the new RD region. (23) and (25) are the mutual covering and packing bounds which are also present in the Theorem 3, respectively. (24) is a generalization of the additional covering bound derived in the Lemma in 12. Note that the common component among decoders  $N \in \mathcal{A}_1$  is the pair  $(U_{\mathcal{A}_1}, V_{\mathcal{A}_1})$ , and similarly for  $\mathcal{A}_2$ . The common component among decoders  $N \in \mathcal{A}_3$  is the pair  $(U_{\mathcal{A}_1}, W_{\mathcal{A}_3, 1, 1})$ , and observe that  $W_{\mathcal{A}_3, 1, 1} = V_{\mathcal{A}_1} + V_{\mathcal{A}_2}$ .

*Proof.* Given a joint distribution  $P_{U, V, X}$ , and codebook and binning rates satisfying the bounds in the theorem we prove achievability of the RD vector in (26).

<sup>1</sup>We have used the script  $\mathcal{A}$  to denote subscripts of random variables throughout the paper. However, the collection  $\{\mathcal{A}_3, \alpha, \beta\}$  is used as the subscript for  $W$  since the random variable is defined using  $\alpha$  and  $\beta$ .

**Codebook Generation:** Fix blocklength  $n$ . For every  $\mathcal{M} \in \mathbf{S}_L$ , independently generate a linear code  $C_{\mathcal{M}}$  with size  $2^{nr_{o,\mathcal{M}}}$ . Also generate two nested coset codes  $C_{\mathcal{A}_j} = (C_i, C_{o,\mathcal{A}_j})$ ,  $j \in \{1, 2\}$  where the inner code has rate  $r_i$  and the outer codes have rates  $r'_{o,\mathcal{A}_j}$ . Define the set of codewords  $C_{o,\mathcal{A}_3} \triangleq C_{o,\mathcal{A}_1} + C_{o,\mathcal{A}_2}$ . The size of  $C_{o,\mathcal{A}_3}$  is  $2^{nr'_{o,\mathcal{A}_3}}$ , where  $r'_{o,\mathcal{A}_3} = r'_{o,\mathcal{A}_1} + r'_{o,\mathcal{A}_2} - r_i$ . For the  $i$ th description bin the codebook  $C_{\mathcal{M}}$  randomly and uniformly with rate  $2^{nr_{\mathcal{M},i}}$ .

**Encoding:** Upon receiving the source vector  $X^n$ , the encoder finds a jointly-typical set of codewords  $c_{\mathcal{M}}$ . Each description carries the *bin-indices* of all of the corresponding codewords. The encoder declares an error if there is no jointly typical set of codewords available.

**Decoding:** Having received the bin-indices from descriptions  $i \in \mathbf{N}$ , decoder  $\mathbf{N}$  tries to find a set of jointly typical codewords  $c_{\mathcal{M}}$ ,  $\mathcal{M} \in \overline{\mathbf{M}}_{\mathbf{N}}$ . If the set of codewords is not unique, the decoder declares error.

In order for the encoder to find a set of jointly typical codewords, the mutual covering bounds (23) and (24) should hold. This is a generalization of the result in lemma 12 and we omit the proof for brevity. The bounds in (25) are the mutual packing bounds at each decoder. □

**Remark 8.** Here we have considered the general case where  $\mathcal{A}_i$  are chosen arbitrarily from  $\mathbf{S}_L$ . It turns out that only certain choices of  $\mathcal{A}_i$  would give non-redundant codebooks and thus provide improvements over the SSC scheme. One can show that the codebooks are redundant if  $\exists N \in \mathcal{A}_1 \cup \mathcal{A}_2, N' \in \mathcal{A}_3$  such that  $N \subset N'$ . For example take  $\mathcal{A}_1 = \{\{1\}, \{3\}\}$ ,  $\mathcal{A}_2 = \{\{2\}\}$  and  $\mathcal{A}_3 = \{\{2, 3\}\}$ .

### 6.3 Stage 3: Reconstruction of a summation of arbitrary number of codebooks

In this section we reconstruct a multi-variate summation of an arbitrary number  $m$  of random variables at one decoder where  $m \in L$  and the summation is with respect to a finite field  $\mathbb{F}_q$ . Following the steps in the previous section, we add  $m$  new codebooks to the original SSC scheme. Let  $\mathbf{M} \triangleq [1, m]$ . The underlying random variables for these codebooks are denoted by  $V_{\mathcal{A}_k}$ ,  $k \in \mathbf{M}$ . The random variable  $V_{\mathcal{A}_k}$  is decoded at decoder  $\mathbf{N}$  if  $\mathcal{A}_k \in \mathbf{M}_{\mathbf{N}}$ . We take the families  $\mathcal{A}_k, k \in \mathbf{M}$  to be distinct. The random variable  $\sum_{i \in \mathbf{M}} V_{\mathcal{A}_k}$  is decoded at decoder  $\mathbf{N}$  if  $\mathcal{A}_{m+1} \in \mathbf{M}_{\mathbf{N}}$ , where  $\mathcal{A}_{m+1}$  is an element of  $\mathbf{S}_L$ . The following theorem describes the achievable RD region:

**Definition 14.** For any  $m \in L$ , and  $m+1$  distinct families  $\mathcal{A}_i \in \mathbf{S}_L, i \in [1, m+1]$ , and for a joint distribution  $P$  on random variables  $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_L, V_{\mathcal{A}_k}, k \in \mathbf{M}$  and  $X$ , where the underlying alphabet for the auxiliary random variables is the field  $\mathbb{F}_q$ , and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_{\mathbf{N}} : \mathbf{U}_{\mathbf{N}} \rightarrow \mathbf{X}, \mathbf{N} \in \mathcal{L}\}$ , the set  $\mathcal{RD}_3(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers  $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i \in \overline{\mathbf{M}}, \mathcal{M} \in \mathbf{S}_L}$  and  $\rho_{o,\mathcal{A}_k,i}, \rho_{o,\mathcal{A}_{m+1},i}, r'_{o,\mathcal{A}_k}, i \in \overline{\mathbf{A}}_k, k \in [1, m+1]$  and  $r_{i,\alpha_j}, J \subset \mathbf{M}$ :

$$H(U_{\mathbf{M}}V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M} \in \mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E} \in \mathbf{E}} (\log q - r'_{o,\mathcal{E}}), \forall \mathbf{M} \subset \mathbf{S}_L, \mathbf{E} \subset \mathbf{A}, \quad (27)$$

$$H(U_{\mathbf{M}}W_{\mathbf{F}}|X) \geq \sum_{\mathcal{M} \in \mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{F} \in \mathbf{F}} (\log q - r'_{o,\mathcal{F}}), \forall \mathbf{M} \subset \mathbf{S}_L, \mathbf{F} \subset \mathbf{B}, \quad (28)$$

$$H([U, V, W]_{\overline{\mathbf{M}}_{\mathbf{N}}} | [U, V, W]_{\overline{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\mathbf{N}} \setminus \overline{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}}} (\log q + \sum_{j \in \overline{\mathbf{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M} \in \mathbf{M}_{\mathbf{N}} \setminus \overline{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}} \\ \cap \{\mathcal{A}_i | i \in [1, m+1]\}}} (\log q + \sum_{j \in \overline{\mathbf{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathbf{N}} \quad (29)$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathbf{N}} = E\{d_{\mathbf{N}}(h_{\mathbf{N}}(U_{\mathbf{N}}, X))\}. \quad (30)$$

where (a)  $\mathbf{A} = \{\mathcal{A}_k, k \in \mathbf{M}\}$ , (b)  $\mathbf{B} = \{(\mathcal{A}_{m+1}, \alpha_{\mathbf{M}}) | \alpha_{\mathbf{M}} \in \mathbb{F}_q^m\}$ , (c)  $r'_{o, \mathcal{A}_{m+1}, \alpha_{\mathbf{M}}} = \sum_{k \in \mathbf{J}} r'_{o, \mathcal{A}_k} - r_{i, \alpha_{\mathbf{J}}}$ ,  $\mathbf{J} = \{k | \alpha_k \neq 0\}$ , (d)  $\sum_{J': J \subset J'} r_{i, \alpha_{J'}} \leq r_{i, \alpha_J}$ ,  $\forall \mathbf{J} \subset \mathbf{M}$ , (e)  $\overline{\mathbf{M}}_{\mathbf{N}} = (\mathbf{M}_{\mathbf{N}}, \{\mathcal{A}_k | \mathcal{A}_k \in \mathbf{M}_{\mathbf{N}}\}, \{(\mathcal{A}_{m+1}, \alpha_{\mathbf{M}}) | \mathcal{A}_{m+1} \in \mathbf{M}_{\mathbf{N}}, \alpha_i = 1, i \in \mathbf{M}\})$ , (f)  $\widehat{\mathbf{M}}_{\mathbf{N}} = \bigcup_{N' \subset \mathbf{N}} \overline{\mathbf{M}}_{N'}$ , (g)  $r_{o, \mathbf{M}} \leq \log q$ , and (h)  $W_{\mathcal{A}_{m+1}, \alpha_{\mathbf{M}}} = \sum_{i \in \mathbf{M}} \alpha_i V_{\mathcal{A}_k}$ .

**Theorem 6.** *The RD vector  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}}$  is achievable for the  $l$ -descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}} \in \mathcal{RD}_3(P, g_{\mathcal{L}})$ .*

Toward proving the theorem we need the following definition.

**Definition 15.** *A set of  $m$  coset codes  $C_{\mathcal{A}_k}^n, k \in \mathbf{M}$  is called an ensemble of nested coset codes with parameter  $(r_{\mathbf{J}})_{\mathbf{J} \subset \mathbf{M}}$  if the size of the intersection  $C_{\mathcal{A}_{\mathbf{J}}} \triangleq \bigcap_{k \in \mathbf{J}} C_{\mathcal{A}_k}$  is equal to  $2^{nr_{\mathbf{J}}}$  for all  $\mathbf{J} \subset \mathbf{M}$ .*

It is straightforward to show that one can always generate an ensemble of nested coset codes  $C_{\mathcal{A}_k}, k \in \mathbf{M}$  with parameter  $(r_{i, \alpha_{\mathbf{J}}})_{\mathbf{J} \subset \mathbf{M}}$  as long as  $\sum_{J': J \subset J'} r_{i, \alpha_{J'}} \leq r_{i, \alpha_J}, \forall \mathbf{J} \subset \mathbf{M}$ . It is enough to choose the rows of the generator matrices of  $C_{\mathcal{A}_k}, k \in \mathbf{J}$  such that they have  $nr_{i, \alpha_{\mathbf{J}}}$  common rows, similar to the case of Figure 9.

*Proof.* We provide an outline of the proof. The codebook generation for codebooks  $C_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathbf{L}}$  is similar to the previous scheme. For random variables  $V_{\mathcal{A}_k}, k \in \mathbf{M}$  we construct an ensemble of nested coset codes  $C_{\mathcal{A}_k}, k \in \mathbf{M}$  with parameter  $(r_{i, \alpha_{\mathbf{J}}})_{\mathbf{J} \subset \mathbf{M}}$ . The encoder chooses a set of codewords from all the codebooks that is jointly typical with the source sequence. The following is a generalized covering lemma which shows that if (27) and (28) is satisfied such a set of codewords exists.

**Definition 16.** *Let  $\mathbb{F}_q$  be a field and define  $\mathcal{M} \triangleq \{\{1\}, \{2\}, \dots, \{m\}\}$ . Consider  $m + 1$  random variables  $X, V_{\{i\}}, i \in \mathbf{M}$ , where  $X$  is defined on an arbitrary finite set  $\mathbf{X}$  and  $V_{\{i\}}$  are defined on  $\mathbb{F}_q$ . Fix a PMF  $P_{X, V_{\mathcal{M}}}$  on  $\mathbf{X} \times \mathbb{F}_q^m$ . A sequence of  $m$ -tuples of codebooks  $(C_{\{i\}}^n)_{\{i\} \in \mathbf{M}}$  is called  $P_{X, V_{\mathcal{M}}}$ -covering if:*

$$\forall \epsilon > 0, P(\{x^n | \exists v_{\mathcal{M}}^n \in A_{\epsilon}^n(V_{\mathcal{M}} | x^n) \cap \prod_{\{i\} \in \mathbf{M}} C_{\{i\}}\}) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

**Lemma 16 (Covering Lemma).** *For any  $P_{X, V_{\mathcal{M}}}$  on  $\mathbf{X} \times \mathbb{F}_q^m$  and rates  $r_{o, \{j\}}, \{j\} \in \mathcal{M}$  satisfying (31)-(33), there exists a sequence of ensemble of nested coset codes  $C_{\mathcal{M}}^n$  with parameter  $(r_{i, \mathbf{J}})_{\mathbf{J} \subset \mathbf{M}}$  which are  $P_{X, V_{\mathcal{M}}}$ -covering.*

$$H(V_{\mathcal{J}} | X) \geq \sum_{\{j\} \in \mathcal{J}} (\log q - r_{o, \{j\}}), \forall \mathcal{J} \subset \mathcal{M} \quad (31)$$

$$H(W_{\mathcal{K}} | X) \geq \sum_{\alpha_{\mathbf{M}} \in \mathcal{K}} (\log q - r_{o, \alpha_{\mathbf{M}}}), \forall \mathcal{J} \subset \mathcal{M}, \mathcal{K} \subset \mathcal{N} \quad (32)$$

$$\sum_{J': J \subset J'} r_{i, \alpha_{J'}} \leq r_{i, \alpha_J}, \forall \mathbf{J} \subset \mathbf{M}, \quad (33)$$

where, (a)  $\mathcal{N} \triangleq \{\alpha_{\mathbf{M}} \in \mathbb{F}_q^m\}$ , (b)  $W_{\alpha_{\mathbf{M}}} \triangleq \sum_{j \in \mathbf{M}} \alpha_j V_{\{j\}}$  and (c)  $r_{o, \alpha_{\mathbf{M}}} \triangleq \sum_{j \in \mathbf{J}} r_{o, \{j\}} - r_{i, \mathbf{J}}, \mathbf{J} = \{k | \alpha_k \neq 0\}$ .

*Proof.* The proof of the lemma follows the same steps as in lemma 12. We provide the intuition behind the proof. Given that there is a set of codewords in the codebooks  $C_{\{j\}}, j \in \mathbf{M}$  which are jointly typical with the source sequence, for any linear combination  $C \triangleq \sum_{j \in \mathbf{M}} \alpha_j C_{\{j\}}$  there is a codeword which is jointly typical with the random variables  $X, V_{\mathcal{M}}, W_{\mathcal{N}}, \forall \mathcal{M}, \mathcal{N}$ . From a PtP perspective, the rate of codebook  $C$  must satisfy (31) and (32). This rate can be calculated by counting the number of rows in the generator matrix of  $C$  which is  $nr_{o, \alpha_{\mathbf{M}}}$ .  $\square$



The packing bounds at each encoder can be written in the same way as in the previous section and are given in (29).  $\mathbf{B}$  is defined such that  $W_{\mathbf{B}}$  is the set of all possible linear combinations of  $V_{\mathcal{A}_k}$ 's. □

#### 6.4 Stage 4: Reconstruction of an Arbitrary Number of Summations of Arbitrary Lengths

In this section for completeness, we provide a coding scheme where we reconstruct multi-variate summations of random variables at an arbitrary number of decoders, and these summations each have arbitrary lengths. Of course, due to the large number of random variables the coding scheme becomes extremely complicated. Let the number of the summations be  $s$ , and for each summation, let the length of the summation be denoted by  $m_i \in L, i \in [1, s]$ . Define the sets  $\mathbf{S} \triangleq [1, s]$  and  $M_i \triangleq [1, m_i], i \in \mathbf{S}$ . Following the steps in the previous sections, we add  $m_i$  new codebooks for each summation. The underlying random variables for these codebooks are denoted by  $V_{\mathcal{A}_{k,i}}, k \in M_i, i \in \mathbf{S}$ . The random variable  $V_{\mathcal{A}_{k,i}}$  is decoded at decoder  $\mathbf{N}$  if  $\mathcal{A}_{k,i} \in \mathbf{M}_{\mathbf{N}}$ . Fix the prime number  $q_i$ . The random variable  $\sum_{j \in M_i} V_{\mathcal{A}_{k,i}}$  is decoded at decoder  $\mathbf{N}$  if  $\mathcal{A}_{m_i+1,i} \in \mathbf{M}_{\mathbf{N}}$ , where the summation is carried out in the finite field  $\mathbb{F}_{q_i}$ . The following theorem describes the achievable RD region.

**Definition 17.** For a joint distribution  $P$  on random variables  $U_{\mathcal{M}}, \mathcal{M} \in \mathbf{S}_{\mathbf{L}}, V_{\mathcal{A}_{k,i}}, k \in M_i, i \in \mathbf{S}$  and  $X$ , where the underlying alphabet the auxiliary random variables is the field  $\mathbb{F}_q$ , and a set of reconstruction functions  $g_{\mathcal{L}} = \{g_{\mathbf{N}} : \mathbf{U}_{\mathbf{N}} \rightarrow \mathbf{X}, \mathbf{N} \in \mathcal{L}\}$ , the set  $\mathcal{RD}_{\text{linear}}(P, g_{\mathcal{L}})$  is defined as the set of RD vectors satisfying the following bounds for some non-negative real numbers  $(\rho_{\mathcal{M},i}, r_{o,\mathcal{M}})_{i \in \widetilde{\mathbf{M}}, \mathcal{M} \in \mathbf{S}_{\mathbf{L}}}$  and  $\rho_{o,\mathcal{A}_{k,i},j,k}, \rho_{o,\mathcal{A}_{m_i+1,i},j,k}, r_{o,\mathcal{A}_{k,i}}, j,k \in \widetilde{A}_{k,i}, k \in [1, m_i + 1], i \in \mathbf{S}$  and  $r_{i,\alpha_{j,i}}, J \subset M_i, i \in \mathbf{S}$ :

$$H(U_{\mathbf{M}}, V_{\mathbf{E}}|X) \geq \sum_{\mathcal{M} \in \mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{E} \in \mathbf{E}} (\log q - r_{o,\mathcal{E}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathbf{L}}, \mathbf{E} \subset \mathbf{A}, \quad (34)$$

$$H(U_{\mathbf{M}}, W_{\mathbf{F}}|X) \geq \sum_{\mathcal{M} \in \mathbf{M}} (\log q - r_{o,\mathcal{M}}) + \sum_{\mathcal{F} \in \mathbf{F}} (\log q - r_{o,\mathcal{F}}), \forall \mathbf{M} \subset \mathbf{S}_{\mathbf{L}}, \mathbf{F} \subset \mathbf{B}, \quad (35)$$

$$H([U, V, W]_{\overline{\mathbf{M}}_{\mathbf{N}}} | [U, V, W]_{\widehat{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\mathbf{N}} \setminus \overline{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}}} (\log q + \sum_{j \in \widetilde{\mathcal{M}}} \rho_{\mathcal{M},j} - r_{o,\mathcal{M}}) + \sum_{\substack{\mathcal{M} \in \mathbf{M}_{\mathbf{N}} \setminus \overline{\mathbf{M}}_{\mathbf{N}} \cup \overline{\mathbf{L}} \\ \cap \{\mathcal{A}_{[i],s} | i \in [1, m_s + 1], s \in \mathbf{S}\}}} (\log q + \sum_{j \in \widetilde{\mathcal{M}}} \rho_{o,\mathcal{M},j} - r'_{o,\mathcal{M}}), \quad \forall \overline{\mathbf{L}} \subset \overline{\mathbf{M}}_{\mathbf{N}} \quad (36)$$

$$R_i = \sum_{\mathcal{M}} \rho_{\mathcal{M},i}, \quad D_{\mathbf{N}} = E\{d_{\mathbf{N}}(h_{\mathbf{N}}(U_{\mathbf{N}}, X))\}. \quad (37)$$

where (a)  $\mathbf{A} = \cup_{i \in \mathbf{S}} \{\mathcal{A}_{k,i} | k \in M_i\}$ , (b)  $\mathbf{B} = \cup_{i \in \mathbf{S}} \{(\mathcal{A}_{m_i+1,i}, \alpha_{M_i,i}) | \alpha_{j,i} \in \mathbb{F}_{q_i}\}$ , (c)  $r_{o,\mathcal{A}_{m_i+1,i}, \alpha_{M_i,i}} = \sum_{k \in J} r_{o,\mathcal{A}_{k,i}} - r_{i,\alpha_{j,i}}, J_i = \{k | \alpha_{k,i} \neq 0\}, i \in \mathbf{S}$ , (d)  $\sum_{J' \subset J} r_{i,\alpha_{j',i}} \leq r_{i,\alpha_{j,i}}, \forall J \subset M_i, i \in \mathbf{S}$ , (e)  $\overline{\mathbf{M}}_{\mathbf{N}} = (\mathbf{M}_{\mathbf{N}}, \{\mathcal{A}_{k,i} | k \in M_i, i \in \mathbf{S}, \mathcal{A}_{k,i} \in \mathbf{M}_{\mathbf{N}}\}, \{(\mathcal{A}_{m_i+1,i}, \alpha_{M_i,i}) | \mathcal{A}_{m_i+1,i} \in \mathbf{M}_{\mathbf{N}}, \alpha_{k,i} = 1\})$ , (f)  $\widehat{\mathbf{M}}_{\mathbf{N}} = \cup_{\mathbf{N}' \subset \mathbf{N}} \overline{\mathbf{M}}_{\mathbf{N}'}$ , (g)  $r_{o,\mathcal{M}} \leq \log q$  and (h)  $W_{\mathcal{A}_{m_i+1,i}, \alpha_{M_i,i}} = \sum_{j=1}^{m_i} \alpha_{j,i} V_{\mathcal{A}_{k,i}}$ .

**Theorem 7.** The RD vector  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}}$  is achievable for the  $l$ -descriptions problem, if there exists a distribution  $P$  and reconstruction functions  $g_{\mathcal{L}}$  such that  $(R_i, D_{\mathbf{N}})_{i \in \mathbf{L}, \mathbf{N} \in \mathcal{L}} \in \mathcal{RD}_{\text{linear}}(P, g_{\mathcal{L}})$ .

*Proof.* This is a straightforward generalization of the previous step, since the proof is similar, it is omitted. □

**Remark 9.** Similar to Theorem 3 one can identify the non-redundant codebooks in the above scheme. One can show that a large number of possible codebooks become redundant in this case as well.

## 7 Conclusion

We provided several improvements over previous coding strategies for the MD problem. First, we showed that the CMSB strategy can be enhanced using additional unstructured quantizers and a new unstructured binning approach. We demonstrated these gains using examples involving binary sources and three descriptions. We provided the resulting RD region for the  $l$ -descriptions problem with arbitrary sources and distortion functions. Additionally, we proved that all of the new codebooks in our scheme are non-redundant for the  $l$ -descriptions problem. In the second part of the paper, we introduced structure in both the quantizer construction as well as in the binning functions. We showed through several examples that the improvements derived in the first part can be extended if structured quantizers and binning functions are utilized. The RD region in the first part of the paper was further improved upon by introducing additional linear coding layers. Lastly, we combined the ideas in the two parts to provide a new and strictly improved achievable RD region for the  $l$ -descriptions problem.

## Appendices

### A Proofs for Section 4

#### A.1 Proof of Lemma 1

*Proof.* Let  $(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}} \in \mathcal{RD}_{SSC}(P_{\mathbf{U}, X})$  and  $(R'_i, D'_N)_{i \in \mathcal{L}, N \in \mathcal{L}} \in \mathcal{RD}_{SSC}(P'_{\mathbf{U}, X})$ . Without loss of generality, assume  $\mathbf{U}_{\{1, \{2, \{3\}\}} = \mathbf{U}'_{\{1, \{2, \{3\}\}}$ . Let  $\tilde{U}_{\{1, \{2, \{3\}\}}$  be defined on  $\mathbf{U}_{\{1, \{2, \{3\}\}} \times \{0, 1\}$ . Also let  $\tilde{U}_{\mathcal{M}} = U_{\mathcal{M}}$  if  $\mathcal{M} \neq \{\{1\}, \{2\}, \{3\}\lambda\}$ . For  $\lambda \in [0, 1]$ , define a new distribution  $\tilde{P}_{\tilde{\mathbf{U}}, X}$  as follows:

$$\tilde{P}_{\tilde{\mathbf{U}}, X}(\tilde{\mathbf{u}}, x) = \begin{cases} \lambda P_{\mathbf{U}, X}(\mathbf{u}, x) & \tilde{u}_{\{1, \{2, \{3\}\}} = (u_{\{1, \{2, \{3\}\}}, 0) \\ (1 - \lambda) P'_{\mathbf{U}, X}(\mathbf{u}, x) & \tilde{u}_{\{1, \{2, \{3\}\}} = (u_{\{1, \{2, \{3\}\}}, 1) \end{cases}$$

Then it is straightforward to check that  $\lambda(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}} + (1 - \lambda)(R'_i, D'_N)_{i \in \mathcal{L}, N \in \mathcal{L}} \in \mathcal{RD}_{SSC}(\tilde{P}_{\tilde{\mathbf{U}}, X})$ .  $\square$

#### A.2 Proof of lemma 2

*Proof.* We provide an outline of the proof. Fix  $M' \in \mathcal{L}$ . Consider a new scheme where the reconstruction function at decoder  $M'$  is defined as  $f_{M'} : \prod_{M \in \mathcal{M}_{M'}} U_M \rightarrow X$  with the rest of the reconstruction functions defined as in Theorem 3. Let the RD vector  $(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}}$  be achievable in the new scheme using the distribution  $P_{U_{\mathcal{S}_L}, X}$  and reconstruction functions  $f_M, g_M, M \in \mathcal{L} \setminus \{N\}$ . We provide a new probability distribution  $P'_{U_{\mathcal{S}_L}, X}$  and reconstruction functions  $g'_M : U_M \rightarrow X, M \in \mathcal{L}$  to show that the RD region given in Theorem 3 contains  $(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}}$ . To construct the probability distribution define  $U'_M = U_M, M \in \mathcal{S}_N \setminus \{M'\}$ , and  $U'_{M'} = (U_{M'}, f_{M'}(U_{\mathcal{M}_{M'}}))$ . As for the reconstruction functions define  $g'_M(U_M) = g_M(U_M), M \in \mathcal{L} \setminus \{N\}$  and  $g'_{M'}(U'_{M'}) = f_{M'}(U_{\mathcal{M}_{M'}})$ . It is straightforward to check that with these parameters, the RD region in theorem 3 contains  $(R_i, D_N)_{i \in \mathcal{L}, N \in \mathcal{L}}$ . Intuitively, since the reconstruction functions are the same, the same distortion is achieved by both schemes. As for the rates, in the first scheme, wherever  $U_{M'}$  is decoded, all of the random variables  $U_{\mathcal{M}_{M'}}$  are also decoded. So, adding a function of these random variables to  $U_{M'}$  does not require additional rate.  $\square$

### A.3 Proof of lemma 4

*Proof.* Let  $U_{\{1\}} = \hat{W}$ ,  $U_{\{1,2\},\{3\}} = W$ ,  $U_{\{1,2\}} = \hat{X}_1$ ,  $U_{\{3\}} = \hat{X}_2$ , where  $\hat{X}_i$  are the reconstructions at decoder  $\{i\}$  in the two user problem in Example 1. Then it is straightforward to check that the RD vector is achievable from Theorem 3. Next, assuming the codebook  $C_{\{1,2\},\{3\}}$  is empty, we consider all of the remaining 16 codebooks in the SSC scheme and show that the RD vector is not achievable.

**Step 1:** In this step, we argue that the only non-trivial codebooks are  $C_{\{1\}}$ ,  $C_{\{3\}}$ ,  $C_{\{1\},\{3\}}$ ,  $C_{\{1,2\}}$  and  $C_{\{2\},\{3\}}$ . Due to the structure of the problem, a number of the codebooks are functionally equivalent, meaning they are decoded at exactly the same decoders. So we can merge these codebooks without any loss. For example, description  $\{2\}$  is only received by decoders  $\{1, 2\}$  and  $\{1, 2, 3\}$ , hence we can merge  $C_{\{2\}}$  into  $C_{\{1,2\}}$  without any loss.  $C_{\{1,3\},\{23\}}$ ,  $C_{\{1,3\}}$ ,  $C_{\{2,3\}}$  and  $C_{\{1,2,3\}}$  are only decoded at decoder  $\{1, 2, 3\}$  so they are redundant from the results in [25].  $C_{\{1\},\{2\}}$  can be merged into  $C_{\{1\}}$  since decoder  $\{2\}$  is not present.  $C_{\{1\},\{2\},\{3\}}$  is equivalent to  $C_{\{1\},\{3\}}$  and can be eliminated.  $C_{\{1,2\},\{1,3\},\{2,3\}}$ ,  $C_{\{1,2\},\{1,3\}}$  and  $C_{\{1,2\},\{2,3\}}$  can be merged into  $C_{\{1,2\}}$ . Finally  $C_{\{2,3\},\{1\}}$  can be merged with  $C_{\{1\}}$ . Also  $C_{\{1,2\}}$  can be merged with  $C_{\{1,2,3\}}$  and is eliminated. So we are left with four codebooks  $C_{\{1\}}$ ,  $C_{\{3\}}$ ,  $C_{\{1\},\{3\}}$ ,  $C_{\{1,2\}}$  and  $C_{\{2\},\{3\}}$ .

**Step 2:** In this step, we show that if we set  $U_{\{1\},\{3\}} = \hat{W}$  and  $U_{\{1\}} = \phi$ , there would be no loss in terms of RD function. The codebooks  $C_{\{1\}}$  and  $C_{\{1\},\{3\}}$  are decodable using description 1. Since decoder  $\{1\}$  is at PtP optimality, these codebooks only carry  $\hat{W}$ . To be more precise there is a Markov chain  $(U_{\{1\}}, U_{\{1\},\{3\}}) \leftrightarrow \hat{W} \leftrightarrow X$ , which we prove in the following lemma.

**Lemma 17.** *In a PtP setup assume the decoder is at optimal PtP RD. It receives variables  $U_{\mathbf{M}}$ , and the reconstruction function is  $f(U_{\mathbf{M}})$ . Then the following Markov chain holds  $U_{\mathbf{M}} \leftrightarrow f(U_{\mathbf{M}}) \leftrightarrow X$ .*

*Proof.*

$$\begin{aligned} R &\geq I(U_{\mathbf{M}}; X) \stackrel{(a)}{=} I(f(U_{\mathbf{M}}), U_{\mathbf{M}}; X) = I(f(U_{\mathbf{M}}); X) + I(U_{\mathbf{M}}; X|f(U_{\mathbf{M}})) \\ &\stackrel{(b)}{\geq} R + I(U_{\mathbf{M}}; X|f(U_{\mathbf{M}})) \Rightarrow I(U_{\mathbf{M}}; X|f(U_{\mathbf{M}})) = 0, \end{aligned}$$

where in (a) we used the fact that  $f(U_{\mathbf{M}})$  is a function of  $U_{\mathbf{M}}$  and in (b) we used the PtP optimality.  $\square$

Since  $\hat{W}$  is decoded both at decoder  $\{1\}$  and  $\{3\}$ , if we replace  $U_{\{1\},\{3\}}$  with  $(U_{\{1\},\{3\}}, \hat{W})$ , the decoders decode the same random variables as before, so no extra rate is required. Also, from the lemma  $(U_{\{1\}}, U_{\{1\},\{3\}}) \leftrightarrow \hat{W} \leftrightarrow X$ . Hence, we conclude that we can set  $U_{\{1\},\{3\}} = \hat{W}$  and  $U_{\{1\}} = \phi$  without any loss in terms of distortion.

**Step 3:** Assume there are random variables  $U_{\{1\},\{3\}}$  and  $U_{\{2\},\{3\}} = (W, U'_{\{2\},\{3\}})$  such that the RD vector is achievable in the SSC scheme. From the Markov chain  $\hat{W} \leftrightarrow W \leftrightarrow X$ , description 1 is not used in the reconstruction in decoders  $\{1, 2\}$ ,  $\{3\}$  and  $\{1, 2, 3\}$ . If we set  $U_{\{1\}} = \phi$ , the distortions constraint in decoders  $\{1, 2\}$ ,  $\{3\}$  and  $\{1, 2, 3\}$  are satisfied. So we have constructed a scheme to send the descriptions at a lower rate (by setting  $U_{\{1\}} = \phi$ ) without any loss in terms of distortion in these three decoders. This contradicts optimality of the random variables chosen for the two user scheme.

### A.4 Proof of lemma 3 for $l > 3$

We have proved that if  $C_{\{12\},\{3\}} = \phi$ , the RD vector is not achievable but if the constraint is lifted the scheme can achieve this RD vector, so the codebook is non-redundant. For the general  $l$ -descriptions problem, we provide an outline of the non-redundancy proof for  $C_{\mathcal{H}}, \mathcal{H} \in \mathbf{S}_L$ . Let  $\{a_{1,i}, a_{2,i}, \dots, a_{n_i,i}\}, i \in$

$[1, k]$  be the elements of  $\mathcal{H}$ . Then to construct an example where  $C_{\mathcal{H}}$  is non-redundant, first consider a set up where for any  $i$ , each set of three decoders  $\{a_{1,i}, a_{2,i}, \dots, a_{n_i,i}\}$  and  $\{a_{1,i+1}, a_{2,i+1}, \dots, a_{n_{i+1},i+1}\}$  and  $\{a_{1,i}, a_{2,i}, \dots, a_{n_i,i}, a_{1,i+1}, a_{2,i+1}, \dots, a_{n_{i+1},i+1}\}$  are as in the two user setup in Example 1. Then there should be a common component between each two of the descriptions. It is straightforward to show that the common components must be the same for all of the decoders, otherwise since the codebooks are independent there would be a rate-loss as explained in the previous section. We ensure that the common component can be decoded only when all descriptions  $a_{1,i}a_{2,i} \dots a_{n_i,i}a_{1,i+1}a_{2,i+1} \dots a_{n_{i+1},i+1}$  are received and not when a subset of the descriptions is received. This is done by adding decoders  $\{a_{1,i}\}$ ,  $\{a_{1,i}, a_{2,i}\}$  through  $\{a_{1,i}, a_{2,i}, \dots, a_{n_i,i}, a_{1,i+1}, a_{2,i+1}, \dots, a_{n_{i+1},i+1}\}$  such that each of them would be at PtP optimality by receiving a refined version of  $W$  (i.e  $\{a_{1,i}\}$  would receive  $\hat{W}$  and  $\{a_{1,i}, a_{2,i}\}$  would receive a refinement of  $\hat{W}$  and so on). In this way the only codebook that can carry  $W$  without rate-loss is  $C_{\mathcal{H}}$ .  $\square$

## A.5 Proof of Lemma 5

*Proof.* Let  $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} > 0$ , description 1 carries  $\hat{W}$  to decoder  $\{1\}$  with rate  $I(\hat{W}; X)$ . Descriptions 2 and 3 send  $W$  to decoders  $\{1, 2\}$  and  $\{1, 3\}$  by sending a refinement on  $C_{\{1,2\},\{1,3\}}$ . In other words  $U_{\{1\}} = \hat{W}$ ,  $U_{\{1,2\},\{1,3\}} = W$ ,  $U_{\{1,2\}} = \hat{X}_1$  and  $U_{\{1,2\}} = \hat{X}_2$  similar to the proof of Lemma 4. Then one can check that the RD vector is achievable using the SSC scheme. Next, assume  $\rho_{\{1,2\},\{1,3\},2} + \rho_{\{1,2\},\{1,3\},3} = 0$ , then  $\rho_{\{1,2\},\{1,3\},i} = 0, i \in \{2, 3\}$ . As in the previous section, we begin by eliminating the redundant codebooks for this communications setting.

**Step 1:** In this step we argue that only the codebooks  $C_{\{1,3\}}$ ,  $C_{\{1,2\}}$   $C_{\{1\}}$  and  $C_{\{1,2\},\{1,3\}}$  are non-trivial. Due to the structure of this communications setting many of the codebooks are functionally the same and can be merged together. The codebooks  $C_{\{1\},\{2\},\{3\}}$ ,  $C_{\{1\},\{3\}}$ ,  $C_{\{1\},\{2\}}$ ,  $C_{\{2,3\},\{1\}}$  are decoded at all four of the decoders and can be merged with  $C_{\{1\}}$ .  $C_{\{1,3\},\{2,3\}}$  can be merged with  $C_{\{1,3\}}$  since decoder  $\{2, 3\}$  is not present, by the same argument  $C_{\{1,2\},\{1,3\},\{2,3\}}$  is concatenated with  $C_{\{1,2\},\{1,3\}}$ , also  $C_{\{1,2\},\{2,3\}}$  and  $C_{\{1,2\},\{3\}}$  are merged with  $C_{\{1,2\}}$ .  $C_{\{1,3\},\{2\}}$  and  $C_{\{2\},\{3\}}$  are combined with  $C_{\{1,2\},\{2,3\}}$ . Lastly since decoders 2 and 3 are not present,  $C_{\{2\}}$  and  $C_{\{3\}}$  can be merged into  $C_{\{1,2\}}$  and  $C_{\{1,3\}}$ , respectively. So only the four codebooks  $C_{\{1,3\}}$ ,  $C_{\{1,2\}}$   $C_{\{1\}}$  and  $C_{\{1,2\},\{1,3\}}$  remain.

**Step 2:** By the same arguments as in step 2 of Lemma 4, we can set  $U_{\{1\}} = \hat{W}$ .

**Step 3:** By assumption, the codebook  $C_{\{1,2\},\{1,3\}}$  is only carried by the first description. However, the codebook is not decoded at decoder  $\{1\}$ . Since the decoder is at PtP optimality,  $C_{\{1,2\},\{1,3\}}$  can't be sent through the first description either (i.e  $\rho_{\{1,2\},\{1,3\},1} = 0$  and  $C_{\{1,2\},\{1,3\}}$  can be eliminated.).

**Step 4:** After Fourier-Motzkin elimination, the covering and packing bounds for the remaining three codebooks give the following inequality,

$$R_1 + R_2 + R_3 \geq I(U_{\{1,2\}}, U_{\{1,3\}}; X|\hat{W}) + I(U_{\{1,3\}}; U_{\{1,2\}}|\hat{W}) + I(\hat{W}; X) \quad (38)$$

By the definition of  $\hat{W}$  we have  $\hat{W} \leftrightarrow W \leftrightarrow X$  and  $I(\hat{W}; X) < I(W; X)$ , so the bound above is strictly larger than the case when  $\hat{W}$  is replaced by  $W$  (i.e. when  $U_{\{1,2\},\{1,3\}} = W$ ). This concludes the proof.  $\square$

## B Proofs for Section 5

### B.1 Proof of Lemma 7

*Proof.*

$$\begin{aligned} \frac{1}{n}E(d_H(\hat{X}^n \oplus_2 \hat{Z}^n, X^n \oplus_2 Z^n)) &= \frac{1}{n}E(w_H(\hat{X}^n \oplus_2 \hat{Z}^n \oplus_2 X^n \oplus_2 Z^n)) \\ &= \frac{1}{n}E(w_H(X^n \oplus_2 \hat{X}^n \oplus_2 Z^n \oplus_2 \hat{Z}^n)) \\ &= \frac{1}{n}E(d_H(X^n \oplus_2 \hat{X}^n, Z^n \oplus_2 \hat{Z}^n)). \end{aligned}$$

Note that  $X^n \oplus_2 \hat{X}^n$  is the quantization noise of quantizing  $X^n$  and  $Z^n \oplus_2 \hat{Z}^n$  is the quantization noise of quantizing  $Z^n$ . Since the source vectors are independent, the noise vectors are also independent and the summation converges to  $\delta * \delta$  (The arguments are similar to the ones given in [28]).

□

### B.2 Proof of Lemma 8

*Proof.* We assume that there exists a probability distribution  $P$  on  $X$  and  $U_{S_L}$  for which the RD vector is achievable using the SSC scheme and arrive at a contradiction. Since all of the decoders are present in this setup, we need to consider the SSC with all the codebooks present, so the proof is more involved than the proofs in the previous section.

**Step 1:** In this step we show that description  $i$ , where  $i = 1, 2$ , does not carry any bin indices for codewords from codebook  $C_{\mathcal{M}}$  if  $\mathcal{M} \notin \mathbf{M}_{\{i\}}$ . Descriptions 1 and 2 only carry indices which are used in the reconstruction at decoders  $\{1\}$  and  $\{2\}$ , respectively. This is true since these two decoders are receiving information at optimal PtP rate-distortion. Note that this does not mean the corresponding codebooks are empty, we can only conclude that no bin indices for the codewords are sent through these descriptions. For example if  $\mathcal{M} = \{\{2\}, \{1, 3\}\}$  and  $i = 1$ , then  $\rho_{\mathcal{M},i} = 0$ .

**Lemma 18.** For  $i \in \{1, 2\}$ , and  $\mathcal{M}$  such that  $\{i\} \notin \mathcal{M}$ ,  $\rho_{\mathcal{M},i} = 0$ .

*Proof.* From optimality at decoder  $\{1\}$  we have the following equality:

$$R_i = I(U_{\mathbf{M}_{\{i\}}}; X, Z) \quad (39)$$

Consider the following covering bound on the random variables  $U_{\mathbf{M}_{\{i\}}}$ :

$$H(U_{\mathbf{M}_{\{i\}}}|X, Z) \geq \sum_{\mathcal{M} \in \mathbf{M}_{\{i\}}} (H(U_{\mathcal{M}}) - r_{\mathcal{M}}), \quad (40)$$

also we have the following packing bound at decoder  $\{i\}$ :

$$H(U_{\mathbf{M}_{\{i\}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\{i\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},i} - r_{\mathcal{M}}), \quad (41)$$

adding (40) and (41) we get:

$$\sum_{\mathcal{M} \in \mathbf{M}_{\{i\}}} \rho_{\mathcal{M},i} \geq I(U_{\mathbf{M}_{\{i\}}}; X, Z), \quad (42)$$

$R_i = \sum_{\mathcal{M} \in \mathcal{S}_L} \rho_{\mathcal{M},i}$ , comparing this equality with (39) completes the proof.  $\square$

**Step 2:** In this step, we show that there are no common codebooks decoded at decoders {1} and {2}. Since decoder {1, 2} receives descriptions 1 and 2 at optimal RD from a PtP perspective, the random variables decoded at decoder {1} must be independent of those decoded at decoder {2}. From the next lemma we have that if  $\mathcal{M} \in \mathbf{M}_{\{1\}} \cap \mathbf{M}_{\{2\}}$  then  $r_{\mathcal{M}} = 0$ .

**Lemma 19.** Consider the setup in Figure 1, let  $(R_1, R_2, D_1, D_2, D_{\{1,2\}})$  be such that  $R_1 + R_2 = RD_{d_{\{1,2\}}}(D_{\{1,2\}})$ , where  $RD_d(D)$  is Shannon's optimal PtP RD function for distortion function  $d$  at point  $D$ . For any distribution  $P_{U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}, U_{\{1,2\}}}$  which achieves this RD vector, the following conditions must hold: 1)  $U_{\{1\}} \perp U_{\{2\}}$  and  $C_{\{1,2\}} = \phi$   
2) If in addition  $R_i = RD_{d_i}(D_i)$ ,  $i \in \{1, 2\}$  then,  $U_{\{1,2\}} \leftrightarrow (U_{\{1\}}, U_{\{2\}}) \leftrightarrow X$ .

*Proof.* Consider the following packing bounds:

$$\text{Dec } \{1\} : H(U_{\{1,2\}}, U_{\{1\}}) \leq H(U_{\{1,2\}}) + H(U_{\{1\}}) + \rho_{\{1,2\},1} + \rho_{\{1,1\}} - r_{\{1,2\}} - r_{\{1\}} \quad (43)$$

$$\text{Dec } \{2\} : H(U_{\{1,2\}}, U_{\{2\}}) \leq H(U_{\{1,2\}}) + H(U_{\{2\}}) + \rho_{\{1,2\},2} + \rho_{\{2,2\}} - r_{\{1,2\}} - r_{\{2\}} \quad (44)$$

$$\text{Dec } \{1, 2\} : H(U_{\{1,2\}} | U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}) \leq H(U_{\{1,2\}}) + \rho_{\{1,2\},1} + \rho_{\{1,2\},2} - r_{\{1,2\}} \quad (45)$$

Also the mutual covering bound:

$$H(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}} | X) \geq H(U_{\{1,2\}}) + H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{1,2\}}) - r_{\{1,2\}} - r_{\{1\}} - r_{\{2\}} - r_{\{1,2\}} \quad (46)$$

Now we add inequalities (43-45) and subtract (46), we get:

$$I(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}, U_{\{1,2\}}; X) + I(U_{\{1\}}; U_{\{2\}} | U_{\{1,2\}}) \leq R_1 + R_2 - r_{\{1,2\}}$$

Using the condition  $R_1 + R_2 = RD_{d_{\{1,2\}}}(D_{\{1,2\}})$  we conclude:

$$I(U_{\{1\}}; U_{\{2\}} | U_{\{1,2\}}) + r_{\{1,2\}} \leq 0 \quad (47)$$

From (47) one may deduce  $C_{\{1,2\}} = \phi$  and  $U_{\{1\}} \perp U_{\{2\}}$ . Furthermore we get:

$$R_1 + R_2 = I(U_{\{1,2\}}, U_{\{1\}}, U_{\{2\}}; X) = I(U_{\{1\}}; X) + I(U_{\{2\}}; X) + I(U_{\{1\}}; U_{\{2\}} | X) + I(U_{\{1,2\}}; X | U_{\{1\}}, U_{\{2\}}),$$

where the right-hand side of the second equality is the sum-rate of the two-descriptions problem. Using the conditions  $R_i = RD_{d_i}(D_i)$ ,  $i \in \{1, 2\}$ , we have:

$$I(U_{\{1\}}; U_{\{2\}} | X) + I(U_{\{1,2\}}; X | U_{\{1\}}, U_{\{2\}}) = 0.$$

So  $I(U_{\{1,2\}}; X | U_{\{1\}}, U_{\{2\}}) = 0$ , which gives the desired Markov chain in (2).  $\square$

Assuming the original scheme achieves the RD vector in the theorem, we give a new scheme which also achieves the RD vector. We propose that the encoder operates as before, but decoder {1, 2} decodes  $U_{\mathcal{M}}$  only if  $\mathcal{M} \in \mathbf{M}_{\{1\}}$  or  $\mathcal{M} \in \mathbf{M}_{\{2\}}$ . It needs to be shown that the RD vector is the same. First we consider the resulting rates. The covering bounds are not changed. The packing bounds are the same at all decoders

other than decoder  $\{1, 2\}$  since the same variables are being decoded at those decoders.  $\mathbf{M}_{\{1\}} \cap \mathbf{M}_{\{2\}} = \phi$ . Let  $\tilde{\mathbf{M}}_{\{1\}}$  and  $\tilde{\mathbf{M}}_{\{2\}}$  be subsets of  $\mathbf{M}_{\{1\}}$  and  $\mathbf{M}_{\{2\}}$ . We need to show that the following packing bound is satisfied:

$$H(U_{\mathbf{M}_{\{1\}}}, U_{\mathbf{M}_{\{2\}}} | U_{\tilde{\mathbf{M}}_{\{1\}}}, U_{\tilde{\mathbf{M}}_{\{2\}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\{1\}} \cup \mathbf{M}_{\{2\}} \setminus \tilde{\mathbf{M}}_{\{1\}} \cup \tilde{\mathbf{M}}_{\{2\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},1} + \rho_{\mathcal{M},2} - r_{\mathcal{M}}) \quad (48)$$

We have the following two packing bounds from decoders  $\{1\}$  and  $\{2\}$ :

$$H(U_{\mathbf{M}_{\{1\}}} | U_{\tilde{\mathbf{M}}_{\{1\}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\{1\}} \setminus \tilde{\mathbf{M}}_{\{1\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},1} - r_{\mathcal{M}}) \quad (49)$$

$$H(U_{\mathbf{M}_{\{2\}}} | U_{\tilde{\mathbf{M}}_{\{2\}}}) \leq \sum_{\mathcal{M} \in \mathbf{M}_{\{2\}} \setminus \tilde{\mathbf{M}}_{\{2\}}} (H(U_{\mathcal{M}}) + \rho_{\mathcal{M},2} - r_{\mathcal{M}}) \quad (50)$$

Note that from arguments in Lemma 19,  $U_{\mathbf{M}_{\{1\}}}$  is independent of  $U_{\mathbf{M}_{\{2\}}}$ . Hence adding (49) and (50), we get (48). This proves that the packing bounds are also the same.

From lemma 19, we have  $U_{\mathbf{M}_{\{1,2\}}} \leftrightarrow U_{\mathbf{M}_{\{1\}}}, U_{\mathbf{M}_{\{2\}}} \leftrightarrow X, Z$ . Lemma 20 shows that the new scheme achieves the same distortions as the previous one.

**Lemma 20.** *Let the random variables  $U, V, X$  be such that  $U \leftrightarrow V \leftrightarrow X$ . Then for an arbitrary distortion function  $f : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$ , there is an optimal reconstruction of  $X$  using  $U$  and  $V$  which is a only function of  $V$ .*

*Proof.* We know that the optimal reconstruction function for  $X$  given  $U$  and  $V$  is given by:

$$g(u, v) = \arg \min_{\hat{x} \in \hat{\mathcal{X}}} \mathbf{E}(f(\hat{x}, X) | u, v) = \arg \min_{\hat{x} \in \hat{\mathcal{X}}} \mathbf{E}(f(\hat{x}, X) | v),$$

which is only a function of  $V$ . □

By these arguments, codebook  $U_{\mathcal{M}}$  is eliminated if  $\mathcal{M} \in \mathbf{M}_{\{1,2\}} \setminus \tilde{\mathbf{M}}_{\{1,2\}}$ . Also in the new scheme,  $U_{\{1,2\},\{1,3\},\{2,3\}}$  and  $U_{\{1,3\},\{2,3\}}$  are functionally similar since by the same arguments as in this step  $U_{\{1,2\},\{1,3\},\{2,3\}}$  is not used in the reconstruction in decoder  $\{1, 2\}$ , so we can eliminate  $C_{\{1,2\},\{1,3\},\{2,3\}}$ . In summary, thus far we have eliminated 7 codebooks.

**Step 3:** We have the following lemma:

**Lemma 21.** *From optimality of rate and distortion at decoders  $\{1, 3\}, \{2, 3\}$  we have:*

$$\rho_{\{2,3\},3} = \rho_{\{1,3\},3} = \rho_{\{2,3\},\{1,3\}} = \rho_{\{1,3\},\{2,3\}} = 0$$

*Proof.* First we argue that  $\rho_{\{2,3\},3} = 0$ . If this is not true, it contradicts optimality at decoder  $\{1, 3\}$ .  $U_{\{2,3\}}$  is not decoded at decoder  $\{1, 3\}$ , but its bin index is carried through description 3. So if the bin index is non-zero, one could reduce  $R_3$  by setting the bin index equal to 0 without increasing distortion at decoder  $\{1, 3\}$ , this contradicts optimality at that decoder. By the same arguments  $\rho_{\{1,3\},3} = 0$ . Now assume  $\rho_{\{2,3\},\{1,3\}} \neq 0$ . We show that this contradicts optimality at decoder  $\{1, 3\}$ .  $U_{\{2,3\},\{1\}}$  is decodable using description 1 (since it is decodable at decoder  $\{1\}$ ). Hence, if we set  $\rho_{\{2,3\},\{1\},3}$  to 0 (i.e. do not send the bin index on description 3), then decoder  $\{1, 3\}$  can still decode  $U_{23,1}$  using description 1. So the distortion is the same at this decoder, but the rate  $R_3$  is reduced which contradicts optimality. By the same arguments,  $\rho_{\{1,3\},\{2,3\}} = 0$ . □

**Step 4:** We proceed by showing that  $r_{\{1,3\}} = r_{\{2,3\}} = 0$ . So far we have shown that none of the descriptions carry the bin indices for these codebooks. Consider the following packing bounds in decoders  $\{1\}$ ,  $\{2, 3\}$  and  $\{1, 3\}$ :

$$\begin{aligned}
H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) &\leq H(U_{\{1\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2,3\},\{1\}}) + R_1 - r_{\{1\}} - r_{\{1\},\{3\}} - r_{\{2,3\},\{1\}} \\
H(U_{\{2\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) &\leq H(U_{\{2\}}) + H(U_{\{3\}}) + H(U_{\{1\},\{3\}}) + \\
H(U_{\{2\},\{3\}}) + H(U_{\{2,3\}}) + H(U_{\{1,3\},\{2\}}) + H(U_{\{2,3\},\{1\}}) + H(U_{\{1,3\},\{2,3\}}) + R_2 + R_3 - r_{\{2\}} - r_{\{3\}} - r_{\{1\},\{3\}} - \\
r_{\{2\},\{3\}} - r_{\{2,3\}} - r_{\{1,3\},\{2\}} - r_{\{2,3\},\{1\}} - r_{\{1,3\},\{2,3\}} \\
H(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) &\leq H(U_{\{1,3\}}) - r_{\{1,3\}}
\end{aligned}$$

We add the above inequalities and subtract the mutual covering bound on all RV's, we get:

$$\begin{aligned}
&H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) + H(U_{\{2\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) \\
&+ H(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) \\
&- H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}, U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, U_{\{1,3\},\{2\}}, U_{\{2,3\},\{1\}}, U_{\{2,3\}}, U_{\{1,3\},\{2,3\}}, U_{\{1,3\}}|X, Z) \\
&\leq H(U_{\{1,3\}}) + H(U_{\{2,3\},\{1\}}) - r_{\{1,3\}} - r_{\{2,3\},\{1\}} + R_1 + R_2 + R_3 \\
&\Rightarrow I(U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}; X, Z) + I(U_{\{1\}}, U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}; X, Z) + \\
I(U_{\{1,3\}}; X, Z|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) &\leq R_1 + R_2 + R_3 \\
&\Rightarrow I(X, Z; U_{\{1,3\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}) = 0.
\end{aligned}$$

This imposes the Markov chain  $U_{\{1,3\}} \leftrightarrow U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}} \leftrightarrow X, Z$ . Hence by the same arguments as in step 2, we can eliminate  $C_{\{1,3\}}$ . Also by the same arguments  $C_{\{2,3\}}$  can be eliminated.

**Step 5:** In this step we eliminate  $C_{\{1\},\{3\}}$  and  $C_{\{2\},\{3\}}$ .

**Lemma 22.** *The following equality holds:*

$$\rho_{\{1\},\{3\},1} = \rho_{\{1\},\{3\},3} = \rho_{\{2\},\{3\},2} = \rho_{\{2\},\{3\},1} = 0$$

*Proof.* Assume  $\rho_{\{1\},\{3\},1} > 0$ . We claim this contradicts optimality at decoder  $\{1, 3\}$ , since  $U_{\{1\},\{3\}}$  can readily be decoded from the bin number carried by description 3, so setting  $\rho_{\{1\},\{3\},1}$  to 0 would decrease rate without increasing distortion. The rest of the proof follows by the same argument.  $\square$

Now consider the following packing bounds at decoders  $\{1\}$ ,  $\{3\}$  and  $\{1, 3\}$  and the mutual covering bound:

$$\begin{aligned}
H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) &\leq H(U_{\{1\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2,3\},\{1\}}) + R_1 - r_{\{1\}} - r_{\{1\},\{3\}} - r_{\{2,3\},\{1\}} \\
H(U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}) &\leq H(U_{\{3\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) + R_3 - \rho_{\{1,3\},\{2,3\},3} - r_{\{3\}} - r_{\{1\},\{3\}} - r_{\{2\},\{3\}} \\
H(U_{\{1,3\},\{2,3\}}, U_{\{1,3\},\{2\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\},\{1\}}) &\leq H(U_{\{1,3\},\{2,3\}}) + H(U_{\{1,3\},\{2\}}) + \rho_{\{1,3\},\{2,3\},3} - r_{\{1,3\},\{2,3\}} \\
H(U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{1,3\},\{2\}}U_{\{2,3\},\{1\}}U_{\{1,3\},\{2,3\}}|X, Z) &\geq H(U_{\{1\}}) + H(U_{\{3\}}) + H(U_{\{1\},\{3\}}) + H(U_{\{2\},\{3\}}) \\
+ H(U_{\{1,3\},\{2\}}) + H(U_{\{2,3\},\{1\}}) + H(U_{\{1,3\},\{2,3\}}) - r_{\{1\}} + r_{\{3\}} - r_{\{1\},\{3\}} - r_{\{2\},\{3\}} - r_{\{1,3\},\{2\}} - r_{\{2,3\},\{1\}} - r_{\{1,3\},\{2,3\}}
\end{aligned}$$

Adding the above packing bounds and subtracting the mutual covering bound we get:

$$H(U_{\{1\}}U_{\{1\},\{3\}}U_{\{2,3\},\{1\}}) + H(U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}) + H(U_{\{1,3\},\{2,3\}}, U_{\{1,3\},\{2\}}|U_{\{1\}}U_{\{3\}}U_{\{1\},\{3\}}U_{\{2\},\{3\}}U_{\{2,3\},\{1\}})$$



$$\begin{aligned}
& -H(U_{\{1\}}U_{\{3\}}U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}U_{\{1,\{3\}\{2\}\}}U_{\{2,\{3\}\{1\}\}}U_{\{1,\{3\}\{2,\{3\}\}\}|X,Z) \leq R_1 + R_3 + H(U_{\{1\}\{3\}} - r_{\{1,\{3\}\}}) \\
& \Rightarrow I(U_{\{1\}}U_{\{1,\{3\}\}}U_{\{2,\{3\}\{1\}\}; U_{\{3\}}U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}) + \\
& I(U_{\{1\}}U_{\{3\}}U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}U_{\{1,\{3\}\{2\}\}}U_{\{2,\{3\}\{1\}\}}U_{\{1,\{3\}\{2,\{3\}\}\}|X,Z) - H(U_{\{1,\{3\}\}}) \leq R_1 + R_3 - r_{\{1,\{3\}\}} \\
& \Rightarrow I(U_{\{1\}}U_{\{2,\{3\}\{1\}\}; U_{\{3\}}U_{\{2,\{3\}\}}|U_{\{1,\{3\}\}}) + r_{\{1,\{3\}\}} \leq 0
\end{aligned}$$

Particularly  $r_{\{1,\{3\}\}} = 0$ , by the same arguments  $r_{\{2,\{3\}\}} = 0$ .

**Step 6:** So far we have shown that only  $C_{\{1\}}, C_{\{2\}}, C_{\{3\}}, C_{\{1,\{3\}\{2\}\}, C_{\{2,\{3\}\{1\}\}}$  and  $C_{\{1,\{3\}\{2,\{3\}\}}$  can be non-trivial. From optimality at decoders  $\{1\}$  and  $\{1, 3\}$  we have the following equalities:

$$R_1 = I(U_{\{1\}}, U_{\{2,\{3\}\{1\}\}; X, Z), R_1 + R_3 = I(U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, U_{\{3\}}, U_{\{1,\{3\}\{2\}\}, U_{\{1,\{3\}\{2,\{3\}\}\}; X, Z) \quad (51)$$

Hence we have:

$$R_3 = I(U_{\{3\}}, U_{\{1,\{3\}\{2\}\}, U_{\{1,\{3\}\{2,\{3\}\}\}; X, Z|U_{\{1\}}, U_{\{2,\{3\}\{1\}\}}) \quad (52)$$

Define the following:

$$N_\delta^1 \triangleq X + h_{\{1\}}(U_{\{1\}}, U_{\{2,\{3\}\{1\}\}}) \quad (53)$$

$$N_{\delta^*\delta}^3 \triangleq X + Z + h_{\{3\}}(U_{\{3\}}) \quad (54)$$

$$N_\delta^{1,3} \triangleq Z + h_{\{1,\{3\}\}}(U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, U_{\{3\}}, U_{\{1,\{3\}\{2\}\}, U_{\{1,\{3\}\{2,\{3\}\}\}}), \quad (55)$$

where  $h_{\{1\}}$  is the reconstruction of  $X$  at decoder  $\{1\}$ ,  $h_{\{3\}}$  is the reconstruction of  $X + Z$  at decoder  $\{3\}$ , and  $h_{\{1,\{3\}\}}$  is the reconstruction of  $Z$  at decoder  $\{1, 3\}$ . Then from (52):

$$\begin{aligned}
R_3 & \geq I(h_{\{1,\{3\}\}}(\cdot), h_{\{3\}}(U_{\{3\}}); X, Z|U_{\{1\}}, U_{\{2,\{3\}\{1\}\}}h_{\{1\}}(\cdot)) \\
& \Rightarrow R_3 \geq I(Z + N_\delta^{1,3}, X + Z + N_{\delta^*\delta}^3; X, Z|U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, X + N_\delta^{(1)}}) \\
& \Rightarrow R_3 \geq H(Z|U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, X + N_\delta^{(1)}}) - H(Z|Z + N_\delta^{1,3}, X + Z + N_{\delta^*\delta}^3, U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, X + N_\delta^{(1)}}) \\
& \stackrel{(a)}{\Rightarrow} R_3 \geq 1 - H(Z|Z + N_\delta^{1,3}, X + Z + N_{\delta^*\delta}^3, U_{\{1\}}, U_{\{2,\{3\}\{1\}\}, X + N_\delta^{(1)}}) \\
& \Rightarrow R_3 \geq 1 - H(Z|Z + N_\delta^{1,3}) \\
& \Rightarrow R_3 \geq 1 - H(N_\delta^{1,3}|Z + N_\delta^{1,3}) \\
& \stackrel{(b)}{\Rightarrow} R_3 \geq 1 - h_b(\delta)
\end{aligned}$$

All the above inequalities must be equality. In particular we have:

$$\begin{aligned}
(a) & \Rightarrow Z \leftrightarrow Z + N_\delta^{1,3} \leftrightarrow X + Z + N_{\delta^*\delta}^3, X + N_\delta^{(1)} \\
& \Rightarrow N_\delta^{1,3} \leftrightarrow Z + N_\delta^{1,3} \leftrightarrow Z + N_{\delta^*\delta}^3 + N_\delta^{(1)} \\
& \Rightarrow N_\delta^{1,3} \leftrightarrow Z + N_\delta^{1,3} \leftrightarrow N_\delta^{1,3} + N_{\delta^*\delta}^3 + N_\delta^{(1)}
\end{aligned}$$

Note that from (b), we can conclude that  $Z$  is independent of  $N_\delta^{1,3}$ , we have  $N_\delta^{1,3}$  and  $N_\delta^{1,3} + N_{\delta^*\delta}^3 + N_\delta^{(1)}$  are independent. Define  $N' \triangleq N_{\delta^*\delta}^3 + N_\delta^{(1)}$ . We have:

$$\begin{aligned}
P(N_\delta^{1,3} + N' = 0) & \stackrel{(a)}{=} P(N_\delta^{1,3} + N' = 0|N_\delta^{1,3} = 0) = P(N' = 0|N_\delta^{1,3} = 0) \\
& \stackrel{(b)}{\Rightarrow} P(N' = 0, N_\delta^{1,3} = 0) = (1 - \delta)(P(N' = 0, N_\delta^{1,3} = 0) + P(N' = 1, N_\delta^{1,3} = 1))
\end{aligned}$$

$$\Rightarrow P(N' = 0, N_\delta^{(1,3)} = 0) = \frac{1-\delta}{\delta} P(N' = 1, N_\delta^{(1,3)} = 1)$$

(a) holds since  $N_\delta^{1,3}$  and  $N_\delta^{1,3} + N_{\delta*\delta}^3 + N_\delta^{(1)}$  are independent. In (b) we have replaced  $P(N_\delta^{(1,3)} + N' = 0)$  by  $P(N' = 0, N_\delta^{(1,3)} = 0) + P(N' = 1, N_\delta^{(1,3)} = 1)$ .

Define  $a \triangleq P(N' = 1, N_\delta^{(1,3)} = 1)$ , then by the same calculations  $P(N' = 1, N_\delta^{(1,3)} = 0) = (1-\delta)(1 - \frac{1}{\delta}a)$ , so  $P(N' = 1) = 1 - \delta + \frac{2\delta-1}{\delta}a$ . Note  $a = P(N' = 1, N_\delta^{(1,3)} = 1) \leq P(N_\delta^{(1,3)} = 1) = \delta$ , hence using  $P(N' = 1) = 1 - \delta + \frac{2\delta-1}{\delta}a$ , we get  $P(N' = 1) \leq \delta$  with equality if and only if  $a = \delta$ . Also note that  $Z + N'$  is available at decoder  $\{1, 3\}$  so  $P(N' = 1) = \delta$  and  $a = \delta$ , otherwise there is a contradiction with optimality of  $h_{\{1,3\}}$ . If  $a = \delta$ , then  $N_\delta^{(1,3)}$  is equal to  $N'$ . So by the same arguments we have:

$$N_{\delta*\delta}^{(3)} = N_\delta^{(1,3)} + N_\delta^{(1)} = N_\delta^{(2,3)} + N_\delta^{(2)},$$

where

$$N_\delta^2 \triangleq Z + h_{\{2\}}(U_{\{2\}}, U_{\{1,3\},\{2\}}) \quad (56)$$

$$N_\delta^{2,3} \triangleq Z + h_{\{2,3\}}(U_{\{2\}}, U_{\{1,3\},\{2\}}, U_{\{3\}}, U_{\{2,3\},\{1\}}, U_{\{1,3\},\{2,3\}}) \quad (57)$$

Since  $N_\delta^{(1)} \perp\!\!\!\perp N_\delta^{(2)}$ ,  $N^{(1)} \perp\!\!\!\perp N^{(1,3)}$  and  $N^{(2)} \perp\!\!\!\perp N^{(2,3)}$ , we have:

$$N_\delta^{(1,3)} = N_\delta^2, N_\delta^{(2,3)} = N_\delta^{(1)}, N_{\delta*\delta}^{(3)} = N_\delta^1 + N_\delta^2$$

We argue that  $C_{\{1,3\},\{2\}}$ ,  $C_{\{2,3\},\{1\}}$  and  $C_{\{1,3\},\{2,3\}}$  can be taken eliminated without any loss in RD. To prove this assume we have a scheme with  $P_{U_{\{1,3\},\{2\}}, U_{\{2,3\},\{1\}}, U_{\{1,3\},\{2,3\}}, U_{\{1\}}, U_{\{2\}}, U_{\{3\}}}$ . Construct new random variables  $\tilde{U}_{\{1\}} = X + N_\delta^{(1)}$ ,  $U_{\{1\}}$ ,  $U_{\{1\},\{2,3\}}$ ,  $\tilde{U}_{\{2\}} = Z + N_\delta^{(2)}$ ,  $U_{\{2\}}$ ,  $U_{\{2\},\{1,3\}}$  and  $\tilde{U}_{\{3\}} = U_{\{3\}}$  and eliminate the rest of the codebooks. From the independence relations above, the packing bounds would stay the same. Since we have merged codebooks, the covering bounds would loosen, and it is straightforward to see that the reconstructions at each decoder are still the same. We are left with four codebooks,  $C_{\{1\}}$ ,  $C_{\{2\}}$  and  $C_{\{3\}}$ . Note that since decoder  $\{1\}$  is only decoding  $C_{\{1\}}$  we must have  $\rho_{\{1\},1} = r_{\{1\}} = R_1$ . This is deduced from the packing bound in decoder  $\{1\}$ :

$$H(U_{\{1\}}) \leq H(U_{\{1\}}) + \rho_{\{1\},1} - r_{\{1\}} \Rightarrow r_{\{1\}} \leq \rho_{\{1\},1}$$

But  $\rho_{\{1\},1} \leq r_{\{1\}}$  so they are equal. The same argument gives  $\rho_{\{2\},2} = r_{\{2\}} = R_2$ , and  $\rho_{\{3\},3} = r_{\{3\}} = R_3$ . Also, from optimality at the joint decoders and lemma 19, we have  $U_i \perp\!\!\!\perp U_j, \forall i \neq j$ .

$$\begin{aligned} H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}|X, Z) &\geq H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{3\}}) - R_1 - R_2 - R_3 \\ &\Rightarrow I(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}; X, Z) + I(U_{\{3\}}; X, Z, U_{\{1\}}, U_{\{2\}}) \leq R_1 + R_2 + R_3 \\ &\Rightarrow I(U_{\{3\}}; X, Z, U_{\{1\}}, U_{\{2\}}) \leq R_3 \end{aligned} \quad (58)$$

Note that  $R_1 + R_3 = I(U_{\{1\}}, U_{\{3\}}; X, Z)$  and  $R_1 = I(U_{\{1\}}; X)$  from optimality at decoders  $\{1\}$  and  $\{1, 3\}$ . So  $R_3 = I(U_{\{3\}}; X, Z|U_{\{1\}})$ . Replacing  $R_3$  into (58), we get  $I(U_{\{3\}}; U_{\{2\}}|U_{\{1\}}, X, Z) = 0$ . So we have the Markov chain  $U_{\{3\}} \leftrightarrow U_{\{1\}}, X, Z \leftrightarrow U_{\{2\}}$ . By the same arguments we can derive the Markov chain  $U_{\{3\}} \leftrightarrow U_{\{2\}}, X, Z \leftrightarrow U_{\{1\}}$ . Using lemma 23 and the previous two Markov chains we get  $U_{\{3\}} \leftrightarrow X, Z \leftrightarrow U_{\{1\}}, U_{\{2\}}$ . Take the Markov chain  $U_{\{3\}} \leftrightarrow X, Z \leftrightarrow U_{\{1\}}$ , along with  $Z \perp\!\!\!\perp X, U_{\{1\}}$  we get  $U_{\{3\}}, Z \leftrightarrow X \leftrightarrow U_{\{1\}}$ . Also from the optimality of the reconstruction of  $X$  at decoders  $\{1\}$  and  $\{1, 3\}$ , we have:

$$I(U_{\{1\}}; X) = I(U_{\{1\}}, U_{\{3\}}; X) \Rightarrow I(U_{\{3\}}; X|U_{\{1\}}) = 0.$$

From the above and  $Z \perp\!\!\!\perp X, U_{\{1\}}$ , we conclude  $U_{\{3\}}, Z \leftrightarrow U_{\{1\}} \leftrightarrow X$ . Applying Lemma 23 we get  $Z, U_{\{3\}} \perp\!\!\!\perp X, U_{\{1\}}$ .

**Lemma 23.** Let  $A, B, C$  and  $D$  be RV's such that  $A \leftrightarrow B, C \leftrightarrow D$  and  $A \leftrightarrow B, D \leftrightarrow C$ , and also assume there is no  $b \in \mathcal{B}$  for which given  $B = b$  there are non-constant functions  $f_b(C)$  and  $g_b(D)$  with  $f_b(C) = g_b(D)$  with probability 1. Then  $A \leftrightarrow B \leftrightarrow C, D$ .

*Proof.* This lemma is a generalization of the one in [22]. We need to show that  $p(A = a|B = b, C = c, D = d) = p(A = a|B = b, C = c', D = d')$  for any  $a, b, c, c', d, d'$ . Note since functions  $f_b$  and  $g_b$  do not exist, it is straightforward to show that there is a finite sequence of pairs  $(c_i, d_i)$  such that  $(c_1, D_{\{1\}}) = (c, d)$  and  $(c_n, d_n) = (c', d')$  with the property that either  $c_i = c_{i+1}$  or  $d_i = d_{i+1}$  and that  $p(B = b, C = c_i, D = d_i) \neq 0$ . Then from the first Markov chain if  $d_i = d_{i+1}$ , we have  $p(A = a|B = b, C = c_i, D = d_i) = p(A = a|B = b, C = c_{i+1}, D = d_{i+1})$ , also if  $c_i = c_{i+1}$  the second Markov chain gives this result. So  $p(A = a|B = b, C = c_i, D = d_i)$  is constant on all of the sequence particularly  $p(A = a|B = b, C = c, D = d) = p(A = a|B = b, C = c', D = d')$ .  $\square$

Let  $g(U_{\{1\}}, U_{\{3\}})$  be the reconstruction of  $Z$  at decoder  $\{1, 3\}$ . We have:

$$\sum_{z, u_{\{1\}}, u_{\{3\}}} p(z, u_{\{1\}}, u_{\{3\}}) d_H(g(u_{\{1\}}, u_{\{3\}}), z) \leq \delta \Rightarrow \sum_{u_{\{1\}}} p(u_{\{1\}}) \sum_{z, u_{\{3\}}} p(z, u_{\{3\}}) d_H(g(u_{\{1\}}, u_{\{3\}}), z) \leq \delta$$

So there is at least one  $u_{\{1\}} \in \mathbf{U}_{\{1\}}$  such that  $\sum_{z, u_{\{3\}}} p(z, u_{\{3\}}) d_H(g(u_{\{1\}}, u_{\{3\}}), z) \leq \delta$ . Let  $g_{u_{\{1\}}}(U_{\{3\}}) = g(u_{\{1\}}, U_{\{3\}})$  be the reconstruction of  $Z$  using  $U_{\{3\}}$ . By the same argument we can find a reconstruction of  $X$  using  $U_{\{3\}}$ , then  $I(U_{\{3\}}; X, Z) \geq 2(1 - h_b(\delta))$  from a PtP perspective which is a contradiction.  $\square$

### B.3 Proof of Lemma 10

*Proof.* We provide an outline of the proof here, the arguments are similar to the ones in the previous proofs.

**Step 1:** Any codebook which is not decoded at decoders  $\{1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}$  and  $\{4\}$  is redundant. This implies that there are at most only 17 codebooks which are non-redundant. These codebooks are  $C_{\{1\}}, C_{\{1\}, \{2, 3\}}, C_{\{1\}, \{3, 4\}}, C_{\{1\}, \{4\}}, C_{\{1\}, \{2, 3\}, \{3, 4\}}, C_{\{1\}, \{4\}, \{2, 3\}}, C_{\{4\}}, C_{\{4\}, \{2, 3\}}, C_{\{4\}, \{1, 2\}}, C_{\{4\}, \{2, 3\}, \{1, 2\}}, C_{\{1, 2\}}, C_{\{2, 3\}}, C_{\{3, 4\}}, C_{\{1, 2\}, \{2, 3\}}, C_{\{1, 2\}, \{3, 4\}}, C_{\{2, 3\}, \{3, 4\}}$  and  $C_{\{1, 2\}, \{2, 3\}, \{3, 4\}}$ .

**Step 2:** In this step we prove that the only non-trivial codebook decoded at decoder  $\{i\}$  is  $C_{\{i\}}$  for  $i = 1, 4$ . All possible codebooks decoded at decoder  $\{1\}$  are  $C_{\{1\}}, C_{\{1\}, \{2, 3\}}, C_{\{1\}, \{3, 4\}}, C_{\{1\}, \{4\}}, C_{\{1\}, \{2, 3\}, \{3, 4\}}$  and  $C_{\{1\}, \{2, 3\}, \{4\}}$ . From optimality at decoder  $\{1, 2\}$ ,  $C_{\{1\}, \{2, 3\}}$  is redundant. The reason is  $\rho_{\{1\}, \{2, 3\}, 2} = 0$  otherwise we can set it to zero without any loss in distortion at decoder  $\{1, 2\}$  which contradicts optimality, also any random variable that description  $\{3\}$  carries must be used in reconstructing  $Z$  at decoder  $\{3, 4\}$  because that decoder is at optimality, which means  $\rho_{\{1\}, \{2, 3\}, 3} = 0$  so the codebook is decoded at decoder  $\{2, 3\}$  but not sent through either description  $\{2\}$  or  $\{3\}$ , from similar arguments as before the codebook is redundant. Same arguments can be provided to deduce redundancy of  $C_{\{1\}, \{3, 4\}}, C_{\{2, 3\}}, C_{\{1\}, \{2, 3\}, \{3, 4\}}$  and  $C_{\{1\}, \{2, 3\}, \{4\}}$ . This implies that only  $C_{\{1\}}$  is decoded at decoder  $\{1\}$  and  $C_{\{4\}}$  at decoder  $\{4\}$ .

**Step 3:** We proceed with eliminating  $C_{\{1, 2\}, \{3, 4\}}$  and  $C_{\{1, 2\}, \{2, 3\}, \{3, 4\}}$ . Using the PtP optimality of decoder  $\{1, 2\}$  we have:

$$\begin{aligned} I(U_{\{1\}}, U_{\{1, 2\}}, U_{\{1, 2\}, \{2, 3\}}, U_{\{1, 2\}, \{3, 4\}}, U_{\{1, 2\}, \{2, 3\}, \{3, 4\}}; X) = \\ R_1 + R_2 \stackrel{(a)}{\geq} I(U_{\{1\}}, U_{\{1, 2\}}, U_{\{1, 2\}, \{2, 3\}}, U_{\{1, 2\}, \{3, 4\}}, U_{\{1, 2\}, \{2, 3\}, \{3, 4\}}; X, Z) \end{aligned}$$

where (a) follows from the usual PtP source coding results. Comparing the LHS with the RHS we conclude the Markov chain  $U_{\{1\}}, U_{\{1, 2\}}, U_{\{1, 2\}, \{2, 3\}}, U_{\{1, 2\}, \{3, 4\}}, U_{\{1, 2\}, \{2, 3\}, \{3, 4\}} \leftrightarrow X \leftrightarrow Z$ . In particular we are interested in  $U_{\{1, 2\}, \{3, 4\}}, U_{\{1, 2\}, \{2, 3\}, \{3, 4\}} \leftrightarrow X \leftrightarrow Z$ . By the same arguments and using the optimality at decoder

$\{3, 4\}$ , we get  $U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \leftrightarrow Z \leftrightarrow X$ . These two Markov chains along with lemma 23 prove  $U_{\{1,2\},\{3,4\}}, U_{\{1,2\},\{2,3\},\{3,4\}} \perp\!\!\!\perp X, Z$ . So these two variables are not used in reconstructing the source and the corresponding codebooks are eliminated.

**Step 4:** The only remaining codebooks are  $C_{\{1\}}, C_{\{4\}}, C_{\{1,2\}}, C_{\{3,4\}}, C_{\{1,2\},\{2,3\}}$  and  $C_{\{2,3\},\{3,4\}}$ . From optimality at decoders  $\{1, 2\}$  and  $\{3, 4\}$  we must have  $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow (X, Z) \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$ , also  $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow X \leftrightarrow Z$  and  $X \leftrightarrow Z \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$ . From lemma 24, we get  $U_{\{1\}}, U_{\{1,2\}}, U_{\{1,2\},\{2,3\}} \leftrightarrow X \leftrightarrow Z \leftrightarrow U_{\{4\}}, U_{\{3,4\}}, U_{\{2,3\},\{3,4\}}$ .

**Lemma 24.** For random variables  $A, B, C, D$ , the three short Markov chains  $A \leftrightarrow (B, C) \leftrightarrow D$ ,  $A \leftrightarrow B \leftrightarrow C$  and  $B \leftrightarrow C \leftrightarrow D$  are equivalent to the long Markov chain  $A \leftrightarrow B \leftrightarrow C \leftrightarrow D$ .

*Proof.* We only need to show that  $A \leftrightarrow B \leftrightarrow D$ , the rest of the implications of the long Markov chain are either direct results of the three short Markov chains or follow by symmetry. For arbitrary  $a, b, d$  we have:

$$\begin{aligned} P(D = d|B = b, A = a) &= \sum_{c \in \mathcal{C}} P(C = c|B = b, A = a)P(D = d|A = a, B = b, C = c) \\ &= \sum_{c \in \mathcal{C}} P(C = c|B = b)P(D = d|B = b, C = c) = P(D = d|B = b) \end{aligned}$$

□

We get an inner bound for  $R_2 + R_3$  at decoder  $\{2, 3\}$ :

$$R_2 + R_3 \geq \min I(U, V; X, Z) = H(X, Z) = 1 + h_b(p),$$

where the minimum is taken over all  $P_{U,V|X,Z}$  for which the long Markov chain  $U \leftrightarrow X \leftrightarrow Z \leftrightarrow V$  is satisfied and  $(U, V)$  produce a lossless reconstruction of  $X + Z$ . This resembles the distributed source coding problem in [15]. So the RD vector can't be achieved using random codes. □

## B.4 Proof of Lemma 12

*Proof.* In this proof we use bold letters to denote vectors and matrices. Fix integers  $n, l, l'$  and  $k$ . Choose the elements of the matrices  $\Delta \mathbf{G}_{l \times n}$ ,  $\Delta \mathbf{G}'_{k' \times n}$  and  $\mathbf{G}_{k \times n}$  and vectors  $\mathbf{B}^n$  and  $\mathbf{B}^m$  randomly and uniformly from  $\mathbb{F}_q$ . The codebooks  $C_o^n$  and  $C_o^m$  are defined as follows:

$$\begin{aligned} C_o &= \{\mathbf{a}\mathbf{G} + \mathbf{m}\Delta \mathbf{G} + \mathbf{B} | \mathbf{a} \in \mathbb{F}_q^k, \mathbf{m} \in \mathbb{F}_q^l\} \\ C'_o &= \{\mathbf{b}\mathbf{G} + \mathbf{m}'\Delta \mathbf{G}' + \mathbf{B}' | \mathbf{b} \in \mathbb{F}_q^k, \mathbf{m}' \in \mathbb{F}_q^{l'}\} \end{aligned}$$

For a typical sequence  $\mathbf{x}$  with respect to  $P_X$ , we define  $\theta(\mathbf{x})$  as the function which counts the number of codewords in  $C_o$  and  $C'_o$  jointly typical with respect to  $P_{XUV}$ :

$$\begin{aligned} \theta(\mathbf{x}) &= \sum_{\mathbf{u} \in C_o, \mathbf{v} \in C'_o} \mathbb{I}\{(\mathbf{u}, \mathbf{v}) \in A_\epsilon^n(U, V|\mathbf{x})\} \\ &= \sum_{\mathbf{m}, \mathbf{m}'} \sum_{\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^k} \sum_{(\mathbf{u}, \mathbf{v}) \in A_\epsilon^n(U, V|\mathbf{x})} \mathbb{I}\{\mathbf{a}\mathbf{G} + \mathbf{m}\Delta \mathbf{G} + \mathbf{B} = \mathbf{u}, \mathbf{b}\mathbf{G} + \mathbf{m}'\Delta \mathbf{G}' + \mathbf{B}' = \mathbf{v}\} \end{aligned}$$

Our goal is to find bounds on  $n, l, l'$  and  $k$  such that  $P(\theta(\mathbf{x}) = 0) \rightarrow 0$  as  $n \rightarrow \infty$ .

For  $\mathbf{a} \in \mathbb{F}_q^k$  and  $\mathbf{m} \in \mathbb{F}_q^l$ , we denote the corresponding codeword as  $g(\mathbf{a}, \mathbf{m}) := \mathbf{a}\mathbf{G} + \mathbf{m}\Delta \mathbf{G} + \mathbf{B}$ . Similarly define  $g'(\mathbf{b}, \mathbf{m}') := \mathbf{b}\mathbf{G} + \mathbf{m}'\Delta \mathbf{G}' + \mathbf{B}'$  for any  $\mathbf{b} \in \mathbb{F}_q^k$  and  $\mathbf{m}' \in \mathbb{F}_q^{l'}$ . The following lemma proves several results on the pairwise independence of the codewords.

**Lemma 25.** *The following hold:*

1.  $g(\mathbf{a}, \mathbf{m})$  and  $g'(\mathbf{b}, \mathbf{m}')$  are distributed uniformly over  $\mathbb{F}_q^n$ .
2. If  $\mathbf{a} \neq \tilde{\mathbf{a}}$ , then  $g(\mathbf{a}, \mathbf{m})$  is independent of  $g(\tilde{\mathbf{a}}, \mathbf{m})$ .
3. If  $\mathbf{b} \neq \tilde{\mathbf{b}}$ , then  $g'(\mathbf{b}, \mathbf{m}')$  is independent of  $g'(\tilde{\mathbf{b}}, \mathbf{m}')$ .
4. If  $\mathbf{B}$  and  $\mathbf{B}'$  are chosen independently and uniformly over  $\mathbb{F}_q^n$ , then  $g(\mathbf{b}, \mathbf{m}')$  and  $g'(\mathbf{a}, \mathbf{m})$  are independent.

*Proof.* Follows from [17], and the fact that  $\mathbf{B}, \mathbf{B}'$  are independent and uniform.  $\square$

We intend to use Chebyshev's inequality to obtain:

$$P\{\theta(\mathbf{X}) = 0\} \leq \frac{4\text{var}\{\theta(\mathbf{X})\}}{\mathbb{E}\{\theta(\mathbf{X})\}^2} \rightarrow 0$$

**Lemma 26.** *For  $\mathbf{X} \in A_\epsilon^{(n)}(X)$  we have the following bound on  $\frac{\text{var}\{\theta(\mathbf{X})\}}{\mathbb{E}\{\theta(\mathbf{X})\}^2}$ :*

$$\begin{aligned} \frac{\text{var}\{\theta(\mathbf{X})\}}{\mathbb{E}\{\theta(\mathbf{X})\}^2} &\leq \frac{q^{2n}}{q^{l+l'}q^{2k}} 2^{-n(H(U,V|X))} + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(U|X))} + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(V|X))} + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(U,V|X) - \max_{i \neq 0} H(U,V|X, V+iU))} \\ &\quad + \frac{q^n}{q^l q^k} 2^{-n(H(U|X))} + \frac{q^n}{q^l q^k} 2^{-n(H(V|X))} + \frac{1}{q^l} + \frac{1}{q^{l'}} + \frac{1}{q^{l+l'}} + \frac{1}{q^{l+k}} + \frac{1}{q^{l'+k}} \end{aligned}$$

*Proof.* We calculate the expected value of  $\theta(\mathbf{X})$  for any  $\mathbf{X} \in A_\epsilon^{(n)}(X)$ :

$$\begin{aligned} \mathbb{E}\{\theta(\mathbf{X})\} &= \sum_{\mathbf{x} \in A_\epsilon^{(n)}(X)} \sum_{\substack{\mathbf{m} \in \mathbb{F}_q^l \\ \mathbf{m}' \in \mathbb{F}_q^{l'}}} \sum_{\mathbf{a} \neq \mathbf{b}} \sum_{(\mathbf{u}, \mathbf{v}) \in A_\epsilon^n(U, V|\mathbf{x})} P(\mathbf{x}) P\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g'(\mathbf{b}, \mathbf{m}') = \mathbf{v}\} \\ &= \sum_{\mathbf{m}, \mathbf{m}'} \sum_{\mathbf{x} \in A(X)} \sum_{\mathbf{a} \neq \mathbf{b}} |A_\epsilon^n(U, V|\mathbf{x})| P(\mathbf{x}) \frac{1}{q^{2n}} = \frac{q^{l+l'} q^{2k}}{q^{2n}} 2^{n(H(U, V|X) + O(\epsilon))} \end{aligned}$$

Also:

$$\mathbb{E}\{\theta(\mathbf{X})^2\} = \sum_{\substack{\mathbf{m}, \tilde{\mathbf{m}} \\ \mathbf{m}', \tilde{\mathbf{m}}'}} \sum_{\mathbf{a}, \tilde{\mathbf{a}}} \sum_{\mathbf{b}, \tilde{\mathbf{b}}} \sum_{(\mathbf{u}, \mathbf{v})} \sum_{(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \in A_\epsilon^n(U, V|\mathbf{x})} P\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}}, \tilde{\mathbf{m}}) = \tilde{\mathbf{u}}, g'(\mathbf{b}, \mathbf{m}') = \mathbf{v}, g'(\tilde{\mathbf{b}}, \tilde{\mathbf{m}}') = \tilde{\mathbf{v}}\}$$

(59)

Using Lemma 25:

$$\begin{aligned} P_S &\triangleq P\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}}, \tilde{\mathbf{m}}) = \tilde{\mathbf{u}}, g(\mathbf{b}, \mathbf{m}') = \mathbf{v}, g(\tilde{\mathbf{b}}, \tilde{\mathbf{m}}') = \tilde{\mathbf{v}}\} \\ &= \frac{1}{q^{2n}} \times P\{g_0(\mathbf{a} - \tilde{\mathbf{a}}, \mathbf{m} - \tilde{\mathbf{m}}) = \mathbf{u} - \tilde{\mathbf{u}}, g'_0(\mathbf{b} - \tilde{\mathbf{b}}, \mathbf{m}' - \tilde{\mathbf{m}}') = \mathbf{v} - \tilde{\mathbf{v}}\} \end{aligned}$$

At this point we have to consider several different cases for the values of  $\mathbf{a}, \tilde{\mathbf{a}}, \mathbf{b}, \tilde{\mathbf{b}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{m}', \tilde{\mathbf{m}}'$ .

- 1)  $\mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'$

$$1.1: \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{2n}} \delta(\mathbf{u} - \tilde{\mathbf{u}}) \delta(\mathbf{v} - \tilde{\mathbf{v}})$$

$$1.2: \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{u} - \tilde{\mathbf{u}})$$

$$1.3: \mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{v} - \tilde{\mathbf{v}})$$

$$1.4: \mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \sum_{\alpha \in \mathbb{F}_q} \frac{1}{q^{3n}} \delta(\mathbf{u} - \tilde{\mathbf{u}} - \alpha(\mathbf{v} - \tilde{\mathbf{v}})) + \frac{1}{q^{4n}} (1 - \sum_{\alpha \in \mathbb{F}_q} \delta(\mathbf{u} - \tilde{\mathbf{u}} - \alpha(\mathbf{v} - \tilde{\mathbf{v}})))$$

$$2) \quad \mathbf{m} \neq \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'$$

$$2.1: \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{v} - \tilde{\mathbf{v}})$$

$$2.2: \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{4n}}$$

$$2.3: \mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{3n}} \delta(\mathbf{v} - \tilde{\mathbf{v}})$$

$$2.4: \mathbf{a} \neq \tilde{\mathbf{a}}, \mathbf{b} \neq \tilde{\mathbf{b}} \Rightarrow P_s = \frac{1}{q^{4n}}$$

Cases when  $\mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' \neq \tilde{\mathbf{m}}'$  and  $\mathbf{m} \neq \tilde{\mathbf{m}}, \mathbf{m}' \neq \tilde{\mathbf{m}}'$  are similarly considered but the derivations are omitted for brevity. Considering cases 1.1–4:

$$\mathbb{E}\{\theta(\mathbf{x})^2 | \mathbf{m} = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'\} = \sum_{\mathbf{m}, \mathbf{m}'} \left[ \sum_{\mathbf{a}=\tilde{\mathbf{a}}} \sum_{\mathbf{b}=\tilde{\mathbf{b}}} \sum_{(\mathbf{u}, \mathbf{v}) \in A_q^n(U, V | \mathbf{x})} \frac{1}{q^{2n}} + \sum_{\mathbf{a}=\tilde{\mathbf{a}}} \sum_{\mathbf{b} \neq \tilde{\mathbf{b}}} \sum_{(\mathbf{u}, \mathbf{v}), (\tilde{\mathbf{u}}, \tilde{\mathbf{v}})} \frac{1}{q^{3n}} \right. \quad (60)$$

$$+ \sum_{\mathbf{a} \neq \tilde{\mathbf{a}}} \sum_{\mathbf{b}=\tilde{\mathbf{b}}} \sum_{(\mathbf{u}, \mathbf{v}), (\tilde{\mathbf{u}}, \tilde{\mathbf{v}})} \frac{1}{q^{3n}} + \sum_{\alpha \in \mathbb{F}_q \setminus \{0\}} \sum_{\mathbf{a} \neq \tilde{\mathbf{a}}} \sum_{\substack{\mathbf{b} \neq \tilde{\mathbf{b}} \\ \mathbf{b} - \tilde{\mathbf{b}} = \alpha(\mathbf{a} - \tilde{\mathbf{a}})}} \sum_{\substack{(\mathbf{u}, \mathbf{v}), (\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \\ \mathbf{v} - \tilde{\mathbf{v}} = \alpha(\mathbf{u} - \tilde{\mathbf{u}})}} \frac{1}{q^{3n}} \quad (61)$$

$$\left. + \sum_{\alpha \in \mathbb{F}_q \setminus \{0\}} \sum_{\mathbf{a} \neq \tilde{\mathbf{a}}} \sum_{\substack{\mathbf{b} \neq \tilde{\mathbf{b}} \\ \mathbf{b} - \tilde{\mathbf{b}} \neq \alpha(\mathbf{a} - \tilde{\mathbf{a}})}} \sum_{\substack{(\mathbf{u}, \mathbf{v}), (\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \\ \mathbf{v} - \tilde{\mathbf{v}} \neq \alpha(\mathbf{u} - \tilde{\mathbf{u}})}} \frac{1}{q^{4n}} \right] \quad (62)$$

Consequently:

$$\begin{aligned} \mathbb{E}\{\theta(\mathbf{X})^2 | \mathbf{m}' = \tilde{\mathbf{m}}, \mathbf{m}' = \tilde{\mathbf{m}}'\} &\leq \frac{q^{l+l'} q^{2k}}{q^{2n}} 2^{n(H(U, V | X))} + \frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U, V | X) + H(V | X, U))} + \frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U, V | X) + H(U | X, V))} + \\ &\frac{q^{l+l'} q^{3k}}{q^{3n}} 2^{n(H(U, V | X) + \max_{\alpha \neq 0} H(U, V | X, V + \alpha U))} + \frac{q^{l+l'} q^{4k}}{q^{4n}} 2^{2n(H(U, V | X))}, \end{aligned}$$

where we have used Lemma 8 in [29] to get the fourth term. After considering all the cases, the only non-redundant bounds are the ones mentioned in the lemma.  $\square$

So, the following bounds need to be satisfied:

$$\begin{aligned} r_o + r'_o &\geq 2 \log q - H(U, V | X) \\ r_o + r'_o - r_i &\geq \log q - \min\{H(U | X), H(V | Z)\} \\ r_o + r'_o - r_i &\geq \log q - H(U, V | X) + \max_{\alpha \neq 0} H(U, V | X, V + \alpha U) \\ r_o &\geq \log q - H(U | X) \\ r'_o &\geq \log q - H(V | X) \\ \min\{r_o, r'_o\} &\geq r_i \end{aligned}$$

Observe that

$$H(U, V|X, V + \alpha U) = H(U, V, V + \alpha U|X) - H(V + \alpha U|X) = H(U, V|X) - H(V + \alpha U|X)$$

□

## B.5 Proof of Lemma 13

*Proof.* The proof follows the same arguments as that of Lemma 12. We provide an outline of the proof. Define the probability of error  $P_e$  as follows:

$$P_e = P(\{(\mathbf{x}, \mathbf{u}, \mathbf{v}) \in \mathcal{X}^n \times C_1 \times C_2 | \exists(\mathbf{u}', \mathbf{v}') \in A_\epsilon^n(U, V) \cap B_2(\mathbf{u}) \times B_2(\mathbf{v})\})$$

We define a new conditional probability of error for any triple  $\mathbf{x}, \mathbf{u}, \mathbf{v} \in A_\epsilon(X, U, V)$ :

$$P_{e|\mathbf{x}, \mathbf{u}, \mathbf{v}} = P(\exists(\mathbf{u}', \mathbf{v}') \in A_\epsilon^n(U, V) \cap B_2(\mathbf{u}) \times B_2(\mathbf{v}) | \mathbf{X} = \mathbf{x}, (\mathbf{u}, \mathbf{v}) \in C_1 \times C_2)$$

Clearly if  $P_{e|\mathbf{x}, \mathbf{u}, \mathbf{v}}$  goes to 0 for all  $\mathbf{x}, \mathbf{u}, \mathbf{v} \in A_\epsilon(X, U, V)$  as  $n \rightarrow \infty$ , then  $P_e$  goes to 0. Also define:  $P_{\mathbf{x}, \mathbf{u}, \mathbf{v}} = P(\{(\mathbf{x}, \mathbf{u}, \mathbf{v}) \in \mathcal{X}^n \times C_1 \times C_2\})$ , and  $P_{e, \mathbf{x}, \mathbf{u}, \mathbf{v}} = P_{e|\mathbf{x}, \mathbf{u}, \mathbf{v}} P_{\mathbf{x}, \mathbf{u}, \mathbf{v}}$ . We have:

$$\begin{aligned} P_{\mathbf{x}, \mathbf{u}, \mathbf{v}} &= \sum_{\mathbf{x} \in A_\epsilon^n(X)} \sum_{\substack{\mathbf{m} \in \mathbb{F}_q^l \\ \mathbf{m}' \in \mathbb{F}_q^{l'}}} \sum_{\mathbf{a} \neq \mathbf{b}} \sum_{(\mathbf{u}, \mathbf{v}) \in A_\epsilon^n(U, V|\mathbf{x})} P(\mathbf{x}) P\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g'(\mathbf{b}, \mathbf{m}') = \mathbf{v}\} \\ &= \sum_{\mathbf{m}, \mathbf{m}'} \sum_{\mathbf{x} \in A(X)} \sum_{\mathbf{a} \neq \mathbf{b}} |A_\epsilon^n(U, V|\mathbf{x})| P(\mathbf{x}) \frac{1}{q^{2n}} = \frac{q^{l+l'} q^{2k}}{q^{2n}} 2^{n(H(U, V|X) + O(\epsilon))} \end{aligned}$$

$$P_{e, \mathbf{x}, \mathbf{u}, \mathbf{v}} = \sum_{\substack{\mathbf{m}, \tilde{\mathbf{m}} \\ \mathbf{m}', \tilde{\mathbf{m}}'}} \sum_{\mathbf{x}} \sum_{\mathbf{a}, \tilde{\mathbf{a}}} \sum_{\mathbf{b}, \tilde{\mathbf{b}}} \sum_{\substack{(\mathbf{u}, \mathbf{v}) \in \\ A_\epsilon^n(U, V|\mathbf{x})}} \sum_{\substack{(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \in \\ A_\epsilon^n(U, V)}} \sum_{b_1 \in [1, 2^{n\rho_1}]} \sum_{b_2 \in [1, 2^{n\rho_2}]} P(\mathbf{x})$$

$$P\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g(\tilde{\mathbf{a}}, \tilde{\mathbf{m}}) = \tilde{\mathbf{u}}, g'(\mathbf{b}, \mathbf{m}') = \mathbf{v}, g'(\tilde{\mathbf{b}}, \tilde{\mathbf{m}}') = \tilde{\mathbf{v}}\} P\{B_1(\mathbf{u}) = B_1(\tilde{\mathbf{u}}) = b_1, B_2(\mathbf{u}) = B_2(\tilde{\mathbf{u}}) = b_2\}$$

Note that the binning is done independently and uniformly, so  $P\{B_1(\mathbf{u}) = B_1(\tilde{\mathbf{u}}) = b_1, B_2(\mathbf{u}) = B_2(\tilde{\mathbf{u}}) = b_2\} = 2^{-2(\rho_1 + \rho_2)}$ . The rest of the summations are the ones which were present in the proof of Lemma 12. Again we have to do a case by case investigation of the summation. The only new bond comes from the case when  $\mathbf{m} = \tilde{\mathbf{m}}$  and  $\mathbf{m}' = \tilde{\mathbf{m}}'$ ,  $a \neq \tilde{a}$ ,  $b \neq \tilde{b}$  and  $a - \tilde{a} = i(b - \tilde{b})$ . We have:

$$\begin{aligned} A &= \sum_{\mathbf{m}, \tilde{\mathbf{m}}} \sum_{\substack{\mathbf{a}, \tilde{\mathbf{a}} \\ \mathbf{a} \neq \tilde{\mathbf{a}}}} \sum_{\substack{\mathbf{b}, \tilde{\mathbf{b}} \\ \mathbf{a} - \tilde{\mathbf{a}} = i(\mathbf{b} - \tilde{\mathbf{b}})}} \sum_{\substack{(\mathbf{u}, \mathbf{v}) \in \\ A_\epsilon^n(U, V|\mathbf{x})}} \sum_{\substack{(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \in \\ A_\epsilon^n(U, V) \\ \mathbf{u} - \tilde{\mathbf{u}} = i(\mathbf{v} - \tilde{\mathbf{v}})}} q^{-3n} 2^{-n(\rho_1 + \rho_2)} \\ &= \frac{q^{l+l'}}{q^{3n}} q^{3k} 2^{nH(U, V|X)} 2^{nH(U, V|U+iV)} 2^{-n(\rho_1 + \rho_2)} \end{aligned}$$

Dividing this last term by  $P_{\mathbf{x}, \mathbf{u}, \mathbf{v}}$ :

$$\frac{A}{P_{\mathbf{x}, \mathbf{u}, \mathbf{v}}} = \frac{q^k}{q^n} 2^{nH(U, V|U+iV)} 2^{-n(\rho_1 + \rho_2)}$$

which goes to 0 if the following is satisfied:

$$r_i - \rho_1 - \rho_2 \leq \log q - H(U, V|U + iV) \quad (63)$$

However as shown in the next lemma the new bound in (63) is redundant.

**Lemma 27.** *The inequality (63) in Lemma 13 is redundant.*

*Proof.* Assume there is a distribution  $P_{X,U,V}$  for which (63) is violated, we show that either (17) or (20) is also violated. Conversely, as long as (17) and (20) are satisfied, (63) is also satisfied. Assume we have:

$$\begin{aligned} (r_o - \rho_1) + (r'_o - \rho_2) &\leq 2 \log q - H(U, V) \\ r_i - \rho_1 - \rho_2 &> \log q - H(U, V|U + iV), \forall i \in \mathbb{F}_q. \end{aligned}$$

Adding the two bounds we get:

$$\begin{aligned} r_o + r'_o - r_i &< \log q - H(U, V) + H(U, V|U + iV) \\ &= \log q - H(U + iV) \leq \log q - H(U + iV|X) \end{aligned}$$

which contradicts (17). □

□

□

## B.6 Proof of Lemma 14

*Proof.* The proof follows the same arguments as in the previous two examples. First we assume there exists a joint distribution  $P_{UX}$  such that the SSC scheme achieves the RD vector, then we arrive at a contradiction by eliminating all codebooks. First note that from our definition of  $P_{X,V_{\{1\}},V_{\{2\}}}$ , direct calculation shows that  $R_1 + R_2 = I(V_{\{1\}}, V_{\{2\}}; X) = 1 - h_b(D_0)$ . This means that decoder  $\{1, 2\}$  is at PtP optimality. Also by the definition of the distortion function  $D_{\{3\}}$ , decoder  $\{3\}$  is at optimal RD.

**Step 1:** From the optimality of decoder  $\{1, 2\}$  and Lemma 19, there can't be any codebook common between decoders  $\{1\}$  and  $\{2\}$ . So  $C_{\{1\},\{2\}}$  and  $C_{\{1\},\{2\},\{3\}}$  are eliminated.

**Step 2:** From optimality of decoder  $\{3\}$ , description 3 can't carry the bin number of any codebook which is not decoded at that decoder. Also description 1 and 2 can't carry the bin numbers of codebooks which are not decoded at  $\{1, 2\}$  because of optimality at this decoder. So codebooks  $C_{\{1,3\},\{2,3\}}$ ,  $C_{\{1,3\}}$  and  $C_{\{2,3\}}$  are not sent on any description and are redundant.

**Step 3:** The codebook  $C_{\{1\},\{2,3\}}$  is not binned by description 2 or 3. Description 3 can't bin the codebook since it is not decoded at decoder  $\{3\}$ , and that decoder is at PtP optimality. Note  $C_{\{1\},\{2,3\}}$  can be decoded using description 1, so any bin information for this codebook that is carried by description 2 is not used at decoder  $\{1, 2\}$ , since decoder  $\{1, 2\}$  is at PtP optimality we must have  $\rho_{\{1\},\{2,3\},2} = 0$ . The codebook is not sent on description 2 or 3, so by the same arguments as in the previous proofs it can't help in the reconstruction at decoder  $\{2, 3\}$  and is redundant. By the same arguments  $C_{\{2\},\{1,3\}}$  is redundant.

**Step 4:** In this step we show that there is no refinement codebook decoded at decoder  $\{1, 2\}$ . This would eliminate  $C_{\{1,2\}}, C_{\{1,2\},\{3\}}, C_{\{1,2\},\{1,3\}}, C_{\{1,2\},\{2,3\}}$  and  $C_{\{1,2\},\{1,3\},\{2,3\}}$ . More precisely we show that the reconstruction at decoder  $\{1, 2\}$  is a function of the reconstructions at decoders  $\{1\}$  and  $\{2\}$ . This means that sending a refinement codebook to decoder  $\{1, 2\}$  will not help in the reconstruction, so the codebook is redundant.

To prove this claim we consider the two user example depicted in Figure [6]. Here all distortions are Hamming distortions. We are interested in achieving the rate distortion vector  $(R_1, R_2, D_{\{1\}}, D_{\{2\}}, D_{\{1,2\}})$  given in (10). Let  $P_{X,U_{\{1,2\}},U_{\{1\}},U_{\{2\}},U_{\{1\},\{2\}}}$  be a distribution on the random variables in the two user SSC achieving this RD vector. Define  $\hat{X}_1, \hat{X}_2$  and  $\hat{X}_{12}$  as the reconstructions at the corresponding codebooks.



**Lemma 28.** *There are only two choices for the joint distribution  $P_{X, \hat{X}_1, \hat{X}_2, \hat{X}_{12}}$ , furthermore in both choices,  $\hat{X}_{12}$  is a function of  $\hat{X}_1$  and  $\hat{X}_2$ .*

*Proof.* As in step 1, from optimality of decoder  $\{1, 2\}$ ,  $C_{\{1,2\}}$  is redundant. Also  $U_{\{1\}}$  and  $U_{\{2\}}$  are independent from Lemma 19. Note that  $\hat{X}_1$  is a function of  $U_{\{1\}}$  and  $\hat{X}_2$  is a function of  $U_{\{2\}}$ , so  $\hat{X}_1 \perp \hat{X}_2$ . We proceed by characterizing  $P_{X, \hat{X}_{12}}$ . Note that decoder  $\{1, 2\}$  is at PtP optimality. It is well-known result that when quantizing a BSS to Hamming distortion  $D_0$  with rate  $1 - h_b(D_0)$ , the reconstruction is uniquely given by  $\hat{X}_{12} = X + N_0$ ,  $N_0 \sim Be(D_0)$  where  $N_0 \perp X$ .  $\hat{X}_1, \hat{X}_2$  and  $\hat{X}_{12}$  are available at decoder  $\{1, 2\}$ , from optimality at this decoder we must have:

$$1 - h_b(D_0) = I(\hat{X}_1, \hat{X}_2, \hat{X}_{12}; X) \geq I(\hat{X}_{12}, X) = 1 - h_b(D_0).$$

So the inequality must be equality, which means  $I(\hat{X}_1, \hat{X}_2; X | \hat{X}_{12}) = 0$ . In other words the Markov chain  $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$  must hold. Using the three facts 1)  $\hat{X}_{12} = X \oplus N_0$ , 2)  $\hat{X}_1 \perp \hat{X}_2$  and 3)  $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$ , we can characterize all possible distributions on  $P_{X, \hat{X}_{12}, \hat{X}_1, \hat{X}_2}$ . Let  $\hat{X}_1 \sim Be(a_1)$  and  $\hat{X}_2 \sim Be(a_2)$ . Then from  $\hat{X}_1 \perp \hat{X}_2$ ,  $P_{\hat{X}_1, \hat{X}_2}$  is fixed. Assume the distribution  $P_{\hat{X}_{12}, \hat{X}_1, \hat{X}_2}$  is as given below: As shown on the table there

$\hat{X}_{12} \backslash \hat{X}_1, \hat{X}_2$	00	01	10	11	Sum
0	$P_{000}$	$P_{001}$	$P_{010}$	$P_{011}$	$\frac{1}{2}$
1	$P_{100}$	$P_{101}$	$P_{110}$	$P_{111}$	$\frac{1}{2}$
Sum	$(1 - a_1)(1 - a_2)$	$(1 - a_1)a_2$	$a_1(1 - a_2)$	$a_1a_2$	

Table 2

are 5 independent linear constraints on  $P_{ijk}$ 's. We have:

$$\begin{aligned} P_{011} &= \frac{1}{2} - P_{000} - P_{001} - P_{010}, & P_{100} &= (1 - a_1)(1 - a_2) - P_{000}, & P_{101} &= (1 - a_1)a_2 - P_{001}, \\ P_{110} &= (1 - a_1)a_2 - P_{010}, & P_{111} &= a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010} \\ a_1 &\in [0, 1], a_2 \in [0, 1], P_{000} \in [0, (1 - a_1)(1 - a_2)], P_{001} \in [0, (1 - a_1)a_2], P_{010} \in [0, a_1(1 - a_2)] \\ P_{000} + P_{001} + P_{010} &\in \left[\frac{1}{2} - a_1a_2, \frac{1}{2}\right] \end{aligned}$$

Using the Markov chain  $\hat{X}_1, \hat{X}_2 \leftrightarrow \hat{X}_{12} \leftrightarrow X$ , we have  $P_{X, \hat{X}_1, \hat{X}_2} = \sum_{\hat{x}_{12}} P_{X | \hat{x}_{12}} P_{\hat{X}_1, \hat{X}_2, \hat{x}_{12}}$ . So  $P_{X, \hat{X}_1, \hat{X}_2}$  is as follows: We can minimize the resulting distortion at decoders 1 and 2 by choosing  $P_{000}, P_{001}$  and  $P_{010}$  optimally. Let  $P_{X, \hat{X}_1, \hat{X}_2}^*$  be the optimal joint distribution, we will show that there are two choices for  $P_{X, \hat{X}_1, \hat{X}_2}^*$ . We have:

$$\begin{aligned} \mathbb{E}(d_H(\hat{X}_1, X)) + \mathbb{E}(d_H(\hat{X}_2, X)) &= P(\hat{X}_1 \neq X) + P(\hat{X}_2 \neq X) \\ &= (P_{X, \hat{X}_1, \hat{X}_2}(0, 0, 1) + P_{X, \hat{X}_1, \hat{X}_2}(1, 0, 1)) + (P_{X, \hat{X}_1, \hat{X}_2}(0, 1, 0) + P_{X, \hat{X}_1, \hat{X}_2}(1, 1, 0)) + 2(P_{X, \hat{X}_1, \hat{X}_2}(0, 1, 1) + P_{X, \hat{X}_1, \hat{X}_2}(1, 0, 0)) \\ &= P_{001} + (1 - a_1)a_2 - P_{001} + P_{010} + (1 - a_2)a_1 - P_{010} + 2D_0(P_{000} + a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010}) \\ &\quad + 2(1 - D_0)(\frac{1}{2} - P_{000} - P_{001} - P_{010} + (1 - a_1)(1 - a_2) - P_{000}) \\ &= (2D_0 - 1)a_1 + (2D_0 - 1)a_2 + 4(2D_0 - 1)P_{000} + 2(2D_0 - 1)P_{001} + 2(2D_0 - 1)P_{010} - 4D_0 + 3. \end{aligned}$$

$\hat{X}_1, \hat{X}_2 \backslash X$	00	01	10	11
0	$(1 - D_0)P_{000} + D_0((1 - a_1)(1 - a_2) - P_{000})$	$(1 - D_0)P_{001} + D_0((1 - a_1)a_2 - P_{001})$	$(1 - D_0)P_{010} + D_0(a_1(1 - a_2) - P_{010})$	$(1 - D_0)(\frac{1}{2} - P_{000} - P_{001} - P_{010}) + D_0(a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010})$
1	$D_0P_{000} + (1 - D_0)((1 - a_1)(1 - a_2) - P_{000})$	$D_0P_{001} + (1 - D_0)((1 - a_1)a_2 - P_{001})$	$D_0P_{010} + (1 - D_0)(a_1(1 - a_2) - P_{010})$	$D_0(\frac{1}{2} - P_{000} - P_{001} - P_{010}) + (1 - D_0)(a_1a_2 - \frac{1}{2} + P_{000} + P_{001} + P_{010})$

Table 3

This is an optimization problem on  $a_1, a_2, P_{000}, P_{001}, P_{010}$  with respect to the constraints:

$$a_1 \in [0, 1], a_2 \in [0, 1], P_{000} \in [0, (1 - a_1)(1 - a_2)], P_{001} \in [0, (1 - a_1)a_2], P_{010} \in [0, a_1(1 - a_2)]$$

$$P_{000} + P_{001} + P_{010} \in [\frac{1}{2} - a_1a_2, \frac{1}{2}].$$

Also note that for fixed  $a_1$  and  $a_2$  the problem becomes a linear optimization problem (otherwise the constraints are not linear). So we fix  $a_1$  and  $a_2$  and optimize  $P_{000}, P_{001}$  and  $P_{010}$  for each value of  $a_1$  and  $a_2$ . In this case the simplex algorithm provides a straightforward solution. We investigate the solution in several different cases:

**Case 1:**  $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$ : Note that in the simplex algorithm, the variable with smallest (most negative) coefficient takes its maximum possible value first. Since  $D_0 < \frac{1}{2}$ ,  $(2D_0 - 1) < 0$ , so the algorithm would first maximize the value of  $P_{000}$ . Since  $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$ , we have  $P_{000}^* = \frac{1}{2}$ . This along with constraint  $P_{000} + P_{001} + P_{010} \in [\frac{1}{2} - a_1a_2, \frac{1}{2}]$  sets  $P_{001}^* = 0$  and  $P_{010}^* = 0$ . So in this case:

$$\begin{aligned} \mathbb{E}(d_H(\hat{X}_1, X)) + \mathbb{E}(d_H(\hat{X}_2, X)) &= (2D_0 - 1)a_1 + (2D_0 - 1)a_2 + 2(2D_0 - 1) - 4D_0 + 3 \\ &= 1 + (2D_0 - 1)(a_1 + a_2). \end{aligned}$$

Now we optimize on  $a_1, a_2$  such that  $(1 - a_1)(1 - a_2) \geq \frac{1}{2}$ . Increasing  $a_1$  or  $a_2$  decreases the distortion so the optimal value is achieved when  $(1 - a_1)(1 - a_2) = \frac{1}{2}$ , so  $a_2 = 1 - \frac{1}{2(1 - a_1)}$ . We have:

$$\mathbb{E}(d_H(\hat{X}_1, X)) + \mathbb{E}(d_H(\hat{X}_2, X)) = 1 + (2D_0 - 1)(a_1 + 1 - \frac{1}{2(1 - a_1)})$$

Optimizing the value of  $a_1$ , we get  $a_1^* = a_2^* = 1 - \frac{\sqrt{2}}{2}$ . These values give  $P_{X, \hat{X}_1, \hat{X}_2} = P_{X, V_{(1)}, V_{(2)}}$ . Also replacing the values in  $P_{\hat{X}_{12}, \hat{X}_1, \hat{X}_2}$ , we get: which shows that  $\hat{X}_{12}$  is a function of  $\hat{X}_1$  and  $\hat{X}_2$ . **Case 2:**  $(1 - a_1)(1 - a_2) < \frac{1}{2}$ ,

$\hat{X}_{12} \backslash \hat{X}_1, \hat{X}_2$	00	01	10	11
0	$\frac{1}{2}$	0	0	0
1	0	$\frac{\sqrt{2}-1}{2}$	$\frac{\sqrt{2}-1}{2}$	$\frac{3-2\sqrt{2}}{2}$

Table 4

$a_1 \leq \frac{1}{2}$ : In this case the simplex method yields the following set of optimal distributions:

$$P_{000}^* = (1 - a_1)(1 - a_2), P_{001}^* = \alpha, P_{010}^* = \frac{1}{2} - (1 - a_1)(1 - a_2) - \alpha, P_{011}^* = 0$$

$$P_{100}^* = 0, P_{101}^* = (1 - a_1)a_2 - \alpha, P_{010}^* = (1 - a_2)a_1 - \frac{1}{2} + (1 - a_1)(1 - a_2) + \alpha, P_{111}^* = a_1a_2.$$

Where  $\alpha \in [a_2 - \frac{1}{2}, \frac{1}{2} - (1 - a_1)(1 - a_2)]$  is an auxiliary variable that does not play a role in the distortion since the coefficients of  $P_{001}^*$  and  $P_{010}^*$  are equal in the distortion formula. We get:

$$E(d_H(\hat{X}_1, X)) + E(d_H(\hat{X}_2, X)) = 1 + (2D_0 - 1)((1 - a_1)(1 - a_2) + a_1a_2).$$

Note that since  $a_1 < \frac{1}{2}$ , the term  $(1 - a_1)(1 - a_2) + a_1a_2$  is decreasing with  $a_2$ , so the distortion is increasing with  $a_2$  and the optimal values are  $a_2^* = \max(0, 1 - \frac{1}{2(1-a_1)})$ , since  $a_1 \leq \frac{1}{2}$ ,  $a_2^* = 1 - \frac{1}{2(1-a_1)}$ , replacing  $a_2^*$  we have:

$$E(d_H(\hat{X}_1, X)) + E(d_H(\hat{X}_2, X)) = 1 + (2D_{\{1\}} - 1)(\frac{1}{2} + a_1(1 - \frac{1}{2(1-a_1)})).$$

Solving for  $a_1$  we get  $a_1 = 1 - \frac{1}{\sqrt{2}}$  and in turn  $a_2 = 1 - \frac{1}{\sqrt{2}}$  as in the previous case.

**Case 3:**  $(1 - a_1)(1 - a_2) < \frac{1}{2}, a_1 > \frac{1}{2}, a_1a_2 < \frac{1}{2}$ : The probabilities are as in the last case with  $\alpha \in [0, \frac{1}{2} - (1 - a_1)(1 - a_2)]$ . The distortion is similar to the last case. Since  $a_1 > \frac{1}{2}$ , the distortion is decreasing in  $a_2$ . So  $a_2^* = \frac{1}{2a_1}$ . Which yields:

$$E(d_H(\hat{X}_1, X)) + E(d_H(\hat{X}_2, X)) = 1 + (2D_{\{1\}} - 1)((1 - a_1)(1 - \frac{1}{2a_1}) + \frac{1}{2}).$$

This would have no solution for optimizing  $a_1$  at the given range.

**Case 4:**  $a_1a_2 > \frac{1}{2}$ : By the same arguments the optimal solution is

$$P_{000}^* = (1 - a_1)(1 - a_2), P_{001}^* = (1 - a_1)a_2, P_{010}^* = (1 - a_2)a_1, P_{011}^* = 0$$

$$P_{100}^* = 0, P_{101}^* = 0, P_{010}^* = 0, P_{111}^* = \frac{1}{2}.$$

Then  $P_{\hat{X}_{12}, \hat{X}_1, \hat{X}_2}^*$  is: which is the second choice for the optimal joint distribution. Note that again  $\hat{X}_{12}$  is a

$\hat{X}_{12} \backslash \hat{X}_1, \hat{X}_2$	00	01	10	11
0	$\frac{3-2\sqrt{2}}{2}$	$\frac{\sqrt{2}-1}{2}$	$\frac{\sqrt{2}-1}{2}$	0
1	0	0	0	$\frac{1}{2}$

Table 5

function of  $\hat{X}_1$  and  $\hat{X}_2$ . □

**Step 5:** We are left with  $C_{\{1\},\{3\}}, C_{\{2\},\{3\}}, C_{\{1\}}, C_{\{2\}}$  and  $C_{\{3\}}$ . Let  $X_i$  be the reconstruction at decoder  $\{i\}$  for  $i \in \{1, 2, 3\}$ .

**Lemma 29.** *The following Markov chains hold:*

$$U_{\{1,\{3\}\}}, U_{\{1\}}, X_1 \perp\!\!\!\perp U_{\{2,\{3\}\}}, U_{\{2\}}, X_2 \quad (64)$$

$$U_{\{1\}}, U_{\{2\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}} \leftrightarrow X_1, X_2 \leftrightarrow X \quad (65)$$

$$U_{\{1,\{3\}\}}, U_{\{1\}} \leftrightarrow X_1 \leftrightarrow X, U_{\{2,\{3\}\}}, U_{\{2\}} \quad (66)$$

$$U_{\{2,\{3\}\}}, U_{\{2\}} \leftrightarrow X_2 \leftrightarrow X, U_{\{1,\{3\}\}}, U_{\{1\}} \quad (67)$$

$$U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, U_{\{3\}} \leftrightarrow X_3 \leftrightarrow X \quad (68)$$

$$X_1, X_2, U_{\{1\}}, U_{\{2\}} \leftrightarrow U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}, X \leftrightarrow U_{\{3\}}, X_3 \quad (69)$$

$$U_{\{1\}} \leftrightarrow U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}, X_1, U_{\{3\}} \leftrightarrow X \quad (70)$$

$$U_{\{2\}} \leftrightarrow U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}, X_2, U_{\{3\}} \leftrightarrow X \quad (71)$$

$$U_{\{3\}} \leftrightarrow U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_3, X_1 \leftrightarrow X \quad (72)$$

$$U_{\{3\}} \leftrightarrow U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_3, X_2 \leftrightarrow X \quad (73)$$

*Proof.* (64) holds from Lemma 19. From the optimality at decoder  $\{1, 2\}$  and step 4 we have:

$$I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, U_{\{1\}}, U_{\{2\}}, X_1, X_2; X) = I(X_1, X_2; X) = 1 - h_b(D_0),$$

which proves (65). Next we prove (66):

$$\begin{aligned} P(U_{\{1,\{3\}\}}, U_{\{1\}}, U_{\{2,\{3\}\}}, U_{\{2\}}, X_1, X) &\stackrel{(a)}{=} \sum_{X_2} P(U_{\{1,\{3\}\}}, U_{\{1\}}, X_1)P(X_2, U_{\{2,\{3\}\}}, U_{\{2\}})P(X|X_1, X_2) \\ &= P(U_{\{1,\{3\}\}}, U_{\{1\}}, X_1)P(U_{\{2,\{3\}\}}, U_{\{2\}}) \sum_{X_2} P(X_2|U_{\{2\}}, U_{\{2,\{3\}\}})P(X|X_1, X_2) \\ &= P(U_{\{1,\{3\}\}}, U_{\{1\}}, X_1)P(U_{\{2,\{3\}\}}, U_{\{2\}})P(X|X_1, U_{\{2\}}, U_{\{2,\{3\}\}}) \\ &\stackrel{(b)}{=} P(U_{\{1,\{3\}\}}, U_{\{1\}}, X_1)P(U_{\{2\}}, U_{\{2,\{3\}\}}, X|X_1) \end{aligned}$$

In (a) we have used (64) and the Markov chain (65), in (b), we have used (64). (67) follows by symmetry. (68) can be proved using optimality at decoder  $\{3\}$  and the argument given in the proof of (64). We proceed with the proof of (69). Consider the following packing bounds at decoder  $\{1, 2\}$  and  $\{3\}$ :

$$\begin{aligned} H(U_{\{1\}}, U_{\{2\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}) &\leq H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{1,\{3\}\}}) + H(U_{\{2,\{3\}\}}) - r_1 - r_2 - r_{1,3} - r_{2,3} + R_1 + R_2 \\ H(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, U_{\{3\}}) &\leq H(U_{\{1,\{3\}\}}) + H(U_{\{2,\{3\}\}}) + H(U_{\{3\}}) - r_{1,3} - r_{2,3} - r_3 + R_3 \end{aligned}$$

And the following covering bounds:

$$\begin{aligned} H(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}|X) &\geq H(U_{\{1\}}) + H(U_{\{2\}}) + H(U_{\{3\}}) \\ &\quad + H(U_{\{1,\{3\}\}}) + H(U_{\{2,\{3\}\}}) - r_1 - r_2 - r_3 - r_{1,3} - r_{2,3} \\ H(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}|X) &\geq H(U_{\{1,\{3\}\}}) + H(U_{\{2,\{3\}\}}) - r_{1,3} - r_{2,3} \end{aligned}$$

Adding all the bounds and simplifying we get:

$$R_1 + R_2 + R_3 \geq I(U_{\{1\}}, U_{\{2\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}; X) + I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, U_{\{3\}}; X) + I(U_{\{1\}}, U_{\{2\}}, U_{\{3\}}|U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X)$$

This resembles the two user sum-rate bound when the first user is sending descriptions 1 and 2 while the second user transmits description 3. From optimality at decoder  $\{1\}2$ ,  $R_1 + R_2 = I(U_{\{1\}}, U_{\{2\}}, U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}; X)$

and optimality at decoder {3} yields  $R_3 = I(U_{\{1,\{3\}}, U_{\{2,\{3\}}, U_{\{3\}}; X)$ . So  $I(U_{\{1\}}, U_{\{2\}}; U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X) = 0$ . This proves (69). We have:

$$\begin{aligned}
& P(U_{\{1,\{3\}}, U_{\{1\}}, X_1, U_{\{2,\{3\}}, U_{\{3\}}, X) \\
&= P(U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(U_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1, U_{\{1\}})P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1, U_{\{1\}}, X) \\
&\stackrel{(a)}{=} P(U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(U_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1, U_{\{1\}}, X) \\
&\stackrel{(b)}{=} P(U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(U_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1)P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X)
\end{aligned}$$

where (a) follows from (66) and Lemma 30 given below. (b) follows from (69). So we have shown that  $U_{\{1\}} \leftrightarrow U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_1 \leftrightarrow X, U_{\{3\}}$ , using Lemma 30 we conclude (70). (71) follows by symmetry. Lastly we prove (72):

$$\begin{aligned}
& P(X, X_1, U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3) \\
&= P(X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3)P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3, X)P(X_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X, U_{\{3\}}, X_3) \\
&\stackrel{a}{=} P(X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3)P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3, X)P(X_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X, X_3) \\
&= P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3, X)P(X, X_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3) \\
&\stackrel{b}{=} P(U_{\{3\}}|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3)P(X, X_1|U_{\{1,\{3\}}, U_{\{2,\{3\}}, X_3)
\end{aligned}$$

where (a) follows form 69. (b) holds because of (68). (72) follows from lemma 30.

**Lemma 30.** For random variables  $A, B, C$  and  $D$  if we have  $A, B \leftrightarrow C \leftrightarrow D$  then  $A \leftrightarrow B, C \leftrightarrow D$ .

*Proof.* We have:

$$P(A, D|B, C) = \frac{P(A, B, C, D)}{P(B, C)} = \frac{P(C)P(A, B|C)P(D|C)}{P(C)P(B|C)} = P(A|BC)P(D|C) = P(A|BC)P(D|BC)$$

□

□

Next we argue that if we set  $U_{\{1\}}$  to be equal to  $X_1$  there would be no change in distortion and the rate does not increase. First consider decoder {1, 3}. The optimal reconstruction function is given by  $\operatorname{argmax}_x(P_{X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, U_{\{1\}}, U_{\{3\}}}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}}))$ . We have:

$$\begin{aligned}
& \operatorname{argmax}_x(P_{X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, U_{\{1\}}, U_{\{3\}}}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}})) \\
&\stackrel{(a)}{=} \operatorname{argmax}_x(P_{X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, U_{\{1\}}, U_{\{3\}}, X_1}(x|u_{1,3}, u_{23}, u_{\{1\}}, u_{\{3\}}, x_1)) \\
&\stackrel{(b)}{=} \operatorname{argmax}_x(P_{X|U_{\{1,\{3\}}, U_{\{2,\{3\}}, U_{\{3\}}, X_1}(x|u_{1,3}, u_{23}, u_{\{3\}}, x_1))
\end{aligned}$$

where in (a) we used the fact that  $x_1$  is a function of  $U_{\{1\}}, U_{\{1,\{3\}}$  and in (b) we use (70). So the distortion won't change at decoder {1, 3}. Also the reconstruction at decoder {1} is  $X_1$  so setting  $U_{\{1\}} = X_1$  won't change the reconstruction at this decoder. At decoder {1, 2} we showed in step 4 that  $X_{12}$  is a function of  $X_1, X_2$  where  $X_2$  is a function of  $U_{\{2,\{3\}}, U_{\{2\}}$ , so setting  $U_{\{1\}} = X_1$  does not change the distortion at this decoder either. The rest of the decoders do not receive  $U_{\{1\}}$ . As for the rate, note that  $X_1$  was reconstructed at

all decoders reconstructing  $U_{\{1\}}$ . So replacing  $U_{\{1\}}$  with  $X_1$  does not require sending any extra information. So we set  $U_{\{1\}} = X_1$  without any loss in distortion and with a potential gain in rate. The same argument combined with the Markov chains (71) sets  $U_{\{2\}} = X_2$ , also using Markov chains (72) and (73) we set  $U_{\{3\}} = X_3$ .

**Lemma 31.** *The following constraints hold:*

$$P_{XX_1X_2} \text{ is fixed and equal to } P_{X,V_{\{1\}},V_{\{2\}}} \text{ in the previous step.} \quad (74)$$

$$P_{XX_3} \text{ is fixed and equal to } P_{X,V_3} \text{ which is the optimizing distribution for decoder } \{3\}. \quad (75)$$

$$U_{\{1\},\{3\}} \leftrightarrow X_1 \leftrightarrow U_{\{2\},\{3\}}, X, X_2 \quad (76)$$

$$U_{\{2\},\{3\}} \leftrightarrow X_2 \leftrightarrow U_{\{1\},\{3\}}, X, X_1 \quad (77)$$

$$U_{\{1\},\{3\}}, U_{\{2\},\{3\}} \leftrightarrow X_3 \leftrightarrow X \quad (78)$$

$$X_1, X_2 \leftrightarrow X, U_{\{1\},\{3\}}, U_{\{2\},\{3\}} \leftrightarrow X_3 \quad (79)$$

*Proof.* (74) was proved in the step 4. (75) follows from PtP optimality at decoder {3}. (76) follows from (66), (77) follows from (67). (78) follows from (68). (79) follows from (69).  $\square$

We proceed by bounding the cardinality of  $U_{\{1\},\{3\}}$  and  $U_{\{2\},\{3\}}$ . Using Lemma 31, the joint distribution between the random variables is given as follows:

$$\begin{aligned} P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X_1, X_2, X_3, X) &= P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)P(X_3|U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X) \\ &= P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2) \frac{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X_3X)}{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}X)} \\ &= P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2) \frac{P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3)P(X_3X)}{\sum_{X_1, X_2} P(U_{\{1\},\{3\}}, X_1)P(U_{\{2\},\{3\}}, X_2)P(X|X_1, X_2)} \end{aligned} \quad (80)$$

Also note that we have the following equality:

$$P(U_{\{1\},\{3\}}, U_{\{2\},\{3\}}, X) = \sum_{X_3} P(X, X_3)P(U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3) = \sum_{X_1, X_2} P(U_{\{1\},\{3\}}|X_1)P(U_{\{2\},\{3\}}|X_2)P(XX_1, X_2)$$

Denote  $P(X, X_1, X_2) = P_{xx_1x_2}$  and  $P(U_{\{1\},\{3\}} = \theta|X_1 = i) = \alpha_i(\theta)$ ,  $\theta \in U_{1,3}$ ,  $i \in \{0, 1\}$  and  $P(U_{\{2\},\{3\}} = \gamma|X_2 = i) = \beta_i(\gamma)$ ,  $\gamma \in U_{2,3}$ ,  $i \in \{0, 1\}$ . We have:

$$\begin{aligned} P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|0)P_{X_3, X}(0, 0) + P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|1)P_{X_3, X}(1, 0) \\ = \alpha_0(\theta)\beta_0(\gamma)P_{000} + \alpha_0(\theta)\beta_1(\gamma)P_{001} + \alpha_1(\theta)\beta_0(\gamma)P_{010} + \alpha_1(\theta)\beta_1(\gamma)P_{011} \end{aligned}$$

$$\begin{aligned} P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|0)P_{X_3, X}(0, 1) + P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|1)P_{X_3, X}(1, 1) \\ = \alpha_0(\theta)\beta_0(\gamma)P_{100} + \alpha_0(\theta)\beta_1(\gamma)P_{101} + \alpha_1(\theta)\beta_0(\gamma)P_{110} + \alpha_1(\theta)\beta_1(\gamma)P_{111} \end{aligned}$$

Using the values given in Table (1), we solve the system of equations:

$$\begin{aligned} P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|0) &= \alpha_0(\theta)\beta_0(\gamma) \\ P_{U_{\{1\},\{3\}}U_{\{2\},\{3\}}|X_3}(\theta, \gamma|1) &= \frac{\sqrt{2}-1}{2}(\alpha_1(\theta)\beta_1(\gamma) - \alpha_0(\theta)\beta_0(\gamma)) + \frac{1}{2}(\alpha_0(\theta)\beta_1(\gamma) + \alpha_1(\theta)\beta_0(\gamma)) \end{aligned}$$

Hence the distribution in 80 is completely determined by  $\alpha_i$  and  $\beta_i$ ,  $i \in \{0, 1\}$ .

**Lemma 32.** Assume there exists  $\alpha_i$  and  $\beta_i$ , such that  $D_{\{1,3\}} \leq D_0$ , then  $I(U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}X_1X_3; X) \geq 1 - h_b(D_0)$ .

*Proof.* The proof follows from Shannon's rate distortion function for PtP source coding.  $\square$

Based on the previous lemma it is enough to show that for every  $\alpha_i$  and  $\beta_i$ ,  $I(U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}X_1X_3; X) < 1 - h_b(D_0)$ , in that case we have a contradiction. We need to maximize  $I(U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}X_1X_3; X)$  as a function of  $\alpha_i$  and  $\beta_i$ . We use the following lemma:

**Lemma 33.** [21] Let  $\mathcal{X}$  be a finite set and  $\mathcal{U}$  be an arbitrary set. Let  $\mathcal{P}(\mathcal{X})$  be a set of pmfs on  $\mathcal{X}$  and  $p(x|u)$  be a collection of pmfs on  $\mathcal{X}$  for every  $u \in \mathcal{U}$ . Let  $g_j, j \in [1, d]$  be real-valued continuous functions on  $\mathcal{P}(\mathcal{X})$ . Then for every  $U \sim F(u)$  defined on  $\mathcal{U}$ , there exists random variable  $U' \sim p(u')$  with cardinality  $|\mathcal{U}'| \leq d$  and a collection of conditional pmfs  $p(u'|x)$  on  $\mathcal{X}$  for every  $u' \in \mathcal{U}'$  such that for every  $j \in [1, d]$ :

$$\int_{\mathcal{U}} g_j(p_{X|U}(x|u))dF(u) = \sum_{u'} g_j(p_{X|U'}(x|u'))p(u')$$

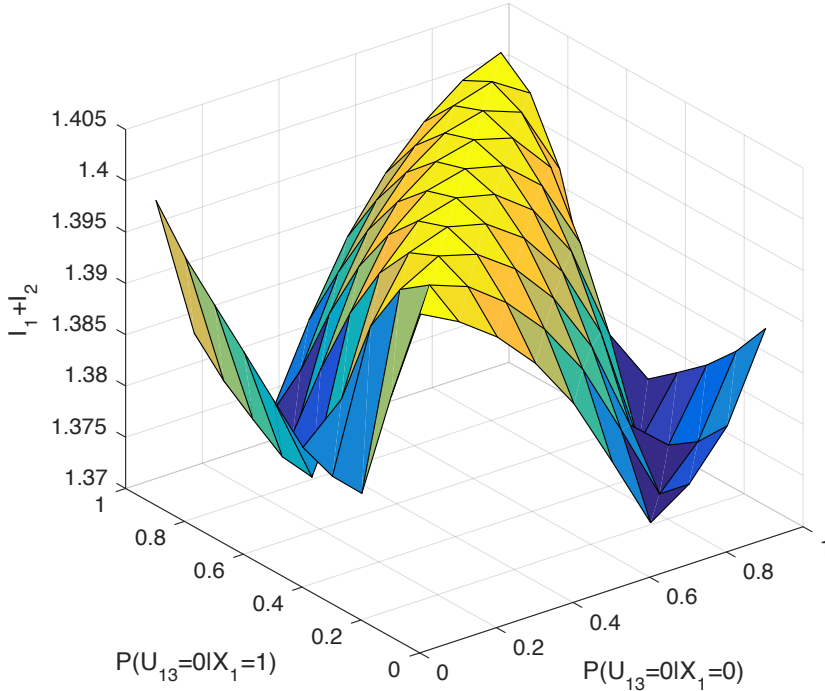


Figure 10: Plot of maximum value of  $I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_1, X_3; X) + I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_2, X_3; X)$

We want to use the lemma to bound cardinality of  $\mathcal{U}_{1,3}$ . Take  $g_1(p_{U_{\{2,\{3\}\}}, X_1, X_2, X_3, X|U_{\{1,\{3\}\}}}) = p_{X_1|U_{\{1,\{3\}\}}}(1|u_{13})$  and  $g_2(p_{U_{\{2,\{3\}\}}, X_1, X_2, X_3, X|U_{\{1,\{3\}\}}}) = H(X|U_{\{2,\{3\}\}}, X_1, X_3, X, U_{\{1,\{3\}\}} = u_{1,3})$ . Note that fixing the expectation on  $g_1$  fixes the joint distribution in (80) and fixing the expectation of  $g_2$  fixes the term we want to minimize. So for any  $U_{\{1,\{3\}\}}$  minimizing  $I(U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}X_1X_3; X)$ , there exists  $U'_{1,3}$  with cardinality at most 2, such that the joint distribution and  $I(U_{\{1,\{3\}\}}U_{\{2,\{3\}\}}X_1X_3; X)$  are the same. So it is enough to search over  $U_{\{1,\{3\}\}}$  with cardinality 2. The same arguments hold for bounding the cardinality of  $\mathcal{U}_{2,3}$ . For this size of random variables, computer-assisted calculation shows that  $I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_1, X_3; X) + I(U_{\{1,\{3\}\}}, U_{\{2,\{3\}\}}, X_2, X_3; X) < 1.42 < 2(1 - h_b(D_0)) = 1.58$  as shown in Figure 10. So we have a contradiction and the SSC does not achieve the RD vector.  $\square$

## C Proofs for Section 6

### C.1 Proof of lemma 15

*Proof.* Index the inequalities in the SSC from 1 to  $K$ . For every inequality in the linear coding region (LCR), there exists a unique inequality in the SSC with the same left hand side, index this inequality with the same index used in the RCR. Let  $I_1 > R$  be a bound resulting from applying FME on the SSC. Assume the bound results from adding inequalities indexed  $i_1, i_2, \dots, i_k$ , it is straightforward to show that adding inequalities with the same indices in the LCR gives the same bound. The reason is that by our construction, the left-hand sides would be the same. In the right-hand side, due to the FME, the terms involving  $r_A$  would be eliminated. Define  $r'_A = r_A - H(U_A)$  and  $r'_{o,A} = r_{o,A} - \log(q)$ , eliminating  $r_A$  is equivalent to eliminating  $r'_A$  or  $r'_{o,A}$ .  $\square$

**Acknowledgment:** The authors would like to thank Prof. Kenneth Rose of UC Santa Barbara and Mohsen Heidari Khoozani of Univ. of Michigan for helpful discussions.

## References

- [1] V. K. Goyal, "Multiple description coding: compression meets the network," in Signal Processing Magazine, IEEE , vol.18, no.5, pp.74-93, Sep 2001
- [2] Y. Wang, A. R. Reibman, and S. Lin, "Multiple Description Coding for Video Delivery," in Proceedings of the IEEE , vol.93, no.1, pp.57-70, Jan. 2005
- [3] L. Ozarow, "On a source-coding problem with two channels and three receivers," Bell Systems Tech. Journal, 59(10):1909-1921, Dec 1980.
- [4] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," IEEE Trans. on Inf. Theory, 31(6):721 - 726, nov 1985.
- [5] A. El Gamal, and T. Cover, "Achievable rates for multiple descriptions," IEEE Trans. in Inf. Theory, vol. IT-28, pp. 851-857, 1982.
- [6] Z. Zhang, and T. Berger, "New results in binary multiple-descriptions," IEEE Trans. on Inf. Theory, vol.33, no.4, pp.502,521, Jul 1987
- [7] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," IEEE Trans. on Inf. Theory, vol. 49, no. 9, pp. 2106-2114, 2003.
- [8] C. Tian, and J. Chen, "New Coding Schemes for the Symmetric K -Description Problem," IEEE Trans. on Inf. Theory, vol.56, no.10, pp.5344,5365, Oct. 2010
- [9] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions-part I:(n, k) source-channel erasure codes," IEEE Trans. on Inf. Theory, vol. 50, no. 1, pp. 47-61, 2004.
- [10] E. Akyol, K. Viswanatha, and K. Rose, "Combinatorial message sharing and random binning for multiple description coding," in IEEE International Symp. on Inf. Theory. IEEE, pp.1371,1375, 1-6 July 2012.



- [11] K. Viswanatha, E. Akyol, and K. Rose, “Combinatorial message sharing for a refined multiple-descriptions achievable region,” in IEEE International Symp. on Inf. Theory. IEEE, pp. 1312-1316., 2011.
- [12] K. Viswanatha, E. Akyol, and K. Rose, “Combinatorial Message Sharing and a New Achievable Region for Multiple Descriptions,” in IEEE Trans. on Information Theory, pp. 769-792, Feb., 2016.
- [13] I. Anderson, “Combinatorics of finite sets,” The Clarendon Press Oxford University Press, New York, 1987.
- [14] I. Csiszár and J. Körner, “Information Theory: Coding Theorems for Discrete Memoryless Systems,” Academic Press Inc. Ltd., 1981.
- [15] J. Körner, and K. Marton, “How to encode the modulo-two sum of binary sources,” IEEE Trans. on Inf. Theory , vol. 25, no. 2, pp. 219-221, 1979.
- [16] A. Padakandla, and S.S. Pradhan, “Achievable rate region for three user discrete broadcast channel based on coset codes,” in IEEE International Symp. on Inf. Theory. IEEE, pp.1277,1281, 7-12 July 2013
- [17] A. Padakandla, A.G. Sahebi, and S.S. Pradhan, “A new achievable rate region for the 3-user discrete memoryless interference channel,” in IEEE International Symp. on Inf. Theory, July, 2012 (also to appear in IEEE Trans. on Information Theory 2016).
- [18] B. Nazer, and M. Gastpar, “Computation over Gaussian Multiple-Access Channels,” in IEEE International Symp. on Inf. Theory. IEEE, pp.2391,2395, 24-29 June 2007
- [19] T. Philosof, and R. Zamir, “On the Loss of Single-Letter Characterization: The Dirty Multiple Access Channel,” IEEE Trans. on Inf. Theory, vol.55, no.6, pp.2442,2454, June 2009
- [20] D. Krithivasan, and S.S. Pradhan, “Distributed Source Coding Using Abelian Group Codes: A New Achievable Rate-Distortion Region,” IEEE Trans. on Inf. Theory, vol.57, no.3, pp.1495,1519, March 2011
- [21] A. El Gamal, and Y. H. Kim, “Network information theory”, Cambridge university press, 2011
- [22] A.B. Wagner, B.G. Kelly and Y. Altuğ, “Distributed Rate-Distortion With Common Components,” IEEE Trans. on Inf. Theory, vol.57, no.7, pp.4035-4057, July 2011
- [23] F. Shirani, and S.S. Pradhan “An achievable rate-distortion region for the multiple descriptions problem,” 2014 IEEE International Symp. on Inf. Theory (ISIT), pp.576-580, June 29 2014-July 4 2014
- [24] D. Kleitman, G. Markowsky, “On Dedekind’s problem: the number of isotone Boolean functions. II”, Transactions of the American Mathematical Society 213: 373-390, 1975
- [25] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, “On the Role of the Refinement Layer in Multiple Description Coding and Scalable Coding,” IEEE Trans. on Inf. Theory, vol.57, no.3, pp.1443-1456, March 2011

- [26] S. S. Pradhan, J. Chou, K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. on Inf. Theory*, vol.49, no.5, pp.1181-1203, May 2003
- [27] R. G. Gallager, "Information Theory and Reliable Communication.," New York: Wiley, 1968.
- [28] F. S. Chaharsooghi, A. G. Sahebi, S. S. Pradhan, "Distributed source coding in absence of common components," in *IEEE International Symp. on Inf. Theory*. IEEE, pp.1362-1366, 7-12 July 2013
- [29] K. Vinodh, V. Lalitha, N. Prakash, P. V. Kumar, S. S. Pradhan, "On the achievable rates of sources having a group alphabet in a distributed source coding setting," in *Communication, Control, and Computing (Allerton)*, 2010 48th Annual Allerton Conference on, pp.479-486, Sept. 29 2010-Oct. 1 2010
- [30] P. Gacs and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 119-162, 1972.
- [31] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal of Applied Mathematics*, vol. 28, no. 1, pp. 100-113, January 1975.