CCIT Report #515 January 2005

Two-Way Joint Source-Channel Coding with a Fidelity Criterion

Alina Maor* and Neri Merhav

January 18, 2005

Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
{alinam@tx, merhav@ee}.technion.ac.il

Abstract

Consider a source, $\{X_i, Y_i\}_{i=1}^{\infty}$, producing independent copies of a pair of jointly distributed RVs. The $\{X_i\}$ part of the process is observed at some location, say A, and is supposed to be reproduced at a different location, say B, where the $\{Y_i\}$ part of the process is observed. Similarly, $\{Y_i\}$ should be reproduced at location A. The communication between the two locations is carried out across two memoryless channels in K iterative bi-directional rounds. In each round, the source components are reconstructed at the other locations based on the information exchanged in all previous rounds and the source component known at that location, and it is desired to find the amount of information that should be exchanged between the two locations in each round, so that the distortions incurred (in each round) will not exceed given thresholds. We first derive a single-letter characterization of achievable rates for a pure source-coding problem with successive refinement. Then, for a joint source-channel coding setting, we prove a separation theorem, asserting that in the limit of long blocks, no optimality is lost by first applying lossy (two-way) successive-refinement source coding, regardless of the channels, and then applying good channel codes to each one of the resulting bitstreams, regardless of the source.

Index terms - two-way communication, source coding, joint source-channel coding, successive refinement, channel capacity.

1 Introduction

In recent years, communication over distributed networks has been developed greatly. Intensive research was carried out in both theoretical and practical directions, analyzing

*This work is part of A. Maor's Ph.D. dissertation.

performance of computer networks, telecommunication (such as broadcast TV) and wireless networks, where the most notable examples are of cellular communication and wireless computer networks. The model, generally used in most of these analyzes is that of one-way communication between different network components, where each component transmits information to its neighbors in one transmission [1].

In many practical scenarios, two or more communication units (for example, sonar systems) have access to some noisy descriptions of information that is of interest to all units. Frequently, such units collaborate and wish to exchange their data reliably, in order to enhance the retrieval of the source information in each of the units. In a one-way communication scenario, each unit would send all its data to all the other units, and their transmission rates would fit the well-known model of communication with side information at the decoder [2] (the source description originally received in each of the units would serve as side information upon receiving transmissions from the other unit). Not only do these multiple one-way communications intuitively seem to be inefficient and wasteful in the sense of using transmission rates higher than required (as is, indeed shown in [3] for a pure source-coding problem), but, unfortunately, in many cases, physical constraints of the units (e.g., limited memory) and of the transmission media (e.g., capacity) do not allow such a one-way communication, and an interactive communication between units is to be implemented instead. In the sequel, we confine ourselves to the case of two such units.

The interactive two-way communication problem (over a noise-free medium) was originally studied by Kaspi [3] in the context of lossy source coding, and to the best of our knowledge, this is the only paper related to the lossy two-way source-coding with a fidelity criterion. The following communication model was studied in [3]: Two encoding/decoding units (codecs) have access to certain sources. Codec 1 has access to a discrete memoryless source (DMS) $\boldsymbol{X} = (X_1, X_2, ...)$ and codec 2 has access to another DMS, $\boldsymbol{Y} = (Y_1, Y_2, ...)$, correlated to $\boldsymbol{X}$. In the first communication step, one of the codecs, say codec 1, initiates the data exchange protocol by transmitting to codec 2 some partial description of $\boldsymbol{X}$ (in the form of a compressed bitstream) based on the fact that codec 2 observes $\boldsymbol{Y}$ as side information [2]. Codec 2, in turn, produces a description of $\boldsymbol{Y}$ based on the received bitstream from codec 1 and based on the fact that codec 1 has $\boldsymbol{X}$ as side information. In the second round, codec 1 transmits an additional portion of a description of $\boldsymbol{X}$ to codec 2, on the basis of the information available to both codecs thus far. Then, codec 2 replies similarly, etc.

This process repeats K times. Kaspi [3] found a single-letter description of the achievable region of all cumulative communication rates after K rounds, i.e., the total rates used by each codec, such that after K steps, each one of the codecs would be able to reproduce the other source within given distortion levels.

Another part of the background for this work has to do with the successive refinement of information. In [4], the (one-way) problem of successive coding is studied for the Wyner-Ziv setting: The encoder transmits a source sequence, $\boldsymbol{X}$, to two decoders in two successive steps. In the first step, a coarse description of the source is transmitted to the first-stage decoder at a relatively low rate, and the reconstruction at the decoder should satisfy a certain distortion constraint. The reconstruction is based on the received bitstream and on side information, correlated to the source, which is available at the decoder. At the second decoder, which has access to additional bits of description of $\boldsymbol{X}$, as well as to the first step bitstream, a higher quality of reconstruction of $\boldsymbol{X}$ is required. In other words, in the second stage, the encoder transmits refinement bits to the second-stage decoder and the decoder reconstructs $\boldsymbol{X}$ based on the bitstreams of both stages and on side information known to it. In general, the second-stage side information differs from this available at the first-stage decoder. Necessary and sufficient conditions are provided, in terms of single-letter formulas, for the achievability of information rates corresponding to given distortion levels of each communication step. For stochastically degraded sources, the two-stage coding scheme is extended to include any finite number of steps. Special attention is devoted in [4] to the case where the side information streams at the decoders are identical, and a notion of successively refinable sources is introduced along with conditions for successive refinability.

In [5], the noise-free setting of [4] is extended into a joint source-channel coding setting, considering communication across independent memoryless channels. Similarly as in [4], the output of the channel corresponding to the coarse (first) description of $\boldsymbol{X}$ is also available to the refinement (second) decoder and each of the decoder may have access to different side information correlated with the source. The main result of [5] is a separation theorem asserting that asymptotically no optimality is lost by first applying lossy successive source coding, regardless of the channels and then applying good channel codes to each of the resulting bitstreams, regardless of the source and the side information.

In this paper, we extend the setups of [5] and [3] in a combined manner: We consider the scenario of [3], where the communication between the two codecs is conducted via noisy channels, and furthermore, unlike in [3], where the reconstruction of both sources is carried out only after K rounds, here, distortion constraints are imposed after each round, and in this sense, it extends the scenario of [5]. To simplify the exposition, we begin with the pure source coding part. We provide a single-letter description of the achievable region of all per-round communication rates. A special case of our setting arises when only codec 1 allowed to transmit and codec 2 is "silent". In this case, the achievable rate-region derived in this paper degenerates to that of [4] for the case of identical side information at both decoders. The result of [3] can be derived from our scheme by relaxing the intermediate distortion constraints and summing the transmission rates used in K rounds in each direction.

We next extend the noise-free setting of two-way lossy source coding with a fidelity criterion into a joint source-channel coding setting. In our model, we consider the problem where the communication between the two codecs goes over two independent memoryless channels. The main result of this paper is a single-letter characterization of the achievable region of distortions achieved in each round. The main feature of this characterization is that it admits a separation principle, which tells that no optimality is lost if at each communication step, one first applies optimum two-way lossy successive-refinement source coding, independently of the channels, and then applies good channel coding [1] for each one of the compressed bitstreams, independently of the source descriptions available at each of the codecs.

It should be pointed out that this separation theorem is not trivial. As is well known, source-channel separation does not always exist even in non-interactive communication (see, e.g., [1, pp. 448–449]). A-fortiori, in our communication model, which is interactive, the optimality of source-channel separation is not self-evident since information flows bidirectionally and in several rounds. This separation theorem also motivates the design of good practical source codes which are independent of channel coding, and gives rise to various algorithmic approaches to the interactive communication problem. Similarly as in the noise-free scenario, if only codec 1 transmits, our results degrade to these of [5] for a special case of identical side information available at both decoders.

The outline of the paper is as follows. In Section 2, we define the notation conventions as well as some terminology used throughout the paper. A formal definition of the problem is provided in Section 3. In Section 4, we give the characterizations of the achievable rate-regions and formulate the coding theorems for the successive-refinement two-way source coding and the joint source-channel coding. As source-coding is a special case of joint source-channel coding, to avoid redundancy, the proofs of the coding theorems are provided jointly in Section 5.

2 Notation Conventions and Preliminaries

We begin by setting up the notation. Throughout this paper, scalar random variables (RVs) will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets, as well as most of the other sets, will be denoted by calligraphic letters. Similarly, random vectors, their realizations, and their alphabets will be denoted, respectively, by boldface capital letters, the corresponding boldface lower case letters, and calligraphic letters, superscripted by the dimensions. The notations x_i^j and X_i^j, where i and j are integers and $i \leq j$, will designate segments $(x_i, ..., x_j)$ and $(X_i, ..., X_j)$, respectively, where for $i = 1$, the the subscript will be omitted. For example, the random vector $\boldsymbol{X} = X^N = X_1^N = (X_1, ..., X_N)$, ($N$-positive integer) may take a specific vector value $\boldsymbol{x} = x^N = x_1^N = (x_1, ..., x_N)$ in $\mathcal{X}^N$, the Nth order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. The cardinality of a finite set $\mathcal{X}$ will be denoted by $|\mathcal{X}|$. For $i > j$, x_i^j (or X_i^j) will be understood as the null string.

Sources and channels will be denoted generically by the letter P subscripted by the name of the random variable and its conditioning, if applicable, e.g., $P_X(x)$ is the probability of $X = x$, $P_{Y|X}(y|x)$ is the conditional probability of $Y = y$ given $X = x$, and so on. Whenever clear from the context, these subscripts will be omitted. The notation E will denote the expectation operator. Information-theoretic quantities will be denoted using the conventional notations [6]-[7]: For a pair of discrete random variables (X, Y) with a joint distribution $P_{XY}(x, y) = P_X(x)P_{Y|X}(y|x)$, the entropy of X will be denoted by $H(X)$, the joint entropy - by $H(X, Y)$, the conditional entropy of Y given X - by $H(Y|X)$, and the mutual information by $I(X; Y)$, where logarithms are defined to the base 2. In this paper, we use the definitions from [7] for all the calculations related to the method of types.

3 System Description and Problem Definition

We refer to the communication system depicted in Fig. 1: Consider a source, $\{X_i, Y_i\}_{i=1}^{\infty}$, producing N independent copies of a pair of RVs, (X, Y), taking values in $\mathcal{X} \times \mathcal{Y}$, and drawn

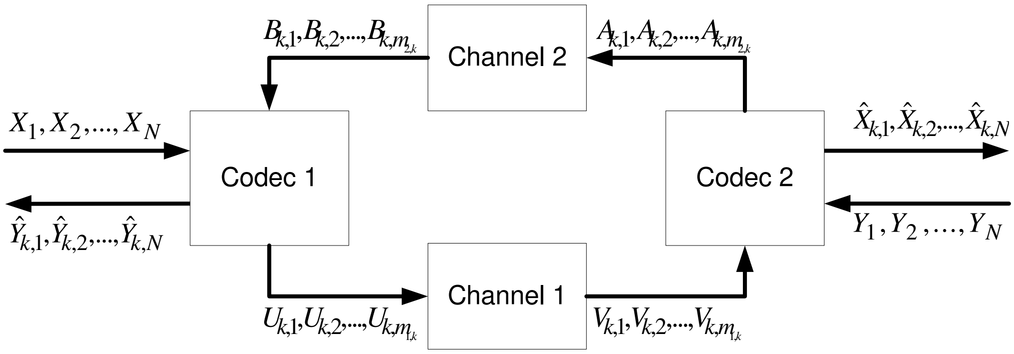

Figure 1: Two-way communication scheme.

under a joint distribution P_{XY}. The $\{X_i\}$ part of the $\{X_i, Y_i\}$ process is observed at the codec 1 and is supposed to be reproduced at the other side (codec 2), where the $\{Y_i\}$ part of the process is observed. Similarly, the process $\{Y_i\}$, which is observed at codec 2, is to be recovered at codec 1. The reproductions take values in the finite sets, $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$, respectively. The communication between the two codecs is carried out over two memoryless channels, channel 1, $P_{V|U}$, from codec 1 to codec 2, and channel 2, $P_{B|A}$, from codec 2 to codec 1. We denote by C_1 and C_2 the capacities of channel 1 and channel 2, respectively.

The following protocol for exchange of information is considered in this paper: Codec 1 starts the information exchange by sending some amount of information to codec 2. We consider block coding only, i.e., an N-vector $\boldsymbol{x}$ is encoded by codec 1 into the block sequence $\boldsymbol{u}_1 = (u_{1,1}, ..., u_{1,m_{1,1}})$ of length $m_{1,1}$. Codec 2 receives a noisy version of $\boldsymbol{u}_1$, denoted by $\boldsymbol{v}_1 = (v_{1,1}, ..., v_{1,m_{1,1}})$, and replies to codec 1 with some information about $\boldsymbol{y}$, which is $\boldsymbol{a}_1 = (a_{1,1}, ..., a_{1,m_{2,1}})$, of length $m_{2,1}$, that is a function of $\boldsymbol{y}$ and $\boldsymbol{v}_1$. This information is distorted by the memoryless channel $P_{B|A}$, and thus, reaches codec 1 as $\boldsymbol{b}_1 = (b_{1,1}, ..., b_{1,m_{2,1}})$. Then, codec 1 sends additional information to codec 2, of length $m_{1,2}$, now basing the generation of $\boldsymbol{u}_2$ on $\boldsymbol{x}$ and $\boldsymbol{b}_1$, etc. The process is repeated K times. In each communication step k, $k = 1, 2, ..., K$, codec 1 uses codewords of length $m_{1,k}$ and codec 2 generates codewords of length $m_{2,k}$, and each codec produces an estimation of the respective part of the source, based on all information currently available. In other words, after step k, codec 1 reconstructs $\hat{\boldsymbol{y}}_k = (\hat{y}_{k,1}, ..., \hat{y}_{k,N}) \in \hat{\mathcal{Y}}^N$ based on $(\boldsymbol{x}, \boldsymbol{b}_1, ..., \boldsymbol{b}_{k-1})$ and codec 2 reproduces $\hat{\boldsymbol{x}}_k = (\hat{x}_{k,1}, ..., \hat{x}_{k,N}) \in \hat{\mathcal{X}}^N$ based on $(\boldsymbol{y}, \boldsymbol{v}_1, ..., \boldsymbol{v}_k)$. The quality of reconstruction after step k is judged in terms of the expectation of an additive distortion measure

$$d_{x,k}(\boldsymbol{x}, \hat{\boldsymbol{x}}_k) = \frac{1}{N}\sum_{l=1}^{N} d_{x,k}(x_l, \hat{x}_{k,l}), \tag{1}$$

and

$$d_{y,k}(\boldsymbol{y}, \hat{\boldsymbol{y}}_k) = \frac{1}{N}\sum_{l=1}^{N} d_{y,k}(y_l, \hat{y}_{k,l}), \tag{2}$$

where $d_{x,k}(x, \hat{x}_k) \geq 0$, $d_{y,k}(y, \hat{y}_k) \geq 0$, $x \in \mathcal{X}$, $\hat{x}_k \in \hat{\mathcal{X}}$, $y \in \mathcal{Y}$, $\hat{y}_k \in \hat{\mathcal{Y}}$, are given bounded distortion measures used in each communication step k.

The source P_{XY} generates block of N independent pairs $\{X_i, Y_i\}$ each T seconds, thus operating at the rate of $\rho_s \triangleq \frac{N}{T}$ source symbols per second. The duration of a block, T, is divided between K communication rounds, with each round consisting of two one-directional transmissions. The two channels operate at constant rates of ρ_{c_1} and ρ_{c_2} channel symbols per second, respectively. For each iteration, $k = 1, 2, ..., K$, we denote the transmission duration from codec 1 to codec 2 by $T_{1,k}$ and from codec 2 to codec 1 by $T_{2,k}$. Thus,

$$T = \sum_{k=1}^{K}(T_{1,k} + T_{2,k}). \tag{3}$$

For the k-th iteration, the operation rates of channel 1 and channel 2 can be given in terms of the block lengths of channel sequences and the duration of the iteration, i.e.,

$$\rho_{c_1} \triangleq \frac{m_{1,k}}{T_{1,k}} = \frac{T}{T_{1,k}} \cdot \frac{m_{1,k}}{T} \triangleq \beta_{1,k} \cdot \frac{m_{1,k}}{T} \tag{4}$$

and

$$\rho_{c_2} \triangleq \frac{m_{2,k}}{T_{2,k}} = \frac{T}{T_{2,k}} \cdot \frac{m_{2,k}}{T} \triangleq \beta_{2,k} \cdot \frac{m_{2,k}}{T}, \tag{5}$$

where $\beta_{1,k}$ and $\beta_{2,k}$ denote the relative duration ratios of each of the one-directional transmissions.

Definition 1. *An $(N, \{m_{1,k}\}_{k=1}^{K}, \{m_{2,k}\}_{k=1}^{K}, \{\Delta_{x,i}\}_{k=1}^{K}, \{\Delta_{y,i}\}_{k=1}^{K})$, K-step joint source-channel code for the source P_{XY} and the channels $P_{V|U}$ and $P_{B|A}$, consists of $2K$ encoder-decoder pairs $\{(f_k, G_k)\}_{k=1}^{K}$ and $\{(g_k, F_k)\}_{k=1}^{K}$:*

$$f_k : \mathcal{X}^N \times \mathcal{B}^{m_{2,1}} \times ... \times \mathcal{B}^{m_{2,k-1}} \rightarrow \mathcal{U}^{m_{1,k}}, \tag{6}$$

$$G_k : \mathcal{Y}^N \times \mathcal{V}^{m_{1,1}} \times ... \times \mathcal{V}^{m_{1,k}} \rightarrow \hat{\mathcal{X}}^N, \tag{7}$$

and

$$g_k : \mathcal{Y}^N \times \mathcal{V}^{m_{1,1}} \times ... \times \mathcal{V}^{m_{1,k}} \rightarrow \mathcal{A}^{m_{2,k}}, \tag{8}$$

$$F_k : \mathcal{X}^N \times \mathcal{B}^{m_{2,1}} \times ... \times \mathcal{B}^{m_{2,k}} \rightarrow \hat{\mathcal{Y}}^N, \tag{9}$$

such that for every $k = 1, 2, ..., K$*,*

$$Ed_{x,k}(\mathbf{X}, G_k(\mathbf{Y}, \mathbf{V}_1, ..., \mathbf{V}_k)) \leq N\Delta_{x,k}, \tag{10}$$

$$Ed_{y,k}(\mathbf{Y}, F_k(\mathbf{X}, \mathbf{B}_1, ..., \mathbf{B}_k)) \leq N\Delta_{y,k}. \tag{11}$$

Definition 2. *The distortions* $\{(\Delta_{x,k}, \Delta_{y,k})\}_{k=1}^K$ *are said to be achievable if for every* $\epsilon > 0$*, and sufficiently large block-lengths* N*,* $\{m_{1,k}\}_{k=1}^K$ *and* $\{m_{2,k}\}_{k=1}^K$ *there exists an* $(N, \{m_{1,k}\}_{k=1}^K, \{m_{2,k}\}_{k=1}^K, \{\Delta_{x,k} + \epsilon\}_{k=1}^K, \{\Delta_{y,k} + \epsilon\}_{k=1}^K)$ *K-step joint source-channel code for the source* P_{XY} *and the channels* $P_{V|U}$*,* $P_{B|A}$*, with operation rates* ρ_s*,* ρ_{c_1} *and* ρ_{c_2}*. The distortion region, denoted* $\mathcal{D}$*, is the closure of the set of all vectors* $\{(\Delta_{x,k}, \Delta_{y,k})\}_{k=1}^K$*.*

The characterization of the achievable region of $\mathcal{D}$ is intimately related to the definition of the successive source coding. Consider then the following definition for the corresponding pure source coding problem (with noise-free channels):

Definition 3. *An* $(N, \{M_{1,k}\}_{i=1}^K, \{M_{2,k}\}_{i=1}^K, \{\Delta_{x,i}\}_{i=1}^K, \{\Delta_{y,i}\}_{i=1}^K)$*, K-step source code with successive refinement for the source* P_{XY}*, consists of* $2K$ *encoder-decoder pairs* $\{(f_k, G_k)\}_{k=1}^K$ *and* $\{(g_k, F_k)\}_{k=1}^K$*:*

$$f_k : \mathcal{X}^N \times \{1, 2, ..., M_{2,1}\} \times ... \times \{1, 2, ..., M_{2,k-1}\} \rightarrow \{1, 2, ..., M_{1,k}\}, \tag{12}$$

$$G_k : \mathcal{Y}^N \times \{1, 2, ..., M_{1,1}\} \times ... \times \{1, 2, ..., M_{1,k}\} \rightarrow \hat{\mathcal{X}}^N, \tag{13}$$

and

$$g_k : \mathcal{Y}^N \times \{1, 2, ..., M_{1,1}\} \times ... \times \{1, 2, ..., M_{1,k}\} \rightarrow \{1, 2, ..., M_{2,k}\}, \tag{14}$$

$$F_k : \mathcal{X}^N \times \{1, 2, ..., M_{2,1}\} \times ... \times \{1, 2, ..., M_{2,k}\} \rightarrow \hat{\mathcal{Y}}^N, \tag{15}$$

such that for every $k = 1, 2, ..., K$*,*

$$Ed_{x,k}(\mathbf{X}, G_k(\mathbf{Y}, f_1(\mathbf{X}), ..., f_k(\mathbf{X}))) \leq N\Delta_{x,k}, \tag{16}$$

$$Ed_{y,k}(\mathbf{Y}, F_k(\mathbf{X}, g_1(\mathbf{Y}), ..., g_k(\mathbf{Y}))) \leq N\Delta_{y,k}. \tag{17}$$

The K rate-pairs $\{(R_{x,k}, R_{y,k})\}_{k=1}^K$ of the $(N, \{M_{1,k}\}_{k=1}^K, \{M_{2,k}\}_{k=1}^K, \{\Delta_{x,k}\}_{k=1}^K, \{\Delta_{y,k}\}_{k=1}^K)$ successive-refinement code are given by

$$R_{x,k} \triangleq \frac{1}{N} \log M_{1,k} \tag{18}$$

and

$$R_{y,k} \triangleq \frac{1}{N} \log M_{2,k} \tag{19}$$

for each $k = 1, 2, ..., K$. Here we adopt the incremental definitions of information rates used in [4] and [5] for one-way successive communication, which are different from those used in [8] and [9], where the achievable rate-region is given via cumulative communication rates after each iteration. The differences between these two approaches are detailed in [4], and, as is thoroughly explained in [4], the characterization of achievable region used in [4] and [5] is more suitable to describe actual systems operating over rate-limited channels (which is the case analyzed in this paper) than that of [8] and [9].

In parallel to Definition 2, we define the following:

Definition 4. *The rate-distortion $(4K)$-tuples $\{(R_{x,k}, R_{y,k}, \Delta_{x,k}, \Delta_{y,k})\}_{k=1}^K$ are said to be achievable if for every $\delta > 0$, $\epsilon > 0$, and a sufficiently large blocklength N, there exists an $(N, \{2^{N(R_{x,k}+\delta)}\}_{k=1}^K, \{2^{N(R_{y,k}+\delta)}\}_{k=1}^K, \{\Delta_{x,k}+\epsilon\}_{k=1}^K, \{\Delta_{y,k}+\epsilon\}_{k=1}^K)$, K-step source code with successive refinement for the source P_{XY}. The collection of all $\{(R_{x,k}, R_{y,k}, \Delta_{x,k}, \Delta_{y,k})\}_{k=1}^K$-achievable rate-distortion quadruples is the achievable rate-distortion region, and is denoted by $\mathcal{RD}$.*

Our first objective is to provide a single-letter characterization of $\mathcal{RD}$ and to propose strategies for (asymptotically) achieving any given point in $\mathcal{RD}$. The main objective of this work, is to provide a single-letter characterization of $\mathcal{D}$, and in particular, to show that any given point in $\mathcal{D}$ can be achieved by separate source coding of the source P_{XY}, achieving rates is $\mathcal{RD}$, used in tandem with an optimal channel code [1, 6, 7] (independently of the source).

Since the characterization of $\mathcal{RD}$ is strongly related to the coding theorem of [3], we now pause to provide a brief description of this theorem. As mentioned in the Introduction, in [3], Kaspi considers the special case of the presented model, where the channels are clean, namely, the problem of pure source coding. The distortion constraints considered in [3] are degenerated to two constraints only $(\Delta_{x,K}, \Delta_{y,K})$, i.e., the reconstruction of the source P_{XY} in each codec is performed only after K steps. In terms of our definition of the problem, the allowed distortions of [3] are given by $\Delta_{x,k} = \infty$ and $\Delta_{y,k} = \infty$ for $k = 1, ..., K-1$, and the distortions $(\Delta_{x,K}, \Delta_{y,K})$ are finite. The main result of [3] (in the notation of the present paper) is given in terms of overall compression rates achievable in the scheme,

$$R_x \triangleq \sum_{k=1}^{K} R_{x,k}, \tag{20}$$

and

$$R_y \triangleq \sum_{k=1}^{K} R_{y,k}, \tag{21}$$

rather than achievable rates in each communication step.

Theorem 1. *[3] A rate-distortion quadruple $(R_x, \Delta_{x,K}, R_y, \Delta_{y,K})$ is achievable if and only if there exists a $(2K)$-tuple of random variables $(Z_1, ..., Z_K, W_1, ..., W_K)$, taking values in finite alphabets $\mathcal{Z}_1, ..., \mathcal{Z}_K, \mathcal{W}_1, ..., \mathcal{W}_K$, respectively, such that the following conditions are satisfied:*

1. The following Markov chains are satisfied:

$$Z_1 \div X \div Y, \tag{22}$$

$$W_1 \div (Y, Z_1) \div X, \tag{23}$$

and for $k = 2, ..., K$,

$$Z_k \div (X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) \div Y, \tag{24}$$

$$W_k \div (Y, Z_1, ..., Z_k, W_1, ..., W_{k-1}) \div X. \tag{25}$$

2. There exist deterministic decoding functions $F_K : \mathcal{X} \times \mathcal{Z}_1 \times ... \times \mathcal{Z}_K \times \mathcal{W}_1 \times ... \times \mathcal{W}_K \to \hat{\mathcal{Y}}$ and $G_K : \mathcal{Y} \times \mathcal{Z}_1 \times ... \times \mathcal{Z}_K \times \mathcal{W}_1 \times ... \times \mathcal{W}_K \to \hat{\mathcal{X}}$, such that

$$Ed_{x,K}(X, G_K(Y, Z_1, ..., Z_K, W_1, ..., W_K)) \leq \Delta_{x,K}, \tag{26}$$

and

$$Ed_{y,K}(Y, F_K(X, Z_1, ..., Z_K, W_1, ..., W_K)) \leq \Delta_{y,K}. \tag{27}$$

3. The rates R_x and R_y satisfy

$$R_x \geq I(X; Z_1, ..., Z_K, W_1, ...W_K|Y), \tag{28}$$

and

$$R_y \geq I(Y; Z_1, ..., Z_K, W_1, ...W_K|X). \tag{29}$$

The necessary sizes of the alphabets of $\{Z_k\}_{k=1}^K$ and $\{W_k\}_{k=1}^K$ are not specified in [3].

4 Main Result

In this section, we consider the problem of two-way joint source-channel coding with a fidelity criterion. For the clarity of exposition, we begin with a pure source coding problem and then extend the setup to include communication over two memoryless rate-limited channels, and formulate coding theorems for each of the cases.

4.1 Source Coding with Successive Refinement

In this subsection, we give a single-letter characterization of $\mathcal{RD}$ for a given source P_{XY} and a K-step two-way interactive coding. Let a distortion $(2K)$-tuple $\boldsymbol{D} \stackrel{\triangle}{=} \{(\Delta_{x,k}, \Delta_{y,k})\}_{k=1}^K$ be given. Define $\mathcal{R}^*(\boldsymbol{D})$ to be the set of $(2K)$ rate-tuples $\{(R_{x,k}, R_{y,k})\}_{k=1}^K$ for which there exists a $(2K)$-tuple of random variables $(Z_1, ..., Z_K, W_1, ..., W_K)$, taking values in finite alphabets, $\mathcal{Z}_1, ..., \mathcal{Z}_K, \mathcal{W}_1, ..., \mathcal{W}_K$, respectively, such that the following conditions are satisfied:

1. The following Markov chains are satisfied:

$$Z_1 \div X \div Y, \tag{30}$$

$$W_1 \div (Y, Z_1) \div X, \tag{31}$$

and for $k = 2, ..., K$,

$$Z_k \div (X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) \div Y, \tag{32}$$

and

$$W_k \div (Y, Z_1, ..., Z_k, W_1, ..., W_{k-1}) \div X. \tag{33}$$

2. There exist $2K$ deterministic decoding functions for $k = 1, 2, ..., K$, $F_k : \mathcal{X} \times \mathcal{Z}_1 \times ... \times \mathcal{Z}_k \times \mathcal{W}_1 \times ... \times \mathcal{W}_k \to \hat{\mathcal{Y}}$ and $G_k : \mathcal{Y} \times \mathcal{Z}_1 \times ... \times \mathcal{Z}_k \times \mathcal{W}_1 \times ... \times \mathcal{W}_k \to \hat{\mathcal{X}}$, such that

$$Ed_{x,k}(X, G_k(Y, Z_1, ..., Z_k, W_1, ..., W_k)) \leq \Delta_{x,k}, \tag{34}$$

and

$$Ed_{y,k}(Y, F_k(X, Z_1, ..., Z_k, W_1, ..., W_k)) \leq \Delta_{y,k}. \tag{35}$$

3. The alphabets $(Z_1, ..., Z_K, W_1, ..., W_K)$ satisfy

$$|\mathcal{Z}_k| \leq (|\mathcal{X}| \cdot |\mathcal{Y}| + 2)^{2k-1} \tag{36}$$

and

$$|\mathcal{W}_k| \leq (|\mathcal{X}| \cdot |\mathcal{Y}| + 2)^{2k}. \tag{37}$$

4. The rates $R_{x,k}$ and $R_{y,k}$ satisfy

$$R_{x,k} \geq I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ...W_{k-1}), \tag{38}$$

and

$$R_{y,k} \geq I(Y; Z_k, W_k | X, Z_1, ..., Z_{k-1}, W_1, ...W_{k-1}). \tag{39}$$

The main result of this subsection is the following:

Theorem 2. *For any DMS P_{XY},*

$$\mathcal{R}(\mathbf{D}) = \mathcal{R}^*(\mathbf{D}). \tag{40}$$

The proof of Theorem 2 appears in Section 5.

Discussion:

First, let us notice that the lower bounds to per-step compression rates $R_{x,k}$ and $R_{y,k}$ given in (38) and (39) equal to

$$R_{x,k} \geq I(X; Z_k | Y, Z_1, ..., Z_{k-1}, W_1, ...W_{k-1}), \tag{41}$$

and

$$R_{y,k} \geq I(Y; W_k | X, Z_1, ..., Z_k, W_1, ...W_{k-1}), \tag{42}$$

as immediately follows from definition of the Markov structures (30)-(33). We provide the compression rates in the form of (38)-(39) in order to emphasize the relation of Theorem 2 to Theorem 1. Specifically, by omitting all the intermediate distortion constraints $\{(\Delta_{x,k}, \Delta_{y,k})\}_{k=1}^{K-1}$ of Theorem 2 (for example, by allowing them all to be infinite), we obtain that Theorem 2 degenerates to Theorem 1, since the cumulative coding rates after K transmissions equal R_x and R_y, as defined in (20) and (21), respectively. Our characterization is more suitable to describe actual systems than that of [3]. Specifically, in the next subsection, we consider a case where the bitstreams of each of the communication steps are transmitted via two separate rate-limited channels. Then, the coding rates of each step are separately compared to the capacities of the corresponding channels, an analysis which is not possible with the original problem characterization of [3].

The results of Theorem 2 relate directly to the problem of one-way successive refinement for the Wyner-Ziv problem, investigated in [4]. Specifically, for the setting defined in this paper, assume that only codec 1 transmits information to codec 2 and codec 2 is "silent". This scenario can be modelled by taking $R_{y,k} = 0$ and allowing $\Delta_{y,k}$ to be infinite for all $k = [1, 2, ..., K]$. Then, we end up with a one-way communication scheme with successive refinement from codec 1 to codec 2, with identical side information Y available at all refinement stages [4]. Note that here, codec 2 represents solely all the per-stage decoders of [4]. Since we can satisfy the above $R_{y,k}$ and $\Delta_{y,k}$ by choosing all auxiliary random variables $\{W_k\}_{k=1}^{K}$ to be constants, under the constraints imposed on $\{Z_k\}_{k=1}^{K}$, we obtain that $\mathcal{R}(\boldsymbol{D})$ turns out to be that of [4].

4.2 Joint Source-Channel Coding

We now present the main result of this paper, which is a necessary and sufficient condition for $\{\Delta_{x,k}, \Delta_{y,k}\}_{k=1}^{K}$ to be the achievable distortion levels of the scheme depicted in Fig. 1.

Theorem 3. *Given a DMS P_{XY}, the distortion levels $\{\Delta_{x,k}, \Delta_{y,k}\}_{k=1}^{K}$ are achievable for the two-way K-step communication over noisy stationary memoryless channels $P_{V|U}$ and $P_{B|A}$ if and only if there exist auxiliary RVs $\{Z_i\}_{i=1}^{K}$ and $\{W_i\}_{i=1}^{K}$, taking values in finite alphabets $\{\mathcal{Z}_i\}_{i=1}^{K}$ and $\{\mathcal{W}_i\}_{i=1}^{K}$, of cardinality denoted by (36) and (37), respectively, and satisfying (30) - (33), and deterministic decoding functions $\{F_k\}_{k=1}^{K}$ and $\{G_k\}_{k=1}^{K}$, satisfying (34) and (35), respectively, such that for all $k = 1, 2, ..., K$,*

$$\rho_s I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ...W_{k-1}) \leq \frac{\rho_{c_1}}{\beta_{1,k}} C_1, \tag{43}$$

$$\rho_s I(Y; Z_k, W_k | X, Z_1, ..., Z_{k-1}, W_1, ...W_{k-1}) \leq \frac{\rho_{c_2}}{\beta_{2,k}} C_2. \tag{44}$$

The similarity between the characterization of the region of achievable distortion levels of Theorem 3 and the characterization of $\mathcal{D}$ is self evident. In fact, the only difference is that in each communication step k, the coding rates $R_{x,k}$ and $R_{y,k}$ of the former are replaced by $\frac{\rho_{c_1}}{\rho_s \beta_{1,k}} C_1$ and $\frac{\rho_{c_2}}{\rho_s \beta_{2,k}} C_2$, respectively. The immediate conclusion from this observation is that the separation principle applies to our model.

The proof of the converse part of Theorem 3 appears in Section 5. The proof of direct part comes directly from considering an asymptotically optimum K-step two-way source code (independent of the channel) followed by a reliable transmission code for each of the channels (independent of the source), i.e., separate source and channel coding. If the distortion level of the two-way source code, presented in Section 4.1, is chosen such that $\rho_s R_{x,k} < \frac{\rho_{c_1}}{\beta_{1,k}} C_1$ and $\rho_s R_{y,k} < \frac{\rho_{c_2}}{\beta_{2,k}} C_2$, one may select constants $R_{s,x,k}$, $R_{c,x,k}$, $R_{s,y,k}$ and $R_{c,y,k}$ such that for every $k = [1, 2, ..., K]$

$$N R_{x,k} < N R_{s,x,k} = m_{1,k} R_{c,x,k} < m_{1,k} C_1 \tag{45}$$

and

$$N R_{y,k} < N R_{s,y,k} = m_{2,k} R_{c,y,k} < m_{2,k} C_2. \tag{46}$$

In each communication step, each codec may then compress the information about X (Y) into $R_{s,x,k}$ ($R_{s,y,k}$) bits per symbol within distortion $\Delta_{x,k}$ ($\Delta_{y,k}$), and then map the resulting $NR_{s,x,k}$-bit ($NR_{s,y,k}$-bit) codewords into channel codewords of the same number of bits $m_{1,k} R_{c,x,k} < m_{1,k} C_1$ ($m_{2,k} R_{c,y,k} < m_{2,k} C_2$). Since $R_{c,x,k} < C_1$ and $R_{c,y,k} < C_2$, there exist reliable channel codes which cause asymptotically negligible additional distortion. Since $\boldsymbol{D}$ can be chosen in this way, such that $\rho_s R_{x,k}$ is arbitrary close to $\frac{\rho_{c_1}}{\beta_{1,k}} C_1$ and $\rho_s R_{y,k}$ to $\frac{\rho_{c_2}}{\beta_{2,k}} C_2$, all distortion levels for which $\rho_s R_{x,k} < \frac{\rho_{c_1}}{\beta_{1,k}} C_1$ and $\rho_s R_{y,k} < \frac{\rho_{c_2}}{\beta_{2,k}} C_2$ are achievable.

5 Proofs

The pure source-coding problem is a special case of the joint source-channel problem. We begin with the direct part of the proof of Theorem 2. We then provide a proof of the converse part of Theorem 3, which includes the converse of Theorem 2 as a special case. Nonetheless, we emphasize the nuances which are special for the noise-free case.

5.1 Proof of the Direct Part of Theorem 2

The proof of the direct part follows the lines of the proof provided in [3]. Given some constants $\epsilon > 0$ and $\delta > 0$, we require the deterministic functions $\{F_k, G_k\}_{k=1}^K$ and RVs X, Y, $\{Z_k\}_{k=1}^K$ and $\{W_k\}_{k=1}^K$ (all of finite cardinality) to satisfy a single-letter distortion constraints (34) and (35) in each step $k = 1, ..., K$. We use the definitions of the variables and functions of [3]. The only difference in notations is that of denoting indexes as a subscript rather than superscript, for example, auxiliary RVs of [3] $\{Z^k\}_{k=1}^K$ and $\{W^k\}_{k=1}^K$ are referred to in our proof as $\{Z_k\}_{k=1}^K$ and $\{W_k\}_{k=1}^K$.

It was shown in [3], that the probability of failure in the coding scheme proposed therein (asymptotically) vanishes. The outline of the coding algorithm is as follows: A random coding tree is generated and kept in both codecs. After codec 1 receives an N-vector $\boldsymbol{x}$ and codec 2 receives an N-vector $\boldsymbol{y}$, the two codecs follow a common path from the root to a leaf. Suppose that just before the k-th transmission from codec 1, both codecs are on a node $2(k-1)$ branches from the root. Codec 1 makes a decision upon the next node and transmits it to codec 2 using $NR_{x,k}$ bits. Both sides move to the next node and codec 2 makes and transmits the next decision using $NR_{y,k}$ bits. The reproduction vectors $\hat{\boldsymbol{X}}_k$ and $\hat{\boldsymbol{Y}}_k$ are generated after both codec 1 and codec 2 make their move to the next nodes.

We show that under the additional constraint added to the original problem of [3], over the ensemble of code trees, at each communication step k, the expected distortions can be made arbitrary close to $\{\Delta_{x,k} + \delta\}_{k=1}^K$, $\{\Delta_{y,k} + \delta\}_{k=1}^K$ and the compression rates can approach $\{R_{x,k} + \epsilon\}_{k=1}^K$ and $\{R_{y,k} + \epsilon\}_{k=1}^K$ as close as desired.

The proof of the achievability of coding rates is based on the properties of δ-strongly joint typicality [7]. It is similar to this of [3] and thus omitted. We rewrite the results of [3] using the Markov chains (30)-(33): The value of $R_{x,k}$ in [3], given by eq. (3.7), equals

$$\begin{aligned} R_{x,k} &= I(X; Z_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) + 2\alpha \\ &= I(X; Z_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) + I(X; W_k | Y, Z_1, ..., Z_k, W_1, ..., W_{k-1}) + 2\alpha \\ &= I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) + 2\alpha, \end{aligned} \tag{47}$$

as immediately follows from the Markov structures (31) and (33). In terms of [3], the component 2α is a payoff for using typical sequences. Similarly, the value of $R_{y,k}$ in [3], is given by

$$
\begin{aligned}
R_{y,k} &= I(Y;W_k|X,Z_1,...,Z_k,W_1,...,W_{k-1}) + 2\widetilde{\alpha} \\
&= I(Y;W_k|X,Z_1,...,Z_k,W_1,...,W_{k-1}) + I(Y;Z_k|X,Z_1,...,Z_{k-1},W_1,...,W_{k-1}) + 2\widetilde{\alpha} \\
&= I(Y;Z_k,W_k|X,Z_1,...,Z_{k-1},W_1,...,W_{k-1}) + 2\widetilde{\alpha}, \qquad (48)
\end{aligned}
$$

as follows from the Markov chains (30) and (32), and $2\widetilde{\alpha}$ is a payoff due to typicality.

To complete the proof, it remains to show that the distortion constraints (34) and (35) are satisfied. We do this by upper bounding the average distortions in each step $k = 1, 2, ..., K$ by means of the typicality properties of the scheme. Specifically:

$$
\begin{aligned}
Ed_{x,k}(\boldsymbol{x}, \hat{\boldsymbol{x}}_k) &= \frac{1}{N}\sum_{i=1}^{N} \sum_{x,y,\{z_{k'},w_{k'}\}_{k'=1}^{k}} \Pr\{(x_i, y_i, \{z_{k',i}, w_{k',i}\}_{k'=1}^{k}) = (x, y, \{z_{k'}, w_{k'}\}_{k'=1}^{k})\} \\
&\cdot \ d_{x,k}(x, F_k(x, \{z_{k'}, w_{k'}\}_{k'=1}^{k})) \qquad (49) \\
&\leq \sum_{x,y,\{z_{k'},w_{k'}\}_{k'=1}^{k}} \Pr\{(x, y, \{z_{k'}, w_{k'}\}_{k'=1}^{k})\} \cdot d_{x,k}(x, \hat{x}_k) + \delta \max_{x,\hat{x}_k}\{d_{x,k}(x, \hat{x}_k)\} \\
&\leq \Delta_{x,k} + \delta \max_{x,\hat{x}_k}\{d_{x,k}(x, \hat{x}_k)\},
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
Ed_{y,k}(\boldsymbol{y}, \hat{\boldsymbol{y}}_k) &= \frac{1}{N}\sum_{i=1}^{N} \sum_{x,y,\{z_{k'},w_{k'}\}_{k'=1}^{k}} \Pr\{(x_i, y_i, \{z_{k',i}, w_{k',i}\}_{k'=1}^{k}) = (x, y, \{z_{k'}, w_{k'}\}_{k'=1}^{k})\} \\
&\cdot \ d_{y,k}(y, G_k(y, \{z_{k'}, w_{k'}\}_{k'=1}^{k})) \qquad (50) \\
&\leq \sum_{x,y,\{z_{k'},w_{k'}\}_{k=1}^{k}} \Pr\{(x, y, \{z_{k'}, w_{k'}\}_{k'=1}^{k})\} \cdot d_{y,k}(y, \hat{y}_k) + \delta \max_{y,\hat{y}_k}\{\rho_{y,k}(y, \hat{y}_k)\} \\
&\leq \Delta_{y,k} + \delta \max_{y,\hat{y}_k}\{d_{y,k}(y, \hat{y}_k)\},
\end{aligned}
$$

which completes the proof of the direct part.

5.2 Proof of the Converse Part of Theorem 3

In order to prove the converse part, we first prove the following lemma which extends the results of Lemma 1 of [3] to stochastic functions:

Lemma 1. *Let J_1, J_2, E_1 and E_2 be random variables with joint probability $P(J_1, J_2, E_1, E_2) = P(J_1, E_1)P(J_2, E_2)$ and suppose that $\{\mathbf{V}_k\}$ and $\{\mathbf{B}_k\}$, $k = [1, 2, ...]$, are any stochastic functions with domain structure $\mathbf{V}_k(J_1, J_2, \mathbf{B}_1, ...\mathbf{B}_{k-1})$ and $\mathbf{B}_k(E_1, E_2, \mathbf{V}_1, ...\mathbf{V}_k)$. Then*

$$I(J_2; E_1|\mathbf{V}_1, ..., \mathbf{V}_k, \mathbf{B}_1, ..., \mathbf{B}_k, J_1, E_2) = 0. \tag{51}$$

Proof of Lemma 1

In the sequel we use the convention that $\boldsymbol{V}_0 = 0$ and $\boldsymbol{B}_0 = 0$. Then,

$$\begin{aligned}
0 \quad &\leq \quad I(J_2; E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k, J_1, E_2) && (52)\\
&= \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k, J_1, E_2) && (53)\\
&- \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k, J_1, E_1, E_2) && (54)\\
&= \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (55)\\
&- \quad I(J_2; \boldsymbol{B}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (56)\\
&- \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_1, E_2) && (57)\\
&+ \quad I(J_2; \boldsymbol{B}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_1, E_2) && (58)\\
&\overset{(a)}{=} \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (59)\\
&- \quad I(J_2; \boldsymbol{B}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (60)\\
&- \quad H(J_2| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_1, E_2) && (61)\\
&= \quad I(J_2; E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (62)\\
&- \quad I(J_2; \boldsymbol{B}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (63)\\
&\leq \quad I(J_2; E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (64)\\
&= \quad H(E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (65)\\
&- \quad I(E_1;\, \boldsymbol{V}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (66)\\
&- \quad H(E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, J_2, E_2) && (67)\\
&+ \quad I(E_1;\, \boldsymbol{V}_k| \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, J_2, E_2) && (68)\\
&\overset{(b)}{\leq} \quad I(J_2; E_1| \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_2) && (69)\\
& \quad\quad ... && (70)\\
&\leq \quad I(J_2; E_1|J_1, E_2) && (71)\\
&\overset{(c)}{=} \quad 0, && (72)
\end{aligned}$$

where:

(a) follows from the Markov chain

$J_2 \div (J_1, E_1, E_2, \boldsymbol{V}_1, ... \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \div \boldsymbol{B}_k$, which implies

$I(J_2; \boldsymbol{B}_k | \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, E_1, E_2) = 0,$

(b) from the Markov chain

$E_1 \div (J_1, J_2, E_2, \boldsymbol{V}_1, ... \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \div \boldsymbol{V}_k,$ which implies

$I(E_1; \boldsymbol{V}_k | \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}, J_1, J_2, E_2) = 0,$

and from non-negativity of mutual information, and

(c) from independence of (J_1, E_1) and (J_2, E_2).

□

Similarly as in [3], the proof of Lemma 1 provides two immediate conclusions:

Corollary 1.

$$I(E_1; \mathbf{V}_k | \mathbf{V}_1, ..., \mathbf{V}_{k-1}, \mathbf{B}_1, ..., \mathbf{B}_{k-1}, J_1, E_2) = 0, \tag{73}$$

and

$$I(J_2; \mathbf{B}_k | \mathbf{V}_1, ..., \mathbf{V}_k, \mathbf{B}_1, ..., \mathbf{B}_{k-1}, J_1, E_2) = 0. \tag{74}$$

We proceed now with the proof of the converse part of Theorem 3. Let $(\{f_k, g_k, \tilde{F}_k, \tilde{G}_k\}_{k=1}^K)$ be given encoder and decoder functions for which $Ed_{x,k}(\boldsymbol{X}, \hat{\boldsymbol{X}}_k) \le N\Delta_{x,k}$ and $Ed_{y,k}(\boldsymbol{Y}, \hat{\boldsymbol{Y}}_k) \le N\Delta_{y,k}$ for each $k = 1, 2, ..., K$. In the proof, for the k-th step of the communication protocol, we examine the mutual informations $I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k)$ and $I(\boldsymbol{Y}, \boldsymbol{V}_1, ..., \boldsymbol{V}_k; \boldsymbol{B}_k)$. We focus on the k-th transmission from codec 1 to codec 2 only, as the proof of the complementary transmission from codec 2 to codec 1 exactly the same.

Due to the physical structure of the communication scheme, there exists a Markov chain $(\boldsymbol{X}, \boldsymbol{B}_1, ...\boldsymbol{B}_{k-1}) \div \boldsymbol{U}_k \div \boldsymbol{V}_k$, and therefore, by the data processing inequality, we obtain

$$I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) \le I(\boldsymbol{U}_k; \boldsymbol{V}_k) \overset{(a)}{\le} m_{1,k} C_1, \tag{75}$$

where (a) follows from the capacity formula for a stationary memoryless channel [1]. On the other hand,

$$I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) \tag{76}$$

$$\overset{(a)}{=} I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) + I(\boldsymbol{Y}; \boldsymbol{V}_k | \boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \tag{77}$$

$$= I(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) \tag{78}$$

$$\overset{(b)}{=} I(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) \tag{79}$$

$$\geq I(\boldsymbol{X};\, \boldsymbol{V}_k|\, \boldsymbol{Y},\, \boldsymbol{V}_1, ...,\, \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \tag{80}$$

$$\stackrel{(c)}{=} I(\boldsymbol{X};\, \boldsymbol{V}_k, \boldsymbol{B}_k|\, \boldsymbol{Y},\, \boldsymbol{V}_1, ...,\, \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \tag{81}$$

$$= \sum_{i=1}^{N} I(X_i;\, \boldsymbol{V}_k, \boldsymbol{B}_k|X_1^{i-1},\, \boldsymbol{Y},\, \boldsymbol{V}_1, ...,\, \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}), \tag{82}$$

where:

(a) follows from the Markov chain $\boldsymbol{Y} \div (\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \div \boldsymbol{V}_k$,

which implies $I(\boldsymbol{Y};\, \boldsymbol{V}_k|\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) = 0$,

(b) from the Markov chain

$\boldsymbol{V}_k \div (\boldsymbol{X},\, \boldsymbol{Y}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \div (\boldsymbol{V}_1, ...,\, \boldsymbol{V}_{k-1})$, and

(c) from the Markov chain

$\boldsymbol{X} \div (\boldsymbol{Y},\, \boldsymbol{V}_1, ...,\, \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) \div \boldsymbol{B}_k$.

Defining auxiliary RVs:

$$Z_{1,i} \stackrel{\triangle}{=} (\boldsymbol{V}_1, X_1^{i-1}, Y_{i+1}^N), \tag{83}$$

for $k = 2, ..., K$

$$Z_{k,i} \stackrel{\triangle}{=} \boldsymbol{V}_k, \tag{84}$$

and for $k = 1, ..., K$

$$W_{k,i} \stackrel{\triangle}{=} \boldsymbol{B}_k, \tag{85}$$

we obtain from (76)-(82) the following lower bound to $I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1};\, \boldsymbol{V}_k)$:

$$I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1};\, \boldsymbol{V}_k) \tag{86}$$

$$\geq \sum_{i=1}^{N} I(X_i; Z_{k,i}, W_{k,i}|Y_i, Y_1^{i-1}, Z_{1,i}, ..., Z_{k-1,i}, W_{1,i}, ..., W_{k-1,i}) \tag{87}$$

$$= \sum_{i=1}^{N} \big[I(X_i; Z_{k,i}, W_{k,i}|Y_i, Z_{1,i}, ..., Z_{k-1,i}, W_{1,i}, ..., W_{k-1,i}) \tag{88}$$

$$+ I(X_i; Y_1^{i-1}|Y_i, Z_{1,i}, ..., Z_{k,i}, W_{1,i}, ..., W_{k,i}) \tag{89}$$

$$- I(X_i; Y_1^{i-1}|Y_i, Z_{1,i}, ..., Z_{k-1,i}, W_{1,i}, ..., W_{k-1,i}) \big]. \tag{90}$$

Now, applying Lemma 1, we can show that the values of (89) and (90) are zero. Specifically, let us examine (89):

$$\begin{aligned}
0 &\leq I(X_i; Y_1^{i-1} | Y_i, Z_{1,i}, ..., Z_{k,i}, W_{1,i}, ..., W_{k,i}) &(91)\\
&\leq I(X_i, X_{i+1}^N; Y_1^{i-1} | Y_i, Z_{1,i}, ..., Z_{k,i}, W_{1,i}, ..., W_{k,i}) &(92)\\
&= I(X_i, X_{i+1}^N; Y_1^{i-1} | Y_i, X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k) &(93)\\
&= 0, &(94)
\end{aligned}$$

where in the last step, we have used Lemma 1 with the substitutions $J_1 \triangleq X_1^{i-1}$, $J_2 \triangleq (X_i, X_{i+1}^N)$, $E_1 \triangleq Y_1^{i-1}$ and $E_2 \triangleq (Y_1, Y_{i+1}^N)$. The proof of equality of (90) to zero follows the same lines. Similarly, the choice of $J_1 \triangleq (X_i, X_1^{i-1})$, $J_2 \triangleq X_{i+1}^N$, $E_1 \triangleq (Y_i, Y_1^{i-1})$ and $E_2 \triangleq Y_{i+1}^N$ serves to prove the rate constraints of the complementary k-th transmission from codec 2 to codec 1. And so, we obtain

$$I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) \geq \sum_{i=1}^{N} \big[I(X_i; Z_{k,i}, W_{k,i} | Y_i, Z_{1,i}, ..., Z_{k-1,i}, W_{1,i}, ..., W_{k-1,i}) \big]. \quad (95)$$

Consider now a time-sharing random variable S, distributed uniformly over $\{1, 2, ..., N\}$, independently of all other random variables in the system, and let us denote $2(k+1)$ random variables:

$$(X, Y, \widetilde{Z}_1, Z_2, ..., Z_K, W_1, ..., W_K) \triangleq (X_S, Y_S, Z_{1,S}, Z_{2,S}, ..., Z_{K,S}, W_{1,S}, ..., W_{K,S}). \quad (96)$$

By denoting $Z_1 \triangleq (\widetilde{Z}_1, S)$, and combining the definitions of (96) with the result of (75), we obtain that

$$\begin{aligned}
m_{1,k} C_1 &\geq I(\boldsymbol{X}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}; \boldsymbol{V}_k) &(97)\\
&\geq N I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}). &(98)
\end{aligned}$$

Of course, a similar analysis can be performed for $I(\boldsymbol{Y}, \boldsymbol{V}_1, ..., \boldsymbol{V}_k; \boldsymbol{B}_k)$, showing that

$$\begin{aligned}
m_{2,k} C_2 &\geq I(\boldsymbol{Y}, \boldsymbol{V}_1, ..., \boldsymbol{V}_k; \boldsymbol{B}_k) &(99)\\
&\geq N I(Y; Z_k, W_k | X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}). &(100)
\end{aligned}$$

Now, from (4) and (5), it follows that

$$\rho_s I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) \leq \frac{\rho_{c_1}}{\beta_{1,k}} C_1 \quad (101)$$

and

$$\rho_s I(Y; Z_k, W_k | X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) \leq \frac{\rho_{c_2}}{\beta_{2,k}} C_2. \quad (102)$$

We pause to adjust the above bounds to prove the necessity part of Theorem 2. In the noise-free case, $f_k = \boldsymbol{U}_k = \boldsymbol{V}_k$ and $g_k = \boldsymbol{A}_k = \boldsymbol{A}_k$. Thus, for each $k = 1, ..., K$, we can lower-bound the intermediate compression rates as follows:

$$\begin{aligned} R_{x,k} &\geq \frac{1}{N} H(\boldsymbol{U}_k) \\ &\geq \frac{1}{N} I(\boldsymbol{X}, \boldsymbol{A}_1, ..., \boldsymbol{A}_{k-1}; \boldsymbol{U}_k) \\ &\geq I(X; Z_k, W_k | Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) \end{aligned} \tag{103}$$

and

$$\begin{aligned} R_{y,k} &\geq \frac{1}{N} H(\boldsymbol{A}_k) \\ &\geq \frac{1}{N} I(\boldsymbol{Y}, \boldsymbol{U}_1, ..., \boldsymbol{U}_k; \boldsymbol{A}_k) \\ &\geq I(Y; Z_k, W_k | X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}), \end{aligned} \tag{104}$$

with the auxiliary RVs from (83)-(85) obtaining the form of

$$Z_{1,i} \stackrel{\triangle}{=} (\boldsymbol{U}_1, X_1^{i-1}, Y_{i+1}^N), \quad Z_{2,i} \stackrel{\triangle}{=} \boldsymbol{U}_2, \quad ... \; Z_{K,i} \stackrel{\triangle}{=} \boldsymbol{U}_K \tag{105}$$

and

$$W_{1,i} \stackrel{\triangle}{=} \boldsymbol{A}_1, \quad W_{2,i} \stackrel{\triangle}{=} \boldsymbol{A}_2, \quad ... \; W_{K,i} \stackrel{\triangle}{=} \boldsymbol{A}_K. \tag{106}$$

The proof that the distortion constraints are met, is based on the fact that in each communication step k, the RVs $\{Z_{k'}\}_{k'=1}^k$ and $\{W_{k'}\}_{k'=1}^k$ contain $(\{\boldsymbol{V}_{k'}\}_{k'=1}^k, X_1^{i-1}, Y_{i+1}^N)$ and $(\{\boldsymbol{B}_{k'}\}_{k'=1}^k, X_1^{i-1}, Y_{i+1}^N)$, respectively. Therefore, we choose following functions F_k and G_k:

$$F_k(Z_1, ..., Z_k, W_1, ..., W_k, X) = \widetilde{F}_{k,S}(\boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k, \boldsymbol{X}), \tag{107}$$

$$G_k(Z_1, ..., Z_k, W_1, ..., W_k, Y) = \widetilde{G}_{k,S}(\boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_k, \boldsymbol{Y}), \tag{108}$$

where, after the k-th step, $\widetilde{F}_{k,S}$ and $\widetilde{G}_{k,S}$ denote the outputs of the Decoders in Codecs 1 and 2, respectively, at time S. Then, for the distortions we have

$$\begin{aligned} &Ed_{x,k}(X, F_k(Z_1, ..., Z_k, W_1, ..., W_k, X)) = \\ &= \frac{1}{N} \sum_{s=1}^{N} Ed_{x,k}(X, F_{k,s}(\boldsymbol{V}(1), ..., \boldsymbol{V}(k), \boldsymbol{B}(1), ..., \boldsymbol{B}(k), \boldsymbol{X})) \leq \Delta_{x,k}, \end{aligned} \tag{109}$$

and similarly,

$$\begin{aligned}
&Ed_{y,k}(Y, G_k(Z_1, ..., Z_k, W_1, ..., W_k, Y)) = \qquad (110)\\
&= \frac{1}{N}\sum_{s=1}^{N} Ed_{y,k}(Y, G_{k,s}(\boldsymbol{V}(1), ..., \boldsymbol{V}(k), \boldsymbol{B}(1), ..., \boldsymbol{B}(k), \boldsymbol{Y})) \leq \Delta_{y,k}.
\end{aligned}$$

It remains to show that the above defined RVs maintain the Markov structures (30) - (33) and that their alphabet sizes are as specified. To prove the Markov structures between $\boldsymbol{X}$, $\boldsymbol{Y}$ and appropriate $\{Z_k\}_{k=1}^K$ and $\{W_k\}_{k=1}^K$, it is enough to prove these for every $i = 1, 2, ..., N$ among X_i, Y_i, $\{Z_{k,i}\}_{k=1}^K$ and $\{W_{k,i}\}_{k=1}^K$. We begin with the proof of $Z_{1,i} \div X_i \div Y_i$, (30):

$$\begin{aligned}
0 &\leq I(Z_{1,i}; Y_i|X_i) && (111)\\
&= I(X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1; Y_i|X_i) && (112)\\
&\leq I(X_1^{i-1}, X_{i+1}^N, Y_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1; Y_i|X_i) && (113)\\
&= I(X_1^{i-1}, X_{i+1}^N, Y_1^{i-1}, Y_{i+1}^N; Y_i|X_i) + I(\boldsymbol{V}_1; Y_i|X_i, X_1^{i-1}, X_{i+1}^N, Y_1^{i-1}, Y_{i+1}^N) && (114)\\
&= 0, && (115)
\end{aligned}$$

where the last equality is because the sources are memoryless, which implies

$$I(X_1^{i-1}, X_{i+1}^N, Y_1^{i-1}, Y_{i+1}^N; Y_i|X_i) = 0, \qquad (116)$$

and because of the Markov chain $Y_i \div (\boldsymbol{X}, Y_1^{i-1}, Y_{i+1}^N) \div \boldsymbol{V}_1$. Similarly, (31) is proved as follows:

$$\begin{aligned}
0 &\leq I(W_{1,i}; X_i|Z_{1,i}, Y_i) && (117)\\
&\leq I(W_{1,i}; X_i, X_{i+1}^N|Z_{1,i}, Y_i) && (118)\\
&= I(\boldsymbol{B}_1; X_i, X_{i+1}^N|X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, Y_i) = 0, && (119)
\end{aligned}$$

due to Corollary 1, (74), for $J_1 \triangleq X_1^{i-1}$, $J_2 \triangleq (X_i, X_{i+1}^N)$, $E_1 \triangleq Y_1^{i-1}$, $E_2 \triangleq (Y_i, Y_{i+1}^N)$.

For $k = 2, ...K$, from Corollary 1, eq. (73), we obtain that for the choice $J_1 \triangleq (X_i, X_1^{i-1})$, $J_2 \triangleq X_{i+1}^N$, $E_1 \triangleq (Y_i, Y_1^{i-1})$, $E_2 \triangleq Y_{i+1}^N$, we have:

$$\begin{aligned}
0 &\leq I(\boldsymbol{V}_k; Y_i|X_i, X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) && (120)\\
&\leq I(\boldsymbol{V}_k; Y_i, Y_1^{i-1}|X_i, X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, ..., \boldsymbol{V}_{k-1}, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) && (121)\\
&= 0. && (122)
\end{aligned}$$

Finally, due to Corollary 1, eq. (74), for $J_1 \triangleq X_1^{i-1}$, $J_2 \triangleq (X_i, X_{i+1}^N)$, $E_1 \triangleq Y_1^{i-1}$, $E_2 \triangleq (Y_i, Y_{i+1}^N)$, we obtain

$$\begin{aligned} 0 &\leq I(\boldsymbol{B}_k; X_i|Y_i, X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) && (123)\\ &\leq I(\boldsymbol{B}_k; X_i, X_{i+1}^N|Y_i, X_1^{i-1}, Y_{i+1}^N, \boldsymbol{V}_1, ..., \boldsymbol{V}_k, \boldsymbol{B}_1, ..., \boldsymbol{B}_{k-1}) && (124)\\ &= 0. && (125) \end{aligned}$$

In order to complete the proof, it is left to show that the cardinality of the alphabets of auxiliary RVs $\{Z_k\}_{k=1}^K$ and $\{W_k\}_{k=1}^K$ is limited. To this end, we will use the support lemma [6], which is based on Carathéodory's theorem, according to which, given J real valued continuous functionals q_j, $j = 1, ..., J$ on the set $\mathcal{P}(\mathcal{X})$ of probability distributions over the alphabets $\mathcal{X}$, and given any probability measure μ on the Borel σ-algebra of $\mathcal{P}(\mathcal{X})$, there exist J elements $Q_1, ...Q_J$ of $\mathcal{P}(\mathcal{X})$ and J non-negative reals, $\alpha_1, ..., \alpha_J$, such that $\sum_{j=1}^J \alpha_j = 1$ and for every $j = 1, ..., J$

$$\int_{\mathcal{P}(\mathcal{X})} q_j(Q)\mu(dQ) = \sum_{i=1}^J \alpha_i q_j(Q_i). \qquad (126)$$

Before we actually apply the support lemma, we first rewrite the relevant conditional mutual informations and the distortion functions in a more convenient form for the use of this lemma, by taking advantage of the Markov structures. We begin with the lower bound to $R_{x,1}$:

$$\begin{aligned} I(X; Z_1, W_1|Y) &\overset{(a)}{=} I(X; Z_1|Y) && (127)\\ &= H(Z_1|Y) - H(Z_1|X, Y) && (128)\\ &\overset{(b)}{=} H(Z_1|Y) - H(Z_1|X) && (129)\\ &= H(Z_1, Y) - H(Y) - H(Z_1, X) + H(X) && (130)\\ &= H(Z_1) + H(Y|Z_1) - H(Y) - H(Z_1) - H(X|Z_1) + H(X) && (131)\\ &= \big[H(X) - H(Y)\big] + \big[H(Y|Z_1) - H(X|Z_1)\big], && (132) \end{aligned}$$

where, (a) follows from the Markov chain $W_1 \div (Y, Z_1) \div X$, and (b) from the Markov chain $Z_1 \div X \div Y$. In the same manner, the lower bound to $R_{y,1}$ obtains a form of

$$\begin{aligned} I(Y; Z_1, W_1|X) &\overset{(a)}{=} I(Y; W_1|X, Z_1) && (133)\\ &= H(W_1|X, Z_1) - H(W_1|X, Y, Z_1) && (134) \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=} \quad H(W_1|X, Z_1) - H(W_1|Y, Z_1) && (135)\\
&= \quad H(W_1, Z_1, X) - H(X, Z_1) - H(W_1, Z_1, Y) + H(Y, Z_1) && (136)\\
&= \quad H(W_1) + H(X, Z_1|W_1) - H(X, Z_1) - \\
&- \quad H(W_1) - H(Y, Z_1|W_1) + H(Y, Z_1) && (137)\\
&= \quad H(Z_1|W_1) + H(X|Z_1, W_1) - H(Z_1) - H(X|Z_1) - \\
&- \quad H(Z_1|W_1) - H(Y|Z_1, W_1) + H(Z_1) + H(Y|Z_1) && (138)\\
&= \quad \big[H(Y|Z_1) - H(X|Z_1)\big] + \big[H(X|Z_1, W_1) - H(Y|Z_1, W_1)\big], && (139)
\end{aligned}
$$

where, (a) follows from the Markov chain $Z_1 \div X \div Y$, and (b) from the Markov chain $X \div (Y, Z_1) \div W_1$. For a general $k \in [1, ..., K]$, by similar considerations, we obtain

$$
\begin{aligned}
&I(X; Z_k, W_k|Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) = && (140)\\
&= \big[H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) - H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1})\big]\\
&+\big[H(Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k) - H(X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k)\big],
\end{aligned}
$$

and

$$
\begin{aligned}
&I(Y; Z_k, W_k|X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) = && (141)\\
&= \big[H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k) - H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k)\big]\\
&+\big[H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k) - H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k)\big].
\end{aligned}
$$

For a given joint distribution of (X, Y), $H(X)$ and $H(Y)$ are both given and unaffected by $\{Z_k\}_{k=1}^{K}$ and $\{W_k\}_{k=1}^{K}$. Therefore, in order to preserve prescribed values of lower bounds to $R_{x,k}$ and $R_{y,k}$ it is sufficient to preserve the associated values of

$$
\begin{aligned}
&\big[H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}) - H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1})\big] && (142)\\
&+\big[H(Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k) - H(X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k)\big],
\end{aligned}
$$

and

$$
\begin{aligned}
&\big[H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k) - H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k)\big] && (143)\\
&+\big[H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k) - H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k)\big].
\end{aligned}
$$

Now, it can be shown that for a certain $k = \tilde{k}$, the values of (142) and (143) are preserved if preserving the values of (142) and (143) for $k = 1, 2, ..., \tilde{k} - 1$. Thus, for each $k = 1, ..., K$, it remains to preserve the values of

$$H(Y, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k) - H(X, Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}|Z_k), \tag{144}$$

and

$$H(X|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k) - H(Y|Z_1, ..., Z_{k-1}, W_1, ..., W_{k-1}, Z_k, W_k). \tag{145}$$

We first invoke the support lemma in order to reduce the alphabet size of Z_1, while preserving the value of $H(Y|Z_1) - H(X|Z_1)$. The alphabets of $\{Z_k\}_{k=2}^K$ and $\{W_k\}_{k=1}^K$ are still kept intact at this step. Define the following functionals of a generic distribution Q over $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X} \times \mathcal{Y}$ is assumed, without loss of generality, to be $\{1, 2, ..., m\}$, $m \stackrel{\triangle}{=} |\mathcal{X}| \cdot |\mathcal{Y}|$:

$$q_i(Q) = Q(x, y),\ i \stackrel{\triangle}{=} (x, y) = 1, 2, ..., m - 1, \tag{146}$$

and

$$q_m(Q) = \sum_{x,y} Q(x, y) \log \frac{\sum_y Q(x, y)}{\sum_x Q(x, y)}. \tag{147}$$

Applying now the support lemma, we find that there exists a random variable Z_1 (jointly distributed with (X, Y)), whose alphabet size is $|Z_1| = m = |\mathcal{X}| \cdot |\mathcal{Y}| \leq |\mathcal{X}| \cdot |\mathcal{Y}| + 2$ and it satisfies simultaneously:

$$\sum_{z_1} \Pr\{Z_1 = z_1\} q_i(P(\cdot|z_1)) = P_{XY}(x, y),\ i = 1, 2, ..., m - 1, \tag{148}$$

and

$$\sum_{z_1} \Pr\{Z_1 = z_1\} q_m(P(\cdot|z_1)) = H(Y|Z_1) - H(X|Z_1). \tag{149}$$

Having found a random variable Z_1, we now proceed to reduce the alphabet of W_1 in a similar manner, where this time, we have $m = |\mathcal{X}| \cdot |\mathcal{Y}| \cdot |\mathcal{Z}_1| - 1$ constraints to preserve the joint distribution of (X, Y, Z_1) just defined, one more constraint to preserve $H(X|Z_1, W_1) - H(Y|Z_1, W_1)$, and two more constraints to preserve the distortions at the first communication round $k = 1$:

$$q_{m+1}(Q) = \min_{\hat{x}} \sum_{X,Y,Z_1} Q(X, Y, Z_1) d_{x,1}(x, \hat{x}), \tag{150}$$

and

$$q_{m+2}(Q) = \min_{\hat{y}} \sum_{X,Y,Z_1} Q(X, Y, Z_1) d_{y,1}(y, \hat{y}). \tag{151}$$

Applying the support lemma, we obtain that W_1 satisfies (145), and also

$$\sum_{w_1} \Pr\{W_1 = w_1\} q_{m+1}(P(\cdot|w_1)) = \min_{G_1} Ed(X, G_1(X, Z_1, W_1)) \tag{152}$$

and

$$\sum_{w_1} \Pr\{W_1 = w_1\} q_{m+1}(P(\cdot|w_1)) = \min_{F_1} Ed(X, F_1(Y, Z_1, W_1)). \tag{153}$$

Thus, the necessary alphabet size of W_1 is upper-bounded by

$$|\mathcal{W}_1| \leq |\mathcal{X}| \cdot |\mathcal{Y}| \cdot |\mathcal{Z}_1| + 2 \leq |\mathcal{X}| \cdot |\mathcal{Y}| \cdot |\mathcal{X}| \cdot |\mathcal{Y}| + 2 \leq (|\mathcal{X}| \cdot |\mathcal{Y}| + 2)^2. \tag{154}$$

Following this step-by-step reduction of the alphabets of $\{Z_k\}_{k=2}^{K}$ and $\{W_k\}_{k=2}^{K-1}$ we prove the existence of auxiliary RVs with necessary alphabet sizes of

$$|\mathcal{Z}_k| \leq (|\mathcal{X}| \cdot |\mathcal{Y}| + 2)^{2k-1} \tag{155}$$

and

$$|\mathcal{W}_k| \leq (|\mathcal{X}| \cdot |\mathcal{Y}| + 2)^{2k}. \tag{156}$$

This completes the proof of the converse part.

References

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

[2] A.D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. on Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, January 1976.

[3] A. Kaspi, "Two-Way Source Coding with a Fidelity Criterion," *IEEE Trans. Inform. Theory*, vol. IT–31, no. 6. pp. 735–740, November 1985.

[4] Y. Steinberg and N. Merhav, "On Successive Refinement to the Wyner-Ziv Problem," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1636–1654, August 2004.

[5] Y. Steinberg and N. Merhav, "On Hierarchical Joint Source-Channel Coding ," submitted to *IEEE Trans. Inform. Theory*, November 2003.

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

[7] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression*, edited by T. Kailath, Prentice - Hall, 1971.

[8] V.N. Koshelev, "On the divisibility of discrete sources with an additive single-letter distortion measure," *Probl. Peredachi Inform.*, vol. 30, no. 1, pp. 31–50, 1994. English translation: vol. 30, no. 1, pp. 27–43, 1994.

[9] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 253–259, January 1994.