PAPER

Multiterminal source coding with complementary delivery*

Akisato KIMURA[†a)] ***and*** **Tomohiko UYEMATSU**[††b)], ***Members***

SUMMARY A coding problem for correlated information sources is investigated. Messages emitted from two correlated sources are jointly encoded, and delivered to two decoders. Each decoder has access to one of the two messages to enable it to reproduce the other message. The rate-distortion function for the coding problem and its interesting properties are clarified.
key words: *multiterminal source coding, complementary delivery, joint encoding, separate decoding*

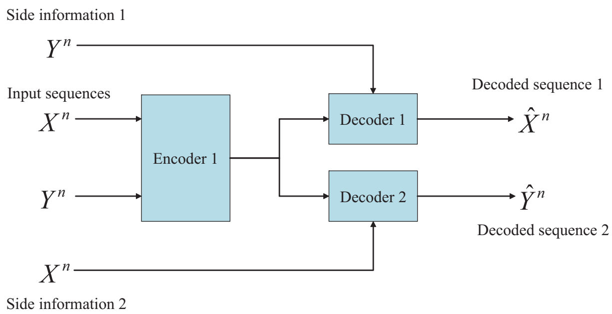

Fig. 1 Complementary delivery network

1. Introduction

Coding problems for correlated information sources were originally investigated by Slepian and Wolf [1]. Corresponding rate-distortion coding problems [2]–[4] and various coding problems (e.g. [5]–[7]) inspired by the work by Slepian and Wolf have been considered. Including the above studies, the main focus in the 1970's was on coding problems with *separate encoding* (each message is separately encoded) and *joint decoding* (several codewords are sent to a decoder and decoded simultaneously).

In contrast, since the 1980's, coding problems that involve *joint encoding* (messages from several sources are encoded at once) and/or *separate decoding* (each message is separately decoded) have been explored. Separate decoding processes have mainly been considered in relation to multiple description (e.g. [8]–[10]), while joint encoding processes can be seen, for example, in the cascading and branching communication systems [11], the triangular communication system [12] and multi-hop networks [13], [14].

Also, a coding problem that involves joint encoding and separate decoding was considered by Willems et al. [15], [16]. The coding system models a communication network via a satellite. Several stations are separately deployed in a field. Every station collects its own target data and wants to share all the target data with the other stations. To accomplish this task, each station transmits the collected data to a satellite, and the satellite broadcasts all the received data back to the stations. Each station utilizes its own target data as side information to reproduce all the other target data. Willems et al. [16] investigated a special case of the above scenario in which three stations were deployed and each station had access to one of three target messages, and determined the minimum *lossless* achievable rate for uplink (from each station to the satellite) and downlink (from the satellite to all the stations) transmissions. Their main result implies that the uplink transmission is equivalent to the traditional Slepian-Wolf coding system [1], and thus the main problem is the downlink part. Henceforth we denote the networks characterized by the downlink transmission as *generalized complementary delivery networks*, and we denote the generalized complementary network with two stations and two target messages as the *complementary delivery network* (Fig. 1). This notation is based on the network structure where each station (i.e. decoder) complements the target messages from the codeword delivered by the satellite (i.e. encoder). Kimura et al. investigated a universal coding problem for the complementary delivery network [17] and the generalized complementary delivery network [18], and proposed an explicit construction of lossless universal codes which attains the optimal error exponent. Also, Kuzuoka et al. [19], [20] simplified the coding scheme by introducing a concept of network coding [21].

The above previous researches considered only the lossless coding problem. In contrast, this paper focuses on the lossy coding problem. The minimum achievable rate given distortion criteria and some interesting

Manuscript received March 23, 2008.
[†]NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan.
[††]Department of Communications and Integrated Systems, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan.
a) E-mail: research@akisato.org
b) E-mail: uyematsu@ieee.org
[*]Some of the material in this paper was presented at the 2006 International Symposium on Information Theory and Its Applications, Seoul, Korea October-November 2006.

properties of the minimum achievable rate are clarified.

This paper is organized as follows. Section 2 provides notations and definitions used throughout in this paper. Section 3 investigates the lossy coding problem for the complementary delivery network, which includes descriptions of the main result and several related properties. The main result can be easily extended to the problem of the generalized complementary delivery networks, which will be discussed in Section 4. Finally, Section 5 provides theorem proofs.

2. Preliminaries

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets. Especially, for any natural number M, we denote $\mathcal{I}_M = \{1, 2, \cdots, M\}$. The cardinality of $\mathcal{X}$ is denoted as $|\mathcal{X}|$. A member of $\mathcal{X}^n$ is written as $x^n = (x_1, x_2, \cdots, x_n)$, and substrings of x^n are written as $x_i^j = (x_i, x_{i+1}, \cdots, x_j)$ for $i \leq j$. A set of all the probability distributions on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$. A discrete memoryless source $(\mathcal{X}, P_X)$ is an infinite sequence $\{X_i\}_{i=1}^{\infty}$ of independent copies of a random variable X taking values in $\mathcal{X}$ with a generic distribution $P_X \in \mathcal{P}(\mathcal{X})$, namely

$$P_{X^n}(x^n) = \prod_{i=1}^{n} P_X(x_i).$$

$\mathcal{P}(\mathcal{X}|P_Y)$ denotes a set of all the probability distributions on $\mathcal{X}$ given a distribution $P_Y \in \mathcal{P}(\mathcal{Y})$. Namely, each member of $\mathcal{P}(\mathcal{X}|P_Y)$ is characterized by $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ as $P_{XY} = P_{X|Y} P_Y$. A source $(\mathcal{X}, P_X)$ can be denoted by referring to its generic distribution P_X or random variable X. For a correlated source (X, Y), $H(X)$, $H(X|Y)$ and $I(X;Y)$ denote the entropy of X, the conditional entropy of X given Y, and the mutual information of X and Y, respectively. Similarly, for a correlated source (X, Y, Z), $I(X;Y|Z)$ denotes the conditional mutual information of X and Y given Z. In the following, all bases of exponentials and logarithms are set at e (the base of the natural logarithm). Let $\widehat{\mathcal{X}}$ stand for a reconstruction alphabet that corresponds to a source X to be encoded, and let $\Delta_X : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \overline{\Delta}_X]$ be a corresponding single-letter distortion function, where $\overline{\Delta}_X < \infty$. The vector distortion function is defined in the usual way, i.e.

$$\Delta_X^n(x^n, \widehat{x}^n) = \frac{1}{n} \sum_{k=1}^{n} \Delta_X(x_k, \widehat{x}_k).$$

3. Complementary delivery

3.1 Problem formulation

Definition 1. (CD (Complementary Delivery) code) *A set $(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})$ of an encoder and decoders is a CD code $(n, M_n, \rho_n^{(X)}, \rho_n^{(Y)})$ for the source (X, Y) if and only if*

$$\begin{aligned}
\varphi_n &: \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{I}_{M_n} \\
\widehat{\varphi}_n^{(1)} &: \mathcal{I}_{M_n} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\
\widehat{\varphi}_n^{(2)} &: \mathcal{I}_{M_n} \times \mathcal{X}^n \to \widehat{\mathcal{Y}}^n, \\
\rho_n^{(X)} &= E\left[\Delta_X^n(X^n, \widehat{\varphi}_n^{(1)}(A_n, Y^n))\right], \\
\rho_n^{(Y)} &= E\left[\Delta_Y^n(Y^n, \widehat{\varphi}_n^{(2)}(A_n, X^n))\right], \\
A_n &= \varphi_n(X^n, Y^n).
\end{aligned}$$

Definition 2. (Lossy CD-achievable rate) *R is a lossy CD-achievable rate of the source (X, Y) for a given distortion pair (D_X, D_Y) if and only if there exists a sequence $\left\{\left(n, M_n, \rho_n^{(X)}, \rho_n^{(Y)}\right)\right\}_{n=1}^{\infty}$ of CD codes for the source (X, Y) such that*

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R, \\
&\limsup_{n\to\infty} \rho_n^{(X)} \leq D_X, \quad \limsup_{n\to\infty} \rho_n^{(Y)} \leq D_Y.
\end{aligned}$$

Definition 3. (Inf lossy CD-achievable rate)

$$\begin{aligned}
R(X, Y|D_X, D_Y) = \inf\{&R | R \textit{ is a lossy} \\
&\textit{CD-achievable rate of } (X, Y) \textit{ for } (D_X, D_Y)\}.
\end{aligned}$$

3.2 Statement of results

Theorem 1. (Lossy coding theorem for CD code)

$$\begin{aligned}
&R(X, Y|D_X, D_Y) \\
&= \min_{P_{U|XY} \in \mathcal{P}_{CD}(\mathcal{U}|P_{XY})} \left[\max\{I(X;U|Y), I(Y;U|X)\}\right],
\end{aligned}$$

where the alphabet $\mathcal{U}$ satisfies

$$|\mathcal{U}| \leq |\mathcal{X} \times \mathcal{Y}| + 2$$

and $\mathcal{P}_{CD}(\mathcal{U}|P_{XY}) \subseteq \mathcal{P}(\mathcal{U}|P_{XY})$ is a set of probability distributions such that there exist functions $\phi_{(1)} : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ and $\phi_{(2)} : \mathcal{U} \times \mathcal{X} \to \widehat{\mathcal{Y}}$ that satisfy

$$\begin{aligned}
D_X &\geq E\left[\Delta_X(X, \phi_{(1)}(U, Y))\right], \\
D_Y &\geq E\left[\Delta_Y(Y, \phi_{(2)}(U, X))\right].
\end{aligned}$$

Several important relationships between Theorem 1 and previously reported results are presented in the following.

Lemma 1. (Compatibility with the result obtained for the lossless coding) *Suppose that $\widehat{\mathcal{X}} = \mathcal{X}$, $\widehat{\mathcal{Y}} = \mathcal{Y}$, $\Delta_X(x, \widehat{x}) = 0$ if and only if $x = \widehat{x}$ and $\Delta_Y(y, \widehat{y}) = 0$ if and only if $y = \widehat{y}$. In this case, the inf achievable rate $R(X, Y|D_X, D_Y)$ for $D_X = D_Y = 0$ is reduced to the minimum achievable rate for the lossless coding.*

$$\begin{aligned}
R(X, Y) &\stackrel{\text{def.}}{=} R(X, Y|D_X = 0, D_Y = 0) \\
&= \max\{H(X|Y), H(Y|X)\},
\end{aligned}$$

which coincides with the result reported by Willems et al. [16].

Proof. Note that if the conditions shown in Lemma 1 satisfy we have

$$\rho_n^{(X)} = 0 \iff \Pr\{X^n \neq \widehat{\varphi}_n^{(1)}(A_n, Y^n))\} = 0,$$
$$\rho_n^{(Y)} = 0 \iff \Pr\{Y^n \neq \widehat{\varphi}_n^{(2)}(A_n, X^n))\} = 0.$$

□

Lemma 2. (Relationship to the conditional rate-distortion function)

$$R(X,Y|D_X = d_1, D_Y) = R_C(Y|X, D_Y),$$
$$R(X,Y|D_X, D_Y = d_2) = R_C(X|Y, D_X)$$

if $d_1 \geq \overline{\Delta}_X$ and $d_2 \geq \overline{\Delta}_Y$, where $R_C(X|Y,D)$ denotes the conditional rate-distortion function [22], namely the minimum achievable rate when X is encoded and reproduced both with the side information Y to guarantee the distortion criterion D.

Proof. It is sufficient to show that first equation. The condition $d_1 \geq \overline{\Delta}_X$ implies that one of the two messages does not have to be reproduced. Therefore, the encoder φ_n sends the codeword only to the decoder $\widehat{\varphi}_n^{(2)}$, which means that the coding rate characterized by the conditional rate-distortion function is an achievable rate.

$$R(X,Y|D_X = d_1, D_Y) \leq R_C(Y|X, D_Y).$$

On the other hand, we have

$$\begin{aligned} & R(X,Y|D_X, D_Y) \\ & \geq \max\{R_C(X|Y, D_X), R_C(Y|X, D_Y)\} \\ & \geq R_C(Y|X, D_Y). \end{aligned}$$

from the result of Theorem 1. □

Lemma 3. (Relationships to the conditional rate-distortion function and Wyner-Ziv rate distortion function)

$$\begin{aligned} & \max\{R_C(X|Y, D_X), R_C(Y|X, D_Y)\} \\ & \leq R(X,Y|D_X, D_Y) \\ & \leq \max\{R_{WZ}(X|Y, D_X), R_{WZ}(Y|X, D_Y)\}, \end{aligned}$$

where $R_{WZ}(X|Y, D_X)$ is the minimum achievable rate for the coding system called the Wyner-Ziv coding system [2], where X is encoded without any side information and reproduced with the side information Y.

Proof. The left inequality was shown in the proof of Lemma 2. The right inequality was shown by Kuzuoka et al. [20]. □

Lemma 3 indicates that there may be some rate losses only for the lossy coding. This property results from the auxiliary random variable U included in the inf achievable rate $R(X,Y|D_X, D_Y)$.

4. Extension to multiple sources

Theorem 1 considered only two correlated sources. However, the theorem can be easily extended to any finite number of correlated sources.

Let $\boldsymbol{X}$ be a set of N discrete memoryless sources

$$\boldsymbol{X} = \{X^{(1)}, X^{(2)}, \cdots, X^{(N)}\},$$

each of which $X^{(i)}$ takes a value in a finite set $\mathcal{X}^{(i)}$ ($i \in \mathcal{I}_N$). For a given subset $\mathcal{S} \subseteq \mathcal{I}_N$ of source indexes, the corresponding subsets of sources, alphabets and its members are denoted by

$$\begin{aligned} \boldsymbol{X}^{(\mathcal{S})} &= \{X^{(i)} | i \in \mathcal{S}\}, \\ \mathcal{X}^{(\mathcal{S})} &= \prod_{i \in \mathcal{S}} \mathcal{X}^{(i)}, \\ \boldsymbol{x}^{(\mathcal{S})} &= \{x^{(i)} \in \mathcal{X}^{(i)} | i \in \mathcal{S}\}. \end{aligned}$$

Similarly, for a given subset $\mathcal{S} \subseteq \mathcal{I}_N$, the n-th Cartesian product of $\mathcal{X}^{(\mathcal{S})}$, its member and the corresponding random variable are written as $\mathcal{X}^{(\mathcal{S})n}$, $\boldsymbol{x}^{(\mathcal{S})n}$ and $\boldsymbol{X}^{(\mathcal{S})n}$, respectively. A substring of $\boldsymbol{x}^{(\mathcal{S})n}$ is written as $\boldsymbol{x}_i^{(\mathcal{S})j}$ for $i \leq j$. With $\mathcal{S} = \mathcal{I}_N$, we denote $\boldsymbol{X}^{(\mathcal{S})n} = \boldsymbol{X}^n$. Also for a given subset $\mathcal{S} \subseteq \mathcal{I}_N$, its complement is denoted by $\mathcal{S}^c = \mathcal{I}_N - \mathcal{S}$.

Here, we introduce the definition and the coding theorem of the *generalized complementary delivery code* which considers multiple correlated sources, multiple encoders and multiple decoders.

Definition 4. (GCD (Generalized Complementary Delivery) code)
A set $(\varphi_n, \widehat{\varphi}_n^{(1)}, \cdots, \widehat{\varphi}_n^{(M)})$ of single encoder and M decoders is a GCD code

$$\left(n, M_n, \{\rho_n^{(j,i)}\}_{j \in \mathcal{I}_M, i \in \mathcal{S}_j}\right)$$

for the source $\boldsymbol{X}$ if and only if for any $j \in \mathcal{I}_M$ and $i \in \mathcal{S}_j \subseteq \mathcal{I}_N$

$$\begin{aligned} \varphi_n &: \mathcal{X}^{(\mathcal{I}_N)n} \to \mathcal{I}_{M_n} \\ \widehat{\varphi}_n^{(j)} &: \mathcal{I}_{M_n} \times \mathcal{X}^{(\mathcal{S}_j^c)n} \to \widehat{\mathcal{X}}^{(\mathcal{S}_j)n}, \\ \rho_n^{(j,i)} &= E\left[\Delta^n_{X^{(i)}}(X^{(i)n}, \widehat{\varphi}_n^{(j;i)}(A_n, \boldsymbol{X}^{(\mathcal{S}_j^c)n}))\right], \\ A_n &= \varphi_n(\boldsymbol{X}^n), \end{aligned}$$

where $\widehat{\varphi}_n^{(j;i)}$ is the output of $\widehat{\varphi}_n^{(j)}$ that corresponds to the reproduction of $X^{(i)n}$.

Definition 5. (Lossy GCD-achievable rate)
R is a lossy GCD-achievable rate of the source $\boldsymbol{X}$ for a given set

$$\boldsymbol{D} = \{D_{j,i}\}_{j \in \mathcal{I}_M, i \in \mathcal{S}_j}$$

of distortion criteria if and only if there exists a sequence

$$\left\{\left(n, M_n, \{\rho_n^{(j,i)}\}_{j\in\mathcal{I}_M, i\in\mathcal{S}_j}\right)\right\}_{n=1}^{\infty}$$

of GCD codes for the source $\boldsymbol{X}$ *such that for any* $j \in \mathcal{I}_M$ *and* $i \in \mathcal{S}_j$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le R,$$
$$\limsup_{n\to\infty} \rho_n^{(j,i)} \le D_{j,i}.$$

Definition 6. (Inf lossy GCD-achievable rate)

$$R(\boldsymbol{X}|\boldsymbol{D}) = \inf\{R | R \textit{ is a lossy GCD-achievable rate of } \boldsymbol{X} \textit{ for } \boldsymbol{D}\}.$$

Theorem 2. (Coding theorem of lossy GCD code)

$$R(\boldsymbol{X}|\boldsymbol{D}) = \min_{P_{U|\boldsymbol{X}}\in\mathcal{P}_{CD}(\mathcal{U}|P_{\boldsymbol{X}})} \max_{j\in\mathcal{I}_M} I\left(\boldsymbol{X}^{(\mathcal{S}_j)}; U \middle| \boldsymbol{X}^{(\mathcal{S}_j^c)}\right),$$

where the alphabet $\mathcal{U}$ *satisfies*

$$|\mathcal{U}| \le |\mathcal{X}^{(\mathcal{I}_N)}| + \sum_{j=1}^{M} |\mathcal{S}_j|,$$

and $\mathcal{P}_{CD}(\mathcal{U}|P_{\boldsymbol{X}}) \subseteq \mathcal{P}(\mathcal{U}|P_{\boldsymbol{X}})$ *is a set of probability distributions such that for any* $j \in \mathcal{I}_M$ *and* $i \in \mathcal{S}_j$ *there exists a function* $\phi_{(j,i)} : \mathcal{U} \times \mathcal{X}^{(\mathcal{S}_j^c)} \to \widehat{\mathcal{X}}^{(i)}$ *that satisfy*

$$D_{j,i} \ge E\left[\Delta_{X^{(i)}}\left(X^{(i)}, \phi_{(j,i)}\left(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}\right)\right)\right].$$

As a typical example, Theorem 2 can be applied to the coding problem formulated by Willems et al. [16]. In this coding system, the encoder sends three messages $\boldsymbol{X} = \{X, Y, Z\}$ to three decoders, and each decoder has access to one of three messages to reproduce the two other messages. Theorem 2 indicates that the inf achievable rate for this coding problem is obtained as

$$\begin{aligned} &R(X,Y,Z|D_1,D_2,D_3) \\ &= \min_{P_{U|XYZ}\in\mathcal{P}_{CD}(\mathcal{U}|P_{XYZ})} \\ &\quad \max\{I(XY;U|Z), I(YZ;U|X), I(XZ;U|Y)\}, \end{aligned}$$

where the alphabet $\mathcal{U}$ satisfies $|\mathcal{U}| \le |\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}|+6$, and $\mathcal{P}_{CD}(\mathcal{U}|P_{XYZ}) \subseteq \mathcal{P}(\mathcal{U}|P_{XYZ})$ is a set of probability distributions such that there exist functions

$$\begin{aligned} &\phi_{(12)} : \mathcal{U}\times\mathcal{X} \to \widehat{\mathcal{Y}}, \quad \phi_{(13)} : \mathcal{U}\times\mathcal{X} \to \widehat{\mathcal{Z}}, \\ &\phi_{(21)} : \mathcal{U}\times\mathcal{Y} \to \widehat{\mathcal{X}}, \quad \phi_{(23)} : \mathcal{U}\times\mathcal{Y} \to \widehat{\mathcal{Z}}, \\ &\phi_{(31)} : \mathcal{U}\times\mathcal{Z} \to \widehat{\mathcal{X}}, \quad \phi_{(32)} : \mathcal{U}\times\mathcal{Z} \to \widehat{\mathcal{Y}} \end{aligned}$$

that satisfy

$$\begin{aligned} D_{12} &\ge E[\Delta_Y(Y, \phi_{(12)}(U,X))], \\ D_{13} &\ge E[\Delta_Z(Z, \phi_{(13)}(U,X))], \\ D_{21} &\ge E[\Delta_X(X, \phi_{(21)}(U,Y))], \\ D_{23} &\ge E[\Delta_Z(Z, \phi_{(23)}(U,Y))], \\ D_{31} &\ge E[\Delta_X(X, \phi_{(31)}(U,Z))], \\ D_{32} &\ge E[\Delta_Y(Y, \phi_{(32)}(U,Z))]. \end{aligned}$$

5. Proof of theorems

5.1 Theorem 1: converse part

Proof.
Let a sequence $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ of CD codes be given that satisfy the conditions of Definitions 1 and 2. From Definition 2, for any $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ and then for all $n \ge n_1(\delta)$, we can obtain

$$\frac{1}{n}\log M_n \le R + \delta.$$

It should be remembered that $A_n = \varphi_n(X^n, Y^n)$. Then, we obtain

$$\begin{aligned} &n(R+\delta) \\ &\ge \log M_n \\ &\ge H(A_n) \\ &\ge H(A_n|Y^n) \\ &= I(X^n; A_n|Y^n) \quad (\because A_n = \varphi_n(X^n, Y^n)) \\ &= H(X^n|Y^n) - H(X^n|A_nY^n) \\ &= \sum_{k=1}^{n}\{H(X_k|Y_k) - H(X_k|A_nX^{k-1}Y^n)\} \\ &= \sum_{k=1}^{n} I(X_k; A_nX^{k-1}Y^{k-1}Y_{k+1}^n|Y_k) \\ &\ge \sum_{k=1}^{n} I(X_k; A_nX^{k-1}Y^{k-1}|Y_k). \end{aligned}$$

Let us define random variables $U_k = A_nX^{k-1}Y^{k-1}$. With these definitions, we have

$$n(R+\delta) \ge \sum_{k=1}^{n} I(X_k; U_k|Y_k).$$

In a similar manner, we obtain

$$n(R+\delta) \ge \sum_{k=1}^{n} I(Y_k; U_k|X_k).$$

Here, let J be a random variable that is independent of (X, Y) and uniformly distributed over the set $\mathcal{I}_n$. We define a random variable $U = (J, U_J)$. This implies that

$$\begin{aligned} &R+\delta \\ &\ge \frac{1}{n}\sum_{k=1}^{n} I(X_k; U_k|Y_k) \\ &= \frac{1}{n}\sum_{k=1}^{n}\{H(X_k|Y_k) - H(X_k|U_kY_k)\} \\ &= \frac{1}{n}\sum_{k=1}^{n}\{H(X_k|Y_k) - H(X_J|U_JY_J, J=k)\} \end{aligned}$$

$$= H(X|Y) - H(X_J|JU_JY_J)$$
$$= H(X|Y) - H(X|UY)$$
$$= I(X;U|Y)$$

and

$$R + \delta \geq I(Y;U|X).$$

Since $\delta > 0$ is arbitrary, we obtain

$$R \geq \max\{I(X;U|Y), I(Y;U|X)\}.$$

We next show the existence of functions $\phi_{(1)}$ and $\phi_{(2)}$ that satisfy the conditions of Theorem 1. From Definition 2, for any $\gamma > 0$, there exists an integer $n_2 = n_2(\gamma)$, and for all $n \geq n_2(\gamma)$, we have

$$\begin{aligned} D_X + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{\varphi}^{(1)}_{n,k}(A_n, Y^n))\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{X}_k)\right], \\ D_Y + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_Y(Y_k, \widehat{\varphi}^{(2)}_{n,k}(A_n, X^n))\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_Y(Y_k, \widehat{Y}_k)\right], \end{aligned}$$

where $\widehat{\varphi}^{(i)}_{n,k}$ ($i = 1, 2,\ k \in \mathcal{I}_n$) is the output of $\widehat{\varphi}^n_{(i)}$ at time k, and

$$\begin{aligned} \widehat{X}_k &= \widehat{\varphi}^{(1)}_{n,k}(A_n, Y^n), \\ \widehat{Y}_k &= \widehat{\varphi}^{(2)}_{n,k}(A_n, X^n). \end{aligned}$$

We note that U_kY_k contains A_nY^k, and U_kX_k contains A_nX^k, which implies that Y^n_{k+1} (resp. X^n_{k+1}) is further needed to generate $\widehat{X}_k$ from U_kY_k (resp. $\widehat{Y}_k$ from U_kX_k). Here, let us define the distribution Q_{k_1,k_2} of $A_nX^{k_1}Y^{k_2}$, namely for any $x^{k_1} \in \mathcal{X}^{k_1}$, $y^{k_2} \in \mathcal{Y}^{k_2}$ and $a_n \in \mathcal{I}_{M^{(1)}_n}$

$$\begin{aligned} &Q_{k_1,k_2}(a_n, x^{k_1}, y^{k_2}) \\ &\stackrel{\text{def.}}{=} \Pr\{\varphi_n(X^n, Y^n) = a_n, X^{k_1} = x^{k_1}, Y^{k_2} = y^{k_2}\} \\ &= \sum_{\substack{(x^n_{k_1+1}, y^n_{k_2+1}) \in \mathcal{X}^{n-k_1} \times \mathcal{Y}^{n-k_2}: \\ \varphi_n(x^n, y^n) = a_n}} P_{X^nY^n}(x^n, y^n). \end{aligned}$$

Also, let $Q^{(1)}_k$ be the distribution of X_k given U_kY_k, namely for any $u_k = a_nx^{k-1}y^{k-1}$

$$Q^{(1)}_k(x_k|u_k, y_k) \stackrel{\text{def.}}{=} \frac{Q_{k,k}(a_n, x^k, y^k)}{Q_{k-1,k}(a_n, x^{k-1}, y^k)},$$

and $Q^{(2)}_k$ be the distribution of Y_k given U_kX_k defined similarly.

$$Q^{(2)}_k(y_k|u_k, x_k) \stackrel{\text{def.}}{=} \frac{Q_{k,k}(a_n, x^k, y^k)}{Q_{k,k-1}(a_n, x^k, y^{k-1})}.$$

Further, let us define $\widetilde{Y}^n_{k+1}(U_k, Y_k)$ (resp. $\widetilde{X}^n_{k+1}(U_k, X_k)$) as a random variable selected to minimize the average distortion between X_k and $\widehat{X}_k$ given U_kY_k (resp. between Y_k and $\widehat{Y}_k$ given U_kX_k), namely

$$\begin{aligned} &\widetilde{Y}^n_{k+1}(U_k, Y_k) \stackrel{\text{def.}}{=} \mathop{\arg\min}_{Y^n_{k+1} \in \mathcal{Y}^{n-k}} \\ &\qquad \sum_{X_k \in \mathcal{X}} Q^{(1)}_k(X_k|U_k, Y_k)\Delta_X(X_k, \widehat{X}_k), \\ &\widetilde{X}^n_{k+1}(U_k, X_k) \stackrel{\text{def.}}{=} \mathop{\arg\min}_{X^n_{k+1} \in \mathcal{X}^{n-k}} \\ &\qquad \sum_{Y_k \in \mathcal{Y}} Q^{(2)}_k(Y_k|U_k, X_k)\Delta_Y(Y_k, \widehat{Y}_k). \end{aligned}$$

We choose the functions $\phi_{(1)}$ and $\phi_{(2)}$ as follows:

$$\begin{aligned} \phi_{(1)k}(U_k, Y_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}^{(1)}_{n,k}(A_n, Y^k * \widetilde{Y}^n_{k+1}(U_k, Y_k)), \\ \phi_{(2)k}(U_k, X_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}^{(2)}_{n,k}(A_n, X^k * \widetilde{X}^n_{k+1}(U_k, X_k)), \\ \phi_{(1)}(U, Y) &\stackrel{\text{def.}}{=} \phi_{(1)J}(U_J, Y), \\ \phi_{(2)}(U, X) &\stackrel{\text{def.}}{=} \phi_{(2)J}(U_J, X) \end{aligned}$$

where $*$ is an operator that represents string concatenation. It is easy to see that

$$\begin{aligned} &E\left[\Delta_X(X_k, \phi_{(1)k}(U_k, Y_k))\right] \\ &= E\left[\Delta_X(X_k, \widehat{\varphi}^{(1)}_{n,k}(A_n, Y^k * \widetilde{Y}^n_{k+1}(U_k, Y_k)))\right] \\ &\leq E\left[\Delta_X(X_k, \widehat{\varphi}^{(1)}_{n,k}(A_n, Y^n))\right] \\ &= E\left[\Delta_X(X_k, \widehat{X}_k)\right] \\ &E\left[\Delta_Y(Y_k, \phi_{(2)k}(U_k, X_k))\right] \\ &\leq E\left[\Delta_Y(Y_k, \widehat{Y}_k)\right]. \end{aligned}$$

This implies

$$\begin{aligned} D_X + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{X}_k)\right] \\ &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \phi_{(1)k}(U_k, Y_k))\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X, \phi_{(1)J}(U_J, Y))|J = k\right] \\ &= E\left[\Delta_X(X, \phi_{(1)}(U, Y))\right], \\ D_Y + \gamma &\geq E\left[\Delta_Y(Y, \phi_{(2)}(U, X))\right]. \end{aligned}$$

Since $\gamma > 0$ is arbitrary, we obtain

$$\begin{aligned} D_X &\geq E\left[\Delta_X(X, \phi_{(1)}(U, Y))\right], \\ D_Y &\geq E\left[\Delta_Y(Y, \phi_{(2)}(U, X))\right]. \end{aligned}$$

It remains to establish that the bound on $|\mathcal{U}|$ specified in Theorem 1 does not affect the determination of the inf achievable rate $R(X,Y|D_X,D_Y)$. To do this, we introduce the support lemma [23, Lemma 3.3.4]. We can see that

$$
\begin{aligned}
&P_{XY}(x,y) = \sum_{u\in\mathcal{U}} P_U(u)P_{XY|U}(x,y|u),\\
&I(X;U|Y) = H(X|Y) - H(X|UY)\\
&= H(X|Y) - \sum_{u\in\mathcal{U}} P_U(u)\\
&\qquad \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u)\log\frac{P_{Y|U}(y|u)}{P_{XY|U}(x,y|u)}\\
&I(Y;U|X) = H(Y|X) - H(Y|UX)\\
&= H(Y|X) - \sum_{u\in\mathcal{U}} P_U(u)\\
&\qquad \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u)\log\frac{P_{X|U}(x|u)}{P_{XY|U}(x,y|u)}\\
&E[\Delta_X(X,\phi_{(1)}(U,Y))]\\
&= \sum_{u\in\mathcal{U}} P_U(u)\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u)\Delta_X(x,\phi_{(1)}(u,y))\\
&\geq \sum_{u\in\mathcal{U}} P_U(u)\sum_{y\in\mathcal{Y}}\min_{\widehat{x}\in\widehat{\mathcal{X}}}\\
&\qquad \sum_{x\in\mathcal{X}} P_{XY|U}(x,y|u)\Delta_X(x,\widehat{x}), \qquad (1)\\
&E[\Delta_Y(Y,\phi_{(2)}(U,X))]\\
&= \sum_{u\in\mathcal{U}} P_U(u)\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u)\Delta_Y(y,\phi_{(2)}(u,x))\\
&\geq \sum_{u\in\mathcal{U}} P_U(u)\sum_{x\in\mathcal{X}}\min_{\widehat{y}\in\widehat{\mathcal{Y}}}\\
&\qquad \sum_{y\in\mathcal{Y}} P_{XY|U}(x,y|u)\Delta_Y(y,\widehat{y}), \qquad (2)
\end{aligned}
$$

where Eq.(1) (resp. Eq.(2)) comes from the fact that for given letters $(u,y)\in\mathcal{U}\times\mathcal{Y}$ (resp. $(u,x)\in\mathcal{U}\times\mathcal{X}$) the output of the function $\phi_{(1)}$ (resp. $\phi_{(2)}$) can be selected so as to minimize the average distortion. We then define the following functions of a generic distribution $Q\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$:

$$
\begin{aligned}
&q_1(Q,(x,y)) = Q(x,y),\\
&q_2(Q) = \max\{q_{2,1}(Q), q_{2,2}(Q)\},\\
&q_{2,1}(Q)\\
&= H(X|Y) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} Q(x,y)\log\frac{\sum_{x'\in\mathcal{X}} Q(x',y)}{Q(x,y)},\\
&q_{2,2}(Q)\\
&= H(Y|X) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} Q(x,y)\log\frac{\sum_{y'\in\mathcal{Y}} Q(x,y')}{Q(x,y)},\\
&q_{3,1}(Q) = \sum_{y\in\mathcal{Y}}\min_{\widehat{x}\in\widehat{\mathcal{X}}}\sum_{x\in\mathcal{X}} Q(x,y)\Delta_X(x,\widehat{x})\\
&q_{3,2}(Q) = \sum_{x\in\mathcal{X}}\min_{\widehat{y}\in\widehat{\mathcal{Y}}}\sum_{y\in\mathcal{Y}} Q(x,y)\Delta_Y(y,\widehat{y})
\end{aligned}
$$

Note that $|\mathcal{X}\times\mathcal{Y}|-1$ functions are necessary to preserve the distribution $Q(x,y)$ and 2 functions to preserve the average distortion characterized by the generic distribution Q. From the support lemma, we can find a generic distribution $\alpha\in\mathcal{P}(\widetilde{\mathcal{U}})$ such that $\widetilde{\mathcal{U}}\subseteq\mathcal{U}$, $|\widetilde{\mathcal{U}}|\leq|\mathcal{X}\times\mathcal{Y}|+2$ and the following equations are simultaneously satisfied:

$$
\begin{aligned}
&\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_1(P_{XY|U}(\cdot|u),(x,y)) = P_{XY}(x,y), \qquad (3)\\
&\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_2(P_{XY|U}(\cdot|u))\\
&\qquad = \max\{I(X;U|Y), I(Y;U|X)\},\\
&\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_{3,1}(P_{XY|U}(\cdot|u))\\
&= \sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)\sum_{y\in\mathcal{Y}}\min_{\widehat{x}\in\widehat{\mathcal{X}}}\sum_{x\in\mathcal{X}} P_{XY|U}(x,y|u)\Delta_X(x,\widehat{x}),\\
&\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_{3,2}(P_{XY|U}(\cdot|u))\\
&= \sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)\sum_{x\in\mathcal{X}}\min_{\widehat{y}\in\widehat{\mathcal{Y}}}\sum_{y\in\mathcal{Y}} P_{XY|U}(x,y|u)\Delta_Y(y,\widehat{y}).
\end{aligned}
$$

Here, let us define functions $\phi^*_{(1)}:\widetilde{\mathcal{U}}\times\mathcal{Y}\to\widehat{\mathcal{X}}$ and $\phi^*_{(2)}:\widetilde{\mathcal{U}}\times\mathcal{X}\to\widehat{\mathcal{Y}}$ that satisfy

$$
\begin{aligned}
\phi^*_{(1)}(u,y) &= \arg\min_{\widehat{x}\in\widehat{\mathcal{X}}}\sum_{x\in\mathcal{X}} P_{XY|U}(x,y|u)\Delta_X(x,\widehat{x}),\\
\phi^*_{(2)}(u,x) &= \arg\min_{\widehat{y}\in\widehat{\mathcal{Y}}}\sum_{y\in\mathcal{Y}} P_{XY|U}(x,y|u)\Delta_Y(y,\widehat{y}).
\end{aligned}
$$

With these definitions, we have

$$
\begin{aligned}
\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_{3,1}(P_{XY|U}(\cdot|u)) &= E[\Delta_X(X,\phi^*_{(1)}(U,Y))],\\
\sum_{u\in\widetilde{\mathcal{U}}}\alpha(u)q_{3,2}(P_{XY|U}(\cdot|u)) &= E[\Delta_Y(Y,\phi^*_{(2)}(U,X))],
\end{aligned}
$$

and

$$
\begin{aligned}
D_1 &\geq E[\Delta_X(X,\phi_{(1)}(U,Y))]\\
&\geq E[\Delta_X(X,\phi^*_{(1)}(U,Y))],\\
D_2 &\geq E[\Delta_Y(Y,\phi_{(2)}(U,X))]\\
&\geq E[\Delta_Y(Y,\phi^*_{(2)}(U,X))].
\end{aligned}
$$

Hence, $\phi^*_{(1)}$ and $\phi^*_{(2)}$ satisfy the conditions of Theorem 1. Further, Eq.(3) implies that there exist a random variable $\widetilde{U}$ and a joint distribution $P_{\widetilde{U}XY}$ that satisfy

$$\alpha(u)P_{XY|U}(x,y|u) = P_{\widetilde{U}XY}(u,x,y)$$

for all $(u,x,y) \in \widetilde{\mathcal{U}} \times \mathcal{X} \times \mathcal{Y}$. The new joint distribution preserves the distribution P_{XY}

$$\begin{aligned}\sum_{u\in\widetilde{\mathcal{U}}} P_{\widetilde{U}XY}(u,x,y) &= \sum_{u\in\widetilde{\mathcal{U}}} \alpha(u)P_{XY|U}(x,y|u) \\ &= P_{XY}(x,y).\end{aligned}$$

This completes the proof of the converse part. □

5.2 Theorem 1: direct part

We begin by establishing some notation and mentioning a few basic facts that will be used hereafter.

Definition 7. (Set of typical sequences)
For any $\delta > 0$, define the set of typical sequences as

$$T_X^n(\delta) = \left\{ x^n \in \mathcal{X}^n : \left| \frac{1}{n}N(x|x^n) - P_X(x) \right| \le \delta \ \forall x \in \mathcal{X} \right\},$$

where $N(x|x^n)$ stands for the number of occurrences of the letter x included in the sequence x^n. A similar convention is used for other random variables. When the dimension is clear from the context, the superscript n will be omitted, e.g. $T_X(\delta)$.

Lemma 4. (Csiszár-Körner [23])
For any $\delta > 0$

$$\Pr\{X^n \in T_X(\delta)\} \ge 1 - \epsilon_n(\delta),$$

where

$$\lim_{n\to\infty} \epsilon_n(\delta) = 0.$$

Lemma 5. (Csiszár-Körner [23, Lemma 1.2.10])
For any $\delta, \delta' > 0$, if $(x^n, y^n) \in T_{XY}(\delta_1)$ then $x^n \in T_X(\delta_1|\mathcal{Y}|)$.

Lemma 6. (Steinberg-Merhav [24])
For any $\delta' > \delta > 0$ and $x^n \in T_X(\delta)$

$$\begin{aligned}&\exp\{-n(I(X;U)+\epsilon_1)\} \\ &\le \sum_{u^n:(u^n,x^n)\in T_{UX}(\delta')} P_U(u^n) \le \exp\{-n(I(X;U)-\epsilon_2)\},\end{aligned}$$

where ϵ_1 is a function of (δ,δ'), ϵ_2 is a function of (δ,δ') and

$$\lim_{\delta,\delta'\to 0} \epsilon_1 = \lim_{\delta,\delta'\to 0} \epsilon_2 = 0.$$

Now, we proceed with the proof of the direct part of Theorem 1.

Proof.
Let a distortion pair (D_X, D_Y) be given, and $P_{U|XY} \in \mathcal{P}_{CD}(\mathcal{U}|P_{XY})$. Fix arbitrary $\gamma, \delta > 0$.

Codeword selection: φ_n
(1) Randomly generate M_U independent codewords $u^n(i) \in \mathcal{U}^n$ $(i \in \mathcal{I}_{M_U})$, each of length n, according to P_U to create a codebook $\mathcal{A}_U = \{u^n(i)\}_{i=1}^{M_U}$.
(2) Partition the codebook $\mathcal{A}_U$ into N_U bins, each containing $L_U = M_U/N_U$ members of $\mathcal{A}_U$. For simplicity, M_U is a multiple of N_U. Let $\mathcal{A}_U(j)$ denote the subset of $\mathcal{A}_U$ whose elements are assigned to bin j $(j \in \mathcal{I}_{N_U})$. Without loss of generality, we define

$$\mathcal{A}_U(j) = \{u^n(i)\}_{i=(j-1)L_U+1}^{jL_U}.$$

Encoding: φ_n
(1) For a given input pair $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequences, the encoder seeks a vector $u^n \in \mathcal{A}_U$ that satisfies $(u^n, x^n, y^n) \in T_{UXY}(k_1\delta)$, where $k_1 > 0$. If there is more than one such vector in the codebook $\mathcal{A}_U$, the first one is chosen. If there is no such vector in the codebook $\mathcal{A}_U$, a default vector is chosen, say $u^n(1)$, and an error is declared. The selected vector is denoted by $u^n(x^n, y^n)$.
(2) The value assigned to the encoder $\varphi_n(\cdot)$ is the bin index to which $u^n(x^n, y^n)$ belongs, that is,

$$\varphi_n(x^n, y^n) = j \quad \text{if } u^n(x^n, y^n) \in \mathcal{A}_U(j).$$

Decoding: $\widehat{\varphi}_n^{(1)}$
(1) The decoder has access to the bin index $j_U \in \mathcal{I}_{N_U}$ received from the encoder and the sequence $y^n \in \mathcal{Y}^n$ of side information.
(2) The decoder seeks a unique vector $u^n \in \mathcal{A}_U(j_U)$ that satisfies $(u^n, y^n) \in T_{UY}(k_2\delta)$, where $k_2 > 0$. This vector is denoted by $\widehat{u}^n(y^n)$. If there is no or more than one vector $u^n \in \mathcal{A}_U(j_U)$ jointly typical with y^n, arbitrary $\widehat{u}^n$ is chosen, and an error is declared.
(3) The reconstruction vector $\widehat{x}^n = (\widehat{x}_1, \widehat{x}_2, \cdots, \widehat{x}_n)$ is given by

$$\widehat{x}_k = \phi_{(1)}(\widehat{u}_k(y^n), \widehat{y}_k) \quad (k \in \mathcal{I}_n),$$

where $\widehat{u}_k(y^n)$ is the k-th element of $u^n(y^n)$.

Decoding: $\widehat{\varphi}_n^{(2)}$
(1) The decoder has access to the bin index $j_U \in \mathcal{I}_{N_U}$ and the sequence $x^n \in \mathcal{X}^n$ of side information.
(2) In a similar manner to $\widehat{\varphi}_n^{(1)}$, the decoder seeks a unique vector $u^n \in \mathcal{A}_U(j_U)$ that satisfies $(u^n, x^n) \in T_{UX}(k_3\delta)$, where $k_3 > 0$, and the reconstruction vector $\widehat{y}^n$ is given by

$$\widehat{y}_k = \phi_{(2)}(\widehat{u}_k(x^n), \widehat{x}_k) \quad (k \in \mathcal{I}_n).$$

Distortion evaluation: $\widehat{\varphi}_n^{(1)}$
For the distortion, we obtain

$$\begin{aligned}&\Delta_X^n(x^n, \widehat{x}^n) \\ &= \frac{1}{n}\sum_{k=1}^n \Delta_X(x_k, \widehat{x}_k) \\ &= \frac{1}{n}\sum_{k=1}^n \Delta_X(x_k, \phi_{(1)}(\widehat{u}_k(y^n), \widehat{y}_k))\end{aligned}$$

$$= \frac{1}{n} \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} N(u,x,y|\widehat{u}^n(y^n),x^n,y^n) \Delta_X(x,\phi_{(1)}(u,y)).$$

We note that $(u^n(x^n,y^n),x^n,y^n) \in T_{UXY}(k_1\delta)$. Also, if no error occurs in the encoding/decoding process, we have $u^n(x^n,y^n) = \widehat{u}^n(y^n)$. In this case, the following inequalities are satisfied:

$$\begin{aligned}
&\Delta_X^n(x^n,\widehat{x}^n) \\
&\le \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} (P_{UXY}(u,x,y)+k_1\delta)\Delta_X(x,\phi_{(1)}(u,y)) \\
&\le E\left[\Delta_X(X,\phi_{(1)}(U,Y))\right] + k_1\delta\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}| \\
&\le D_X + k_1\delta\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|.
\end{aligned}$$

We denote error probabilities in the encoding/decoding process as P_e^n. Then, the average distortion can be bounded as

$$\begin{aligned}
&E\left[\Delta_X^n(X^n,\widehat{X}^n)\right] \\
&\le (1-P_e^n)(D_X + k_1\delta\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|) + P_e^n\overline{\Delta}_X.
\end{aligned}$$

Since $\delta > 0$ is arbitrarily small for a sufficiently large n, if P_e^n vanishes as $n \to \infty$, we can obtain

$$\limsup_{n\to\infty} E\left[\Delta_X^n(X^n,\widehat{X}^n)\right] \le D_X.$$

<u>*Distortion evaluation:* $\widehat{\varphi}_n^{(2)}$</u>

We can obtain

$$\limsup_{n\to\infty} E\left[\Delta_Y^n(Y^n,\widehat{Y}^n)\right] \le D_Y$$

in a similar manner to $\widehat{\varphi}_n^{(1)}$.

<u>*Error evaluation:* φ_n</u>

If there is no $u^n \in \mathcal{A}_U$ that satisfies $(u^n,x^n,y^n) \in T_{UXY}(k_1\delta)$, an encoding error has occurred. This event is denoted as

$$E_{11} \overset{\text{def.}}{=} \bigcap_{i=1}^{M_U} \{(u^n(i),x^n,y^n) \notin T_{UXY}(k_1\delta)\}.$$

Here, let us define

$$E_0 \overset{\text{def.}}{=} \{(x^n,y^n) \in T_{XY}(k_0\delta)\},$$

where $k_0 > 0$. From Lemma 4, $\Pr\{E_0^c\} \to 0$ as $n \to \infty$. Then, we have

$$\begin{aligned}
\Pr\{E_1\} &\le \Pr\{E_1 \cup E_0^c\} \\
&= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_1\},
\end{aligned}$$

$$\begin{aligned}
&\Pr\{E_0 \cap E_1\} \\
&\le \sum_{(x^n,y^n)\in T_{XY}(k_0\delta)} P_{XY}(x^n,y^n) \\
&\quad \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U^n(i),x^n,y^n) \notin T_{UXY}(k_1\delta)\} \middle| x^n,y^n \right\} \\
&= \sum_{(x^n,y^n)\in T_{XY}(k_0\delta)} P_{XY}(x^n,y^n) \\
&\quad \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U^n(i),x^n,y^n) \notin T_{UXY}(k_1\delta)\} \right\} \\
&\qquad (\because u^n(i) \text{ is selected independently of } (x^n,y^n)) \\
&\le \sum_{(x^n,y^n)\in T_{XY}(k_0\delta)} P_{XY}(x^n,y^n) \\
&\quad [1-\exp\{-n(I(XY;U)+\epsilon_u)\}]^{M_U} \\
&\qquad (\because \text{Lemma 6}) \\
&\le \sum_{(x^n,y^n)\in T_{XY}(k_0\delta)} P_{XY}(x^n,y^n) \\
&\quad \exp\left[-M_U \exp\{-n(I(XY;U)+\epsilon_u)\}\right], \\
&\qquad (\because (1-a)^n \le \exp(-an))
\end{aligned}$$

where ϵ_u is a function of $(k_1\delta, k_0\delta)$. By setting M_U, k_1 and k_0 as

$$M_U \ge \exp\{n(I(XY;U)+m_1\gamma)\},\ m_1 > 0,$$

$m_1\gamma > \epsilon_u$ and $k_1 < k_0$, we have $\lim_{n\to\infty} \Pr\{E_1\} = 0$.

<u>*Error evaluation:* $\widehat{\varphi}_n^{(1)}$</u>

If there is no or more than one $u^n \in \mathcal{A}_U(j_U)$ such that $(u^n,y^n) \in T_{UY}(k_2\delta)$, a decoding error is declared. This event is classified into two cases.

(1) First case: $(u^n(x^n,y^n),y^n) \notin T_{UY}(k_2\delta)$. However, this error does not occur by setting k_2 as $k_2 > k_1|\mathcal{X}|$ because $(u^n(x^n,y^n),x^n,y^n) \in T_{UXY}(k_1\delta)$ and Lemma 5.

(2) Second case: If there exists $u^n \in \mathcal{A}_U(j_U)$, $u^n \ne u^n(x^n,y^n)$ such that $(u^n,y^n) \in T_{UY}(k_2\delta)$. This event is denoted as

$$E_2 \overset{\text{def.}}{=} \bigcup_{u^n\in\mathcal{A}_U(j_U),u^n\ne u^n(x^n,y^n)} \{(u^n,y^n) \in T_{UY}(k_2\delta)\}.$$

Let $i(j,k)$ be the index i of k-th $u^n(i)$ in $\mathcal{A}_U(j)$, namely from the definition of $\mathcal{A}_U(j)$ we have

$$i(j,k) = (j-1)L_U + k.$$

Since if $(x^n,y^n) \in T_{XY}(k_0\delta)$ then $y^n \in T_Y(k_0\delta|\mathcal{X}|)$, we have

$$\begin{aligned}
\Pr\{E_2\} &\le \Pr\{E_2 \cup E_0^c\} \\
&= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_2\},
\end{aligned}$$

$$\begin{aligned}
&\Pr\{E_0 \cap E_2\} \\
&\le \sum_{k=1}^{L_U} \sum_{y^n\in T_Y(k_0\delta|\mathcal{X}|)} P_Y(y^n) \\
&\qquad \Pr\{(U^n(i(j_U,k)),y^n) \in T_{UY}(k_2\delta)\} \\
&\qquad (\because u^n(i) \text{ is selected independently of } y^n) \\
&\le L_U \exp\{-n(I(Y;U)-\epsilon)\}, \\
&\qquad (\because \text{Lemma 6})
\end{aligned}$$

where ϵ is a function of $(k_0\delta\|\mathcal{X}\|, k_2\delta)$. By setting L_U and k_2 as

$$L_U \le \exp\{n(I(Y;U) - l_1\gamma)\},\ l_1 > 0,$$

$l_1\gamma > \epsilon$ and $k_0|\mathcal{X}| < k_2$, we have $\lim_{n\to\infty} \Pr\{E_2\} = 0$.

Error evaluation: $\widehat{\varphi}_n^{(2)}$

This is almost the same as the case of $\widehat{\varphi}_n^{(1)}$. We have to set

$$L_U \le \exp\{n(I(X;U) - l_2\gamma)\},\ l_2 > 0$$

to vanish the encoding/decoding errors.

Rate evaluation: φ_n

The encoder sends the indexes of the bin using

$$\begin{aligned} R &= \frac{1}{n}\log N_U \\ &= \frac{1}{n}\log\frac{M_U}{L_U} \\ &\ge I(XY;U) + m_1\gamma \\ &\quad - \min\{I(Y;U) - l_1\gamma, I(X;U) - l_2\gamma\} \\ &= \max\{I(X;U|Y) + l_1\gamma, I(Y;U|X) + l_2\gamma\} + m_1\gamma \end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as $\max\{I(X;U|Y), I(Y;U|X)\}$.

This completes the proof of Theorem 1. □

5.3 Theorem 2: converse part

Proof.

The proof of Theorem 2 is quite similar to that of Theorem 1. Let a sequence $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \cdots, \widehat{\varphi}_n^{(M)})\}_{n=1}^{\infty}$ of GCD codes be given that satisfy the conditions of Definitions 4 and 5. From Definition 5, for any $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ such that for all $n \ge n_1(\delta)$, we can obtain

$$\frac{1}{n}\log M_n \le R + \delta.$$

In a similar manner to Theorem 1 we obtain

$$n(R+\delta) \ge \sum_{k=1}^{n} I(\boldsymbol{X}_k^{(\mathcal{S}_j^c)}; A_n\boldsymbol{X}^{k-1}|\boldsymbol{X}_k^{(\mathcal{S}_j^c)}).$$

Let us define random variables $U_k = A_n\boldsymbol{X}^{k-1}$, and let J be a random variable that is independent of $\boldsymbol{X}$ and uniformly distributed over the set $\mathcal{I}_n$. We define a random variable $U = (J, U_J)$. This implies that for every $j \in \mathcal{I}_M$

$$R + \delta \ge I(\boldsymbol{X}^{(\mathcal{S}_j)}; U|\boldsymbol{X}^{(\mathcal{S}_j^c)}).$$

Since $\delta > 0$ is arbitrary for a sufficiently large n, we obtain

$$R \ge \max_{j\in\mathcal{I}_M} I(\boldsymbol{X}^{(\mathcal{S}_j)}; U|\boldsymbol{X}^{(\mathcal{S}_j^c)}).$$

We next show the existence of functions $\phi_{(j,i)}$ ($j \in \mathcal{I}_M,\ i \in \mathcal{S}_j$) that satisfy the conditions of Theorem 2. From Definition 5, for any $\gamma > 0$ there exists an integer $n_2 = n_2(\gamma)$ such that for all $n \ge n_2(\gamma)$

$$\begin{aligned} &D_{j,i} + \gamma \\ &\ge \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_{X^{(i)}}(X_k^{(i)}, \widehat{\varphi}_{n,k}^{(j;i)}(A_n, \boldsymbol{X}^{(\mathcal{S}_j^c)n}))\right], \end{aligned}$$

where $\widehat{\varphi}_{n,k}^{(j;i)}$ ($k \in \mathcal{I}_n$) is the output of $\widehat{\varphi}_n^{(j;i)}$ at time k. We note that $U_k\boldsymbol{X}_k^{(\mathcal{S}_j^c)}$ contains $A_n\boldsymbol{X}^{(\mathcal{S}_j^c)k}$, which implies that $\boldsymbol{X}_{k+1}^{(\mathcal{S}_j^c)n}$ is further needed to generate $\widehat{\boldsymbol{X}}_k^{(\mathcal{S}_j)}$ from $U_k\boldsymbol{X}_k^{(\mathcal{S}_j^c)}$. Here, let us define the distribution Q_{k_1,k_2} of $A_n\boldsymbol{X}^{(\mathcal{S}_j)k_1}\boldsymbol{X}^{(\mathcal{S}_j^c)k_2}$, namely for any $\boldsymbol{x}^{(\mathcal{S}_j)k_1} \in \mathcal{X}^{(\mathcal{S}_j)k_1}$, $\boldsymbol{x}^{(\mathcal{S}_j^c)k_2} \in \mathcal{X}^{(\mathcal{S}_j^c)k_2}$ and $a_n \in \mathcal{I}_{M_n}$

$$\begin{aligned} &Q_{k_1,k_2}(a_n, \boldsymbol{x}^{(\mathcal{S}_j)k_1}, \boldsymbol{x}^{(\mathcal{S}_j^c)k_2}) \\ &\stackrel{\text{def.}}{=} \Pr\{\varphi_n(\boldsymbol{X}^n) = a_n, \\ &\qquad \boldsymbol{X}^{(\mathcal{S}_j)k_1} = \boldsymbol{x}^{(\mathcal{S}_j)k_1}, \boldsymbol{X}^{(\mathcal{S}_j^c)k_2} = \boldsymbol{x}^{(\mathcal{S}_j^c)k_2}\} \\ &= \sum_{\substack{\left(\boldsymbol{x}_{k_1+1}^{(\mathcal{S}_j)n}, \boldsymbol{x}_{k_2+1}^{(\mathcal{S}_j^c)n}\right) \in \mathcal{X}^{(\mathcal{S}_j)n-k_1} \times \mathcal{X}^{(\mathcal{S}_j^c)n-k_2}: \\ \varphi_n(\boldsymbol{x}^{(\mathcal{I}_N)n}) = a_n}} P_{\boldsymbol{X}^n}(\boldsymbol{x}^{(\mathcal{I}_N)n}). \end{aligned}$$

Also, let $Q_k^{(j)}$ be the distribution of $\boldsymbol{X}_k^{(\mathcal{S}_j)}$ given $U_k\boldsymbol{X}_k^{(\mathcal{S}_j^c)}$, namely for any $u_k = a_n\boldsymbol{x}^{(\mathcal{I}_N)k-1}$

$$\begin{aligned} &Q_k^{(j)}(\boldsymbol{x}_k^{(\mathcal{S}_j)}|u_k, \boldsymbol{x}_k^{(\mathcal{S}_j^c)}) \\ &\stackrel{\text{def.}}{=} \frac{Q_{k,k}(a_n, \boldsymbol{x}^{(\mathcal{S}_j)k}, \boldsymbol{x}^{(\mathcal{S}_j^c)k})}{Q_{k-1,k}(a_n, \boldsymbol{x}^{(\mathcal{S}_j)k-1}, \boldsymbol{x}^{(\mathcal{S}_j^c)k})}. \end{aligned}$$

Further, let us define $\widetilde{\boldsymbol{X}}_{k+1}^{(\mathcal{S}_j^c)n}(U_k, \boldsymbol{X}_k^{(\mathcal{S}_j^c)}, i)$ as random variables selected to minimize the average distortion between $X_k^{(i)}$ and the output of $\widehat{\varphi}_{n,k}^{(j;i)}$ ($i \in \mathcal{S}_j$) given $U_k\boldsymbol{X}_k^{(\mathcal{S}_j^c)}$, namely

$$\begin{aligned} &\widetilde{\boldsymbol{X}}_{k+1}^{(\mathcal{S}_j^c)n}(U_k, \boldsymbol{X}_k^{(\mathcal{S}_j^c)}, i) \stackrel{\text{def.}}{=} \mathop{\arg\min}_{\boldsymbol{X}_{k+1}^{(\mathcal{S}_j^c)n} \in \mathcal{X}^{(\mathcal{S}_j^c)n-k}} \\ &\sum_{\boldsymbol{X}_k^{(\mathcal{S}_j)} \in \mathcal{X}^{(\mathcal{S}_j)}} Q_k^{(j)}(\boldsymbol{X}_k^{(\mathcal{S}_j)}|U_k\boldsymbol{X}_k^{(\mathcal{S}_j^c)}) \\ &\qquad \Delta_{X^{(i)}}(\boldsymbol{X}_k^{(i)}, \widehat{\varphi}_{n,k}^{(j;i)}(A_n, \boldsymbol{X}^{(\mathcal{S}_j^c)n})). \end{aligned}$$

We choose the functions $\phi_{(j;i)}$ as follows:

$$\begin{aligned} &\phi_{(j;i)k}(U_k, \boldsymbol{X}_k^{(\mathcal{S}_j^c)}) \\ &\stackrel{\text{def.}}{=} \widehat{\varphi}_{n,k}^{(j;i)}(A_n, \boldsymbol{X}^{(\mathcal{S}_j^c)k} * \widetilde{\boldsymbol{X}}_{k+1}^{(\mathcal{S}_j^c)n}(U_k, \boldsymbol{X}_k^{(\mathcal{S}_j^c)}, i)), \\ &\phi_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}) \stackrel{\text{def.}}{=} \phi_{(j;i)J}(U_J, \boldsymbol{X}^{(\mathcal{S}_j^c)}) \end{aligned}$$

In a similar way to Theorem 1, we obtain

$$
\begin{aligned}
&D_{j,i} + \gamma \\
&\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_{X^{(i)}}(X_k^{(i)}, \widehat{\varphi}_{n,k}^{(j;i)}(A_n, \boldsymbol{X}^{(\mathcal{S}_j^c)n}))\right] \\
&\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_{X^{(i)}}(X_k^{(i)}, \phi_{(j;i)k}(U_k, \boldsymbol{X}_k^{(\mathcal{S}_j^c)}))\right] \\
&= E\left[\Delta_{X^{(i)}}(X^{(i)}, \phi_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))\right].
\end{aligned}
$$

Since $\gamma > 0$ is arbitrary for a sufficiently large n, we obtain

$$D_{j,i} \geq E\left[\Delta_{X^{(i)}}(X^{(i)}, \phi_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))\right].$$

It remains to establish that the bound on $|\mathcal{U}|$ specified in Theorem 2 does not affect the determination of the inf achievable rate $R(\boldsymbol{X}|\boldsymbol{D})$. In a similar way to Theorem 1, we then define the following functions of a generic distribution $Q \in \mathcal{P}(\mathcal{X}^{(\mathcal{I}_N)})$:

$$
\begin{aligned}
&q_1(Q, \boldsymbol{x}^{(\mathcal{I}_N)}) = Q(\boldsymbol{x}^{(\mathcal{I}_N)}) \\
&q_2(Q) = \max_{j\in\mathcal{I}_M} q_{2,j}(Q), \\
&q_{2,j}(Q) = H(\boldsymbol{X}^{(\mathcal{S}_j)}|\boldsymbol{X}^{(\mathcal{S}_j^c)}) \\
&\quad - \sum_{\boldsymbol{x}^{(\mathcal{I}_N)}\in\mathcal{X}^{(\mathcal{I}_N)}} Q(\boldsymbol{x}^{(\mathcal{I}_N)}) \log \frac{\sum_{\widetilde{\boldsymbol{x}}^{(\mathcal{S}_j)}\in\mathcal{X}^{(\mathcal{S}_j)}} Q(\widetilde{\boldsymbol{x}}^{(\mathcal{S}_j)}, \boldsymbol{x}^{(\mathcal{S}_j^c)})}{Q(\boldsymbol{x}^{(\mathcal{I}_N)})}, \\
&q_{3,m(j,i)}(Q) = \sum_{\boldsymbol{x}^{(\mathcal{S}_j^c)}\in\mathcal{X}^{(\mathcal{S}_j^c)}} \min_{\widehat{x}^{(i)}\in\widehat{\mathcal{X}}^{(i)}} \\
&\quad \sum_{\boldsymbol{x}^{(\mathcal{S}_j)}\in\mathcal{X}^{(\mathcal{S}_j)}} Q(\boldsymbol{x}^{(\mathcal{I}_N)})\Delta_{X^{(i)}}(x^{(i)}, \widehat{x}^{(i)}),
\end{aligned}
$$

where $j \in \mathcal{I}_M$ $i \in \mathcal{S}_j$ and $m(j,i)$ denotes the serial number of the source X_i contained in the index set $\mathcal{S}_j$ defined as follows:

$$m(j,i) \stackrel{\text{def.}}{=} \left|\{\widetilde{i} \in \mathcal{S}_j | \widetilde{i} \leq i\}\right| + \sum_{\widetilde{j}=1}^{j-1} |\mathcal{S}_j|.$$

Note that $|\mathcal{X}^{(\mathcal{I}_N)}| - 1$ functions are needed to preserve the distribution $Q(\boldsymbol{x}^{(\mathcal{I}_N)})$, and $\sum_{j=1}^{M} |\mathcal{S}_j|$ functions to preserve the average distortion characterized by the generic distribution Q. From the support lemma, we can find a generic distribution $\alpha \in \mathcal{P}(\widetilde{\mathcal{U}})$ such that $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$,

$$|\widetilde{\mathcal{U}}| \leq |\mathcal{X}^{(\mathcal{I}_N)}| + \sum_{j=1}^{M} |\mathcal{S}_j|$$

and the following equations are simultaneously satisfied:

$$\sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) q_1(P_{\boldsymbol{X}|U}(\cdot|u), \boldsymbol{x}^{(\mathcal{I}_N)}) = P_{\boldsymbol{X}}(\boldsymbol{x}^{(\mathcal{I}_N)}), \quad (4)$$

$$
\begin{aligned}
&\sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) q_2(P_{\boldsymbol{X}|U}(\cdot|u)) \\
&\qquad = \max_{j\in\mathcal{I}_M} I(\boldsymbol{X}^{(\mathcal{S}_j)}; U|\boldsymbol{X}^{(\mathcal{S}_j^c)}), \\
&\sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) q_{3,m(j,i)}(P_{\boldsymbol{X}|U}(\cdot|u)) \\
&= \sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) \sum_{\boldsymbol{x}^{(\mathcal{S}_j^c)}\in\mathcal{X}^{(\mathcal{S}_j^c)}} \min_{\widehat{x}^{(i)}\in\widehat{\mathcal{X}}^{(i)}} \\
&\quad \sum_{\boldsymbol{x}^{(\mathcal{S}_j)}\in\mathcal{X}^{(\mathcal{S}_j)}} P_{\boldsymbol{X}|U}(\boldsymbol{x}^{(\mathcal{I}_N)}|u)\Delta_{X^{(i)}}(x^{(i)}, \widehat{x}^{(i)}).
\end{aligned}
$$

Here, let us define functions $\phi^*_{(j;i)} : \widetilde{\mathcal{U}} \times \mathcal{X}^{(\mathcal{S}_j^c)} \to \widehat{\mathcal{X}}^{(i)}$ $(j \in \mathcal{I}_M, i \in \mathcal{S}_j)$ that satisfy

$$
\begin{aligned}
&\phi^*_{(j;i)}(u, \boldsymbol{x}^{(\mathcal{S}_j^c)}) = \arg\min_{\widehat{x}^{(i)}\in\widehat{\mathcal{X}}^{(i)}} \\
&\quad \sum_{\boldsymbol{x}^{(\mathcal{S}_j)}\in\mathcal{X}^{(\mathcal{S}_j)}} P_{\boldsymbol{X}|U}(\boldsymbol{x}^{(\mathcal{I}_N)}|u)\Delta_{X^{(i)}}(x^{(i)}, \widehat{x}^{(i)}).
\end{aligned}
$$

With these definitions, we have

$$
\begin{aligned}
&\sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) q_{3,m(j,i)}(P_{\boldsymbol{X}|U}(\cdot|u)) \\
&\quad = E[\Delta_{X^{(i)}}(X^{(i)}, \phi^*_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))]
\end{aligned}
$$

and

$$
\begin{aligned}
D_{j,i} &\geq E[\Delta_{X^{(i)}}(X^{(i)}, \phi_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))] \\
&\geq E[\Delta_{X^{(i)}}(X^{(i)}, \phi^*_{(j;i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))].
\end{aligned}
$$

Hence, $\phi^*_{(j;i)}$ satisfies the conditions of Theorem 2. Further, Eq.(4) implies that there exist a random variable $\widetilde{U}$ and a joint distribution $P_{\widetilde{U}\boldsymbol{X}}$ that satisfy

$$\alpha(u) P_{\boldsymbol{X}|U}(\boldsymbol{x}^{(\mathcal{I}_N)}|u) = P_{\widetilde{U}\boldsymbol{X}}(u, \boldsymbol{x}^{(\mathcal{I}_N)})$$

for all $(u, \boldsymbol{x}^{(\mathcal{I}_N)}) \in \widetilde{\mathcal{U}} \times \mathcal{X}^{(\mathcal{I}_N)}$. The new joint distribution preserves the distribution $P_{\boldsymbol{X}}$

$$
\begin{aligned}
\sum_{u\in\widetilde{\mathcal{U}}} P_{\widetilde{U}\boldsymbol{X}}(u, \boldsymbol{x}^{(\mathcal{I}_N)}) &= \sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) P_{\boldsymbol{X}|U}(\boldsymbol{x}^{(\mathcal{I}_N)}|u) \\
&= P_{\boldsymbol{X}}(\boldsymbol{x}^{(\mathcal{I}_N)}).
\end{aligned}
$$

This completes the proof of the converse part of Theorem 2. □

5.4 Theorem 2: direct part

Proof.
The proof of Theorem 2 is quite similar to that of Theorem 1. Let a set $\boldsymbol{D}$ of distortion criteria be given, and $P_{U|\boldsymbol{X}} \in \mathcal{P}_{CD}(\mathcal{U}|P_{\boldsymbol{X}})$. Fix arbitrary $\gamma, \delta > 0$.

<u>Codeword selection: φ_n</u>
The same way as Theorem 1.

Encoding: φ_n
Almost the same way as Theorem 1.
(1) For an input set $\boldsymbol{x}^{(\mathcal{I}_N)n} \in \mathcal{X}^{(\mathcal{I}_N)n}$ of sequences, the encoder seeks a vector $u^n(i) \in \mathcal{A}_U$ such that $(u^n(i), \boldsymbol{x}^{(\mathcal{I}_N)n}) \in T_{U\boldsymbol{X}}(k_1\delta)$, where $k_1 > 0$. The selected vector is denoted by $u^n(\boldsymbol{x}^{(\mathcal{I}_N)n})$.
(2) The value assigned to the encoder $\varphi_n(\cdot)$ is the bin index to which $u^n(\boldsymbol{x}^{(\mathcal{I}_N)n})$ belongs, that is,

$$\varphi_n(\boldsymbol{x}^{(\mathcal{I}_N)n}) = j, \quad u^n(\boldsymbol{x}^{(\mathcal{I}_N)n}) \in \mathcal{A}_U(j).$$

Decoding: $\widehat{\varphi}_n^{(j)}$
Almost the same way as Theorem 1.
(1) The decoder has access to the indexes j_U received from the encoder φ_n and the sequence set $\boldsymbol{x}^{(\mathcal{S}_j^c)n} \in \mathcal{X}^{(\mathcal{S}_j^c)n}$.
(2) The decoder seeks a unique vector $u^n \in \mathcal{A}_U(j_U)$ such that $(u^n, \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \in T_{U\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_{2,j}\delta)$, where $k_{2,j} > 0$. This vector is denoted by $\widehat{u}^n(\boldsymbol{x}^{(\mathcal{S}_j^c)n})$.
(3) The reconstruction vector $\widehat{\boldsymbol{x}}^{(\mathcal{S}_j)n}$ is given by

$$\begin{aligned}
\widehat{\boldsymbol{x}}^{(\mathcal{S}_j)n} &= \{\widehat{x}^{(i;j)n} \mid i \in \mathcal{S}_j\}, \\
\widehat{x}^{(i;j)n} &= (\widehat{x}_1^{(i;j)}, \cdots, \widehat{x}_n^{(i;j)}), \\
\widehat{x}_k^{(i;j)} &= \phi_{(j;i)}(\widehat{u}_k(\boldsymbol{x}^{(\mathcal{S}_j^c)n}), \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \quad (k \in \mathcal{I}_n),
\end{aligned}$$

where $\widehat{u}_k(\boldsymbol{x}^{(\mathcal{S}_j^c)n})$ is the k-th element of $u^n(\boldsymbol{x}^{(\mathcal{S}_j^c)n})$.
Distortion evaluation: $\widehat{\varphi}_n^{(j)}$
In the same way as Theorem 1, we obtain

$$\begin{aligned}
&\Delta^n_{X^{(i)}}(x^{(i)n}, \widehat{x}^{(i;j)n}) \\
&= \frac{1}{n} \sum_{(u, x^{(\mathcal{I}_N)}) \in \mathcal{U} \times \mathcal{X}^{(\mathcal{I}_N)}} N(u, \boldsymbol{x}^{(\mathcal{I}_N)} | \widehat{u}^n(\boldsymbol{x}^{(\mathcal{S}_j^c)n}), \boldsymbol{x}^{(\mathcal{I}_N)n}) \\
&\qquad \Delta_{X^{(i)}}(x^{(i)}, \phi_{(j;i)}(u, \boldsymbol{x}^{(\mathcal{S}_j^c)})) \\
&\le \sum_{(u, \boldsymbol{x}^{(\mathcal{I}_N)}) \in \mathcal{U} \times \mathcal{X}^{(\mathcal{I}_N)}} (P_{U\boldsymbol{X}}(u, \boldsymbol{x}^{(\mathcal{I}_N)}) + k_1\delta) \\
&\qquad \Delta_{X^{(i)}}(x^{(i)}, \phi_{(j,i)}(u, \boldsymbol{x}^{(\mathcal{S}_j^c)})) \\
&\le E\left[\Delta_{X^{(i)}}(X^{(i)}, \phi_{(j,i)}(U, \boldsymbol{X}^{(\mathcal{S}_j^c)}))\right] \\
&\qquad + k_1\delta\overline{\Delta}_{X^{(i)}} |\mathcal{U} \times \mathcal{X}^{(\mathcal{I}_N)}| \\
&\le D_{j,i} + k_1\delta\overline{\Delta}_{X^{(i)}} |\mathcal{U} \times \mathcal{X}^{(\mathcal{I}_N)}|.
\end{aligned}$$

We denote error probabilities in the encoding/decoding process as P_e^n. Then, the average distortion can be bounded as

$$\begin{aligned}
&E\left[\Delta^n_{X^{(i)}}(X^{(i)n}, \widehat{X}^{(i;j)n})\right] \\
&\le (1 - P_e^n)(D_{j,i} + k_1\delta\overline{\Delta}_{X^{(i)}} |\mathcal{U} \times \mathcal{X}^{(\mathcal{I}_N)}|) + P_e^n \overline{\Delta}_{X^{(i)}}.
\end{aligned}$$

Since $\delta > 0$ is arbitrarily small for a sufficiently large n, if P_e^n vanishes as $n \to \infty$, we can obtain

$$\limsup_{n \to \infty} E\left[\Delta^n_{X^{(i)}}(X^{(i)n}, \widehat{X}^{(i;j)n})\right] \le D_{j,i}.$$

Error evaluation: φ_n
If there is no $u^n \in \mathcal{A}_U$ such that $(u^n, \boldsymbol{x}^{(\mathcal{I}_N)n}) \in T_{U\boldsymbol{X}}(k_1\delta)$, an encoding error has occurred. This event is denoted as

$$E_1 \stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_U} \left\{(u^n(i), \boldsymbol{x}^{(\mathcal{I}_N)n}) \notin T_{U\boldsymbol{X}}(k_1\delta)\right\}.$$

Here, let us define

$$E_0 \stackrel{\text{def.}}{=} \{(\boldsymbol{x}^{(\mathcal{I}_N)n}) \in T_{\boldsymbol{X}}(k_0\delta)\},$$

where $k_0 > 0$. From Lemma 4, $\Pr\{E_0^c\} \to 0$ as $n \to \infty$. Then, in a similar manner to Theorem 1, we have

$$\begin{aligned}
\Pr\{E_1\} &\le \Pr\{E_1 \cup E_0^c\} \\
&= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_1\},
\end{aligned}$$

$$\Pr\{E_0 \cap E_1\} \to 0 \quad (n \to \infty)$$

by setting M_U, k_1 and k_0 as

$$M_U \ge \exp\{n(I(\boldsymbol{X}; U) + m_1\gamma)\}, \ m_1 > 0,$$

$m_1\gamma > \epsilon_u = \epsilon_u(k_1\delta, k_0\delta)$ and $k_1 < k_0$.

Error evaluation: $\widehat{\varphi}_n^{(j)}$
If there is no or more than one $u^n_{(i)} \in \mathcal{A}_U(j_U)$ such that $(u^n_{(i)}, \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \in T_{U\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_2\delta)$, a decoding error is declared. This event is classified into two cases.
(1) First case:

$$(u^n(\boldsymbol{x}^{(\mathcal{I}_N)n}), \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \notin T_{U\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_2\delta).$$

However, this error does not occur by setting k_2 as $k_2 > k_1|\mathcal{X}^{(\mathcal{S}_j)}|$ because

$$(u^n(\boldsymbol{x}^{(\mathcal{I}_N)n}), \boldsymbol{x}^{(\mathcal{I}_N)n}) \in T_{U\boldsymbol{X}}(k_1\delta)$$

and Lemma 5.
(2) Second case: If there exists $u^n \in \mathcal{A}_U(j_U)$, $u^n \ne u^n(\boldsymbol{x}^{(\mathcal{I}_N)n})$ such that $(u^n, \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \in T_{U\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_2\delta)$. This event is denoted as

$$E_2 \stackrel{\text{def.}}{=} \bigcup_{\substack{u^n \in \mathcal{A}_U(j_U) \\ u^n \ne u^n(\boldsymbol{x}^{(\mathcal{I}_N)n})}} \{(u^n, \boldsymbol{x}^{(\mathcal{S}_j^c)n}) \in T_{U\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_2\delta)\}.$$

Note that if $(\boldsymbol{x}^{(\mathcal{I}_N)n}) \in T_{\boldsymbol{X}}(k_0\delta)$ then

$$\boldsymbol{x}^{(\mathcal{S}_j^c)n} \in T_{\boldsymbol{X}^{(\mathcal{S}_j^c)}}(k_0\delta|\mathcal{X}^{(\mathcal{S}_j)}|).$$

Therefore, we have

$$\begin{aligned}
\Pr\{E_2\} &\le \Pr\{E_2 \cup E_0^c\} \\
&= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_2\},
\end{aligned}$$

$$\Pr\{E_0 \cap E_2\} \to 0 \quad (n \to \infty)$$

in a similar manner to Theorem 1 by setting L_U, k_2 as

$$L_U \le \exp\{n(I(\boldsymbol{X}^{(\mathcal{S}_j^c)}; U) - l_{1j}\gamma)\}, \ l_1 > 0,$$

$l_{1j}\gamma > \epsilon = \epsilon(k_0\delta|\mathcal{X}^{(\mathcal{S}_j)}|, k_2\delta)$ and $k_0|\mathcal{X}^{(\mathcal{S}_j)}| < k_2$.

Rate evaluation: φ_n

The encoder sends the indexes of the bin using

$$\begin{aligned} R &= \frac{1}{n}\log N_U \\ &= \frac{1}{n}\log\frac{M_U}{L_U} \\ &\geq I(\boldsymbol{X}; U) + m_1\gamma - \min_{j\in\mathcal{I}_M}\{I(\boldsymbol{X}^{(\mathcal{S}_j^c)}; U) - l_{1j}\gamma\} \\ &= \max_{j\in\mathcal{S}_j}\{I(\boldsymbol{X}^{(\mathcal{S}_j)}; U|\boldsymbol{X}^{(\mathcal{S}_j^c)}) + l_{1j}\gamma\} + m_1\gamma \end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as $\max_{j\in\mathcal{S}_j} I(\boldsymbol{X}^{(\mathcal{S}_j)}; U|\boldsymbol{X}^{(\mathcal{S}_j^c)})$.

This completes the proof of Theorem 2. □

Acknowledgments

The authors wish to thank Prof. Ryutaroh Matsumoto of Tokyo Institute of Technology for his support. The authors also thank Dr. Yoshinobu Tonomura, Dr. Hiromi Nakaiwa, Dr. Shoji Makino, Dr. Junji Yamato and Dr. Kunio Kashino of NTT Communication Science Laboratories for their help. Lastly, the authors thank the associate editor and the anonymous reviewers for their constructive remarks and suggestions, which led to the improvement of this work.

References

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, Vol.19, No.4, pp.471–480, July 1973.
[2] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inf. Theory, Vol.22, No.1, pp.1–10, January 1976.
[3] T. Berger, "Multiterminal source coding," in The information theory approach to communications, ed. G. Longo, CISM Courses and Lectures, Vol.229, pp.171–231, Springer-Verlag, 1978.
[4] S.Y. Tung, Multiterminal source coding, Ph.D. thesis, School of Electrical Engineering, Cornell University, May 1978.
[5] A.D. Wyner, "On source coding with side information at the decoder," IEEE Trans. Inf. Theory, Vol.21, No.3, pp.294–300, May 1975.
[6] A. Sgarro, "Source coding with side information at several decoders," IEEE Trans. Inf. Theory, Vol.23, No.2, pp.179–182, March 1977.
[7] J. Körner and K. Marton, "Images of a set via two channels and their role in multi-user communication," IEEE Trans. Inf. Theory, Vol.23, No.6, pp.751–761, November 1975.
[8] H.S. Witsenhausen, "On source networks with minimal breakdown degradation," Bell Systems Technical Journal, Vol.59, No.6, pp.1083–1087, July 1980.
[9] J. Wolf, A.D. Wyner, and J. Ziv, "Source coding for multiple descriptions," Bell Systems Technical Journal, Vol.59, No.8, pp.1417–1426, October 1980.
[10] A.A. ElGamal and T.M. Cover, "Achievable rates for multiple descriptions," IEEE Trans. Inf. Theory, Vol.28, No.6, pp.851–857, November 1982.
[11] H. Yamamoto, "Source coding theory for cascade and branching communication systems," IEEE Trans. Inf. Theory, Vol.27, No.3, pp.299–308, May 1981.
[12] H. Yamamoto, "Source coding theory for a triangular communication system," IEEE Trans. Inf. Theory, Vol.42, No.3, pp.848–853, May 1996.
[13] W. Gu and M. Effros, "Source coding for a multiple network," Proc. Data Compression Conference (DCC), p.461, 2005.
[14] W. Gu and M. Effros, "Source coding for a simple multiple network," Proc. IEEE International Symposium on Information Theory (ISIT), pp.2335–2339, 2005.
[15] F.M.J. Willems, J.K. Wolf, and A.D. Wyner, "Communicating via a processing broadcast satellite," Proc. of the 1989 IEEE/CAM Information Theory Workshop, June 1989.
[16] A.D. Wyner, J.K. Wolf, and F.M.J. Willems, "Communicating via a processing broadcast satellite," IEEE Trans. Inf. Theory, Vol.48, No.6, pp.1243–1249, June 2002.
[17] A. Kimura, T. Uyematsu, and S. Kuzuoka, "Universal coding for correlated sources with complementary delivery," IEICE Trans. Fundamentals, Vol.E90-A, No.9, pp.1840–1847, September 2007.
[18] A. Kimura, T. Uyematsu, S. Kuzuoka, and S. Watanabe, "Universal coding for correlated sources with generalized complementary delivery networks," IEEE Trans. Inf. Theory, 2007. submitted.
[19] S. Kuzuoka, A. Kimura, and T. Uyematsu, "Simple coding schemes for lossless and lossy complementary delivery problems," Proc. Shannon Theory Workshop (STW), pp.43–50, September 2007.
[20] S. Kuzuoka, A. Kimura, and T. Uyematsu, "Simple coding schemes for lossless and lossy complementary delivery problems," IEEE Trans. Inf. Theory, 2008. submitted [Online] Available: http://jp.arxiv.org/abs/0802.1567.
[21] S.R. Li, R.W. Yeung, and N. Cai, "Linear network coding," IEEE Trans. Inf. Theory, Vol.49, No.2, pp.371–381, February 2003.
[22] T. Berger, Rate distortion theory: A mathematical basis for data compression, Prentice Hall, Englewood Cliffs, NJ, 1971.
[23] I. Csiszár and J. Körner, Information theory: Coding theorems for discrete memoryless systems, Academic Press, New York, 1981.
[24] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," IEEE Trans. Inf. Theory, Vol.50, No.8, pp.1636–1654, August 2004.

Akisato Kimura received B.E., M.E. and D.E. degrees from Tokyo Institute of Technology in 1998, 2000 and 2007, respectively. In 2000, he joined NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Japan, where he is currently a research scientist in Media Information Laboratory. He has been working on multimedia information retrieval, perceptual image processing and multiterminal information theory. His research interests are in the areas of pattern recognition, computer vision and information theory.

Tomohiko Uyematsu received B.E., M.E. and D.E. degrees from Tokyo Institute of Technology in 1982, 1984 and 1988, respectively. From 1984 to 1992, he was with the Department of Electrical and Electronics Engineering of Tokyo Institute of Technology, first as a research associate, next as a lecturer, and lastly as an associate professor. From 1992 to 1997, he was with School of Information Science of Japan Advanced Institute of Science and Technology as an associate professor, and currently he is a professor in the Department of Communications and Integrated Systems. In 1992 and 1996, he was a visiting researcher at the Centre National de la Recherche Scientifique, France and Delft University of Technology, Netherlands, respectively. He received the Shinohara Memorial Young Engineer Award in 1989, and the Best Paper Award in 1993, 1996, 2002 and 2007 all from IEICE. His current research interests are in the areas of information theory, especially Shannon theory and multiterminal information theory. Dr. Uyematsu is a senior member of IEEE.