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 presence 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 presence 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.

![Block diagram of the first coding problem](image)

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

Figure 2 shows a block diagram of the second coding problem, which is quite 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 $X$ and $Y$ be finite sets, $|X|$ be the cardinality of $X$ and $I_M = \{1, 2, \cdots, M\}$. A member of $X^n$ 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 $x^j_i = (x_i, x_{i+1}, \cdots, x_j)$ for $i \leq j$. When the dimension is clear from the content, vectors will be denoted by boldface letters, i.e., $\mathbf{x} \in \mathcal{X}^n$. $\mathcal{M}(\mathcal{X})$ 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 $P_Y \in \mathcal{M}(Y)$, namely each member of $\mathcal{M}(\mathcal{X}|P_Y)$ is characterized by $P_{XY} \in \mathcal{M}(X \times Y)$ as $P_{XY} = P_{X|Y}P_Y$. A discrete memoryless source $(\mathcal{X}, P_X)$ is an infinite sequence $X = \{X_i\}_{i=1}^{\infty}$ 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

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

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 $\tilde{X}$ stand for a reconstruction alphabet, and let $\Delta : \mathcal{X} \times \tilde{X} \to [0, \infty)$ be a single-letter distortion function. The vector distortion function is defined in the usual way, i.e.

$$\Delta_X^n(x, \tilde{x}) = \frac{1}{n} \sum_{k=1}^{n} \Delta_X(x_k, \tilde{x}_k). \quad (2)$$
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 \((\varphi_n^0, \varphi_n^1, \varphi_n^2, \hat{\varphi}_n^1, \hat{\varphi}_n^2)\) 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

\[
\begin{align*}
\varphi_n^1 &: X^n \rightarrow I_{M_n^1}^{(1)} \\
\varphi_n^0 &: Y_n \times I_{M_n^1}^{(1)} \rightarrow I_{M_n^0}^{(0)} \\
\varphi_n^2 &: X^n \rightarrow I_{M_n^2}^{(2)} \\
\hat{\varphi}_n^1 &: I_{M_n^0}^{(1)} \times Y_n \rightarrow \hat{X}_n \\
\hat{\varphi}_n^2 &: I_{M_n^0}^{(0)} \times I_{M_n^2}^{(2)} \times Y_n \rightarrow \hat{X}_n,
\end{align*}
\]

where

\[
\begin{align*}
\rho_n^1 &= E\left[\Delta^n(X^n, \hat{X}_n^{(1)})\right], & \rho_n^2 &= E\left[\Delta^n(X^n, \hat{X}_n^{(2)})\right], \\
A_n^{(1)} &= \varphi_n^1(X^n), & A_n^{(0)} &= \varphi_n^0\left(A_n^{(1)}, Y_n\right), & A_n^{(2)} &= \varphi_n^2(X^n), \\
\hat{X}_n^{(1)} &= \hat{\varphi}_n^1\left(A_n^{(1)}, Y_n\right), & \hat{X}_n^{(2)} &= \hat{\varphi}_n^2\left(A_n^{(0)}, A_n^{(2)}, Y_n\right).
\end{align*}
\]

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

**Definition 3.** (CR-achievable rate region)

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

**2.2 Statement of results**

First, we state the main theorem.

**Theorem 1.** (Coding theorem of CR code)

\[
\mathcal{R}_c(X, Y | D_1, D_2) \subseteq \{ (R_0, R_1, R_2) : \]
\[
R_1 \geq I(X; UV|Y), \\
R_0 \geq I(X; V|Y), \\
R_2 \geq I(X; W|VY) \} \quad \text{(outer bound)}
\]

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
  \[
  |\mathcal{U}| \leq |\mathcal{X}| + 1, \\
  |W| \leq |\mathcal{U} \times \mathcal{X}| + 1, \\
  |V| \leq |\mathcal{U} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y}| + 3,
\]

- The following Markov chains are satisfied:
  \[
  U \rightarrow X \rightarrow Y, \\
  W \rightarrow UVX \rightarrow Y,
\]

- there exist functions \(\phi_{(1)} : \mathcal{U} \times \mathcal{Y} \rightarrow \hat{X}\) and \(\phi_{(2)} : \mathcal{V} \times \mathcal{W} \times \mathcal{Y} \rightarrow \hat{X}\) which satisfy
  \[
  D_1 \geq E[\Delta(X, \phi_{(1)}(U, Y))], \\
  D_2 \geq E[\Delta(X, \phi_{(2)}(V, W, Y))].
\]

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

\[
UV \rightarrow X \rightarrow Y, \\
W \rightarrow VX \rightarrow UY.
\]
Remark. The Markov condition $W \rightarrow VX \rightarrow UY$ is tight compared with the condition $W \rightarrow UVX \rightarrow Y$ because $I(W; Y|UVX) \leq I(W; UY|VX)$.

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

$$P_{UVWXY}(u,v,w,x,y) = P_{XY}(x,y)P_{U|X}(u|x)P_{V|UXY}(v|u,x,y)P_{W|VX}(w|v,x).$$

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

$$P_{UVWXY}(u,v,w,x,y) = P_{XY}(x,y)P_{U|X}(u|x)P_{V|UY}(v|u,y)P_{W|VX}(w|v,x).$$

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.

$$R_c(X|D_1, D_2) = \{(R_0, R_1, R_2) : R_1 \geq I(X; \hat{X}_1 V), R_0 \geq I(X; V), R_2 \geq I(X; \hat{X}_2 | V)\}$$

where the random variable $V$ taking a value in $V$ is selected such that $|V| \leq |X| + 2$. 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 $\{\varphi(i), \hat{\varphi}(i)\}_{i \in \mathcal{I}_N}$ of encoders and decoders, and $(N-1)$ cascading encoders $\{\varphi(0)\}_{i \in \mathcal{I}_{N-1}}$. Let $D = \{D_i\}_{i \in \mathcal{I}_N}$ be a set of distortion criteria, each of which corresponds to the decoder $\hat{\varphi}(i)$. We define the CR-achievable rate region $R_c(X,Y|D)$ of the source $(X,Y)$ for given distortion criteria $D$ 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 $R_0$ ($j \in \mathcal{I}_{N-1}$) corresponds to the rate of the cascading encoder $\varphi(0)$. Let $U^{(i)}$ ($i \in \mathcal{I}_N$) be an auxiliary random variable that takes a value in a finite set $U_i$, 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 $S \subseteq \mathcal{I}_N$ and $T_1, T_2 \subseteq \mathcal{I}_{N-1}$, let us define

$$U^{(S)} = \{U^{(i)} : i \in S\},$$

$$U = U^{(\mathcal{I}_N)}.$$
Figure 4: Diagram of 4-stage CR code (without side information for simplicity)

\[
\begin{align*}
V^{(i,T_2)} &= \{ V^{(i,j)} : j \in T_2 \}, \\
V^{(T_1,T_2)} &= \{ V^{(i,j)} : i \in T_2 \}, \\
V &= \bigcup_{k \in I_{N-1}} V^{(k,I_{N-1}-I_k-1)}, \\
\mathcal{U}(S) &= \prod_{i \in S} U^{(i)}, \\
\mathcal{V}^{(i,T_2)} &= \prod_{j \in T_2} V^{(i,j)}, \\
\mathcal{V}^{(T_1,T_2)} &= \prod_{i \in T_1} V^{(i,T_2)}.
\end{align*}
\]

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

\[
R_c(X,Y|D) \subseteq \{(R_0, R_i) \in I_N : R_i \geq I(X; U^{(i)} V^{(i,I_{N-1}-I_i-1)} | V^{(I_{i-1},I_{N-1}-I_{i-2})} Y), R_0 \geq I(X; V^{(I_j,I_{N-1}-I_j-1)} | Y)\} \quad \text{(outer bound)}
\]

where the random variables \(U\) and \(V\) whose alphabets are \(\mathcal{U}(I_N)\) and \(\prod_{k \in I_{N-1}} V^{(k,I_{N-1}-I_k-1)}\), respectively, are selected such that

- the alphabet sizes are bounded as

\[
|\mathcal{U}^{(i)}| \\
\leq |\mathcal{U}^{(I_i-1)}| \times \prod_{k=1}^{i-1} V^{(k,I_{N-1}-I_k-1)} \times |X| + 2,
\]
\( \{i \in \mathcal{I}_N\} \)

\[ |Y(i,j)| \leq \left| U(i) \times \prod_{k=1}^{i-1} \left\{ V(k,\mathcal{I}_{N-1-\mathcal{I}_k-1}) \right\} \times V(i,\mathcal{I}_{N-1-\mathcal{I}_j}) \times X \times Y \right| + (3j + 2), \ (i, j \in \mathcal{I}_{N-1}) \]

- The following Markov chain is satisfied:
  \[ U(i) \rightarrow V(\mathcal{I}_{i-1},\mathcal{I}_{N-1-\mathcal{I}_{i-2}}) X \rightarrow U(\mathcal{I}_{i-1}) Y \]

- there exist functions \( \{ \phi(i) : U(i) \times V(\mathcal{I}_{i-1},\mathcal{I}_{N-1-\mathcal{I}_{i-2}}) \times Y \rightarrow \hat{X} \}_{i \in \mathcal{I}_N} \) which satisfy

\[ D_i \geq E \left[ \Delta(X, \phi(i)(U(i)), V(\mathcal{I}_{i-1},\mathcal{I}_{N-1-\mathcal{I}_{i-2}}), Y) \right]. \]

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

\[ U(i) V(i,\mathcal{I}_{N-1-\mathcal{I}_i}) \rightarrow V(\mathcal{I}_{i-1},\mathcal{I}_{N-1-\mathcal{I}_{i-2}}) X \rightarrow U(\mathcal{I}_{i-1}) Y \]

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 \( \{ \varphi_n^{(0)}, \varphi_n^{(1)}, \varphi_n^{(2)}, \hat{\varphi}_n^{(1)}, \hat{\varphi}_n^{(2)} \} \) 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

\[
\varphi_n^{(0)} : X^n \to \mathcal{I}_{M_n^{(0)}} \\
\varphi_n^{(1)} : \mathcal{I}_{M_n^{(1)}} \times Y^n \to \mathcal{I}_{M_n^{(0)}} \\
\varphi_n^{(2)} : \mathcal{I}_{M_n^{(0)}} \times X^n \to \mathcal{I}_{M_n^{(2)}} \\
\hat{\varphi}_n^{(1)} : \mathcal{I}_{M_n^{(1)}} \times Y^n \to \hat{X}^n, \\
\hat{\varphi}_n^{(2)} : \mathcal{I}_{M_n^{(2)}} \times Y^n \to \hat{X}^n, \\
\rho_n^{(i)} = E \left[ \Delta^n(X^n, \hat{X}_n) \right], \ (i = 1, 2)
\]

where
\[ A_n^{(1)} = \varphi_n^{(1)}(X^n), \quad A_n^{(0)} = \varphi_n^{(0)}(A_n^{(1)}, Y^n), \]
\[ A_n^{(2)} = \varphi_n^{(2)}(A_n^{(0)}, X^n), \]
\[ \hat{X}_n^{(1)} = \hat{\varphi}_n^{(1)}(A_n^{(1)}, Y^n), \quad \hat{X}_n^{(2)} = \hat{\varphi}_n^{(2)}(A_n^{(2)}, Y^n). \]

**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 \(\{(n, M_n^{(i)}, M_n^{(1)}, M_n^{(2)}, \rho_n^{(1)}, \rho_n^{(2)})\}_{n=1}^{\infty}\) for the source \((X, Y)\) such that

\[
\limsup_{n \to \infty} \frac{1}{n} \log M_n^{(i)} \leq R_i, \quad (i = 0, 1, 2)
\]

\[
\limsup_{n \to \infty} \rho_n^{(j)} \leq D_j, \quad (j = 1, 2)
\]

**Definition 6.** (FR-achievable rate region)

\[
R_f(X, Y | D_1, D_2) = \{(R_0, R_1, R_2) : (R_0, R_1, R_2) \text{ is an FR-achievable rate triad of } (X, Y) \text{ for } (D_1, D_2)\}.
\]

### 3.2 Statement of results

First, we state the main theorem.

**Theorem 2.** (Coding theorem of FR code)

\[
R_f(X, Y | D_1, D_2) \subseteq \{(R_0, R_1, R_2) : \]

\[
R_1 \geq I(X; U | Y),
\]

\[
R_0 \geq I(Y; V | UX),
\]

\[
R_2 \geq I(X; W | VY) \} \quad \text{(outer bound)}
\]

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

- the alphabet sizes are bounded as
  \[
  |\mathcal{U}| \leq |X| + 2
  \]
  \[
  |\mathcal{V}| \leq |\mathcal{U} \times \mathcal{Y}| + 1
  \]
  \[
  |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V} \times X| + 1
  \]

- the following Markov chains are satisfied:
  \[
  U \rightarrow X \rightarrow Y;
  \]
  \[
  V \rightarrow UY \rightarrow X;
  \]
  \[
  W \rightarrow VX \rightarrow UY;
  \]
there exist functions $\phi_1 : U \times Y \rightarrow \hat{X}$ and $\phi_2 : W \times Y \rightarrow \hat{X}$ which satisfy
\[
D_1 \geq E[\Delta(X, \phi_1(U, Y))], \\
D_2 \geq E[\Delta(X, \phi_2(W, Y))].
\]

An inner bound is obtained in the same functional forms, while the Markov chains are replaced as
\[
U \rightarrow X \rightarrow Y, \\
V \rightarrow Y \rightarrow UX, \\
W \rightarrow VX \rightarrow UY.
\]

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

Remark. The Markov conditions $U \rightarrow X \rightarrow Y$, $V \rightarrow UY \rightarrow X$ and $W \rightarrow VX \rightarrow UY$ are equivalent to the condition that the joint distribution $P_{UVWXY}$ satisfies
\[
P_{UVWXY}(u, v, w, x, y) \\
= P_{XY}(x, y)P_{U|X}(u|x)P_{V|UY}(v|u, y)P_{W|VX}(w|v, x).
\]

In a similar manner, the Markov conditions $U \rightarrow X \rightarrow Y$, $V \rightarrow Y \rightarrow UX$ and $W \rightarrow VX \rightarrow UY$ are equivalent to the condition that the joint distribution $P_{UVWXY}$ satisfies
\[
P_{UVWXY}(u, v, w, x, y) \\
= P_{XY}(x, y)P_{U|X}(u|x)P_{V|Y}(v|y)P_{W|VX}(w|v, x).
\]

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.
\[
\mathcal{R}_f(X|D_1, D_2) = \{(R_0, R_1, R_2) : \\
R_1 \geq I(X; \hat{X}^{(1)}), \\
R_2 \geq I(X; \hat{X}^{(2)})\}.
\]

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 \(\{\varphi(i), \hat{\varphi}(i)\}_{i \in \mathcal{I}_N}\) of encoders and decoders, and \(N - 1\) feedback encoders \(\{\varphi_{(0j)}\}_{j \in \mathcal{I}_{N-1}}\). Let \(D = \{D_i\}_{i \in \mathcal{I}_N}\) be a set of distortion criteria, each of which corresponds to the decoder \(\hat{\varphi}(i)\). We define the FR-achievable rate region \(\mathcal{R}_f(X,Y|D)\) of the source \((X,Y)\) for given distortion criteria \(D\) 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

\[
\mathcal{U}^{(S)} = \{U^{(i)} : i \in \mathcal{S}\},
\]

\[
\mathcal{U} = \mathcal{U}^{(\mathcal{I}_N)},
\]

\[
\mathcal{V}^{(T)} = \{V^{(j)} : j \in \mathcal{T}\},
\]

\[
\mathcal{V} = \mathcal{V}^{(\mathcal{I}_{N-1})},
\]

\[
\mathcal{U}^{(\mathcal{I}_N)} = \prod_{i \in \mathcal{I}_N} \mathcal{U}^{(i)},
\]

\[
\mathcal{V}^{(\mathcal{I}_{N-1})} = \prod_{j \in \mathcal{I}_{N-1}} \mathcal{V}^{(j)}.
\]

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

\[
\mathcal{R}_f(X,Y|D) \subseteq \{(R_{0j}, R_i)_{i \in \mathcal{I}_N, j \in \mathcal{I}_{N-1}} : \]

\[
R_{0j} \geq I(Y; V^{(i)}|U^{(i)}X),
\]

\[
R_i \geq I(X; U^{(i)}|V^{(i-1)}Y) \quad \text{(outer bound)}
\]

where the random variables \(\mathcal{U}\) and \(\mathcal{V}\) 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

\[
|\mathcal{U}^{(i)}| \leq |\mathcal{U}^{(\mathcal{I}_{N-1})} \times \mathcal{V}^{(\mathcal{I}_{N-1})} \times \mathcal{X}| + 2, \quad (i \in \mathcal{I}_N)
\]

\[
|\mathcal{V}^{(i)}| \leq |\mathcal{U}^{(\mathcal{I}_N)} \times \mathcal{V}^{(\mathcal{I}_{N-1})} \times \mathcal{Y}| + 1, \quad (i \in \mathcal{I}_{N-1})
\]

- the following Markov chains are satisfied:

\[
U^{(i)} \to V^{(i-1)}X \to U^{(\mathcal{I}_{i-1})}V^{(\mathcal{I}_{i-1})}Y, \quad (i \in \mathcal{I}_N)
\]

\[
V^{(i)} \to U^{(i)}Y \to U^{(\mathcal{I}_{i-1})}V^{(\mathcal{I}_{i-1})}X, \quad (i \in \mathcal{I}_{N-1}),
\]

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

\[
D_i \geq E\left[\Delta(X, \phi_{(i)}(U^{(i)}, Y))\right].
\]
An inner bound is obtained with the same functional forms, while the Markov chains are replaced as

\[ U^{(i)} \rightarrow V^{(i-1)}X \rightarrow U^{(I_{i-1})}V^{(I_{i-2})}Y, \quad (i \in I_N) \]
\[ V^{(i)} \rightarrow Y \rightarrow U^{(I_i)}V^{(I_{i-1})}X, \quad (i \in I_{N-1}) \]

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 \((\varphi^n_{(01)}, \varphi^n_{(02)}, \varphi^n_{(1)}, \varphi^n_{(2)}, \hat{\varphi}^n_{(1)}, \hat{\varphi}^n_{(2)})\) 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

\[ \varphi^n_{(1)} : X^n \rightarrow I_{M^n_{(1)}} \]
\[ \varphi^n_{(01)} : I_{M^n_{(1)}} \times Y^n \rightarrow I_{M^n_{(01)}} \]
\[ \varphi^n_{(02)} : I_{M^n_{(1)}} \times Y^n \rightarrow I_{M^n_{(02)}} \]
\[ \varphi^n_{(2)} : I_{M^n_{(01)}} \times A^n \rightarrow I_{M^n_{(2)}} \]
\[ \hat{\varphi}^n_{(1)} : I_{M^n_{(1)}} \times Y^n \rightarrow \hat{X}^n, \]
\[ \hat{\varphi}^n_{(2)} : I_{M^n_{(02)}} \times I_{M^n_{(2)}} \times Y^n \rightarrow \hat{X}^n, \]

where

\[ \rho^n_{(i)} = E\left[ \Delta^n(X^n, \hat{X}^n_{(i)}) \right], \quad (i = 1, 2) \]
\[ A^n_{(1)} = \varphi^n_{(1)}(X^n), \quad A^n_{(01)} = \varphi^n_{(01)}(A^n_{(1)}, Y^n), \]
\[ A^n_{(02)} = \varphi^n_{(02)}(A^n_{(1)}, Y^n), \quad A^n_{(2)} = \varphi^n_{(2)}(A^n_{(01)}, X^n), \]
\[ \hat{X}^n_{(1)} = \hat{\varphi}^n_{(1)}(A^n_{(1)}, Y^n), \quad \hat{X}^n_{(2)} = \hat{\varphi}^n_{(2)}(A^n_{(01)}, A^n_{(2)}, Y^n). \]

**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 \(\{n, M^n_{(01)}, M^n_{(02)}, M^n_{(1)}, M^n_{(2)}, \rho^n_{(1)}, \rho^n_{(2)}\}\) for the source \((X, Y)\) such that

\[ \limsup_{n \to \infty} \frac{1}{n} \log M^n_{(i)} \leq R_i, \quad (i = 01, 02, 1, 2) \]
\[ \limsup_{n \to \infty} \rho_n^{(j)} \leq D_j. \quad (j = 1, 2) \]

**Definition 9.** (CFR-achievable rate region)

\[ R_{cf}(X,Y|D_1,D_2) = \{(R_{01}, R_{02}, R_1, R_2) : \]
\[ (R_{01}, R_{02}, R_1, R_2) \text{ is a CFR-achievable rate quadruplet of } (X,Y) \text{ for } (D_1,D_2)\}. \]

### 4.2 Statement of results

First, we state the main theorem.

**Theorem 3.** (Coding theorem of CFR code)

\[ R_{cf}(X,Y|D_1,D_2) \subseteq \{(R_{01}, R_{02}, R_1, R_2) : \]
\[ R_1 \geq I(X;U|V(2)Y), \]
\[ R_{01} \geq I(Y;V^{(1)}|UV(2)X), \]
\[ R_{02} \geq I(X;V^{(2)}Y), \]
\[ R_2 \geq I(X;W|V^{(1)}V^{(2)}Y) \} \quad \text{(outer bound)} \]

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

- the alphabet sizes are bounded as
  \[ |U| \leq |X| + 2, \]
  \[ |V^{(2)}| \leq |U \times X \times Y| + 4, \]
  \[ |V^{(1)}| \leq |U \times V^{(2)} \times Y| + 1, \]
  \[ |W| \leq |U \times V^{(1)} \times V^{(2)} \times X| + 1, \]

- the following Markov chains are satisfied:
  \[ U \to X \to Y, \]
  \[ V^{(1)} \to UV^{(2)}Y \to X, \]
  \[ W \to V^{(1)}V^{(2)}X \to UY, \]

- there exist functions \( \phi_{(1)} : U \times Y \to \hat{X} \) and \( \phi_{(2)} : W \times V^{(2)} \times Y \to \hat{X} \) which satisfy
  \[ D_1 \geq E[\Delta(X,\phi_{(1)}(U,Y))], \]
  \[ D_2 \geq E[\Delta(X,\phi_{(2)}(W,V^{(2)},Y))]. \]
An inner bound is obtained in the same functional forms, while the Markov chains are replaced as

\[
\begin{align*}
UV^{(2)} & \rightarrow X \rightarrow Y, \\
V^{(1)} & \rightarrow V^{(2)}Y \rightarrow UX, \\
W & \rightarrow V^{(1)}V^{(2)}X \rightarrow UY.
\end{align*}
\]

Remark. The Markov conditions

\[
\begin{align*}
U & \rightarrow X \rightarrow Y, \\
V^{(1)} & \rightarrow UV^{(2)}Y \rightarrow UX, \\
W & \rightarrow V^{(1)}V^{(2)}X \rightarrow UY,
\end{align*}
\]

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

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

Also, the Markov conditions

\[
\begin{align*}
U & \rightarrow X \rightarrow Y, \\
V^{(1)} & \rightarrow V^{(2)}Y \rightarrow UX, \\
W & \rightarrow V^{(1)}V^{(2)}X \rightarrow UY.
\end{align*}
\]

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

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

5 Proofs of theorems

5.1 Theorem 1: converse

Proof. Let a sequence \(\{(\varphi_{(0)}^{\delta}, \varphi_{(1)}^{\delta}, \varphi_{(2)}^{\delta}, \tilde{\varphi}_{(1)}^{\delta}, \tilde{\varphi}_{(2)}^{\delta})\}_{n=1}^{\infty}\) 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

\[
\frac{1}{n} \log M_k^{(i)} \leq R_i + \delta. \quad (i = 0, 1, 2)
\]
Let us remind you that $A_n^{(1)} = \varphi^n_1(X^n)$, $A_n^{(0)} = \varphi^n_0(A_n^{(1)}, Y^n)$ and $A_n^{(2)} = \varphi^n_2(X^n)$. First, we evaluate Eq. (3) for $i = 1$. We obtain

\[
n(R_1 + \delta) \\
\geq \log M_n^{(1)} \\
\geq H(A_n^{(1)}) \\
\geq H(A_n^{(1)}|Y^n) \\
= I(X^n; A_n^{(1)}|Y^n) \quad (\because A_n^{(1)} = \varphi_1^n(X^n)) \\
= I(X^n; A_n^{(1)} A_n^{(0)}|Y^n) \quad (\because A_n^{(0)} = \varphi_0^n(A_n^{(1)}, Y^n)) \\
= H(X^n|Y^n) - H(X^n|A_n^{(1)} A_n^{(0)} Y^n) \\
= \sum_{k=1}^n \left( H(X_k|Y_k) - H(X_k|A_n^{(1)} A_n^{(0)} X_k^{k-1} Y_n^{k+1} Y_k) \right) \\
= \sum_{k=1}^n I(X_k; A_n^{(1)} A_n^{(0)} X_k^{k-1} Y_k^{k-1} Y_n^{k+1} Y_k|Y_k).
\]

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

\[
I(Y_k; U_k|X_k) \\
= I(Y_k; A_n^{(1)} X_k^{k-1} Y_k^{k-1} Y_n^{k+1} Y_k|X_k) \\
\leq I(Y_k; A_n^{(1)} X_k^{k-1} X_k^{k+1} Y_k^{k-1} Y_n^{k+1} Y_k|X_k) \\
= I(Y_k; X_k^{k-1} X_k^{k+1} Y_k^{k-1} Y_n^{k+1} Y_k|X_k) \quad (\because A_n^{(1)} = \varphi_1^n(X^n)) \\
\leq I(X_k Y_k; X_k^{k-1} X_k^{k+1} Y_k^{k-1} Y_n^{k+1} Y_k) \\
= 0.
\]

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 $U \rightarrow X \rightarrow Y$ still holds, and we have

\[
R_1 + \delta \\
\geq \frac{1}{n} \sum_{k=1}^n I(X_k; U_k V_k|Y_k) \\
\geq \frac{1}{n} \sum_{k=1}^n \left\{ H(X_k|Y_k) - H(X_k Y_k U_k V_k) + H(Y_k|U_k V_k) \right\}
\]

15
\[
H(X|Y) + \frac{1}{n} \sum_{k=1}^{n} \{ -H(X,J|U_jV_j, J = k) + H(Y|U_jV_j, J = k) \} \\
= H(X|Y) - H(XY|U_jV_j) + H(Y|U_jV_j) \\
= H(X|Y) - H(XY|UV) + H(Y|UV) \\
= I(X; UV|Y).
\]

Since \( \delta > 0 \) is arbitrary, we obtain
\[
R \geq I(X; UV|Y).
\]

Next, we evaluate Eq. (3) for \( i = 0 \). We obtain
\[
n(R_0 + \delta) \\
\geq \log M_n \quad \quad \quad \quad (22) \\
\geq H(A_n^{(0)}) \quad \quad \quad \quad (23) \\
\geq H(A_n^{(0)}|Y^n) \quad \quad \quad \quad (24) \\
= I(X^n; A_n^{(0)}|Y^n) \quad \quad \quad \quad (25) \\
= H(X^n|Y^n) - H(X^n|A_n^{(0)}Y^n) \quad \quad \quad \quad (26) \\
= \sum_{k=1}^{n} \left\{ H(X_k|Y_k) - H(X_k|A_n^{(0)}X^{k-1}Y^n) \right\} \quad \quad \quad \quad (27) \\
= \sum_{k=1}^{n} I(X_k; A_n^{(0)}X^{k-1}Y^{n_k-1}Y_{k+1}^n|Y_k) \quad \quad \quad \quad (28) \\
= \sum_{k=1}^{n} I(X_k; V_k|Y_k) \quad \quad \quad \quad (29)
\]

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
\[
n(R_2 + \delta) \\
\geq \log M_n \quad \quad \quad \quad (30) \\
\geq H(A_n^{(2)}) \quad \quad \quad \quad (31) \\
\geq H(A_n^{(2)}|A_n^{(0)}Y^n) \quad \quad \quad \quad (32) \\
= I(X^n; A_n^{(2)}|A_n^{(0)}Y^n) \quad \quad \quad \quad (33) \\
= \sum_{k=1}^{n} I(X_k; A_n^{(2)}|A_n^{(0)}X^{k-1}Y^n) \quad \quad \quad \quad (34)
\]

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 \( W_k \rightarrow U_kV_kX_k \rightarrow Y_k \) because
\[
I(Y_k; W_k|U_kV_kX_k)
\]
The bound (34) becomes

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

In the same way as the above discussion, we have

\[ R_2 \geq I(X; W | V Y) \]

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

\[ D_1 + \gamma \geq E \left[ \Delta^n(X^n, \hat{X}^{(1)}_k) \right] = \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{X}^{(1)}_k) \right], \]

\[ D_2 + \gamma \geq E \left[ \Delta^n(X^n, \hat{X}^{(2)}_k) \right] = \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{X}^{(2)}_k) \right], \]

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

\[
\hat{x}^{(1)}_k = \tilde{\phi}^{(1)}_{(k)}(\hat{x}^{(1)}_0(x), y), \\
\hat{x}^{(2)}_k = \tilde{\phi}^{(2)}_{(k)}(\hat{x}^{(2)}_0(x), y, \hat{x}^{(1)}_0(x), y).
\]

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

\[
\phi_{(1)}(U_k, Y_k) \overset{\text{def.}}{=} \tilde{\phi}^{(1)}_{(k)}(A^{(1)}_n, Y^n) = \hat{X}^{(1)}_k, \quad (44)
\]

\[
\phi_{(2)}(V_k, W_k, Y_k) \overset{\text{def.}}{=} \tilde{\phi}^{(2)}_{(k)}(A^{(0)}_n, A^{(2)}_n, Y^n) = \hat{X}^{(2)}_k, \quad (45)
\]

\[
\phi_{(1)}(U, Y) \overset{\text{def.}}{=} \phi_{(1)}(U, Y), \quad (46)
\]
φ(2)(V, W, Y) \overset{\text{def.}}{=} φ(2), J(V, W, Y).

This implies that

\[
D_1 + \gamma \geq \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{X}_{(1)k}) \right]
\]

(48)

\[
= \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \tilde{\varphi}(1)k(A_{1}^{(1)}, Y^n)) \right]
\]

(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)

\[
D_2 + \gamma \geq E \left[ \Delta(X, \phi(2)(V, W, Y)) \right]
\]

(52)

Since γ > 0 is arbitrary, we get

\[
D_1 \geq E \left[ \Delta(X, \phi(1)(U, Y)) \right],
\]

(54)

\[
D_2 \geq E \left[ \Delta(X, \phi(2)(V, W, Y)) \right].
\]

(55)

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

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

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 \(Q \in M(X)\):

\[
q_x(Q) = Q(x), \quad (x \in I_{m_x} - 1)
\]

(56)

\[
q_{m_x}(Q) = H(X|Y)
\]

\[- \sum_{(v, x, y) \in V \times X \times Y} Q(x)P_{V|X}(v, y|x) \log \frac{Q(x)P_{V|X}(v, y|x)}{\sum_{x' \in X} Q(x')P_{V|X}(v, y|x')}\]

(57)

\[
q_{m_x+1}(Q) = \sum_{(v, x, y) \in V \times X \times Y} Q(x)P_{V|X}(v, y|x)\Delta(x, \phi(1)(u, v, y))
\]

(58)
for a given $u \in \mathcal{U}$, \hspace{1cm} (59) \\
(60)

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

\[
\sum_{u \in \tilde{U}} P_{\tilde{U}}(u)q_u(Q_u) = P_X(x) \ (x \in \mathcal{X}_{m_x-1}), \tag{61}
\]
\[
\sum_{u \in \tilde{U}} P_{\tilde{U}}(u)q_{m_x}(Q_u) = I(X;UV|Y), \tag{62}
\]
\[
\sum_{u \in \tilde{U}} P_{\tilde{U}}(u)q_{m_x+1}(Q_u) = E[\Delta(X, \varphi_1(U, V, Y))]. \tag{63}
\]

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

\[
P_{\tilde{U}VWXY}(u, v, w, x, y) \ \overset{\text{def.}}{=} \ P_{\tilde{U}}(u)Q_v(x)P_{Y|X}(y|x)P_{VW|XY}(v, w|x, y). \tag{64}
\]

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

\[
\sum_{u \in \tilde{U}} P_{\tilde{U}VWXY}(u, v, w, x, y) = \sum_{u \in \tilde{U}} P_{\tilde{U}}(u)Q_v(x)P_{Y|X}(y|x)P_{VW|XY}(v, w|x, y) \overset{\text{def.}}{=} P_Y|X(y|x), \tag{65}
\]

Also, we obtain

\[
P_{Y|\tilde{U}X}(y|u, x) = \frac{\sum_{(v, w,y) \in V \times W \times Y} P_{\tilde{U}VWXY}(u, v, w, x, y)}{\sum_{(v, w,y) \in V \times W \times Y} P_{\tilde{U}VWXY}(u, v, w, x, y)} \overset{\text{def.}}{=} \frac{P_{\tilde{U}}(u)Q_v(x)P_{Y|X}(y|x)}{P_{\tilde{U}}(u)Q_v(x)} \overset{\text{def.}}{=} P_{Y|X}(y|x) \tag{66}
\]

which indicates that the Markov chain $\tilde{U} \rightarrow X \rightarrow Y$ still remains.
Having found such a random variable $U$, we can reduce the alphabet size of $W$. In this time, we have $|U \times X| - 1$ constraints to preserve the distribution $P_{\tilde{U}X}$ just defined, and two more constraints for preserving $I(X; W| \mathcal{V}Y)$ and $E[\Delta(X, \phi(2)(V, W, Y))]$. Therefore, we can find a random variable $\tilde{W}$ taking a value in $\tilde{W} \subseteq \tilde{W}$ with a generic distribution $P_{\tilde{W}} \in \mathcal{M}(\tilde{W})$ and distributions $Q_w \in \mathcal{M}(\tilde{U} \times \mathcal{X})$ ($w \in \tilde{W}$) such that $|\tilde{W}| \leq |\tilde{U} \times \mathcal{X}| + 1$. A joint distribution $P_{\tilde{U}\tilde{W}XY}(u, v, w, x, y)$ is defined as the following equation.

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

$$
\sum_{w \in \tilde{W}} P_{\tilde{U}\tilde{W}XY}(u, v, w, x, y) = P_{\tilde{W}}(w)Q_w(u, x)P_{V|\tilde{U}X}(v|u, x)P_{Y|\tilde{V}X}(y|u, v, x)\quad(69)
$$

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

In a similar manner, we can reduce the alphabet size of $V$. In this time, we have $|U \times W \times X| \times Y - 1$ constraints to preserve the distribution $P_{UXWY}$ just defined, and 4 more constraints for preserving $I(X; UV|Y)$, $I(X; V|Y)$, $I(X; W|Y)$, 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 $y \in \mathcal{Y}^n$ and $\delta > 0$, define the set of typical sequences as

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

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

20
For any $\delta, \delta'$, Lemma 2. (Csiszar-Körner [4, Lemma 1.2.10])

\[ \Pr\{X^n \in T_X(\delta)\} \geq 1 - \epsilon_n^{(1)}(\delta), \]
\[ \Pr\{X^n \in T_X(\delta')\} \geq 1 - \epsilon_n^{(2)}(\delta, \delta') \quad \forall y \in T_Y(\delta'), \]

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

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

\[ \frac{1}{n} \log |T_X(\delta)| - H(X) \leq \epsilon_1, \]
\[ \frac{1}{n} \log |T_X(\delta')| - H(X) \leq \epsilon_2 \quad \forall y \in T_Y(\delta'), \]

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.**

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

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

**Lemma 5.** (Steinberg-Merhav[5])

\[ \exp\{-n(I(X;U) + \epsilon_1)\} \leq \sum_{u \in U} P_U(u) \leq \exp\{-n(I(X;U) - \epsilon_2)\} \]

for any $x \in T_X(\delta)$ 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].
Lemma 6. (Steinberg-Merhav[5])

\[ \exp\{-n(I(X;V|U) + \epsilon_1)\} \]
Let a distortion pair $(D_1, D_2)$ be given, and let $U$, $V$, $W$ and $P_{UVWXY}$ satisfy the conditions that define $R_c(X,Y|D_1, D_2)$. Fix arbitrary $\gamma, \delta > 0$.

**Proof.**

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

**Codeword selection: $\varphi_n^{(1)}$**

(1) Randomly generate $M_V$ independent codewords $A_V = \{v_i\}_{i=1}^{M_V}$, $v_i \in V^n$, each of length $n$, according to the distribution $P_V$.

(2) For each $v \in A_V$, randomly generate $M_U$ independent codewords $A_U(v) = \{u_i(v)\}_{i=1}^{M_U}$, $u_i(v) \in U^n$, each of length $n$, according to the distribution $P_{U|V}$.

(3) Divide the codebook $A_V$ into $N_V$ bins, each containing $L_V = M_V/N_V$ members of $A_V$. Let $A_V(j)$ denote the set of elements $v \in A_V$ assigned to bin $j \in I_{N_V}$.

(4) Divide the codebook $A_U(v)$ into $N_U$ bins, each containing $L_U = M_U/N_U$ members of $A_U(v)$. Let $A_U(v,j)$ denote the set of elements $u \in A_U(v)$ assigned to bin $j \in I_{N_U}$.

**Codeword selection: $\varphi_n^{(2)}$**

Unnecessary ($A_V$ will be used as the codeword set).

**Codeword selection: $\varphi_n^{(3)}$**

(1) For each $v \in A_V$, randomly generate $M_W$ independent codewords $A_W(v) = \{w_i(v)\}_{i=1}^{M_W}$, $w_i(v) \in W^n$, each of length $n$, according to the distribution $P_{W|V}$.

(2) Divide the codebook $A_W(v)$ into $N_W$ bins, each containing $L_W = M_W/N_W$ members of $A_W(v)$. Let $A_W(v,j)$ denote the set of elements $w \in A_W(v)$ assigned to bin $j \in I_{N_W}$.

**Encoding: $\varphi_n^{(4)}$**

(1) For a given $x \in A^n$, the encoder first seeks a vector $v_i \in A_V$ such that $(v_i, x) \in T_{XY}(k_1\delta)$, $k_1 > 0$. If there is more than one such vector in $A_V$, the

\[
\sum_{\nu(x, v_1, x) \in T_{UVX}(\delta') \left| \nu(x, v_1, x) \right| < \cdot \]
first one is chosen. If there is no such vector in $A_V$, a default vector is chosen, say $v_1$, and an error is declared. Denote the selected vector by $v(x)$.

(2) The encoder next seeks a vector $u_i \in A_U(v)$ such that $v = v(x)$ and $(u_i, v, x) \in T_{VW}(k_2\delta)$, $k_2 > 0$. If there is more than one such vector in $A_U(v)$, the first one is chosen. If there is no such vector in $A_U(v)$, a default vector is chosen, and an error is declared. Denote the selected vector by $u(v, x)$.

(3) The value assigned to the encoder $\varphi^n_{(1)}(\cdot)$ is the pair of the bin numbers to which $v(x)$ and $u(v(x), x)$ belong, that is,

$$\varphi^n_{(1)}(x) = j_vM_U + j_u,$$

$$v(x) \in A_V(j_v), \quad u(v(x), x) \in A_U(v(x), j_u).$$

**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 $x \in X^n$ and $v = v(x)$, the encoder seeks a vector $w_i \in A_W(v)$ such that $(v, w_i, x) \in T_{VW}(k_3\delta)$, $k_3 > 0$. If there is more than one such vector in $A_W(v)$, the first one is chosen. If there is no such vector in $A_W(v)$, a default vector is chosen, and an error is declared. Denote the selected vector by $w(v, x)$.

(2) The value assigned to the encoder $\varphi^n_{(2)}(\cdot)$ is the bin number to which $w(v(x), x)$ belong, that is,

$$\varphi^n_{(2)}(x) = j_w, \quad w(v(x), x) \in A_W(v(x), j_w).$$

**Decoding: $\tilde{x}^n_{(1)}$**

(1) The decoder has access to the index $j_u, j_v$ and the vector $y \in Y^n$.

(2) It first seeks a unique vector $v \in A_V(j_v)$ such that $(v, y) \in T_{VY}(k_4\delta)$, $k_4 > 0$. Denote this vector $\tilde{v}(y)$. If there is no or more than one such vectors $v \in A_V(j_v)$, an arbitrary $\tilde{v}$ is chosen, and an error is declared.

(3) It next seeks a unique vector $u \in A_U(\tilde{v}(y), j_u)$ such that $(u, \tilde{v}(y), y) \in T_{UY}(k_5\delta)$, $k_5 > 0$. Denote this vector $\tilde{u}(\tilde{v}(y), y)$.

(4) The reconstruction vector $\tilde{x}^n_{(1)}(x) = (\tilde{x}^n_{(1)1}, \tilde{x}^n_{(1)2}, \cdots, \tilde{x}^n_{(1)n})$ is given by

$$\tilde{x}^n_{(1)k} = \phi_{(1)}(\tilde{u}_{(1)k}(\tilde{v}(y), y), y_k) \quad (k \in I_n).$$

**Decoding: $\tilde{x}^n_{(2)}$**

(1) The decoder has access to the index $j_v, j_w$ and the vector $y \in Y^n$.

(2) It first seeks a unique vector $v \in A_V(j_v)$ in the same way as $\tilde{x}^n_{(1)}$.

(3) It then seeks a unique vector $w \in A_W(\tilde{v}(y), j_w)$ such that $(\tilde{v}(y), w, y) \in T_{VW}(k_6\delta)$, $k_6 > 0$. Denote this vector $\tilde{w}(\tilde{v}(y), y)$. If there is no or more than one such vectors $w \in A_W(\tilde{v}(y), j_w)$, an arbitrary $\tilde{w}$ is chosen, and an error is
declared.
(4) The reconstruction vector \( \hat{x}_{(2)} = (\hat{x}_{(2)1}, \hat{x}_{(2)2}, \cdots, \hat{x}_{(2)n}) \) is given by
\[
\hat{x}_{(2)k} = \phi_{(2)}(\hat{v}_k(y), \hat{w}_k(\hat{v}(y), y), y_k) \quad (k \in \mathcal{I}_n).
\] (106)

\textit{Distortion evaluation:} \( \hat{\varphi}_n^{(1)} \)

(1) For the distortion, we obtain
\[
\Delta^n(x, \hat{x}_{(1)}) = \frac{1}{n} \sum_{k=1}^{n} \Delta(x_k, \hat{x}_{(1)k})
\] (107)
\[
= \frac{1}{n} \sum_{k=1}^{n} \Delta(x_k, \phi_{(1)}(\hat{u}_k(\hat{v}(y), y), \hat{v}_k(y), y_k))
\] (108)
\[
= \frac{1}{n} \sum_{(u,x,y) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}} N(u, x, y | \hat{u}(\hat{v}(y), y), \hat{v}_k(y), y_k) \Delta(x, \phi_{(1)}(u, y))
\] (109)

If no error occurs in the encoding/decoding processes and \((u(v(x), x), x, y) \in T_{UXY}(k\delta)\), then the following inequalities satisfy:
\[
\Delta^n(x, \hat{x}_{(1)}) \leq \sum_{(u,x,y) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}} (P_{UXY}(u, x, y) + k\delta) \Delta(x, \phi_{(1)}(u, y))
\] (110)
\[
\leq E[\Delta(x, \phi_{(1)}(u, y))] + k\delta \sum |\mathcal{U} \times \mathcal{X} \times \mathcal{Y}|
\] (111)
\[
\leq D_1 + k\delta \sum |\mathcal{U} \times \mathcal{X} \times \mathcal{Y}|
\] (112)

where
\[
\sum \overset{\text{def.}}{=} \max_{(x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}} \Delta(x, \hat{x}) < \infty.
\] (113)

Let us define
\[
E_1 \overset{\text{def.}}{=} \{ (u(v(x), x), v(x), x, y) \notin T_{UXY}(k\delta) \},
\] (114)
and let us denote error probabilities in the encoding/decoding processes as \( P^n_{e} \).

Then, the average distortion can be bounded as
\[
E \left[ \Delta^n(X^n, \hat{X}^{n}_{(1)}) \right] \leq (1 - P^n_{e} - \Pr\{E_1\})(D_1 + k\delta \sum |\mathcal{U} \times \mathcal{X} \times \mathcal{Y}|) + (P^n_{e} + \Pr\{E_1\}) \sum.
\] (115)
Since $\delta > 0$ can be arbitrarily small for a sufficiently large $n$. Therefore, if $P^m_n$ and $\Pr\{E_1\}$ vanish as $n \to \infty$, then we can obtain

$$
\limsup_{n \to \infty} E \left[ \Delta_n^m \{ X^n, \hat{X}^n \} \right] \leq D_1.
$$

**Distortion evaluation:** $\hat{\varphi}^n_{(2)}$

In a similar manner as $\hat{\varphi}^n_{(1)}$, we can obtain

$$
\limsup_{n \to \infty} E \left[ \Delta^n(X^n, \hat{X}^n_{(2)}) \right] \leq D_2.
$$

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

(1) If there is no $v_i \in A_v$ such that $(v_i, x) \in T_{VX}(k_1 \delta)$, an encoding error occurs. This event is denoted as

$$
E_2 \overset{\text{def.}}{=} \bigcap_{i=1}^{M_V} \left\{ (v_i, x) \notin T_{VX}(k_1 \delta) \right\}.
$$

Here, let us define

$$
E_0 \overset{\text{def.}}{=} \left\{ (x, y) \in T_{XY}(k_0 \delta) \right\}, \quad k_0 > 0.
$$

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

$$
\Pr\{E_2\} \leq \Pr\{E_0 \cup E_2\} \leq \Pr\{E_0\} + \Pr\{E_0 \cap E_2\}
$$

$$
\Pr\{E_0 \cap E_2\}
$$

$$
\leq \sum_{x \in T_X(k_0 \delta | \gamma|)} P_X(x) \Pr \left\{ \bigcap_{i=1}^{M_V} \left\{ (v_i^n, x) \notin T_{VX}(k_1 \delta) \right\} \bigg| x \right\}
$$

$$
= \sum_{x \in T_X(k_0 \delta | \gamma|)} P_X(x) \Pr \left\{ \bigcap_{i=1}^{M_V} \left\{ (v_i^n, x) \notin T_{VX}(k_1 \delta) \right\} \right\}
$$

$$
\left(\because v_i \text{ is selected independently of } x\right)
$$

$$
\leq \sum_{x \in T_X(k_0 \delta | \gamma|)} P_X(x) \left[ 1 - \exp \left\{ -n(I(X; V) + \epsilon_v) \right\} \right]^{M_V}
$$

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

$$
\leq \sum_{x \in T_X(k_0 \delta | \gamma|)} P_X(x) \exp \left\{ -M_V \exp \left\{ -n(I(X; V) + \epsilon_v) \right\} \right\}
$$

$$
\left(\because (1 - a)^n \leq \exp(-an) \text{ for } 0 \leq a \leq 1 \right)
$$

$$
\leq \exp \left\{ -M_V \exp \left\{ -n(I(X; V) + \epsilon_v) \right\} \right\}.
$$
By setting $M_V$, $k_0$ and $k_1$ as

$$M_V \geq \exp\{n(I(X; V) + m_1)\}, \ m_1 > 0,$$ \hspace{1cm} (127)

$m_1 > \epsilon = \epsilon_n(k_0|\mathcal{Y}||, k_1\delta)$ and $k_0|\mathcal{Y}| < k_1$, we have $\lim_{n \to \infty} \Pr\{E_2\} = 0$.

(2) If there is no $u_i \in \mathcal{A}_U(v(x))$ such that $(u_i, v(x), x) \in T_UV_X(k\delta)$, an encoding error occurs. This event is denoted as

$$E_3 \quad \text{def.} = \bigcap_{i=1}^{M_U} \{(u_i, v(x), x) \not\in T_UV_X(k\delta)\}.$$ \hspace{1cm} (128)

Here, we have

$$\Pr\{E_3\} \leq \Pr\{E_2 \cup E_3\} \leq \Pr\{E_2\} + \Pr\{E_2^c \cap E_3\}.$$ \hspace{1cm} (129)

Since $(v(x), x) \in T_Y(k\delta)$ if $E_2$ does not occur, then we have

$$\Pr\{E_2^c \cap E_3\} \leq \sum_{(v, x) \in T_Y(k\delta)} P_{VX}(v, x) \Pr\left\{\bigcap_{i=1}^{M_U} \{(U_i^v, v, x) \not\in T_UV_X(k\delta)\}\right\}.$$ \hspace{1cm} (131)

$$= \sum_{(v, x) \in T_Y(k\delta)} P_{VX}(v, x) \Pr\left\{\bigcap_{i=1}^{M_U} \{(U_i^v, v, x) \not\in T_UV_X(k\delta)\}\right\}.$$ \hspace{1cm} (132)

($; u_i$ is selected independently of $x$)

$$\leq \sum_{(v, x) \in T_Y(k\delta)} P_{VX}(v, x) \left[1 - \exp\{-n(I(X; U|V) + \epsilon_{u|v})\}\right]^{M_U}.$$ \hspace{1cm} (133)

($; \text{Lemma 6}$)

$$\leq \sum_{(v, x) \in T_Y(k\delta)} P_{VX}(v, x) \exp\{-M_U \exp\{-n(I(X; U|V) + \epsilon_{u|v})\}\}.$$ \hspace{1cm} (134)

($; (1 - a)^n \leq \exp(-an)$)

$$\leq \exp\{-M_U \exp\{-n(I(X; U|V) + \epsilon_{u|v})\}\}.$$ \hspace{1cm} (135)

By setting $M_U$, $k_1$ and $k_2$ as

$$M_U \geq \exp\{n(I(X; U|V) + m_2)\}, \ m_2 > 0,$$ \hspace{1cm} (136)

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

**Error evaluation:** $\Phi(0)$

Any encoding errors do not occur because the encoder does not generate any codewords.
Error evaluation: $\phi_n^a(2)$

(2) If there is no $w_i \in A_W(v(x))$ such that $(v(x), w_i, x) \in T_{VWX}(k_3\delta)$, an encoding error occurs. In almost the same way as $\phi_n^a(1)$, the probability such that this event occurs vanishes as $n \to \infty$ by setting $M_W$ as

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

Error evaluation: $\hat{\phi}_n^a(1)$

(1) If there is no or more than one $v_i \in A_V(j_v)$ such that $\phi_n^a(1)(x) = j_v M_U + j_v$ and $(v_i, y) \in T_{VY}(k_4\delta)$, an decoding error is declared. This event is classified into two cases.

(1-1) The first case: $(v(x), y) \notin T_{VY}(k_4\delta)$. From Lemma 3, this event is included in

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

Here, we have

$$\Pr\{E_4\} \leq \Pr\{E_2 \cup E_4\}$$ (139)

$$= \Pr\{E_2\} + \Pr\{E_2^c \cap E_4\}.$$ (140)

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

$$\Pr\{E_2^c \cap E_4\} \leq \sum_{(v, x) \in T_{VX}(k_1\delta)} P_{VX}(v, x) \Pr\{(v, x, Y^n) \notin T_{VXY}(k_4\delta/|Y|)|v, x\}$$ (141)

$$\to 0 \quad (n \to \infty).$$ (142)

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

$$E_5 \overset{\text{def.}}{=} \left\{ \bigcup_{v \in A_V(j_v), v \neq v(x)} \{(v, y) \in T_{VY}(k_4\delta)\} \right\},$$ (143)

Let $i_v(j, k)$ be the index $i$ of $k$-th $v_i$, which belongs to $A_V(j_v)$. Since $(x, y) \in T_{XY}(k_0\delta) \Rightarrow y \in T_Y(k_0\delta/|X|)$ from Lemma 3, we have

$$\Pr\{E_5\} \leq \Pr\{E_6^c \cup E_5\}$$ (144)

$$= \Pr\{E_6^c\} + \Pr\{E_0 \cap E_5\}.$$ (145)

$$\Pr\{E_0 \cap E_5\}$$
By setting $L_V$, $k_0$ and $k_1$ as

$$L_V \leq \exp\{n(I(Y;V) - l_1\gamma)\}, \quad l_1 > 0,$$

(149)

$l_1 > \epsilon_2 = \epsilon_2(k_0|X|,k_0\delta)$ and $k_0|X| < k_4$, we have $\Pr\{E_3\} \to 0 \ (n \to \infty)$.

(2) If there is no or more than one $u_i \in A_U(v(x),j_a)$ such that $y_i(x) = j_v M_U + j_u$ and $(u_i,v(y),y) \in T_{UVY}(k_5\delta)$, an decoding error is declared. This event is classified into two cases.

(2-1) The first case: $(u(v(x),x),v(x),y) \notin T_{UVY}(k_5\delta)$. From Lemma 3, this event is included in

$$E_6 \overset{\text{def.}}{=} \{(u(v(x),x),v(x),x,y) \notin T_{UVY}(k_5\delta/|Y|)\}. \quad (150)$$

Here, we have

$$\Pr\{E_6\} \leq \Pr\{E_3 \cup E_6\} \quad (151)$$

$$= \Pr\{E_3\} + \Pr\{E_3' \cap E_6\}. \quad (152)$$

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

$$\Pr\{E_3' \cap E_6\} \leq \sum_{u,v,x} \Pr_{UVX}(u,v,x) \Pr\{(u,v,x,Y^n) \notin T_{UVX}(k_5\delta/|Y|)|u,v,x\} \quad (153)$$

$$\to 0 \ (n \to \infty). \quad (154)$$

Furthermore, $\lim_{n \to \infty} \Pr\{E_1\} = 0$ by setting $k_7$ and $k_5$ as $k_7 < k_5/|Y|$.

(2-2) The second case: There exists $u' \in A_U(v(x),j_a)$, $u' \neq u(v(x),x)$ such that $(u',v(x),y) \in T_{UVY}(k_5\delta)$. This event is denoted as

$$E_7 \overset{\text{def.}}{=} \left\{ \bigcup_{u \in A_U(v(x),j_a), \ u \neq u(v(x),x)} \{(u,v(x),y) \in T_{UVY}(k_5\delta)\} \right\}, \quad (155)$$

Let $i_u(j,k)$ be the index $i$ of $k$-th $u_i$, which belongs to $A_U(v(x),j)$. Here, we have

$$\Pr\{E_7\} \leq \Pr\{E_4 \cup E_7\} \quad (156)$$

29
\[ \Pr\{E_4\} + \Pr\{E_6^c \cap E_7\}. \] (157)

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

\[ \Pr\{E_6^c \cap E_7\} \]
\[ \leq \sum_{(v, y) \in T_{VY}(k \delta)} \Pr\left\{U_{u,v,k}^n, v, y \in T_{UVY}(k \delta)\right\} \]
\[ \leq \sum_{(v, y) \in T_{VY}(k \delta)} P_{VY}(v, y) \exp\{-n(I(Y; U|V) - \epsilon_2)\} \]
\[ (\because: \text{Lemma 5}) \]
\[ \leq L_U \exp\{-n(I(Y; U|V) - \epsilon_2)\} \]
\[ (158) \]
\[ (159) \]
\[ (160) \]

By setting \(L_U, k_4\) and \(k_5\) as

\[ L_U \leq \exp\{n(I(Y; U|V) - l_2 \gamma)\}, \quad l_2 > 0, \]
\[ l_2 \gamma > \epsilon_2 = \epsilon_2(k_4 \delta, k_5 \delta) \text{ and } k_4 < k_5, \text{ we have } \Pr\{E_7\} \to 0 \text{ (} n \to \infty). \]

Error evaluation: \(\hat{\varphi}(2)\)

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

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

to vanish the decoding errors.

Rate evaluation: \(\varphi(1)\)

The encoder sends the indexes of the bin using

\[ \frac{1}{n} \log N_H N_U \]
\[ = \frac{1}{n} \log M_V M_U \]
\[ \geq I(X; V) + m_1 \gamma - I(Y; V) + l_1 \gamma + I(X; U|V) + m_2 \gamma - I(Y; U|V) + l_2 \gamma \]
\[ \geq I(XY; V) - I(Y; V) + I(XY; U|V) - I(Y; U|V) + (m_1 + m_2 + l_1 + l_2) \gamma \]
\[ (\because: UV \to X \to Y) \]
\[ = I(X; VY) + I(X; U|VY) + (m_1 + m_2 + l_1 + l_2) \gamma \]
\[ = I(X; UV|Y) + (m_1 + m_2 + l_1 + l_2) \gamma \]
\[ (165) \]
\[ (166) \]
\[ (167) \]

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

\[ \frac{1}{n} \log N_H N_U \geq I(X; UV|Y). \]
\[ (168) \]
Rate evaluation: $\varphi^n(0)$

The encoder sends the indexes of the bin using

$$\frac{1}{n} \log N_V = \frac{1}{n} \log \frac{M_V}{L_V}$$

$$\geq I(X; V) + m_1 \gamma - I(Y; V) + l_1 \gamma$$

$$= I(XY; V) - I(Y; V) + (m_1 + l_1) \gamma$$

$$(: V \rightarrow X \rightarrow Y)$$

$$= I(X; V|Y) + (m_1 + l_1) \gamma$$

(170)

(171)

(172)

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

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

(173)

Rate evaluation: $\varphi^n(2)$

The encoder sends the indexes of the bin using

$$\frac{1}{n} \log N_W = \frac{1}{n} \log \frac{M_W}{L_W}$$

$$\geq I(X; W|V) + m_3 \gamma - I(Y; W|V) + l_3 \gamma$$

$$= I(XY; W|V) - I(Y; W|V) + (m_3 + l_3) \gamma$$

$$(: VW \rightarrow X \rightarrow Y)$$

$$= I(X; W|VY) + (m_3 + l_3) \gamma$$

(175)

(176)

(177)

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

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

(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 $A \rightarrow D \rightarrow B$, $B \rightarrow D \rightarrow C$ or $C \rightarrow D \rightarrow A$ form Markov chains. Then,

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

(179)

Remark. This lemma is an extended version of Massay’s lemma.
Proof. (1) If $A \rightarrow D \rightarrow B$ forms a Markov chain,

$$I(A; B|D) = 0 \leq I(A; B|CD).$$

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

$$I(A; B|CD) = I(A; BC|D) - I(A; C|D) \geq I(A; B|D)$$

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

$$I(A; B|CD) = I(A; BC|D) - I(A; C|D) \geq I(A; B|D)$$

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

Proof.

Let a sequence $\{ (\varphi^n_0, \varphi^n_1, \varphi^n_2, \hat{\varphi}^n_1, \hat{\varphi}^n_2) \}_{n=1}^\infty$ of FR codes be given to satisfy the conditions of Definitions 4 and 5. From Definition 5, for $\delta > 0$ there exists an integer $n_1(\delta)$, and then for all $n \geq n_1(\delta)$, we can obtain

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

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

$$R_1 + \delta$$

(188)

$$\geq \log M_n^{(1)}$$

(189)

$$\geq H(A_n^{(1)})$$

(190)

$$\geq H(A_n^{(1)}|Y^n)$$

(191)

$$= I(X^n; A_n^{(1)}|Y^n) \quad (\because A_n^{(1)} = \varphi^n_1(X^n))$$

(192)

$$= H(X^n|Y^n) - H(X^n|A_n^{(1)}Y^n)$$

(193)

$$= \sum_{k=1}^n \left\{ H(X_k|Y_k) - H(X_k|A_n^{(1)}X^{k-1}Y^n) \right\}$$

(194)

$$= \sum_{k=1}^n I(X_k; A_n^{(1)}X^{k-1}Y^n_k|Y_1Y^n_{k-1})$$

(195)

$$\geq \sum_{k=1}^n I(X_k; A_n^{(1)}X^{k-1}Y^n_{k+1}|Y_k)$$
Let us define the random variables \( U_k = A_n^{(1)} X^{k-1} Y^n_{k+1} \). With these definitions, we have the Markov structure \( U_k \rightarrow X_k \rightarrow Y_k \) because

\[
I(Y_k; U_k | X_k) = I(Y_k; A_n^{(1)} X^{k-1} Y^n_{k+1} | X_k) \\
\leq I(Y_k; A_n^{(1)} X^{k-1} Y^n_{k+1} Y^{k-1} Y^n_{k+1} | X_k) \\
= I(Y_k; X^{k-1} Y^n_{k+1} Y^{k-1} Y^n_{k+1} | X_k) \quad (\because A_n^{(1)} = \varphi_n^{(1)}(X^n)) \\
\leq I(X_k Y_k; X^{k-1} Y^n_{k+1} Y^{k-1} Y^n_{k+1}) \\
= 0.
\] (196)

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 \( U \rightarrow X \rightarrow Y \) still holds, and we have

\[
R_1 + \delta \geq I(X; U | Y). 
\] (201)

Since \( \delta > 0 \) is arbitrary, we obtain

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

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

\[
n(R_0 + \delta) \\
\geq \log M_n^{(0)} \\
\geq H(A_n^{(0)}) \\
\geq H(A_n^{(0)} | A_n^{(1)} X^n) \\
= I(Y^n; A_n^{(0)} | A_n^{(1)} X^n) \quad (\because A_n^{(0)} = \varphi_n^{(0)}(A_n^{(1)}, Y^n)) \\
= H(Y^n | A_n^{(0)} X^n) - H(Y^n | A_n^{(0)} A_n^{(1)} X^n) \\
= H(Y^n | X^n) - H(Y^n | A_n^{(0)} A_n^{(1)} X^n) \quad (\because A_n^{(1)} = \varphi_n^{(1)}(X^n)) \\
= \sum_{k=1}^{n} \left\{ H(Y_k; K_k) - H(Y_k; A_n^{(1)} A_n^{(0)} Y^n_{k+1} X^n) \right\} \\
= \sum_{k=1}^{n} I(Y_k; A_n^{(1)} A_n^{(0)} X^{k-1} Y^n_{k+1} Y^n_{k+1} | X_k) \\
\geq \sum_{k=1}^{n} I(Y_k; A_n^{(1)} A_n^{(0)} X^{k-1} Y^n_{k+1} | X_k) \\
\geq \sum_{k=1}^{n} I(Y_k; A_n^{(0)} | A_n^{(1)} X^n Y^n_{k+1})
\]

(208)
Let us define the random variables $V_k = A^{(1)}_k A^{(0)}_n X^{k-1} Y^n_{k+1}$. With these definitions, we have the Markov structure $V_k \rightarrow U_k Y_k \rightarrow X_k$ because

$$I(X_k; V_k | U_k Y_k) = I(X_k; A^{(0)}_n | A^{(1)}_k X^{k-1} Y^n_k)$$

$$\leq I(X_k; A^{(0)}_n Y^{k-1} | A^{(1)}_k X^{k-1} Y^n_k)$$

$$= I(X_k; Y^{k-1} | A^{(1)}_k X^{k-1} Y^n_k) \quad (\because A^{(0)}_n = \varphi^{(0)}_n(A^{(1)}_k, Y^n))$$

$$\leq I(A^{(1)}_n X_k; Y^{k-1} | X^{k-1} Y^n_k)$$

$$= I(X^n_k; Y^{k-1} | X^{k-1} Y^n_k) \quad (\because A^{(1)}_n = \varphi^{(1)}_n(X^n))$$

$$\leq I(X^n_k Y^n_k; X^{k-1} Y^{k-1})$$

$$= 0.$$  

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 $V \rightarrow U Y \rightarrow X$ and

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

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

$$n(R_2 + \delta) \geq \log M^{(2)}_n$$

$$\geq H(A^{(2)}_n)$$

$$\geq H(A^{(2)}_n | A^{(1)}_n A^{(0)}_n Y^n)$$

$$= I(X^n; A^{(2)}_n | A^{(1)}_n A^{(0)}_n Y^n) \quad (\because A^{(2)}_n = \varphi^{(2)}_n(A^{(0)}_n, X^n))$$

$$= \sum_{k=1}^n I(X_k; A^{(2)}_n | A^{(1)}_n A^{(0)}_n X^{k-1} Y^n)$$

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

$$I(A; C | D) = I(X_k; Y^{k-1} | A^{(1)}_n A^{(0)}_n X^{k-1} Y^n_k)$$

$$\leq I(X_k; A^{(0)}_n Y^{k-1} | A^{(1)}_n X^{k-1} Y^n_k)$$

$$= I(X_k; Y^{k-1} | A^{(1)}_n X^{k-1} Y^n_k) \quad (\because A^{(0)}_n = \varphi^{(0)}_n(A^{(1)}_n, Y^n))$$

$$\leq I(A^{(1)}_n X_k; Y^{k-1} | X^{k-1} Y^n_k)$$

34
\begin{align*}
  &= I(A_n^{(1)}X^n_k;Y^{k-1}|X^{k-1}Y^n_k) \\
  &= I(X^n_k;Y^{k-1}|X^{k-1}Y^n_k) \quad (\because A_n^{(1)} = \varphi_n^{(1)}(X^n)) \quad (230) \\
  &= I(X^n_k Y^n_k;X^{k-1}Y^{k-1}) \quad (231) \\
  &= 0. \quad (232)
\end{align*}

From Lemma 8, the bound (224) becomes

\begin{align*}
  n(R_2 + \delta) &\geq \sum_{k=1}^{n} I(A;B|CD) \quad (233) \\
  &\geq \sum_{k=1}^{n} I(A;B|D) \quad (234) \\
  &= \sum_{k=1}^{n} I(X_k;A_n^{(2)}|A_n^{(1)}A^{(0)}_nX^{k-1}Y^n_k). \quad (235)
\end{align*}

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

\begin{align*}
  I(U_kY_k;W_k|V_kX_k) &= I(Y_k;A_n^{(2)}|A_n^{(1)}A^{(0)}_nX^{k}Y^n_{k+1}) \quad (236) \\
  &\leq I(Y_k;X^n_{k+1}|A_n^{(1)}A^{(0)}_nX^{k}Y^n_{k+1}) \quad (237) \\
  &= I(Y_k;X^n_{k+1}|A_n^{(1)}A^{(0)}_nX^{k}Y^n_{k+1}) \quad (\because A_n^{(2)} = \varphi_n^{(2)}(A_n^{(1)},X^n)) \quad (238) \\
  &\leq I(A^{(0)}_nY_k;X^n_{k+1}|A_n^{(1)}A^{(0)}_nX^{k}Y^n_{k+1}) \quad (239) \\
  &\leq I(A^{(0)}_nY^{k};X^n_{k+1}|A_n^{(1)}X^{k}Y^n_{k+1}) \quad (240) \\
  &= I(Y^{k};X^n_{k+1}|A_n^{(1)}X^{k}Y^n_{k+1}) \quad (\because A^{(0)}_n = \varphi_n^{(0)}(A_n^{(1)},Y^n)) \quad (241) \\
  &\leq I(Y^{k};A^{(1)}_nX^n_{k+1}|X^{k}Y^n_{k+1}) \quad (242) \\
  &= I(Y^{k};X^n_{k+1}|X^{k}Y^n_{k+1}) \quad (\because A^{(1)}_n = \varphi_n^{(1)}(X^n)) \quad (243) \\
  &\leq I(X^{k}Y^{k};X^n_{k+1}Y^n_{k+1}) \quad (244) \\
  &= 0. \quad (245)
\end{align*}

Substituting \( V_k \) and \( W_k \) into (235), we have

\begin{align*}
  R_2 &\geq \sum_{k=1}^{n} I(X_k;W_k|V_kY_k)
\end{align*}

In the same way as the above discussion, we have the Markov structure \( W \rightarrow VX \rightarrow Y \) and

\begin{align*}
  R_2 &\geq I(X;W|VY).
\end{align*}

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
We note that $\hat{n}$ where choose the functions $\hat{X}$ be a random variable selected to minimize the average distortion between $X$ and $Q$ given $W$. The $k$th term $Y_{\hat{k}}Y_k$ contains a concatenation of sequences. From the definitions, we have

$$\Delta^n(X^n, X(1)n) = \frac{1}{n} \sum_{k=1}^{n} \Delta(X_k, \hat{X}(1)k),$$

$$\Delta^n(X^n, X(2)n) = \frac{1}{n} \sum_{k=1}^{n} \Delta(X_k, \hat{X}(2)k),$$

Now, we denote by $\hat{X}(1)k$ and $\hat{X}(2)k$ be a random variable selected to minimize the average distortion between $X_k$ and $\hat{X}(1)k$ given $U_kY_k$, and let $Y_{\hat{k}-1}(W_k, Y_k)$ be a random variable selected to minimize the average distortion between $X_k$ and $\hat{X}(2)k$ given $W_kY_k$, namely

$$Y_{\hat{k}-1}(U_k, Y_k) \overset{\text{def.}}{=} \arg\min_{Y^{k-1} \in \mathcal{X}^{k-1}} \sum_{X_k \in \mathcal{X}} Q_k^{(1)}(X_k|U_kY_k) \Delta(X_k, \hat{X}(1)k(A_n^{(1)}, Y^n)),$$

$$Y_{\hat{k}-1}(W_k, Y_k) \overset{\text{def.}}{=} \arg\min_{Y^{k-1} \in \mathcal{X}^{k-1}} \sum_{X_k \in \mathcal{X}} Q_k^{(2)}(X_k|W_kY_k) \Delta(X_k, \hat{X}(2)k(A_n^{(2)}, Y^n)),$$

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

$$Q_k^{(1)}(x|uY) = Q_k^{(1)}(x|\hat{A}, y \ast \hat{y}_{k+1})$$

$$= \sum_{(x_{n+1}^{k-1}y_{n+1}) \in \mathcal{X}^{n-k} \times \mathcal{Y}^{k-1}} P_{XY}(x^n, y^n)$$

$$= \sum_{(x_{n+1}^{k-1}y_{n+1}) \in \mathcal{X}^{n-k} \times \mathcal{Y}^{k-1}, \phi_n^{(1)}(x^n) = \hat{A}, x = x_{k-1}y_{k}}, y_{n+1} = y_{k+1}$$

$$\text{for } u = \hat{A}g_{k+1}, x \in \mathcal{X}, y \in \mathcal{Y},$$

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

$$\phi(1)k(U_k, Y_k) \overset{\text{def.}}{=} \hat{\phi}(1)k(A_n^{(1)}, Y_{\hat{k}-1}(U_k, Y_k) \ast Y_{\hat{k}}) = \hat{X}(1)k,$$

$$\phi(2)k(W_k, Y_k) \overset{\text{def.}}{=} \hat{\phi}(2)k(A_n^{(2)}, Y_{\hat{k}-1}(W_k, Y_k) \ast Y_{\hat{k}}) = \hat{X}(2)k,$$

$$\phi(1)(U, Y) \overset{\text{def.}}{=} \phi(1)(1)(U, Y),$$

$$\phi(2)(W, Y) \overset{\text{def.}}{=} \phi(2)(1)(W, Y).$$
\[ \phi_{(2)}(W, Y) \overset{\text{def.}}{=} \phi_{(2) J}(W_J, Y). \] (249)

This implies that

\[
D_1 + \gamma \geq \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{X}_{(1)k}) \right] \tag{250}
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{\varphi}_{(1)k}(A^{(1)}_n, Y^n)) \right] \tag{251}
\]

\[
\geq \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \hat{\varphi}_{(1)k}(A^{(1)}_n, Y^{k-1}(U_k, Y_k) \ast Y_k^{n})) \right] \tag{252}
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X_k, \phi_{(1)k}(U_k, Y_k)) \right] \tag{253}
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} E \left[ \Delta(X, \phi_{(1)J}(U_J, Y)) | J = k \right] \tag{254}
\]

\[
D_2 + \gamma \geq E \left[ \Delta(X, \phi_{(2)}(W, Y)) \right]. \tag{255}
\]

Since \( \gamma > 0 \) is arbitrary, we get

\[
D_1 \geq E \left[ \Delta(X, \phi_{(1)}(U, Y)) \right], \tag{257}
\]

\[
D_2 \geq E \left[ \Delta(X, \phi_{(2)}(W, Y)) \right]. \tag{258}
\]

It remains to establish that the cardinality bounds on \(|U|, |V|, |W|\) specified in Theorem 2 does not affect the region \( R_f(X, Y | D_1, D_2) \). 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. \( \square \)

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 \( R_f(X, Y | D_1, D_2) \). Fix arbitrary \( \gamma, \delta > 0 \).

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

1. Randomly generate \( M_U \) independent codewords \( \mathcal{A}_U = \{ u_i \}_{i=1}^{M_U} \), \( u_i \in U^n \), 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 \( u \in \mathcal{A}_U \) assigned to bin \( j \) \((j \in \mathcal{I}_{N_U})\).

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

In the same way as \( \varphi_{(1)}^n \), generate \( \mathcal{A}_V = \{ v_i \}_{i=1}^{M_V} \) according to the distribution
$P_V$, and divide $A_V$ into $N_V$ bins. Let $A_V(j)$ denote the set of elements $v \in A_V$ assigned to bin $j$ ($j \in I_{N_V}$). Note that $A_V$ does not depend on any $u \in A_U$.

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

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

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

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

$$
\varphi^n_{(1)}(x) = j_u, \quad u(x) \in A_U(j_u).
$$

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

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

$$
\hat{x}_{(1)k} = \phi_{(1)}(\hat{u}(y), y_k) \quad (k \in I_n).
$$

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

1. For a given $y \in Y^n$, the encoder seeks a vector $v \in A_V$ such that $(v, y) \in T_{VY}(k_3 \delta)$, $k_3 > 0$. If there is more than one such vector in $A_V$, the first one is chosen. If there is no such vector in $A_V$, a default vector is chosen, say $v_1$, and an error is declared. Denote the selected vector by $v(y)$.
2. The value assigned to the encoder $\varphi^n_{(0)}(\cdot, \cdot)$ is the bin number to which $v(y)$ belongs, that is,

$$
\varphi^n_{(0)}(j_v, x) = j_v, \quad v(y) \in A_V(j_v).
$$

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

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

1. The encoder $\hat{\varphi}^n_{(2)}$ has access to the index $j_v$ and the vector $x \in X^n$.
2. It seeks a unique vector $v \in A_V(j_v)$ such that $(u(x), v, x) \in T_{UX}(k_4 \delta)$, $k_4 > 0$. Denote this vector $\hat{v}(u(x), x)$. If there is no such vector or more than
one such vector \( v \in A_{V}(j_{w}) \), an arbitrary \( \hat{v} \) is chosen, and an error is declared.

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

1. For a given \( x \in X^{n} \) and \( \hat{v} = \hat{v}(u(x), x) \), the encoder seeks a vector \( w_{t} \in A_{W}(\hat{v}) \) such that \( (\hat{v}, w_{t}, x) \in T_{VWY}(k_{5} \delta), k_{5} > 0 \). If there is more than one such vector in \( A_{W}(\hat{v}) \), the first one is chosen. If there is a such vector in \( A_{W}(\hat{v}) \), a default vector is chosen, and an error is declared. Denote the selected vector by \( w(\hat{v}, x) \).

2. The value assigned to the encoder \( \varphi_{(2)}^{n}(\cdot, \cdot) \) is the bin number to which \( w(\hat{v}, x) \) belong, that is,

\[
\varphi_{(2)}^{n}(j_{w}, x) = j_{w}, \quad w(\hat{v}, x) \in A_{W}(\hat{v}, j_{w}).
\]

**Decoding:** \( \hat{\varphi}_{(2)}^{n} \)

1. The decoder has access to the index \( j_{w} \) and the vector \( y \in Y^{n} \).

2. It seeks a unique vector \( w \in A_{W}(v(y), j_{w}) \) such that \( (v(y), w, y) \in T_{VWY}(k_{5} \delta), k_{5} > 0 \). Denote this vector \( w(\hat{v}(v(y), y)) \). If there is no such vector or more than one such vector \( w \in A_{W}(v(y), j_{w}) \), an arbitrary \( \hat{w} \) is chosen, and an error is declared.

3. The reconstruction vector \( \hat{x}_{(2)} = (\hat{x}_{(2)1}, \hat{x}_{(2)2}, \cdots, \hat{x}_{(2)n}) \) is given by

\[
\hat{x}_{(2)k} = \phi_{(2)}(\hat{w}(v(y), y), y_{k}) \quad (k \in \mathcal{I}_{n}).
\]

**Distortion evaluation:** \( \hat{\varphi}_{(1)}^{n} \)

Almost the same way as Distortion evaluation: \( \varphi_{(1)}^{n} \) of Theorem 1. If probabilities of encoding/decoding errors and \( (u(x), x, y) \notin T_{UXY}(k_{7} \delta) \) vanish as \( n \rightarrow \infty \), we obtain

\[
\limsup_{n \rightarrow \infty} E \left[ \Delta_{X}^{n}(X^{n}, \hat{X}_{(1)}^{n}) \right] \leq D_{1}.
\]

**Distortion evaluation:** \( \hat{\varphi}_{(2)}^{n} \)

In a similar manner to \( \varphi_{(1)}^{n} \), if the probabilities of encoding/decoding errors and \( (w(\hat{v}, x), x, y) \notin T_{WXY}(k_{5} \delta) \) vanish as \( n \rightarrow \infty \), we obtain

\[
\limsup_{n \rightarrow \infty} E \left[ \Delta_{Y}^{n}(X^{n}, \hat{X}_{(2)}^{n}) \right] \leq D_{2}.
\]

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

If there is no \( u_{i} \in A_{U} \) such that \( (u_{i}, x) \in T_{UX}(k_{1} \delta) \), an encoding error occurs. In the same way as Error evaluation: \( \varphi_{(1)}^{n} \) of Theorem 1, the probability of the event vanishes as \( n \rightarrow \infty \) by setting

\[
M_{U} \geq \exp\{n(I(X; U) + m_{1} \gamma)\}, \quad m_{1} > 0.
\]
Error evaluation: $\tilde{\varphi}^n_{(1)}$

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

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

(2) The second case: There exists $u' \in A_U(j_u)$ such that $u' \neq u(x)$ and $(u', y) \in T_{UY}(k_3\delta)$. In the same way as Error evaluation: $\tilde{\varphi}^n_{(1)}$ of Theorem 1, the probability of the event vanishes as $n \rightarrow \infty$ by setting

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

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

If there is no $v_i \in A_V$ such that $(v_i, y) \in T_{VY}(k_3\delta)$, a decoding error occurs. In the same way as $\tilde{\varphi}^n_{(1)}$, the probability of the event vanishes as $n \rightarrow \infty$ by setting

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

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

(1) If there is no or more than one $v_i \in A_V(j_v)$ such that $\varphi^n_{(0)}(j_v, y) = j_v$ and $(v_i, x) \in T_{UVX}(k_4\delta)$, an encoding error is declared. This event is classified into two cases.

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

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

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

(2) If there is no $w_i \in A_W(v(y))$ such that $(v(y), w_i, x) \in T_{VWX}(k_5\delta)$, 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 $n \rightarrow \infty$ by setting $M_W$ as

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

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

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

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

40
to vanish the decoding errors. Through the discussion, we can also obtain \( \Pr\{W^n(\tilde{V}^n(X^n),X^n) \in T_{WXY}(k_8 \delta) \} \to 0 \ (n \to \infty) \) by setting \( k_8 \) and \( k_6 \) as \( k_8 < k_6/|X| \), which is necessary to show Eq. (265).

**Rate evaluation: \( \varphi^n_{(1)} \)**
The encoder sends the indexes of the bin using
\[
\frac{1}{n} \log N_U = \frac{1}{n} \log \frac{M_U}{L_U} \geq I(X;U) + m_1 \gamma - I(Y;U) + l_1 \gamma \tag{273}
\]
\[
= I(XY;U) - I(Y;U) + (m_1 + l_1) \gamma \tag{274}
\]
\[
\therefore U \to X \to Y \tag{275}
\]
bits per letter. Since \( \gamma > 0 \) is arbitrary, we obtain the coding rate as
\[
\frac{1}{n} \log N_U \geq I(X;U|Y). \tag{276}
\]

**Rate evaluation: \( \varphi^n_{(0)} \)**
The encoder sends the indexes of the bin using
\[
\frac{1}{n} \log N_V = \frac{1}{n} \log \frac{M_V}{L_V} \geq I(Y;V) + m_2 \gamma - I(U;V) + l_2 \gamma \tag{278}
\]
\[
= I(UXY;V) - I(U;V) + (m_2 + l_2) \gamma \tag{279}
\]
\[
\therefore V \to Y \to UX \tag{280}
\]
bits per letter. Since \( \gamma > 0 \) is arbitrary, we obtain the coding rate as
\[
\frac{1}{n} \log N_V \geq I(Y;V|UX). \tag{281}
\]

**Rate evaluation: \( \varphi^n_{(2)} \)**
The encoder sends the indexes of the bin using
\[
\frac{1}{n} \log N_W = \frac{1}{n} \log \frac{M_W}{L_W} \geq I(X;W|V) + m_3 \gamma - I(Y;W|V) + l_3 \gamma \tag{283}
\]
\[
= I(XY;W|V) - I(Y;W|V) + (m_3 + l_3) \gamma \tag{284}
\]
\[
\therefore W \to VX \to Y \tag{285}
\]
bits per letter. Since \( \gamma > 0 \) is arbitrary, we obtain the coding rate as
\[
\frac{1}{n} \log N_W \geq I(X;W|YY). \tag{286}
\]

This completes the proof of the direct part of Theorem 2. \( \square \)
Theorem 3: converse

Proof. Let a sequence \( \left\{ (\varphi_n^{(01)}, \varphi_n^{(02)}, \varphi_n^{(1)}, \varphi_n^{(2)}), \hat{\varphi}_n^{(1)}, \hat{\varphi}_n^{(2)} \right\}_{n=1}^{\infty} \) 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

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

\[
n(R_1 + \delta) \geq \sum_{k=1}^{n} I(X_k; A_n^{(1)}A_n^{(02)}X_{k-1}Y_{k+1}^{n}|Y_k) \geq \sum_{k=1}^{n} I(X_k; A_n^{(1)}A_n^{(02)}X_{k-1}Y_{k+1}^{n}|Y_k)
\] (288)

(289)

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

\[
n(R_1 + \delta) \geq \sum_{k=1}^{n} I(X_k; U_k V^{(2)}_k|Y_k).
\]

Introducing a random variable \( J \), the Markov condition \( U \rightarrow X \rightarrow Y \) still holds, and we have

\[
R_1 + \delta \geq I(X; UV^{(2)}|Y).
\] (290)

Since \( \delta > 0 \) is arbitrary, we obtain

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

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

\[
n(R_{01} + \delta) \geq \log M_n^{(01)} \geq H(A_n^{(01)}) \geq H(A_n^{(01)}|A_n^{(1)}A_n^{(02)}X^n) \geq I(Y^n, A_n^{(01)}|A_n^{(1)}A_n^{(02)}X^n) \quad (\because A_n^{(0)} = \varphi_n^{(01)}(A_n^{(1)}, Y^n))
\] (291)

(292)

(293)

(294)
From Lemma 8, the bound (295) becomes

\[ n(R_0 + \delta) \geq \sum_{k=1}^{n} I(Y_k; A_n^{(01)}A_n^{(02)}X^nY_{k+1}^{n}) \]  

(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 \( D = A_n^{(02)}X^nY_{k+1}^{n} \). Then, we have

\[
I(A; C|D) \\
= I(Y_k; X_{k+1}^{n}A_n^{(01)}A_n^{(02)}X^nY_{k+1}^{n}) \\
\leq I(A_n^{(02)}Y_k; X_{k+1}^{n}A_n^{(01)}X^nY_{k+1}^{n}) \\
\leq I(A_n^{(02)}Y_k, X_{k+1}^{n}|A_n^{(01)}X^nY_{k+1}^{n}) \\
= I(Y_k; X_{k+1}^{n}|A_n^{(01)}X^nY_{k+1}^{n}) \quad (\because A_n^{(02)} = \varphi_n^{(02)}(A_n^{(01)}, Y^n)) \\
\leq I(Y_k; A_n^{(01)}X_{k+1}^{n}X^nY_{k+1}^{n}) \quad (\because A_n^{(01)} = \varphi_n^{(1)}(X^n)) \\
\leq I(X^nY_{k+1}^{n}X_{k+1}^{n}Y_{k+1}^{n}) = 0. \\
\]

(296)

(297)

(298)

(299)

(300)

(301)

(302)

(303)

From Lemma 8, the bound (295) becomes

\[ n(R_0 + \delta) \geq \sum_{k=1}^{n} I(Y_k; A_n^{(01)}A_n^{(02)}X^nY_{k+1}^{n}) \]  

(304)

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

\[
I(X_k; V_k^{(1)}|U_kV_k^{(2)}Y_k) \\
= I(X_k; A_n^{(01)}A_n^{(02)}X^nY_{k+1}^{n}) \\
\leq I(X_k; A_n^{(01)}A_n^{(02)}X^{n-1}Y_k^n) \\
\leq I(X_k; A_n^{(01)}A_n^{(02)}Y_{k-1}^{n}|A_n^{(1)}X^{n-1}Y_k^n) \\
= I(X_k; Y_{k-1}^{n}|A_n^{(1)}X^{n-1}Y_k^n) \quad (\because A_n^{(01)} = \varphi_n^{(01)}(A_n^{(1)}, Y^n), A_n^{(02)} = \varphi_n^{(02)}(A_n^{(1)}, Y^n)) \\
\leq I(A_n^{(1)}X_k; Y_{k-1}^{n}|X^{n-1}Y_k^n) \quad (310)
\leq I(A_n^{(1)}X_k; Y_{k-1}^{n}|X^{n-1}Y_k^n) \quad (311)
= I(X_k^n; Y_{k-1}^{n}|X^{n-1}Y_k^n) \quad (\because A_n^{(1)} = \varphi_n^{(1)}(X^n)) \\
\leq I(X_k^nY_k^n; X^{n-1}Y_{k-1}^{n}) \quad (312)
= 0. \\
\]

(305)

(306)

(307)

(308)

(309)

(310)

(311)

(312)

(313)

(314)

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

\[ n(R_0 + \delta) \geq \sum_{k=1}^{n} I(Y_k; V_k^{(1)}|U_kV_k^{(2)}X_k). \]
In the same way as the above discussion, we have the Markov structure $V^{(1)} \rightarrow UV^{(2)}Y \rightarrow X$ and

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

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

$$n(R_2 + \delta) \geq \log M^{(2)}$$

$$\geq H(A^{(2)})$$

$$\geq H (A_n^{(2)}|A_n^{(1)} A_n^{(01)} A_n^{(02)} Y^n)$$

$$= I(X^n; A_n^{(2)}|A_n^{(1)} A_n^{(01)} A_n^{(02)} Y^n) \quad (\because A_n^{(2)} = \phi_n^{(2)}(A_n^{(01)}, X^n))$$

$$= \sum_{k=1}^{n} I(X_k; A_n^{(2)}|A_n^{(1)} A_n^{(01)} A_n^{(02)} X_k^{k-1} Y^n)$$

(319)

Here, we would like to apply Lemma 8 again. Let $A = X_k, B = A_n^{(2)}, C = Y_k^{k-1}$ and $D = A_n^{(01)} A_n^{(02)} X_k^{k-1} Y_k^n$. 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

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

(320)

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

$$I(U_k Y_k; W_k|V_k^{(1)} V_k^{(2)} X_k)$$

$$= I(Y_k; A_n^{(2)}|A_n^{(1)} A_n^{(01)} A_n^{(02)} X_k^{k-1} Y_k^n)$$

$$\leq I(Y_k; A_n^{(2)} X_k^{n+1}|A_n^{(1)} A_n^{(01)} A_n^{(02)} X_k^{k-1} Y_k^{n+1})$$

$$= I(Y_k; A_n^{(01)} A_n^{(02)} Y_k; X_k^{n+1}|A_n^{(1)} X_k^{k-1} Y_k^{n+1})$$

$$\leq I(A_n^{(01)} A_n^{(02)} Y_k; X_k^{n+1}|A_n^{(1)} X_k^{k-1} Y_k^{n+1})$$

$$= I(Y_k; X_k^{n+1}|A_n^{(1)} X_k^{k-1} Y_k^n) \quad (\because A_n^{(1)} = \phi_n^{(1)}(X^n) \text{ for } i = 01, 02)$$

$$\leq I(Y_k; X_k^{n+1}|X_k^{k-1} Y_k^n)$$

$$= I(Y_k; X_k^{n+1}) \quad (\because A_n^{(1)} = \phi_n^{(1)}(X^n))$$

(328)

$$\leq I(X_k Y_k; X_k^{n+1})$$

(329)

$$= 0.$$  

(330)

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

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

44
In the same way as the above discussion, we have the Markov structure \( W \rightarrow V^{(1)} V^{(2)} X \rightarrow U Y \) 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 \( |U|, |V^{(1)}|, |V^{(2)}| \) and \( |W| \) 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


