Multiterminal source coding for cascading and feedback refinement systems

Akisato Kimura Tomohiko Uyematsu *

2006.11.14

No. AK-TR-2006-03

Abstract

Lossy coding problems are investigated for some communication systems in the presense of cascading and/or feedback information channels from decoders so as to refine reproduction messages. This framework provides different types of refinement structures from so-called successive refinement. Three different types of communication systems are considered, i.e. refinement systems in the presense of a cascading channel, a feedback channel, and both channels. Outer and inner bounds of achievable rate-distortion regions for those problems are obtained.

Key Words: multiterminal source coding, scalable coding, side information, cascading, refinement, feedback

*The authors are with Department of Communications and Integrated Systems, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan. Akisato Kimura is also with NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan. E-mail: akisato@eye.brl.ntt.co.jp, uematsu@it.ss.titech.ac.jp

1 Introduction

We consider some coding problems for correlated information sources. The situations we investigate here involve cascading and feedback transmission from decoders. Three types of communication systems are investigated.

Figure 1 shows a block diagram of the first coding problem. There are two encoders, both of which observe a message from a source X and deliver the message to corresponding decoders. These two decoders has access to side information Y. A cascading channel is placed from one decoder to the other, and some amount of information is sent via the channel to refine a reproducted message.

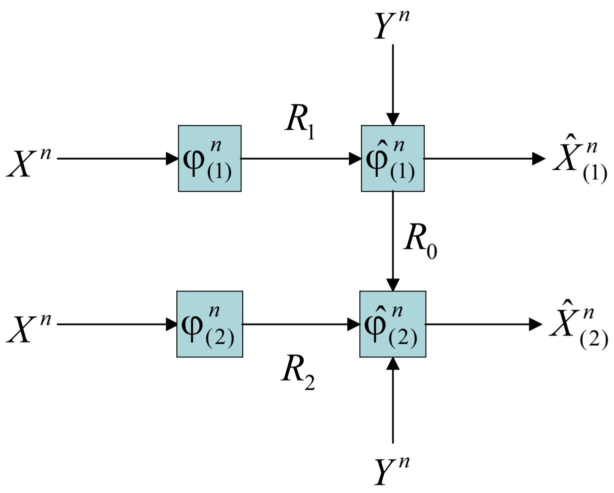

Figure 1: Cascading refinement system (cascading channel is placed)

Figure 2 shows a block diagram of the second coding problem, which is quite a similar to the first one except a channel from the first decoder. The channel delivers feedback information from the first decoder to the second encoder.

Figure 3 shows a block diagram of the third coding problem, which involves cascading and feedback channels. This communication system models a certain type of information retrieval with index structures. We will show the detail in another technical report [1].

In this setting, we consider rate-distortion problems and clarify outer bounds and inner bounds of the achievable rate-distortion regions. These bounds coincide for some special cases.

1.1 Preliminaries

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, $|\mathcal{X}|$ be the cardinality of $\mathcal{X}$ and $\mathcal{I}_M = \{1, 2, \cdots, M\}$. A member of $\mathcal{X}^n$ is written as $x^n = (x_1, x_2, \cdots, x_n)$, and substrings of x^n are

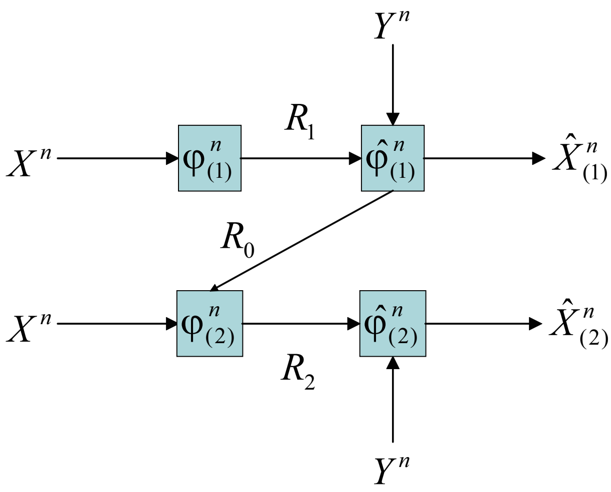

Figure 2: Cascading refinement system (feedback channel is placed)

written as $x_i^j = (x_i, x_{i+1}, \cdots, x_j)$ for $i \le j$. When the dimension is clear from the content, vectors will be denoted by boldface letters, i.e., $\boldsymbol{x} \in \mathcal{X}^n$. $\mathcal{M}(\mathcal{X})$ denotes the set of all probability distributions on $\mathcal{X}$. Also, $\mathcal{M}(\mathcal{X}|P_Y)$ denotes the set of all probability distributions on $\mathcal{X}$ given a distribution $P_Y \in \mathcal{M}(\mathcal{Y})$, namely each member of $\mathcal{M}(\mathcal{X}|P_Y)$ is characterized by $P_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ as $P_{XY} = P_{X|Y}P_Y$. A discrete memoryless source $(\mathcal{X}, P_X)$ is an infinite sequence $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ of independent copies of a random variable X taking values in $\mathcal{X}$ with a generic distribution $P_X \in \mathcal{M}(\mathcal{X})$, namely

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

We will denote a source $(\mathcal{X}, P_X)$ by referring to its generic distribution P_X or random variable X. For a correlated source (X, Y), $H(X|Y)$ denotes a conditional entropy of X given Y. Similarly, for a correlated source (X, Y, Z), $I(X; Y|Z)$ denotes a conditional mutual information between X and Y given Z. A similar convention is used for other random variables and vectors. In the following, all bases of exponentials and logarithms are set at e (the base of the natural logarithm). We are interested in coding the source X. Let $\widehat{\mathcal{X}}$ stand for a reconstruction alphabet, and let $\Delta : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \infty)$ be a single-letter distortion function. The vector distortion function is defined in the usual way, i.e.

$$\Delta_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) \quad = \quad \frac{1}{n} \sum_{k=1}^{n} \Delta_X(x_k, \widehat{x}_k). \tag{2}$$

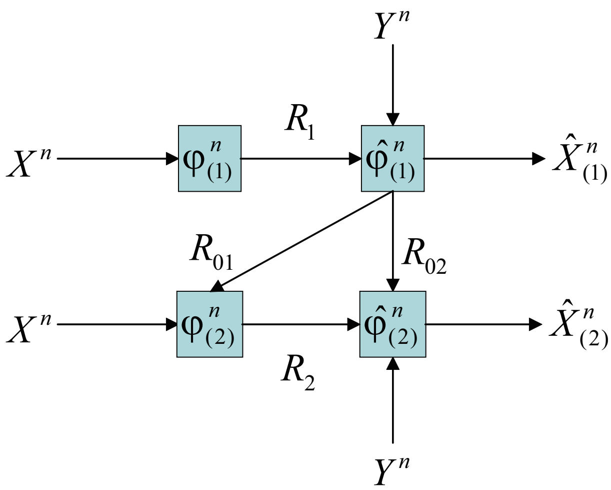

Figure 3: Cascading refinement system (both cascading and feedback channels are placed)

2 Refinement with cascading channels

2.1 Problem formulation

Definition 1. (CR (Cascading Refinement) code)
A set $(\varphi^n_{(0)},\ \varphi^n_{(1)},\ \varphi^n_{(2)},\ \widehat{\varphi}^n_{(1)},\ \widehat{\varphi}^n_{(2)})$ of encoders and decoders is a CR code $(n, M_n^{(0)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})$ for the source (X, Y) if and only if

$$\begin{aligned}
\varphi^n_{(1)} &: \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}} \\
\varphi^n_{(0)} &: \mathcal{Y}^n \times \mathcal{I}_{M_n^{(1)}} \to \mathcal{I}_{M_n^{(0)}} \\
\varphi^n_{(2)} &: \mathcal{X}^n \to \mathcal{I}_{M_n^{(2)}} \\
\widehat{\varphi}^n_{(1)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\
\widehat{\varphi}^n_{(2)} &: \mathcal{I}_{M_n^{(0)}} \times \mathcal{I}_{M_n^{(2)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n,
\end{aligned}$$

where

$$\begin{aligned}
&\rho_n^{(1)} = E\left[\Delta^n(X^n, \widehat{X}^n_{(1)})\right], \quad \rho_n^{(2)} = E\left[\Delta^n(X^n, \widehat{X}^n_{(2)})\right], \\
&A_n^{(1)} = \varphi^n_{(1)}(X^n),\ A_n^{(0)} = \varphi^n_{(0)}(A_n^{(1)}, Y^n),\ A_n^{(2)} = \varphi^n_{(2)}(X^n), \\
&\widehat{X}^n_{(1)} = \widehat{\varphi}^n_{(1)}(A_n^{(1)}, Y^n), \widehat{X}^n_{(2)} = \widehat{\varphi}^n_{(2)}(A_n^{(0)}, A_n^{(2)}, Y^n).
\end{aligned}$$

Definition 2. (CR-achievable rate triad)
(R_0, R_1, R_2) is a CR-achievable rate triad of the source (X, Y) for a given

distortion pair (D_1, D_2) if and only if there exists a sequence of CR codes $\{(n, M_n^{(0)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})\}_{n=1}^{\infty}$ for the source (X, Y) such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(i)} \le R_i, \quad (i = 0, 1, 2)$$
$$\limsup_{n\to\infty} \rho_n^{(j)} \le D_j. \quad (j = 1, 2)$$

Definition 3. (CR-achievable rate region)

$$\mathcal{R}_c(X, Y|D_1, D_2) = \{(R_0, R_1, R_2) : (R_0, R_1, R_2) \text{ is a CR-achievable rate triad of } (X, Y) \text{ for } (D_1, D_2)\}.$$

2.2 Statement of results

First, we state the main theorem.

Theorem 1. (Coding theorem of CR code)

$$\begin{aligned} \mathcal{R}_c(X, Y|D_1, D_2) \subseteq \big\{&(R_0, R_1, R_2) : \\ &R_1 \ge I(X; UV|Y), \\ &R_0 \ge I(X; V|Y), \\ &R_2 \ge I(X; W|VY)\big\} \quad \text{(outer bound)} \end{aligned}$$

where the random variables U, V and W whose alphabets are $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, respectively, are selected such that

- *the alphabet sizes are bounded as*

$$\begin{aligned} |\mathcal{U}| &\le |\mathcal{X}| + 1, \\ |\mathcal{W}| &\le |\mathcal{U} \times \mathcal{X}| + 1, \\ |\mathcal{V}| &\le |\mathcal{U} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y}| + 3, \end{aligned}$$

- *The following Markov chains are satisfied:*

$$\begin{aligned} &U \to X \to Y, \\ &W \to UVX \to Y, \end{aligned}$$

- *there exist functions $\phi_{(1)} : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ and $\phi_{(2)} : \mathcal{V} \times \mathcal{W} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ which safisfy*

$$\begin{aligned} D_1 &\ge E\left[\Delta(X, \phi_{(1)}(U, Y))\right], \\ D_2 &\ge E\left[\Delta(X, \phi_{(2)}(V, W, Y))\right]. \end{aligned}$$

An inner bound is obtained with the same functional forms, while the Markov chain is replaced as

$$\begin{aligned} &UV \to X \to Y, \\ &W \to VX \to UY. \end{aligned}$$

Remark . *The Markov condition* $W \to VX \to UY$ *is tight compared with the condition* $W \to UVX \to Y$ *because* $I(W;Y|UVX) \le I(W;UY|VX)$.

Remark . *The Markov conditions* $U \to X \to Y$ *and* $W \to UVX \to Y$ *are equivalent to the condition that the joint distribution* P_{UVWXY} *satisfies*

$$\begin{aligned} &P_{UVWXY}(u,v,w,x,y) \\ &\quad = P_{XY}(x,y)P_{U|X}(u|x)P_{V|UXY}(v|u,x,y)P_{W|UVX}(w|u,v,x). \end{aligned}$$

In a similar manner, the Markov conditions $UV \to X \to Y$ *and* $W \to VX \to UY$ *are equivalent to the condition that the joint distribution* P_{UVWXY} *satisfies*

$$\begin{aligned} &P_{UVWXY}(u,v,w,x,y) \\ &\quad = \quad P_{XY}(x,y)P_{U|X}(u|x)P_{V|UY}(v|u,y)P_{W|VX}(w|v,x). \end{aligned}$$

From Theorem 1, we can obtain the following properties.

Corollary 1. (Compatibility with a known result)
If side information Y *is not available at both of two decoders, the outer bound indicated in Theorem 1 coincides with the inner bound, i.e.*

$$\begin{aligned} \mathcal{R}_c(X|D_1,D_2) = \big\{(R_0,R_1,R_2) :& \\ R_1 &\ge I(X;\widehat{X}_{(1)}V), \\ R_0 &\ge I(X;V), \\ R_2 &\ge I(X;\widehat{X}_{(2)}|V)\big\} \end{aligned}$$

where the random variable V *taking a value in* $\mathcal{V}$ *is selected such that* $|\mathcal{V}| \le |\mathcal{X}| + 2$*. This rate region coincides with the one indicated by Yamamoto [2].*

In the above discussions, we have considered only two-stage refinement. However, they can be easily extended to communication systems with multi-stage refinement. Let us consider N pairs $\{\varphi_{(i)}, \widehat{\varphi}_{(i)}\}_{i\in\mathcal{I}_N}$ of encoders and decoders, and $(N-1)$ cascading encoders $\{\varphi_{(0j)}\}_{j\in\mathcal{I}_{N-1}}$. Let $\boldsymbol{D} = \{D_i\}_{i\in\mathcal{I}_N}$ be a set of distortion criteria, each of which corresponds to the decoder $\widehat{\varphi}_{(i)}$. We define the CR-achievable rate region $\mathcal{R}_c(X,Y|\boldsymbol{D})$ of the source (X,Y) for given distortion criteria $\boldsymbol{D}$ in the same way as for the two-stage refinement system, where R_i $(i \in \mathcal{I}_N)$ corresponds to the rate of the encoder $\varphi_{(i)}$, and R_{0j} $(j \in \mathcal{I}_{N-1})$ corresponds to the rate of the cascading encoder $\varphi_{(0j)}$. Let $U^{(i)}$ $(i \in \mathcal{I}_N)$ be an auxiliary random variable that takes a value in a finite set $\mathcal{U}^{(i)}$, and $V^{(i,j)}$ $(i,j \in \mathcal{I}_{N-1})$ be an auxiliary random variable that takes values in a finite set $\mathcal{V}^{(i,j)}$. For some $\mathcal{S} \subseteq \mathcal{I}_N$ and $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{I}_{N-1}$, let us define

$$\begin{aligned} \boldsymbol{U}^{(\mathcal{S})} &= \{U^{(i)} : i \in \mathcal{S}\}, \\ \boldsymbol{U} &= \boldsymbol{U}^{(\mathcal{I}_N)}, \end{aligned}$$

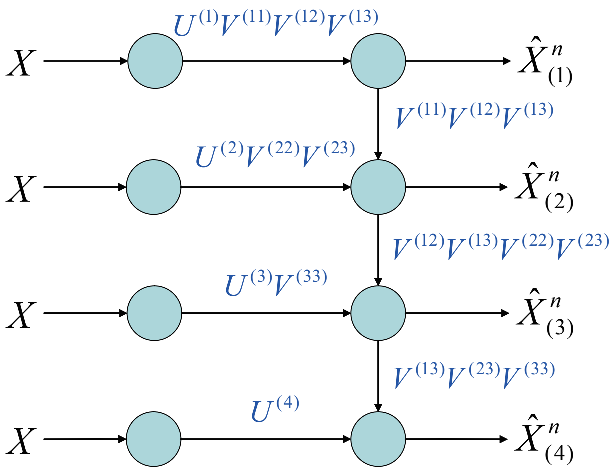

Figure 4: Diagram of 4-stage CR code (without side information for simplicity)

$$
\begin{aligned}
\boldsymbol{V}^{(i,\mathcal{T}_2)} &= \{V^{(i,j)} : j \in \mathcal{T}_2\}, \\
\boldsymbol{V}^{(\mathcal{T}_1,\mathcal{T}_2)} &= \{V^{(i,j)} \in \boldsymbol{V}^{(i,\mathcal{T}_2)} : i \in \mathcal{T}_2\}, \\
\boldsymbol{V} &= \bigcup_{k \in \mathcal{I}_{N-1}} \boldsymbol{V}^{(k,\mathcal{I}_{N-1}-\mathcal{I}_{k-1})}, \\
\mathcal{U}^{(\mathcal{S})} &= \prod_{i \in \mathcal{S}} \mathcal{U}^{(i)}, \\
\mathcal{V}^{(i,\mathcal{T}_2)} &= \prod_{j \in \mathcal{T}_2} \mathcal{V}^{(i,j)}, \\
\mathcal{V}^{(\mathcal{T}_1,\mathcal{T}_2)} &= \prod_{i \in \mathcal{T}_1} \mathcal{V}^{(i,\mathcal{T}_2)}.
\end{aligned}
$$

Corollary 2. (Coding theorem of CR code with multi-stage cascading)

$$
\begin{aligned}
\mathcal{R}_c(X,Y|\boldsymbol{D}) \subseteq \big\{ & (R_{0j}, R_i)_{i \in \mathcal{I}_N, j \in \mathcal{I}_{N-1}} : \\
& R_i \geq I(X; U^{(i)} \boldsymbol{V}^{(i,\mathcal{I}_{N-1}-\mathcal{I}_{i-1})} | \boldsymbol{V}^{(\mathcal{I}_{i-1},\mathcal{I}_{N-1}-\mathcal{I}_{i-2})} Y), \\
& R_{0j} \geq I(X; \boldsymbol{V}^{(\mathcal{I}_j,\mathcal{I}_{N-1}-\mathcal{I}_{j-1})} | Y) \big\} \quad \text{(outer bound)}
\end{aligned}
$$

where the random variables $\boldsymbol{U}$ and $\boldsymbol{V}$ whose alphabets are $\mathcal{U}^{(\mathcal{I}_N)}$ and $\prod_{k \in \mathcal{I}_{N-1}} \mathcal{V}^{(k,\mathcal{I}_{N-1}-\mathcal{I}_{k-1})}$, respectively, are selected such that

- *the alphabet sizes are bounded as*

$$
\begin{aligned}
& |\mathcal{U}^{(i)}| \\
\leq \quad & \left| \mathcal{U}^{(\mathcal{I}_{i-1})} \times \prod_{k=1}^{i-1} \mathcal{V}^{(k,\mathcal{I}_{N-1}-\mathcal{I}_{k-1})} \times \mathcal{X} \right| + 2,
\end{aligned}
$$

$$(i \in \mathcal{I}_N)$$

$$\begin{aligned} |\mathcal{V}^{(i,j)}| & \\ \leq & \left| \mathcal{U}^{(i)} \times \prod_{k=1}^{i-1} \left\{ \mathcal{V}^{(k,\mathcal{I}_{N-1}-\mathcal{I}_{k-1})} \right\} \times \mathcal{V}^{(i,\mathcal{I}_{N-1}-\mathcal{I}_j)} \right. \\ & \left. \times \mathcal{X} \times \mathcal{Y} \right| + (3j+2), \ (i, j \in \mathcal{I}_{N-1}) \end{aligned}$$

- *The following Markov chain is satisfied:*

$$\boldsymbol{U}^{(i)} \to \boldsymbol{V}^{(\mathcal{I}_{i-1},\mathcal{I}_{N-1}-I_{i-2})} X \to \boldsymbol{U}^{(\mathcal{I}_{i-1})} Y$$

- *there exist functions* $\{\phi_{(i)} : \mathcal{U}^{(i)} \times \mathcal{V}^{(\mathcal{I}_{i-1},\mathcal{I}_{N-1}-\mathcal{I}_{i-2})} \times \mathcal{V}^{(i,\mathcal{I}_{N-1}-\mathcal{I}_{i-1})} \times \mathcal{Y} \to \widehat{\mathcal{X}}\}_{i \in \mathcal{I}_N}$ *which safisfy*

$$\begin{aligned} D_i \ \geq \ & E\big[\Delta(X, \phi_{(i)}(U^{(i)}, \\ & \boldsymbol{V}^{(\mathcal{I}_{i-1},\mathcal{I}_{N-1}-\mathcal{I}_{i-2})}, \boldsymbol{V}^{(i,\mathcal{I}_{N-1}-\mathcal{I}_{i-1})}, Y))\big]. \end{aligned}$$

An inner bound is obtained with the same functional forms, while the Markov chains are replaced as

$$\boldsymbol{U}^{(i)} \boldsymbol{V}^{(i,\mathcal{I}_{N-1}-I_{i-1})} \to \boldsymbol{V}^{(\mathcal{I}_{i-1},\mathcal{I}_{N-1}-I_{i-2})} X \to \boldsymbol{U}^{(\mathcal{I}_{i-1})} Y$$

3 Refinement with feedback channels

Next, a *feedback refinement system* is considered, which indicates the refinement system with a feedback channel from the first decoder back to the second decoder.

3.1 Problem formulation

Definition 4. (FR (Feedback Refinement) code)
A set $(\varphi^n_{(0)}, \varphi^n_{(1)}, \varphi^n_{(2)}, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})$ of encoders and decoders is an FR code $(n, M_n^{(0)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})$ for the source (X, Y) if and only if

$$\begin{aligned} \varphi^n_{(1)} &: \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}} \\ \varphi^n_{(0)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \mathcal{I}_{M_n^{(0)}} \\ \varphi^n_{(2)} &: \mathcal{I}_{M_n^{(0)}} \times \mathcal{X}^n \to \mathcal{I}_{M_n^{(2)}} \\ \widehat{\varphi}^n_{(1)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\ \widehat{\varphi}^n_{(2)} &: \mathcal{I}_{M_n^{(2)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \end{aligned}$$

where

$$\rho_n^{(i)} = E\left[\Delta^n(X^n, \widehat{X}^n_{(i)})\right], \quad (i = 1, 2)$$

$$
\begin{aligned}
&A_n^{(1)} = \varphi_{(1)}^n(X^n),\ A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n),\\
&A_n^{(2)} = \varphi_{(2)}^n(A_n^{(0)}, X^n),\\
&\widehat{X}_{(1)}^n = \widehat{\varphi}_{(1)}^n(A_n^{(1)}, Y^n),\ \widehat{X}_{(2)}^n = \widehat{\varphi}_{(2)}^n(A_n^{(2)}, Y^n).
\end{aligned}
$$

Definition 5. (FR-achievable rate triad)
(R_0, R_1, R_2) is an FR-achievable rate triad of the source (X, Y) for a given distortion pair (D_1, D_2) if and only if there exists a sequence of FR codes $\{(n, M_n^{(0)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})\}_{n=1}^{\infty}$ for the source (X, Y) such that

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(i)} \le R_i, \quad (i = 0, 1, 2)\\
&\limsup_{n\to\infty} \rho_n^{(j)} \le D_j. \quad (j = 1, 2)
\end{aligned}
$$

Definition 6. (FR-achievable rate region)

$$
\begin{aligned}
\mathcal{R}_f(X, Y|D_1, D_2) = \{&(R_0, R_1, R_2) : (R_0, R_1, R_2) \text{ is}\\
&\text{an FR-achievable rate triad of } (X, Y) \text{ for } (D_1, D_2)\}.
\end{aligned}
$$

3.2 Statement of results

First, we state the main theorem.

Theorem 2. (Coding theorem of FR code)

$$
\begin{aligned}
\mathcal{R}_f(X, Y|D_1, D_2) \subseteq \big\{&(R_0, R_1, R_2) :\\
&R_1 \ge I(X; U|Y),\\
&R_0 \ge I(Y; V|UX),\\
&R_2 \ge I(X; W|VY)\big\} \quad \text{(outer bound)}
\end{aligned}
$$

where the random variables U, V and W whose alphabets are $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, respectively, are selected such that

- *the alphabet sizes are bounded as*

$$
\begin{aligned}
|\mathcal{U}| &\le |\mathcal{X}| + 2,\\
|\mathcal{V}| &\le |\mathcal{U} \times \mathcal{Y}| + 1,\\
|\mathcal{W}| &\le |\mathcal{U} \times \mathcal{V} \times \mathcal{X}| + 1,
\end{aligned}
$$

- *the following Markov chains are satisfied:*

$$
\begin{aligned}
&U \to X \to Y,\\
&V \to UY \to X,\\
&W \to VX \to UY,
\end{aligned}
$$

- *there exist functions* $\phi_{(1)} : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ *and* $\phi_{(2)} : \mathcal{W} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ *which safisfy*

$$\begin{aligned} D_1 &\geq E\left[\Delta(X, \phi_{(1)}(U, Y))\right], \\ D_2 &\geq E\left[\Delta(X, \phi_{(2)}(W, Y))\right]. \end{aligned}$$

An inner bound is obtained in the same functional forms, while the Markov chains are replaced as

$$\begin{aligned} &U \to X \to Y, \\ &V \to Y \to UX, \\ &W \to VX \to UY. \end{aligned}$$

Remark . *The Markov condition* $V \to Y \to UX$ *is tight compared with the condition* $V \to UY \to X$ *because* $I(V; X|UY) \leq I(V; UX|Y)$.

Remark . *The Markov conditions* $U \to X \to Y$, $V \to UY \to X$ *and* $W \to VX \to UY$ *are equivalent to the condition that the joint distribution* P_{UVWXY} *satisfies*

$$\begin{aligned} &P_{UVWXY}(u, v, w, x, y) \\ &\quad = P_{XY}(x, y) P_{U|X}(u|x) P_{V|UY}(v|u, y) P_{W|VX}(w|v, x). \end{aligned}$$

In a similar manner, the Markov conditions $U \to X \to Y$, $V \to Y \to UX$ *and* $W \to VX \to UY$ *are equivalent to the condition that the joint distribution* P_{UVWXY} *satisfies*

$$\begin{aligned} &P_{UVWXY}(u, v, w, x, y) \\ &\quad = P_{XY}(x, y) P_{U|X}(u|x) P_{V|Y}(v|y) P_{W|VX}(w|v, x). \end{aligned}$$

Theorem 2 indicates that feedback information contributes to refining reproduction messages, in contrast to the result for a lossless configuration reported by Yang et al. [3].

From Theorem 2, we can obtain the following properties.

Corollary 3. (Coding theorem for some special cases)
If side information Y *is not available at both of two decoders, the outer bound indicated in Theorem 2 coincides with the inner bound, i.e.*

$$\begin{aligned} \mathcal{R}_f(X|D_1, D_2) = \big\{ &(R_0, R_1, R_2) : \\ &R_1 \geq I(X; \widehat{X}^{(1)}), \\ &R_2 \geq I(X; \widehat{X}^{(2)}) \big\}. \end{aligned}$$

This shows that feedback information is of no use for refining in the absence of side information.

The above discussions can be easily extended to communication systems with multi-stage refinement. Let us consider N pairs $\{\varphi_{(i)}, \widehat{\varphi}_{(i)}\}_{i\in\mathcal{I}_N}$ of encoders and decoders, and $N-1$ feedback encoders $\{\varphi_{(0j)}\}_{j\in\mathcal{I}_{N-1}}$. Let $\boldsymbol{D} = \{D_i\}_{i\in\mathcal{I}_N}$ be a set of distortion criteria, each of which corresponds to the decoder $\widehat{\varphi}_{(i)}$. We define the FR-achievable rate region $\mathcal{R}_f(X,Y|\boldsymbol{D})$ of the source (X,Y) for given distortion criteria $\boldsymbol{D}$ in the same way as for the two-stage refinement system, where R_i $(i\in\mathcal{I}_N)$ corresponds to the rate of the encoder $\varphi_{(i)}$, and $R_{(0j)}$ $(j\in\mathcal{I}_{N-1})$ corresponds to the rate of the feedback encoder $\varphi_{(0j)}$. Let $U^{(i)}$ $(i\in\mathcal{I}_N)$ and $V^{(j)}$ $(j\in\mathcal{I}_{N-1})$ be auxiliary random variables that take values in finite sets $\mathcal{U}^{(i)}$ and $\mathcal{V}^{(j)}$, respectively. For some $\mathcal{S}\subseteq\mathcal{I}_N$ and $\mathcal{T}\subseteq\mathcal{I}_{N-1}$, let us define

$$
\begin{aligned}
\boldsymbol{U}^{(\mathcal{S})} &= \{U^{(i)} : i\in\mathcal{S}\},\\
\boldsymbol{U} &= \boldsymbol{U}^{(\mathcal{I}_N)},\\
\boldsymbol{V}^{(\mathcal{T})} &= \{V^{(j)} : j\in\mathcal{T}\},\\
\boldsymbol{V} &= \boldsymbol{V}^{(\mathcal{I}_{N-1})},\\
\mathcal{U}^{(\mathcal{I}_N)} &= \prod_{i\in\mathcal{I}_N}\mathcal{U}^{(i)},\\
\mathcal{V}^{(\mathcal{I}_{N-1})} &= \prod_{j\in\mathcal{I}_{N-1}}\mathcal{V}^{(j)}.
\end{aligned}
$$

Corollary 4. (Coding theorem of FR code with multi-stage feedback)

$$
\begin{aligned}
&\mathcal{R}_f(X,Y|\boldsymbol{D}) \subseteq \big\{(R_{0j}, R_i)_{i\in\mathcal{I}_N, j\in\mathcal{I}_{N-1}} :\\
&\quad R_{0j} \geq I(Y;V^{(i)}|U^{(i)}X),\\
&\quad R_i \geq I(X;U^{(i)}|V^{(i-1)}Y)\} \quad \text{(outer bound)}
\end{aligned}
$$

where the random variables $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *whose alphabets are* $\mathcal{U}^{(\mathcal{I}_N)}$ *and* $\mathcal{V}^{(\mathcal{I}_{N-1})}$*, respectively, are selected such that*

- *the alphabet sizes are bounded as*

$$
\begin{aligned}
|\mathcal{U}^{(i)}| &\leq |\mathcal{U}^{(\mathcal{I}_{i-1})}\times\mathcal{V}^{(\mathcal{I}_{i-1})}\times\mathcal{X}|+2, \quad (i\in\mathcal{I}_N)\\
|\mathcal{V}^{(i)}| &\leq |\mathcal{U}^{(\mathcal{I}_{i})}\times\mathcal{V}^{(\mathcal{I}_{i-1})}\times\mathcal{Y}|+1, \quad (i\in\mathcal{I}_{N-1})
\end{aligned}
$$

- *the following Markov chains are satisfied:*

$$
\begin{aligned}
&U^{(i)} \to V^{(i-1)}X \to \boldsymbol{U}^{(\mathcal{I}_{i-1})}\boldsymbol{V}^{(\mathcal{I}_{i-2})}Y, \quad (i\in\mathcal{I}_N)\\
&V^{(i)} \to U^{(i)}Y \to \boldsymbol{U}^{(\mathcal{I}_{i-1})}\boldsymbol{V}^{(\mathcal{I}_{i-1})}X, \quad (i\in\mathcal{I}_{N-1}),
\end{aligned}
$$

- *there exist functions* $\{\phi_{(i)} : \mathcal{U}^{(i)}\times\mathcal{Y}\to\widehat{\mathcal{X}}\}_{i\in\mathcal{I}_N}$ *which safisfy*

$$
D_i \geq E\left[\Delta(X,\phi_{(i)}(U^{(i)},Y))\right].
$$

An inner bound is obtained with the same functional forms, while the Markov chains are replaced as

$$U^{(i)} \to V^{(i-1)}X \to \boldsymbol{U}^{(\mathcal{I}_{i-1})}\boldsymbol{V}^{(\mathcal{I}_{i-2})}Y, \quad (i \in \mathcal{I}_N)$$
$$V^{(i)} \to Y \to \boldsymbol{U}^{(\mathcal{I}_i)}\boldsymbol{V}^{(\mathcal{I}_{i-1})}X, \quad (i \in \mathcal{I}_{N-1})$$

4 Refinement with cascading and feedback channels

Lastly, a *cascading and feedback refinement system* is considered, which denotes a refinement system with a cascading channel from the first decoder to the second decoder and a feedback channel from the first decoder to the second decoder.

4.1 Problem formulation

Definition 7. (CFR (Cascading and Feedback Refinement) code)
A set $(\varphi^n_{(01)}, \varphi^n_{(02)}, \varphi^n_{(1)}, \varphi^n_{(2)}, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})$ of encoders and decoders is a CFR code $(n, M_n^{(01)}, M_n^{(02)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})$ for the source (X, Y) if and only if

$$\begin{aligned}
\varphi^n_{(1)} &: \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}} \\
\varphi^n_{(01)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \mathcal{I}_{M_n^{(01)}} \\
\varphi^n_{(02)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \mathcal{I}_{M_n^{(02)}} \\
\varphi^n_{(2)} &: \mathcal{I}_{M_n^{(01)}} \times \mathcal{X}^n \to \mathcal{I}_{M_n^{(2)}} \\
\widehat{\varphi}^n_{(1)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\
\widehat{\varphi}^n_{(2)} &: \mathcal{I}_{M_n^{(02)}} \times \mathcal{I}_{M_n^{(2)}} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n,
\end{aligned}$$

where

$$\begin{aligned}
&\rho_n^{(i)} = E\left[\Delta^n(X^n, \widehat{X}^n_{(i)})\right], \quad (i = 1, 2) \\
&A_n^{(1)} = \varphi^n_{(1)}(X^n),\ A_n^{(01)} = \varphi^n_{(01)}(A_n^{(1)}, Y^n), \\
&A_n^{(02)} = \varphi^n_{(02)}(A_n^{(1)}, Y^n),\ A_n^{(2)} = \varphi^n_{(2)}(A_n^{(01)}, X^n), \\
&\widehat{X}^n_{(1)} = \widehat{\varphi}^n_{(1)}(A_n^{(1)}, Y^n),\ \widehat{X}^n_{(2)} = \widehat{\varphi}^n_{(2)}(A_n^{(01)}, A_n^{(2)}, Y^n).
\end{aligned}$$

Definition 8. (CFR-achievable rate quadruplet)
$(R_{01}, R_{02}, R_1, R_2)$ is a CFR-achievable rate quadruplet of the source (X, Y) for a given distortion pair (D_1, D_2) if and only if there exists a sequence of CFR codes $\{(n, M_n^{(01)}, M_n^{(02)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})\}_{n=1}^{\infty}$ for the source (X, Y) such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(i)} \le R_i, \quad (i = 01, 02, 1, 2)$$

$$\limsup_{n\to\infty} \rho_n^{(j)} \le D_j. \quad (j=1,2)$$

Definition 9. (CFR-achievable rate region)

$$\begin{aligned}\mathcal{R}_{cf}(X,Y|D_1,D_2) = \{&(R_{01},R_{02},R_1,R_2) : \\ &(R_{01},R_{02},R_1,R_2) \text{ is a CFR-achievable rate} \\ &\text{quadruplet of } (X,Y) \text{ for } (D_1,D_2)\}.\end{aligned}$$

4.2 Statement of results

First, we state the main theorem.

Theorem 3. (Coding theorem of CFR code)

$$\begin{aligned}\mathcal{R}_{cf}(X,Y|D_1,D_2) \subseteq \big\{&(R_{01},R_{02},R_1,R_2) : \\ &R_1 \ge I(X;UV^{(2)}|Y), \\ &R_{01} \ge I(Y;V^{(1)}|UV^{(2)}X), \\ &R_{02} \ge I(X;V^{(2)}|Y), \\ &R_2 \ge I(X;W|V^{(1)}V^{(2)}Y)\big\} \quad \text{(outer bound)}\end{aligned}$$

where the random variables U, $V^{(1)}$, $V^{(2)}$ and W whose alphabets are $\mathcal{U}$, $\mathcal{V}^{(1)}$, $\mathcal{V}^{(2)}$ and $\mathcal{W}$, respectively, are selected such that

- *the alphabet sizes are bounded as*

$$\begin{aligned}|\mathcal{U}| &\le |\mathcal{X}|+2, \\ |\mathcal{V}^{(2)}| &\le |\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|+4, \\ |\mathcal{V}^{(1)}| &\le |\mathcal{U}\times\mathcal{V}^{(2)}\times\mathcal{Y}|+1, \\ |\mathcal{W}| &\le |\mathcal{U}\times\mathcal{V}^{(1)}\times\mathcal{V}^{(2)}\times\mathcal{X}|+1,\end{aligned}$$

- *the following Markov chains are satisfied:*

$$\begin{aligned}U \to X &\to Y, \\ V^{(1)} \to UV^{(2)}Y &\to X, \\ W \to V^{(1)}V^{(2)}X &\to UY,\end{aligned}$$

- *there exist functions $\phi_{(1)} : \mathcal{U}\times\mathcal{Y} \to \widehat{\mathcal{X}}$ and $\phi_{(2)} : \mathcal{W}\times\mathcal{V}^{(2)}\times\mathcal{Y} \to \widehat{\mathcal{X}}$ which safisfy*

$$\begin{aligned}D_1 &\ge E\left[\Delta(X,\phi_{(1)}(U,Y))\right], \\ D_2 &\ge E\left[\Delta(X,\phi_{(2)}(W,V^{(2)},Y))\right].\end{aligned}$$

An inner bound is obtained in the same functional forms, while the Markov chains are replaced as

$$\begin{aligned} UV^{(2)} &\to X \to Y, \\ V^{(1)} &\to V^{(2)}Y \to UX, \\ W &\to V^{(1)}V^{(2)}X \to UY. \end{aligned}$$

Remark . *The Markov conditions*

$$\begin{aligned} U &\to X \to Y, \\ V^{(1)} &\to UV^{(2)}Y \to X, \\ W &\to V^{(1)}V^{(2)}X \to UY, \end{aligned}$$

are equivalent to the condition that the joint distribution $P_{UV^{(1)}V^{(2)}WXY}$ satisfies

$$\begin{aligned} & P_{UV^{(1)}V^{(2)}WXY}(u, v^{(1)}, v^{(2)}, w, x, y) \\ &= P_{XY}(x, y)P_{U|X}(u|x)P_{V^{(2)}|UXY}(v^{(2)}|u, x, y) \\ &\quad P_{V^{(1)}|UV^{(2)}Y}(v^{(1)}|u, v^{(2)}, y) \\ &\quad P_{W|V^{(1)}V^{(2)}X}(w|v^{(1)}, v^{(2)}, x). \end{aligned}$$

Also, the Markov conditions

$$\begin{aligned} U &\to X \to Y, \\ V^{(1)} &\to V^{(2)}Y \to UX, \\ W &\to V^{(1)}V^{(2)}X \to UY \end{aligned}$$

are equivalent to the condition that the joint distribution $P_{UV^{(1)}V^{(2)}WXY}$ satisfies

$$\begin{aligned} & P_{UV^{(1)}V^{(2)}WXY}(u, v^{(1)}, v^{(2)}, w, x, y) \\ &= P_{XY}(x, y)P_{U|X}(u|x)P_{V^{(2)}|UXY}(v^{(2)}|u, x, y) \\ &\quad P_{V^{(1)}|V^{(2)}Y}(v^{(1)}|v^{(2)}, y) \\ &\quad P_{W|V^{(1)}V^{(2)}X}(w|v^{(1)}, v^{(2)}, x) \end{aligned}$$

5 Proofs of theorems

5.1 Theorem 1: converse

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

$$\frac{1}{n} \log M_n^{(i)} \quad \leq \quad R_i + \delta. \quad (i = 0, 1, 2) \tag{3}$$

Let us remind you that $A_n^{(1)} = \varphi_{(1)}^n(X^n)$, $A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)$ and $A_n^{(2)} = \varphi_{(2)}^n(X^n)$. First, we evaluate Eq. (3) for $i = 1$. We obtain

$$
\begin{aligned}
& n(R_1 + \delta) \\
& \quad \geq \log M_n^{(1)} && (4) \\
& \quad \geq H(A_n^{(1)}) && (5) \\
& \quad \geq H(A_n^{(1)}|Y^n) && (6) \\
& \quad = I(X^n; A_n^{(1)}|Y^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (7) \\
& \quad = I(X^n; A_n^{(1)} A_n^{(0)}|Y^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)) && (8) \\
& \quad = H(X^n|Y^n) - H(X^n|A_n^{(1)} A_n^{(0)} Y^n) && (9) \\
& \quad = \sum_{k=1}^{n} \Big\{ H(X_k|Y_k) - H(X_k|A_n^{(1)} A_n^{(0)} X^{k-1} Y^n) \Big\} && (10) \\
& \quad = \sum_{k=1}^{n} I(X_k; A_n^{(1)} A_n^{(0)} X^{k-1} Y^{k-1} Y_{k+1}^n | Y_k) && (11)
\end{aligned}
$$

Let us define the random variables $U_k = A_n^{(1)} X^{k-1} Y^{k-1} Y_{k+1}^n$ and $V_k = A_n^{(0)} X^{k-1} Y^{k-1} Y_{k+1}^n$. With these definitions, we have the Markov structure $U_k \to X_k \to Y_k$ because

$$
\begin{aligned}
& I(Y_k; U_k|X_k) \\
& \quad = I(Y_k; A_n^{(1)} X^{k-1} Y^{k-1} Y_{k+1}^n | X_k) && (12) \\
& \quad \leq I(Y_k; A_n^{(1)} X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n | X_k) && (13) \\
& \quad = I(Y_k; X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n | X_k) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (14) \\
& \quad \leq I(X_k Y_k; X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n) && (15) \\
& \quad = 0. && (16)
\end{aligned}
$$

Substituting U_k and V_k into (11), we obtain

$$
n(R_1 + \delta) \geq \sum_{k=1}^{n} I(X_k; U_k V_k | Y_k).
$$

Here, let J be a random variable, independent of X and Y, and uniformly distributed over the set $\mathcal{I}_n$. Define the random variables $U = (J, U_J)$ and $V = (J, V_J)$. The Markov condition $U \to X \to Y$ still holds, and we have

$$
\begin{aligned}
& R_1 + \delta \\
& \quad \geq \frac{1}{n} \sum_{k=1}^{n} I(X_k; U_k V_k | Y_k) && (17) \\
& \quad = \frac{1}{n} \sum_{k=1}^{n} \{ H(X_k|Y_k) - H(X_k Y_k | U_k V_k) + H(Y_k | U_k V_k) \} && (18)
\end{aligned}
$$

$$
\begin{aligned}
&= H(X|Y) + \frac{1}{n}\sum_{k=1}^{n}\bigl\{-H(X_JY_J|U_JV_J, J=k) + H(Y_J|U_JV_J, J=k)\bigr\} \\
&= H(X|Y) - H(XY|JU_JV_J) + H(Y|JU_JV_J) && (19) \\
&= H(X|Y) - H(XY|UV) + H(Y|UV) && (20) \\
&= I(X;UV|Y). && (21)
\end{aligned}
$$

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

$$
R \geq I(X;UV|Y).
$$

Next, we evaluate Eq. (3) for $i = 0$. We obtain

$$
\begin{aligned}
&n(R_0 + \delta) \\
&\geq \log M_n^{(0)} && (22) \\
&\geq H(A_n^{(0)}) && (23) \\
&\geq H(A_n^{(0)}|Y^n) && (24) \\
&= I(X^n; A_n^{(0)}|Y^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(\varphi_{(1)}^n(X^n), Y^n)) && (25) \\
&= H(X^n|Y^n) - H(X^n|A_n^{(0)}Y^n) && (26) \\
&= \sum_{k=1}^{n}\Bigl\{H(X_k|Y_k) - H(X_k|A_n^{(0)}X^{k-1}Y^n)\Bigr\} && (27) \\
&= \sum_{k=1}^{n} I(X_k; A_n^{(0)}X^{k-1}Y^{k-1}Y_{k+1}^n|Y_k) && (28) \\
&= \sum_{k=1}^{n} I(X_k; V_k|Y_k) && (29)
\end{aligned}
$$

In the same way as the above discussion, we have

$$
R_0 \geq I(X;V|Y).
$$

Lastly, we evaluate Eq. (3) for $i = 2$. We obtain

$$
\begin{aligned}
&n(R_2 + \delta) \\
&\geq \log M_n^{(2)} && (30) \\
&\geq H(A_n^{(2)}) && (31) \\
&\geq H(A_n^{(2)}|A_n^{(0)}Y^n) && (32) \\
&= I(X^n; A_n^{(2)}|A_n^{(0)}Y^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(X^n)) && (33) \\
&= \sum_{k=1}^{n} I(X_k; A_n^{(2)}|A_n^{(0)}X^{k-1}Y^n) && (34)
\end{aligned}
$$

Let us define the random variable $W_k = A_n^{(2)}X^{k-1}Y^{k-1}Y_{k+1}^n$. With these definitions, we have the Markov structure $W_k \to U_kV_kX_k \to Y_k$ because

$$
I(Y_k; W_k|U_kV_kX_k)
$$

$$
\begin{aligned}
&= I(Y_k; A_n^{(2)} | A_n^{(1)} A_n^{(0)} X^k Y^{k-1} Y_{k+1}^n) && (35)\\
&\leq I(Y_k; A_n^{(2)} X_{k+1}^n | A_n^{(1)} A_n^{(0)} X^k Y^{k-1} Y_{k+1}^n) && (36)\\
&= I(Y_k; X_{k+1}^n | A_n^{(1)} A_n^{(0)} X^k Y^{k-1} Y_{k+1}^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(X^n)) && (37)\\
&\leq I(A_n^{(0)} Y_k; X_{k+1}^n | A_n^{(1)} X^k Y^{k-1} Y_{k+1}^n) && (38)\\
&= I(Y_k; X_{k+1}^n | A_n^{(1)} X^k Y^{k-1} Y_{k+1}^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, X^n)) && (39)\\
&\leq I(Y_k; A_n^{(1)} X_{k+1}^n | X^k Y^{k-1} Y_{k+1}^n) && (40)\\
&= I(Y_k; X_{k+1}^n | X^k Y^{k-1} Y_{k+1}^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (41)\\
&= I(X^k Y^k; X_{k+1}^n Y_{k+1}^n) && (42)\\
&= 0. && (43)
\end{aligned}
$$

The bound (34) becomes

$$
R_2 \quad \geq \quad \sum_{k=1}^{n} I(X_k; W_k | V_k Y_k)
$$

In the same way as the above discussion, we have

$$
R_2 \quad \geq \quad I(X; W | V Y).
$$

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

$$
\begin{aligned}
D_1 + \gamma &\geq E\left[\Delta^n(X^n, \widehat{X}_{(1)}^n)\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k, \widehat{X}_{(1)k})\right],\\
D_2 + \gamma &\geq E\left[\Delta^n(X^n, \widehat{X}_{(2)}^n)\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k, \widehat{X}_{(2)k})\right],
\end{aligned}
$$

Now, we denote by $\widehat{\varphi}_{(i)k}$ $(i = 1, 2)$ the output of $\widehat{\varphi}_{(i)}^n$ at time k $(k \in \mathcal{I}_n)$, namely

$$
\begin{aligned}
\widehat{x}_{(1)k} &= \widehat{\varphi}_{(1)k}(\varphi_{(1)}^n(\boldsymbol{x}), \boldsymbol{y}),\\
\widehat{x}_{(2)k} &= \widehat{\varphi}_{(2)k}(\varphi_{(0)}^n(\varphi_{(1)}^n(\boldsymbol{x}), \boldsymbol{y}), \varphi_{(2)}^n(\boldsymbol{x}), \boldsymbol{y}).
\end{aligned}
$$

We note that $U_k Y_k$ contains $A_n^{(1)} Y^n$, and $V_k W_k Y_k$ contains $A_n^{(0)} A_n^{(2)} Y^n$. Therefore we choose the functions $\phi_{(1)}$ and $\phi_{(2)}$ as follows:

$$
\begin{aligned}
\phi_{(1)k}(U_k, Y_k) &\overset{\text{def.}}{=} \widehat{\varphi}_{(1)k}(A_n^{(1)}, Y^n) = \widehat{X}_{(1)k}, && (44)\\
\phi_{(2)k}(V_k, W_k, Y_k) &\overset{\text{def.}}{=} \widehat{\varphi}_{(2)k}(A_n^{(0)}, A_n^{(2)}, Y^n) = \widehat{X}_{(2)k}, && (45)\\
\phi_{(1)}(U, Y) &\overset{\text{def.}}{=} \phi_{(1)J}(U_J, Y), && (46)
\end{aligned}
$$

$$\phi_{(2)}(V,W,Y) \quad \stackrel{\text{def.}}{=} \quad \phi_{(2)J}(V_J,W_J,Y). \tag{47}$$

This implies that

$$\begin{aligned}
D_1+\gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{X}_{(1)k})\right] && (48)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{\varphi}_{(1)k}(A_n^{(1)},Y^n))\right] && (49)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\phi_{(1)k}(U_k,Y_k))\right] && (50)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X,\phi_{(1)J}(U_J,Y))|J=k\right] && (51)\\
&= E\left[\Delta(X,\phi_{(1)}(U,Y))\right], && (52)\\
D_2+\gamma &\geq E\left[\Delta(X,\phi_{(2)}(V,W,Y))\right]. && (53)
\end{aligned}$$

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

$$\begin{aligned}
D_1 &\geq E\left[\Delta(X,\phi_{(1)}(U,Y))\right], && (54)\\
D_2 &\geq E\left[\Delta(X,\phi_{(2)}(V,W,Y))\right]. && (55)
\end{aligned}$$

It remains to establish that the bounds on $|\mathcal{U}|$, $|\mathcal{V}|$ and $|\mathcal{W}|$ specified in Theorem 1 does not affect the region $\mathcal{R}_c(X,Y|D_1,D_2)$. Namely, the proposition that have to be proved follows:

Proposition 1. *Let us define a set $\mathcal{P}\subseteq\mathcal{M}(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{X}\times\mathcal{Y})$ of generic distributions satisfying the conditions defined in Theorem 1. Given a distribution $P_{UVWXY}\in\mathcal{P}$, there exist auxiliary random variables $\widetilde{U}$, $\widetilde{V}$ and $\widetilde{W}$ taking values in $\widetilde{\mathcal{U}}\subseteq\mathcal{U}$, $\widetilde{\mathcal{V}}\subseteq\mathcal{V}$ and $\widetilde{\mathcal{W}}\subseteq\mathcal{W}$, respectively, and a corresponding joint distribution $P_{\widetilde{U}\widetilde{V}\widetilde{W}XY}\in\mathcal{P}$.*

To do this, we introduce the support lemma [4, Lemma 3.3.4]. We first reduce the alphabet size of U. Define the following functional forms of a generic distribution $Q\in\mathcal{M}(\mathcal{X})$:

$$\begin{aligned}
q_x(Q) &= Q(x), \quad (x\in\mathcal{I}_{m_x-1}) && (56)\\
q_{m_x}(Q) &= H(X|Y)\\
&\quad - \sum_{(v,x,y)\in\mathcal{V}\times\mathcal{X}\times\mathcal{Y}} Q(x)P_{VY|X}(v,y|x)\log\frac{\sum_{x'\in\mathcal{X}}Q(x')P_{VY|X}(v,y|x')}{Q(x)P_{VY|X}(v,y|x)} && (57)\\
q_{m_x+1}(Q) &= \sum_{(v,x,y)\in\mathcal{V}\times\mathcal{X}\times\mathcal{Y}} Q(x)P_{VY|X}(v,y|x)\Delta(x,\phi_{(1)}(u,v,y)) && (58)
\end{aligned}$$

$$\text{for a given } u \in \mathcal{U}, \tag{59}$$
$$\tag{60}$$

where $m_x = |\mathcal{X}|$. From the support lemma, we can find a random variable $\widetilde{U}$ taking values in $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$ with a generic distribution $P_{\widetilde{U}} \in \mathcal{M}(\widetilde{\mathcal{U}})$ and distributions $Q_u \in \mathcal{M}(\mathcal{X})$ $(u \in \widetilde{\mathcal{U}})$ such that $|\widetilde{\mathcal{U}}| \leq m_x + 1$ and the following equations are simultaneously satisfied:

$$\sum_{u \in \widetilde{\mathcal{U}}} P_{\widetilde{U}}(u) q_x(Q_u) = P_X(x) \quad (x \in \mathcal{I}_{m_x - 1}), \tag{61}$$
$$\sum_{u \in \widetilde{\mathcal{U}}} P_{\widetilde{U}}(u) q_{m_x}(Q_u) = I(X; UV|Y), \tag{62}$$
$$\sum_{u \in \widetilde{\mathcal{U}}} P_{\widetilde{V}}(u) q_{m_x+1}(Q_u) = E[\Delta(X, \phi_{(1)}(U, V, Y))]. \tag{63}$$

Here, let us define a joint distribution $P_{\widetilde{U}VWXY}(u, v, w, x, y) \in \mathcal{M}(\widetilde{\mathcal{U}} \times \mathcal{V} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y})$ as

$$P_{\widetilde{U}VWXY}(u, v, w, x, y) \stackrel{\text{def.}}{=} P_{\widetilde{U}}(u) Q_v(x) P_{Y|X}(y|x) P_{VW|XY}(v, w|x, y). \tag{64}$$

From the definition of $P_{\widetilde{U}VWXY}(u, v, w, x, y)$, the distribution of (V, W, X, Y) has been preserved since

$$\begin{aligned}
&\sum_{u \in \widetilde{\mathcal{U}}} P_{\widetilde{U}VWXY}(u, v, w, x, y) \\
&= P_{Y|X}(y|x) P_{VW|XY}(v, w|x, y) \sum_{u \in \widetilde{\mathcal{U}}} P_{\widetilde{U}}(u) Q_u(x) \\
&= P_{Y|X}(y|x) P_{VW|XY}(v, w|x, y) P_X(x) \quad (\because \text{Eq.}(61)) \\
&= P_{VWXY}(v, w, x, y).
\end{aligned}$$

Also, we obtain

$$P_{Y|\widetilde{U}X}(y|u, x)$$
$$\stackrel{\text{def.}}{=} \frac{\sum_{(v,w) \in \mathcal{V} \times \mathcal{W}} P_{\widetilde{U}VWXY}(u, v, w, x, y)}{\sum_{(v,w,y) \in \mathcal{V} \times \mathcal{W} \times \mathcal{Y}} P_{\widetilde{U}VWXY}(u, v, w, x, y)} \tag{65}$$
$$= \frac{P_{\widetilde{U}}(u) Q_v(x) P_{Y|X}(y|x)}{P_{\widetilde{U}}(u) Q_v(x)} \tag{66}$$
$$= P_{Y|X}(y|x), \tag{67}$$

which indicates that the Markov chain $\widetilde{U} \to X \to Y$ still remains.

Having found such a random variable U, we can reduce the alphabet size of W. In this time, we have $|\mathcal{U} \times \mathcal{X}| - 1$ constraints to preserve the distribution

$$P_{\widetilde{U}X} \stackrel{\text{def.}}{=} \sum_{(v,w,y) \in \mathcal{V} \times \mathcal{W} \times \mathcal{Y}} P_{\widetilde{U}VWXY}(u, v, w, x, y) \tag{68}$$

just defined, and two more constraints for preserving $I(X; W|VY)$ and $E[\Delta(X, \phi_{(2)}(V, W, Y))]$. Therefore, we can find a random variable $\widetilde{W}$ taking a value in $\widetilde{\mathcal{W}} \subseteq \mathcal{W}$ with a generic distribution $P_{\widetilde{W}} \in \mathcal{M}(\widetilde{\mathcal{W}})$ and distributions $Q_w \in \mathcal{M}(\widetilde{\mathcal{U} \times \mathcal{X}})$ $(w \in \widetilde{\mathcal{W}})$ such that $|\widetilde{\mathcal{W}}| \leq |\widetilde{\mathcal{U}} \times \mathcal{X}| + 1$. A joint distribution $P_{\widetilde{U}V\widetilde{W}XY}(u, v, w, x, y)$ is defined as the following equation.

$$P_{\widetilde{U}V\widetilde{W}XY}(u, v, w, x, y) = P_{\widetilde{W}}(w) Q_w(u, x) P_{V|\widetilde{U}X}(v|u, x) P_{Y|\widetilde{U}VX}(y|u, v, x) \tag{69}$$

From the definition of $P_{\widetilde{U}V\widetilde{W}XY}(u, v, w, x, y)$, the distribution of $(\widetilde{U}, V, X, Y)$ has been preserved since

$$\begin{aligned}
&\sum_{w \in \widetilde{\mathcal{W}}} P_{\widetilde{U}V\widetilde{W}XY}(u, v, w, x, y) && (70)\\
&= P_{V|\widetilde{U}X}(v|u, x) P_{Y|\widetilde{U}VX}(y|u, v, x) \sum_{w \in \widetilde{\mathcal{W}}} P_{\widetilde{W}}(w) Q_w(u, x) && (71)\\
&= P_{V|\widetilde{U}X}(v|u, x) P_{Y|\widetilde{U}VX}(y|u, v, x) P_{\widetilde{U}X}(u, x) && (72)\\
&= P_{\widetilde{U}VXY}(u, v, x, y). && (73)
\end{aligned}$$

This implies that the Markov chain $\widetilde{U} \to X \to Y$ still remains. Also, we can show that the Markov chain $\widetilde{W} \to \widetilde{U}VX \to Y$ still remains.

In a similar manner, we can reduce the alphabet size of V. In this time, we have $|\mathcal{U} \times \mathcal{W} \times \mathcal{X}| \times \mathcal{Y} - 1$ constraints to preserve the distribution P_{UWXY} just defined, and 4 more constraints for preserving $I(X; UV|Y)$, $I(X; V|Y)$, $I(X; W|VY)$, and $E[\Delta(X, \phi_{(2)}(V, W, Y))]$.

This completes the proof of the converse part. □

5.2 Theorem 1: direct part

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

Definition 10. (Set of typical sequences)
For $\boldsymbol{y} \in \mathcal{Y}^n$ and $\delta > 0$, define the set of typical sequences as

$$T_X^n(\delta) = \left\{ \boldsymbol{x} \in \mathcal{X}^n : \left| \frac{1}{n} N(x|\boldsymbol{x}) - P_X(x) \right| \leq \delta \ \forall x \in \mathcal{X} \right\}. \tag{74}$$

$$T_{X|\boldsymbol{y}}^n(\delta) = \Big\{ \boldsymbol{x} \in \mathcal{X}^n :$$

$$\left|\frac{1}{n}N(x,y|\boldsymbol{x},\boldsymbol{y})-\frac{1}{n}N(y|\boldsymbol{y})P_{X|Y}(x|y)\right|\leq\delta\ \forall(x,y)\in\mathcal{X}\times\mathcal{Y}\Bigg\}. \quad (75)$$

A similar convention is used for other random variables. When the dimension is clear from the content, the superscript n will be omitted, e.g. $T_X(\delta)$.

Lemma 1. (Csiszar-Körner [4])

$$\begin{aligned}\Pr\{X^n\in T_X(\delta)\} &\geq 1-\epsilon_n^{(1)}(\delta), &(76)\\ \Pr\{X^n\in T_{X|\boldsymbol{y}}(\delta)\} &\geq 1-\epsilon_n^{(2)}(\delta,\delta')\quad \forall\boldsymbol{y}\in T_Y(\delta'), &(77)\end{aligned}$$

where $\epsilon_n^{(1)}(\delta),\epsilon_n^{(2)}(\delta,\delta')\to 0\ (n\to\infty)$.

Lemma 2. (Csiszar-Körner[4, Lemma 1.2.10])
For any $\delta,\delta'>0$,

$$\begin{aligned}\boldsymbol{x}\in T_X(\delta_1)\ \text{and}\ \boldsymbol{y}\in T_{Y|\boldsymbol{x}}(\delta_2) &\Rightarrow (\boldsymbol{x},\boldsymbol{y})\in T_{XY}(\delta_1+\delta_2), &(78)\\ (\boldsymbol{x},\boldsymbol{y})\in T_{XY}(\delta_1) &\Rightarrow \boldsymbol{x}\in T_X(\delta_1|\mathcal{Y}|), &(79)\\ (\boldsymbol{x},\boldsymbol{y})\in T_{XY}(\delta_1)\ \text{and}\ \boldsymbol{x}\in T_X(\delta_2) &\Rightarrow \boldsymbol{y}\in T_{Y|\boldsymbol{x}}(\delta_1+\delta_2). &(80)\end{aligned}$$

Remark . *Although Eq.(80) of Lemma 2 is not shown in [4], we can easily obtain the property in the same way as the others.*

Lemma 3. (Csiszar-Körner [4, Lemma 1.2.13])

$$\begin{aligned}\left|\frac{1}{n}\log|T_X(\delta)|-H(X)\right| &\leq \epsilon_1, &(81)\\ \left|\frac{1}{n}\log|T_{X|\boldsymbol{y}}(\delta')|-H(X|Y)\right| &\leq \epsilon_2\quad \boldsymbol{y}\in T_Y(\delta'), &(82)\end{aligned}$$

where $\epsilon_1=\epsilon_1(\delta)$, $\epsilon_2=\epsilon_2(\delta,\delta')$ *and* $\epsilon_1,\epsilon_2\to 0\ (\delta,\delta'\to 0)$.

Lemma 4.

$$\left|-\frac{1}{k}\log P_X(\boldsymbol{x})-H(X)\right| \leq \epsilon\quad \forall\boldsymbol{x}\in T_X(\delta), \quad (83)$$

where $\epsilon=\epsilon(\delta)$ *and* $\epsilon\to 0\ (\delta\to 0)$.

Lemma 5. (Steinberg-Merhav[5])

$$\exp\{-n(I(X;U)+\epsilon_1)\}\leq\sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta')}P_U(\boldsymbol{u})\leq\exp\{-n(I(X;U)-\epsilon_2)\} \quad (84)$$

for any $\boldsymbol{x}\in T_X(\delta)$ *and* $\delta'>\delta$, *where* $\epsilon_1=\epsilon_1(\delta,\delta')$, $\epsilon_2=\epsilon_2(\delta,\delta')$, *and* $\epsilon_1,\epsilon_2\to 0$ $(\delta,\delta'\to 0)$

Remark . *Steinberg and Merhav [5] utilized Lemma 5 as a result of Csiszar and Körner [4] without any proofs. However, the lemma has not been shown in [4].*

Proof. First, we show the right inequality.

$$
\begin{aligned}
&\sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta')} P_U(\boldsymbol{u}) \\
&= \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta'),\boldsymbol{u}\in T_U(\delta'|\mathcal{X}|)} P_U(\boldsymbol{u}) \quad (\because \text{Lemma 2 Eq.(79)}) && (85)\\
&\leq \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta'),\boldsymbol{u}\in T_U(\delta'|\mathcal{X}|)} \exp\{-n(H(U)-\epsilon_1')\} \quad (\because \text{Lemma 4}) && (86)\\
&= \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta')} \exp\{-n(H(U)-\epsilon_1')\} \quad (\because \text{Lemma 2 Eq.(79)}) && (87)\\
&\leq \sum_{\boldsymbol{u}\in T_{U|\boldsymbol{x}}(\delta'+\delta)} \exp\{-n(H(U)-\epsilon_1')\} \quad (\because \text{Lemma 2 Eq.(80)}) && (88)\\
&\leq \exp\{n(H(U|X)+\epsilon_2')\}\exp\{-n(H(U)-\epsilon_1')\} \quad (\because \text{Lemma 3}) && (89)\\
&= \exp\{-n(I(X;U)-(\epsilon_1'+\epsilon_2'))\} && (90)\\
&= \exp\{-n(I(X;U)-\epsilon_2)\} && (91)
\end{aligned}
$$

Next, we show the left inequality. Noting that from Lemma 2 for $\delta' > \delta > 0$

$$
\boldsymbol{u} \in T_{U|\boldsymbol{x}}(\delta'-\delta) \text{ and } \boldsymbol{x} \in T_X(\delta) \quad \Rightarrow \quad (\boldsymbol{u},\boldsymbol{x}) \in T_{UX}(\delta'), \qquad (92)
$$

we obtain

$$
\begin{aligned}
&\sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta')} P_U(\boldsymbol{u}) \\
&= \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta'),\boldsymbol{u}\in T_U(\delta'|\mathcal{X}|)} P_U(\boldsymbol{u}) \quad (\because \text{Lemma 2 Eq.(79)}) && (93)\\
&\geq \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta'),\boldsymbol{u}\in T_U(\delta'|\mathcal{X}|)} \exp\{-n(H(U)+\epsilon_1')\} \quad (\because \text{Lemma 4}) && (94)\\
&= \sum_{\boldsymbol{u}:(\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta')} \exp\{-n(H(U)+\epsilon_1')\} \quad (\because \text{Lemma 2 Eq.(79)}) && (95)\\
&\geq \sum_{\boldsymbol{u}\in T_{U|\boldsymbol{x}}(\delta'-\delta)} \exp\{-n(H(U)+\epsilon_1')\} \quad (\because \text{Eq.(92)}) && (96)\\
&\geq \exp\{n(H(U|X)-\epsilon_3')\}\exp\{-n(H(U)+\epsilon_1')\} \quad (\because \text{Lemma 3}) && (97)\\
&= \exp\{-n(I(X;U)+(\epsilon_1'+\epsilon_3'))\} && (98)\\
&= \exp\{-n(I(X;U)+\epsilon_1)\} && (99)
\end{aligned}
$$

□

Lemma 6. (Steinberg-Merhav[5])

$$
\exp\{-n(I(X;V|U)+\epsilon_1)\}
$$

$$\leq \sum_{\boldsymbol{v}:(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x})\in T_{UVX}(\delta')} P_{V|U}(\boldsymbol{v}|\boldsymbol{u}) \leq \exp\{-n(I(X;V|U)-\epsilon_2)\} \quad (100)$$

for any $(\boldsymbol{u},\boldsymbol{x}) \in T_{UX}(\delta)$ *and* $\delta' > \delta$*, where* $\epsilon_1 = \epsilon_1(\delta,\delta')$*,* $\epsilon_2 = \epsilon_2(\delta,\delta')$*, and* $\epsilon_1, \epsilon_2 \to 0 \quad (\delta, \delta' \to 0)$

Remark . *Steinberg and Merhav [5] also utilized Lemma 6 without any proofs. We can prove the lemma in a similar manner to Lemma 5.*

Lemma 7. (Markov lemma [6, Lemma 14.8.1])
Let (X,Y,Z) *form a Markov chain* $X \to Y \to Z$*. If* X^n *is emitted from* $\prod_{k=1}^{n} P_{X|Y}$ *given* $(\boldsymbol{y},\boldsymbol{z}) \in T_{YZ}(\delta)$*, then*

$$\Pr\{(X^n,\boldsymbol{y},\boldsymbol{z}) \in T_{XYZ}(\delta)\} \quad > \quad 1-\delta \quad (101)$$

for a sufficiently large n *and* $\delta > 0$*.*

Here, we proceed the proof of the achievability part of Theorem 1.

Proof.
Let a distortion pair (D_1, D_2) be given, and let U, V, W and P_{UVWXY} satisfy the conditions that define $\mathcal{R}_c(X,Y|D_1,D_2)$. Fix arbitrary $\gamma, \delta > 0$.

Codeword selection: $\varphi^n_{(1)}$

(1) Randomly generate M_V independent codewords $\mathcal{A}_V = \{\boldsymbol{v}_i\}_{i=1}^{M_V}$, $\boldsymbol{v}_i \in \mathcal{V}^n$, each of length n, according to the distribution P_V.
(2) For each $\boldsymbol{v} \in \mathcal{A}_V$, randomly generate M_U independent codewords $\mathcal{A}_U(\boldsymbol{v}) = \{\boldsymbol{u}_i(\boldsymbol{v})\}_{i=1}^{M_U}$, $\boldsymbol{u}_i(\boldsymbol{v}) \in \mathcal{U}^n$, each of length n, according to the distribution $P_{U|V}$.
(3) Divide the codebook $\mathcal{A}_V$ into N_V bins, each containing $L_V = M_V/N_V$ members of $\mathcal{A}_V$. Let $\mathcal{A}_V(j)$ denote the set of elements $\boldsymbol{v} \in \mathcal{A}_V$ assigned to bin j $(j \in \mathcal{I}_{N_V})$.
(4) Divide the codebook $\mathcal{A}_U(\boldsymbol{v})$ into N_U bins, each containing $L_U = M_U/N_U$ members of $\mathcal{A}_U(\boldsymbol{v})$. Let $\mathcal{A}_U(\boldsymbol{v}, j)$ denote the set of elements $\boldsymbol{u} \in \mathcal{A}_U(\boldsymbol{v})$ assigned to bin j $(j \in \mathcal{I}_{N_U})$.

Codeword selection: $\varphi^n_{(0)}$

Unnecessary ($\mathcal{A}_V$ will be used as the codeword set).

Codeword selection: $\varphi^n_{(2)}$

(1) For each $\boldsymbol{v} \in \mathcal{A}_V$, randomly generate M_W independent codewords $\mathcal{A}_W(\boldsymbol{v}) = \{\boldsymbol{w}_i(\boldsymbol{v})\}_{i=1}^{M_W}$, $\boldsymbol{w}_i(\boldsymbol{v}) \in \mathcal{W}^n$, each of length n, according to the distribution $P_{W|V}$.
(2) Divide the codebook $\mathcal{A}_W(\boldsymbol{v})$ into N_W bins, each containing $L_W = M_W/N_W$ members of $\mathcal{A}_W(\boldsymbol{v})$. Let $\mathcal{A}_W(\boldsymbol{v}, j)$ denote the set of elements $\boldsymbol{w} \in \mathcal{A}_W(\boldsymbol{v})$ assigned to bin j $(j \in \mathcal{I}_{N_W})$.

Encoding: $\varphi^n_{(1)}$

(1) For a given $\boldsymbol{x} \in \mathcal{X}^n$, the encoder first seeks a vector $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{x}) \in T_{VX}(k_1\delta)$, $k_1 > 0$. If there is more than one such vector in $\mathcal{A}_V$, the

first one is chosen. If there is no such vector in $\mathcal{A}_V$, a default vector is chosen, say $\boldsymbol{v}_1$, and an error is declared. Denote the selected vector by $\boldsymbol{v}(\boldsymbol{x})$.
(2) The encoder next seeks a vector $\boldsymbol{u}_i \in \mathcal{A}_U(\boldsymbol{v})$ such that $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x})$ and $(\boldsymbol{u}_i, \boldsymbol{v}, \boldsymbol{x}) \in T_{UVX}(k_2\delta)$, $k_2 > 0$. If there is more than one such vector in $\mathcal{A}_U(\boldsymbol{v})$, the first one is chosen. If there is no such vector in $\mathcal{A}_U(\boldsymbol{v})$, a default vector is chosen, and an error is declared. Denote the selected vector by $\boldsymbol{u}(\boldsymbol{v}, \boldsymbol{x})$.
(3) The value assigned to the encoder $\varphi^n_{(1)}(\cdot)$ is the pair of the bin numbers to which $\boldsymbol{v}(\boldsymbol{x})$ and $\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x})$ belong, that is,

$$\varphi^n_{(1)}(\boldsymbol{x}) = j_v M_U + j_u, \tag{102}$$

$$\boldsymbol{v}(\boldsymbol{x}) \in \mathcal{A}_V(j_v), \quad \boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in \mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}), j_u). \tag{103}$$

Encoding: $\varphi^n_{(0)}$

The value assigned to the encoder $\varphi^n_{(0)}(\cdot)$ is the index j_v received from $\varphi^n_{(1)}$.

Encoding: $\varphi^n_{(2)}$

(1) For a given $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x})$, the encoder seeks a vector $\boldsymbol{w}_i \in \mathcal{A}_W(\boldsymbol{v})$ such that $(\boldsymbol{v}, \boldsymbol{w}_i, \boldsymbol{x}) \in T_{VWX}(k_3\delta)$, $k_3 > 0$. If there is more than one such vector in $\mathcal{A}_W(\boldsymbol{v})$, the first one is chosen. If there is no such vector in $\mathcal{A}_W(\boldsymbol{v})$, a default vector is chosen, and an error is declared. Denote the selected vector by $\boldsymbol{w}(\boldsymbol{v}, \boldsymbol{x})$.
(2) The value assigned to the encoder $\varphi^n_{(2)}(\cdot)$ is the bin number to which $\boldsymbol{w}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x})$ belong, that is,

$$\varphi^n_{(2)}(\boldsymbol{x}) = j_w, \quad \boldsymbol{w}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in \mathcal{A}_W(\boldsymbol{v}(\boldsymbol{x}), j_w). \tag{104}$$

Decoding: $\widehat{\varphi}^n_{(1)}$

(1) The decoder has access to the index j_u, j_v and the vector $\boldsymbol{y} \in \mathcal{Y}^n$.
(2) It first seeks a unique vector $\boldsymbol{v} \in \mathcal{A}_V(j_v)$ such that $(\boldsymbol{v}, \boldsymbol{y}) \in T_{VY}(k_4\delta)$, $k_4 > 0$. Denote this vector $\widehat{\boldsymbol{v}}(\boldsymbol{y})$. If there is no or more than one such vectors $\boldsymbol{v} \in \mathcal{A}_V(j_v)$, an arbitrary $\widehat{\boldsymbol{v}}$ is chosen, and an error is declared.
(3) It next seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(\widehat{\boldsymbol{v}}(\boldsymbol{y}), j_u)$ such that $(\boldsymbol{u}, \widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}) \in T_{UVY}(k_5\delta)$, $k_5 > 0$. Denote this vector $\widehat{\boldsymbol{u}}(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y})$. If there is no or more than one such vectors $\boldsymbol{u} \in \mathcal{A}_U(\widehat{\boldsymbol{v}}(\boldsymbol{y}), j_u)$, an arbitrary $\widehat{\boldsymbol{u}}$ is chosen, and an error is declared.
(4) The reconstruction vector $\widehat{\boldsymbol{x}}_{(1)} = (\widehat{x}_{(1)1}, \widehat{x}_{(1)2}, \cdots, \widehat{x}_{(1)n})$ is given by

$$\widehat{x}_{(1)k} = \phi_{(1)}(\widehat{u}_k(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}), y_k) \quad (k \in \mathcal{I}_n). \tag{105}$$

Decoding: $\widehat{\varphi}^n_{(2)}$

(1) The decoder has access to the index j_v, j_w and the vector $\boldsymbol{y} \in \mathcal{Y}^n$.
(2) It first seeks a unique vector $\boldsymbol{v} \in \mathcal{A}_V(j_v)$ in the same way as $\widehat{\varphi}^n_{(1)}$
(3) It then seeks a unique vector $\boldsymbol{w} \in \mathcal{A}_W(\widehat{\boldsymbol{v}}(\boldsymbol{y}), j_w)$ such that $(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{w}, \boldsymbol{y}) \in T_{VWY}(k_6\delta)$, $k_6 > 0$. Denote this vector $\widehat{\boldsymbol{w}}(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y})$. If there is no or more than one such vectors $\boldsymbol{w} \in \mathcal{A}_W(\widehat{\boldsymbol{v}}(\boldsymbol{y}), j_w)$, an arbitrary $\widehat{\boldsymbol{w}}$ is chosen, and an error is

declared.
(4) The reconstruction vector $\widehat{\boldsymbol{x}}_{(2)} = (\widehat{x}_{(2)1}, \widehat{x}_{(2)2}, \cdots, \widehat{x}_{(2)n})$ is given by

$$\widehat{x}_{(2)k} \quad = \quad \phi_{(2)}(\widehat{v}_k(\boldsymbol{y}), \widehat{w}_k(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}), y_k) \quad (k \in \mathcal{I}_n). \tag{106}$$

<u>*Distortion evaluation:* $\widehat{\varphi}^n_{(1)}$</u>
(1) For the distortion, we obtain

$$\begin{aligned}
&\Delta^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}_{(1)}) \\
&= \frac{1}{n}\sum_{k=1}^{n} \Delta(x_k, \widehat{x}_{(1)k}) && (107)\\
&= \frac{1}{n}\sum_{k=1}^{n} \Delta(x_k, \phi_{(1)}(\widehat{u}_k(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}), \widehat{v}_k(\boldsymbol{y}), y_k)) && (108)\\
&= \frac{1}{n}\sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} N(u,x,y|\widehat{\boldsymbol{u}}(\widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y})\Delta(x, \phi_{(1)}(u,y)) && (109)
\end{aligned}$$

If no error occurs in the encoding/decoding processes and $(\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \in T_{UXY}(k_7\delta)$, then the following inequalities satisfy:

$$\begin{aligned}
&\Delta^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}_{(1)}) \\
&\leq \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} (P_{UXY}(u,x,y) + k_7\delta)\Delta(x, \phi_{(1)}(u,y)) && (110)\\
&\qquad (\because \text{Definition 10}) \\
&\leq E\left[\Delta(x, \phi_{(1)}(u,y))\right] + k_7\delta\overline{\Delta}|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}| && (111)\\
&\leq D_1 + k_7\delta\overline{\Delta}|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|, && (112)
\end{aligned}$$

where

$$\overline{\Delta} \stackrel{\text{def.}}{=} \max_{(x,\hat{x})\in\mathcal{X}\times\widehat{\mathcal{X}}} \Delta(x, \widehat{x}) < \infty. \tag{113}$$

Let us define

$$E_1 \stackrel{\text{def.}}{=} \left\{(\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}), \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T_{UVXY}(k_7\delta)\right\}, \tag{114}$$

and let us denote error probabilities in the encoding/decoding processes as P_e^n. Then, the average distortion can be bounded as

$$\begin{aligned}
&E\left[\Delta^n(X^n, \widehat{X}^n_{(1)})\right] \\
&\leq (1 - P_e^n - \Pr\{E_1\})(D_1 + k_7\delta\overline{\Delta}|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|) + (P_e^n + \Pr\{E_1\})\overline{\Delta}, && (115)
\end{aligned}$$

Since $\delta > 0$ can be arbitrarily small for a sufficiently large n. Therefore, if P_e^n and $\Pr\{E_1\}$ vanish as $n \to \infty$, then we can obtain

$$\limsup_{n\to\infty} E\left[\Delta_X^n(X^n, \widehat{X}^n)\right] \quad \le \quad D_1. \tag{116}$$

Distortion evaluation: $\widehat{\varphi}_{(2)}^n$

In a similar manner as $\widehat{\varphi}_{(1)}^n$, we can obtain

$$\limsup_{n\to\infty} E\left[\Delta^n(X^n, \widehat{X}_{(2)}^n)\right] \quad \le \quad D_2. \tag{117}$$

Error evaluation: $\varphi_{(1)}^n$

(1) If there is no $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{x}) \in T_{VX}(k_1\delta)$, an encoding error occurs. This event is denoted as

$$E_2 \quad \stackrel{\text{def.}}{=} \quad \bigcap_{i=1}^{M_V} \{(\boldsymbol{v}_i, \boldsymbol{x}) \notin T_{VX}(k_1\delta)\}\,. \tag{118}$$

Here, let us define

$$E_0 \quad \stackrel{\text{def.}}{=} \quad \big\{(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}(k_0\delta)\big\}, \quad k_0 > 0. \tag{119}$$

From Lemma 1, $\Pr\{E_0^c\} \to 0$ as $n \to \infty$. Also, we note that $(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}(k_0\delta)$ $\Rightarrow \boldsymbol{x} \in T_X(k_0\delta|\mathcal{Y}|)$ from Lemma 3. Then, we have

$$\begin{aligned} \Pr\{E_2\} \quad &\le \quad \Pr\{E_0^c \cup E_2\} && (120)\\ &= \quad \Pr\{E_0^c\} + \Pr\{E_0 \cap E_2\} && (121) \end{aligned}$$

$$\begin{aligned} &\Pr\{E_0 \cap E_2\}\\ &\le \sum_{\boldsymbol{x}\in T_X(k_0\delta|\mathcal{Y}|)} P_X(\boldsymbol{x}) \Pr\left\{\bigcap_{i=1}^{M_V} \{(V_i^n, \boldsymbol{x}) \notin T_{VX}(k_1\delta)\} \middle| \boldsymbol{x}\right\} && (122)\\ &= \sum_{\boldsymbol{x}\in T_X(k_0\delta|\mathcal{Y}|)} P_X(\boldsymbol{x}) \Pr\left\{\bigcap_{i=1}^{M_V} \{(V_i^n, \boldsymbol{x}) \notin T_{VX}(k_1\delta)\}\right\} && (123)\\ &\qquad (\because \boldsymbol{v}_i \text{ is selected independently of } \boldsymbol{x})\\ &\le \sum_{\boldsymbol{x}\in T_X(k_0\delta|\mathcal{Y}|)} P_X(\boldsymbol{x}) \left[1 - \exp\{-n(I(X;V) + \epsilon_v)\}\right]^{M_V} && (124)\\ &\qquad (\because \text{Lemma } 5)\\ &\le \sum_{\boldsymbol{x}\in T_X(k_0\delta|\mathcal{Y}|)} P_X(\boldsymbol{x}) \exp\left[-M_V \exp\{-n(I(X;V) + \epsilon_v)\}\right] && (125)\\ &\qquad (\because (1-a)^n \le \exp(-an) \text{ for } 0 \le a \le 1)\\ &\le \exp\left[-M_V \exp\{-n(I(X;V) + \epsilon_v)\}\right]. && (126) \end{aligned}$$

By setting M_V, k_0 and k_1 as

$$M_V \geq \exp\{n(I(X;V)+m_1\gamma)\},\ m_1>0, \tag{127}$$

$m_1\gamma > \epsilon_v = \epsilon_v(k_0\delta|\mathcal{Y}|, k_1\delta)$ and $k_0|\mathcal{Y}| < k_1$, we have $\lim_{n\to\infty} \Pr\{E_2\} = 0$.
(2) If there is no $\boldsymbol{u}_i \in \mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}))$ such that $(\boldsymbol{u}_i, \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in T_{UVX}(k_2\delta)$, an encoding error occurs. This event is denoted as

$$E_3 \stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_U} \{(\boldsymbol{u}_i, \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \notin T_{UVX}(k_2\delta)\}. \tag{128}$$

Here, we have

$$\begin{aligned} \Pr\{E_3\} &\leq \Pr\{E_2 \cup E_3\} && (129)\\ &= \Pr\{E_2\} + \Pr\{E_2^c \cap E_3\}. && (130)\end{aligned}$$

Since $(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in T_{VX}(k_1\delta)$ if E_2 does not occur, then we have

$$\begin{aligned}
&\Pr\{E_2^c \cap E_3\}\\
&\leq \sum_{\substack{(\boldsymbol{v},\boldsymbol{x})\in T_{VX}(k_1\delta)\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x})}} P_{VX}(\boldsymbol{v},\boldsymbol{x}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{v}, \boldsymbol{x}) \notin T_{UVX}(k_2\delta)\} \middle| \boldsymbol{v}, \boldsymbol{x} \right\} && (131)\\
&= \sum_{\substack{(\boldsymbol{v},\boldsymbol{x})\in T_{VX}(k_1\delta)\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x})}} P_{VX}(\boldsymbol{v},\boldsymbol{x}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{v}, \boldsymbol{x}) \notin T_{UVX}(k_2\delta)\} \middle| \boldsymbol{v} \right\} && (132)\\
&\qquad (\because \boldsymbol{u}_i \text{ is selected independently of } \boldsymbol{x})\\
&\leq \sum_{\substack{(\boldsymbol{v},\boldsymbol{x})\in T_{VX}(k_1\delta)\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x})}} P_{VX}(\boldsymbol{v},\boldsymbol{x}) \left[1 - \exp\left\{-n(I(X;U|V)+\epsilon_{u|v})\right\}\right]^{M_U}\\
&\qquad (\because \text{Lemma 6}) && (133)\\
&\leq \sum_{\substack{(\boldsymbol{v},\boldsymbol{x})\in T_{VX}(k_1\delta)\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x})}} P_{VX}(\boldsymbol{v},\boldsymbol{x}) \exp\left[-M_U \exp\left\{-n(I(X;U|V)+\epsilon_{u|v})\right\}\right] && (134)\\
&\qquad (\because (1-a)^n \leq \exp(-an))\\
&\leq \exp\left[-M_U \exp\left\{-n(I(X;U|V)+\epsilon_{u|v})\right\}\right]. && (135)
\end{aligned}$$

By setting M_U, k_1 and k_2 as

$$M_U \geq \exp\{n(I(X;U|V)+m_2\gamma)\},\ m_2>0, \tag{136}$$

$m_2\gamma > \epsilon_{u|v} = \epsilon_{u|v}(k_1\delta, k_2\delta)$ and $k_1 < k_2$, we have $\lim_{n\to\infty} \Pr\{E_3\} = 0$.

Error evaluation: $\varphi_{(0)}^n$

Any encoding errors do not occur because the encoder does not generate any codewords.

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

(2) If there is no $\boldsymbol{w}_i \in \mathcal{A}_W(\boldsymbol{v}(\boldsymbol{x}))$ such that $(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{w}_i, \boldsymbol{x}) \in T_{VWX}(k_3\delta)$, an encoding error occurs. In almost the same way as $\varphi^n_{(1)}$, the probability such that this event occurs vanishes as $n \to \infty$ by setting M_W as

$$M_W \geq \exp\{n(I(X;W|V) + m_3\gamma)\},\ m_3 > 0. \tag{137}$$

Error evaluation: $\widehat{\varphi}^n_{(1)}$

(1) If there is no or more than one $\boldsymbol{v}_i \in \mathcal{A}_V(j_v)$ such that $\varphi^n_{(1)}(\boldsymbol{x}) = j_v M_U + j_u$ and $(\boldsymbol{v}_i, \boldsymbol{y}) \in T_{VY}(k_4\delta)$, an decoding error is declared. This event is classified into two cases.
(1-1) The first case: $(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{y}) \notin T_{VY}(k_4\delta)$. From Lemma 3, this event is included in

$$E_4 \stackrel{\text{def.}}{=} \{(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T_{VXY}(k_4\delta/|\mathcal{Y}|)\}. \tag{138}$$

Here, we have

$$\begin{aligned}\Pr\{E_4\} &\leq \Pr\{E_2 \cup E_4\} &(139)\\ &= \Pr\{E_2\} + \Pr\{E_2^c \cap E_4\}. &(140)\end{aligned}$$

We note that $(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in T_{VX}(k_1\delta)$ if E_2 does not occur and $V \to X \to Y$ forms a Markov chain. Therefore, from Lemma 7, we have

$$\begin{aligned}&\Pr\{E_2^c \cap E_4\}\\ &\leq \sum_{\substack{(\boldsymbol{v},\boldsymbol{x}) \in T_{VX}(k_1\delta)\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x})}} P_{VX}(\boldsymbol{v},\boldsymbol{x}) \Pr\{(\boldsymbol{v},\boldsymbol{x},Y^n) \notin T_{VXY}(k_4\delta/|\mathcal{Y}|)|\boldsymbol{v},\boldsymbol{x}\} &(141)\\ &\to 0 \quad (n \to \infty). &(142)\end{aligned}$$

(1-2) The second case: There exists $\boldsymbol{v}' \in \mathcal{A}_V(j_v)$ such that $\boldsymbol{v}' \neq \boldsymbol{v}(\boldsymbol{x})$ and $(\boldsymbol{v}', \boldsymbol{y}) \in T_{VY}(k_4\delta)$. This event is denoted as

$$E_5 \stackrel{\text{def.}}{=} \left\{ \bigcup_{\boldsymbol{v} \in \mathcal{A}_V(j_v), \boldsymbol{v} \neq \boldsymbol{v}(\boldsymbol{x})} \{(\boldsymbol{v}, \boldsymbol{y}) \in T_{VY}(k_4\delta)\} \right\}, \tag{143}$$

Let $i_v(j,k)$ be the index i of k-th $\boldsymbol{v}_i$, which belongs to $\mathcal{A}_V(j)$. Since $(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}(k_0\delta) \Rightarrow \boldsymbol{y} \in T_Y(k_0\delta|\mathcal{X}|)$ from Lemma 3, we have

$$\begin{aligned}\Pr\{E_5\} &\leq \Pr\{E_0^c \cup E_5\} &(144)\\ &= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_5\} &(145)\end{aligned}$$

$$\Pr\{E_0 \cap E_5\}$$

$$
\begin{aligned}
&\le \sum_{k=1}^{|\mathcal{A}_V(j_v)|} \sum_{\boldsymbol{y}\in T_Y(k_0\delta|\mathcal{X}|)} P_Y(\boldsymbol{y}) \Pr\left\{(V^n_{i_v(j_v,k)}, \boldsymbol{y}) \in T_{VY}(k_4\delta)\right\} && (146)\\
&\quad (\because \boldsymbol{v}_i \text{ is selected independently of } \boldsymbol{y})\\
&\le |\mathcal{A}_V(j_v)| \exp\{-n(I(Y;V)-\epsilon_2)\} && (147)\\
&\quad (\because \text{Lemma 5})\\
&= L_V \exp\{-n(I(Y;V)-\epsilon_2)\} && (148)
\end{aligned}
$$

By setting L_V, k_0 and k_4 as

$$L_V \le \exp\{n(I(Y;V) - l_1\gamma)\},\ l_1 > 0, \tag{149}$$

$l_1\gamma > \epsilon_2 = \epsilon_2(k_0\delta|\mathcal{X}|, k_4\delta)$ and $k_0|\mathcal{X}| < k_4$, we have $\Pr\{E_5\} \to 0$ $(n \to \infty)$.
(2) If there is no or more than one $\boldsymbol{u}_i \in \mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}), j_u)$ such that $\varphi^n_{(1)}(\boldsymbol{x}) = j_v M_U + j_u$ and $(\boldsymbol{u}_i, \widehat{\boldsymbol{v}}(\boldsymbol{y}), \boldsymbol{y}) \in T_{UVY}(k_5\delta)$, an decoding error is declared. This event is classified into two cases.
(2-1) The first case: $(\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}), \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{y}) \notin T_{UVY}(k_5\delta)$. From Lemma 3, this event is included in

$$E_6 \stackrel{\text{def.}}{=} \{(\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}), \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T_{UVXY}(k_5\delta/|\mathcal{Y}|)\}. \tag{150}$$

Here, we have

$$
\begin{aligned}
\Pr\{E_6\} &\le \Pr\{E_3 \cup E_6\} && (151)\\
&= \Pr\{E_3\} + \Pr\{E_3^c \cap E_6\}. && (152)
\end{aligned}
$$

We note that $(\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}), \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x}) \in T_{UVX}(k_2\delta)$ if E_3 does not occur and $UV \to X \to Y$ forms a Markov chain. Therefore, from Lemma 7 we have

$$
\begin{aligned}
&\Pr\{E_3^c \cap E_6\}\\
&\le \sum_{\substack{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x})\in T_{UVX}(k_2\delta),\\ \boldsymbol{v}=\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{u}=\boldsymbol{u}(\boldsymbol{v},\boldsymbol{x})}} P_{UVX}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}) \Pr\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x},Y^n) \notin T_{UVXY}(k_5\delta/|\mathcal{Y}|)|\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}\} && (153)\\
&\to 0 \quad (n \to \infty). && (154)
\end{aligned}
$$

Furthermore, $\lim_{n\to\infty} \Pr\{E_1\} = 0$ by setting k_7 and k_5 as $k_7 < k_5/|\mathcal{Y}|$.
(2-2) The second case: There exists $\boldsymbol{u}' \in \mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}), j_u)$, $\boldsymbol{u}' \ne \boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{x})$ such that $(\boldsymbol{u}', \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{y}) \in T_{UVY}(k_5\delta)$. This event is denoted as

$$E_7 \stackrel{\text{def.}}{=} \left\{ \bigcup_{\substack{\boldsymbol{u}\in\mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}),j_u),\\ \boldsymbol{u}\ne\boldsymbol{u}(\boldsymbol{v}(\boldsymbol{x}),\boldsymbol{x})}} \{(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{x}), \boldsymbol{y}) \in T_{UVY}(k_5\delta)\} \right\}, \tag{155}$$

Let $i_u(j,k)$ be the index i of k-th $\boldsymbol{u}_i$, which belongs to $\mathcal{A}_U(\boldsymbol{v}(\boldsymbol{x}), j)$. Here, we have

$$\Pr\{E_7\} \le \Pr\{E_4 \cup E_7\} \tag{156}$$

$$= \quad \Pr\{E_4\} + \Pr\{E_4^c \cap E_7\}. \tag{157}$$

Noting that $(\boldsymbol{v}(\boldsymbol{x}), \boldsymbol{y}) \in T_{VY}(k_4\delta)$ if E_4 does not occur, we have

$$\Pr\{E_4^c \cap E_7\}$$

$$\leq \quad \sum_{(\boldsymbol{v},\boldsymbol{y}) \in T_{VY}(k_4\delta)} P_{VY}(\boldsymbol{v},\boldsymbol{y}) \sum_{k=1}^{|\mathcal{A}_U(\boldsymbol{v},j_u)|} \Pr\left\{(U^n_{i_u(j_u,k)}, \boldsymbol{v}, \boldsymbol{y}) \in T_{UVY}(k_5\delta) | \boldsymbol{v}\right\}$$

$$(\because \boldsymbol{u}_i \text{ is selected independently of } \boldsymbol{y}) \tag{158}$$

$$\leq \quad \sum_{(\boldsymbol{v},\boldsymbol{y}) \in T_{VY}(k_4\delta)} P_{VY}(\boldsymbol{v},\boldsymbol{y}) |\mathcal{A}_U(\boldsymbol{v},j_u)| \exp\{-n(I(Y;U|V) - \epsilon_2)\} \tag{159}$$

$$(\because \text{Lemma 5})$$

$$\leq \quad L_U \exp\{-n(I(Y;U|V) - \epsilon_2)\} \tag{160}$$

By setting L_U, k_4 and k_5 as

$$L_U \leq \exp\{n(I(Y;U|V) - l_2\gamma)\}, \ l_2 > 0, \tag{161}$$

$l_2\gamma > \epsilon_2 = \epsilon_2(k_4\delta, k_5\delta)$ and $k_4 < k_5$, we have $\Pr\{E_7\} \to 0$ $(n \to \infty)$.

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

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

$$L_W \quad \leq \quad \exp\{n(I(Y;W|V) - l_3\gamma)\}, \quad l_3 > 0 \tag{162}$$

to vanish the decoding errors.

Rate evaluation: $\varphi^n_{(1)}$

The encoder sends the indexes of the bin using

$$\frac{1}{n} \log N_V N_U$$

$$= \quad \frac{1}{n} \log \frac{M_V}{L_V} \frac{M_U}{L_U} \tag{163}$$

$$\geq \quad I(X;V) + m_1\gamma - I(Y;V) + l_1\gamma + I(X;U|V) + m_2\gamma - I(Y;U|V) + l_2\gamma \tag{164}$$

$$= \quad I(XY;V) - I(Y;V) + I(XY;U|V) - I(Y;U|V) + (m_1 + m_2 + l_1 + l_2)\gamma \tag{165}$$

$$(\because UV \to X \to Y)$$

$$= \quad I(X;V|Y) + I(X;U|VY) + (m_1 + m_2 + l_1 + l_2)\gamma \tag{166}$$

$$= \quad I(X;UV|Y) + (m_1 + m_2 + l_1 + l_2)\gamma \tag{167}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n} \log N_V N_U \quad \geq \quad I(X;UV|Y). \tag{168}$$

Rate evaluation: $\varphi^n_{(0)}$

The encoder sends the indexes of the bin using

$$\begin{aligned}
\frac{1}{n}\log N_V &= \frac{1}{n}\log\frac{M_V}{L_V} && (169)\\
&\geq I(X;V)+m_1\gamma - I(Y;V)+l_1\gamma && (170)\\
&= I(XY;V)-I(Y;V)+(m_1+l_1)\gamma && (171)\\
&\quad (\because V\to X\to Y)\\
&= I(X;V|Y)+(m_1+l_1)\gamma && (172)
\end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n}\log N_V \geq I(X;V|Y). \quad (173)$$

Rate evaluation: $\varphi^n_{(2)}$

The encoder sends the indexes of the bin using

$$\begin{aligned}
\frac{1}{n}\log N_W &= \frac{1}{n}\log\frac{M_W}{L_W} && (174)\\
&\geq I(X;W|V)+m_3\gamma - I(Y;W|V)+l_3\gamma && (175)\\
&= I(XY;W|V)-I(Y;W|V)+(m_3+l_3)\gamma && (176)\\
&\quad (\because VW\to X\to Y)\\
&= I(X;W|VY)+(m_3+l_3)\gamma && (177)
\end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n}\log N_W \geq I(X;W|VY). \quad (178)$$

This completes the proof of the direct part of Theorem 1. □

5.3 Theorem 2: converse

We begin by introducing a lemma that will be used hereafter.

Lemma 8.
Let A, B, C, D be random variables such that at least one of $A\to D\to B$, $B\to D\to C$ or $C\to D\to A$ form Markov chains. Then,

$$I(A;B|D) \leq I(A;B|CD). \quad (179)$$

Remark . *This lemma is an extended version of Massay's lemma.*

Proof. (1) If $A \to D \to B$ forms a Markov chain,

$$I(A;B|D) \quad = \quad 0 \leq I(A;B|CD). \tag{180}$$

(2) If $B \to D \to C$ forms a Markov chain, $I(B;C|D) = 0$. Therefore,

$$\begin{aligned} I(A;B|CD) &= I(AC;B|D) - I(C;B|D) & (181)\\ &= I(AC;B|D) & (182)\\ &\geq I(A;B|D) & (183)\end{aligned}$$

(3) If $C \to D \to A$ forms a Markov chain, $I(C;A|D) = 0$. Therefore,

$$\begin{aligned} I(A;B|CD) &= I(A;BC|D) - I(A;C|D) & (184)\\ &= I(A;BC|D) & (185)\\ &\geq I(A;B|D) & (186)\end{aligned}$$

□

Here, we proceed the proof of the converse part of Theorem 2.

Proof.
Let a sequence $\{(\varphi^n_{(0)}, \varphi^n_{(1)}, \varphi^n_{(2)}, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ of FR codes be given to satisfy the conditions of Definitions 4 and 5. From Definition 5, for $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$, and then for all $n \geq n_1(\delta)$, we can obtain

$$\frac{1}{n}\log M_n^{(i)} \quad \leq \quad R_i + \delta. \quad (i = 0, 1, 2) \tag{187}$$

Please remember $A_n^{(1)} = \varphi^n_{(1)}(X^n)$, $A_n^{(0)} = \varphi^n_{(0)}(A_n^{(1)}, Y^n)$ and $A_n^{(2)} = \varphi^n_{(2)}(A_n^{(0)}, X^n)$. First, we evaluate Eq. (187) for $i = 1$. We obtain

$$\begin{aligned} & n(R_1 + \delta) \\ &\geq \log M_n^{(1)} & (188)\\ &\geq H(A_n^{(1)}) & (189)\\ &\geq H(A_n^{(1)}|Y^n) & (190)\\ &= I(X^n; A_n^{(1)}|Y^n) \quad (\because A_n^{(1)} = \varphi^n_{(1)}(X^n)) & (191)\\ &= H(X^n|Y^n) - H(X^n|A_n^{(1)}Y^n) & (192)\\ &= \sum_{k=1}^{n}\Big\{H(X_k|Y_k) - H(X_k|A_n^{(1)}X^{k-1}Y^n)\Big\} & (193)\\ &= \sum_{k=1}^{n} I(X_k; A_n^{(1)}X^{k-1}Y^{k-1}Y_{k+1}^n|Y_k) & (194)\\ &\geq \sum_{k=1}^{n} I(X_k; A_n^{(1)}X^{k-1}Y_{k+1}^n|Y_k) & (195)\end{aligned}$$

Let us define the random variables $U_k = A_n^{(1)} X^{k-1} Y_{k+1}^n$. With these definitions, we have the Markov structure $U_k \to X_k \to Y_k$ because

$$\begin{aligned}
& I(Y_k; U_k|X_k) \\
&\quad = I(Y_k; A_n^{(1)} X^{k-1} Y_{k+1}^n | X_k) && (196)\\
&\quad \leq I(Y_k; A_n^{(1)} X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n | X_k) && (197)\\
&\quad = I(Y_k; X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n | X_k) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (198)\\
&\quad \leq I(X_k Y_k; X^{k-1} X_{k+1}^n Y^{k-1} Y_{k+1}^n) && (199)\\
&\quad = 0. && (200)
\end{aligned}$$

Substituting U_k into (195), we obtain

$$n(R_1 + \delta) \geq \sum_{k=1}^n I(X_k; U_k|Y_k).$$

By introducing the random variable J in the same way as the proof shown in Section 5.1, the Markov condition $U \to X \to Y$ still holds, and we have

$$R_1 + \delta \geq I(X; U|Y). \tag{201}$$

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

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

Next, we evaluate Eq. (187) for $i = 0$. We obtain

$$\begin{aligned}
& n(R_0 + \delta) \\
&\quad \geq \log M_n^{(0)} && (202)\\
&\quad \geq H(A_n^{(0)}) && (203)\\
&\quad \geq H(A_n^{(0)} | A_n^{(1)} X^n) && (204)\\
&\quad = I(Y^n; A_n^{(0)} | A_n^{(1)} X^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)) && (205)\\
&\quad = H(Y^n | A_n^{(1)} X^n) - H(Y^n | A_n^{(1)} A_n^{(0)} X^n) && (206)\\
&\quad = H(Y^n | X^n) - H(Y^n | A_n^{(1)} A_n^{(0)} X^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (207)\\
&\quad = \sum_{k=1}^n \left\{ H(Y_k|K_k) - H(Y_k | A_n^{(1)} A_n^{(0)} Y_{k+1}^n X^n) \right\} && (208)\\
&\quad = \sum_{k=1}^n I(Y_k; A_n^{(1)} A_n^{(0)} X^{k-1} X_{k+1}^n Y_{k+1}^n | X_k) && (209)\\
&\quad \geq \sum_{k=1}^n I(Y_k; A_n^{(1)} A_n^{(0)} X^{k-1} Y_{k+1}^n | X_k) && (210)\\
&\quad \geq \sum_{k=1}^n I(Y_k; A_n^{(0)} | A_n^{(1)} X^k Y_{k+1}^n) && (211)
\end{aligned}$$

Let us define the random variables $V_k = A_n^{(1)} A_n^{(0)} X^{k-1} Y_{k+1}^n$. With these definitions, we have the Markov structure $V_k \to U_k Y_k \to X_k$ because

$$\begin{aligned}
& I(X_k; V_k | U_k Y_k) \\
&= I(X_k; A_n^{(0)} | A_n^{(1)} X^{k-1} Y_k^n) && (212) \\
&\leq I(X_k; A_n^{(0)} Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) && (213) \\
&= I(X_k; Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)) && (214) \\
&\leq I(A_n^{(1)} X_k; Y^{k-1} | X^{k-1} Y_k^n) && (215) \\
&\leq I(A_n^{(1)} X_k^n; Y^{k-1} | X^{k-1} Y_k^n) && (216) \\
&= I(X_k^n; Y^{k-1} | X^{k-1} Y_k^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (217) \\
&\leq I(X_k^n Y_k^n; X^{k-1} Y^{k-1}) && (218) \\
&= 0. && (219)
\end{aligned}$$

Substituting U_k and V_k into (211), we obtain

$$n(R_0 + \delta) \geq \sum_{k=1}^{n} I(Y_k; V_k | U_k X_k).$$

In the same way as the above discussion, we have the Markov structure $V \to UY \to X$ and

$$R_0 \geq I(Y; V | UX).$$

Lastly, we evaluate Eq. (211) for $i = 2$. We obtain

$$\begin{aligned}
& n(R_2 + \delta) \\
&\geq \log M_n^{(2)} && (220) \\
&\geq H(A_n^{(2)}) && (221) \\
&\geq H(A_n^{(2)} | A_n^{(1)} A_n^{(0)} Y^n) && (222) \\
&= I(X^n; A_n^{(2)} | A_n^{(1)} A_n^{(0)} Y^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(A_n^{(0)}, X^n)) && (223) \\
&= \sum_{k=1}^{n} I(X_k; A_n^{(2)} | A_n^{(1)} A_n^{(0)} X^{k-1} Y^n) && (224)
\end{aligned}$$

Here, we will apply Lemma 8. Let us set $A = X_k$, $B = A_n^{(2)}$, $C = Y^{k-1}$ and $D = A_n^{(1)} A_n^{(0)} X^{k-1} Y_k^n$. Then, we have

$$\begin{aligned}
& I(A; C | D) \\
&= I(X_k; Y^{k-1} | A_n^{(1)} A_n^{(0)} X^{k-1} Y_k^n) && (225) \\
&\leq I(X_k; A_n^{(0)} Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) && (226) \\
&= I(X_k; Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)) && (227) \\
&\leq I(A_n^{(1)} X_k; Y^{k-1} | X^{k-1} Y_k^n) && (228)
\end{aligned}$$

$$
\begin{aligned}
&\leq I(A_n^{(1)}X_k^n; Y^{k-1}|X^{k-1}Y_k^n) && (229)\\
&= I(X_k^n; Y^{k-1}|X^{k-1}Y_k^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (230)\\
&= I(X_k^nY_k^n; X^{k-1}Y^{k-1}) && (231)\\
&= 0. && (232)
\end{aligned}
$$

From Lemma 8, the bound (224) becomes

$$
\begin{aligned}
n(R_2+\delta) &\geq \sum_{k=1}^{n} I(A;B|CD) && (233)\\
&\geq \sum_{k=1}^{n} I(A;B|D) && (234)\\
&= \sum_{k=1}^{n} I(X_k; A_n^{(2)}|A_n^{(1)}A_n^{(0)}X^{k-1}Y_k^n). && (235)
\end{aligned}
$$

Let us define the random variable $W_k = A_n^{(2)}X^{k-1}Y_{k+1}^n$. With these definitions, we have the Markov structure $W_k \to V_kX_k \to U_kY_k$ because

$$
\begin{aligned}
&I(U_kY_k; W_k|V_kX_k)\\
&= I(Y_k; A_n^{(2)}|A_n^{(1)}A_n^{(0)}X^kY_{k+1}^n) && (236)\\
&\leq I(Y_k; A_n^{(2)}X_{k+1}^n|A_n^{(1)}A_n^{(0)}X^kY_{k+1}^n) && (237)\\
&= I(Y_k; X_{k+1}^n|A_n^{(1)}A_n^{(0)}X^kY_{k+1}^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(A_n^{(0)}, X^n)) && (238)\\
&\leq I(A_n^{(0)}Y_k; X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) && (239)\\
&\leq I(A_n^{(0)}Y^k; X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) && (240)\\
&= I(Y^k; X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) \quad (\because A_n^{(0)} = \varphi_{(0)}^n(A_n^{(1)}, Y^n)) && (241)\\
&\leq I(Y^k; A_n^{(1)}X_{k+1}^n|X^kY_{k+1}^n) && (242)\\
&= I(Y^k; X_{k+1}^n|X^kY_{k+1}^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) && (243)\\
&\leq I(X^kY^k; X_{k+1}^nY_{k+1}^n) && (244)\\
&= 0. && (245)
\end{aligned}
$$

Substituting V_k and W_k into (235), we have

$$
R_2 \geq \sum_{k=1}^{n} I(X_k; W_k|V_kY_k)
$$

In the same way as the above discussion, we have the Markov structure $W \to VX \to Y$ and

$$
R_2 \geq I(X; W|VY).
$$

We next show the existence of functions $\phi_{(1)}$ and $\phi_{(2)}$ that satisfy the conditions of Theorem 2. From Definition 5, for any $\gamma > 0$ there exists an integer

$n_2 = n_2(\gamma)$, and for all $n \geq n_2(\gamma)$, we have

$$
\begin{aligned}
D_1+\gamma &\geq E\left[\Delta^n(X^n,\widehat{X}^n_{(1)})\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{X}_{(1)k})\right],\\
D_2+\gamma &\geq E\left[\Delta^n(X^n,\widehat{X}^n_{(2)})\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{X}_{(2)k})\right],
\end{aligned}
$$

Now, we denote by $\widehat{\varphi}_{(i)k}$ $(i=1,2)$ the output of $\widehat{\varphi}_{(i)}$ at time k $(k \in \mathcal{I}_n)$, namely

$$
\begin{aligned}
\widehat{x}_{(1)k} &= \widehat{\varphi}_{(1)k}(\varphi^n_{(1)}(\boldsymbol{x}),\boldsymbol{y}),\\
\widehat{x}_{(2)k} &= \widehat{\varphi}_{(2)k}(\varphi^n_{(2)}(\varphi^n_{(0)}(\varphi^n_{(1)}(\boldsymbol{x}),\boldsymbol{y}),\boldsymbol{x}),\boldsymbol{y}).
\end{aligned}
$$

We note that U_kY_k contains $A^{(1)}_n Y^n_k$, and W_kY_k contains $A^{(2)}_n Y^n_k$. Let $Y^{k-1}(U_k,Y_k)$ be a random variable selected to minimize the average distortion between X_k and $\widehat{X}_{(1)k}$ given U_kY_k, and let $Y^{k-1}(W_k,Y_k)$ be a random variable selected to minimize the average distortion between X_k and $\widehat{X}_{(2)k}$ given W_kY_k, namely

$$
\begin{aligned}
Y^{k-1}(U_k,Y_k) &\overset{\text{def.}}{=} \mathop{\arg\min}_{Y^{k-1}\in\mathcal{Y}^{k-1}} \sum_{X_k\in\mathcal{X}} Q^{(1)}_k(X_k|U_kY_k)\Delta(X_k,\widehat{\varphi}_{(1)k}(A^{(1)}_n,Y^n)),\\
Y^{k-1}(W_k,Y_k) &\overset{\text{def.}}{=} \mathop{\arg\min}_{Y^{k-1}\in\mathcal{Y}^{k-1}} \sum_{X_k\in\mathcal{X}} Q^{(2)}_k(X_k|W_kY_k)\Delta(X_k,\widehat{\varphi}_{(2)k}(A^{(2)}_n,Y^n)),
\end{aligned}
$$

where $Q^{(1)}_k$ is the distribution of X_k given U_kY_k, and $Q^{(2)}_k$ is the distribution of X_k given W_kY_k, e.g.

$$
\begin{aligned}
Q^{(1)}_k(x|uy) &= Q^{(1)}_k\left(x\,\middle|\,\widehat{A},y*\widehat{y}^n_{k+1}\right)\\
&= \frac{\displaystyle\sum_{\substack{(x^n_{k+1},y^{k-1})\in\mathcal{X}^{n-k}\times\mathcal{Y}^{k-1}:\\ \varphi^n_{(1)}(x^n)=\widehat{A},\ x^k=\widehat{x}^{k-1}*x,\ y^n_k=y*\widehat{y}^n_{k+1}}} P_{XY}(x^n,y^n)}{\displaystyle\sum_{\substack{(x^n_k,y^{k-1})\in\mathcal{X}^n\times\mathcal{Y}^{k-1}:\\ \varphi^n_{(1)}(x^n)=\widehat{A},\ x^{k-1}=\widehat{x}^{k-1},\ y^n_k=y*\widehat{y}^n_{k+1}}} P_{XY}(x^n,y^n)},\\
&\quad \text{for } u=\widehat{A}\widehat{y}^n_{k+1}, x\in\mathcal{X}, y\in\mathcal{Y},
\end{aligned}
$$

where $*$ stands for a concatination of sequences. From the above definitions, we choose the functions $\phi_{(1)}$ and $\phi_{(2)}$ as follows:

$$
\phi_{(1)k}(U_k,Y_k) \overset{\text{def.}}{=} \widehat{\varphi}_{(1)k}(A^{(1)}_n,Y^{k-1}(U_k,Y_k)*Y^n_k)=\widehat{X}_{(1)k}, \tag{246}
$$

$$
\phi_{(2)k}(W_k,Y_k) \overset{\text{def.}}{=} \widehat{\varphi}_{(2)k}(A^{(2)}_n,Y^{k-1}(W_k,Y_k)*Y^n_k)=\widehat{X}_{(2)k}, \tag{247}
$$

$$
\phi_{(1)}(U,Y) \overset{\text{def.}}{=} \phi_{(1)J}(U_J,Y), \tag{248}
$$

$$\phi_{(2)}(W,Y) \overset{\text{def.}}{=} \phi_{(2)J}(W_J,Y). \tag{249}$$

This implies that

$$\begin{aligned}
D_1+\gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{X}_{(1)k})\right] && (250)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{\varphi}_{(1)k}(A_n^{(1)},Y^n))\right] && (251)\\
&\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\widehat{\varphi}_{(1)k}(A_n^{(1)},Y^{k-1}(U_k,Y_k)*Y_k^n))\right] && (252)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X_k,\phi_{(1)k}(U_k,Y_k))\right] && (253)\\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta(X,\phi_{(1)J}(U_J,Y))|J=k\right] && (254)\\
&= E\left[\Delta(X,\phi_{(1)}(U,Y))\right], && (255)\\
D_2+\gamma &\geq E\left[\Delta(X,\phi_{(2)}(W,Y))\right]. && (256)
\end{aligned}$$

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

$$\begin{aligned}
D_1 &\geq E\left[\Delta(X,\phi_{(1)}(U,Y))\right], && (257)\\
D_2 &\geq E\left[\Delta(X,\phi_{(2)}(W,Y))\right]. && (258)
\end{aligned}$$

It remains to establish that the cardinality bounds on $|\mathcal{U}|$, $|\mathcal{V}|$ and $|\mathcal{W}|$ specified in Theorem 2 does not affect the region $\mathcal{R}_f(X,Y|D_1,D_2)$. However, we can derive the bounds in a similar manner to the proof of Theorem 1.

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

5.4 Theorem 2: direct part

Proof.
Let a distortion pair (D_1,D_2) be given, and let U, V, W and P_{UVWXY} satisfy the conditions that define $\mathcal{R}_f(X,Y|D_1,D_2)$. Fix arbitrary $\gamma,\delta>0$.

Codeword selection: $\varphi_{(1)}^n$

(1) Randomly generate M_U independent codewords $\mathcal{A}_U=\{\boldsymbol{u}_i\}_{i=1}^{M_U}$, $\boldsymbol{u}_i\in\mathcal{U}^n$, each of length n, according to the distribution P_U.
(2) Divide the codebook $\mathcal{A}_U$ into N_U bins, each containing $L_U=M_U/N_U$ members of $\mathcal{A}_U$. Let $\mathcal{A}_U(j)$ denote the set of elements $\boldsymbol{u}\in\mathcal{A}_U$ assigned to bin j $(j\in\mathcal{I}_{N_U})$.

Codeword selection: $\varphi_{(0)}^n$

In the same way as $\varphi_{(1)}^n$, generate $\mathcal{A}_V=\{\boldsymbol{v}_i\}_{i=1}^{M_V}$ according to the distribution

P_V, and divide $\mathcal{A}_V$ into N_V bins. Let $\mathcal{A}_V(j)$ denote the set of elements $\boldsymbol{v} \in \mathcal{A}_V$ assigned to bin j ($j \in \mathcal{I}_{N_V}$). Note that $\mathcal{A}_V$ does not depend on any $\boldsymbol{u} \in \mathcal{A}_U$.

Codeword selection: $\varphi^n_{(2)}$

(1) For each $\boldsymbol{v} \in \mathcal{A}_V$, randomly generate M_W independent codewords $\mathcal{A}_W(\boldsymbol{v}) = \{\boldsymbol{w}_i(\boldsymbol{v})\}_{i=1}^{M_W}$, $\boldsymbol{w}_i(\boldsymbol{v}) \in \mathcal{W}^n$, each of length n, according to the distribution $P_{W|V}$.
(2) Divide the codebook $\mathcal{A}_W(\boldsymbol{v})$ into N_W bins, each containing $L_W = M_W/N_W$ members of $\mathcal{A}_W(\boldsymbol{v})$. Let $\mathcal{A}_W(\boldsymbol{v}, j)$ denote the set of elements $\boldsymbol{w} \in \mathcal{A}_W(\boldsymbol{v})$ assigned to bin j ($j \in \mathcal{I}_{N_W}$).

Encoding: $\varphi^n_{(1)}$

(1) For a given $\boldsymbol{x} \in \mathcal{X}^n$, the encoder seeks a vector $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{x}) \in T_{UX}(k_1\delta)$, $k_1 > 0$. If there is more than one such vector in $\mathcal{A}_U$, the first one is chosen. If there is no such vector in $\mathcal{A}_U$, a default vector is chosen, say $\boldsymbol{u}_1$, and an error is declared. Denote the selected vector by $\boldsymbol{u}(\boldsymbol{x})$.
(2) The value assigned to the encoder $\varphi^n_{(1)}(\cdot)$ is the bin number to which $\boldsymbol{u}(\boldsymbol{x})$ belongs, that is,

$$\varphi^n_{(1)}(\boldsymbol{x}) \quad = \quad j_u, \quad \boldsymbol{u}(\boldsymbol{x}) \in \mathcal{A}_U(j_u). \tag{259}$$

Decoding: $\widehat{\varphi}^n_{(1)}$

(1) The decoder has access to the index j_u and the vector $\boldsymbol{y} \in \mathcal{Y}^n$.
(2) It seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(j_u)$ such that $(\boldsymbol{u}, \boldsymbol{y}) \in T_{UY}(k_2\delta)$, $k_2 > 0$. Denote this vector $\widehat{\boldsymbol{u}}(\boldsymbol{y})$. If there is no such vector or more than one such vector $\boldsymbol{u} \in \mathcal{A}_U(j_u)$, an arbitrary $\widehat{\boldsymbol{u}}$ is chosen, and an error is declared.
(3) The reconstruction vector $\widehat{\boldsymbol{x}}_{(1)} = (\widehat{x}_{(1)1}, \widehat{x}_{(1)2}, \cdots, \widehat{x}_{(1)n})$ is given by

$$\widehat{x}_{(1)k} \quad = \quad \phi_{(1)}(\widehat{u}_k(\boldsymbol{y}), y_k) \quad (k \in \mathcal{I}_n). \tag{260}$$

Encoding: $\varphi^n_{(0)}$

(1) For a given $\boldsymbol{y} \in \mathcal{Y}^n$, the encoder seeks a vector $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{y}) \in T_{VY}(k_3\delta)$, $k_3 > 0$. If there is more than one such vector in $\mathcal{A}_V$, the first one is chosen. If there is no such vector in $\mathcal{A}_V$, a default vector is chosen, say $\boldsymbol{v}_1$, and an error is declared. Denote the selected vector by $\boldsymbol{v}(\boldsymbol{y})$.
(2) The value assigned to the encoder $\varphi^n_{(0)}(\cdot, \cdot)$ is the bin number to which $\boldsymbol{v}(\boldsymbol{y})$ belongs, that is,

$$\varphi^n_{(0)}(j_u, \boldsymbol{x}) \quad = \quad j_v, \quad \boldsymbol{v}(\boldsymbol{y}) \in \mathcal{A}_V(j_v). \tag{261}$$

It should be noted that outputs of $\varphi^n_{(0)}$ do not depend on the codeword j_u of $\varphi^n_{(1)}$ albeit the codeword j_u is available.

Decoding: $\varphi^n_{(2)}$

(1) The encoder $\varphi^n_{(2)}$ has access to the index j_v and the vector $\boldsymbol{x} \in \mathcal{X}^n$.
(2) It seeks a unique vector $\boldsymbol{v} \in \mathcal{A}_V(j_v)$ such that $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{v}, \boldsymbol{x}) \in T_{UVX}(k_4\delta)$, $k_4 > 0$. Denote this vector $\widehat{\boldsymbol{v}}(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{x})$. If there is no such vector or more than

one such vector $\boldsymbol{v} \in \mathcal{A}_V(j_v)$, an arbitrary $\widehat{\boldsymbol{v}}$ is chosen, and an error is declared.

Encoding: $\varphi_{(2)}^n$

(1) For a given $\boldsymbol{x} \in \mathcal{X}^n$ and $\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{v}}(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{x})$, the encoder seeks a vector $\boldsymbol{w}_i \in \mathcal{A}_W(\widehat{\boldsymbol{v}})$ such that $(\widehat{\boldsymbol{v}}, \boldsymbol{w}_i, \boldsymbol{x}) \in T_{VWX}(k_5\delta)$, $k_5 > 0$. If there is more than one such vector in $\mathcal{A}_W(\widehat{\boldsymbol{v}})$, the first one is chosen. If there is no such vector in $\mathcal{A}_W(\widehat{\boldsymbol{v}})$, a default vector is chosen, and an error is declared. Denote the selected vector by $\boldsymbol{w}(\widehat{\boldsymbol{v}}, \boldsymbol{x})$.
(2) The value assigned to the encoder $\varphi_{(2)}^n(\cdot, \cdot)$ is the bin number to which $\boldsymbol{w}(\widehat{\boldsymbol{v}}, \boldsymbol{x})$ belong, that is,

$$\varphi_{(2)}^n(j_v, \boldsymbol{x}) \quad = \quad j_w, \quad \boldsymbol{w}(\widehat{\boldsymbol{v}}, \boldsymbol{x}) \in \mathcal{A}_W(\widehat{\boldsymbol{v}}, j_w). \tag{262}$$

Decoding: $\widehat{\varphi}_{(2)}^n$

(1) The decoder has access to the index j_w and the vector $\boldsymbol{y} \in \mathcal{Y}^n$.
(2) It seeks a unique vector $\boldsymbol{w} \in \mathcal{A}_W(\boldsymbol{v}(\boldsymbol{y}), j_w)$ such that $(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{w}, \boldsymbol{y}) \in T_{VWY}(k_6\delta)$, $k_6 > 0$. Denote this vector $\widehat{\boldsymbol{w}}(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{y})$. If there is no such vector or more than one such vector $\boldsymbol{w} \in \mathcal{A}_W(\boldsymbol{v}(\boldsymbol{y}), j_w)$, an arbitrary $\widehat{\boldsymbol{w}}$ is chosen, and an error is declared.
(3) The reconstruction vector $\widehat{\boldsymbol{x}}_{(2)} = (\widehat{x}_{(2)1}, \widehat{x}_{(2)2}, \cdots, \widehat{x}_{(2)n})$ is given by

$$\widehat{x}_{(2)k} \quad = \quad \phi_{(2)}(\widehat{w}_k(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{y}), y_k) \quad (k \in \mathcal{I}_n). \tag{263}$$

Distortion evaluation: $\widehat{\varphi}_{(1)}^n$

Almost the same way as *Distortion evaluation:* $\widehat{\varphi}_{(1)}^n$ of Theorem 1. If probabilities of encoding/decoding errors and $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T_{UXY}(k_7\delta)$ vanish as $n \to \infty$, we obtain

$$\limsup_{n \to \infty} E\left[\Delta_X^n(X^n, \widehat{X}_{(1)}^n)\right] \quad \leq \quad D_1. \tag{264}$$

Distortion evaluation: $\widehat{\varphi}_{(2)}^n$

In a similar manner to $\widehat{\varphi}_{(1)}^n$, if the probabilities of encoding/decoding errors and $(\boldsymbol{w}(\widehat{\boldsymbol{v}}, \boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T_{WXY}(k_8\delta)$ vanish as $n \to \infty$, we obtain

$$\limsup_{n \to \infty} E\left[\Delta^n(X^n, \widehat{X}_{(2)}^n)\right] \quad \leq \quad D_2. \tag{265}$$

Error evaluation: $\varphi_{(1)}^n$

If there is no $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{x}) \in T_{UX}(k_1\delta)$, an encoding error occurs. In the same way as *Error evaluation:* $\varphi_{(1)}^n$ of Theorem 1, the probability of the event vanishes as $n \to \infty$ by setting

$$M_U \geq \exp\{n(I(X;U) + m_1\gamma)\}, \; m_1 > 0. \tag{266}$$

Error evaluation: $\widehat{\varphi}^n_{(1)}$

If there is no or more than one $\boldsymbol{u}_i \in \mathcal{A}_U(j_u)$ such that $\varphi^n_{(1)}(\boldsymbol{x}) = j_u$ and $(\boldsymbol{u}_i, \boldsymbol{y}) \in T_{UY}(k_2\delta)$, an decoding error is declared. This event is classified into two cases.

(1) The first case: $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{y}) \notin T_{UY}(k_2\delta)$. In the same way as *Error evaluation:* $\widehat{\varphi}^n_{(1)}$ of Theorem 1, the probability of the event vanishes as $n \to \infty$ by introducing the Markov lemma. Through the discussion, we can also obtain $\Pr\{(U^n(X^n), X^n, Y^n) \notin T_{UXY}(k_7\delta)\} \to 0$ $(n \to \infty)$ by setting k_7 and k_2 as $k_7 < k_2/|\mathcal{X}|$, which is necessary to show Eq. (264).

(2) The second case: There exists $\boldsymbol{u}' \in \mathcal{A}_U(j_u)$ such that $\boldsymbol{u}' \neq \boldsymbol{u}(\boldsymbol{x})$ and $(\boldsymbol{u}', \boldsymbol{y}) \in T_{UY}(k_2\delta)$. In the same way as *Error evaluation:* $\widehat{\varphi}^n_{(1)}$ of Theorem 1, the probability of the event vanishes as $n \to \infty$ by setting

$$L_U \leq \exp\{n(I(Y;U) - l_1\gamma)\},\ l_1 > 0. \tag{267}$$

Error evaluation: $\varphi^n_{(0)}$

If there is no $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{y}) \in T_{VY}(k_3\delta)$, an encoding error occurs. In the same way as $\widehat{\varphi}^n_{(1)}$, the probability of the event vanishes as $n \to \infty$ by setting

$$M_V \geq \exp\{n(I(Y;V) + m_2\gamma)\},\ m_2 > 0. \tag{268}$$

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

(1) If there is no or more than one $\boldsymbol{v}_i \in \mathcal{A}_V(j_v)$ such that $\varphi^n_{(0)}(j_u, \boldsymbol{y}) = j_v$ and $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{v}_i, \boldsymbol{x}) \in T_{UVX}(k_4\delta)$, an decoding error is declared. This event is classified into two cases.

(1-1) The first case: $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \notin T_{UVX}(k_4\delta)$. In the same way as $\widehat{\varphi}^n_{(1)}$, the probability of the event vanishes as $n \to \infty$ by introducing the Markov lemma.

(1-2) The second case: There exists $\boldsymbol{v}' \in \mathcal{A}_V(j_v)$ such that $\boldsymbol{v}' \neq \boldsymbol{v}(\boldsymbol{v})$ and $(\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{v}', \boldsymbol{x}) \in T_{UVX}(k_4\delta)$. In the same way as $\widehat{\varphi}^n_{(1)}$, the probability of the event vanishes as $n \to \infty$ by setting

$$L_V \leq \exp\{n(I(UX;V) - l_2\gamma)\},\ l_2 > 0. \tag{269}$$

(2) If there is no $\boldsymbol{w}_i \in \mathcal{A}_W(\boldsymbol{v}(\boldsymbol{y}))$ such that $(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{w}_i, \boldsymbol{x}) \in T_{VWX}(k_5\delta)$, an encoding error occurs. In almost the same way as *Error evaluation:* $\varphi^n_{(2)}$ of Theorem 1, the probability of this event vanishes as $n \to \infty$ by setting M_W as

$$M_W \geq \exp\{n(I(X;W|V) + m_3\gamma)\},\ m_3 > 0. \tag{270}$$

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

This is almost the same as *Error evaluation:* $\varphi^n_{(1)}$ of Theorem 1,. We have to set

$$L_W \quad \leq \quad \exp\{n(I(Y;W|V) - l_3\gamma)\}, \quad l_3 > 0 \tag{271}$$

to vanish the decoding errors. Through the discussion, we can also obtain $\Pr\{(W^n(\widehat{V}^n(X^n), X^n), X^n, Y^n) \notin T_{WXY}(k_8\delta)\} \to 0$ $(n \to \infty)$ by setting k_8 and k_6 as $k_8 < k_6/|\mathcal{X}|$, which is necessary to show Eq. (265).

Rate evaluation: $\varphi^n_{(1)}$

The encoder sends the indexes of the bin using

$$\begin{aligned}
\frac{1}{n}\log N_U &= \frac{1}{n}\log\frac{M_U}{L_U} && (272)\\
&\geq I(X;U) + m_1\gamma - I(Y;U) + l_1\gamma && (273)\\
&= I(XY;U) - I(Y;U) + (m_1 + l_1)\gamma && (274)\\
&\qquad (\because U \to X \to Y) \\
&= I(X;U|Y) + (m_1 + l_1)\gamma && (275)
\end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n}\log N_U \geq I(X;U|Y). \qquad (276)$$

Rate evaluation: $\varphi^n_{(0)}$

The encoder sends the indexes of the bin using

$$\begin{aligned}
\frac{1}{n}\log N_V &= \frac{1}{n}\log\frac{M_V}{L_V} && (277)\\
&\geq I(Y;V) + m_2\gamma - I(UX;V) + l_2\gamma && (278)\\
&= I(UXY;V) - I(UX;V) + (m_2 + l_2)\gamma && (279)\\
&\qquad (\because V \to Y \to UX) \\
&= I(Y;V|UX) + (m_2 + l_2)\gamma && (280)
\end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n}\log N_V \geq I(Y;V|UX). \qquad (281)$$

Rate evaluation: $\varphi^n_{(2)}$

The encoder sends the indexes of the bin using

$$\begin{aligned}
\frac{1}{n}\log N_W &= \frac{1}{n}\log\frac{M_W}{L_W} && (282)\\
&\geq I(X;W|V) + m_3\gamma - I(Y;W|V) + l_3\gamma && (283)\\
&= I(XY;W|V) - I(Y;W|V) + (m_3 + l_3)\gamma && (284)\\
&\qquad (\because W \to VX \to Y) \\
&= I(X;W|VY) + (m_3 + l_3)\gamma && (285)
\end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as

$$\frac{1}{n}\log N_W \geq I(X;W|VY). \qquad (286)$$

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

5.5 Theorem 3: converse

Proof.
Let a sequence $\{(\varphi^n_{(01)}, \varphi^n_{(02)}, \varphi^n_{(1)}, \varphi^n_{(2)}, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ of CFR codes be given to satisfy the conditions of Definitions 7 and 8. From Definition 8, for $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$, and then for all $n \geq n_1(\delta)$, we can obtain

$$\frac{1}{n}\log M_n^{(i)} \quad \leq \quad R_i + \delta. \quad (i = 01, 02, 1, 2) \tag{287}$$

Please remember $A_n^{(1)} = \varphi^n_{(1)}(X^n)$, $A_n^{(01)} = \varphi^n_{(01)}(A_n^{(1)}, Y^n)$, $A_n^{(02)} = \varphi^n_{(02)}(A_n^{(1)}, Y^n)$ and $A_n^{(2)} = \varphi^n_{(2)}(A_n^{(01)}, X^n)$. First, we evaluate Eq. (287) for $i = 1$. In a similar manner as the proof of the converse part for Theorem 1, we obtain

$$\begin{aligned} & n(R_1 + \delta) \\ & \geq \sum_{k=1}^{n} I(X_k; A_n^{(1)} A_n^{(02)} X^{k-1} Y^{k-1} Y_{k+1}^n | Y_k) \quad (288) \\ & \geq \sum_{k=1}^{n} I(X_k; A_n^{(1)} A_n^{(02)} X^{k-1} Y_{k+1}^n | Y_k) \quad (289) \end{aligned}$$

Let us define the random variables $U_k = A_n^{(1)} X^{k-1} Y_{k+1}^n$ and $V_k^{(2)} = A_n^{(02)} X^{k-1} Y_{k+1}^n$. With these definitions, we have the Markov structure $U_k \to X_k \to Y_k$ in a similar manner as Section 5.1, and we obtain

$$n(R_1 + \delta) \quad \geq \quad \sum_{k=1}^{n} I(X_k; U_k V_k | Y_k).$$

Introducting a random variable J, the Markov condition $U \to X \to Y$ still holds, and we have

$$R_1 + \delta \quad \geq \quad I(X; UV^{(2)} | Y). \tag{290}$$

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

$$R_1 \quad \geq \quad I(X; UV^{(2)} | Y).$$

The evaluation of Eq. (287) for $i = 02$ is almost the same as the proof of Theorem 1. Notice that $A_n^{(0)}$ is replaced as $A_n^{(02)}$ here. Next, we evaluate Eq. (287) for $i = 01$. We obtain

$$\begin{aligned} & n(R_{01} + \delta) \\ & \geq \log M_n^{(01)} \quad (291) \\ & \geq H(A_n^{(01)}) \quad (292) \\ & \geq H(A_n^{(01)} | A_n^{(1)} A_n^{(02)} X^n) \quad (293) \\ & = I(Y^n; A_n^{(01)} | A_n^{(1)} A_n^{(02)} X^n) \quad (\because A_n^{(01)} = \varphi^n_{(01)}(A_n^{(1)}, Y^n)) \quad (294) \end{aligned}$$

$$= \sum_{k=1}^{n} I(Y_k; A_n^{(01)} | A_n^{(1)} A_n^{(02)} X^n Y_{k+1}^n) \tag{295}$$

Here, we would like to apply Lemma 8 (See the proof of Theorem 2). Let $A = Y_k$, $B = A_n^{(01)}$, $C = X_{k+1}^n$ and $D = A_n^{(1)} A_n^{(02)} X^k Y_{k+1}^n$. Then, we have

$$\begin{aligned}
& I(A; C|D) \\
&= I(Y_k; X_{k+1}^n | A_n^{(1)} A_n^{(02)} X^k Y_{k+1}^n) && (296) \\
&\leq I(A_n^{(02)} Y_k; X_{k+1}^n | A_n^{(1)} X^k Y_{k+1}^n) && (297) \\
&\leq I(A_n^{(02)} Y^k; X_{k+1}^n | A_n^{(1)} X^k Y_{k+1}^n) && (298) \\
&= I(Y^k; X_{k+1}^n | A_n^{(1)} X^k Y_{k+1}^n) \quad (\because A_n^{(02)} = \varphi_{(02)}^n (A_n^{(1)}, Y^n)) && (299) \\
&\leq I(Y^k; A_n^{(1)} X_{k+1}^n | X^k Y_{k+1}^n) && (300) \\
&= I(Y^k; X_{k+1}^n | X^k Y_{k+1}^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n (X^n)) && (301) \\
&\leq I(X^k Y^k; X_{k+1}^n Y_{k+1}^n) && (302) \\
&= 0. && (303)
\end{aligned}$$

From Lemma 8, the bound (295) becomes

$$n(R_{01} + \delta) \geq \sum_{k=1}^{n} I(Y_k; A_n^{(01)} | A_n^{(1)} A_n^{(02)} X^k Y_{k+1}^n) \tag{304}$$

Let us define the random variable $V_k^{(1)} = A_n^{(1)} A_n^{(01)} A_n^{(02)} X^{k-1} Y_{k+1}^n$. With this definition, we have the Markov structure $V_k^{(1)} \to U_k V_k^{(2)} Y_k \to X_k$ because

$$\begin{aligned}
& I(X_k; V_k^{(1)} | U_k V_k^{(2)} Y_k) \\
&= I(X_k; A_n^{(01)} | A_n^{(1)} A_n^{(02)} X^{k-1} Y_k^n) && (305) \\
&\leq I(X_k; A_n^{(01)} A_n^{(02)} | A_n^{(1)} X^{k-1} Y_k^n) && (306) \\
&\leq I(X_k; A_n^{(01)} A_n^{(02)} Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) && (307) \\
&= I(X_k; Y^{k-1} | A_n^{(1)} X^{k-1} Y_k^n) && (308) \\
&\quad (\because A_n^{(01)} = \varphi_{(01)}^n (A_n^{(1)}, Y^n), A_n^{(02)} = \varphi_{(02)}^n (A_n^{(1)}, Y^n)) && (309) \\
&\leq I(A_n^{(1)} X_k; Y^{k-1} | X^{k-1} Y_k^n) && (310) \\
&\leq I(A_n^{(1)} X_k^n; Y^{k-1} | X^{k-1} Y_k^n) && (311) \\
&= I(X_k^n; Y^{k-1} | X^{k-1} Y_k^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n (X^n)) && (312) \\
&\leq I(X_k^n Y_k^n; X^{k-1} Y^{k-1}) && (313) \\
&= 0. && (314)
\end{aligned}$$

Substituting U_k, $V_k^{(1)}$ and $V_k^{(2)}$ into (304), we obtain

$$n(R_{01} + \delta) \geq \sum_{k=1}^{n} I(Y_k; V_k^{(1)} | U_k V_k^{(2)} X_k).$$

In the same way as the above discussion, we have the Markov structure $V^{(1)} \to UV^{(2)}Y \to X$ and

$$R_{01} \geq I(Y;V^{(1)}|UV^{(2)}X).$$

Lastly, we evaluate Eq. (3). We obtain

$$\begin{aligned}
&n(R_2+\delta)\\
&\geq \log M_n^{(2)} &(315)\\
&\geq H(A_n^{(2)}) &(316)\\
&\geq H(A_n^{(2)}|A_n^{(1)}A_n^{(01)}A_n^{(02)}Y^n) &(317)\\
&= I(X^n;A_n^{(2)}|A_n^{(1)}A_n^{(01)}A_n^{(02)}Y^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(A_n^{(01)},X^n)) &(318)\\
&= \sum_{k=1}^n I(X_k;A_n^{(2)}|A_n^{(1)}A_n^{(01)}A_n^{(02)}X^{k-1}Y^n) &(319)
\end{aligned}$$

Here, we would like to apply Lemma 8 again. Let $A = X_k$, $B = A_n^{(2)}$, $C = Y^{k-1}$ and $D = A_n^{(1)}A_n^{(01)}A_n^{(02)}X^{k-1}Y_k^n$. In the same way as the proof of Theorem 2 (Section 5.3), we have $I(A;C|D) = 0$. Notice that $A_n^{(0)}$ is replaced as $A_n^{(01)}A_n^{(02)}$ here. Therefore, from Lemma 8, the bound (319) becomes

$$n(R_2+\delta) \geq \sum_{k=1}^n I(X_k;A_n^{(2)}|A_n^{(1)}A_n^{(01)}A_n^{(02)}X^{k-1}Y_k^n). \quad (320)$$

Let us define the random variable $W_k = A_n^{(2)}X^{k-1}Y_{k+1}^n$. With this definition, we have the Markov structure $W_k \to V_k^{(1)}V_k^{(2)}X_k \to U_kY_k$ because

$$\begin{aligned}
&I(U_kY_k;W_k|V_k^{(1)}V_k^{(2)}X_k)\\
&= I(Y_k;A_n^{(2)}|A_n^{(1)}A_n^{(01)}A_n^{(02)}X^kY_{k+1}^n) &(321)\\
&\leq I(Y_k;A_n^{(2)}X_{k+1}^n|A_n^{(1)}A_n^{(01)}A_n^{(02)}X^kY_{k+1}^n) &(322)\\
&= I(Y_k;X_{k+1}^n|A_n^{(1)}A_n^{(01)}A_n^{(02)}X^kY_{k+1}^n) \quad (\because A_n^{(2)} = \varphi_{(2)}^n(A_n^{(01)},X^n)) &(323)\\
&\leq I(A_n^{(01)}A_n^{(02)}Y_k;X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) &(324)\\
&\leq I(A_n^{(01)}A_n^{(02)}Y^k;X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) &(325)\\
&= I(Y^k;X_{k+1}^n|A_n^{(1)}X^kY_{k+1}^n) \quad (\because A_n^{(i)} = \varphi_{(i)}^n(A_n^{(1)},Y^n) \text{ for } i = 01, 02) &(326)\\
&\leq I(Y^k;A_n^{(1)}X_{k+1}^n|X^kY_{k+1}^n) &(327)\\
&= I(Y^k;X_{k+1}^n|X^kY_{k+1}^n) \quad (\because A_n^{(1)} = \varphi_{(1)}^n(X^n)) &(328)\\
&\leq I(X^kY^k;X_{k+1}^nY_{k+1}^n) &(329)\\
&= 0. &(330)
\end{aligned}$$

Substituting U_k, $V_k^{(1)}$, $V_k^{(2)}$ and W_k into Eq. (320), we have

$$R_2 \geq \sum_{k=1}^n I(X_k;W_k|V_k^{(1)}V_k^{(2)}Y_k)$$

In the same way as the above discussion, we have the Markov structure $W \to V^{(1)}V^{(2)}X \to UY$ and

$$R_2 \geq I(X;W|V^{(1)}V^{(2)}Y).$$

Showing the existence of functions $\phi_{(1)}$ and $\phi_{(2)}$, and deriving the bounds on $|\mathcal{U}|$, $|\mathcal{V}^{(1)}|$, $|\mathcal{V}^{(2)}|$ and $|\mathcal{W}|$ are similar to the proof of Theorems 1 and 2.

This completes the proof of the converse part of Theorem 3. □

5.6 Theorem 3: direct part

Proof. We would like to omit the proof because this would be almost the same way as those of Theorems 1 and 2. □

References

[1] A. Kimura and T. Uyematsu, "Information-theoretical analysis of index searching: revised," Symposium on Information Theory and Its Applications (SITA), November 2006. to appear.

[2] H. Yamamoto, "Source coding theory for a triangular communication system," IEEE Trans. Inform. Theory, Vol.42, No.3, pp.848–853, May 1996.

[3] E. Yang, D. He, T. Uyematsu, and R.W. Yeung, "String matching-based universal source coding for source networks with asymptotically zero feedback," Proc. IEEE International Symposium on Information Theory (ISIT), pp.2349–2353, September 2005.

[4] I. Csiszar and J. Körner, Information theory: Coding theorems for discrete memoryless systems, Academic Press, New York, 1981.

[5] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," IEEE Trans. Inform. Theory, Vol.50, No.8, pp.1636–1654, August 2004.

[6] T. Berger, Rate distortion theory: A mathematical basis for data compression, Prentice Hall, Englewood Cliffs, NJ, 1971.