Coding theorems for correlated sources with cooperative encoding

Akisato Kimura

Supervisor: Prof. Tomohiko Uyematsu

Department of Communications and Integrated Systems
Graduate School of Science and Engineering
Tokyo Institute of Technology

Abstract

This thesis deals with multiterminal source coding problems for a general framework of coding systems, called *coding systems with cooperation*, where there are some linkages among encoders and decoders. Especially, the main focus of this thesis is encoder cooperation. Two types of coding systems are investigated that incorporate encoder cooperation: the Slepian-Wolf coding system with linkages (called the SWL system) and the complementary delivery coding system.

The SWL system involves some mutual linkages between two encoders of the coding system investigated by Slepian and Wolf (called the SW system) that involves two separate encoders and one common decoder. Especially, some special cases are considered, where the coding rate for the mutual linkage between two encoders is negligibly small. The main results in this thesis shows that the achievable rate region of the SWL system equals that of the SW system when considering fixed-length coding, while weak variable-length coding makes the achievable rate region of the SWL system larger than that of the SW system. This implies that encoder cooperation may improve the coding rate.

The complementary delivery coding system contrasts with the SW system in the sense of cooperation, which means that the complementary delivery coding system consists of a common encoder and separate decoders, while the SW system includes separate encoders and a common decoder. Especially, in the complementary delivery coding system, each decoder has access to some of encoded messages to enable the decoder to reproduce the other messages from a common codeword emitted from the common encoder. First, the minimum achievable rate for lossy coding is clarified, which implies that encoder cooperation may increase the coding rate. Next, universal coding schemes for lossless coding are proposed. Explicit constructions of universal lossless codes and the bounds of the error probabilities are clarified by using methods of types and the graph-theoretical analysis.

Contents

Chapter 1

Introduction

1.1 Background

Information theory, or more specifically *Shannon theory* was founded in 1948 by Claude Shannon in his seminal work "A Mathematical Theory of Communication" [53]. The central research issues of his work involve *source coding* which aims to establish the limits to possible data compression under certain criteria, and *channel coding* which aims to establish the limits to reliable data transmission over noisy channels. Shannon established in his paper [53] the most fundamental results for each research issue: Shannon's source coding theorem states that on average the number of bits necessary to represent the data generated via an uncertain event is given by its entropy. Shannon's channel coding theorem shows that the reliable communication over noisy channels is possible if on average the logarithm of the number of messages is below a certain threshold called the *channel capacity*.

The classical information theory assumes that input data is generated from a single data source, and sent via a communication channel from a single transmitter to a single receiver. On the other hand, *multiterminal information theory*, which is the basis of this thesis, extends the classical information theory to deal with multiple data sources, multiple communication channels, multiple transmitters and/or multiple receivers, as shown in Figure 1.1. Multiterminal information theory is directly associated with many communication schemes such as mobile communication (e.g. [77]), broadcasting via satellites (e.g. [72]), MIMO (e.g. [56, 21, 23]) or sensor networks (e.g. [48]). Also, relationships with other applications have also been reported such as scalable video coding (e.g. [19]), pattern recognition [65, 64] or information retrieval [57, 41]. The central research issues of the multiterminal information theory include *multiterminal source coding* and *multiterminal channel coding*, each of which corresponds to source coding and channel coding in the classical information theory. This thesis especially focuses on multiterminal source coding.

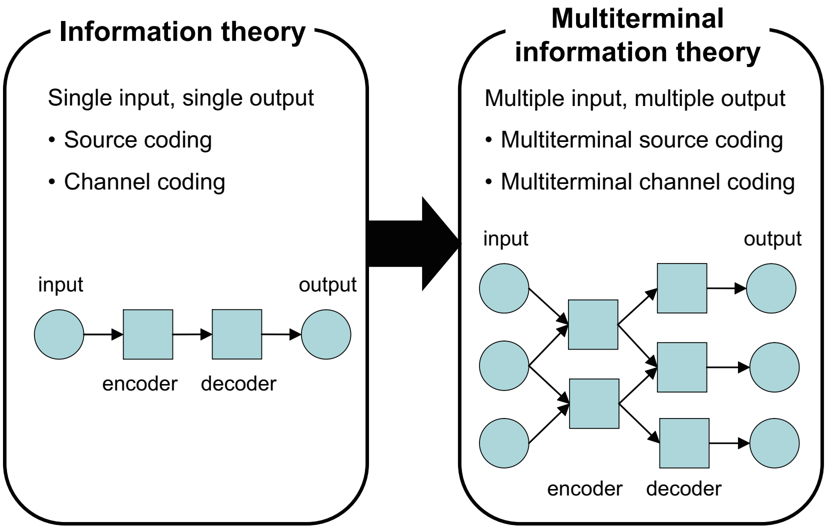

Figure 1.1: Multiterminal information theory

1.2 Related work

1.2.1 Lossless source coding

The work by Slepian and Wolf [54] is regarded as a substantial starting point of multiterminal source coding. In the coding system considered by Slepian and Wolf, called the *Slepian-Wolf coding system (SW system)*, two sequences of length n emitted from correlated information sources are separately encoded to nR_1 and nR_2 bit codewords, and sent to a single receiver which has to output original sequence pairs with small probability of error. They established the *achievable rate region*, namely the closure of the set which consists of the rate pair (R_1, R_2) such that the error probability of decoding can be made arbitrarily small by letting n to be large. After the original proof of their coding theorem, Cover [9] showed a simple proof based on the random coding argument called *bin coding*, and extended their result to stationary ergodic sources. Miyake and Kanaya [45] further extended the coding theorem to the class of non-ergodic or non-stationary sources called *general sources* by using the method of *information spectrum* developed by Han and Verdú [32, 30].

Many variations derived from the SW system have been also investigated. Wyner [71], Ahlswede and Körner [3] investigated the coding system where one of two sequences does not have to be reconstructed. Körner and Marton [40] considered another coding system, where codewords can be used not only to recover the original sequence but also as *side information* to recover other sequences. Wyner [71] and Sgarro [52] treated other types of coding systems where several codewords are used as side information to reconstruct some other original sequences. Csiszár, Körner [12], Han and Kobayashi [31] presented a unified treatment of a large class of coding systems that include all of the above mentioned coding systems. Ericson and Körner [20] studied the case not included in their unified framework, where one of two encoders can observe not only the sequence from its own sources but also the codeword generated from the other encoder. Oohama [47] investigated another variation of the SW system, where each encoder can observe the codeword generated from the other encoder.

1.2.2 Lossy source coding

Researches shown in Section 1.2.1 examined the minimum number of bits of codewords per data sample such that the probability of decoding errors converges to 0 as the number of data samples tends to infinity. On the other hand, in the theory of *lossy source coding*, or source coding with a *fidelity criterion*, we are interested in determining the minimum number of bits of codewords per data sample, subject to a constraint on the distortion in reconstructing the data from its codeword. Wyner and Ziv [73] first investigated a multiterminal lossy coding problem. They considered a special case of the SW system, where one of two sequences is directly transmitted to the receiver, and clarified the *rate-distortion function* (i.e. the minimum achievable rate for a given distortion criterion) for this coding system. Several variations and extensions have been also investigated. Iwata and Muramatsu [34] extended their results to general sources. Gastpar [24] investigated an extended coding system, where a decoder of the SW system has access to side information, and clarified the achievable rate-distortion region (i.e. the achievable rate region for given distortion criteria) in the case where two sources are conditionally independent given the side information source. Heegard, Berger [33] and Kaspi [35] clarified the rate-distortion function of another extended coding system, where the encoder does not know whether side information is available at the decoder or not. Yamamoto [74] proposed another type of coding systems called the *cascade and branching communication systems*, and clarified the achievable rate-distortion regions for several special cases. Gu and Effros [27, 28] extended the results by Yamamoto to consider side information messages only at a decoder. Yamamoto [75] presented a new coding system called the *triangular communication system*, which is an extended version of the simple cascade communication system.

The problem of deriving the explicit achievable rate region for the SW system still remains as one of the most famous open problems in multiterminal source coding. Berger and Tung [6, 58] first clarified inner and outer bounds of the achievable rate-distortion region for the SW system, in which distortion criteria are imposed on the reconstruction of both sequences. Several attempts have been made to improve inner and outer bounds [4, 63]. Berger and Yeung [7] clarified the achievable rate-distortion region for a special case of the coding problem investigated by Berger and Tung, where one of two messages is required to be recovered almost perfectly. Kaspi and Berger [38, 36] investigated a variation of the SW system, where one of two encoders can observe not only the sequence from its own sources but also the codeword generated from the other encoder.

Research issues that have been lively investigated in lossy source coding include *multiple description* and *successive refinement*:

The problem of multiple description was first posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv and Ozarow (see [17]), where a common encoder sends two descriptions (i.e. codewords) of a sequence to decoders, and one of the descriptions may or may not be lost before it gets to the decoder. Primary contributions to this problem have been presented in the beginning of 1980's [67, 69, 68]. Later on, El Gamal and Cover [17] obtained an inner bound of the achievable rate-distortion region for the multiple description, and showed that it is tight for deterministic distortion measures in which every decoder is required to recover a function of the sequence perfectly. Ozarow [50] clarified that the inner bound obtained by El Gamal and Cover is also tight for Gaussian sources with the square error distortion. Ahlswede [1], Zhang and Berger [78] showed that the inner bound of El Gamal and Cover is tight for the case of no excess rate for the joint description, namely the case that the sum of the coding rates for two descriptions is required to equal the rate-distortion function for the simple source coding (cf. [5]). Fu and Yeung [22] clarified that the inner bound of El Gamal and Cover is also tight if a decoder that receives one of two descriptions is required to recover a function of the message.

The problem of successive refinement was originally formulated by Koshelev [43], and later by Equitz and Cover [18, 19] as a special case of the multiple description problem. In the original setting of successive refinement, the encoder is assumed to operate in two stages: In the first stage the encoder transmits the first description of a message over a relatively rate-limited channel. In the second stage, the encoder transmits the secondary description of the message to provide a more accurate reproduction of the message. It is expected that every source is *successive refinable*, namely coding rates in two stages simultaneously lie on the rate-distortion function for the simple source coding. Koshelev [43], Equitz and Cover [18, 19] showed necessary and sufficient conditions for a source to be successive refinable. Rimoldi [51] first characterized the achievable rate-distortion region for successive refinement, and extended his own result to coding systems with a finite number of refinement stages. Effros [15] extended the results by Rimoldi to a class of stationary sources. Viswanathan and Berger [61] investigated a generalization of the successive refinement problem, where the second stage involves describing another correlated message as opposed to improving the description of the same message, and characterized the achievable rate-distortion region for this problem. Steinberg and Merhav [55] considered another extension of the successive refinement problem, where each decoder has access to correlated side information, and clarified the achievable rate-distortion region for degraded side information, i.e. the side information for the second stage can be considered to be of better quality than that of the first stage.

1.2.3 Universal source coding

Meanwhile, a *universal coding* problem for those systems was first investigated by Csiszár and Körner [12]. Universal coding aims to constitute source coding schemes that (asymptotically) establish the limit of coding rates without any stochastic knowledges. Universal coding problems are not only interesting in their own right but also very important in terms of practical applications. Subsequent work including the work by Csiszár and Körner has mainly focused on the SW system. Csiszár and Körner [12] showed the existence of universal codes by introducing a random coding argument and minimum entropy decoding. Later, Csiszár [11], Ahlswede and Dueck [2] presented explicit constructions of universal codes that attain the *error exponent* (the logarithm of the error probability) clarified by Csiszár and Körner. Their constructions utilized linear codes that were usually incorporated into channel coding schemes. Oohama [49] clarified

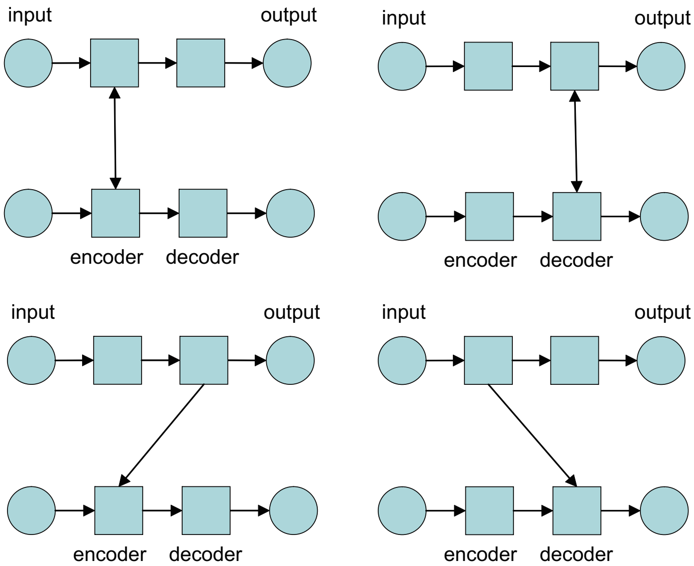

Figure 1.2: Four types of "coding systems with cooperation" [upper left] coding system with cooperative encoders [upper right] coding system with cooperative decoders [lower left] coding system with feedback channels [lower right] coding system with feed-forward channels

a better error exponent than the exponent clarified by Csiszár, Ahlswede and Dueck. Uyematsu [60] improved the work by Csiszár and Körner in terms of the complexity in encoding and decoding to a polynomial order of the length of messages.

The above discussion implies that previous researches have mainly focused on the SW system since it appears to be difficult to construct universal codes for most of the other coding systems. For example, Muramatsu [46] showed that we cannot construct asymptotically optimal fixed-to-variable length (FV) universal codes for the coding system investigated by Wyner and Ziv [73] in terms of the coding rate. A few exception include the work by Effros [16] for the coding system of multistage successive refinement.

1.3 Main issues

This thesis aims to clarify fundamental limits and to construct universal codes of multiterminal source coding for a general framework that involve the SW system, which we call *coding systems with cooperation*. Figure 1.3 shows coding systems included in this framework. As you can see in Figure 1.3, coding systems with cooperation are classified into four types of coding systems that incorporate several linkages among encoders and decoders. The main research issues of the previous work focused on coding systems shown in the upper right of Figure 1.3, which involves cooperative decoders. In this type of coding systems, decoders can share some of information extracted from received codewords. The most typical example is the SW system [54, 6, 58, 4, 63]. Its special cases [73, 34, 24, 7, 38, 36] and extensions [9, 45, 71, 3, 12, 31, 47] are also frequently investigated. Moreover, the triangular communication system [75] and its related systems [74, 27, 28] are also included in this framework. Coding systems shown in the lower right of Figure 1.3 are also frequently investigated, which involves feed-forward channels from encoders to decoders. The most representative examples include successive refinement [43, 18, 19, 51, 15] and related coding systems such as successive refinement with side information [55], sequential coding [61], and source coding models for information retrieval [57]. On the contrary, coding systems shown in the left side of Figure 1.3 have not been fully investigated. The upper left of Figure 1.3 involves cooperation between encoders. Although coding problems of multiple description have been frequently considered [67, 69, 68, 17, 50, 1, 78, 22], a few other studies deal with this framework (e.g. Ericson-Körner [20], Kaspi-Berger [38, 36], Oohama [47]). The lower left of Figure 1.3 represents coding systems with feedback channels from decoders to encoders. Although feedback channels are frequently investigated in problems of channel coding, there are a few studies in source coding (e.g. Yang et al. [76], Kaspi [37]).

To this end, this thesis focuses on two kinds of coding systems that incorporate cooperation with encoders.

First, a coding system that extends the SW system is presented, where there are some mutual linkages between two encoders. This coding system, called the *SW system with linkage (SWL system)*, was originally investigated by Oohama [47], where each encoder can observe the codeword generated from the other encoder. However, his setting allowed each encoder always to observe the codeword of the other encoder. This setting is rather different from the original coding problem of the SW system. On the contrary, in this thesis another case is considered, where the coding rate of the mutual linkage between two encoders is negligible. The above setting can be regarded as a generalization of the original SW system allowing two encoders to communicate with zero rate. This thesis shows that this mutual linkage is enough to effectively reduce the coding rate for a certain class of sources even though the rate of the mutual linkage is zero.

Next, a coding system that contrasts with the SW system in terms of cooperation is investigated, which we call the *complementary delivery coding system*. In this coding system, the encoder observes messages emitted from two correlated sources, and delivers these messages to other locations (i.e. decoders). Each decoder has access one of these messages, and therefore wants to reproduce the other message. Csiszár and Körner [12] considered a general framework that includes the complementary delivery coding system, and clarified the minimum achievable rate (cf. Section VII in [12]) for lossless coding. Also, Willems et al. [66, 72] independently investigated a coding problem where two users are physically separated but communicate with each other via a satellite, and the determined the minimum achievable rate when transmitting to and from the satellite. Their coding system includes the complementary delivery coding problem as a special case.

Unlike these previous studies, lossy coding and universal coding are the focus of attention in this thesis. First, a lossy coding problem for the complementary delivery coding system is investigated, and the rate-distortion function for this problem is clarified. This result implies that there may be some rate losses only for lossy coding when two messages are sent simultaneously to two decoders.

Next, universal coding schemes are proposed for the complementary delivery coding system and its extension called *generalized complementary delivery coding systems*. Explicit constructions of universal lossless codes and the bounds of the error probabilities are clarified via *methods of types* developed by Csiszár and Körner [13] and graph-theoretical analyses.

This thesis is organized as follows: Chapter 2 provided notations, definitions and fundamental lemmas which will be used in the rest of this thesis. Chapter 3 deals with the SWL coding system and clarified the achievable rate regions. Chapter 4 and 5 treat the complementary delivery coding system. Chapter 4 investigates lossy coding problems and clarifies the rate-distortion function for this coding system. Chapter 5 considers universal lossless coding and presents explicit constructions of universal codes and clarifies bounds of the error probabilities. Chapter 6 states concluding remarks and problems that remains to be solved.

Chapter 2

Preliminaries

2.1 Basic definitions

2.1.1 Sequences and distributions

Random variables are denoted by capital letters such as X, and their sample values (resp. alphabets) by the corresponding small letters (resp. calligraphic letters) such as x (resp. $\mathcal{X}$), except as otherwise noted. Especially, let $\mathcal{B}$ be a binary alphabet, $\mathcal{R}$ be a set of real numbers, $\mathcal{R}^+$ be a set of positive real numbers including 0, $\mathcal{N}$ be a set of integers, $\mathcal{N}^+$ be a set of positive integers including 0, and $\mathcal{I}_M = \{1, 2, \cdots, M\}$ for an integer M. Let $\mathcal{B}^*$ be a set of all finite sequences in the alphabet $\mathcal{B}$, $|\mathcal{X}|$ be the cardinality of $\mathcal{X}$. For a subset $\mathcal{A}$ of $\mathcal{X}$, its complement is denoted by $\mathcal{A}^c = \mathcal{X} - \mathcal{A}$. A member of $\mathcal{X}^n$ is written as $x^n = (x_1, x_2, \cdots, x_n)$, and substrings of x^n are written as $x_i^j = (x_i, x_{i+1}, \cdots, x_j) \quad (i \le j)$. When the dimension is clear from the content, vectors will be denoted by boldface letters, i.e., $\boldsymbol{x} \in \mathcal{X}^n$. Also, $*$ represents an operator of sequence concatenation, e.g. $x^n = x^i * x_{i+1}^n \quad (i \le n)$. A function $l : \mathcal{B}^* \to \mathcal{N}^+$ stands for the *length function*, which outputs the length of a binary sequence.

$\mathcal{P}(\mathcal{X})$ represents the set of all probability distributions on $\mathcal{X}$. The distribution of a random variable X taking values in $\mathcal{X}$ is written as P_X, which is a member of $\mathcal{P}(\mathcal{X})$. In a similar manner, the distribution of a set (X, Y) of random variables taking values in $\mathcal{X} \times \mathcal{Y}$ is written as $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Also, $\mathcal{P}(\mathcal{X}|P_Y)$ denotes the set of all probability distributions on $\mathcal{X}$ given a distribution $P_Y \in \mathcal{P}(\mathcal{Y})$, namely each member $P_{X|Y} \in \mathcal{P}(\mathcal{X}|P_Y)$ is characterized by $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ as $P_{XY} = P_{X|Y}P_Y$. When every letter x_i of a sequence $\boldsymbol{x} \in \mathcal{X}^n$ is emitted from a distribution $P_X \in \mathcal{P}(\mathcal{X})$, the distribution of the sequence $\boldsymbol{x}$ is denoted by

$$P_X^n(\boldsymbol{x}) \stackrel{\text{def.}}{=} \prod_{i=1}^{n} P_X(x_i).$$

Random variables X, Y and Z are said to form a *Markov chain*, denoted by $X \to Y \to Z$, if the joint distribution of (X, Y, Z) can be written as

$$P_{XYZ}(x, y, z) = P_X(x)P_{Y|X}(y|x)P_{Z|Y}(z|y).$$

Note that $X \to Y \to Z$ implies $Z \to Y \to X$. For a random variable X and a function $f : \mathcal{X} \to \mathcal{R}$, the expected value of a random variable $f(X)$ is written as $E[f(X)]$.

$$E[f(X)] \quad \stackrel{\text{def.}}{=} \quad \sum_{x \in \mathcal{X}} P_X(x) f(x).$$

Let $\widehat{\mathcal{X}}$ stand for a reproduction alphabet that corresponds to the source alphabet $\mathcal{X}$, and $d_X^n : \mathcal{X}^n \to \widehat{\mathcal{X}}$ stand for a distortion function between an original sequence $\boldsymbol{x} \in \mathcal{X}^n$ and its reproduction sequence $\widehat{\boldsymbol{x}} \in \widehat{\mathcal{X}}^n$. A distortion function d_X^n is called *additive* if d_X^n can be expressed by a single-letter distortion function $d_X : \mathcal{X} \to \widehat{\mathcal{X}}$ as

$$d^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) \quad = \quad \sum_{k=1}^{n} d_X(x_k, \widehat{x}_k).$$

2.1.2 Information sources

A *stationary ergodic source*, or simply an *ergodic source*, is defined as a stationary ergodic process

$$\boldsymbol{X} = \{X_j\}_{j=1}^{\infty}$$

of random variables X_j $(j = 1, 2, \cdots)$, each of which takes values in $\mathcal{X}$. A stationary ergodic source can be defined also for a stationary ergodic process

$$(\boldsymbol{X}, \boldsymbol{Y}) = \{(X_j, Y_j)\}_{j=1}^{\infty}$$

of a set of random variables (X_j, Y_j) $(j = 1, 2, \cdots)$ each of which takes values in $\mathcal{X} \times \mathcal{Y}$.

A *discrete memoryless source (DMS)* is one of the most intuitive special cases of ergodic sources, which is defined as an infinite sequence $\boldsymbol{X}$ of independent copies of a random variable X taking a value in $\mathcal{X}$ with a generic distribution $P_X \in \mathcal{P}(\mathcal{X})$, namely

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

A DMS can be defined also for an infinite sequence $(\boldsymbol{X}, \boldsymbol{Y})$ of independent copies of a set of random variables (X, Y). A DMS will be denoted by referring to its random variable X.

A *general source* [32, 30] is defined as an infinite sequence

$$\boldsymbol{X} = \left\{ X^n = (X_1^{(n)}, \cdots, X_n^{(n)}) \right\}_{n=1}^{\infty}$$

of n-dimensional random variables, where each component random variable $X_i^{(n)}$ $(1 \le i \le n)$ takes values in $\mathcal{X}$. Similarly, a general source can be defined also for an infinite sequence $(\boldsymbol{X}, \boldsymbol{Y})$ of a set of n-dimensional random variables. It should be noted here that each component of (X^n, Y^n) may change depending on the block length n. This implies that the sequence $(\boldsymbol{X}, \boldsymbol{Y})$ is quite general in the sense that it may not satisfy even the consistency condition, where the consistency condition means that for any integers m, n such that $m < n$ it holds that the distribution of $(X_i^{(m)}, Y_i^{(m)})$ equals that of $(X_i^{(n)}, Y_i^{(n)})$ for $i = 1, 2, \cdots, m$. The class of general sources thus defined covers a very wide range of sources including all stationary and/or non-ergodic sources.

A typical example of general sources which are not included in the class of stationary ergodic sources is a *mixed source* $\boldsymbol{X}$, which is defined by the following distribution:

$$P_{X^n}(\boldsymbol{x}) \stackrel{\text{def.}}{=} \alpha P_n^{(1)}(\boldsymbol{x}) + (1-\alpha) P_n^{(2)}(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathcal{X}^n,$$

where $0 < \alpha < 1$ is a constant and $P_n^{(i)} \in \mathcal{P}(\mathcal{X}^n)$ $(i = 1, 2)$ is a distribution of the stationary ergodic process $X^{(i)n} = \{X_j^{(i)}\}_{j=1}^n$. Further, the following notation is introduced

$$\boldsymbol{X}^{(i)} \stackrel{\text{def.}}{=} \left\{X_j^{(i)}\right\}_{j=1}^{\infty} \qquad (i = 1, 2).$$

Similarly, a mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ can be defined by the following distribution:

$$P_{X^n Y^n}(\boldsymbol{x}, \boldsymbol{y}) \stackrel{\text{def.}}{=} \alpha P_n^{(1)}(\boldsymbol{x}, \boldsymbol{y}) + (1-\alpha) P_n^{(2)}(\boldsymbol{x}, \boldsymbol{y}) \qquad \forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n,$$

where $P_n^{(i)} \in \mathcal{P}(\mathcal{X}^n \times \mathcal{Y}^n)$ $(i = 1, 2)$ are distributions of the stationary ergodic process $(X^{(i)n}, Y^{(i)n}) = \{(X_j^{(i)}, Y_j^{(i)})\}_{j=1}^n$.

Suppose that $(X^{(1)}, X^{(2)}, \cdots, X^{(N_s)})$ is a set of N_s DMSs, where each DMS $X^{(i)}$ takes values in the alphabet $\mathcal{X}^{(i)}$ $(i \in \mathcal{I}_{N_s})$. For a subset $\mathcal{S} \subseteq \mathcal{I}_{N_s}$ of the source indices, the corresponding subset of sources, sample sequences and alphabets are respectively written as

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

In what follows, all bases of exponentials and logarithms are set at 2.

2.1.3 Information measures

For a random variable X which takes values in $\mathcal{X}$, $H(X)$ denotes the *Shannon entropy*, or simply the *entropy* of X, defined by the following equation:

$$H(X) \stackrel{\text{def.}}{=} \sum_{x \in \mathcal{X}} P_X(x) \log \frac{1}{P_X(x)}.$$

For a set (X, Y) of random variables which takes values in $\mathcal{X} \times \mathcal{Y}$, $H(X, Y)$ and $H(X|Y)$ denote the *joint entropy* of (X, Y) and the *conditional entropy* of X given Y, respectively, defined as follows:

$$H(X, Y) \stackrel{\text{def.}}{=} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} P_{XY}(x, y) \log \frac{1}{P_{XY}(x, y)},$$

$$H(X|Y) \stackrel{\text{def.}}{=} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x,y)\log\frac{1}{P_{X|Y}(x|y)}.$$

For a set (X, Y, Z) of random variables taking values in $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$, $I(X;Y)$ and $I(X;Y|Z)$ denote the *mutual information* between X and Y, and the *conditional mutual information* between X and Y given Z, respectively, defined as follows:

$$I(X;Y) \stackrel{\text{def.}}{=} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)},$$
$$I(X;Y|Z) \stackrel{\text{def.}}{=} \sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} P_{XYZ}(x,y,z)\log\frac{P_{XY|Z}(x,y|z)}{P_{X|Z}(x|z)P_{Y|Z}(y|z)}.$$

For a generic distribution $P\in\mathcal{P}(\mathcal{Y})$ and a stochastic matrix $W:\mathcal{X}\to\mathcal{Y}$, $H(P)$ and $H(W|P)$ also denote the entropy of Y the conditional entropy of X given Y, where $P=P_Y$, $W=P_{X|Y}$. For two generic distributions $P,Q\in\mathcal{P}(\mathcal{X})$, $D(P\|Q)$ denotes the *Kullback-Leibler divergence*, or simply the *divergence* between P and Q, defined as follows:

$$D(P\|Q) \stackrel{\text{def.}}{=} \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}.$$

Similarly, for two stochastic matrices V, $W:\mathcal{X}\to\mathcal{Y}$ and a generic distribution $P\in\mathcal{P}(\mathcal{X})$, $D(V\|W|P)$ denotes the *conditional Kullback-Leibler divergence*, or simply the *conditional divergence* between V and W given P, defined as

$$\begin{aligned} D(V\|W|P) &\stackrel{\text{def.}}{=} \sum_{x\in\mathcal{X}} P(x)D(V(\cdot|x)\|W(\cdot|x)) \\ &= \sum_{x\in\mathcal{X}} P(x)\sum_{y\in\mathcal{Y}} V(y|x)\log\frac{V(y|x)}{W(y|x)} \end{aligned}$$

The above information measures satisfy the following properties, which are well known as *chain rules*:

$$\begin{aligned} H(X,Y) &= H(X)+H(Y|X) \\ I(X,Y;Z) &= I(X;Z)+I(Y;Z|X) \\ D(P_{XY}\|Q_{XY}) &= D(P_X\|Q_X)+D(P_{Y|X}\|Q_{Y|X}|P_X) \end{aligned}$$

For an ergodic source $\boldsymbol{X}$, the *entropy rate* $H(\boldsymbol{X})$ of $\boldsymbol{X}$ is defined as

$$H(\boldsymbol{X}) \stackrel{\text{def.}}{=} \lim_{n\to\infty}\frac{1}{n}H(X^n).$$

Similarly, for an ergodic source $(\boldsymbol{X},\boldsymbol{Y})$, a *joint entropy rate* $H(\boldsymbol{X},\boldsymbol{Y})$, a *conditional entropy rate* $H(\boldsymbol{X}|\boldsymbol{Y})$ and a *mutual information rate* $I(\boldsymbol{X};\boldsymbol{Y})$ are defined as

$$\begin{aligned} H(\boldsymbol{X},\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \lim_{n\to\infty}\frac{1}{n}H(X^n,Y^n), \\ H(\boldsymbol{X}|\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \lim_{n\to\infty}\frac{1}{n}H(X^n|Y^n), \end{aligned}$$

$$I(\boldsymbol{X};\boldsymbol{Y}) \stackrel{\text{def.}}{=} \lim_{n\to\infty}\frac{1}{n}I(X^n;Y^n),$$

respectively. The entropy rate also satisfies the chain rule:

$$H(\boldsymbol{X},\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}|\boldsymbol{X}).$$

For any DMS (X,Y), the entropy rates and the mutual information rates are reduced to ordinary information measures such as entropies and mutual information, as follows:

$$\begin{aligned} H(\boldsymbol{X}) &= H(X), \\ H(\boldsymbol{X},\boldsymbol{Y}) &= H(X,Y), \\ H(\boldsymbol{X}|\boldsymbol{Y}) &= H(X|Y), \\ I(\boldsymbol{X};\boldsymbol{Y}) &= I(X;Y). \end{aligned}$$

For a general source $\boldsymbol{X}$, the *entropy spectrum sup* $\overline{H}(\boldsymbol{X})$ [30] of $\boldsymbol{X}$ is defined as

$$\overline{H}(\boldsymbol{X}) \stackrel{\text{def.}}{=} \inf\left\{\alpha : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > \alpha\right\} = 0\right\}.$$

Similarly, for a general source $(\boldsymbol{X},\boldsymbol{Y})$, the *joint entropy spectrum sup* $\overline{H}(\boldsymbol{X},\boldsymbol{Y})$ of $(\boldsymbol{X},\boldsymbol{Y})$, and the *conditional entropy spectrum sup* $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$ of $\boldsymbol{X}$ given $\boldsymbol{Y}$ are respectively defined as follows:

$$\begin{aligned} \overline{H}(\boldsymbol{X},\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \inf\left\{\alpha : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^nY^n}(X^n,Y^n)} > \alpha\right\} = 0\right\}, \\ \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \inf\left\{\alpha : \lim_{n\to\infty}\Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^nY^n}(X^n|Y^n)} > \alpha\right\} = 0\right\}. \end{aligned}$$

The *sup entropy rate* $\widetilde{H}(\boldsymbol{X})$ [30] of a general source $\boldsymbol{X}$ is defined as

$$\widetilde{H}(\boldsymbol{X}) \stackrel{\text{def.}}{=} \limsup_{n\to\infty}\frac{1}{n}H(X^n).$$

Similarly, the *sup joint entropy rate* $\widetilde{H}(\boldsymbol{X},\boldsymbol{Y})$ and the *sup conditional entropy rate* $\widetilde{H}(\boldsymbol{X}|\boldsymbol{Y})$ are defined as

$$\begin{aligned} \widetilde{H}(\boldsymbol{X},\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \limsup_{n\to\infty}\frac{1}{n}H(X^n,Y^n), \\ \widetilde{H}(\boldsymbol{X}|\boldsymbol{Y}) &\stackrel{\text{def.}}{=} \limsup_{n\to\infty}\frac{1}{n}H(X^n|Y^n). \end{aligned}$$

Moreover, let me introduce another type of entropy rates. The *weak sup entropy rate* $H^*(\boldsymbol{X})$ of a general source $\boldsymbol{X}$ is defined as

$$H^*(\boldsymbol{X}) \stackrel{\text{def.}}{=} \lim_{\epsilon\to+0}\limsup_{n\to\infty}\frac{1}{n}H_{[\epsilon]}(X^n),$$

where $H_{[\epsilon]}(X)$ is the ϵ-*entropy* of a random variable X for a give constant $\epsilon > 0$, defined as

$$H_{[\epsilon]}(X) \stackrel{\text{def.}}{=} \inf_{A\subseteq\mathcal{X}:\Pr\{X\in A\}\geq 1-\epsilon} H(X|A).$$

When a correlated ergodic source $(\boldsymbol{X},\boldsymbol{Y})$ is considered, the entropy spectrum sum, the sup entropy rate and the weak sup entropy rate all equal the entropy rate:

Lemma 2.1.1. (Han [30])
For any ergodic source $(\boldsymbol{X},\boldsymbol{Y})$

$$\begin{aligned}\overline{H}(\boldsymbol{X}) &= \widetilde{H}(\boldsymbol{X}) = H^*(\boldsymbol{X}) = H(\boldsymbol{X}),\\ \overline{H}(\boldsymbol{X},\boldsymbol{Y}) &= \widetilde{H}(\boldsymbol{X},\boldsymbol{Y}) = H(\boldsymbol{X},\boldsymbol{Y}),\\ \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) &= \widetilde{H}(\boldsymbol{X}|\boldsymbol{Y}) = H(\boldsymbol{X}|\boldsymbol{Y}).\end{aligned}$$

On the contrary, when a source $(\boldsymbol{X},\boldsymbol{Y})$ is a mixed source, those information measures differ from each other.

Lemma 2.1.2. (Han [30])
For any mixed source $(\boldsymbol{X},\boldsymbol{Y})$

$$\begin{aligned}\overline{H}(\boldsymbol{X}) &= \max\{H(\boldsymbol{X}^{(1)}), H(\boldsymbol{X}^{(2)})\},\\ \widetilde{H}(\boldsymbol{X}) &= H^*(\boldsymbol{X}) = \alpha H(\boldsymbol{X}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}),\\ \overline{H}(\boldsymbol{X},\boldsymbol{Y}) &= \max\{H(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}), H(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)})\},\\ \widetilde{H}(\boldsymbol{X},\boldsymbol{Y}) &= H^*(\boldsymbol{X},\boldsymbol{Y}) = \alpha H(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)}),\\ \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) &= \max\{H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}), H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)})\},\\ \widetilde{H}(\boldsymbol{X}|\boldsymbol{Y}) &= \alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}),\end{aligned}$$

where $(\boldsymbol{X}^{(j)},\boldsymbol{Y}^{(j)})$ *is an ergodic source whose distribution is given by* $\{P_n^{(j)}\}_{n=1}^{\infty}$ $(j=1,2)$*.*

2.1.4 Fundamental coding theorems

In the following, several fundamental coding theorems are introduced, which clarify the operational meaning of the information measures defined in Section 2.1.3.

First, let us introduce coding theorems for lossless coding.

Lemma 2.1.3. (e.g. [10, Theorem 5.3.1])
A set $\mathcal{A}\subseteq\mathcal{B}^*$ *is called a prefix set if no element of the set is a prefix of any other element. Also, a code* $f:\mathcal{X}\to\mathcal{B}^*$ *is called a prefix code if its range* $f(\mathcal{X})$ *is a prefix set. The expected length of any prefix code for a random variable* X *must satisfy*

$$E[l(f(X))] \quad\geq\quad H(X),$$

with equality if and only if

$$\exp\{-l(f(x))\} \quad=\quad P_X(x) \quad \forall x\in\mathcal{X}.$$

Lemma 2.1.4. (Coding theorem of fixed-length lossless codes for general sources [32])
A fixed-length lossless code is defined as a sequence $\{(\phi_n,\widehat{\phi}_n)\}_{n=1}^{\infty}$ *of codes* $(\phi_n,\widehat{\phi}_n)$ *that satisfies*

$$\phi_n:\mathcal{X}^n\to\mathcal{I}_{M_n},$$

$$\begin{aligned}
&\widehat{\phi}_n : \mathcal{I}_{M_n} \to \mathcal{X}^n, \\
&\lim_{n\to\infty} \Pr\{\widehat{\phi}_n(\phi_n(X^n)) \neq X^n\} = 0.
\end{aligned}$$

There exists a fixed-length lossless code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq \overline{H}(\boldsymbol{X}).$$

Also, any fixed-length lossless code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *must satisfy*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \geq \overline{H}(\boldsymbol{X}).$$

Lemma 2.1.5. (Coding theorem of variable-length lossless codes for general sources [30, Theorem 1.10])
A variable-length lossless code is defined as a sequence $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *of codes* $(\phi_n, \widehat{\phi}_n)$ *that satisfies*

$$\begin{aligned}
&\phi_n : \mathcal{X}^n \to \mathcal{B}^*, \\
&\widehat{\phi}_n : \phi_n(\mathcal{X}^n) \to \mathcal{X}^n, \\
&\widehat{\phi}_n(\phi_n(\boldsymbol{x})) = \boldsymbol{x} \quad \forall \boldsymbol{x} \in \mathcal{X}^n.
\end{aligned}$$

There exists a variable-length lossless code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(X^n))] \leq \widetilde{H}(\boldsymbol{X}).$$

Also, any variable-length lossless code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *must satisfy*

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(X^n))] \geq \widetilde{H}(\boldsymbol{X}).$$

Lemma 2.1.6. (Coding theorem of weak variable-length (lossless) codes for general sources [29][30, Theorem 1.12])
A weak variable-length (lossless) code is defined as a sequence $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *of codes* $(\phi_n, \widehat{\phi}_n)$ *that satisfies*

$$\begin{aligned}
&\phi_n : \mathcal{X}^n \to \mathcal{B}^*, \\
&\widehat{\phi}_n : \phi_n(\mathcal{X}^n) \to \mathcal{X}^n, \\
&\lim_{n\to\infty} \Pr\{\widehat{\phi}_n(\phi_n(X^n)) \neq X^n\} = 0.
\end{aligned}$$

There exists a weak variable-length code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(X^n))] \leq H^*(\boldsymbol{X}).$$

Also, any weak variable-length code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a general source* $\boldsymbol{X}$ *must satisfy*

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(X^n))] \geq H^*(\boldsymbol{X}). \tag{2.1}$$

From the above discussions, $\overline{H}(\boldsymbol{X})$, $\widetilde{H}(\boldsymbol{X})$ and $H^*(\boldsymbol{X})$ each represents the minimum achievable rate of fixed-length coding, variable-length coding and weak variable-length coding, respectively.

Next, let us introduce coding theorems for lossy coding.

Lemma 2.1.7. (Coding theorem of fixed-length lossy codes [5])
A fixed-length lossy code is defined as a sequence $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *of codes* $(\phi_n, \widehat{\phi}_n)$ *that satisfies*

$$\begin{aligned}
&\phi_n : \mathcal{X}^n \to \mathcal{I}_{M_n}, \\
&\widehat{\phi}_n : \mathcal{I}_{M_n} \to \widehat{\mathcal{X}}^n, \\
&E[\Delta_X(X^n, \widehat{\phi}_n(\phi_n(X^n)))] \leq D,
\end{aligned}$$

for a given distortion criterion $D > 0$*. There exists a fixed-length lossy code* $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a DMS* X *and a given distortion criterion* D *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R(X|D),$$

where $R(X|D)$ *is the rate-distortion function defined as*

$$R(X|D) \stackrel{\text{def.}}{=} \min_{\substack{P_{\widehat{X}|X} \in \mathcal{P}(\widehat{\mathcal{X}}|P_X): \\ E[\Delta_X(X, \widehat{X})] \leq D}} I(X; \widehat{X}). \tag{2.2}$$

Also, any fixed-length lossy code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *for a DMS* X *and a given distortion criterion* D *must satisfy*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \geq R(X|D).$$

Lemma 2.1.8. (Coding theorem of fixed-length lossy codes with common side information [5])
A fixed-length lossy code with common side information is defined as a sequence $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *of codes* $(\phi_n, \widehat{\phi}_n)$ *that satisfies*

$$\begin{aligned}
&\phi_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{I}_{M_n}, \\
&\widehat{\phi}_n : \mathcal{I}_{M_n} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\
&E[\Delta_X(X^n, \widehat{\phi}_n(\phi_n(X^n, Y^n), Y^n))] \leq D,
\end{aligned}$$

for a given distortion criterion $D > 0$*. There exists a fixed-length lossy code* $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ *with common side information for a DMS* (X, Y) *and a given distortion criterion* D *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R_C(X|Y, D),$$

where $R_C(X|Y,D)$ is the conditional rate-distortion function defined as

$$R(X|Y,D) \stackrel{\text{def.}}{=} \min_{\substack{P_{\widehat{X}|XY} \in \mathcal{P}(\widehat{\mathcal{X}}|P_{XY}): \\ E[\Delta_X(X,\widehat{X})] \le D}} I(X;\widehat{X}|Y).$$

Also, any fixed-length code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ with common side information for a DMS (X,Y) and a distortion criterion must satisfy

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \ge R_C(X|Y,D).$$

Lemma 2.1.9. (Coding theorem of fixed-length lossy codes with decoder side information [73])
A fixed-length lossy code with decoder side information is defined as a sequence $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ of codes $(\phi_n, \widehat{\phi}_n)$ that satisfies

$$\begin{aligned} &\phi_n : \mathcal{X}^n \to \mathcal{I}_{M_n}, \\ &\widehat{\phi}_n : \mathcal{I}_{M_n} \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\ &E[\Delta_X(X^n, \widehat{\phi}_n(\phi_n(X^n), Y^n))] \le D, \end{aligned}$$

for a given distortion criterion $D > 0$. There exists a fixed-length lossy code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ with decoder side information for a DMS (X,Y) and a given distortion criterion that satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R_{WV}(X|Y,D),$$

where $R_{WV}(X|Y,D)$ is called the Wyner-Ziv rate-distortion function, which is defined as follows:

$$R_{WV}(X|Y,D) \stackrel{\text{def.}}{=} \min_{P_{U|XY} \in \mathcal{P}_{WV}(\mathcal{U}|P_{XY})} I(X;U|Y),$$

where the alphabet $\mathcal{U}$ satisfies $|\mathcal{U}| \le |\mathcal{X}| + 1$, and $\mathcal{P}_{WV}(\mathcal{U}|P_{XY})$ is a subset of $\mathcal{P}(\mathcal{U}|P_{XY})$ such that there exists a function $\psi : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ satisfying

$$E[\Delta_X(X, \psi(U,Y))] \quad \le \quad D.$$

Also, any fixed-length code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ with decoder side information for a DMS (X,Y) and a given distortion criterion D must satisfy

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \ge R_{WV}(X|Y,D).$$

2.2 Types of sequences

Let us define a *type* of a sequence $\boldsymbol{x} \in \mathcal{X}^n$ as the distribution $Q_{\boldsymbol{x}} \in \mathcal{P}(\mathcal{X})$ where $Q_{\boldsymbol{x}}(a)$ is the relative frequency of the symbol a in $\boldsymbol{x}$, i.e.

$$Q_{\boldsymbol{x}}(a) = \frac{1}{n} N(a|\boldsymbol{x}) \quad \forall a \in \mathcal{X},$$

where $N(a|\boldsymbol{x})$ represents the number of occurrences of the letter a in the sequence $\boldsymbol{x}$. The joint type $Q_{\boldsymbol{x}^{(\mathcal{S})}} \in \mathcal{P}(\mathcal{X}^{(\mathcal{S})})$ of the set $\boldsymbol{x}^{(\mathcal{S})} \in \mathcal{X}^{(\mathcal{S})}$ of sequences is similarly defined by

$$\begin{aligned} Q_{\boldsymbol{x}^{(\mathcal{S})}}(a_{i_1}, a_{i_2}, \cdots, a_{i_{|\mathcal{S}|}}) &= \frac{1}{n} N\left(a_{i_1}, a_{i_2}, \cdots, a_{i_{|\mathcal{S}|}} | \boldsymbol{x}^{(\mathcal{S})}\right) \\ & \quad \forall (a_{i_1}, a_{i_2}, \cdots, a_{i_{|\mathcal{S}|}}) \in \mathcal{X}^{(\mathcal{S})}. \end{aligned}$$

Further, for every $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$, if $(\boldsymbol{x}, \boldsymbol{y})$ has a joint type $Q_{\boldsymbol{x},\boldsymbol{y}}$ that satisfies

$$Q_{\boldsymbol{x},\boldsymbol{y}}(a, b) = Q_{\boldsymbol{x}}(a) V(b|a),$$

we shall say that $\boldsymbol{y}$ has a *conditional type* V given $\boldsymbol{x}$. The set of types of sequences in $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. Further, for every type $Q \in \mathcal{P}_n(\mathcal{X})$ and some sequence $\boldsymbol{x}$ of type Q, the set of conditional types of sequences in $\mathcal{Y}^n$ given $\boldsymbol{x}$ is denoted by $\mathcal{V}_n(\mathcal{Y}|Q)$. For every type $Q \in \mathcal{P}_n(\mathcal{X})$, the set of sequences of type Q in $\mathcal{X}^n$ is denoted by T_Q^n:

$$T_Q^n \stackrel{\text{def.}}{=} \{\boldsymbol{x} \in \mathcal{X}^n | Q_{\boldsymbol{x}} = Q\}.$$

Similarly, for every $\boldsymbol{x} \in T_Q^n$ and $V \in \mathcal{V}_n(\mathcal{Y}|Q)$, the set of sequences of conditional type V given $\boldsymbol{x}$ is denoted by $T_V^n(\boldsymbol{x})$:

$$T_V^n(\boldsymbol{x}) \stackrel{\text{def.}}{=} \{\boldsymbol{y} \in \mathcal{Y}^n | Q(x)V(y|x) = Q_{\boldsymbol{x},\boldsymbol{y}}(x, y),\ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}\}.$$

Hereafter, we call $T_V^n(\boldsymbol{x})$ a *V-shell*.

Here, let us introduce several important characteristics of types.

Lemma 2.2.1. (Type counting lemma [13, Lemma 2.2])

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}.$$

Lemma 2.2.2. (Size of V-shells [13, Lemma 2.5])
For every sequence $\boldsymbol{x} \in \mathcal{X}^n$ and stochastic matrix $V : \mathcal{X} \to \mathcal{Y}$ such that the corresponding V-shell $T_V^n(\boldsymbol{x})$ is not empty, we have

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{nH(V|P_X)\} \leq |T_V(\boldsymbol{x})| \leq \exp\{nH(V|P_X)\}.$$

Lemma 2.2.3. (Probability of types [13, Lemma 2.6])
For every type $Q \in \mathcal{P}_n(\mathcal{X})$ and every distribution $P_X \in \mathcal{P}(\mathcal{X})$, we have

$$\begin{aligned} & P_X^n(\boldsymbol{x}) = \exp\{-n(D(Q\|P_X) + H(Q))\} \quad \text{if } \boldsymbol{x} \in T_Q^n, \\ & (n+1)^{-|\mathcal{X}|} \exp\{-nD(Q\|P_X)\} \leq P_X^n(T_Q) \leq \exp\{-nD(Q\|P_X)\}. \end{aligned}$$

2.3 Typical sequences

Let us introduce a concept of *typical sequences* with respect to a distribution $P_X \in \mathcal{P}(\mathcal{X})$ of a random variable X. For $\delta > 0$, the set of typical sequences is defined as

$$T_X^n(\delta) \stackrel{\text{def.}}{=} \{\boldsymbol{x} \in \mathcal{X}^n : D(Q_{\boldsymbol{x}}\|P_X) \leq \delta\}.$$

Similarly, the set of *conditional typical sequences* with respect to a conditional distribution $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|P_X)$ of a random variable Y given a sequence $\boldsymbol{x} \in \mathcal{X}^n$ is defined as follows:

$$T^n_{Y|\boldsymbol{x}}(\delta) \stackrel{\text{def.}}{=} \left\{ \boldsymbol{y} \in \mathcal{Y}^n : D(V_{\boldsymbol{y}|\boldsymbol{x}} \| P_{Y|X} | P_{\boldsymbol{x}}) \leq \delta \right\},$$

where $V_{\boldsymbol{y}|\boldsymbol{x}}$ is the conditional type of the sequence $\boldsymbol{y} \in \mathcal{Y}^n$ given $\boldsymbol{x} \in \mathcal{X}^n$. Here let us introduce several fundamental lemmas for typical sequences:

Lemma 2.3.1. [59, Theorem 2.6]
For any $\boldsymbol{x} \in T^n_X(\delta)$

$$\begin{aligned} |P_{\boldsymbol{x}}(x) - P_X| &\leq \sqrt{2\delta} \quad \forall x \in \mathcal{X}, \\ P_{\boldsymbol{x}}(x) &= 0 \quad \textit{if } P_X(x) = 0. \end{aligned}$$

Also, for any $\boldsymbol{y} \in T^n_{Y|\boldsymbol{x}}(\delta)$

$$\begin{aligned} |P_{\boldsymbol{xy}}(x,y) - P_{XY}(x,y)| &\leq \sqrt{2\delta} \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}, \\ P_{\boldsymbol{xy}}(x,y) &= 0 \quad \textit{if } P_{Y|X}(y|x) = 0. \end{aligned}$$

Lemma 2.3.2. (Uyematsu [59, Theorem 2.7])
For any δ *satisfying* $0 \leq \delta \leq \frac{1}{8}$, *if* $\boldsymbol{x} \in T^n_X(\delta)$ *then*

$$\left| -\frac{1}{n} \log P^n_X(\boldsymbol{x}) - H(X) \right| \leq \delta - \sqrt{2\delta} \log \frac{\sqrt{2\delta}}{|\mathcal{X}|}.$$

Also, for any $\boldsymbol{x} \in T^n_X(\delta)$ *and any* δ' *satisfying* $0 \leq \delta' \leq \frac{1}{8}$, *if* $\boldsymbol{y} \in T^n_{Y|\boldsymbol{x}}(\delta)$ *then*

$$\left| -\frac{1}{n} \log P^n_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) - H(Y|X) \right| \leq \delta' - \sqrt{2\delta'} \log \frac{2\delta'}{|\mathcal{X} \times \mathcal{Y}|} + \sqrt{2\delta} \log |\mathcal{Y}|.$$

Lemma 2.3.3. (Wolfowitz [70], Csiszar-Körner[13, Lemma 2.12])
For a given $\delta > 0$ *and any DMS* X,

$$\Pr\{X^n \in T^n_X(\delta)\} \geq 1 - \exp\left\{ -n \left(\delta - \frac{|\mathcal{X}| \log(n+1)}{n} \right) \right\}.$$

Also, any $\boldsymbol{x} \in \mathcal{X}^n$ *and any DMS* (X, Y),

$$\Pr\{Y^n \in T^n_{Y|\boldsymbol{x}}(\delta) | \boldsymbol{x}\} \geq 1 - \exp\left\{ -n \left(\delta - \frac{|\mathcal{X} \times \mathcal{Y}| \log(n+1)}{n} \right) \right\}.$$

Remark 2.3.1. *Lemmas 2.3.1, 2.3.2 and 2.3.3 can be obtained even when a sequence* $\{\delta_n\}_{n=1}^{\infty}$ *is introduced instead of a given* $\delta > 0$, *provided that*

$$\left. \begin{aligned} \lim_{n\to\infty} \delta_n &= 0, \\ \lim_{n\to\infty} n\delta_n &= \infty. \end{aligned} \right\} \tag{2.3}$$

Lemma 2.3.4. (Wolfowitz [70], Csiszar-Körner[13, Lemma 2.13])
For given sequences $\{(\delta_n, \delta'_n)\}_{n=1}^{\infty}$ *satisfying Eq. (2.3) and any pair* (X, Y) *of random variables, there exists a sequence* $\{\epsilon_n = \epsilon_n(|\mathcal{X}|, |\mathcal{Y}|, \delta_n, \delta'_n)\}_{n=1}^{\infty}$ *that satisfies*

$$\begin{aligned} \left| \frac{1}{n} \log |T_X^n(\delta_n)| - H(X) \right| &\leq \epsilon_n, \\ \left| \frac{1}{n} \log |T_{Y|\boldsymbol{x}}^n(\delta'_n)| - H(Y|X) \right| &\leq \epsilon_n, \quad \boldsymbol{x} \in T_X^n(\delta_n), \\ \lim_{n\to\infty} \epsilon_n &= 0. \end{aligned}$$

Lemma 2.3.5. (Csiszar and Körner[13, Lemma 1.2.10])
For any $\delta, \delta' > 0$,

$$\boldsymbol{x} \in T_X^n(\delta_1) \text{ and } \boldsymbol{y} \in T_{Y|\boldsymbol{x}}^n(\delta_2) \Rightarrow (\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}^n(\delta_1 + \delta_2), \tag{2.4}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}^n(\delta_1) \Rightarrow \boldsymbol{x} \in T_X^n(\delta_1 |\mathcal{Y}|), \tag{2.5}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \in T_{XY}^n(\delta_1) \text{ and } \boldsymbol{x} \in T_X^n(\delta_2) \Rightarrow \boldsymbol{y} \in T_{Y|\boldsymbol{x}}^n(\delta_1 + \delta_2). \tag{2.6}$$

Remark 2.3.2. *Although Eq.(2.6) of Lemma 2.3.5 is not shown in [13], this property can be easily obtained in the same way as the others.*

Lemma 2.3.6. (Markov lemma [58, 6] [13, Lemma 1.2.10])
Let random variables X*,* Y *and* Z *form a Markov chain* $X \to Y \to Z$*. If for given* $\delta > 0$ *and* $(\boldsymbol{y}, \boldsymbol{z}) \in T_{YZ}^n(\delta)$ *a random variable* X^n *is selected according to the distribution* $P_{X|Y}^n$*, then*

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

for an arbitrary $\epsilon > 0$ *and a sufficiently large* n*.*

Next, let us introduce *weak typical sequences* with respect to a distribution $P_{X^n} \in \mathcal{P}(\mathcal{X}^n)$ of a random variable X^n. For $\delta > 0$, the set of weak typical sequences is defines as

$$\widehat{T}_{\boldsymbol{X}}^n(\delta) \stackrel{\text{def.}}{=} \left\{ \boldsymbol{x} \in \mathcal{X}^n : \left| \frac{1}{n} \log \frac{1}{P_{X^n}(\boldsymbol{x})} - H(\boldsymbol{X}) \right| \leq \delta \right\}.$$

The following lemma shows a fundamental property of weak typical sequences for ergodic sources, which corresponds to Lemma 2.3.3.

Lemma 2.3.7. (Shannon-McMillan-Breiman Theorem [10, Theorem 16.8.1])
For any ergodic source $(\boldsymbol{X}, \boldsymbol{Y})$

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} &= H(\boldsymbol{X}) \quad a.s., \\ \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{Y^n}(Y^n)} &= H(\boldsymbol{Y}) \quad a.s., \\ \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{X^nY^n}(X^n, Y^n)} &= H(\boldsymbol{X}, \boldsymbol{Y}) \quad a.s. \end{aligned}$$

These properties imply that for any $\epsilon, \delta > 0$ *there exists an integer* $n_0 = n_0(\epsilon, \delta, \boldsymbol{X}, \boldsymbol{Y})$ *such that for all* $n \geq n_0(\epsilon, \delta, \boldsymbol{X}, \boldsymbol{Y})$

$$\Pr\left\{X^n \in \widehat{T}^n_{\boldsymbol{X}}(\epsilon)\right\} \leq \delta,$$
$$\Pr\left\{Y^n \in \widehat{T}^n_{\boldsymbol{Y}}(\epsilon)\right\} \leq \delta,$$
$$\Pr\left\{(X^n, Y^n) \in \widehat{T}^n_{\boldsymbol{XY}}(\epsilon)\right\} \leq \delta,$$

simultaneously hold.

2.4 Graph coloring

Let us introduce several notations and lemmas related to graph coloring.

An *undirected graph*, or simply *graph*, is denoted as $G = (V_G, E_G)$, where V_G is a set of vertices and E_G is a set of edges. The *degree* $\Delta(v)$ of a vertex $v \in V_G$ is the number of other vertices connected to by edges, and the degree $\Delta(G)$ of a graph G is the maximum number of degrees of vertices $v \in V_G$, namely

$$\Delta(G) = \max_{v \in V_G} \Delta(v).$$

A simple graph where an edge connects every pair of vertices is called as a *complete graph*. A complete subgraph is called as a *clique*, and the largest degree of cliques in a graph G is called the *clique number* $\omega(G)$ of the graph G.

An *vertex coloring*, or simply *coloring* of a graph G is that no two adjacent vertices are assigned the same symbol, and the number of symbols necessary for vertex coloring is called the *chromatic number* $\chi(G)$. Similarly, an *edge coloring* of a graph G is that no two adjacent edges are assigned the same symbol, and the number of symbols necessary for edge coloring is called the *edge chromatic number* $\chi'(G)$.

The following lemmas show well known bounds of the chromatic number and the edge chromatic number.

Lemma 2.4.1. (Brooks [8, 14])

$$\omega(G) \leq \chi(G) \leq \Delta(G)$$

unless G *is a complete graph or an odd cycle (a cycle graph that consists of odd number of vertexes).*

Lemma 2.4.2. (Vizing [62, 14])

$$\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1.$$

Lemma 2.4.3. (König [39, 14])
If a graph G *is bipartite, then*

$$\chi'(G) = \Delta(G).$$

Chapter 3

Slepian-Wolf coding with linked encoders

3.1 Introduction

This chapter investigates coding problems for the SWL coding system, where there are some mutual linkages between encoders of the SW system. The SWL system was originally investigated by Oohama [47], where each encoder can observe the codeword generated from the other encoder. However, his setting is rather different from the original coding problem of the SW system. On the contrary, this chapter considers another case where the coding rate of the mutual linkage between two encoders is negligibly small. This setting can be regarded as a generalization of the original SW system allowing two encoders to communicate with *zero rate*. This chapter shows that this mutual linkage is enough to effectively reduce the coding rate even though the rate of the mutual linkage is zero.

The fixed-length coding for general sources is first considered, and its achievable rate region is clarified, which equals the achievable rate region of the SW system. This implies that the linkage does not reduce the coding rate for fixed-length codes. Next, the weak variable-length coding is considered. The achievable rate region for mixed sources characterized by two ergodic sources is clarified, which is strictly wider than that for fixed-length codes. This result contrasts with that for the fixed-length coding. Even though the rate of the mutual linkage is zero, this linkage is enough to distinguish for which ergodic source the input sequence is typical, and therefore effectively reduces the coding rate. Further, the universal coding for DMSs is investigated in the SWL system, and it is shown that an arbitrary pair of coding rates in the achievable rate region depending on the source can be attained by the weak variable-length code. In case of universal coding, the linkage is used to estimate the probability distribution of the source, and gives drastic flexibility to weak variable-length coding.

The organization of this chapter follows: Section 3.2 deals with fixed-length coding problems for SW and SWL coding systems, and shows their achievable rate regions. Section 3.3 discusses weak variable-length coding problems for the SWL system, and presents the main results without proofs, and Section 3.4

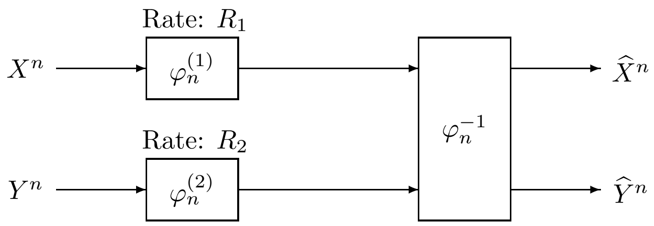

Figure 3.1: Slepian-Wolf coding system

gives their proofs.

3.2 Fixed-length coding problems

3.2.1 Slepian-Wolf coding system

Slepian and Wolf [54] studied the coding problem for two correlated sources, where two sequences from correlated sources are separately encoded, sent to a single decoder which has to output original sequence pairs (Figure 3.1). We call this data compression system *the Slepian-Wolf system (the SW system).*

Definition 3.2.1. (Fixed-length SW code)
A sequence $\{(\varphi_n^{(1)}, \varphi_n^{(2)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$ of codes $(\varphi_n^{(1)}, \varphi_n^{(2)}, \widehat{\varphi}_n)$ is called a *fixed-length SW code (f-SW code)* for the source $(\boldsymbol{X}, \boldsymbol{Y})$, if and only if

$$\begin{aligned}
&\varphi_n^{(1)} : \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}},\\
&\varphi_n^{(2)} : \mathcal{Y}^n \to \mathcal{I}_{M_n^{(2)}},\\
&\widehat{\varphi}_n : \mathcal{I}_{M_n^{(1)}} \times \mathcal{I}_{M_n^{(2)}} \to \mathcal{X}^n \times \mathcal{Y}^n,\\
&\lim_{n\to\infty} \Pr\{\widehat{\varphi}_n(\varphi_n^{(1)}(X^n), \varphi_n^{(2)}(Y^n)) \neq (X^n, Y^n)\} = 0. \qquad (3.1)
\end{aligned}$$

Definition 3.2.2. (f-SW achievable rate pair)
A rate pair (R_1, R_2) is a *f-SW achievable rate pair* of the source $(\boldsymbol{X}, \boldsymbol{Y})$, if and only if there exists a SW code for the source $(\boldsymbol{X}, \boldsymbol{Y})$, which satisfies

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(1)} \le R_1,\\
\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(2)} \le R_2.
\end{aligned}$$

Definition 3.2.3. (f-SW achievable rate region)
The *f-SW achievable rate region* $\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\begin{aligned}
&\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y})\\
&\quad \overset{\text{def.}}{=} \{(R_1, R_2) : (R_1, R_2) \text{ is a f-SW achievable rate pair of } (\boldsymbol{X}, \boldsymbol{Y}).\}.
\end{aligned}$$

Miyake and Kanaya [45] investigated the SW system for correlated general sources and clarified the f-SW achievable rate region as follows:

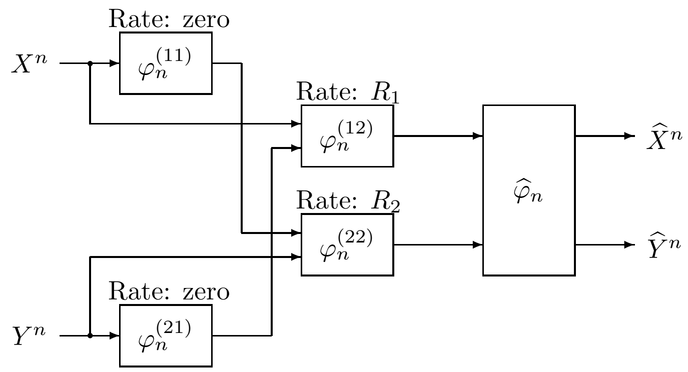

Figure 3.2: Slepian-Wolf coding system with linked encoders

Theorem 3.2.1. (Miyake and Kanaya [45])
For any correlated general source $(\boldsymbol{X}, \boldsymbol{Y})$,

$$\begin{aligned}\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y}) = \{(R_1, R_2) \quad : \quad & R_1 \geq \overline{H}(\boldsymbol{X}|\boldsymbol{Y}),\ R_2 \geq \overline{H}(\boldsymbol{Y}|\boldsymbol{X}), \\ & R_1 + R_2 \geq \overline{H}(\boldsymbol{X}, \boldsymbol{Y})\},\end{aligned}$$

The next corollary can be obtained immediately from Lemmas 2.1.1 and 2.1.2.

Corollary 3.2.1. *If* $(\boldsymbol{X}, \boldsymbol{Y})$ *is an ergodic source, then*

$$\begin{aligned}\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y}) = \{(R_1, R_2) \quad : \quad & R_1 \geq H(\boldsymbol{X}|\boldsymbol{Y}),\ R_2 \geq H(\boldsymbol{Y}|\boldsymbol{X}), \\ & R_1 + R_2 \geq H(\boldsymbol{X}, \boldsymbol{Y})\}.\end{aligned}$$

Further, if $(\boldsymbol{X}, \boldsymbol{Y})$ *is a mixed source, then*

$$\begin{aligned}\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y}) = \big\{(R_1, R_2) \quad : \quad & R_1 \geq \max\Big(H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}), H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)})\Big), \\ & R_2 \geq \max\Big(H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}), H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)})\Big), \\ & R_1 + R_2 \geq \max\Big(H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}), H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})\Big)\Big\}.\end{aligned}$$

3.2.2 Slepian-Wolf coding system with linked encoders

Oohama [47] considered the *the SWL system* in the sense of an SW system having the *linkage* of two encoders. First, the fixed-length coding for the SWL system (called *f-SWL system*) is defined.

Definition 3.2.4. (Fixed-length SWL code (f-SWL code))
A sequence

$$\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$$

of codes

$$(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)$$

is called a *fixed-length SWL code (f-SWL code)* for the source $(\boldsymbol{X}, \boldsymbol{Y})$, if and only if

$$
\begin{aligned}
&\varphi_n^{(11)} : \mathcal{X}^n \to \mathcal{I}_{M_n^{(11)}},\\
&\varphi_n^{(12)} : \mathcal{X}^n \times \mathcal{I}_{M_n^{(21)}} \to \mathcal{I}_{M_n^{(12)}},\\
&\varphi_n^{(21)} : \mathcal{Y}^n \to \mathcal{I}_{M_n^{(21)}},\\
&\varphi_n^{(22)} : \mathcal{Y}^n \times \mathcal{I}_{M_n^{(11)}} \to \mathcal{I}_{M_n^{(22)}},\\
&\widehat{\varphi}_n : \mathcal{I}_{M_n^{(12)}} \times \mathcal{I}_{M_n^{(22)}} \to \mathcal{X}^n \times \mathcal{Y}^n,\\
&\lim_{n\to\infty} \Pr\{\widehat{\varphi}_n(A_n^{(1)}, A_n^{(2)}) \neq (X^n, Y^n)\} = 0,
\end{aligned}
$$

where

$$
\left.
\begin{aligned}
A_n^{(1)} &\stackrel{\text{def.}}{=} \varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)),\\
A_n^{(2)} &\stackrel{\text{def.}}{=} \varphi_n^{(22)}(Y^n, \varphi_n^{(11)}(X^n)).
\end{aligned}
\right\}
\tag{3.2}
$$

Definition 3.2.5. (f-SWL achievable rate pair)
A rate pair (R_1, R_2) is a *f-SWL achievable rate pair* of the source $(\boldsymbol{X}, \boldsymbol{Y})$, if and only if there exists a f-SWL code for the source $(\boldsymbol{X}, \boldsymbol{Y})$ which satisfies

$$
\left.
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\log M_n^{(11)} = 0,\\
&\lim_{n\to\infty} \frac{1}{n}\log M_n^{(21)} = 0,\\
&\limsup_{n\to\infty} \frac{1}{n}\log M_n^{(12)} \le R_1,\\
&\limsup_{n\to\infty} \frac{1}{n}\log M_n^{(22)} \le R_2.
\end{aligned}
\right\}
\tag{3.3}
$$

It should be noted that both rates of the first encoders $\varphi_n^{(11)}$ and $\varphi_n^{(21)}$ are zero. This implies that the outputs of the first encoders can be provided to the decoder with zero-rate as suitable prefixes in the outputs of the second encoders $\varphi_n^{(12)}$ and $\varphi_n^{(22)}$. Hence, even if the outputs of the first encoders are directly provided to the decoder, the admissible rate region remains the same. On the other hand, Oohama [47] considered the opposite case where both rates of the second encoders $\varphi_n^{(12)}$ and $\varphi_n^{(22)}$ are zero, and the decoder can see the outputs of the first encoders.

Definition 3.2.6. (f-SWL achievable rate region)
The f-SWL achievable rate region is defined as

$$
\begin{aligned}
&\mathcal{R}_{SWL}(\boldsymbol{X}, \boldsymbol{Y})\\
&\stackrel{\text{def.}}{=} \{(R_1, R_2) : (R_1, R_2) \text{ is an f-SWL achievable rate pair of } (\boldsymbol{X}, \boldsymbol{Y})\}.
\end{aligned}
$$

The next theorem clarifies the f-SWL achievable rate region.

Theorem 3.2.2. (Coding theorem for the f-SWL system)
For any general source $(\boldsymbol{X},\boldsymbol{Y})$,

$$\begin{aligned}&\mathcal{R}_{SWL}(\boldsymbol{X},\boldsymbol{Y})=\mathcal{R}_{SW}(\boldsymbol{X},\boldsymbol{Y})\\&\quad=\ \{(R_1,R_2):R_1\geq\overline{H}(\boldsymbol{X}|\boldsymbol{Y}),\ R_2\geq\overline{H}(\boldsymbol{Y}|\boldsymbol{X}),\ R_1+R_2\geq\overline{H}(\boldsymbol{X},\boldsymbol{Y})\}.\end{aligned}$$

This theorem implies that the achievable rate region does not expand when fixed-length coding is considered even if there are mutual linkages between two encoders.

3.3 Weak variable-length coding problems

3.3.1 Slepian-Wolf coding system with linked encoders

This section defines a weak variable-length code for the SWL system (called *the wv-SWL system*).

Definition 3.3.1. (Weak variable-length SWL code (wv-SWL code))
A sequence

$$\{(\varphi_n^{(11)},\varphi_n^{(12)},\varphi_n^{(21)},\varphi_n^{(22)},\widehat{\varphi}_n)\}_{n=1}^{\infty}$$

of codes

$$(\varphi_n^{(11)},\varphi_n^{(12)},\varphi_n^{(21)},\varphi_n^{(22)},\widehat{\varphi}_n)$$

is called a *wv-SWL code* for the source $(\boldsymbol{X},\boldsymbol{Y})$, if and only if

$$\begin{aligned}&\varphi_n^{(11)}:\mathcal{X}^n\to\mathcal{B}^*,\\&\varphi_n^{(12)}:\mathcal{X}^n\times\varphi_n^{(21)}(\mathcal{Y}^n)\to\mathcal{B}^*,\\&\varphi_n^{(21)}:\mathcal{Y}^n\to\mathcal{B}^*,\\&\varphi_n^{(22)}:\mathcal{Y}^n\times\varphi_n^{(11)}(\mathcal{X}^n)\to\mathcal{B}^*,\\&\widehat{\varphi}_n:\varphi_n^{(12)}(\mathcal{X}^n,\varphi_n^{(21)}(\mathcal{Y}^n))\times\varphi_n^{(22)}(\mathcal{Y}^n,\varphi_n^{(11)}(\mathcal{X}^n))\to\mathcal{X}^n\times\mathcal{Y}^n,\\&\lim_{n\to\infty}\Pr\{\widehat{\varphi}_n(A_n^{(1)},A_n^{(2)})\neq(X^n,Y^n)\}=0\end{aligned}\tag{3.4}$$

and the images of $\varphi_n^{(11)}$, $\varphi_n^{(12)}$, $\varphi_n^{(21)}$ and $\varphi_n^{(22)}$ are all prefix sets, where $A_n^{(1)}$ and $A_n^{(2)}$ are defined in Eq. (3.2).

Definition 3.3.2. (wv-SWL achievable rate pair)
A rate pair (R_1,R_2) is a *wv-SWL achievable rate pair* of the source $(\boldsymbol{X},\boldsymbol{Y})$, if and only if there exists a wv-SWL code for the source $(\boldsymbol{X},\boldsymbol{Y})$ which satisfies

$$\left.\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}E[l(\varphi_n^{(11)}(X^n))]&=0,\\\limsup_{n\to\infty}\frac{1}{n}E[l(\varphi_n^{(21)}(Y^n))]&=0,\\\limsup_{n\to\infty}\frac{1}{n}E\left[l(\varphi_n^{(12)}(X^n,\varphi_n^{(21)}(Y^n)))\right]&\leq R_1,\\\limsup_{n\to\infty}\frac{1}{n}E\left[l(\varphi_n^{(22)}(Y^n,\varphi_n^{(11)}(X^n)))\right]&\leq R_2.\end{aligned}\right\}\tag{3.5}$$

Definition 3.3.3. (wv-SWL achievable rate region)
The *wv-SWL achievable rate region* $\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y})$ is defined as

$$\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y}) \stackrel{\text{def.}}{=} \{(R_1,R_2) : (R_1,R_2) \text{ is a wv-SWL achievable rate pair of } (\boldsymbol{X},\boldsymbol{Y})\}.$$

3.3.2 Main results

This section clarifies the wv-SWL rate region for mixed sources. The next theorem is our main result.

Theorem 3.3.1. *If* $(\boldsymbol{X},\boldsymbol{Y})$ *is a correlated mixed source, then*

$$\begin{aligned}\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y}) = \Big\{(R_1,R_2) :\ & R_1 \geq \alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}),\\ & R_2 \geq \alpha H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}) + (1-\alpha)H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)}),\\ & R_1+R_2 \geq \alpha H(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)})\Big\}.\end{aligned}$$

Theorems 3.3.1 and 3.2.2 imply that the wv-SWL region is strictly wider than the f-SWL region, i.e.

$$\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y}) \supset \mathcal{R}_{SWL}(\boldsymbol{X},\boldsymbol{Y}) = \mathcal{R}_{SW}(\boldsymbol{X},\boldsymbol{Y})$$

for any mixed source $(\boldsymbol{X},\boldsymbol{Y})$. This implies that wv-SWL codes can achieve strictly lower coding rate than f-SWL codes.

It is instructive to note here how to construct the wv-SWL code. For a given rate $(R_1,R_2) \in \mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y})$, two rate pairs $(R_1^{(1)},R_2^{(1)}) \in \mathcal{R}_{SW}(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)})$ and $(R_1^{(2)},R_2^{(2)}) \in \mathcal{R}_{SW}(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)})$ can be found such that

$$\begin{aligned} R_1 &= \alpha R_1^{(1)} + (1-\alpha)R_1^{(2)},\\ R_2 &= \alpha R_2^{(1)} + (1-\alpha)R_2^{(2)}.\end{aligned}$$

Then, two SW codes are prepared, one of which is $\{(f_n^{(1)},f_n^{(2)},\widehat{f}_n)\}_{n=1}^\infty$ for the source $(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)})$ with rate $(R_1^{(1)},R_2^{(1)})$, and the other of which is $\{(g_n^{(1)},g_n^{(2)},\widehat{g}_n)\}_{n=1}^\infty$ for the source $(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)})$ with rate $(R_1^{(2)},R_2^{(2)})$. The encoders $\varphi_n^{(11)}$ and $\varphi_n^{(21)}$ send the first N_n (e.g. $N_n = \log n$) symbols of each input sequence of length n to the other encoder. Sharing the pair of sequences of length N_n, one of the SW codes, i.e. $(f_n^{(1)},f_n^{(2)},\widehat{f}_n)$ or $(g_n^{(1)},g_n^{(2)},\widehat{g}_n)$ can be selected depending on for which source the shared pair of sequences is typical. Then, the encoders $\varphi_n^{(12)}$ and $\varphi_n^{(22)}$ send to the decoder the first N_n symbols the input sequences and the codewords of the selected SW code. Since the decoder can have the knowledge of which SW code the encoders employ, the estimate of the input pair of sequences can be obtained by using the corresponding decoder.

As a special case of Theorem 3.3.1, the wv-SWL achievable rate region for ergodic sources can be immediately obtained.

Corollary 3.3.1. *If* $(\boldsymbol{X},\boldsymbol{Y})$ *is an ergodic source, then*

$$\begin{aligned}\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y}) = \{(R_1,R_2)\ :\ & R_1 \geq H(\boldsymbol{X}|\boldsymbol{Y}),\ R_2 \geq H(\boldsymbol{Y}|\boldsymbol{X}),\\ & R_1+R_2 \geq H(\boldsymbol{X},\boldsymbol{Y})\}.\end{aligned}$$

Theorem 3.2.2 and Corollary 3.3.1 imply that

$$\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y}) = \mathcal{R}_{SWL}(\boldsymbol{X},\boldsymbol{Y})$$

for any ergodic source $(\boldsymbol{X},\boldsymbol{Y})$. Hence, wv-SWL codes cannot improve the coding rate for ergodic sources compared with f-SWL codes.

3.3.3 Related results

The next theorem shows that for a restricted class of ergodic sources, the rate pair in the region $\mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y})$ can be achieved without any linkage of encoders.

Theorem 3.3.2. *Assume that a mixed source $(\boldsymbol{X},\boldsymbol{Y})$ satisfies both $H(\boldsymbol{X}^{(1)}) \neq H(\boldsymbol{X}^{(2)})$ and $H(\boldsymbol{Y}^{(1)}) \neq H(\boldsymbol{Y}^{(2)})$. Then, for any $(R_1,R_2) \in \mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y})$, we can construct a wv-SWL code*

$$\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$$

such that

$$\varphi_n^{(11)}(\boldsymbol{x}) = \varphi_n^{(21)}(\boldsymbol{y}) = \lambda \text{ (null string)},$$

for any $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ and positive integer n.

This theorem indicates that for a restricted class of mixed sources, the wv-SWL region can be achieved by the SW code. Further, in such a case, $\mathcal{R}_{SW}(\boldsymbol{X},\boldsymbol{Y}) \subset \mathcal{R}^*_{SWL}(\boldsymbol{X},\boldsymbol{Y})$, that is, a weak variable-length code can achieve smaller rate than the fixed-length code for the SW system.

Theorems 3.3.1 and 3.3.2 only considered the mixture of two ergodic sources, but it can be easily extended to the mixture of any finite number of ergodic sources. This implies the next corollary which shows a simple version of universal coding for the SWL system.

Corollary 3.3.2. *For a given set $\{(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})\}_{n=1}^{m}$ of ergodic sources and any set $\{(R_1^{(i)},R_2^{(i)})\}_{n=1}^{m}$ of rate pairs which satisfies*

$$R_1^{(i)} \geq H(\boldsymbol{X}^{(i)}|\boldsymbol{Y}^{(i)}), \quad R_2^{(i)} \geq H(\boldsymbol{Y}^{(i)}|\boldsymbol{X}^{(i)}), \quad R_1^{(i)} + R_2^{(i)} \geq H(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)}),$$

there exists a wv-SWL code $\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$ such that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(11)}(X^{(i)n}))] &= 0,\\
\lim_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(21)}(Y^{(i)n}))] &= 0,\\
\limsup_{n\to\infty} \frac{1}{n} E[l(A_n^{(1)}(i))] &\leq R_1^{(i)},\\
\limsup_{n\to\infty} \frac{1}{n} E[l(A_n^{(2)}(i))] &\leq R_2^{(i)}
\end{aligned}$$

for $i = 1, 2, \cdots, m$, where

$$\begin{aligned}
A_n^{(1)}(i) &\stackrel{\text{def.}}{=} \varphi_n^{(12)}(X^{(i)n}, \varphi_n^{(21)}(Y^{(i)n})),\\
A_n^{(2)}(i) &\stackrel{\text{def.}}{=} \varphi_n^{(22)}(Y^{(i)n}, \varphi_n^{(11)}(X^{(i)n})).
\end{aligned}$$

This corollary shows that only if the source is known to belong to the given set S, a pair of sequences from the source $(X^{(i)}, Y^{(i)})$ can be encoded with a rate pair $(R_1^{(i)}, R_2^{(i)})$ which is an arbitrary point in the achievable rate region of the source $(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})$. Though this corollary is valid for a finite set of sources, this result is rather different from the conventional universal coding for SW systems with fixed length codes [13]. The next theorem shows that this property of the weak variable-length universal coding is strengthened for DMSs.

Theorem 3.3.3. *Let S be a set of DMSs. Further, for every DMS $(X, Y) \in S$, we correspond a rate pair $(R_1(P_{XY}), R_2(P_{XY}))$ which is an inner point of the SW region $\mathcal{R}_{SW}(X, Y)$. We assume that $R_1(P_{XY})$ and $R_2(P_{XY})$ are continuous functions of P_{XY}. Then, there exists a universal wv-SWL code*

$$\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$$

such that for any DMS $(X, Y) \in S$

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(11)}(X^n))] &= 0, \\
\lim_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(21)}(Y^n))] &= 0, \\
\limsup_{n\to\infty} \frac{1}{n} E[l(A_n^{(1)})] &\leq R_1(P_{XY}), \\
\limsup_{n\to\infty} \frac{1}{n} E[l(A_n^{(1)})] &\leq R_2(P_{XY}),
\end{aligned}$$

where $A_n^{(1)}$ and $A_n^{(2)}$ are defined in Eq. (3.2).

This corollary shows that for each DMS, the rate pair may be arbitrarily specified in its SW region, and that the rate pair is achievable by a universal wv-SWL code. Without any linkage of encoders, only (universal) fixed-length coding is realized. However, using the linkage of encoders, variable-length coding can be realized depending on the source.

3.4 Proof of theorems

In this section, P_n denotes the distribution $P_{X^nY^n}$ of the mixed source $(\boldsymbol{X}, \boldsymbol{Y})$, and $P_n^{(i)}$ denotes the distribution $P_{X^{(i)n}Y^{(i)n}}$ of the ergodic source $(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})$ $(i = 1, 2)$, for simplicity.

3.4.1 Proof of Theorem 3.2.2

Proof. The proof of the achievability part is obvious from Theorem 3.2.1. Therefore, we shall only prove the converse part for the f-SWL system under the condition

$$\left.\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(11)} M_n^{(12)} &\leq R_1 \\
\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(21)} M_n^{(22)} &\leq R_2
\end{aligned}\right\} \quad (3.6)$$

which is weaker than the condition (3.3).

First, for any $\gamma > 0$, we define the sets $\tilde{T}_n^{(i)}$ $(i = 1, 2, 3)$ and S_n by

$$
\begin{aligned}
\tilde{T}_n^{(1)} &\stackrel{\text{def.}}{=} \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x}|\boldsymbol{y})} \geq \frac{1}{n}\log M_n^{(1)} + \gamma\right\}, \\
\tilde{T}_n^{(2)} &\stackrel{\text{def.}}{=} \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y}|\boldsymbol{x})} \geq \frac{1}{n}\log M_n^{(2)} + \gamma\right\}, \\
\tilde{T}_n^{(3)} &\stackrel{\text{def.}}{=} \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} \geq \frac{1}{n}\log M_n^{(1)}M_n^{(2)} + \gamma\right\}, \\
S_n &\stackrel{\text{def.}}{=} \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \widehat{\varphi}_n(\varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y}), \varphi_n^{(2)}(\boldsymbol{x}, \boldsymbol{y})) = (\boldsymbol{x}, \boldsymbol{y})\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y}) &\stackrel{\text{def.}}{=} (\varphi_n^{(11)}(\boldsymbol{x}), \varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y}))), \\
\varphi_n^{(2)}(\boldsymbol{x}, \boldsymbol{y}) &\stackrel{\text{def.}}{=} (\varphi_n^{(21)}(\boldsymbol{y}), \varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x}))), \\
M_n^{(1)} &\stackrel{\text{def.}}{=} M_n^{(11)}M_n^{(12)}, \\
M_n^{(2)} &\stackrel{\text{def.}}{=} M_n^{(21)}M_n^{(22)}.
\end{aligned}
$$

By letting

$$e_n \stackrel{\text{def.}}{=} \Pr\{(X^n, Y^n) \notin S_n\},$$

we obtain

$$
\begin{aligned}
&\Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)}\} \\
&= \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n\} + \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n^c\} \\
&\leq \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n\} + \Pr\{(X^n, Y^n) \notin S_n\} \\
&= \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n\} + e_n. \qquad (3.7)
\end{aligned}
$$

Note that if $(\boldsymbol{x}, \boldsymbol{y}) \in \tilde{T}_n^{(1)}$ then $P_n(\boldsymbol{x}|\boldsymbol{y}) \leq \exp(-n\gamma)/M_n^{(1)}$. Hence, the first term of Eq. (3.7) can be evaluated as

$$
\begin{aligned}
\Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n\} &= \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \tilde{T}_n^{(1)} \cap S_n} P_n(\boldsymbol{x}, \boldsymbol{y}) \\
&\leq \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \tilde{T}_n^{(1)} \cap S_n} P_n(\boldsymbol{y})\frac{\exp(-n\gamma)}{M_n^{(1)}} \\
&\leq \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in S_n} P_n(\boldsymbol{y})\frac{\exp(-n\gamma)}{M_n^{(1)}} \\
&= \sum_{\boldsymbol{y} \in \mathcal{Y}^n} P_n(\boldsymbol{y})|S_n(\boldsymbol{y})|\frac{\exp(-n\gamma)}{M_n^{(1)}}, \qquad (3.8)
\end{aligned}
$$

where

$$S_n(\boldsymbol{y}) \stackrel{\text{def.}}{=} \{\boldsymbol{x} \in \mathcal{X}^n : (\boldsymbol{x}, \boldsymbol{y}) \in S_n\}.$$

Since

$$\varphi_n^{(1)}(\boldsymbol{x},\boldsymbol{y}) = (\varphi_n^{(11)}(\boldsymbol{x}), \varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})))$$

and $\varphi_n^{(2)}$ is a function of $\boldsymbol{y}$ and $\varphi^{(11)}$, $\varphi_n^{(2)}$ is a function of $\boldsymbol{y}$ and $\varphi_n^{(1)}$. Thus, for a given $\boldsymbol{y} \in \mathcal{Y}^n$, the number of sequences $\boldsymbol{x} \in \mathcal{X}^n$ which satisfy

$$\widehat{\varphi}_n(\varphi_n^{(1)}(\boldsymbol{x},\boldsymbol{y}), \varphi_n^{(2)}(\boldsymbol{x},\boldsymbol{y})) = (\boldsymbol{x},\boldsymbol{y})$$

is at most the number of codewords of $\varphi_n^{(1)}$. Then, we have

$$|S_n(\boldsymbol{y})| \le |\varphi_n^{(1)}(\mathcal{X}^n, \mathcal{Y}^n)| \le M_n^{(1)} \quad \forall \boldsymbol{y} \in \mathcal{Y}^n.$$

Substituting this inequality into Eq. (3.8), we obtain

$$\Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)} \cap S_n\} \le \exp(-n\gamma).$$

Hence (3.7) can be rewritten as

$$\Pr\{(X^n, Y^n) \in \tilde{T}_n^{(1)}\} \le e_n + \exp(-n\gamma). \tag{3.9}$$

In a similar manner, we obtain

$$\begin{aligned} \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(2)}\} &\le e_n + \exp(-n\gamma), & (3.10)\\ \Pr\{(X^n, Y^n) \in \tilde{T}_n^{(3)}\} &\le e_n + \exp(-n\gamma). & (3.11) \end{aligned}$$

Now, let (R_1, R_2) be admissible for the f-SWL system. Then, there exists a fixed-length SWL code such that for any $\gamma > 0$ and sufficiently large n,

$$\begin{aligned} \frac{1}{n}\log M_n^{(1)} &\le R_1 + \gamma, & (3.12)\\ \frac{1}{n}\log M_n^{(2)} &\le R_2 + \gamma, & (3.13)\\ \lim_{n\to\infty} e_n &= 0. & (3.14) \end{aligned}$$

Substituting Eqs. (3.12)-(3.13) into Eqs. (3.9)-(3.11), we have

$$\begin{aligned} e_n &\ge P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x}|\boldsymbol{y})} \ge R_1 + 2\gamma\right\} - \exp(-n\gamma),\\ e_n &\ge P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y}|\boldsymbol{x})} \ge R_2 + 2\gamma\right\} - \exp(-n\gamma),\\ e_n &\ge P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} \ge R_1 + R_2 + 3\gamma\right\} - \exp(-n\gamma). \end{aligned}$$

Then, according to Eq. (3.14),

$$\begin{aligned} \lim_{n\to\infty} P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x}|\boldsymbol{y})} \ge R_1 + 2\gamma\right\} &= 0,\\ \lim_{n\to\infty} P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y}|\boldsymbol{x})} \ge R_2 + 2\gamma\right\} &= 0,\\ \lim_{n\to\infty} P_n\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} \ge R_1 + R_2 + 3\gamma\right\} &= 0. \end{aligned}$$

From the definition of the entropy spectrum sup, we must have

$$\begin{aligned} R_1 + 2\gamma &\geq \overline{H}(\boldsymbol{X}|\boldsymbol{Y}), \\ R_2 + 2\gamma &\geq \overline{H}(\boldsymbol{Y}|\boldsymbol{X}), \\ R_1 + R_2 + 3\gamma &\geq \overline{H}(\boldsymbol{X}, \boldsymbol{Y}). \end{aligned}$$

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

$$\begin{aligned} R_1 &\geq \overline{H}(\boldsymbol{X}|\boldsymbol{Y}), \\ R_2 &\geq \overline{H}(\boldsymbol{Y}|\boldsymbol{X}), \\ R_1 + R_2 &\geq \overline{H}(\boldsymbol{X}, \boldsymbol{Y}). \end{aligned}$$

This completes the proof of Theorem 3.2.2. □

3.4.2 Proof of Theorem 3.3.1

[Converse part]

Proof. Here, it should be noted that from Lemma 2.1.2

$$\begin{aligned} \widetilde{H}(\boldsymbol{X}) = H^*(\boldsymbol{X}) &= \alpha H(\boldsymbol{X}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)}), \\ \widetilde{H}(\boldsymbol{Y}) = H^*(\boldsymbol{Y}) &= \alpha H(\boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{Y}^{(2)}), \\ \widetilde{H}(\boldsymbol{X}, \boldsymbol{Y}) = H^*(\boldsymbol{X}, \boldsymbol{Y}) &= \alpha H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)}), \\ \widetilde{H}(\boldsymbol{X}|\boldsymbol{Y}) = H^*(\boldsymbol{X}|\boldsymbol{Y}) &= \alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}), \\ \widetilde{H}(\boldsymbol{Y}|\boldsymbol{X}) = H^*(\boldsymbol{Y}|\boldsymbol{X}) &= \alpha H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}) + (1-\alpha) H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)}), \end{aligned}$$

where

$$H^*(\boldsymbol{Y}|\boldsymbol{X}) \stackrel{\text{def.}}{=} H^*(\boldsymbol{X}, \boldsymbol{Y}) - H^*(\boldsymbol{X}), \tag{3.15}$$

$$H^*(\boldsymbol{X}|\boldsymbol{Y}) \stackrel{\text{def.}}{=} H^*(\boldsymbol{X}, \boldsymbol{Y}) - H^*(\boldsymbol{Y}), \tag{3.16}$$

According to Lemma 2.1.5, there exists a variable-length code $\{(\phi_n, \widehat{\phi}_n)\}_{n=1}^{\infty}$ for the mixed source $\boldsymbol{Y}$ such that the encoder $\phi_n : \mathcal{Y}^n \to \mathcal{B}^*$ and the decoder $\widehat{\phi}_n : \mathcal{B}^* \to \mathcal{Y}^n$ satisfy

$$\limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(Y^n))] \leq \widetilde{H}(\boldsymbol{Y}) = H^*(\boldsymbol{Y}), \tag{3.17}$$

$$\widehat{\phi}_n(\phi_n(\boldsymbol{y})) = \boldsymbol{y} \qquad \forall \boldsymbol{y} \in \mathcal{Y}^n. \tag{3.18}$$

Then, for a given wv-SWL code

$$\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty},$$

we construct a sequence of codes $\{(\psi_n, \psi_n^{-1})\}_{n=1}^{\infty}$ ($\psi_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{B}^*$, $\widehat{\psi}_n : \mathcal{B}^* \to \mathcal{X}^n \times \mathcal{Y}^n$) for the mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ as follows (see also Figure 3.2):

$$\left.\begin{aligned} \psi_n(\boldsymbol{x}, \boldsymbol{y}) &\stackrel{\text{def.}}{=} s_1 * s_2 * s_3, \\ \widehat{\psi}_n(s_1 * s_2 * s_3) &\stackrel{\text{def.}}{=} \widehat{\varphi}_n(s_2, \varphi_n^{(22)}(\widehat{\phi}_n(s_3), s_1)) \end{aligned}\right\} \tag{3.19}$$

for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, where

$$
\begin{aligned}
s_1 &\stackrel{\text{def.}}{=} \varphi_n^{(11)}(\boldsymbol{x}), \\
s_2 &\stackrel{\text{def.}}{=} \varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})), \\
s_3 &\stackrel{\text{def.}}{=} \phi_n(\boldsymbol{y}).
\end{aligned}
$$

Since the images of $\varphi_n^{(11)}$, $\varphi_n^{(12)}$, $\varphi_n^{(21)}$ and ϕ_n are all prefix sets, the image of ψ_n is also a prefix set. Further, from Eqs. (3.4)(3.18), the error probability of this code can be bounded as

$$
\begin{aligned}
&\Pr\{\widehat{\psi}_n(\psi_n(X^n, Y^n)) \neq (X^n, Y^n)\} \\
&= \Pr\{\widehat{\varphi}_n(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)), \varphi_n^{(22)}(Y^n, \varphi_n^{(11)}(X^n))) \neq (X^n, Y^n)\} \\
&\to 0 \quad (n \to \infty).
\end{aligned}
$$

This implies that $\{(\psi_n, \psi_n^{-1})\}_{n=1}^{\infty}$ is a weak variable-length code. Hence, according to Lemma 2.1.6, it must satisfy

$$
\limsup_{n\to\infty} \frac{1}{n} E[l(\psi_n(X^n, Y^n))] \geq H^*(\boldsymbol{X}, \boldsymbol{Y}).
$$

Hence, we have

$$
\begin{aligned}
H^*(\boldsymbol{X}, \boldsymbol{Y}) &\leq \limsup_{n\to\infty} \frac{1}{n} E[l(\psi_n(X^n, Y^n))] \\
&\stackrel{(a)}{=} \limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(11)}(X^n)) + l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n))) + l(\phi_n(Y^n))] \\
&\leq \limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(11)}(X^n))] + \limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] \\
&\quad + \limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(Y^n))] \\
&\stackrel{(b)}{=} \limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] + \limsup_{n\to\infty} \frac{1}{n} E[l(\phi_n(Y^n))] \\
&\stackrel{(c)}{\leq} \limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] + H^*(\boldsymbol{Y}),
\end{aligned}
$$

where the equality (a) comes from Eq. (3.19), (b) from Eq. (3.5) and (c) from Eq. (3.17). This implies

$$
\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] &\geq H^*(\boldsymbol{X}, \boldsymbol{Y}) - H^*(\boldsymbol{Y}) \\
&= H^*(\boldsymbol{X}|\boldsymbol{Y}),
\end{aligned}
$$

where the last equality follows from Eq. (3.16). Therefore, if a wv-SWL code $\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$ satisfies

$$
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] \leq R_1,
$$

then $R_1 \geq H^*(\boldsymbol{X}|\boldsymbol{Y})$. In a similar manner, if

$$
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(22)}(Y^n, \varphi_n^{(11)}(X^n)))] \leq R_2,
$$

then $R_2 \geq H^*(\boldsymbol{Y}|\boldsymbol{X})$. Further, $R_1 + R_2 \geq H^*(\boldsymbol{X}, \boldsymbol{Y})$ is obvious from Lemma 2.1.6. This completes the proof of the converse part. □

[Achievability part]

Proof. First, we consider the case where two ergodic sources $(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)})$ and $(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})$ cannot be discriminated by the entropy rate, i.e. a mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ satisfies the following three conditions:

$$
\begin{aligned}
H(\boldsymbol{X}^{(1)}) &= H(\boldsymbol{X}^{(2)}), \\
H(\boldsymbol{Y}^{(1)}) &= H(\boldsymbol{Y}^{(2)}), \\
H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}) &= H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)}).
\end{aligned}
$$

In this case, we have $\mathcal{R}_{SW}(\boldsymbol{X}, \boldsymbol{Y}) = \mathcal{R}^*_{SWL}(\boldsymbol{X}, \boldsymbol{Y})$. Hence, for a given mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ and any rate pair $(R_1, R_2) \in \mathcal{R}^*_{SWL}(\boldsymbol{X}, \boldsymbol{Y})$, there exists an SW code $\{(\varphi_n^{(1)}, \varphi_n^{(2)}, \varphi_n^{(-1)})\}_{n=1}^{\infty}$ which satisfies

$$
\begin{aligned}
\limsup_{n \to \infty} \frac{1}{n} \log M_n^{(1)} &\leq R_1, \\
\limsup_{n \to \infty} \frac{1}{n} \log M_n^{(2)} &\leq R_2.
\end{aligned}
$$

This shows the existence of wv-SWL codes for any $(R_1, R_2) \in \mathcal{R}^*_{SWL}(\boldsymbol{X}, \boldsymbol{Y})$, since a SW code is a special case of wv-SWL codes.

Next, we consider the case where two ergodic sources $(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)})$ and $(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})$ can be discriminated by the entropy rate, i.e. a mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ satisfies at least one of the following conditions (1) – (3):

$$
\left.
\begin{array}{ll}
(1) & H(\boldsymbol{X}^{(1)}) \neq H(\boldsymbol{X}^{(2)}) \\
(2) & H(\boldsymbol{Y}^{(1)}) \neq H(\boldsymbol{Y}^{(2)}) \\
(3) & H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}) \neq H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})
\end{array}
\right\} \tag{3.20}
$$

Here, we introduce the following fundamental lemma.

Lemma 3.4.1. *For a mixed source $(\boldsymbol{X}, \boldsymbol{Y})$ and any $\gamma > 0$, we have*

$$
P_n^{(i)} \Bigg\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\
\left| \frac{1}{n} \log \frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} - \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} \right| \leq \max(\gamma, c_0/n) \Bigg\} \geq 1 - \exp(-n\gamma), \tag{3.21}
$$

for $i = 1, 2$ and any integer $n > 0$, where $c_0 = -\log \min(\alpha, 1 - \alpha)$.

The proof of Lemma 3.4.1 is given in the last part of this chapter.

(Step 1) Preliminaries

For any subset $B_n \subset \mathcal{X}^n \times \mathcal{Y}^n$, we introduce the notations

$$
\Pr\{(X^n, Y^n) \in B_n\} \stackrel{\text{def.}}{=} \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in B_n} P_n(\boldsymbol{x}, \boldsymbol{y}),
$$

$$\Pr\{(X^{(i)n}, Y^{(i)n}) \in B_n\} \stackrel{\text{def.}}{=} \sum_{(\boldsymbol{x},\boldsymbol{y})\in B_n} P_n^{(i)}(\boldsymbol{x},\boldsymbol{y}) \qquad (i=1,2).$$

By the above notations, we immediately have

$$\begin{aligned}
&\Pr\{(X^n, Y^n) \in B_n\}\\
&= \sum_{(\boldsymbol{x},\boldsymbol{y})\in B_n} P_n(\boldsymbol{x},\boldsymbol{y})\\
&= \alpha \sum_{(\boldsymbol{x},\boldsymbol{y})\in B_n} P_n^{(1)}(\boldsymbol{x},\boldsymbol{y}) + (1-\alpha) \sum_{(\boldsymbol{x},\boldsymbol{y})\in B_n} P_n^{(2)}(\boldsymbol{x},\boldsymbol{y})\\
&= \alpha \Pr\{(X^{(1)n}, Y^{(1)n}) \in B_n\} + (1-\alpha)\Pr\{(X^{(2)n}, Y^{(2)n}) \in B_n\}. \qquad (3.22)
\end{aligned}$$

From (3.20), we can choose a positive number ε such that

$$\begin{aligned}
0 < 3\varepsilon < \ & \max\{|H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}) - H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})|,\\
& |H(\boldsymbol{X}^{(1)}) - H(\boldsymbol{X}^{(2)})|,\ |H(\boldsymbol{Y}^{(1)}) - H(\boldsymbol{Y}^{(2)})|\}. \qquad (3.23)
\end{aligned}$$

Then, we define subsets $T_n^{(i)}$ $(i=1,2)$ of $\mathcal{X}^n \times \mathcal{Y}^n$ by

$$\begin{aligned}
T_n^{(i)} \stackrel{\text{def.}}{=} \Bigg\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \quad & \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - H(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})\right| \le \varepsilon,\\
& \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(i)})\right| \le \varepsilon,\\
& \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(i)})\right| \le \varepsilon, \Bigg\}. \qquad (3.24)
\end{aligned}$$

It should be noted that $T_n^{(i)}$ is characterized not by $P_n^{(i)}(\boldsymbol{x},\boldsymbol{y})$ but by $P_n(\boldsymbol{x},\boldsymbol{y})$. According to Eqs. (3.23)(3.24), it is easy to see that $T_n^{(1)} \cap T_n^{(2)} = \emptyset$, i.e.

$$\Pr\{(X^n, Y^n) \in T_n^{(1)} \cap T_n^{(2)}\} = 0 \qquad (3.25)$$

Then, from Eq. (3.22), we immediately obtain

$$\begin{aligned}
&\Pr\{(X^n, Y^n) \notin T_n^{(1)} \cup T_n^{(2)}\}\\
&= \alpha\Pr\{(X^{(1)n}, Y^{(1)n}) \notin T_n^{(1)} \cup T_n^{(2)}\} + (1-\alpha)\Pr\{(X^{(2)n}, Y^{(2)n}) \notin T_n^{(1)} \cup T_n^{(2)}\}\\
&\le \alpha\Pr\{(X^{(1)n}, Y^{(1)n}) \notin T_n^{(1)}\} + (1-\alpha)\Pr\{(X^{(2)n}, Y^{(2)n}) \notin T_n^{(2)}\}. \qquad (3.26)
\end{aligned}$$

From the definition of $T_n^{(i)}$, we have

$$\begin{aligned}
&\Pr\{(X^{(i)n}, Y^{(i)n}) \notin T_n^{(i)}\}\\
&\le P_n^{(i)}\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - H(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})\right| > \varepsilon\right\}\\
&\quad + P_n^{(i)}\left\{\boldsymbol{x} \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(i)})\right| > \varepsilon\right\}\\
&\quad + P_n^{(i)}\left\{\boldsymbol{y} \in \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(i)})\right| > \varepsilon\right\}. \qquad (3.27)
\end{aligned}$$

According to Lemma 2.3.7 and Lemma 3.4.1 with $\gamma = \varepsilon/2$, for any $\delta > 0$ and sufficiently large n, the first term in Eq. (3.27) can be bounded by

$$
\begin{aligned}
& P_n^{(i)}\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - H(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})\right| > \varepsilon\right\} \\
\leq\ & P_n^{(i)}\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - H(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})\right| > \varepsilon\right. \\
& \qquad \left. \text{and } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - \frac{1}{n}\log\frac{1}{P_n^{(i)}(\boldsymbol{x},\boldsymbol{y})}\right| \leq \max(\varepsilon/2, c_0/n)\right\} \\
& + P_n^{(i)}\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n :\right. \\
& \qquad \left. \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x},\boldsymbol{y})} - \frac{1}{n}\log\frac{1}{P_n^{(i)}(\boldsymbol{x},\boldsymbol{y})}\right| > \max(\varepsilon/2, c_0/n)\right\} \\
\leq\ & P_n^{(i)}\left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n^{(i)}(\boldsymbol{x},\boldsymbol{y})} - H(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})\right| > \varepsilon/2\right\} \\
& + \exp\left(-\frac{n\varepsilon}{2}\right) \\
\leq\ & \delta.
\end{aligned}
$$

In a similar manner, the second and third terms in Eq. (3.27) satisfy

$$
\begin{aligned}
P_n^{(i)}\left\{\boldsymbol{x} \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(i)})\right| > \varepsilon\right\} &\leq \delta, \\
P_n^{(i)}\left\{\boldsymbol{y} \in \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(i)})\right| > \varepsilon\right\} &\leq \delta,
\end{aligned}
$$

for sufficiently large n. Since $\delta > 0$ can be chosen arbitrarily small, the right hand side of Eq. (3.27) vanishes, that is,

$$\lim_{n\to\infty} \Pr\{(X^{(i)n}, Y^{(i)n}) \notin T_n^{(i)}\} = 0 \qquad (i = 1, 2). \tag{3.28}$$

Substituting Eq. (3.28) into Eq. (3.26), we obtain

$$\lim_{n\to\infty} \Pr\{(X^n, Y^n) \notin T_n^{(1)} \cup T_n^{(2)}\} = 0. \tag{3.29}$$

(Step 2) Determination of rate pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$

Suppose that we are given a rate pair (R_1, R_2) which satisfies

$$
\begin{aligned}
R_1 &\geq \alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}), \\
R_2 &\geq \alpha H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}) + (1-\alpha) H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)}), \\
R_1 + R_2 &\geq \alpha H(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)}).
\end{aligned}
$$

It is easy to see that there exists a pair $(\tilde{R}_1, \tilde{R}_2)$, $c_1 \geq 0$ and $c_2 \geq 0$ such that

$$\tilde{R}_1 + \tilde{R}_2 = \alpha H(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}) + (1-\alpha) H(\boldsymbol{X}^{(2)},\boldsymbol{Y}^{(2)})$$

$$\alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}) \le \tilde{R}_1 \le \alpha H(\boldsymbol{X}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}),$$

and (R_1, R_2) can be written as

$$\begin{aligned} R_1 &= \tilde{R}_1 + c_1, \\ R_2 &= \tilde{R}_2 + c_2. \end{aligned}$$

Further, there exists $0 \le \beta \le 1$ such that

$$\begin{aligned} \tilde{R}_1 &= \beta(\alpha H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)})) \\ &\quad +(1-\beta)(\alpha H(\boldsymbol{X}^{(1)}) + (1-\alpha)H(\boldsymbol{X}^{(2)})) \\ \tilde{R}_2 &= \beta(\alpha H(\boldsymbol{Y}^{(1)}) + (1-\alpha)H(\boldsymbol{Y}^{(2)})) \\ &\quad +(1-\beta)(\alpha H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}) + (1-\alpha)H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)})). \end{aligned}$$

By using c_1 and c_2 and β, define two rate pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$ as

$$\begin{aligned} R_1^{(1)} &\stackrel{\text{def.}}{=} \beta H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}) + (1-\beta)H(\boldsymbol{Y}^{(1)}) + c_1, \\ R_2^{(1)} &\stackrel{\text{def.}}{=} \beta H(\boldsymbol{Y}^{(1)}) + (1-\beta)H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}) + c_2, \\ R_1^{(2)} &\stackrel{\text{def.}}{=} \beta H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}) + (1-\beta)H(\boldsymbol{Y}^{(2)}) + c_1, \\ R_2^{(2)} &\stackrel{\text{def.}}{=} \beta H(\boldsymbol{Y}^{(2)}) + (1-\beta)H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)}) + c_2. \end{aligned}$$

We can easily confirm that

$$\alpha R_1^{(1)} + (1-\alpha)R_1^{(2)} = R_1, \tag{3.30}$$
$$\alpha R_2^{(1)} + (1-\alpha)R_2^{(2)} = R_2, \tag{3.31}$$

and

$$\left.\begin{array}{rr} R_1^{(1)} \ge H(\boldsymbol{X}^{(1)}|\boldsymbol{Y}^{(1)}), & R_1^{(2)} \ge H(\boldsymbol{X}^{(2)}|\boldsymbol{Y}^{(2)}), \\ R_2^{(1)} \ge H(\boldsymbol{Y}^{(1)}|\boldsymbol{X}^{(1)}), & R_2^{(2)} \ge H(\boldsymbol{Y}^{(2)}|\boldsymbol{X}^{(2)}), \\ R_1^{(1)} + R_2^{(1)} \ge H(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}), & R_1^{(2)} + R_2^{(2)} \ge H(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)}). \end{array}\right\} \tag{3.32}$$

(Step 3) Construction of wv-SWL code

From Eq. (3.32) and Corollary 3.2.1, we can see $(R_1^{(1)}, R_2^{(1)}) \in \mathcal{R}_{SW}(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)})$ and $(R_1^{(2)}, R_2^{(2)}) \in \mathcal{R}_{SW}(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})$. Hence, there exists an SW code $\{(f_n^{(1)}, f_n^{(2)}, \widehat{f}_n)\}_{n=1}^{\infty}$ for the ergodic source $(\boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)})$, where

$$\begin{aligned} f_n^{(1)} &: \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}(f_n)}, \\ f_n^{(2)} &: \mathcal{Y}^n \to \mathcal{I}_{M_n^{(2)}(f_n)}, \end{aligned}$$

and

$$\left.\begin{array}{l} \displaystyle\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(1)}(f_n) \le R_1^{(1)}, \\ \displaystyle\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(2)}(f_n) \le R_2^{(1)}, \\ \displaystyle\lim_{n\to\infty} \Pr\{\widehat{f}_n(f_n^{(1)}(X^{(1)n}), f_n^{(2)}(Y^{(1)n})) \ne (X^{(1)n}, Y^{(1)n})\} = 0. \end{array}\right\} \tag{3.33}$$

Similarly, there exists an SW code $\{(g_n^{(1)}, g_n^{(2)}, \widehat{g}_n)\}_{n=1}^{\infty}$ for the ergodic source $(\boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)})$, where

$$g_n^{(1)} : \mathcal{X}^n \to \mathcal{I}_{M_n^{(1)}(g_n)},$$
$$g_n^{(2)} : \mathcal{Y}^n \to \mathcal{I}_{M_n^{(2)}(g_n)},$$

and

$$\left.\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(1)}(g_n) \le R_1^{(2)}, \\ &\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(2)}(g_n) \le R_2^{(2)}, \\ &\lim_{n\to\infty} \Pr\{\widehat{g}_n(g_n^{(1)}(X^{(2)n}), g_n^{(2)}(Y^{(2)n})) \ne (X^{(2)n}, Y^{(2)n})\} = 0. \end{aligned}\right\} \tag{3.34}$$

Further, since $\mathcal{X}$ is finite, for any positive integer m, we can easily construct a binary fixed-length code $(\phi_m^{(1)}, \widehat{\phi}_m^{(1)})$ for the source $\boldsymbol{X}$ such that

$$E[\ell(\phi_m^{(1)}(\boldsymbol{x}))] < m \log |\mathcal{X}| + 1 \qquad \forall \boldsymbol{x} \in \mathcal{X}^m,$$
$$\widehat{\phi}_m^{(1)}(\phi_m^{(1)}(\boldsymbol{x})) = \boldsymbol{x} \qquad \forall \boldsymbol{x} \in \mathcal{X}^m.$$

A binary fixed-length code $(\phi_n^{(2)}, \widehat{\phi}_n^{(2)})$ for the source $\boldsymbol{Y}$ can be constructed similarly.

Now, define the sequence of integers $\{N_n\}_{n=1}^{\infty}$ such that

$$0 < N_n \le n, \quad \lim_{n\to\infty} N_n = \infty, \quad \lim_{n\to\infty} \frac{N_n}{n} = 0. \tag{3.35}$$

Then, we construct a wv-SWL code $\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \widehat{\varphi}_n)\}_{n=1}^{\infty}$ as follows:

$\varphi_n^{(11)}$: $\varphi_n^{(11)}(\boldsymbol{x}) \stackrel{\text{def.}}{=} \phi_{N_n}^{(1)}(\boldsymbol{x}_1)$, where $\boldsymbol{x}_1 \in \mathcal{X}^{N_n}$ is the first N_n symbols of $\boldsymbol{x}$.

$\varphi_n^{(21)}$: $\varphi_n^{(21)}(\boldsymbol{y}) \stackrel{\text{def.}}{=} \phi_{N_n}^{(2)}(\boldsymbol{y}_1)$, where $\boldsymbol{y}_1 \in \mathcal{Y}^{N_n}$ is the first N_n symbols of $\boldsymbol{y}$.

$\varphi_n^{(12)}$: Decode $\boldsymbol{y}_1$ from a given codeword $\varphi_n^{(21)}(\boldsymbol{y})$, then assign the codeword in the following manner:

If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}$, then $\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})) \stackrel{\text{def.}}{=} \varphi_n^{(21)}(\boldsymbol{y}) * f_n^{(1)}(\boldsymbol{x})$.
If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}$, then $\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})) \stackrel{\text{def.}}{=} \varphi_n^{(21)}(\boldsymbol{y}) * g_n^{(1)}(\boldsymbol{x})$.
Otherwise, $\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})) \stackrel{\text{def.}}{=} \lambda$.

$\varphi_n^{(22)}$: Decode $\boldsymbol{x}_1$ from a given codeword $\varphi_n^{(11)}(\boldsymbol{x})$, then assign the codeword in the following manner:

If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}$, then $\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x})) \stackrel{\text{def.}}{=} \varphi_n^{(11)}(\boldsymbol{x}) * f_n^{(2)}(\boldsymbol{y})$.
If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}$, then $\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x})) \stackrel{\text{def.}}{=} \varphi_n^{(11)}(\boldsymbol{x}) * g_n^{(2)}(\boldsymbol{y})$.
Otherwise, $\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x})) \stackrel{\text{def.}}{=} \lambda$.

$\widehat{\varphi}_n$: Since both $\varphi_n^{(11)}$ and $\varphi_n^{(21)}$ are prefix codes, for given $\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y}))$ and $\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x}))$, we have $\varphi_n^{(11)}(\boldsymbol{x})$, $\varphi_n^{(21)}(\boldsymbol{y})$, and either $(f_n^{(1)}(\boldsymbol{x}), f_n^{(2)}(\boldsymbol{y}))$ or $(g_n^{(1)}(\boldsymbol{x}), g_n^{(2)}(\boldsymbol{y}))$. We first decode $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ from $(\varphi_n^{(11)}(\boldsymbol{x}), \varphi_n^{(21)}(\boldsymbol{y}))$ and, output an estimate $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in \mathcal{X}^n \times \mathcal{Y}^n$ in the following manner:

If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}$, then $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = \widehat{f}_n(f_n^{(1)}(\boldsymbol{x}), f_n^{(2)}(\boldsymbol{y}))$.
If $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}$, then $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = \widehat{g}_n(g_n^{(1)}(\boldsymbol{x}), g_n^{(2)}(\boldsymbol{y}))$.
Otherwise, we declare an error.

The proposed code cannot reproduce the original sequence pair in the following cases:

(i) $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}$ and $\widehat{f}_n(f_n^{(1)}(\boldsymbol{x}), f_n^{(2)}(\boldsymbol{y})) \neq (\boldsymbol{x}, \boldsymbol{y})$

(ii) $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}$ and $\widehat{g}_n^{-1}(g_n^{(1)}(\boldsymbol{x}), g_n^{(2)}(\boldsymbol{y})) \neq (\boldsymbol{x}, \boldsymbol{y})$

(iii) $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)}$ or $(\boldsymbol{x}_1, \boldsymbol{y}_1) \notin T_{N_n}^{(1)} \cup T_{N_n}^{(2)}$

According to Eqs. (3.1)(3.22)(3.28), the probability of the event (i) can be bounded as

$$
\begin{aligned}
&\Pr\{(X^{N_n}, Y^{N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c} \text{ and } \widehat{f}_n(f_n^{(1)}(X^n), f_n^{(2)}(Y^n)) \neq (X^n, Y^n)\} \\
&\quad = \alpha \Pr\{(X^{(1)N_n}, Y^{(1)N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c} \\
&\qquad \text{and } \widehat{f}_n(f_n^{(1)}(X^{(1)n}), f_n^{(2)}(Y^{(1)n})) \neq (X^{(1)n}, Y^{(1)n})\} \\
&\qquad +(1-\alpha)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c} \\
&\qquad \text{and } \widehat{f}_n(f_n^{(1)}(X^{(2)n}), f_n^{(2)}(Y^{(2)n})) \neq (X^{(2)n}, Y^{(2)n})\} \\
&\quad \leq \alpha \Pr\{\widehat{f}_n(f_n^{(1)}(X^{(1)n}), f_n^{(2)}(Y^{(1)n})) \neq (X^{(1)n}, Y^{(1)n})\} \\
&\qquad +(1-\alpha)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}\} \\
&\quad \leq \alpha \Pr\{\widehat{f}_n(f_n^{(1)}(X^{(1)n}), f_n^{(2)}(Y^{(1)n})) \neq (X^{(1)n}, Y^{(1)n})\} \\
&\qquad +(1-\alpha)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \notin T_{N_n}^{(2)}\} \\
&\quad \to 0 \qquad (n \to \infty).
\end{aligned}
$$

In a similar manner, the probability of the event (ii) vanishes as $n \to \infty$. Further, the probability of the event (iii) vanishes as $n \to \infty$ due to Eqs. (3.25)(3.29). Therefore, the probability of decoding error vanishes as $n \to \infty$.

(Step 4) Evaluation of the average length of codeword

Lastly, we investigate the average length of codeword for the proposed code. For any $\delta > 0$ and sufficiently large n, according to Eqs. (3.22)(3.33)(3.34), we have

$$
\begin{aligned}
&\frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] \\
&\quad = \frac{1}{n} E[l(\varphi_n^{(11)}(X^n))] \\
&\qquad + \left(\frac{1}{n} \log M_n^{(1)}(f_n)\right) \times \Pr\{(X^{N_n}, Y^{N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}\}
\end{aligned}
$$

$$
\begin{aligned}
& + \left(\frac{1}{n}\log M_n^{(1)}(g_n)\right) \times \Pr\{(X^{N_n}, Y^{N_n}) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}\} \\
\le\ & \frac{1}{n}(N_n \log|\mathcal{X}| + 1) \\
& +(R_1^{(1)} + \delta)\Pr\{(X^{N_n}, Y^{N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}\} \\
& +(R_1^{(2)} + \delta)\Pr\{(X^{N_n}, Y^{N_n}) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}\} \\
=\ & \frac{1}{n}(N_n \log|\mathcal{X}| + 1) \\
& +\alpha(R_1^{(1)} + \delta)\Pr\{(X^{(1)N_n}, Y^{(1)N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}\} \\
& +(1-\alpha)(R_1^{(1)} + \delta)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \in T_{N_n}^{(1)} \cap T_{N_n}^{(2)c}\} \\
& +\alpha(R_1^{(2)} + \delta)\Pr\{(X^{(1)N_n}, Y^{(1)N_n}) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}\} \\
& +(1-\alpha)(R_1^{(2)} + \delta)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \in T_{N_n}^{(2)} \cap T_{N_n}^{(1)c}\} \\
\le\ & \frac{1}{n}(N_n \log|\mathcal{X}| + 1) \\
& +\alpha(R_1^{(1)} + \delta)\Pr\{(X^{(1)N_n}, Y^{(1)N_n}) \in T_{N_n}^{(1)}\} \\
& +(1-\alpha)(R_1^{(1)} + \delta)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \notin T_{N_n}^{(2)}\} \\
& +\alpha(R_1^{(2)} + \delta)\Pr\{(X^{(1)N_n}, Y^{(1)N_n}) \notin T_{N_n}^{(1)}\} \\
& +(1-\alpha)(R_1^{(2)} + \delta)\Pr\{(X^{(2)N_n}, Y^{(2)N_n}) \in T_{N_n}^{(2)}\}.
\end{aligned}
$$

By using Eqs. (3.28)(3.35), we obtain

$$
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] \le \alpha R_1^{(1)} + (1-\alpha)R_1^{(2)} + \delta.
$$

In a similar manner,

$$
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(22)}(Y^n, \varphi_n^{(11)}(X^n)))] \le \alpha R_2^{(1)} + (1-\alpha)R_2^{(2)} + \delta.
$$

Since $\delta > 0$ is arbitrary, we can conclude

$$
\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(12)}(X^n, \varphi_n^{(21)}(Y^n)))] &\le \alpha R_1^{(1)} + (1-\alpha)R_1^{(2)} = R_1, \\
\limsup_{n\to\infty} \frac{1}{n} E[l(\varphi_n^{(22)}(Y^n, \varphi_n^{(11)}(X^n)))] &\le \alpha R_2^{(1)} + (1-\alpha)R_2^{(2)} = R_2,
\end{aligned}
$$

where the last equalities come from Eqs. (3.30)(3.31). Hence, (R_1, R_2) is achievable for the wv-SWL system. This completes the proof of Theorem 3.3.1. □

The extension of the theorem to the mixture of any finite number of ergodic sources is immediate. Further, the proof of Corollary 3.3.2 can be done similarly by considering a fixed length SW code for each source $(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})$ with a rate pair $(R_1^{(i)}, R_2^{(i)})$ $(i = 1, 2, \cdots, m)$ instead of $\{(f_n^{(1)}, f_n^{(2)}, \widehat{f}_n)\}_{n=1}^{\infty}$ and $\{(g_n^{(1)}, g_n^{(2)}, \widehat{g}_n)\}_{n=1}^{\infty}$ in Step 2 and Step 3.

3.4.3 Proof of Theorem 3.3.2

Proof. By choosing a positive number ε such that

$$0 < 3\varepsilon < \min\{|H(\boldsymbol{X}^{(1)}) - H(\boldsymbol{X}^{(2)})|, |H(\boldsymbol{Y}^{(1)}) - H(\boldsymbol{Y}^{(2)})|\}$$

instead of (3.23), we have

$$\left\{\boldsymbol{x} \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(1)})\right| \le \varepsilon \text{ and } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(2)})\right| \le \varepsilon\right\}$$
$$\left\{\boldsymbol{y} \in \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(1)})\right| \le \varepsilon \text{ and } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(2)})\right| \le \varepsilon\right\},$$

are both empty sets, which implies that we can determine whether $(\boldsymbol{x}, \boldsymbol{y}) \in T_n^{(1)}$ or $(\boldsymbol{x}, \boldsymbol{y}) \in T_n^{(2)}$ by seeing only $\boldsymbol{x}$ or $\boldsymbol{y}$. Hence, by using the codes $\{(f_n^{(1)}, f_n^{(2)}, \widehat{f}_n)\}_{n=1}^{\infty}$ and $\{(g_n^{(1)}, g_n^{(2)}, \widehat{g}_n)\}_{n=1}^{\infty}$ described in the proof of Theorem 3.3.1, we construct a wv-SWL code $\{(\varphi_n^{(11)}, \varphi_n^{(12)}, \varphi_n^{(21)}, \varphi_n^{(22)}, \varphi_n^{-1})\}_{n=1}^{\infty}$ as follows:

$$\varphi_n^{(11)}(\boldsymbol{x}) \stackrel{\text{def.}}{=} \lambda$$
$$\varphi_n^{(21)}(\boldsymbol{y}) \stackrel{\text{def.}}{=} \lambda$$
$$\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})) \stackrel{\text{def.}}{=} \begin{cases} 0 * f_n^{(1)}(\boldsymbol{x}) & \text{if } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(1)})\right| \le \varepsilon, \\ 1 * g_n^{(1)}(\boldsymbol{x}) & \text{if } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{x})} - H(\boldsymbol{X}^{(2)})\right| \le \varepsilon, \\ \lambda & \text{otherwise,} \end{cases}$$
$$\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x})) \stackrel{\text{def.}}{=} \begin{cases} 0 * f_n^{(2)}(\boldsymbol{y}) & \text{if } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(1)})\right| \le \varepsilon, \\ 1 * g_n^{(2)}(\boldsymbol{y}) & \text{if } \left|\frac{1}{n}\log\frac{1}{P_n(\boldsymbol{y})} - H(\boldsymbol{Y}^{(2)})\right| \le \varepsilon, \\ \lambda & \text{otherwise.} \end{cases}$$

Lastly we describe the decoder $\widehat{\varphi}_n$. For a given pair of codewords $s_1 * s_2 = \varphi_n^{(12)}(\boldsymbol{x}, \lambda)$ and $s_3 * s_4 = \varphi_n^{(22)}(\boldsymbol{y}, \lambda)$ with $s_1, s_3 \in \mathcal{B}$ and $s_2, s_4 \in \mathcal{B}^*$, we output an estimate $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in \mathcal{X}^n \times \mathcal{Y}^n$ as follows:

If $s_1 = s_3 = 0$ then $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = \widehat{f}_n(s_2, s_4)$.
If $s_1 = s_3 = 1$ then $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = \widehat{g}_n(s_2, s_4)$.
Otherwise, we declare an error.

By using this code, in a similar manner as the proof of Theorem 3.3.1, we can show that the rate pair (R_1, R_2) is achievable for the wv-SWL system. □

3.4.4 Proof of Theorem 3.3.3

Proof. We only show the construction of universal code. The proof of the theorem can be done in a similar manner as the proof of Theorem 3.3.1.

The encoder $\varphi_n^{(11)}$ and $\varphi_n^{(21)}$ are the same ones which defined in the proof of Theorem 3.3.1. After sharing the pair (x^{N_n}, y^{N_n}) of sequences between two encoders, each encoder calculates the joint type $Q_{XY} = Q_{x^{N_n} y^{N_n}}$ of the shared sequence. For any $\delta > 0$ and any type $Q \in \mathcal{P}_{N_n}(\mathcal{X} \times \mathcal{Y})$, there exists a *universal* SW code $(f_Q^{(1)}, f_Q^{(2)}, \widehat{f}_Q)$ with block length n and a rate pair $(R_1(Q)+\delta, R_2(Q)+\delta)$ [13]. In the above notation, it should be noted that the joint probability Q determines only the rate of the universal SW code. Then, we define the second encoders as

$$\begin{aligned}
\varphi_n^{(12)}(\boldsymbol{x}, \varphi_n^{(21)}(\boldsymbol{y})) &= \varphi_n^{(21)}(\boldsymbol{y}) * f_{Q_{XY}}^{(1)}(\boldsymbol{x}), \\
\varphi_n^{(22)}(\boldsymbol{y}, \varphi_n^{(11)}(\boldsymbol{x})) &= \varphi_n^{(11)}(\boldsymbol{x}) * f_{Q_{XY}}^{(2)}(\boldsymbol{y}),
\end{aligned}$$

where $Q = Q_{x^{N_n} y^{N_n}}$ denotes the type of the shared sequence. This implies that we choose the universal SW code depending on the joint type Q. Since the decoder can have the knowledge of the type Q, it can output the estimate $\widehat{f}_Q(f_Q^{(1)}(\boldsymbol{x}), f_Q^{(2)}(\boldsymbol{y}))$.

□

3.4.5 Proof of Lemma 3.4.1

Proof. It is easy to see that for any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ and $i = 1, 2$

$$\frac{1}{n} \log \frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} \leq \frac{1}{n} \log \frac{1}{\min(\alpha, 1-\alpha) P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} = \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} + \frac{c_0}{n}. \tag{3.36}$$

On the other hand, for $i = 1, 2$ we have

$$\begin{aligned}
& P_n^{(i)} \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} \log \frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} - \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} \leq -\gamma \right\} \\
&= \sum_{\substack{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y}) \leq P_n(\boldsymbol{x}, \boldsymbol{y}) \exp(-n\gamma)}} P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y}) \\
&\leq \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n :} P_n(\boldsymbol{x}, \boldsymbol{y}) \exp(-n\gamma) \\
&\leq \exp(-n\gamma).
\end{aligned} \tag{3.37}$$

Then, by combining (3.36) and (3.37), we obtain

$$\begin{aligned}
& P_n^{(i)} \Big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\
& \qquad \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} - \gamma \leq \frac{1}{n} \log \frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} \leq \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} + \frac{c_0}{n} \Big\} \\
&= P_n^{(i)} \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} \log \frac{1}{P_n(\boldsymbol{x}, \boldsymbol{y})} - \frac{1}{n} \log \frac{1}{P_n^{(i)}(\boldsymbol{x}, \boldsymbol{y})} \geq -\gamma \right\} \\
&\geq 1 - \exp(-n\gamma).
\end{aligned}$$

This completes the proof of Lemma 3.4.1. □

Chapter 4

Complementary delivery of information

4.1 Introduction

This chapter deals with a multiterminal source coding system that we call *complementary delivery coding system*. Figure 4.1 shows a block diagram of the complementary delivery problem. The encoder observes messages emitted from two correlated sources, and delivers these messages to other locations (i.e. decoders). Each decoder has access to one of two messages, and therefore wants to reproduce the other message. This coding system models a communication network via a satellite [72]. The satellite processes messages from two users and broadcasts one common message to the users. Each user reconstructs the data of the other user based on the broadcast message and its own stored data. Csiszár and Körner considered a general framework that includes the complementary delivery coding system, and clarified the minimum achievable rate for lossless coding. Figure 4.1 can be rewritten in the format of Csiszár and Körner, as shown in Figure 4.2. Willems et al. [66, 72] independently investigated the lossless coding problem where several users are physically separated but communicate with each other via a satellite, and determined the minimum coding rate in transmitting to and from the satellite. The complementary delivery coding system is a special case of the system of Willems et al. which only considers the downstream part from the satellite. This chapter extends their results to lossy coding, and clarify the rate-distortion function and some interesting properties.

An extended version of complementary delivery coding is also examined (Fig. 4.3), where only an encoded sequence of a message is available as side information at each decoder. The coding system models a sensor network with three physically separated sensors and two destinations. Two of these three sensors can observe one of two information sources and transmit individual sensor readings to one of two destinations, while the other sensor can observe both of these two sources and transmit the sensor reading to both destinations. Both lossless and lossy configurations are considered in this setting. Inner and outer bounds of the achievable rate region for given distortion criteria will be clarified.

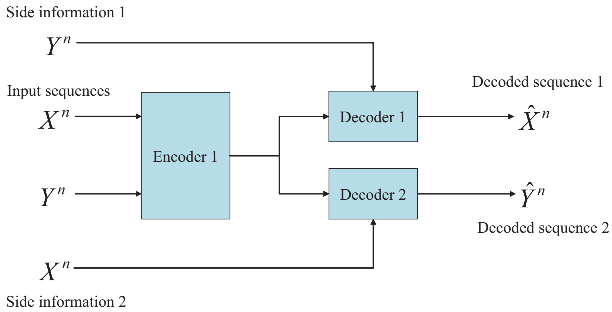

Figure 4.1: The complementary delivery coding system (Full side information is available at decoders)

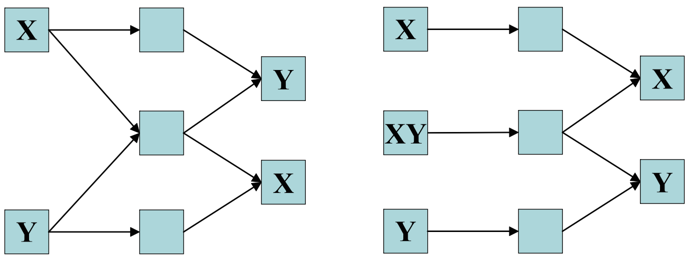

Figure 4.2: The complementary delivery coding system in the format of Csiszár and Körner (Left) original source network (Right) normal source network

4.2 Problem formulation

First, this section formulates a coding problem of the complementary delivery coding system, and shows a fundamental bound of the coding rate when considering lossless coding.

Definition 4.2.1. (Fixed-to-fixed fully-informed complementary delivery (f-FCD) code)
A sequence $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ of codes $(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})$ is an f-FCD code if and only if

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

Definition 4.2.2. (f-FCD achievable rate)
R is an f-FCD achievable rate of the source $(\boldsymbol{X}, \boldsymbol{Y})$ for a given pair (D_1, D_2) of distortion criteria if and only if there exists an f-FCD code such that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R, \\
&\limsup_{n\to\infty} \frac{1}{n} e_n^{(1)} \le D_1, \ \ \limsup_{n\to\infty} \frac{1}{n} e_n^{(2)} \le D_2,
\end{aligned}$$

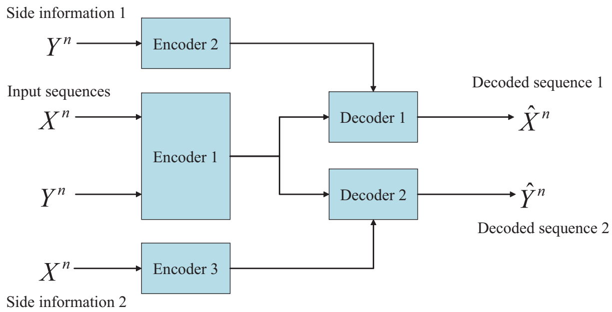

Figure 4.3: Complementary delivery coding (Partial side information is available at decoders)

where

$$
\begin{aligned}
e_n^{(1)} &\stackrel{\text{def.}}{=} E\left[d_X^n(X^n, \widehat{X}^n)\right], \\
e_n^{(2)} &\stackrel{\text{def.}}{=} E\left[d_Y^n(Y^n, \widehat{Y}^n)\right], \\
\widehat{X}^n &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(1)}(\varphi_n(X^n, Y^n), Y^n), \\
\widehat{Y}^n &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(2)}(\varphi_n(X^n, Y^n), X^n).
\end{aligned}
$$

Definition 4.2.3. (Inf f-FCD achievable rate)

$$
\begin{aligned}
&R_F(\boldsymbol{X}, \boldsymbol{Y}|D_1, D_2) \\
&\stackrel{\text{def.}}{=} \inf\{R : R \text{ is an f-FCD achievable rate of } (\boldsymbol{X}, \boldsymbol{Y}) \text{ for } (D_1, D_2)\}.
\end{aligned}
$$

Definition 4.2.4. (Lossless f-FCD achievable rate)
Let $\widehat{\mathcal{X}} = \mathcal{X}$ and $\widehat{\mathcal{Y}} = \mathcal{Y}$. Define the distortion function $d_X^n : \mathcal{X}^n \times \widehat{\mathcal{X}}^n \to \{0, 1\}$ is defined as

$$
d_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \begin{cases} 0 & \text{if } \boldsymbol{x} = \widehat{\boldsymbol{x}}, \\ 1 & \text{if } \boldsymbol{x} \neq \widehat{\boldsymbol{x}}. \end{cases}
$$

$d_Y^n : \mathcal{Y}^n \times \widehat{\mathcal{Y}}^n \to \{0, 1\}$ is defined similarly. Then, R is a lossless f-FCD achievable rate of the source $(\boldsymbol{X}, \boldsymbol{Y})$ if R is an f-FCD achievable rate of the source $(\boldsymbol{X}, \boldsymbol{Y})$ for a pair $(D_1 = 0, D_2 = 0)$ of distortion criteria.

Definition 4.2.5. (Inf lossless f-FCD achievable rate)

$$
R_F(\boldsymbol{X}, \boldsymbol{Y}) \stackrel{\text{def.}}{=} \inf\{R : R \text{ is an lossless f-FCD achievable rate of } (\boldsymbol{X}, \boldsymbol{Y})\}.
$$

As a special case of the results of Csiszár, Körner [12], and Willems et al. [66, 72], we can immediately obtain the closed form of $R_F(\boldsymbol{X}, \boldsymbol{Y})$.

Theorem 4.2.1. (Coding theorem of lossless f-FCD code)
For any DMS (X, Y), we have

$$
R_F(X, Y) = \max\{H(X|Y), H(Y|X)\}
$$

4.3 Main results

Theorem 4.3.1. (Coding theorem of lossy f-FCD code)
For any DMS (X, Y)*,*

$$R_F(X,Y|D_1,D_2) = \min_{P_{U|XY} \in \mathcal{P}_F(\mathcal{U}|P_{XY})} \left[\max\{I(X;U|Y), I(Y;U|X)\}\right],$$

where the alphabet $\mathcal{U}$ *satisfies* $|\mathcal{U}| \le |\mathcal{X} \times \mathcal{Y}| + 2$ *and* $\mathcal{P}_F(\mathcal{U}|P_{XY})$ *is a subset of* $\mathcal{P}(\mathcal{U}|P_{XY})$ *such that there exist functions* $\phi^{(1)} : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ *and* $\phi^{(2)} : \mathcal{U} \times \mathcal{X} \to \widehat{\mathcal{Y}}$ *satisfying*

$$\begin{aligned} D_1 &\ge E\left[\Delta_X(X, \phi^{(1)}(U,Y))\right], \\ D_2 &\ge E\left[\Delta_Y(Y, \phi^{(2)}(U,X))\right]. \end{aligned}$$

From the definition of $R_F(X,Y|D_1,D_2)$ and Theorem 4.3.1, the following corollary can be easily obtained:

Corollary 4.3.1.

$$\begin{aligned} &\max\{R_C(X|Y,D_1), R_C(Y|X,D_2)\} \\ &\quad \le R_F(X,Y|D_1,D_2) \\ &\quad \le \max\{R_{WV}(X|Y,D_1), R_{WV}(Y|X,D_2)\}, \end{aligned} \tag{4.1}$$

where $R_C(X|Y,D)$ *is the conditional rate-distortion function defined in Lemma 2.1.8, and* $R_{WV}(X|Y,D)$ *is the Wyner-Ziv rate-distortion function defined in Lemma 2.1.9.*

When one of the two messages does not need to be reproduced, $R_F(X,Y|D_1,D_2)$ is always reduced to the conditional rate-distortion function, namely

$$\begin{aligned} R_F(X,Y|D_1 = d_1, D_2) &= R_{C1}(Y|X,D_2), \\ R_F(X,Y|D_2, D_2 = d_2) &= R_{C1}(X|Y,D_1) \end{aligned}$$

if $d_1 \ge \overline{\Delta}_X$ and $d_2 \ge \overline{\Delta}_Y$, where

$$\begin{aligned} \overline{\Delta}_X &\stackrel{\text{def.}}{=} \max_{(x,\hat{x}) \in \mathcal{X} \times \widehat{\mathcal{X}}} \Delta_X(x,\widehat{x}) < \infty, \\ \overline{\Delta}_Y &\stackrel{\text{def.}}{=} \max_{(y,\hat{y}) \in \mathcal{Y} \times \widehat{\mathcal{Y}}} \Delta_Y(y,\widehat{y}) < \infty. \end{aligned}$$

The above discussion indicates that there may be some rate losses when sending common codewords to two decoders only for lossy coding. These properties result from the auxiliary random variable U, which characterizes the rate-distortion function $R_F(X,Y|D_1,D_2)$, although messages used as side information are available at both the encoder and decoder.

Theorem 4.3.1 has considered only two sources. However, these propositions can be easily extended to any finite number of sources. A set of distortion criteria is denoted by

$$\boldsymbol{D} = \{D_i^{(j)}\}_{j \in \mathcal{I}_{N_d}, i \in \mathcal{S}_j},$$

where $\mathcal{S}_j \subseteq \mathcal{I}_{N_s}$, N_s is the number of sources and N_d is the number of decoders. The inf f-FCD achievable rate $R_F(\boldsymbol{X}|\boldsymbol{D})$ of the source $\boldsymbol{X} = \boldsymbol{X}^{(\mathcal{I}_{N_s})}$ for given distortion criteria $\boldsymbol{D}$ is defined in the same way as for two correlated sources. Each decoder $\widehat{\varphi}_n^{(j)}$ has access to $\boldsymbol{X}^{(\mathcal{S}_j^c)}$ as side information and therefore has to recover $\boldsymbol{X}^{(\mathcal{S}_j)}$ $(j \in \mathcal{I}_{N_d}, \mathcal{S}_j \subseteq \mathcal{I}_{N_s})$ such that the following criteria are satisfied for any $i \in \mathcal{I}_{N_s}$ and $j \in \mathcal{I}_{N_d}$:

$$D_i^{(j)} \quad \geq \quad E\left[\Delta_{X^{(i)}}(X^{(i)}, \widehat{X}^{(i;j)})\right], \tag{4.2}$$

where

$$\begin{aligned}\widehat{\boldsymbol{X}}^{(\mathcal{S}_j)} &\stackrel{\text{def.}}{=} \{\widehat{X}^{(i;j)}\}_{i\in\mathcal{S}_j} \\ &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(j)}(\varphi_n(\boldsymbol{X}^n), \boldsymbol{X}^{(\mathcal{S}_j^c)}).\end{aligned}$$

Corollary 4.3.2. (Coding theorem of f-FCD code for any finite number of sources)
(1) *If there exists a pair $(j_1, j_2) \in \mathcal{I}_{N_d} \times \mathcal{I}_{N_d}$ such that $\mathcal{S}_{j_1} \not\subseteq \mathcal{S}_{j_2}$ and $\mathcal{S}_{j_1} \not\supseteq \mathcal{S}_{j_2}$, the inf f-FCD achievable rate for a DMS $\boldsymbol{X} = \boldsymbol{X}^{(\mathcal{I}_{N_s})}$ can be obtained as*

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

where the alphabet $\mathcal{U}$ satisfies $|\mathcal{U}| \leq |\mathcal{X}^{(\mathcal{I}_{N_s})}| + N_d$, and $\mathcal{P}_F(\mathcal{U}|P_{\boldsymbol{X}}) \subseteq \mathcal{P}(\mathcal{U}|P_{\boldsymbol{X}})$ is a set of probability distributions such that there exist functions $\phi^{(j,i)} : \mathcal{U} \times \mathcal{X}^{(\mathcal{S}_j^c)} \to \widehat{\mathcal{X}}^{(i)}$ for $i \in \mathcal{S}_j$ that satisfy

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

(2) *Otherwise, the inf CD-achievable rate is*

$$\begin{aligned}&R_F(\boldsymbol{X}|\boldsymbol{D}) \\ &= \min_{P_{\widehat{\boldsymbol{X}}|\boldsymbol{X}} \in \mathcal{P}_F(\widehat{\mathcal{X}}^{(\mathcal{I}_{N_s})}|P_{\boldsymbol{X}})} \max_{j\in\mathcal{I}_{N_d}} I\left(\boldsymbol{X}^{(\mathcal{S}_j)}; \widehat{\boldsymbol{X}}^{(\mathcal{S}_j)} \middle| \boldsymbol{X}^{(\mathcal{S}_j^c)}\right),\end{aligned}$$

where $\mathcal{P}_F(\widehat{\mathcal{X}}^{(\mathcal{I}_{N_s})}|P_{\boldsymbol{X}}) \subseteq \mathcal{P}(\widehat{\mathcal{X}}^{(\mathcal{I}_{N_s})}|P_{\boldsymbol{X}})$ is a set of probability distributions such that the distortion criteria Eq.(4.2) are satisfied.

As a typical example, Corollary 4.3.2 can apply to the coding problem formulated by Willems et al. [72], where the encoder sends three messages to three decoders, and each decoder has access to one of three messages to reproduce the other two messages. Corollary 4.3.2 indicates that the inf achievable rate for this coding problem is obtained as

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

where the alphabet $\mathcal{U}$ satisfies $|\mathcal{U}| \leq |\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}| + 3$, and $\mathcal{P}_F(\mathcal{U}|P_{XYZ}) \subseteq \mathcal{P}(\mathcal{U}|P_{XYZ})$ is a set of probability distributions such that there exist functions

$$\phi^{(1,Y)} : \mathcal{U} \times \mathcal{X} \to \widehat{\mathcal{Y}}, \quad \phi^{(1,Z)} : \mathcal{U} \times \mathcal{X} \to \widehat{\mathcal{Z}},$$

$$\phi^{(2,X)}: \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}, \quad \phi^{(2,Z)}: \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{Z}},$$
$$\phi^{(3,X)}: \mathcal{U} \times \mathcal{Z} \to \widehat{\mathcal{X}}, \quad \phi^{(3,Y)}: \mathcal{U} \times \mathcal{Z} \to \widehat{\mathcal{Y}}$$

that satisfy

$$\begin{aligned}
D_Y^{(1)} &\geq E[\Delta_Y(Y, \phi^{(1,Y)}(U,X))], \\
D_Z^{(1)} &\geq E[\Delta_Z(Z, \phi^{(1,Z)}(U,X))], \\
D_X^{(2)} &\geq E[\Delta_X(X, \phi^{(2,X)}(U,Y))], \\
D_Z^{(2)} &\geq E[\Delta_Z(Z, \phi^{(2,Z)}(U,Y))], \\
D_X^{(3)} &\geq E[\Delta_X(X, \phi^{(3,X)}(U,Z))], \\
D_Y^{(3)} &\geq E[\Delta_Y(Y, \phi^{(3,Y)}(U,Z))].
\end{aligned}$$

4.4 Complementary delivery with partial side information

4.4.1 Problem formulation

This section examines an extended version of the complementary delivery coding, where only an encoded sequence of a message is available as side information at each decoder.

Definition 4.4.1. (Fixed-length partially-informed complementary delivery (f-PCD) code)
A sequence

$$\{(\varphi_n^{(1)}, \varphi_n^{(2)}, \varphi_n^{(3)}, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$$

of codes

$$(\varphi_n^{(1)}, \varphi_n^{(2)}, \varphi_n^{(3)}, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})$$

is an f-PCD code if and only if

$$\begin{aligned}
\varphi_n^{(1)} &: \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{I}_{M_n^{(1)}}, \\
\varphi_n^{(2)} &: \mathcal{Y}^n \to \mathcal{I}_{M_n^{(2)}}, \quad \varphi_n^{(3)}: \mathcal{X}^n \to \mathcal{I}_{M_n^{(3)}}, \\
\widehat{\varphi}_n^{(1)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{I}_{M_n^{(2)}} \to \widehat{\mathcal{X}}^n, \\
\widehat{\varphi}_n^{(2)} &: \mathcal{I}_{M_n^{(1)}} \times \mathcal{I}_{M_n^{(3)}} \to \widehat{\mathcal{Y}}^n.
\end{aligned}$$

Definition 4.4.2. (f-PCD achievable rate triad)
(R_1, R_2, R_3) is an f-PCD achievable rate triad of the source $(\boldsymbol{X}, \boldsymbol{Y})$ for a pair (D_1, D_2) of distortion criteria if and only if there exists an f-PCD code such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^{(i)} \leq R_i, \ (i = 1, 2, 3)$$
$$\limsup_{n\to\infty} \frac{1}{n} e_n^{(1)} \leq D_1, \ \limsup_{n\to\infty} \frac{1}{n} e_n^{(2)} \leq D_2.$$

where

$$\begin{aligned} e_n^{(1)} &= E\left[\Delta_X^n(X^n, \widehat{\varphi}_n^{(1)}(A_n^{(1)}, A_n^{(2)}))\right], \\ e_n^{(2)} &= E\left[\Delta_Y^n(Y^n, \widehat{\varphi}_n^{(2)}(A_n^{(1)}, A_n^{(3)}))\right], \\ A_n^{(1)} &= \varphi_n^{(1)}(X^n, Y^n), \\ A_n^{(2)} &= \varphi_n^{(2)}(Y^n), \quad A_n^{(3)} = \varphi_n^{(3)}(X^n). \end{aligned}$$

Definition 4.4.3. (f-PCD achievable rate region)

$$\begin{aligned} &\mathcal{R}_P(\boldsymbol{X}, \boldsymbol{Y}|D_1, D_2) \\ &\quad = \{(R_1, R_2, R_3) : \\ &\qquad (R_1, R_2, R_3) \text{ is an f-PCD achievable rate triad of } (\boldsymbol{X}, \boldsymbol{Y}) \text{ for } (D_1, D_2)\}. \end{aligned}$$

A lossless f-PCD achievable rate triad (resp. a lossless f-PCD achievable rate region $\mathcal{R}_P(\boldsymbol{X}, \boldsymbol{Y})$) is defined similarly to the lossless f-FCD achievable rate (resp. the inf lossless f-FCD achievable rate $R_F(\boldsymbol{X}, \boldsymbol{Y})$).

4.4.2 Statement of results

As a special case of the results of Csiszár and Körner, we can obtain the closed form of the lossless f-PCD achievable rate region.

Theorem 4.4.1. (Coding theorem of lossless f-PCD code)
For any DMS (X, Y)

$$\begin{aligned} &\mathcal{R}_P(X, Y) = \\ &\quad \big\{ \ (R_1, R_2, R_3) : \exists P_{VW|XY} \in \mathcal{P}_P(\mathcal{V} \times \mathcal{W}|P_{XY}), \\ &\qquad R_1 \geq \max\{H(X|V), I(Y|W)\}, \ R_2 \geq I(Y; V), \ R_3 \geq I(X; W)\big\}, \end{aligned}$$

where $\mathcal{P}_P(\mathcal{V} \times \mathcal{W}|P_{XY})$ is a subset of $\mathcal{P}(\mathcal{V} \times \mathcal{W}|P_{XY})$ such that $V \to Y \to X$ and $W \to X \to Y$ form Markov chains.

The main result of this subsection is the following, which shows inner and outer bounds of the f-PCD achievable rate region $\mathcal{R}_P(\boldsymbol{X}, \boldsymbol{Y}|D_1, D_2)$.

Theorem 4.4.2. (Coding theorem of lossy f-PCD code)
For any DMS (X, Y)

$$\begin{aligned} &\mathcal{R}_P(X, Y|D_1, D_2) \subseteq \quad \textit{(outer bound)} \\ &\quad \big\{ \ (R_1, R_2, R_3) : \exists P_{UVW|XY} \in \underline{\mathcal{P}}_P(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY}), \\ &\qquad R_1 \geq \max\{I(X; U|V), I(Y; U|W)\}, \ R_2 \geq I(Y; V), \ R_3 \geq I(X; W)\big\}, \\ &\mathcal{R}_P(X, Y|D_1, D_2) \supseteq \quad \textit{(inner bound)} \\ &\quad \big\{ \ (R_1, R_2, R_3) : \exists P_{UVW|XY} \in \overline{\mathcal{P}}_P(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY}), \\ &\qquad R_1 \geq \max\{I(VX; U), I(WY; U)\} - \min\{I(V; U), I(W; U)\}, \\ &\qquad R_2 \geq I(Y; V), \ R_3 \geq I(X; W)\big\}, \end{aligned}$$

where $\underline{\mathcal{P}}_P(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY})$ is a subset of $\mathcal{P}(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY})$ such that $V \to Y \to X$ and $W \to X \to Y$ form Markov chains and there exist functions $\phi^{(1)} : \mathcal{U} \times \mathcal{V} \to \widehat{\mathcal{X}}$ and $\phi^{(2)} : \mathcal{U} \times \mathcal{W} \to \widehat{\mathcal{Y}}$ satisfying

$$D_1 \geq E\left[\Delta_X(X, \phi^{(1)}(U, V))\right],$$

$$D_2 \geq E\left[\Delta_Y(Y, \phi^{(2)}(U, W))\right],$$

and $\overline{\mathcal{P}}_P(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY})$ *is a subset of* $\underline{\mathcal{P}}_P(\mathcal{U} \times \mathcal{V} \times \mathcal{W}|P_{XY})$ *whose Markov chains are replaced to* $V \to Y \to X \to W$*.*

Remark 4.4.1. *Bounds for the alphabet sizes* $|\mathcal{U}|$*,* $|\mathcal{V}|$ *and* $|\mathcal{W}|$ *have not been clarified yet.*

The following corollary shows that Theorem 4.3.1 can be derived as a special case of Theorem 4.4.2.

Corollary 4.4.1. *For any DMS* (X, Y)*, if* $r_2 \geq H(Y)$ *and* $r_3 \geq H(X)$*, then*

$$\begin{aligned} & R_P(X, Y|D_1, D_2, R_2 = r_2, R_3 = r_3) \\ & \stackrel{\text{def.}}{=} \inf\{R : (R, r_2, r_3) \in \mathcal{R}_P(X, Y|D_1, D_2)\} \\ & = R_F(X, Y|D_1, D_2). \end{aligned}$$

The above corollary provides a typical example whereby the inner bound indicated in Theorem 4.4.2 coincides with the outer bound. Another example occurs when $\mathcal{X} = \mathcal{Y}$, $P_{XY}(p, q) = P_{XY}(q, p)\ \forall p, q \in \mathcal{X}$ and $D_1 = D_2$. In this case, we can identify Y as X because $P_X(p) = P_Y(p)\ \forall p \in \mathcal{X}$, and identify W as V because $D_1 = D_2$. Therefore, the inner bound is reduced to the outer bound, namely

$$\begin{aligned} \mathcal{R}_P(X, Y|D_1, D_2) = \{&(R_1, R_2, R_3) : \\ & R_1 \geq I(X; U|V) = I(Y; U|W), \\ & R_2 \geq I(Y; V),\ R_3 \geq I(X; W) = I(Y; V)\}. \end{aligned}$$

Also, we have the following properties:

Corollary 4.4.2.

$$\begin{aligned} R_P(X, Y|D_1 = d_1, D_2, R_2 = r_2, R_3 = r_3) &= R_{C2}(Y|X, D_2, R_2 = r_3), \\ R_P(X, Y|D_1, D_2 = d_2, R_2 = r_2, R_3 = r_3) &= R_{C2}(X|Y, D_1, R_2 = r_2), \end{aligned}$$

for $r_2, r_3 > 0$ *if* $d_1 \geq \overline{\Delta}_X$ *and* $d_2 \geq \overline{\Delta}_Y$*, where* $\mathcal{R}_{C2}(X|Y, D)$ *is the minimum achievable rate region when* X *is encoded and reproduced with encoded sequences from* Y *to guarantee the distortion level* D *[26], and*

$$R_{C2}(X|Y, D, R_2 = r) \stackrel{\text{def.}}{=} \inf\{R : (R, r) \in \mathcal{R}_{C2}(X|Y, D)\}.$$

These properties indicate that the rate region $\mathcal{R}_P(X, Y|D_1, D_2)$ is always reduced to $R_{C2}(X|Y, D_1)$ or $R_{C2}(Y|X, D_2)$ when one of the two messages does not need to be reproduced. The properties also agree with the result of sequential coding reported by Viswanathan and Berger [61]. The only difference between the two results is that one of the two messages does not need to be reproduced in the situation of PCD codes.

4.5 Proof of theorems

4.5.1 Proof of Theorem 4.3.1: converse part

Proof.
Let an f-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ be given that satisfies the conditions of Definitions 4.2.1 and 4.2.2. From Definition 4.2.2, for any $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ and then for all $n \geq n_1(\delta)$, we can obtain

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

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

$$\begin{aligned}
n(R+\delta) &\geq \log M_n \\
&\geq H(A_n) \\
&\geq H(A_n|Y^n) \\
&= I(X^n; A_n|Y^n) \qquad (4.3) \\
&= H(X^n|Y^n) - H(X^n|A_nY^n) \\
&= \sum_{k=1}^{n} \{H(X_k|Y_k) - H(X_k|A_nX^{k-1}Y^n)\} \\
&= \sum_{k=1}^{n} I(X_k; A_nX^{k-1}Y^{k-1}Y_{k+1}^n|Y_k) \\
&\geq \sum_{k=1}^{n} I(X_k; A_nX^{k-1}Y^{k-1}|Y_k),
\end{aligned}$$

where Eq. (4.3) comes from the equation $A_n = \varphi_n(X^n, Y^n)$. Let us define random variables $U_k = A_nX^{k-1}Y^{k-1}$. With these definitions, we have

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

In a similar manner, we obtain

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

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

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

$$
\begin{aligned}
&= H(X|Y) - H(X|JU_JY) \\
&= H(X|Y) - H(X|UY) \\
&= I(X;U|Y)
\end{aligned}
$$

and

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

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

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

We next show the existence of functions $\phi^{(1)}$ and $\phi^{(2)}$ that satisfy the conditions of Theorem 4.3.1. From Definition 4.2.2, for any $\gamma > 0$, there exists an integer $n_2 = n_2(\gamma)$, and for all $n \geq n_2(\gamma)$, we have

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

where $\widehat{\varphi}^{(i)}_{n;k}$ $(i = 1, 2,\ k \in \mathcal{I}_n)$ is the output of $\widehat{\varphi}^{(i)}_n$ at time k. We note that U_kY_k contains A_nY^k, and U_kX_k contains A_nX^k. Let $Y^n_{k+1}(U_k, Y_k)$ be a random variable selected to minimize the average distortion between X_k and $\widehat{X}_k$ given U_kY_k, and let $X^n_{k+1}(U_k, X_k)$ be a random variable selected to minimize the average distortion between Y_k and $\widehat{Y}_k$ given U_kX_k, namely

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

where $Q^{(1)}_k$ is the distribution of X_k given U_kY_k, namely for any $x^k \in \mathcal{X}^k$, $y^k \in \mathcal{Y}^k$, $a_n \in \mathcal{I}_{M^{(1)}_n}$ and $u_k = a_n x^{k-1} y^{k-1}$

$$
\begin{aligned}
Q^{(1)}_k(x_k|u_k, y_k) &\stackrel{\text{def.}}{=} Q^{(1)}_k(x_k|a_n, x^{k-1}, y^k) \\
&= \frac{Q^{(1)}_k(a_n, x^k, y^k)}{Q^{(1)}_k(a_n, x^{k-1}, y^k)} \\
Q^{(1)}_k(a_n, x^k, y^k) &\stackrel{\text{def.}}{=} \Pr\{\varphi_n(X^n, Y^n) = a_n, X^k = x^k, Y^k = y^k\} \\
&= \sum_{\substack{(x^n_{k+1}, y^n_{k+1}) \in \mathcal{X}^{n-k} \times \mathcal{Y}^{n-k}: \\ \varphi_n(x^n, y^n) = a_n}} P^n_{XY}(x^n, y^n), \\
Q^{(1)}_k(a_n, x^{k-1}, y^k) &\stackrel{\text{def.}}{=} \Pr\{\varphi_n(X^n, Y^n) = a_n, X^{k-1} = x^{k-1}, Y^k = y^k\} \\
&= \sum_{\substack{(x^n_k, y^n_{k+1}) \in \mathcal{X}^{n-k+1} \times \mathcal{Y}^{n-k}: \\ \varphi_n(x^n, y^n) = a_n}} P^n_{XY}(x^n, y^n),
\end{aligned}
$$

and $Q_k^{(2)}$ is the distribution of Y_k given $U_k X_k$, defined in a similar manner to $Q_k^{(1)}$. We choose the functions $\phi^{(1)}$ and $\phi^{(2)}$ as follows:

$$
\begin{aligned}
\phi_k^{(1)}(U_k, Y_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}_{n;k}^{(1)}(A_n, Y^k * Y_{k+1}^n(U_k, Y_k)), \\
\phi_k^{(2)}(U_k, X_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}_{n;k}^{(2)}(A_n, X^k * X_{k+1}^n(U_k, X_k)), \\
\phi^{(1)}(U, Y) &\stackrel{\text{def.}}{=} \phi_J^{(1)}(U_J, Y), \\
\phi^{(2)}(U, X) &\stackrel{\text{def.}}{=} \phi_J^{(2)}(U_J, X).
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
E\left[\Delta(X_k, \phi_k^{(1)}(U_k, Y_k))\right] &= E\left[\Delta(X_k, \widehat{\varphi}_{n;k}^{(1)}(A_n, Y^k * Y_{k+1}^n(U_k, Y_k)))\right] \\
&\leq E\left[\Delta(X_k, \widehat{\varphi}_{n;k}^{(1)}(A_n, Y^n))\right] \\
E\left[\Delta(Y_k, \phi_k^{(2)}(U_k, X_k))\right] &= E\left[\Delta(Y_k, \widehat{\varphi}_{n;k}^{(2)}(A_n, X^k * X_{k+1}^n(U_k, X_k)))\right] \\
&\leq E\left[\Delta(Y_k, \widehat{\varphi}_{n;k}^{(2)}(A_n, X^n))\right].
\end{aligned}
$$

This implies

$$
\begin{aligned}
D_1 + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{X}_k)\right] \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{\varphi}_{n;k}^{(1)}(A_n, Y^n))\right] \\
&\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \phi_k^{(1)}(U_k, Y_k))\right] \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X, \phi_J^{(1)}(U_J, Y))|J = k\right] \\
&= E\left[\Delta_X(X, \phi^{(1)}(U, Y))\right], \\
D_2 + \gamma &\geq E\left[\Delta_Y(Y, \phi^{(2)}(U, X))\right].
\end{aligned}
$$

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

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

It remains to establish that the bound on $|\mathcal{U}|$ specified in Theorem 4.3.1 does not affect the determination of the inf achievable rate $R_F(X, Y|D_1, D_2)$. To do this, we introduce the support lemma [13, Lemma 3.3.4]. We can see that

$$
P_{XY}(x, y) = \sum_{u \in \mathcal{U}} P_U(u) P_{XY|U}(x, y|u),
$$

$$
I(X; U|Y)
$$

$$
\begin{aligned}
&= H(X|Y) - H(X|UY) \\
&= H(X|Y) - \sum_{u\in\mathcal{U}} P_U(u) \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u) \log \frac{P_{Y|U}(y|u)}{P_{XY|U}(x,y|u)}
\end{aligned}
$$

$I(Y;U|X)$

$$
\begin{aligned}
&= H(Y|X) - H(Y|UX) \\
&= H(Y|X) - \sum_{u\in\mathcal{U}} P_U(u) \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u) \log \frac{P_{X|U}(x|u)}{P_{XY|U}(x,y|u)}
\end{aligned}
$$

$E[\Delta_X(X, \phi^{(1)}(U,Y))]$

$$
\begin{aligned}
&= \sum_{u\in\mathcal{U}} P_U(u) \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u) \Delta_X(x, \phi^{(1)}(u,y)) \\
&\geq \sum_{u\in\mathcal{U}} P_U(u) \sum_{y\in\mathcal{Y}} \min_{\widehat{x}\in\widehat{\mathcal{X}}} \sum_{x\in\mathcal{X}} P_{XY|U}(x,y|u) \Delta_X(x, \widehat{x}), \qquad (4.4)
\end{aligned}
$$

$E[\Delta_Y(Y, \phi^{(2)}(U,X))]$

$$
\begin{aligned}
&= \sum_{u\in\mathcal{U}} P_U(u) \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY|U}(x,y|u) \Delta_Y(y, \phi^{(2)}(u,x)) \\
&\geq \sum_{u\in\mathcal{U}} P_U(u) \sum_{x\in\mathcal{X}} \min_{\widehat{y}\in\widehat{\mathcal{Y}}} \sum_{y\in\mathcal{Y}} P_{XY|U}(x,y|u) \Delta_Y(y, \widehat{y}), \qquad (4.5)
\end{aligned}
$$

where Eq.(4.4) (resp. Eq.(4.5)) come from the fact that for given letters $(u,y) \in \mathcal{U}\times\mathcal{Y}$ (resp. $(u,x) \in \mathcal{U}\times\mathcal{X}$) the output of the function $\phi^{(1)}$ (resp. $\phi^{(2)}$) can be selected so as to minimize the average distortion. We then define the following functions q_i ($i \in \mathcal{I}_{m_y+2}$) of a generic distribution $Q \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, where $m_x = |\mathcal{X}|$, $m_y = |\mathcal{Y}|$, $m = m_x \times m_y$:

$$
\begin{aligned}
q_i(Q) &= Q(x,y), \ (i = xm_y + y \in \mathcal{I}_{m-1}) \\
q_m(Q) &= \max\{q_{m,1}(Q), q_{m,2}(Q)\}, \\
q_{m,1}(Q) &= H(X|Y) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} Q(x,y) \log \frac{\sum_{x'\in\mathcal{X}} Q(x',y)}{Q(x,y)}, \\
q_{m,2}(Q) &= H(Y|X) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} Q(x,y) \log \frac{\sum_{y'\in\mathcal{Y}} Q(x,y')}{Q(x,y)}, \\
q_{m+1}(Q) &= \sum_{y\in\mathcal{Y}} \min_{\widehat{x}\in\widehat{\mathcal{X}}} \sum_{x\in\mathcal{X}} Q(x,y)\Delta_X(x,\widehat{x}), \\
q_{m+2}(Q) &= \sum_{x\in\mathcal{X}} \min_{\widehat{y}\in\widehat{\mathcal{Y}}} \sum_{y\in\mathcal{Y}} Q(x,y)\Delta_Y(y,\widehat{y}).
\end{aligned}
$$

From the support lemma, we can find a generic distribution $\alpha \in \mathcal{P}(\widetilde{\mathcal{U}})$ such that $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$, $|\widetilde{\mathcal{U}}| \leq |\mathcal{X}\times\mathcal{Y}| + 2$ and the following equations are simultaneously satisfied:

$$
\sum_{u\in\widetilde{\mathcal{U}}} \alpha(u) q_i(P_{XY|U}(\cdot|u)) = P_{XY}(x,y) \quad (i = xm_y + y), \qquad (4.6)
$$

$$\sum_{u \in \widetilde{\mathcal{U}}} \alpha(u) q_m(P_{XY|U}(\cdot|u)) = \max\{I(X;U|Y), I(Y;U|X)\},$$
$$\sum_{u \in \widetilde{\mathcal{U}}} \alpha(u) q_{m+1}(P_{XY|U}(\cdot|u)) = \sum_{u \in \widetilde{\mathcal{U}}} \alpha(u) \sum_{y \in \mathcal{Y}} \min_{\widehat{x} \in \widehat{\mathcal{X}}} \sum_{x \in \mathcal{X}} P_{XY|U}(x,y|u) \Delta_X(x,\widehat{x}),$$
$$\sum_{u \in \widetilde{\mathcal{U}}} \alpha(u) q_{m+2}(P_{XY|U}(\cdot|u)) = \sum_{u \in \widetilde{\mathcal{U}}} \alpha(u) \sum_{x \in \mathcal{X}} \min_{\widehat{y} \in \widehat{\mathcal{Y}}} \sum_{y \in \mathcal{Y}} P_{XY|U}(x,y|u) \Delta_Y(y,\widehat{y}).$$

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

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

With these definitions, we have

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

and therefore

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

Also, Eq. (4.6) implies that there exists a random variable $\widetilde{U}$ taking values in $\widetilde{\mathcal{U}}$ and a joint distribution $P_{\widetilde{U}XY} \in \mathcal{P}(\widetilde{\mathcal{U}} \times \mathcal{X} \times \mathcal{Y})$ which satisfy

$$\alpha(u) P_{XY|U}(x,y|u) \quad = \quad P_{\widetilde{U}XY}(u,x,y) \quad \forall (u,x,y) \in \widetilde{\mathcal{U}} \times \mathcal{X} \times \mathcal{Y}.$$

The new joint distribution $P_{\widetilde{U}XY}$ preserves the distribution P_{XY} because

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

This completes the proof of the converse part. □

4.5.2 Proof of Theorem 4.3.1: direct part

We begin by mentioning a basic fact that will be used hereafter.

Lemma 4.5.1. (Steinberg-Merhav [55])
For given sequences $\{(\delta_n^{(1)}, \delta_n^{(2)})\}_{n=1}^{\infty}$ *each of which satisfies Eq. (2.3) and* $\delta_n^{(1)} <$

$\delta_n^{(2)}$ *for all* n*, and any* $\boldsymbol{x} \in T_X^n(\delta_n^{(1)})$*, there exist sequences* $\{(\epsilon_n^{(1)}, \epsilon_n^{(2)})\}_{n=1}^{\infty}$ *that satisfy*

$$\begin{aligned}\epsilon_n^{(1)} &= \epsilon_n^{(1)}(|\mathcal{U}|, |\mathcal{X}|, \delta_n^{(1)}, \delta_n^{(2)}), \\ \epsilon_n^{(2)} &= \epsilon_n^{(2)}(|\mathcal{U}|, |\mathcal{X}|, \delta_n^{(1)}, \delta_n^{(2)}), \\ \lim_{n\to\infty} \epsilon_n^{(1)} &= \lim_{n\to\infty} \epsilon_n^{(2)} = 0\end{aligned}$$

and

$$\exp\{-n(I(X;U) + \epsilon_n^{(1)})\} \leq \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n: \\ (\boldsymbol{u},\boldsymbol{x})\in T_{UX}^n(\delta_n^{(2)})}} P_U(\boldsymbol{u}) \leq \exp\{-n(I(X;U) - \epsilon_n^{(2)})\}.$$

Remark 4.5.1. *Steinberg and Merhav [55] utilized Lemma 4.5.1 as a result of Csiszar and Körner [13] without any proofs. However, the lemma has not been shown in [13].*

The proof of Lemma 4.5.1 is presented in the last part of this chapter.

Now, we proceed with the proof of the direct part of Theorem 4.3.1.

Proof.
Let a distortion pair (D_1, D_2) be given, and $P_{U|XY} \in \mathcal{P}_F(\mathcal{U}|P_{XY})$. Fix arbitrary $\gamma > 0$ and $\{\delta_n\}_{n=1}^{\infty}$ satisfying Eq. (2.3).

Codeword selection: φ_n

(1) Randomly generate M_U independent codewords $\boldsymbol{u}_i \in \mathcal{U}^n$ ($i \in \mathcal{I}_{M_U}$) according to the distribution P_U. These codewords comprises a codebook $\mathcal{A}_U = \{\boldsymbol{u}_i\}_{i=1}^{M_U}$.

(2) Partition the codebook $\mathcal{A}_U$ into N_U bins, each containing $L_U = M_U/N_U$ members of $\mathcal{A}_U$. Let $\mathcal{A}_U(j)$ denote the elements $\boldsymbol{u} \in \mathcal{A}_U$ assigned to bin j ($j \in \mathcal{I}_{N_U}$). Without loss of generality, we define

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

Encoding: φ_n

(1) For a given input pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequences, the encoder seeks a vector $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{x}, \boldsymbol{y}) \in T_{UXY}^n(2\delta_n)$. If there is more than one such vector in the codebook $\mathcal{A}_U$, the first one is chosen. If there is no such vector in the codebook $\mathcal{A}_U$, a default vector is chosen, say $\boldsymbol{u}_1$, and an encoding error is declared. The selected vector is denoted by $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})$.

(2) The value assigned to the encoder $\varphi_n(\cdot)$ is the bin index to which $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})$ belongs, that is,

$$\varphi_n(\boldsymbol{x}, \boldsymbol{y}) = j, \quad \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_U(j).$$

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

(1) The decoder has access to the indexes $j_U = \varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y})$ and the sequence $\boldsymbol{y} \in \mathcal{Y}^n$.

(2) The decoder seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ such that $(\boldsymbol{u}, \boldsymbol{y}) \in T_{UY}^n(3|\mathcal{X}|\delta_n)$. The selected vector is denoted by $\widehat{\boldsymbol{u}}(\boldsymbol{y})$. If there is no or more than one vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ jointly typical with $\boldsymbol{y}$, arbitrary $\widehat{\boldsymbol{u}}$ is chosen, and a decoding error

is declared.
(3) The reconstruction vector $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \widehat{x}_2, \cdots, \widehat{x}_n)$ is given by

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

<u>*Decoding:* $\widehat{\varphi}_n^{(2)}$</u>
(1) The decoder has access to the indexes $j_U = \varphi_n(\boldsymbol{x}, \boldsymbol{y})$ and the sequence $\boldsymbol{x} \in \mathcal{X}^n$.
(2) In a similar manner to $\widehat{\varphi}_n^{(1)}$, the decoder seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ such that $(\boldsymbol{u}, \boldsymbol{x}) \in T^n_{UX}(3|\mathcal{Y}|\delta_n)$, and the reconstruction vector $\widehat{\boldsymbol{y}}$ is given by

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

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

$$\begin{aligned}
\Delta^n_X(\boldsymbol{x}, \widehat{\boldsymbol{x}}) &= \frac{1}{n}\sum_{k=1}^{n} \Delta_X(x_k, \widehat{x}_k) \\
&= \frac{1}{n}\sum_{k=1}^{n} \Delta_X(x_k, \phi^{(1)}(\widehat{u}_k(\boldsymbol{y}), \widehat{y}_k)) \\
&= \frac{1}{n}\sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} N(u,x,y|\widehat{\boldsymbol{u}}(\boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y})\Delta_X(x, \phi^{(1)}(u,y)).
\end{aligned}$$

We note that $(\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y}) \in T^n_{UXY}(2\delta_n)$ from the encoding procedure. Also, if no error occurs in the encoding and decoding processes, we have $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}) = \widehat{\boldsymbol{u}}(\boldsymbol{y})$. In this case, the following inequalities are satisfied:

$$\begin{aligned}
\Delta^n_X(\boldsymbol{x}, \widehat{\boldsymbol{x}}) &\le \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} (P_{UXY}(u,x,y) + 2\delta_n)\Delta_X(x, \phi^{(1)}(u,y)) \\
&\le E\left[\Delta_X(X, \phi^{(1)}(U,Y))\right] + 2\delta_n\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}| \\
&\le D_1 + 2\delta_n\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|.
\end{aligned}$$

Denoting the error probability in the encoding and decoding processes as P_e^n, the average distortion can be bounded as

$$E\left[\Delta^n_X(X^n, \widehat{X}^n)\right] \le (1 - P_e^n)(D_1 + 2\delta_n\overline{\Delta}_X|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}|) + P_e^n\overline{\Delta}_X.$$

If P_e^n vanishes as $n \to \infty$, we can obtain

$$\limsup_{n\to\infty} E\left[\Delta^n_X(X^n, \widehat{X}^n)\right] \le D_1.$$

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

$$\limsup_{n\to\infty} E\left[\Delta^n_Y(Y^n, \widehat{Y}^n)\right] \le D_2$$

in a similar manner to $\widehat{\varphi}_n^{(1)}$, if P_e^n vanishes as $n \to \infty$.

Error evaluation: φ_n

If there is no $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{x}, \boldsymbol{y}) \in T^n_{UXY}(2\delta_n)$, an encoding error has occurred. This event is denoted as

$$E_1 \stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_U} \{(\boldsymbol{u}_i, \boldsymbol{x}, \boldsymbol{y}) \notin T^n_{UXY}(2\delta_n)\},$$

Here, let us define

$$E_0 \stackrel{\text{def.}}{=} \{(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{XY}(\delta_n)\}.$$

From Lemma 2.3.3, we have $\Pr\{E_0^c\} \to 0$ as $n \to \infty$. Then, we obtain

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

$$\begin{aligned}
&\Pr\{E_0 \cap E_1\} \\
&\le \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T^n_{XY}(\delta_n)} P_{XY}(\boldsymbol{x}, \boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{x}, \boldsymbol{y}) \notin T^n_{UXY}(2\delta_n)\} \middle| \boldsymbol{x}, \boldsymbol{y} \right\} \\
&= \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T^n_{XY}(\delta_n)} P_{XY}(\boldsymbol{x}, \boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{x}, \boldsymbol{y}) \notin T^n_{UXY}(2\delta_n)\} \right\} \qquad (4.7) \\
&\le \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T^n_{XY}(\delta_n)} P_{XY}(\boldsymbol{x}, \boldsymbol{y}) \left[1 - \exp\left\{ -n(I(XY;U) + \epsilon_n^{(1)}) \right\} \right]^{M_U} \qquad (4.8) \\
&\le \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T^n_{XY}(\delta_n)} P_{XY}(\boldsymbol{x}, \boldsymbol{y}) \exp\left[-M_U \exp\left\{ -n(I(XY;U) + \epsilon_n^{(1)}) \right\} \right] \qquad (4.9) \\
&\le \exp\left[-M_U \exp\left\{ -n(I(XY;U) + \epsilon_n^{(1)}) \right\} \right],
\end{aligned}$$

where Eq.(4.7) comes from the fact that $\boldsymbol{u}_i$ is selected independently of $(\boldsymbol{x}, \boldsymbol{y})$, Eq.(4.8) from Lemma 4.5.1 Eq.(4.9) from the equation $(1-a)^n \le \exp(-an)$, and

$$\begin{aligned}\epsilon_n^{(1)} &= \epsilon_n^{(1)}(|\mathcal{X} \times \mathcal{Y}|, |\mathcal{U}|, \delta_n, 2\delta_n) \\ &\to 0 \quad (n \to \infty).\end{aligned}$$

By setting M_U as

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

$m_1\gamma > \epsilon_n^{(1)}$, we have

$$\lim_{n\to\infty} \Pr\{E_1\} = 0$$

.

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 $(\boldsymbol{u}_i, \boldsymbol{y}) \in T_{UY}(3|\mathcal{X}|\delta_n)$,

a decoding error is declared. This event is classified into two cases.
(1) The first case: $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{y}) \notin T^n_{UY}(3|\mathcal{X}|\delta_n)$. However, this error does not occur because $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{x},\boldsymbol{y}) \in T^n_{UXY}(2\delta_n)$ and therefore $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{y}) \in T^n_{UXY}(2|\mathcal{X}|\delta_n)$ from Lemma 2.3.5.
(2) The second case: If there exists $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ such that $\boldsymbol{u} \neq \boldsymbol{u}(\boldsymbol{x},\boldsymbol{y})$ and $(\boldsymbol{u},\boldsymbol{y}) \in T^n_{UY}(3|\mathcal{X}|\delta_n)$. This event is denoted as

$$E_2 \stackrel{\text{def.}}{=} \bigcup_{\boldsymbol{u}\in\mathcal{A}_U(j_U),\boldsymbol{u}\neq\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y})} \{(\boldsymbol{u},\boldsymbol{y}) \in T^n_{UY}(3|\mathcal{X}|\delta_n)\}.$$

Let $i(j,k)$ be the index i of k-th $\boldsymbol{u}_i$, which belongs to $\mathcal{A}_U(j)$. Since if $(\boldsymbol{x},\boldsymbol{y}) \in T^n_{XY}(\delta_n)$ then $\boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta_n)$, we have

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

$$\begin{aligned}&\Pr\{E_0 \cap E_2\}\\ &\leq \sum_{k=1}^{|\mathcal{A}_U(j_U)|} \sum_{\boldsymbol{y}\in T^n_Y(|\mathcal{X}|\delta_n)} P_Y(\boldsymbol{y}) \Pr\left\{(U^n_{i(j_U,k)},\boldsymbol{y}) \in T^n_{UY}(3|\mathcal{X}|\delta_n)\right\} \quad (4.10)\\ &\leq \sum_{k=1}^{|\mathcal{A}_U(j_U)|} \sum_{\boldsymbol{y}\in T^n_Y(|\mathcal{X}|\delta_n)} P_Y(\boldsymbol{y}) \exp\{-n(I(Y;U) - \epsilon_n^{(2)})\} \quad (4.11)\\ &\leq |\mathcal{A}_U(j_U)| \exp\{-n(I(Y;U) - \epsilon_n^{(2)})\}\\ &= L_U \exp\{-n(I(Y;U) - \epsilon)\}.\end{aligned}$$

where Eq.(4.10) comes from the fact that $\boldsymbol{u}_i$ is selected independently of $\boldsymbol{y}$, Eq.(4.11) from Lemma 4.5.1, and

$$\begin{aligned}\epsilon_n^{(2)} &= \epsilon_n^{(2)}(|\mathcal{Y}|,|\mathcal{U}|,|\mathcal{X}|\delta_n, 3|\mathcal{X}|\delta_n)\\ &\to 0 \quad (n \to \infty).\end{aligned}$$

By setting L_U as

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

$l_{11}\gamma > \epsilon_n^{(2)}$, we have

$$\lim_{n\to\infty} \Pr\{E_2\} = 0$$

.

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

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

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

to vanish the encoding and decoding errors.

Rate evaluation: $\varphi_n^{(1)}$
The encoder sends the bin indexes using

$$\begin{aligned} R_1 = \frac{1}{n}\log N_U & \\ &= \frac{1}{n}\log\frac{M_U}{L_U} \\ &\geq I(XY;U) + m_1\gamma - \min\{I(Y;U) - l_1\gamma, I(X;U) - l_2\gamma\} \\ &= \max\{I(X;U|Y) + l_1\gamma, I(Y;U|X) + l_2\gamma\} + m_1\gamma \end{aligned}$$

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

This completes the proof of Theorem 4.3.1. □

4.5.3 Proof of Theorem 4.4.2: converse part

Proof.
Let an f-PCD code

$$\left\{(\varphi_n^{(1)}, \varphi_n^{(2)}, \varphi_n^{(3)}, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\right\}_{n=1}^{\infty}$$

be given that satisfy the conditions of Definitions 4.4.1 and 4.4.2. From Definition 4.4.2, for $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ and then for all $n \geq n_1(\delta)$, we can obtain

$$\frac{1}{n}\log M_n^{(i)} \quad \leq \quad R_i + \delta \quad (i = 1, 2, 3). \tag{4.12}$$

First, we evaluate Eq. (4.12) for $i = 1$. Let us define $A_n^{(1)} = \varphi_n^{(1)}(X^n, Y^n)$, $A_n^{(2)} = \varphi_n^{(2)}(Y^n)$ and $A_n^{(3)} = \varphi_n^{(3)}(X^n)$. Then, we obtain

$$\begin{aligned} n(R_1 + \delta) &\geq \log M_n^{(1)} \\ &\geq H(A_n^{(1)}) \\ &\geq H(A_n^{(1)}|A_n^{(2)}) \\ &= I(X^nY^n; A_n^{(1)}|A_n^{(2)}) \qquad (4.13) \\ &= \sum_{k=1}^{n} I(X_kY_k; A_n^{(1)}|A_n^{(2)}X^{k-1}Y^{k-1}) \\ &\geq \sum_{k=1}^{n} I(X_k; A_n^{(1)}|A_n^{(2)}X^{k-1}Y^{k-1}), \qquad (4.14) \end{aligned}$$

where Eq. (4.13) comes from the fact that $A_n^{(1)} = \varphi_n^{(1)}(X^n, Y^n)$. Let us define the random variables $U_k = A_n^{(1)}$ and $V_k = A_n^{(2)}X^{k-1}Y^{k-1}$. With these definitions, we have the Markov structure $V_k \to Y_k \to X_k$ because

$$\begin{aligned} I(V_k; X_k|Y_k) &= I(A_n^{(2)}X^{k-1}Y^{k-1}; X_k|Y_k) \\ &\leq I(A_n^{(2)}X^{k-1}Y^{k-1}Y_{k+1}^n; X_k|Y_k) \\ &= I(X^{k-1}Y^{k-1}Y_{k+1}^n; X_k|Y_k) \qquad (4.15) \\ &\leq I(X^{k-1}Y^{k-1}Y_{k+1}^n; X_kY_k) \end{aligned}$$

$$= \quad 0,$$

where Eq. (4.15) comes from the fact that $A_n^{(2)} = \varphi_n^{(2)}(Y^n)$. Substituting U_k and V_k into Eq. (4.14), we obtain

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

In a similar manner, by letting $W_k = A_n^{(3)} X^{k-1} Y^{k-1}$ we obtain

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

and the Markov structure $W_k \to X_k \to Y_k$. Here, let J be a random variable, independent of X and Y, and uniformly distributed over the set $\mathcal{I}_n$. We define the random variables $U = (J, U_J)$, $V = (J, V_J)$ and $W = (J, W_J)$. The Markov structures $V \to Y \to X$ and $W \to X \to Y$ still hold, and we have

$$\begin{aligned}
R_1 + \delta &\geq \frac{1}{n} \sum_{k=1}^{n} I(X_k; U_k|V_k) \\
&= \frac{1}{n} \sum_{k=1}^{n} \{H(X_k|V_k) - H(X_k|U_k V_k)\} \\
&= \frac{1}{n} \sum_{k=1}^{n} \{H(X|V_J, J = k) - H(X|U_J V_J, J = k)\} \\
&= H(X|JV_J) - H(X|JU_J V_J) \\
&= H(X|V) - H(X|UV) \\
&= I(X; U|V)
\end{aligned}$$

and

$$R_1 + \delta \geq I(Y; U|W).$$

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

$$R_1 \geq \max\{I(X; U|V), I(Y; U|W)\}.$$

Next, we evaluate Eq. (4.12) for $i = 2$.

$$\begin{aligned}
n(R_2 + \delta) &\geq \log M_n^{(2)} \\
&\geq H(A_n^{(2)}) \\
&\geq I(X^n Y^n; A_n^{(2)}) \\
&= \sum_{k=1}^{n} I(X_k Y_k; A_n^{(2)} X^{k-1} Y^{k-1}) \\
&\geq \sum_{k=1}^{n} I(Y_k; A_n^{(2)} X^{k-1} Y^{k-1}) \\
&= \sum_{k=1}^{n} I(Y_k; V_k).
\end{aligned}$$

In the same way as with the above discussion, we have $R_2 \geq I(Y;V)$. In a similar manner, we also obtain $R_3 \geq I(X;W)$.

We next show the existence of functions $\phi^{(1)}$ and $\phi^{(2)}$ that satisfy the conditions of Theorem 4.4.2. From Definition 4.4.2, for any $\gamma > 0$, there exists an integer $n_2 = n_2(\gamma)$, and for all $n \geq n_2(\gamma)$, we have

$$\begin{aligned} D_1 + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{\varphi}^{(1)}_{n;k}(A_n^{(1)}, A_n^{(2)}))\right], \\ D_2 + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_Y(Y_k, \widehat{\varphi}^{(2)}_{n;k}(A_n^{(1)}, A_n^{(3)}))\right]. \end{aligned}$$

Now, we denote the output of $\widehat{\varphi}^n_{(j)}$ at time k ($k \in \mathcal{I}_n$) by $\widehat{\varphi}^{(j)}_{n;k}$ ($j = 1, 2$). From the fact that $U_k V_k$ contains $A_n^{(1)} A_n^{(2)}$, and $U_k W_k$ contains $A_n^{(1)} A_n^{(3)}$, we choose the functions $\phi^{(1)}$ and $\phi^{(2)}$ as follows:

$$\begin{aligned} \phi^{(1)}_k(U_k, V_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}^{(1)}_{n;k}(A_n^{(1)}, A_n^{(2)}), \\ \phi^{(2)}_k(U_k, W_k) &\stackrel{\text{def.}}{=} \widehat{\varphi}^{(2)}_{n;k}(A_n^{(1)}, A_n^{(3)}), \\ \phi^{(1)}(U, V) &\stackrel{\text{def.}}{=} \phi^{(1)}_J(U_J, V_J), \\ \phi^{(2)}(U, W) &\stackrel{\text{def.}}{=} \phi^{(2)}_J(U_J, W_J). \end{aligned}$$

This implies that

$$\begin{aligned} D_1 + \gamma &\geq \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \widehat{\varphi}^{(1)}_{n;k}(A_n^{(1)}, A_n^{(2)}))\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X_k, \phi^{(1)}_k(U_k, V_k))\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[\Delta_X(X, \phi^{(1)}_J(U_J, V_J))|J = k\right] \\ &= E\left[\Delta_X(X, \phi^{(1)}(U, V))\right] \end{aligned}$$

and

$$D_2 + \gamma \geq E\left[\Delta_Y(Y, \phi^{(2)}(U, W))\right].$$

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

$$\begin{aligned} D_1 &\geq E\left[\Delta_X(X, \phi^{(1)}(U, V))\right], \\ D_2 &\geq E\left[\Delta_Y(Y, \phi^{(2)}(U, W))\right]. \end{aligned}$$

This completes the proof of the converse part. □

4.5.4 Proof of Theorem 4.4.2: direct part

Proof.
Let a distortion pair (D_1, D_2) be given, and let U, V, W and $P_{UVW|XY} \in \overline{\mathcal{P}}_P(\mathcal{U}\times\mathcal{V}\times\mathcal{W}|P_{XY})$. Fix arbitrary $\gamma > 0$ and $\{\delta_n\}_{n=1}^{\infty}$ satisfying the conditions shown in Remark 2.3.1.

Codeword selection: $\varphi_n^{(2)}$ $(\varphi_n^{(3)})$

Randomly generate M_V (resp. M_W) independent codewords $\mathcal{A}_V = \{\boldsymbol{v}_i\}_{i=1}^{M_V}$, $\boldsymbol{v}_i \in \mathcal{V}^n$ (resp. $\mathcal{A}_W = \{\boldsymbol{w}_i\}_{i=1}^{M_W}$, $\boldsymbol{w}_i \in \mathcal{W}^n$) according to P_V (resp. P_W).

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

(1) Randomly generate M_U independent codewords $\mathcal{A}_U = \{\boldsymbol{u}_i\}_{i=1}^{M_U}$, $\boldsymbol{u}_i \in \mathcal{U}^n$ according to P_U.
(2) Partition the codebook $\mathcal{A}_U$ into N_U bins, each containing $L_U = M_U/N_U$ members of $\mathcal{A}_U$. Let $\mathcal{A}_U(j)$ denote the elements $\boldsymbol{u} \in \mathcal{A}_U$ assigned to bin j $(j \in \mathcal{I}_{N_U})$. Without loss of generality, we define

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

Encoding: $\varphi_n^{(2)}$

(1) For a given $\boldsymbol{y} \in \mathcal{Y}^n$, the encoder seeks a vector $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{y}) \in T_{VY}^n(2|\mathcal{X}|\delta_n)$. If there is more than one such vector in the codebook $\mathcal{A}_V$, the first one is chosen. If there is no such vector in the codebook $\mathcal{A}_V$, a default vector is chosen, say $\boldsymbol{v}_1$, and an error is declared. We denote the selected vector by $\boldsymbol{v}(\boldsymbol{y})$.
(2) The value assigned to the encoder $\varphi_n^{(2)}(\cdot)$ is the index of the selected vector, that is,

$$\varphi_n^{(2)}(\boldsymbol{y}) \quad = \quad i, \quad \boldsymbol{v}(\boldsymbol{y}) = \boldsymbol{v}_i.$$

Encoding: $\varphi_n^{(3)}$

In a similar manner to $\varphi_n^{(2)}$, the encoder seeks a vector $\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{w}_i \in \mathcal{A}_W$ such that $(\boldsymbol{w}_i, \boldsymbol{x}) \in T_{WX}^n(2|\mathcal{Y}|\delta_n)$, and the value assigned to the encoder $\varphi_n^{(3)}(\cdot)$ is the index of the selected vector.

$$\varphi_n^{(3)}(\boldsymbol{x}) \quad = \quad i, \quad \boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{w}_i.$$

Encoding: $\varphi_n^{(1)}$

(1) For a given $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, the encoder seeks a vector $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \in T_{UVX}^n(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$ and $(\boldsymbol{u}_i, \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{y}) \in T_{UWY}^n(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$. If there is more than one such vector in the codebook $\mathcal{A}_U$, the first one is chosen. If there is no such vector in the codebook $\mathcal{A}_U$, a default vector is chosen, say $\boldsymbol{u}_1$, and an error is declared. The selected vector is denoted by $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})$.
(2) The value assigned to the encoder $\varphi_n^{(1)}(\cdot)$ is the bin index to which $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})$ belongs, that is,

$$\varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y}) \quad = \quad j, \quad \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_U(j).$$

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

(1) The decoder has access to the indexes $j_U = \varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y})$ and $j_V = \varphi_n^{(2)}(\boldsymbol{y})$.

(2) We can recover the unique vector $\widehat{\boldsymbol{v}} = \boldsymbol{v}(\boldsymbol{y}) = \boldsymbol{v}_{j_V} \in \mathcal{A}_V(j_V)$ from the index j_V.
(3) The decoder seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ such that $(\boldsymbol{u}, \widehat{\boldsymbol{v}}) \in T^n_{UV}(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)$. This vector is denoted by $\widehat{\boldsymbol{u}}(\widehat{\boldsymbol{v}})$. If there is no or more than one vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ jointly typical with $\widehat{\boldsymbol{v}}$, arbitrary $\widehat{\boldsymbol{u}}$ is chosen, and an error is declared.
(4) The reconstruction vector $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \widehat{x}_2, \cdots, \widehat{x}_n)$ is given by

$$\widehat{x}_k = \phi^{(1)}(\widehat{u}_k(\widehat{\boldsymbol{v}}), \widehat{v}_k) \quad (k \in \mathcal{I}_n).$$

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

(1) The decoder has access to the indexes $j_U = \varphi_n^{(1)}(\boldsymbol{x}, \boldsymbol{y})$ and $j_W = \varphi_n^{(3)}(\boldsymbol{x})$.
(2) In a similar manner to $\widehat{\varphi}_n^{(1)}$, the decoder seeks a unique vector $\boldsymbol{u} \in \mathcal{A}_U(j_U)$ such that $(\boldsymbol{u}, \widehat{\boldsymbol{w}} = \boldsymbol{w}_{j_V}) \in T^n_{UW}(4|\mathcal{Y}||\mathcal{X} \times \mathcal{Y}|\delta_n)$, and the reconstruction vector $\widehat{\boldsymbol{y}}$ is given by

$$\widehat{y}_k = \phi^{(1)}(\widehat{u}_k(\widehat{\boldsymbol{w}}), \widehat{w}_k) \quad (k \in \mathcal{I}_n).$$

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

For the distortion, we obtain

$$\begin{aligned}
\Delta_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) &= \frac{1}{n}\sum_{k=1}^n \Delta_X(x_k, \widehat{x}_k) \\
&= \frac{1}{n}\sum_{k=1}^n \Delta_X(x_k, \phi^{(1)}(\widehat{u}_k(\widehat{\boldsymbol{v}}), \widehat{v}_k)) \\
&= \frac{1}{n}\sum_{(u,v,x)\in\mathcal{U}\times\mathcal{V}\times\mathcal{X}} N(u, v, x|\widehat{\boldsymbol{u}}(\widehat{\boldsymbol{v}}), \widehat{\boldsymbol{v}}, \boldsymbol{x})\Delta_X(x, \phi^{(1)}(u, v)).
\end{aligned}$$

Noting that $(\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \in T^n_{UVX}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$, if no error occurs in the encoding/decoding process $(\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) = (\widehat{\boldsymbol{u}}(\widehat{\boldsymbol{v}}), \widehat{\boldsymbol{v}}, \boldsymbol{x})$, and therefore the following inequalities are satisfied:

$$\begin{aligned}
\Delta_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) &\leq \sum_{(u,v,x)\in\mathcal{U}\times\mathcal{V}\times\mathcal{X}} (P_{UVX}(u, v, x) + 3|\mathcal{X} \times \mathcal{Y}|\delta_n)\Delta_X(x, \phi^{(1)}(u, v)) \\
&\leq E\left[\Delta_X(X, \phi^{(1)}(U, V))\right] + 3|\mathcal{X} \times \mathcal{Y}|\delta_n\overline{\Delta}_X|\mathcal{U} \times \mathcal{V} \times \mathcal{X}| \\
&\leq D_1 + 3|\mathcal{X} \times \mathcal{Y}|\delta_n\overline{\Delta}_X|\mathcal{U} \times \mathcal{V} \times \mathcal{X}|.
\end{aligned}$$

Denoting error probabilities in the encoding/decoding process as P_e^n, the average distortion can be bounded as

$$E\left[\Delta_X^n(X^n, \widehat{X}^n)\right] \leq (1 - P_e^n)(D_1 + 3|\mathcal{X} \times \mathcal{Y}|\delta_n\overline{\Delta}_X|\mathcal{U} \times \mathcal{V} \times \mathcal{X}|) + P_e^n\overline{\Delta}_X.$$

If P_e^n vanishes as $n \to \infty$, we can obtain

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

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

We can obtain

$$\limsup_{n\to\infty} E\left[\Delta_Y^n(Y^n, \widehat{Y}^n)\right] \leq D_2$$

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

Error evaluation: $\varphi_n^{(2)}$

If there is no $\boldsymbol{v}_i \in \mathcal{A}_V$ such that $(\boldsymbol{v}_i, \boldsymbol{y}) \in T^n_{VY}(2|\mathcal{X}|\delta_n)$, an encoding error has occurred. This event is denoted as

$$E_1 \stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_V} \{(\boldsymbol{v}_i, \boldsymbol{y}) \notin T^n_{VY}(2|\mathcal{X}|\delta_n)\}.$$

Here, let us define

$$E_0 \stackrel{\text{def.}}{=} \{(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{XY}(\delta_n)\}.$$

From Lemma 2.3.3, $\Pr\{E_0^c\} \to 0$ as $n \to \infty$. Also, we note that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{XY}(\delta_n) \Rightarrow \boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta_n)$ from Lemma 2.3.5. Then, we have

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

$$\begin{aligned}
&\Pr\{E_0 \cap E_1\} \\
&\leq \sum_{\boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta_n)} P_Y(\boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_V} \{(V_i^n, \boldsymbol{y}) \notin T^n_{VY}(2|\mathcal{X}|\delta_n)\} \middle| \boldsymbol{y} \right\} \\
&= \sum_{\boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta_n)} P_Y(\boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_V} \{(V_i^n, \boldsymbol{y}) \notin T^n_{VY}(2|\mathcal{X}|\delta_n)\} \right\} \qquad (4.16) \\
&\leq \sum_{\boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta)} P_Y(\boldsymbol{y}) \left[1 - \exp\left\{-n(I(Y;V) + \epsilon_n^{(1)})\right\}\right]^{M_V} \qquad (4.17) \\
&\leq \sum_{\boldsymbol{y} \in T^n_Y(|\mathcal{X}|\delta_n)} P_Y(\boldsymbol{y}) \exp\left[-M_V \exp\left\{-n(I(Y;V) + \epsilon_n^{(1)})\right\}\right], \qquad (4.18)
\end{aligned}$$

where Eq.(4.16) comes from the fact that $\boldsymbol{v}_i$ is selected independently of $\boldsymbol{y}$, Eq.(4.17) from Lemma 4.5.1, Eq.(4.18) from the equation $(1-a)^n \leq \exp(-an)$, and

$$\begin{aligned}\epsilon_n^{(1)} &= \epsilon_n^{(1)}(|\mathcal{Y}|, |\mathcal{V}|, |\mathcal{X}|\delta_n, 2|\mathcal{X}|\delta_n) \\ &\to \infty \quad (n \to \infty).\end{aligned}$$

By setting M_V as

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

$m_1\gamma > \epsilon_n^{(1)}$, we have

$$\lim_{n\to\infty} \Pr\{E_1\} = 0$$

.

Error evaluation: $\varphi_n^{(3)}$

If there is no $\boldsymbol{w}_i \in \mathcal{A}_W$ such that $(\boldsymbol{w}_i, \boldsymbol{x}) \in T^n_{WX}(2|\mathcal{Y}|\delta_n)$, an encoding error has occurred. The probability of this event vanishes as $n \to \infty$ by setting M_W as

$$M_W \geq \exp\{n(I(X;W) + m_2\gamma)\},\ m_2 > 0. \qquad (4.19)$$

$\underline{\textit{Error evaluation: } \varphi_n^{(1)}}$

If there is no $\boldsymbol{u}_i \in \mathcal{A}_U$ such that $(\boldsymbol{u}_i, \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \in T^n_{UVX}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$ and $(\boldsymbol{u}_i, \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{y}) \in T^n_{UWY}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$, an encoding error has occurred. This event is denoted as

$$\begin{aligned} E_2 &\stackrel{\text{def.}}{=} E_{21} \cup E_{22}, \\ E_{21} &\stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_U} \left\{(\boldsymbol{u}_i, \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \notin T^n_{UVX}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)\right\}, \\ E_{22} &\stackrel{\text{def.}}{=} \bigcap_{i=1}^{M_U} \left\{(\boldsymbol{u}_i, \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{y}) \notin T^n_{UWY}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)\right\}. \end{aligned}$$

Here, let us define

$$\begin{aligned} E_{31} &\stackrel{\text{def.}}{=} \{(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y}) \notin T^n_{VXY}(2|\mathcal{X}|\delta_n)\}, \\ E_{32} &\stackrel{\text{def.}}{=} \{(\boldsymbol{w}(\boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y}) \notin T^n_{WXY}(2|\mathcal{Y}|\delta_n)\}. \end{aligned}$$

Since $V \to Y \to X$ forms a Markov chain and $(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{y}) \in T^n_{VY}(2|\mathcal{X}|\delta_n)$, we obtain

$$\begin{aligned} \Pr\{E_{31}\} &\le \Pr\{E_0^c \cup E_{31}\} \\ &= \Pr\{E_0^c\} + \Pr\{E_0 \cap E_{31}\} \\ &\to 0 \quad (n \to \infty) \end{aligned}$$

from Lemma 2.3.6. We also note that $(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y}) \in T^n_{VXY}(2|\mathcal{X}|\delta_n) \Rightarrow (\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \in T^n_{VX}(2|\mathcal{X} \times \mathcal{Y}|\delta_n)$ from Lemma 2.3.5. Then, we have

$$\begin{aligned} \Pr\{E_{21}\} &\le \Pr\{E_{31} \cup E_{21}\} \\ &= \Pr\{E_{31}\} + \Pr\{E_{31}^c \cap E_{21}\}, \end{aligned}$$

$$\begin{aligned} &\Pr\{E_{31}^c \cap E_{21}\} \\ &\le \sum_{(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \in T^n_{VXY}(2|\mathcal{X}|\delta_n)} P_{VXY}(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{v}, \boldsymbol{x}) \notin T^n_{UVX}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)\} \,\middle|\, \boldsymbol{v}, \boldsymbol{x} \right\} \\ &= \sum_{(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \in T^n_{VXY}(2|\mathcal{X}|\delta_n)} P_{VXY}(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \Pr\left\{ \bigcap_{i=1}^{M_U} \{(U_i^n, \boldsymbol{v}, \boldsymbol{x}) \notin T^n_{UVX}(3|\mathcal{X} \times \mathcal{Y}|\delta_n)\} \right\} && (4.20) \\ &\le \sum_{(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \in T^n_{VXY}(2|\mathcal{X}|\delta_n)} P_{VXY}(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \left[1 - \exp\left\{-n(I(VX;U) + \epsilon_n^{(2)})\right\}\right]^{M_U} && (4.21) \\ & && (4.22) \\ &\le \sum_{(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \in T^n_{VXY}(2|\mathcal{X}|\delta_n)} P_{VXY}(\boldsymbol{v},\boldsymbol{x},\boldsymbol{y}) \exp\left[-M_U \exp\left\{-n(I(VX;U) + \epsilon_n^{(2)})\right\}\right], && (4.23) \\ & && (4.24) \end{aligned}$$

where Eq.(4.20) comes from the fact that $\boldsymbol{u}_i$ is selected independently of $(\boldsymbol{v}, \boldsymbol{x})$, Eq.(4.22) from Lemma 4.5.1, Eq.(4.24) from the equation $(1-a)^n \le \exp(-an)$, and

$$\epsilon_n^{(2)} = \epsilon_n^{(2)}(|\mathcal{V} \times \mathcal{X}|, |\mathcal{U}|, 2|\mathcal{X} \times \mathcal{Y}|\delta_n, 3|\mathcal{X} \times \mathcal{Y}|\delta_n)$$

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

By setting M_U as

$$M_U \geq \exp\{n(I(VX;U) + m_{21}\gamma)\},\ m_{21} > 0,$$

$m_{21}\gamma > \epsilon_n^{(2)}$, we have

$$\lim_{n\to\infty} \Pr\{E_{21}\} = 0.$$

In a similar manner, we have

$$\lim_{n\to\infty} \Pr\{E_{22}\} = 0$$

by setting M_U as

$$M_U \geq \exp\{n(I(WY;U) + m_{22}\gamma)\},\ m_{22} > 0.$$

<u>*Error evaluation:* $\widehat{\varphi}_n^{(1)}$</u>
If there is no or more than one $\boldsymbol{u}_i \in \mathcal{A}_U(j_U)$ such that $(\boldsymbol{u}_i, \widehat{\boldsymbol{v}}) \in T_{UV}^n(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)$, a decoding error is declared. This event is classified into two cases.
(1) The first case: $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{v}(\boldsymbol{y})) \notin T_{UV}^n(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)$. However, this error does not occur because $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{v}(\boldsymbol{y}), \boldsymbol{x}) \in T_{UVX}^n(3|\mathcal{X} \times \mathcal{Y}|\delta_n)$ and therefore $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{v}(\boldsymbol{y})) \in T_{UV}^n(3|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)$ from Lemma 2.3.5.
(2) The second case: If there exists $\boldsymbol{u} \in \mathcal{A}_U(j_U)$, $\boldsymbol{u} \neq \boldsymbol{u}(\boldsymbol{x},\boldsymbol{y})$ such that $(\boldsymbol{u}, \widehat{\boldsymbol{v}}) \in T_{UV}^n(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)$. This event is denoted as

$$E_4 \stackrel{\text{def.}}{=} \bigcup_{\boldsymbol{u}\in\mathcal{A}_U(j_U), \boldsymbol{u}\neq\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y})} \{(\boldsymbol{u}, \widehat{\boldsymbol{v}}) \in T_{UV}^n(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)\}.$$

Let $i(j,k)$ be the index i of k-th $\boldsymbol{u}_i$, which belongs to $\mathcal{A}_U(j)$. Since $(\boldsymbol{v}(\boldsymbol{y}), \boldsymbol{y}) \in T_{VY}^n(2|\mathcal{X}|\delta_n)$ and therefore $\boldsymbol{v}(\boldsymbol{y}) \in T_V^n(2|\mathcal{X} \times \mathcal{Y}|\delta_n)$ from Lemma 2.3.5, we have

$$\begin{aligned}
\Pr\{E_4\} &\leq \sum_{k=1}^{|\mathcal{A}_U(j_U)|} \sum_{(\boldsymbol{v},\boldsymbol{y})\in T_{VY}^n(2|\mathcal{X}|\delta_n)} \\
&\quad P_{VY}(\boldsymbol{v},\boldsymbol{y}) \Pr\left\{(U_{i(j_U,k)}^n, \boldsymbol{v}) \in T_{UV}^n(4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n)\right\} \quad (4.25) \\
&\leq \sum_{k=1}^{|\mathcal{A}_U(j_U)|} \sum_{(\boldsymbol{v},\boldsymbol{y})\in T_{VY}^n(2|\mathcal{X}|\delta_n)} P_{VY}(\boldsymbol{v},\boldsymbol{y}) \exp\{-n(I(V;U) - \epsilon_n^{(3)})\} \quad (4.26) \\
&\leq |\mathcal{A}_U(j_U)| \exp\{-n(I(V;U) - \epsilon_n^{(3)})\} \\
&= L_U \exp\{-n(I(V;U) - \epsilon_n^{(3)})\}.
\end{aligned}$$

where Eq.(4.25) comes from the fact that $\boldsymbol{u}_i$ is selected independently of $(\boldsymbol{v},\boldsymbol{y})$, Eq.(4.26) from Lemma 4.5.1, and

$$\begin{aligned}
\epsilon_n^{(3)} &= \epsilon_n^{(3)}(|\mathcal{V}, |\mathcal{U}|, 2|\mathcal{X} \times \mathcal{Y}|\delta_n, 4|\mathcal{X}||\mathcal{X} \times \mathcal{Y}|\delta_n) \\
&\to \infty \quad (n \to \infty).
\end{aligned}$$

By setting L_U as

$$L_U \leq \exp\{n(I(V;U) - l_{21}\gamma)\},\ l_{21} > 0,$$

$l_{21}\gamma > \epsilon_n^{(3)}$, we have

$$\lim_{n\to\infty} \Pr\{E_4\} = 0$$

.

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

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

$$L_U \quad \le \quad \exp\{n(I(W;U) - l_{22}\gamma)\},\ l_{22} > 0$$

to vanish the encoding/decoding errors.

<u>*Rate evaluation:* $\varphi_n^{(2)}$</u>

The encoder sends the indexes of the bin using

$$R_2 = \frac{1}{n}\log M_V \quad \ge \quad I(Y;V) + m_1\gamma$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as $1/n \log M_V \ge I(Y;V)$.

<u>*Rate evaluation:* $\varphi_n^{(3)}$</u>

In the same way as $\varphi_n^{(2)}$, we obtain the coding rate as $R_3 = 1/n \log M_W \ge I(X;W)$.

<u>*Rate evaluation:* $\varphi_n^{(1)}$</u>

The encoder sends the indexes of the bin using

$$\begin{aligned} R_1 &= \frac{1}{n}\log N_U \\ &= \frac{1}{n}\log\frac{M_U}{L_U} \\ &\ge \max\{I(VX;U) + m_{21}\gamma, I(WY;U) + m_{22}\gamma\} \\ &\quad - \min\{I(V;U) - l_{21}\gamma, I(W;U) - l_{22}\gamma\} \end{aligned}$$

bits per letter. Since $\gamma > 0$ is arbitrary, we obtain the coding rate as $\max\{I(VX;U), I(WY;U)\} - \min\{I(V;U), I(W;U)\}$.

This completes the proof of Theorem 4.4.2. □

4.5.5 Proof of Lemma 4.5.1

Proof. First, we show the right inequality. For any $\boldsymbol{x} \in T_X^n(\delta_n^{(1)})$, we have

$$\begin{aligned} &\sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T_{UX}(\delta_n^{(2)})}} P_U^n(\boldsymbol{u}) \\ &= \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T_{UX}^n(\delta_n^{(2)}),\boldsymbol{u}\in T_U^n(|\mathcal{X}|\delta_n^{(2)})}} P_U^n(\boldsymbol{u}) \qquad (4.27) \\ &\le \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T_{UX}^n(\delta_n^{(2)}),\boldsymbol{u}\in T_U^n(|\mathcal{X}|\delta_n^{(2)})}} \exp\{-n(H(U) - \widetilde{\epsilon}_n^{(1)})\} \qquad (4.28) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_n^{(2)})}} \exp\{-n(H(U)-\widetilde{\epsilon}_n^{(1)})\} && (4.29)\\
&\le \sum_{\boldsymbol{u}\in T^n_{U|\boldsymbol{x}}(\delta_n^{(1)}+\delta_n^{(2)})} \exp\{-n(H(U)-\widetilde{\epsilon}_n^{(1)})\} && (4.30)\\
&\le \exp\{n(H(U|X)+\widetilde{\epsilon}_n^{(2)})\}\exp\{-n(H(U)-\widetilde{\epsilon}_n^{(1)})\} && (4.31)\\
&= \exp\{-n(I(X;U)-(\widetilde{\epsilon}_n^{(1)}+\widetilde{\epsilon}_n^{(2)}))\}\\
&= \exp\{-n(I(X;U)-\epsilon_n^{(2)})\}
\end{aligned}
$$

where Eqs. (4.27)(4.29) (4.30) come from Lemma 2.3.5, Eq. (4.28) from Lemma 2.3.2, Eq. (4.31) from Lemma 2.3.4, and

$$
\begin{aligned}
\widetilde{\epsilon}_n^{(1)} &= \widetilde{\epsilon}_n^{(1)}(|\mathcal{X}|\delta_n^{(2)},|\mathcal{U}|),\\
\widetilde{\epsilon}_n^{(2)} &= \widetilde{\epsilon}_n^{(2)}(|\mathcal{U}|,|\mathcal{X}|,\delta_n^{(1)},\delta_n^{(1)}+\delta_n^{(2)}),\\
\epsilon_n^{(2)} &= \widetilde{\epsilon}_n^{(1)}+\widetilde{\epsilon}_n^{(2)}.
\end{aligned}
$$

Next, we show the left inequality. From Lemma 2.3.5 for $\delta_2>\delta_1>0$ we have

$$
\boldsymbol{u}\in T^n_{U|\boldsymbol{x}}(\delta_2-\delta_1) \text{ and } \boldsymbol{x}\in T^n_X(\delta_1) \quad\Rightarrow\quad (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_2). \tag{4.32}
$$

Then, we obtain

$$
\begin{aligned}
&\sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_n^{(2)})}} P_U(\boldsymbol{u})\\
&= \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_n^{(2)}),\boldsymbol{u}\in T^n_U(|\mathcal{X}|\delta_n^{(2)})}} P_U(\boldsymbol{u}) && (4.33)\\
&\ge \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_n^{(2)}),\boldsymbol{u}\in T^n_U(|\mathcal{X}|\delta_n^{(2)})}} \exp\{-n(H(U)+\widetilde{\epsilon}_n^{(1)})\} && (4.34)\\
&= \sum_{\substack{\boldsymbol{u}\in\mathcal{U}^n:\\ (\boldsymbol{u},\boldsymbol{x})\in T^n_{UX}(\delta_n^{(2)})}} \exp\{-n(H(U)+\widetilde{\epsilon}_n^{(1)})\} && (4.35)\\
&\ge \sum_{\boldsymbol{u}\in T_{U|\boldsymbol{x}}-n(\delta_n^{(2)}-\delta_n^{(1)})} \exp\{-n(H(U)+\widetilde{\epsilon}_n^{(1)})\} && (4.36)\\
&\ge \exp\{n(H(U|X)-\widetilde{\epsilon}_n^{(3)})\}\exp\{-n(H(U)+\widetilde{\epsilon}_n^{(1)})\} && (4.37)\\
&= \exp\{-n(I(X;U)+(\widetilde{\epsilon}_n^{(1)}+\widetilde{\epsilon}_n^{(3)}))\}\\
&= \exp\{-n(I(X;U)+\epsilon_n^{(1)})\},
\end{aligned}
$$

where Eqs. (4.33)(4.35) comes from Lemma 2.3.5, Eq. (4.34) from Lemma 2.3.2, Eq. (4.36) from Eq. (4.32), Eq. (4.37) from Lemma 2.3.4,

$$
\begin{aligned}
\widetilde{\epsilon}_n^{(3)} &= \widetilde{\epsilon}_n^{(2)}(|\mathcal{U}|,|\mathcal{X}|,\delta_n^{(1)},\delta_n^{(2)}-\delta_n^{(1)}),\\
\epsilon_n^{(1)} &= \widetilde{\epsilon}_n^{(1)}+\widetilde{\epsilon}_n^{(3)}.
\end{aligned}
$$

This completes the proof of Lemma 4.5.1. □

Chapter 5

Universal coding for complementary delivery

5.1 Introduction

This chapter deals with a universal coding problem for the complementary delivery coding system investigated in Chapter 4. Although Chapter 4 considered lossy configurations, this chapter considers a lossless configuration. An explicit construction of fixed-to-fixed length (FF) universal codes is presented. In this construction, a codebook can be converted into a bipartite graph, and therefore encoding can be regarded as edge coloring of the bipartite graphs. The upper and lower bounds of the error probabilities are clarified by methods of types. Next, a construction of fixed-to-variable (FV) universal codes is considered. This can also be constructed in a similar manner to FF codes. Overflow and underflow probabilities are also evaluated.

The complementary delivery coding system can be easily extended to coding systems that comprise multiple messages and decoders: Messages from multiple correlated sources are jointly encoded, and each decoder has access to some of the messages to enable the decoder to reproduce the other messages. Henceforth we call such kind of coding systems *generalized complementary delivery coding systems*. Willems et al. [66, 72] considered an example of the generalized complementary delivery coding system that includes three sources and three decoders. Fig. 5.1 shows a block diagram of the coding systems investigated by Willems et al. This chapter also deals with a universal coding scheme for generalized complementary delivery coding systems. First, an explicit construction of FF universal codes is presented. In this construction, a codebook can be expressed as a certain kind of graphs, and therefore encoding is equivalent to vertex coloring of the graphs. When the number of decoders is 2, the graph can be translated into an equivalent bipartite graph, and therefore the encoding scheme equals one for the (original) complementary delivery coding system. The upper and lower bounds of the error probabilities are also clarified by methods of types. Next, a construction of FV universal codes is considered. This can also be constructed in a similar manner to FF codes. Overflow and underflow probabilities are also evaluated. FV universal codes can be also constructed in a similar manner.

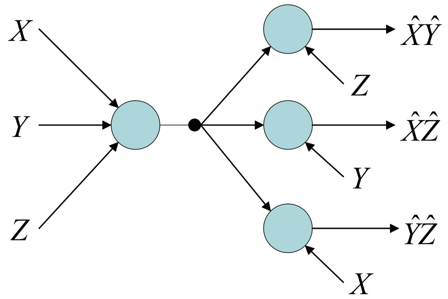

Figure 5.1: The generalized complementary delivery coding system investigated by Willems et al.

5.2 Fixed-length coding for two sources

5.2.1 Code construction

This section shows an explicit construction of universal lossless codes for the complementary delivery coding system defined by Definition 4.2.1. The coding scheme is as follows:
[Encoding]

1. Determine a set $\mathcal{T}_n(R)$ of joint types as

$$\begin{aligned}\mathcal{T}_n(R) \quad = \quad & \{Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) : \\ & \max\{H(V|Q_X), H(W|Q_Y)\} \leq R, \\ & Q_{XY} = Q_X V = Q_Y W, \\ & V \in \mathcal{V}_n(\mathcal{Y}|Q_X), W \in \mathcal{V}_n(\mathcal{X}|Q_Y)\},\end{aligned}$$

   where $R > 0$ is a given coding rate. We note that the joint type Q_{XY} specifies the types Q_X, Q_Y, and the conditional types V and W.

2. Create a table (henceforth we call this a *coding table*, see Figure 5.2 left) for each joint type $Q_{XY} \in \mathcal{T}_n(R)$. Each row of the coding table corresponds to a sequence $\boldsymbol{x} \in T^n_{Q_X}$, and each column corresponds to a sequence $\boldsymbol{y} \in T^n_{Q_Y}$.

3. Mark cells that correspond to sequence pairs $(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{Q_{XY}}$ (see Figure 5.2 middle). Codewords will be given only to sequence pairs that correspond to marked cells.

4. Fill the marked cells with $\exp(nR)$ different symbols such that each symbol occurs at most once in each row and at most once in each column. An example of symbol filling is shown on the right in Figure 5.2 right.

5. For a given pair of sequences $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ with the joint type Q_{XY}, if $Q_{XY} \in \mathcal{T}_n(R)$, the index assigned to the joint type Q_{XY} of $(\boldsymbol{x}, \boldsymbol{y})$ is the first part of the codeword, and the symbol filling the cell of $(\boldsymbol{x}, \boldsymbol{y})$ in the coding table of Q_{XY} is determined as the second part of the codeword. For the sequence pairs $(\boldsymbol{x}, \boldsymbol{y})$ whose joint type Q_{XY} does not belong to $\mathcal{T}_n(R)$, the corresponding codeword is determined arbitrarily and an encoding error is declared.

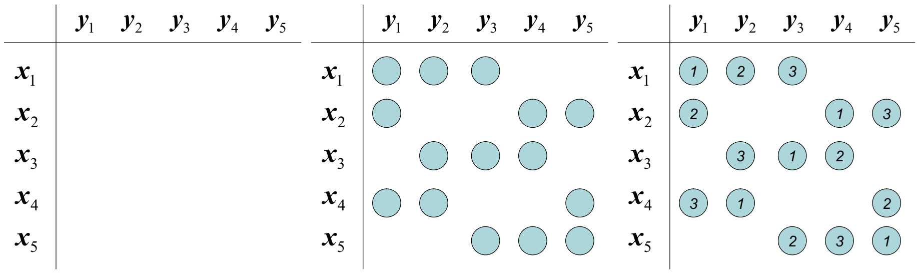

Figure 5.2: Example of coding scheme (left) Coding table (middle) Positions where codewords will be provided (right) Provided codewords

[Decoding: $\widehat{\varphi}_n^{(1)}$] (Almost the same as for $\widehat{\varphi}_n^{(2)}$)

1. Find the coding table of the type $\widehat{Q}_{XY}$ that corresponds to the first part of the received codeword. The decoder can find the coding table used in the encoding scheme if no encoding error occurs. In this case, $\widehat{Q}_{XY}$ should be Q_{XY}.

2. Find the cell filled with the second part of the received codeword from the column of the side information sequence $\boldsymbol{y} \in T^n_{Q_Y}$. The sequence $\widehat{\boldsymbol{x}} \in T^n_{Q_X}$ that corresponds to the row of the cell found in this step is reproduced.

First, we show the existence of such coding tables. To do this, we introduce the following two lemmas.

Lemma 5.2.1. *For a given coding table of a joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, the number of marked cells in every row of the coding table, $N_y(Q_{XY})$, is a constant value that is less than* $\exp(nR)$*, and the number of marked cells in every column of the coding table, $N_x(Q_{XY})$, is also a constant value of less than* $\exp(nR)$*, both of which depend solely on the joint type Q_{XY}.*

Proof. Note that the number of marked cells in each row equals the cardinality of the V-shell $T^n_V(\boldsymbol{x})$ for the sequence $\boldsymbol{x} \in T^n_{Q_X}$ that corresponds to the row. The cardinality of V-shells $T^n_V(\boldsymbol{x})$ is constant for a given joint type Q_{XY} and any sequences $\boldsymbol{x} \in T^n_{Q_X}$, this cardinality is bounded as follows:

$$\begin{aligned} N_y(Q_{XY}) &= |T^n_V(\boldsymbol{x})| \\ &\le \exp\{nH(V|Q_X)\} \\ &\le \exp(nR), \end{aligned} \tag{5.1}$$

where Eq. (5.1) comes from Lemma 2.2.2. In the same way, the number of marked cells in each column equals the cardinality of the V-shell $T_W(\boldsymbol{y})$ for the sequence $\boldsymbol{y} \in T^n_{Q_Y}$ that corresponds to the column, and therefore it can be bounded as

$$N_y(Q_{XY}) = |T^n_W(\boldsymbol{y})| \quad \le \quad \exp(nR).$$

This concludes the proof of Lemma 5.2.1. $\square$

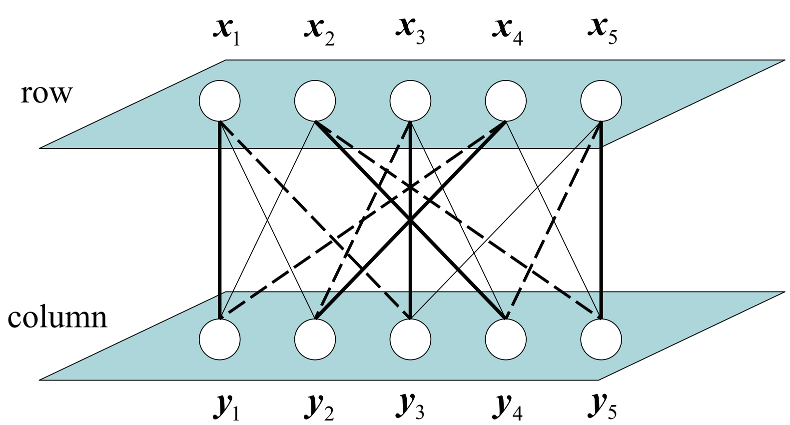

Figure 5.3: Example of a bipartite graph ($m_x = m_y = 5$, $n_x = n_y = 3$, equivalent to the table in Fig. 5.2 right)

Lemma 5.2.2. *For given integers m_x, m_y, n_x and n_y that satisfy $m_x \geq n_x$ and $m_y \geq n_y$, there exists an $m_x \times m_y$ table filled with $\max(n_x, n_y)$ different symbols such that*

- *at most n_y cells are filled with a certain symbol for each row (blank cells are possible),*
- *at most n_x cells are filled with a certain symbol for each column (blank cells are possible),*
- *each symbol occurs at most once in each row and at most once in each column.*

Proof. The table mentioned in this lemma is equivalent to a bipartite graph such that

- each node in one set corresponds to a row in the table, and each node in the other set corresponds to a column in the table,
- each edge corresponds to a cell in the table, to which a certain symbol is assigned
- $\max(n_x, n_y)$ different symbols are assigned to edges,
- each symbol occurs at most once on edges linked to each node.

Figure 5.3 shows an example of such a graph. Since the degree of the above bipartite graph is $\max(n_x, n_y)$, Lemma 2.4.3 ensures the existence of the above bipartite graph. This concludes the proof of Lemma 5.2.2. □

From Lemmas 5.2.1 and 5.2.2, we can easily show the existence of coding tables by setting $m_x = |T^n_{Q_X}|$, $m_y = |T^n_{Q_Y}|$, $n_x = |T^n_W(\boldsymbol{y})|$ and $n_y = |T^n_V(\boldsymbol{x})|$ in Lemma 5.2.2.

5.2.2 Coding theorems

We can obtain the following theorem for universal lossless f-FCD codes constructed in Section 5.2.1. The proof is similar to that of the theorem of universal coding for a single source.

Theorem 5.2.1. *For a given real number $R > 0$, there exists a universal lossless f-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ such that for any integer $n \geq 1$ and any DMS (X, Y) with a generic distribution $P_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$*

$$\frac{1}{n} \log M_n \leq R + \epsilon_n(1), \tag{5.2}$$

$$e_n^{(1)} + e_n^{(2)} \leq \exp\left\{ -n \left(-\epsilon_n(2) + \min_{Q_{XY} \in \mathcal{T}_n^c(R)} D(Q_{XY} \| P_{XY}) \right) \right\}, \tag{5.3}$$

where

$$\begin{aligned} \epsilon_n(N) &\stackrel{\text{def.}}{=} \frac{1}{n}\{|\mathcal{X} \times \mathcal{Y}| \log(n+1) + \log N\} \\ &\to 0 \quad (n \to \infty). \end{aligned} \tag{5.4}$$

Note that $\mathcal{T}_n^c(R) = \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) - \mathcal{T}_n(R)$.

Proof. Lemmas 5.2.1 and 5.2.2 ensure the existence of a coding table for every joint type $Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. From the coding scheme, the size of the codeword set is bounded as

$$\begin{aligned} M_n &\leq |\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \exp(nR) \\ &\leq (n+1)^{|\mathcal{X} \times \mathcal{Y}|} \exp(nR), \end{aligned}$$

where the last inequality comes from Lemma 2.2.1. This implies Eq.(5.2). Next, we evaluate decoding error probabilities. Since every sequence pair $(\boldsymbol{x}, \boldsymbol{y}) \in T_{Q_{XY}}^n$ that satisfies $Q_{XY} \in \mathcal{T}_n(R)$ is reproduced correctly at the decoder, the sum of error probabilities is bounded as

$$\begin{aligned} & e_n^{(1)} + e_n^{(2)} \\ \leq\ & 2 \Pr\left\{ (X^n, Y^n) \in T_{Q_{XY}}^n : Q_{XY} \in \mathcal{T}_n^c(R) \right\} \\ \leq\ & 2 \sum_{Q_{XY} \in \mathcal{T}_n^c(R)} \exp\{-nD(Q_{XY} \| P_{XY})\} \qquad (5.5) \\ \leq\ & 2 \sum_{Q_{XY} \in \mathcal{T}_n^c(R)} \exp\left\{ -n \min_{Q_{XY} \in \mathcal{T}_n^c(R)} D(Q_{XY} \| P_{XY}) \right\} \\ \leq\ & 2(n+1)^{|\mathcal{X} \times \mathcal{Y}|} \exp\left\{ -n \min_{Q_{XY} \in \mathcal{T}_n^c(R)} D(Q_{XY} \| P_{XY}) \right\} \qquad (5.6) \\ \leq\ & \exp\left\{ -n \left(-\epsilon_n(2) + \min_{Q_{XY} \in \mathcal{T}_n^c(R)} D(Q_{XY} \| P_{XY}) \right) \right\}, \end{aligned}$$

where Eq.(5.5) comes from Lemma 2.2.3 and Eq.(5.6) comes from Lemma 2.2.1. This completes the proof of Theorem 5.2.1. □

We can see that for any real value $R \geq R_f(X, Y)$ we have

$$\min_{Q_{XY} \in \mathcal{T}_n^c(R)} D(Q_{XY} \| P_{XY}) > 0.$$

This implies that any real value $R \geq R_f(X,Y)$ is *a universal lossless f-FCD achievable rate* of (X,Y), namely, there exists a universal lossless f-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R,$$
$$\lim_{n\to\infty} e_n^{(1)} = \lim_{n\to\infty} e_n^{(2)} = 0.$$

The following converse theorem indicates that the error exponent obtained in Theorem 5.2.1 is tight.

Theorem 5.2.2. *Any lossless f-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *for the DMS* (X,Y) *must satisfy*

$$e_n^{(1)} + e_n^{(2)} \quad \geq \quad \exp\left\{-n\left(\epsilon_n(2) + \min_{Q_{XY}\in\mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{XY}\|P_{XY})\right)\right\}$$

for any integer $n \geq 1$ *and a given coding rate* $R = 1/n\log M_n > 0$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

Proof. Note that the number of sequences to be decoded correctly for each decoder are at most $\exp(nR)$. Here, let us consider a joint type $Q_{XY} \in \mathcal{T}_n^c(R+\epsilon_n(2))$. Lemma 2.2.2 and the definition of $\mathcal{T}_n^c(R+\epsilon_n(2))$ imply that for $(\boldsymbol{x},\boldsymbol{y}) \in T_{Q_{XY}}^n$ we have

$$\begin{aligned}
&\max\{|T_V^n(\boldsymbol{x})|, |T_W^n(\boldsymbol{y})|\} \\
&\quad \geq \quad (n+1)^{-|\mathcal{X}\times\mathcal{Y}|}\max[\exp\{nH(V|Q_X)\}, \exp\{nH(W|Q_Y)\}] \\
&\quad \geq \quad (n+1)^{-|\mathcal{X}\times\mathcal{Y}|}\exp\{n(R+\epsilon_n(2))\} \\
&\quad = \quad 2\exp(nR).
\end{aligned}$$

Therefore, at least half of the sequences in the V-shell $T_V^n(\boldsymbol{x})$ will not be decoded correctly at the decoder $\varphi_n^{(2)}$, or at least half of sequences in the V-shell $T_W^n(\boldsymbol{y})$ will not be decoded correctly at the decoder $\varphi_n^{(1)}$. Thus, the sum of error probabilities is bounded as

$$\begin{aligned}
&e_n^{(1)} + e_n^{(2)} \\
&\quad \geq \quad \frac{1}{2}\sum_{Q_{XY}\in\mathcal{T}_n^c(R+\epsilon_n(2))} \Pr\{(X^n,Y^n) \in T_{Q_{XY}}\} \\
&\quad \geq \quad \frac{1}{2}(n+1)^{-|\mathcal{X}\times\mathcal{Y}|}\sum_{Q_{XY}\in\mathcal{T}_n^c(R+\epsilon_n(2))} \exp\{-nD(Q_{XY}\|P_{XY})\} \qquad (5.7) \\
&\quad \geq \quad \frac{1}{2}(n+1)^{-|\mathcal{X}\times\mathcal{Y}|}\exp\left\{-n\min_{Q_{XY}\in\mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{XY}\|P_{XY})\right\} \\
&\quad = \quad \exp\left\{-n\left(\epsilon_n(2) + \min_{Q_{XY}\in\mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{XY}\|P_{XY})\right)\right\},
\end{aligned}$$

where Eq. (5.7) comes from Lemma 2.2.3. This concludes the proof of Theorem 5.2.2. □

The following corollary is directly derived from Theorems 5.2.1 and 5.2.2.

Corollary 5.2.1. *For a given real number $R > 0$, there exists a universal lossless f-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ such that for any DMS (X, Y)*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R,$$
$$\lim_{n\to\infty} -\frac{1}{n} \log(e_n^{(1)} + e_n^{(2)}) = \min_{Q_{XY}\in\mathcal{T}^c(R)} D(Q_{XY}\|P_{XY}),$$

where

$$\begin{aligned}\mathcal{T}(R) &= \{Q_{XY} \in \mathcal{M}(\mathcal{X}\times\mathcal{Y}) : \\ &\quad \max\{H(V|Q_X), H(W|Q_Y)\} \le R, \\ &\quad Q_{XY} = Q_X V = Q_Y W, \\ &\quad V \in \mathcal{M}(\mathcal{Y}|Q_X), W \in \mathcal{M}(\mathcal{X}|Q_Y)\}.\end{aligned}$$

In a similar manner, we can investigate the probability such that the original sequence pair is correctly reproduced. The following theorem shows the lower bound of the probability that can be attained by the proposed coding scheme.

Theorem 5.2.3. *For a given real number $R > 0$, there exists a universal lossless f-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ such that for any integer $n \ge 1$ and any DMS (X, Y)*

$$\begin{aligned}\frac{1}{n}\log M_n &\le R + \epsilon_n(1), \\ 1 - (e_n^{(1)} + e_n^{(2)}) &\ge \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{XY}\in\mathcal{T}_n(R)} D(Q_{XY}\|P_{XY})\right)\right\},\end{aligned}$$

where $\epsilon_n(N)$ is defined in Eq. (5.4).

Proof. The first inequality is derived in the same way as the proof of Theorem 5.2.1. Next, we evaluate the probability such that the original sequence pair is correctly reproduced. Since every sequence pair $(\boldsymbol{x}, \boldsymbol{y}) \in T_{Q_{XY}}^n$ that satisfies $Q_{XY} \in \mathcal{T}_n(R)$ is reproduced correctly at the decoder, the sum of the probabilities is bounded as

$$\begin{aligned}&1 - (e_n^{(1)} + e_n^{(2)}) \\ &\quad\ge \Pr\left\{(X^n, Y^n) \in T_{Q_{XY}}^n : Q_{XY} \in \mathcal{T}_n(R)\right\} \\ &\quad\ge \sum_{Q_{XY}\in\mathcal{T}^n(R)} (n+1)^{-|\mathcal{X}\times\mathcal{Y}|} \exp\{-nD(Q_{XY}\|P_{XY})\} \qquad (5.8) \\ &\quad\ge (n+1)^{-|\mathcal{X}\times\mathcal{Y}|} \exp\left\{-n \min_{Q_{XY}\in\mathcal{T}^n(R)} D(Q_{XY}\|P_{XY})\right\} \\ &\quad= \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{XY}\in\mathcal{T}_n(R)} D(Q_{XY}\|P_{XY})\right)\right\},\end{aligned}$$

where Eq. (5.8) comes from Lemma 2.2.3. This completes the proof of Theorem 5.2.3. □

The following converse theorem indicates that the error exponent obtained in Theorem 5.2.3 might not be tight.

Theorem 5.2.4. *Any lossless f-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *for the DMS* (X, Y) *must satisfy*

$$
\begin{aligned}
& 1-(e_n^{(1)}+e_n^{(2)}) \\
& \quad \leq \quad \exp\Big\{-n\Big(-\epsilon_n(1)+\min_{Q_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} \\
& \qquad\qquad |\max(H(V|Q_X),H(W|Q_Y))-(R+\epsilon_n(1))|^{+}+D(Q_{XY}\|P_{XY})\Big)\Big\}
\end{aligned}
$$

for any integer $n \geq 1$ *and a given coding rate* $R = 1/n \log M_n > 0$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

Proof. Note that the number of sequences to be decoded correctly for each decoder are at most $\exp(nR)$. Here, let us consider a joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ such that $Q_{XY} = Q_X V = Q_Y W$, $V \in \mathcal{M}(\mathcal{Y}|Q_X)$ and $W \in \mathcal{M}(\mathcal{X}|Q_Y)\}$. The ratio $r_c(Q_{XY})$ of sequences in the sequence set $T_{Q_{XY}}$ such that the sequences are correctly reproduced is at most

$$
\begin{aligned}
& r_c(Q_{XY}) \\
& \quad \leq \quad \min\left\{\min\left(\frac{\exp(nR)}{|T_V^n(\boldsymbol{x})|},\frac{\exp(nR)}{|T_W^n(\boldsymbol{y})|}\right),1\right\} \\
& \quad \leq \quad \min\Big[\exp(nR)\cdot(n+1)^{|\mathcal{X}\times\mathcal{Y}|}\exp\{-n\max(H(V|Q_X),H(W|Q_Y))\},1\Big] && (5.9) \\
& && (5.10) \\
& \quad = \quad \min\Big[\exp\left\{n\left(R+\gamma_n\right)\right\}\exp\{-n\max(H(V|Q_X),H(W|Q_Y))\},1\Big] \\
& \quad = \quad \exp\Big\{-n\left|\max\{H(V|Q_X),H(W|Q_Y)\}-(R+\gamma_n)\right|^{+}\Big\}
\end{aligned}
$$

where Eq. (5.10) comes from Lemma 2.2.2. Therefore, the probability $P_c(Q_{XY})$ such that the original sequence pair with type Q_{XY} is correctly reproduced is bounded as

$$
\begin{aligned}
& P_c(Q_{XY}) \\
& \quad \leq \quad r_c(Q_{XY})\Pr\{(X^n,Y^n)\in T_{Q_{XY}}^n\} \\
& \quad \leq \quad \exp\big\{-n\big|\max(H(V|Q_X),H(W|Q_Y))-(R+\gamma_n)\big|^{+}+D(Q_{XY}\|P_{XY})\big\}.
\end{aligned}
$$

Thus, the sum of the probabilities is obtained as

$$
\begin{aligned}
& 1-(e_n^{(1)}+e_n^{(2)}) \\
& \quad \leq \sum_{Q_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} P_c(Q_{XY}) \\
& \quad \leq \sum_{Q_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} \exp\big[-n\big|\max\{H(V|Q_X),H(W|Q_Y)\}-(R+\gamma_n)\big|^{+}+D(Q_{XY}\|P_{XY})\big] \\
& \quad \leq \exp\big[-n\min_{Q_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})}\big(\big|\max\{H(V|Q_X),H(W|Q_Y)\} \\
& \qquad\qquad -(R+\gamma_n)\big|^{+}+D(Q_{XY}\|P_{XY})\big)\big].
\end{aligned}
$$

This completes the proof of Theorem 5.2.4. □

We can see that for any real value $R \geq R_f(X,Y)$ we have

$$\begin{aligned}
&\min_{Q_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} \left[\left| \max\{H(V|Q_X), H(W|Q_Y)\} - R \right|^+ + D(Q_{XY} \| P_{XY})) \right] \\
&= \left| \max(H(P_{Y|X}|P_X), H(P_{X|Y}|P_Y)) - R \right|^+ \\
&= 0.
\end{aligned}$$

On the other hand, for any real value $R < R_f(X,Y)$ we have

$$\begin{aligned}
&\min_{Q_{XY} \in \mathcal{T}(R)} D(Q_{XY} \| P_{XY}) \\
&\geq \min_{Q_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})} \left(\left| \max(H(V|Q_X), H(W|Q_Y)) - R \right|^+ + D(Q_{XY} \| P_{XY})) \right\} \\
&\geq 0.
\end{aligned}$$

This implies that the error exponent obtained in Theorem 5.2.3 might not be tight.

5.3 Variable-length coding for two sources

This section discusses variable-length coding for the complementary delivery coding system, and shows an explicit construction of universal variable-length codes. The coding scheme is similar to that of fixed-length codes, and also utilizes the coding tables defined in Section 5.2.1.

5.3.1 Formulation

Definition 5.3.1. (Lossless fixed-to-variable fully-informed complementary delivery (lossless v-FCD) code)
A sequence $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ of codes $(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})$ is a lossless v-FCD code for the source $(\boldsymbol{X}, \boldsymbol{Y})$ if and only if

$$\begin{aligned}
\varphi_n &: \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{B}^* \\
\widehat{\varphi}_n^{(1)} &: \varphi_n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{Y}^n \to \mathcal{X}^n, \\
\widehat{\varphi}_n^{(2)} &: \varphi_n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{X}^n \to \mathcal{Y}^n, \\
e_n^{(1)} &= \Pr\left\{ X^n \neq \widehat{X}^n \right\} = 0, \\
e_n^{(2)} &= \Pr\left\{ Y^n \neq \widehat{Y}^n \right\} = 0,
\end{aligned}$$

where

$$\begin{aligned}
\widehat{X}^n &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(1)}(\varphi_n(X^n, Y^n), Y^n), \\
\widehat{Y}^n &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(2)}(\varphi_n(X^n, Y^n), X^n),
\end{aligned}$$

and the image of φ_n is a prefix set.

Definition 5.3.2. (Lossless v-FCD achievable rate)
R is a lossless v-FCD achievable rate of the source $(\boldsymbol{X}, \boldsymbol{Y})$ if and only if there

exists a lossless v-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ for the source $(\boldsymbol{X}, \boldsymbol{Y})$ such that

$$\limsup_{n\to\infty} \frac{1}{n} E\left[l(\varphi_n(X^n, Y^n))\right] \quad \leq \quad R,$$

where $l(\cdot) : \mathcal{B}^* \to \mathcal{R}$ is a length function.

Definition 5.3.3. (Inf lossless v-FCD achievable rate)

$$R_v(\boldsymbol{X}, \boldsymbol{Y}) = \inf\{R : \\ R \text{ is a lossless v-FCD achievable rate of } (\boldsymbol{X}, \boldsymbol{Y})\}.$$

5.3.2 Code construction

We can construct universal lossless v-FCD codes in a similar manner to universal lossless f-FCD codes. Note that the coding rate depends on the type of sequence pair to be encoded, whereas the coding rate is fixed beforehand for fixed-length coding. The coding scheme is described as follows:
[Encoding]

1. Create a coding table for each joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ in the same way as Step 2 of Section 5.2.1.
2. Mark cells that correspond to sequence pairs $(\boldsymbol{x}, \boldsymbol{y}) \in T_{Q_{XY}}$.
3. Fill the marked cells on the coding table with different $\max\{|T_V^n(\boldsymbol{x})|, |T_W^n(\boldsymbol{y})|\}$ symbols such that each symbol occurs at most once in each row and at most once in each column, where $\boldsymbol{x} \in T_{Q_X}^n$, $\boldsymbol{y} \in T_{Q_Y}^n$.
4. For a given pair of sequences $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, the number (index) assigned to the joint type Q_{XY} of $(\boldsymbol{x}, \boldsymbol{y})$ is the first part of the codeword, and the symbol filling the cell of $(\boldsymbol{x}, \boldsymbol{y})$ in the coding table of Q_{XY} is determined as the second part of the codeword.

[Decoding]
Decoding can be accomplished in almost the same way as the fixed-length coding. Note that the decoder can always find the coding table used in the encoding scheme.

5.3.3 Coding theorems

We begin by showing a theorem for (non-universal) variable-length coding, which indicates that the inf coding rate of variable-length coding is the same as that of fixed-length coding.

Theorem 5.3.1. (Coding theorem of lossless v-FCD code)
For any DMS (X, Y), we have

$$R_v(X, Y) \quad = \quad R_f(X, Y) = \max\{H(X|Y), H(Y|X)\}.$$

Proof.
[Direct part]
We can apply an *achievable* lossless f-FCD codes (fixed-length codes). The encoder φ_n assigns the same codeword as that of the fixed-length code to a sequence pair $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ that is correctly reproduced by the fixed-length code. Otherwise, the encoder sends the sequence pair itself as a codeword.

The above lossless v-FCD code can always reproduce the original sequence pair at the decoders, and it attains the desired coding rate.

[Converse part]
We can prove the converse part in a similar manner to that for fixed-length coding. Let a lossless v-FCD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ be given that satisfies the conditions of Definitions 5.3.1 and 5.3.2. From Definition 5.3.2, for any $\delta > 0$ there exists an integer $n_1 = n_1(\delta)$ and then for all $n \geq n_1(\delta)$, we can obtain

$$\frac{1}{n} E[l(\varphi_n(X^n, Y^n))] \quad \leq \quad R + \delta. \tag{5.11}$$

Here, let us define $A_n = \varphi_n(X^n, Y^n)$. Since the decoder $\widehat{\varphi}_n^{(1)}$ can always reproduce the original sequence X^n from the received codeword A_n and side information Y^n, we can see that

$$H(X^n | A_n Y^n) \quad = \quad 0. \tag{5.12}$$

Similarly, we can obtain

$$H(Y^n | A_n X^n) \quad = \quad 0.$$

Substituting A_n into Eq.(5.11), we have

$$\begin{aligned} n(R+\delta) &\geq E[l(A_n)] \\ &\geq H(A_n) &(5.13)\\ &\geq H(A_n | Y^n) \\ &\geq I(X^n; A_n | Y^n) \\ &= H(X^n | Y^n), &(5.14) \end{aligned}$$

where Eq. (5.13) comes from Lemma 2.1.3, and Eq. (5.14) from Eq. (5.12). Since we can select an arbitrarily small $\delta > 0$ for a sufficient large n, we can obtain

$$R \quad \geq \quad \frac{1}{n} H(X^n | Y^n) = H(X|Y).$$

In the same way, we also obtain

$$R \quad \geq \quad H(Y|X).$$

This completes the proof of Theorem 5.3.1. □

The following direct theorem for universal coding indicates that the coding scheme presented in Section 5.3.2 can achieve the inf achievable rate.

Theorem 5.3.2. *There exists a universal lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *such that for any integer* $n \geq 1$ *and any DMS* (X, Y)*, the overflow probability* $\overline{\rho}_n(R)$*, namely the probability that the length of a codeword exceeds a given real number* $R > 0$*, is bounded as*

$$\begin{aligned}\overline{\rho}_n(R) &\stackrel{\text{def.}}{=} \Pr\{l(\varphi_n(X^n, Y^n)) > nR\} \\ &\leq \exp\left\{-n\left(-\epsilon_n(1) + \min_{Q_{XY} \in \mathcal{T}_n^c(R-\epsilon_n(1))} D(Q_{XY}\|P_{XY})\right)\right\},\end{aligned}$$

where $\epsilon_n(N)$ *is defined in Eq. (5.4). This implies that there exists a universal lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} l(\varphi_n(X^n, Y^n)) \leq R_v(X, Y) \quad a.s. \tag{5.15}$$

Proof. The overflow probability can be obtained in the same way as an upper bound of the error probability of lossless f-FCD codes, which has been shown in the proof of Theorem 5.2.1. Thus, we have

$$\sum_{n=1}^{\infty} \Pr\left\{\frac{1}{n} l(\varphi_n(X^n, Y^n)) > R_v(X, Y) + \delta\right\} < \infty$$

for a given $\delta > 0$. From Borel-Cantelli's lemma [25, Lemma 4.6.3], we immediately obtain Eq.(5.15). This completes the proof of Theorem 5.3.2. □

The following converse theorem for variable-length coding indicates that the exponent of the overflow probability obtained in Theorem 5.3.2 is tight. This can be easily obtained in almost the same way as Theorem 5.2.2.

Theorem 5.3.3. *Any lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *for the DMS* (X, Y) *must satisfy*

$$\begin{aligned}&\overline{\rho}_n(R) \\ &\quad \geq \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{XY} \in \mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{XY}\|P_{XY})\right)\right\}.\end{aligned}$$

for a given real number $R > 0$ *and any integer* $n \geq 1$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

The following corollary is directly derived from Theorems 5.3.2 and 5.3.3.

Corollary 5.3.1. *There exists a universal lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *such that for any DMS* (X, Y)

$$\limsup_{n\to\infty} \frac{1}{n} l(\varphi_n(X^n, Y^n)) \leq R_v(X, Y) \quad a.s.$$

$$\lim_{n\to\infty} -\frac{1}{n} \log \overline{\rho}_n(R) = \min_{Q_{XY} \in \mathcal{T}^c(R)} D(Q_{XY}\|P_{XY})$$

Next, we investigate the underflow probability, namely the probability that the length of a codeword falls below a given real number $R > 0$. For this purpose, we present the following two theorems. The proofs are almost the same as those of Theorems 5.2.3 and 5.2.4.

Theorem 5.3.4. *There exists a universal lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *such that for any integer* $n \geq 1$ *and any DMS* (X, Y)*, the underflow probability* $\underline{\rho}_n(R)$ *is bounded as*

$$\begin{aligned}\underline{\rho}_n(R) &\stackrel{\text{def.}}{=} \Pr\{l(\varphi_n(X^n, Y^n)) < nR\} \\ &\leq \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{XY}\in\mathcal{T}_n(R-\epsilon_n(2))} D(Q_{XY}\|P_{XY})\right)\right\},\end{aligned}$$

where $\epsilon_n(N)$ *is defined in Eq. (5.4). This implies that there exists a universal lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *that satisfies*

$$\liminf_{n\to\infty} \frac{1}{n} l(\varphi_n(X^n, Y^n)) \quad \geq \quad R_v(X, Y) \quad a.s. \tag{5.16}$$

Theorem 5.3.5. *Any lossless v-FCD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *for the DMS* (X, Y) *must satisfy*

$$\begin{aligned}&\underline{\rho}_n(R) \\ &\leq \exp\Big\{-n\Big(-\epsilon_n(1) + \min_{Q_{XY}\in\mathcal{P}_n(\mathcal{X}\times\mathcal{Y})} \\ &\qquad |\max(H(V|Q_X), H(W|Q_Y)) - (R + \epsilon_n(1))|^+ + D(Q_{XY}\|P_{XY})\Big)\Big\}\end{aligned}$$

for a given real number $R > 0$ *and any integer* $n \geq 1$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

5.4 Fixed-length coding for multiple sources

5.4.1 Problem formulation

This section formulates the generalized complementary delivery coding problem, and shows the fundamental bound of the coding rate.

First, we formulate the coding problem. Fig. 5.4 represents a generalized complementary delivery coding system formulated in the following. This system is composed of N_s sources, one encoder and N_d decoders.

Definition 5.4.1. (Fixed-to-fixed fully-informed generalized complementary delivery (f-FGCD) code)
A sequence $\left\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\right\}_{n=1}^{\infty}$ of codes $(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})$ is an f-FGCD code for indices $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_j\}_{j=1}^{N_d}$ of side information sequences if and only if

$$\begin{aligned}&\varphi_n : \left\{\mathcal{X}^{(\mathcal{I}_{N_s})}\right\}^n \to \mathcal{I}_{M_n} \\ &\widehat{\varphi}_n^{(j)} : \mathcal{I}_{M_n} \times \left\{\mathcal{X}^{(\mathcal{S}_j^c)}\right\}^n \to \left\{\mathcal{X}^{(\mathcal{S}_j)}\right\}^n \quad \forall j_1, j_2 \in \mathcal{I}_{N_d}, \\ &\mathcal{S}_{j_1} \neq \mathcal{S}_{j_2} \quad \forall j_1, j_2 \in \mathcal{I}_{N_d},\end{aligned}$$

where $\mathcal{S}_j \subseteq \mathcal{I}_{N_s}$ $\forall j \in \mathcal{I}_{N_d}$. Since indices $\boldsymbol{\mathcal{S}}$ of side information sequences determine the coding system, in the following we identify $\boldsymbol{\mathcal{S}}$ with the coding system.

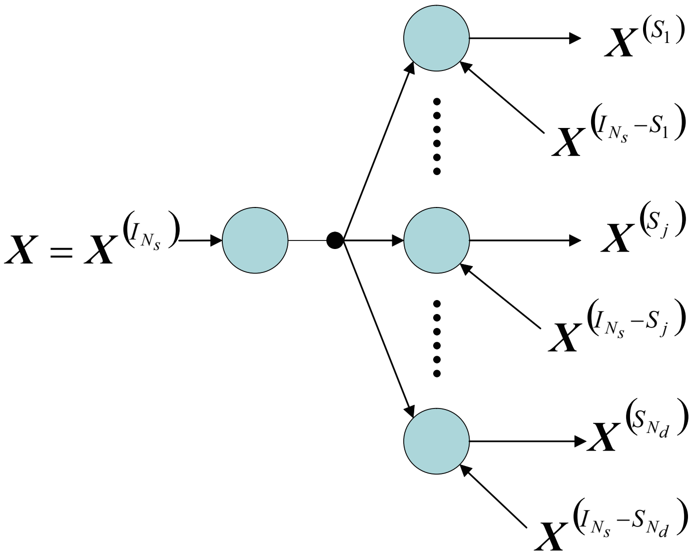

Figure 5.4: The generalized complementary delivery coding system

Definition 5.4.2. (Lossless f-FGCD achievable rate)
R is a lossless f-FGCD achievable rate of the source $\boldsymbol{X} = \boldsymbol{X}^{(\mathcal{I}_{N_s})}$ for the system $\boldsymbol{\mathcal{S}}$ if and only if there exists an f-FGCD code

$$\left\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\right\}_{n=1}^{\infty}, \boldsymbol{\mathcal{S}}$$

for the system $\boldsymbol{\mathcal{S}}$ such that

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log M_n &\leq R, \\ \lim_{n\to\infty} e_n^{(j)} &= 0 \quad \forall j \in \mathcal{I}_{N_d}, \end{aligned}$$

where

$$\begin{aligned} \widehat{\boldsymbol{X}}^{(\mathcal{S}_j)n} &\stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(j)}(\varphi_n(\boldsymbol{X}^n), \boldsymbol{X}^{(\mathcal{S}_j^c)n}), \\ e_n^{(j)} &= \Pr\left\{\boldsymbol{X}^{(\mathcal{S}_j)n} \neq \widehat{\boldsymbol{X}}^{(\mathcal{S}_j)n}\right\}. \end{aligned}$$

Definition 5.4.3. (Inf lossless f-FGCD achievable rate)

$$\begin{aligned} & R_F(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \\ &= \inf\{R | R \text{ is a lossless f-FGCD achievable rate of } \boldsymbol{X} \text{ for } \boldsymbol{\mathcal{S}}\}. \end{aligned}$$

Willems et al. [66, 72] investigated a coding problem where three users are physically separated but communicate with each other via a satellite, and determined the minimum coding rate when transmitting to and from the satellite. The downstream part of the coding system investigated by Willems et al. corresponds to the generalized complementary coding system, where $N_s = N_d = 3$, $\mathcal{S}_1 = \{1\}$, $\mathcal{S}_2 = \{2\}$ and $\mathcal{S}_3 = \{3\}$. Therefore, the following coding theorem can be obtained from the previous work by Willems et al.:

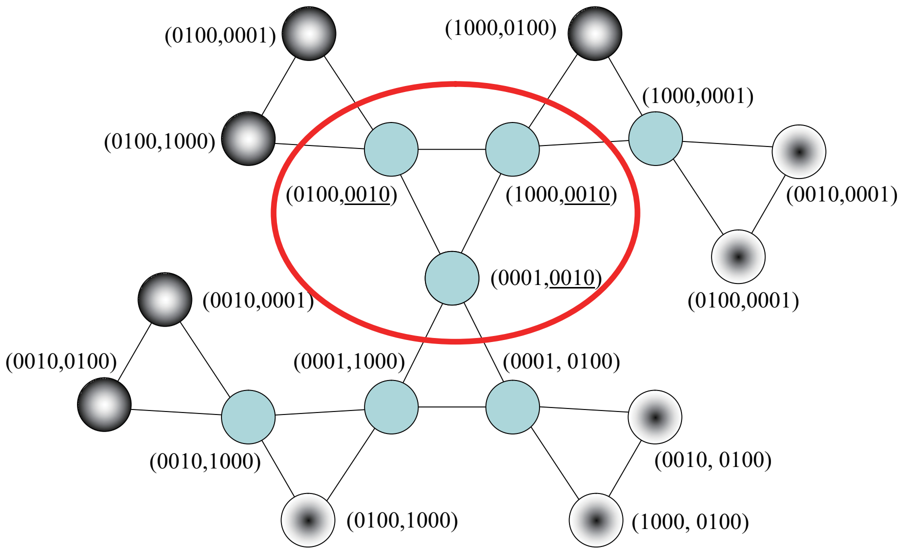

Figure 5.5: Example of coding graph, where each vertex whose center is colored with gray corresponds to another vertex whose verge is colorer with gray (Also see 2) of the encoding scheme)

Theorem 5.4.1. (Coding theorem of lossless f-FGCD codes for three users [72])
If $N_s = N_d = 3$, $\mathcal{S}_1 = \{1, 2\}$, $\mathcal{S}_2 = \{1, 3\}$ *and* $\mathcal{S}_3 = \{2, 3\}$, *then for any DMS* (X, Y)

$$R_F(X, Y, Z|\boldsymbol{\mathcal{S}}) \;=\; \max\{H(X, Y|Z), H(Y, Z|X), H(X, Z|Y)\}$$

It is easy to extend Theorem 5.4.1 to the following coding theorem for general cases:

Theorem 5.4.2. (Coding theorem of lossless f-FGCD codes for general cases) *For any DMS* (X, Y), *we have*

$$R_F(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \;=\; \max_{j \in \mathcal{I}_{N_d}} H\left(\boldsymbol{X}^{(\mathcal{S}_j)} \middle| \boldsymbol{X}^{(\mathcal{S}_j^c)}\right)$$

5.4.2 Code construction

This section shows an explicit construction of universal codes for the generalized complementary delivery coding system. The construction is similar to that for the (original) complementary delivery coding system. The proposed coding scheme is described as follows:

[Encoding]

1. Determine a set $\mathcal{T}_n(R)$ of joint types as

$$\mathcal{T}_n(R) \;=\; \{Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})}) :$$

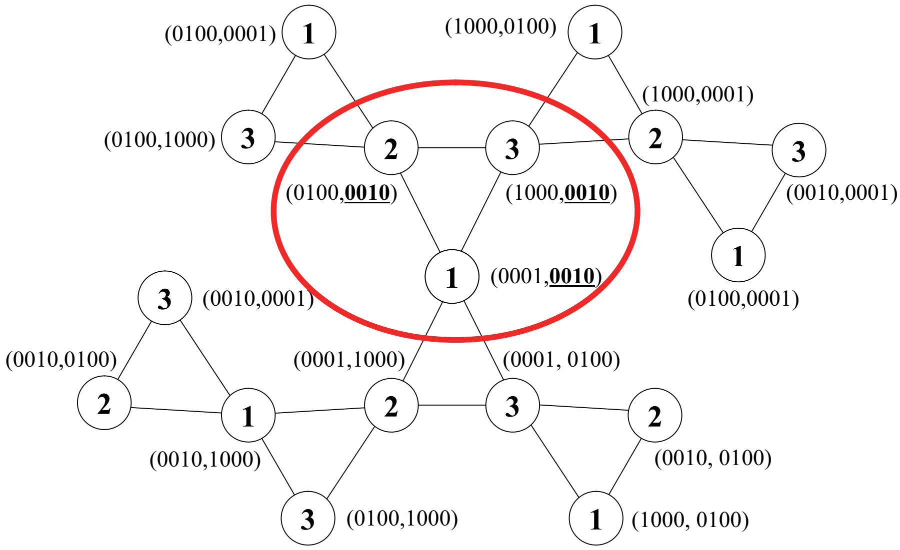

Figure 5.6: Provided codewords for coding graph of Fig. 5.5

$$\max_{j \in \mathcal{I}_{N_d}} \{H(V_j|Q_j)\} \leq R, \quad Q_{\boldsymbol{X}} = Q_j V_j,$$
$$Q_j \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{S}_j^c)}), V_j \in \mathcal{V}_n(\mathcal{X}^{(\mathcal{S}_j)}|Q_j) \quad \forall j \in \mathcal{I}_{N_d}\},$$

where $R > 0$ is a given coding rate. We note that the joint type $Q_{\boldsymbol{X}}$ specifies the types Q_j and the conditional types V_j $(j \in \mathcal{I}_{N_d})$.

2. Create a graph for each joint type $Q_{\boldsymbol{X}} \in \mathcal{T}_n(R)$ as follows: Each vertex of the graph corresponds to a sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_{Q_{\boldsymbol{X}}}$ (henceforth we denote a vertex by referring its corresponding sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})}$). An edge is placed between every pair of vertices whose subsequences $\boldsymbol{x}^{(\mathcal{S}_j)}$ are included in the set $T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})$ for some j, where $\boldsymbol{x}^{(\mathcal{S}_j^c)} \in T^n_{Q_j}$. In the following, we denote this graph the *coding graph* $G(Q_{\boldsymbol{X}})$. An example of coding graphs is shown in Fig. 5.5, where the (original) complementary delivery coding system is considered, $n = 4$, and Q is defined as follows:

$$Q(0,0) = 2/4, \quad Q(0,1) = 1/4,$$
$$Q(1,0) = 1/4, \quad Q(1,1) = 0/4.$$

The large circle in Fig. 5.5 corresponds to a sequence set $T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})$, where $\mathcal{S}_j^c = \{2\}$ and $\boldsymbol{x}^{(\mathcal{S}_j^c)} = \{0001\}$.

3. Assign a symbol to each vertex of the coding graph $G(Q_{\boldsymbol{X}})$ so that the same symbol are not assigned to any pairs of adjacent vertices. An example of codeword assignments to the coding graph in Figure 5.5 is given as Figure 5.6.

4. For an input sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_{Q_{\boldsymbol{X}}}$, if $Q_{\boldsymbol{X}} \in \mathcal{T}_n(R)$, the index assigned to the joint type $Q_{\boldsymbol{X}}$ is the first part of the codeword, and the

symbol assigned to the corresponding vertex of the coding graph is determined as the second part of the codeword. For a sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_{Q_{\boldsymbol{X}}}$ such that $Q_{\boldsymbol{X}} \notin \mathcal{T}_n(R)$, the corresponding codeword is determined arbitrarily and an encoding error is declared.

[Decoding: $\widehat{\varphi}^{(j)}_n$]

1. From the first part of the received codeword, find the coding graph $\widehat{G}(\widehat{Q}_{\boldsymbol{X}})$ of the joint type $\widehat{Q}_{\boldsymbol{X}}$. If no encoding error occurs, then $\widehat{Q}_{\boldsymbol{X}}$ should be $Q_{\boldsymbol{X}}$, and therefore the decoder $\widehat{\varphi}^{(j)}_n$ can find the coding graph $\widehat{G}(\widehat{Q}_{\boldsymbol{X}}) = G(Q_{\boldsymbol{X}})$ used in the encoding scheme.

2. For a given side information $\boldsymbol{x}^{(\mathcal{S}^c_j)}$ and the joint type $Q_{\boldsymbol{X}}$, find the vertex of the coding graph $G(Q_{\boldsymbol{X}})$, to which the second part of the received codeword is assigned. Such vertex is searched from the clique that corresponds to the set $T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})$. Note that the conditional type V_j has been determined by $\widehat{Q}_{\boldsymbol{X}} = Q_{\boldsymbol{X}}$. The sequence set $\widehat{\boldsymbol{x}}^{(\mathcal{S}_j)} \in T^n_{Q_j}$ found in this step is reproduced.

It should be noted that the above coding scheme is universal since this does not depend on a distribution $P_{\boldsymbol{X}}$ of a source $\boldsymbol{X}$.

The coding rate of the above proposed coding scheme is determined by the chromatic number of the coding graph $G(Q_{\boldsymbol{X}})$. To this end, we introduce the following lemmas.

Lemma 5.4.1. *The coding graph $G(Q)$ of the joint type $Q = Q_{\boldsymbol{X}}$ satisfies the following properties:*

1. *Every vertex $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_Q$ belongs to N_d cliques, each of which corresponds to the vertex set*

   $$T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)}). \quad (j \in \mathcal{I}_{N_d})$$

2. *The vertex $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_Q$ has no edges from vertices not included in the vertex sets $T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})$ $(j \in \mathcal{I}_{N_d})$.*

3. *From the coding graph $G(Q)$ of a given joint type $Q \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})$, both the clique number $\omega(G(Q))$ and the degree $\Delta(G(Q))$ of the coding graph $G(Q)$ are constant and obtained as follows:*

$$\begin{aligned} \omega(G(Q)) &= \max_{j \in \mathcal{I}_{N_d}} |T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})|, \\ \Delta(G(Q)) &= \sum_{j \in \mathcal{I}_{N_d}} |T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})|, \end{aligned}$$

Proof. 1) 2) Easily obtained from the first and second steps of the above encoding scheme.
3) Easily obtained from the above properties. □

Lemma 5.4.2. *The chromatic number of the coding graph $G(Q)$ of the joint type $Q \in \mathcal{T}_n(R)$ is bounded as*

$$\chi(G(Q)) \leq N_d \exp(nR).$$

Proof. This property is directly derived from Lemmas 2.2.2, 2.4.1 and 5.4.1 as follows:

$$\begin{aligned}\chi(G(Q)) &\le \Delta(G(Q)) & (5.17)\\ &= \sum_{j\in\mathcal{I}_{N_d}} |T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})| & (5.18)\\ &\le \sum_{j\in\mathcal{I}_{N_d}} \exp\{nH(V_j|Q_j)\} & (5.19)\\ &\le N_d \max_{j\in\mathcal{I}_{N_d}} \exp\{nH(V_j|Q_j)\}\\ &\le N_d\exp(nR). & (5.20)\end{aligned}$$

where Eq. (5.17) comes from Lemma 2.4.1, Eq. (5.18) from Lemma 5.4.1, Eq. (5.19) from Lemma 2.2.2, and Eq. (5.20) from the definition of $\mathcal{T}_n(R)$. This concludes the proof of Lemma 5.4.2. □

From the above discussions, we obtain

$$\omega(G(Q)) \le \chi(G(Q)) \le \Delta(G(Q)) \le N_d\exp(nR).$$

5.4.3 Coding theorems

We can obtain the following theorem for the universal lossless f-FGCD codes constructed in Section 5.4.2.

Theorem 5.4.3. *For a given real number $R>0$, there exists a universal lossless f-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}^{(j)}_n\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any integer $n\ge 1$ and any DMS $\boldsymbol{X}=\boldsymbol{X}^{(I_{N_s})}$ with a generic distribution $P_{\boldsymbol{X}}\in\mathcal{M}(\mathcal{X}^{(\mathcal{I}_{N_s})})$

$$\begin{aligned}\frac{1}{n}\log M_n &\le R+\epsilon_n(N_d), & (5.21)\\ \sum_{j=1}^{N_d} e^{(j)}_n &\le \exp\left\{-n\left(-\epsilon_n(N_d)+\min_{Q_{\boldsymbol{X}}\in\mathcal{T}^c_n(R)} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right)\right\},\end{aligned}$$

where $\epsilon_n(N)$ is defined in Eq. (5.4).

Proof. From the coding scheme, the size of the codeword set is bounded as

$$\begin{aligned}M_n &\le |\mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})|\cdot N_d\exp(nR)\\ &\le N_d(n+1)^{|\mathcal{X}^{(\mathcal{I}_{N_s})}|}\exp(nR),\end{aligned}$$

where the second inequality comes from Lemma 2.2.1. This implies Eq. (5.21). Next, we evaluate decoding error probabilities. Since every sequence set $\boldsymbol{X}^n\in T^n_{\widetilde{Q}_{\boldsymbol{X}}}$ that satisfies $\widetilde{Q}_{\boldsymbol{X}}\in\mathcal{T}_n(R)$ is reproduced correctly at the decoder, the sum of error probabilities is bounded as

$$\sum_{j=1}^{N_d} e^{(j)}_n \le N_d\Pr\left\{\boldsymbol{X}^n\in T^n_{\widetilde{Q}_{\boldsymbol{X}}} : \widetilde{Q}_{\boldsymbol{X}}\in\mathcal{T}^c_n(R)\right\}$$

$$
\begin{aligned}
&\le N_d \sum_{\widetilde{Q}_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} \exp\{-nD(\widetilde{Q}_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\} && (5.22)\\
&\le N_d \sum_{\widetilde{Q}_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} \exp\left\{-n \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right\} \\
&\le N_d (n+1)^{|\mathcal{X}^{(\mathcal{I}_{N_s})}|} \exp\left\{-n \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right\} && (5.23)\\
&= \exp\left\{-n\left(-\epsilon_n(N_d) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\},
\end{aligned}
$$

where Eq. (5.22) comes from Lemma 2.2.3, and Eq. (5.23) from Lemma 2.2.1. This completes the proof of Theorem 5.4.3. □

We can see that for any real value $R \ge R_F(\boldsymbol{X}|\boldsymbol{\mathcal{S}})$ we have

$$
\min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}}) \quad > \quad 0.
$$

This implies that any real value $R \ge R_F(\boldsymbol{X}|\boldsymbol{\mathcal{S}})$ is *a universal lossless f-FGCD achievable rate* for the system $\boldsymbol{\mathcal{S}}$, namely, there exists a universal lossless f-FGCD code for the system $\boldsymbol{\mathcal{S}}$ that satisfies the conditions shown in Definition 5.4.2.

The following converse theorem indicates that the error exponent obtained in Theorem 5.4.3 is tight.

Theorem 5.4.4. *Any lossless f-FGCD code*

$$
\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}
$$

for the system $\boldsymbol{\mathcal{S}}$ and the DMS $\boldsymbol{X}$ must satisfy

$$
\sum_{j=1}^{N_d} e_n^{(j)} \quad \ge \quad \exp\left\{-n\left(\epsilon_n(2) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\}
$$

for any integer $n \ge 1$ and a given coding rate $R = 1/n \log M_n > 0$, where $\epsilon_n(N)$ is defined in Eq. (5.4).

Proof. Note that the number of sequences to be decoded correctly for each decoder are at most $\exp(nR)$. Here, let us consider a joint type $Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R + \epsilon_n(2))$. The definition of $\mathcal{T}_n^c(R+\epsilon_n(2))$ and Lemma 2.2.2 imply that for $(\boldsymbol{x}^{(\mathcal{I}_{N_s})}) \in T^n_{Q_{\boldsymbol{X}}}$ we have

$$
\begin{aligned}
\max_{j \in \mathcal{I}_{N_d}} \{|T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})|\} &\ge (n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|} \max_{j \in \mathcal{I}_{N_d}} \exp\{nH(V_j|Q_j)\} && (5.24)\\
&\ge (n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|} \exp\{n(R + \epsilon_n(2))\} && (5.25)\\
&= 2\exp(nR),
\end{aligned}
$$

where Eq. (5.24) comes from Lemma 2.2.2, and Eq. (5.25) from the definition of $\mathcal{T}_n^c(R + \epsilon_n(2))$. Therefore, at least half of the sequence set in the V-shell

$T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}^c_j)})$ will not be decoded correctly at the decoder $\widehat{\varphi}^{(j)}_n$. Thus, the sum of the error probabilities is bounded as

$$\begin{aligned}
\sum_{j\in\mathcal{I}_{N_d}} e^{(j)}_n &\geq \frac{1}{2}\sum_{Q_{\boldsymbol{X}}\in\mathcal{T}^c_n(R+\epsilon_n(2))}\Pr\{\boldsymbol{X}^n\in T^n_{Q_{\boldsymbol{X}}}\} \\
&\geq \frac{1}{2}(n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|}\sum_{Q_{\boldsymbol{X}}\in\mathcal{T}^c_n(R+\epsilon_n(2))}\exp\{-nD(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\} \quad (5.26)\\
&\geq \frac{1}{2}(n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|}\exp\left\{-n\min_{Q_{\boldsymbol{X}}\in\mathcal{T}^c_n(R+\epsilon_n(2))}D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right\} \\
&= \exp\left\{-n\left(\epsilon_n(2)+\min_{Q_{\boldsymbol{X}}\in\mathcal{T}^c_n(R+\epsilon_n(2))}D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right)\right\},
\end{aligned}$$

where Eq. (5.26) comes from Lemma 2.2.3. This concludes the proof of Theorem 5.4.4. □

The following corollary is directly derived from Theorems 5.4.3 and 5.4.4. This shows the asymptotic optimality of the proposed coding scheme.

Corollary 5.4.1. *For a given real number $R > 0$, there exists a universal lossless f-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}^{(j)}_n\}^{N_d}_{j=1})\}^{\infty}_{n=1}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any DMS $\boldsymbol{X}$

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n \leq R,$$

$$\lim_{n\to\infty} -\frac{1}{n}\log\sum_{j\in\mathcal{I}_{N_d}} e^{(j)}_n = \min_{Q_{\boldsymbol{X}}\in\mathcal{T}^c(R)} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}}),$$

where

$$\begin{aligned}
\mathcal{T}(R) = \{&Q_{\boldsymbol{X}}\in\mathcal{M}(\mathcal{X}^{(\mathcal{I}_{N_s})}): \\
&\max_{j\in\mathcal{I}_{N_d}} H(V_j|Q_j)\leq R, \quad Q_{\boldsymbol{X}} = Q_jV_j, \\
&Q_j\in\mathcal{P}_n(\mathcal{X}^{(\mathcal{S}^c_j)}), V_j\in\mathcal{V}(\mathcal{X}^{(\mathcal{S}_j)}|Q_j), \forall j\in\mathcal{I}_{N_d}\}.
\end{aligned}$$

In a similar manner, we can investigate a probability such that the original sequence set is correctly reproduced. The following theorem shows the lower bound of the probability of correct decoding that can be attained by the proposed coding scheme.

Theorem 5.4.5. *For a given real number $R > 0$, there exists a universal lossless f-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}^{(j)}_n\}^{N_d}_{j=1})\}^{\infty}_{n=1}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any integer $n \geq 1$ and any DMS $\boldsymbol{X}$

$$\frac{1}{n}\log M_n \leq R+\epsilon_n(N_d), \quad (5.27)$$

$$1 - \sum_{j=1}^{N_d} e_n^{(j)} \quad \geq \quad \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\},$$

where $\epsilon_n(N)$ *is defined in Eq. (5.4).*

Proof. Eq. (5.27) is derived in the same way as the proof of Theorem 5.4.3. Next, we evaluate the probability such that the original sequence set is correctly reproduced. Since every sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_{Q_{\boldsymbol{X}}}$ that satisfies $Q_{\boldsymbol{X}} \in \mathcal{T}_n(R)$ is reproduced correctly at the decoder, the sum of the probabilities is bounded as

$$\begin{aligned}
1 - \sum_{j=1}^{N_d} e_n^{(j)} \quad &\geq \quad \Pr\left\{\boldsymbol{X}^n \in T^n_{Q_{\boldsymbol{X}}} : Q_{\boldsymbol{X}} \in \mathcal{T}_n(R)\right\} \\
&\geq \quad \sum_{Q_{\boldsymbol{X}} \in \mathcal{T}^n(R)} (n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|} \exp\{-nD(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\} \qquad (5.28) \\
&\geq \quad (n+1)^{-|\mathcal{X}^{(\mathcal{I}_{N_s})}|} \exp\left\{-n \min_{Q_{\boldsymbol{X}} \in \mathcal{T}^n(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right\} \\
&= \quad \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\},
\end{aligned}$$

where Eq. (5.28) comes from Lemma 2.2.3. This completes the proof of Theorem 5.4.5. □

The following converse theorem indicates that the error exponent obtained in Theorem 5.4.5 might not be tight.

Theorem 5.4.6. *Any lossless f-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *and the DMS* $\boldsymbol{X}$ *must satisfy*

$$1 - \sum_{j=1}^{N_d} e_n^{(j)} \leq \exp\Big[-n\Big\{-\epsilon_n(1) + \min_{Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})} \left(\left|\max_{j \in \mathcal{I}_{N_d}} H(V_j|Q_j) - (R + \epsilon_n(1))\right|^+ + D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\Big\}\Big]$$

for any integer $n \geq 1$ *and a given coding rate* $R = 1/n \log M_n > 0$*, where*

$$\left.\begin{array}{l} Q_{\boldsymbol{X}} = Q_j V_j, \quad \forall j \in \mathcal{I}_{N_d} \\ Q_j \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{S}_j^c)}), \quad V_j \in \mathcal{V}(\mathcal{X}^{(\mathcal{S}_j)}|Q_j) \end{array}\right\} \qquad (5.29)$$

where $\epsilon_n(N)$ *is defined in Eq. (5.4).*

Proof. Note that the number of sequences to be decoded correctly for each decoder are at most $\exp(nR)$. Here, let us consider a joint type $Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})$ that satisfies Eq. (5.29). The ratio $r_c(Q_{\boldsymbol{X}})$ of sequences in the sequence set $T_{Q_{\boldsymbol{X}}}$ such that the sequences are correctly reproduced is at most

$$r_c(Q_{\boldsymbol{X}})$$

$$\begin{aligned}
&\le \min\left\{\min_{j\in\mathcal{I}_{N_d}}\left(\frac{\exp(nR)}{|T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})|}\right),1\right\}\\
&\le \min\Big[\exp(nR)\cdot(n+1)^{|\mathcal{X}^{(\mathcal{I}^{N_s})}|}\exp\{-n\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)\},1\Big] \quad (5.30)\\
&= \min\Big[\exp\{-n\{\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)\}-(R+\epsilon_n(1))\},1\Big]\\
&= \exp\left\{-n\left|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\right|^+\right\}
\end{aligned}$$

where Eq. (5.30) comes from Lemma 2.2.2. Therefore, the probability $P_c(Q_{\boldsymbol{X}})$ such that the original sequence pair with type $Q_{\boldsymbol{X}}$ is correctly reproduced is bounded as

$$\begin{aligned}
&P_c(Q_{\boldsymbol{X}})\\
&\quad\le r_c(Q_{\boldsymbol{X}})\Pr\{\boldsymbol{X}^n\in T^n_{Q_{\boldsymbol{X}}}\}\\
&\quad\le \exp\big\{-n\big|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\big|^+ + D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\big\}, \quad (5.31)
\end{aligned}$$

where Eq. (5.31) comes from Lemma 2.2.3. Thus, the sum of the probabilities is obtained as

$$\begin{aligned}
&1-\sum_{j=1}^{N_d}e_n^{(j)}\\
&\le \sum_{Q_{\boldsymbol{X}}\in\mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})}P_c(Q_{\boldsymbol{X}})\\
&\le \sum_{Q_{\boldsymbol{X}}\in\mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})}\exp\big\{-n\big|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\big|^+ + D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\big\}\\
&\le (n+1)^{|\mathcal{X}^{(\mathcal{I}_{N_s})}|}\exp\big\{-n\min_{Q_{\boldsymbol{X}}\in\mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})}\\
&\quad\big(\big|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\big|^+ + D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\big)\big\}. \quad (5.32)\\
&= \exp\Big[-n\Big\{-\epsilon_n(1)+\min_{Q_{\boldsymbol{X}}\in\mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})}\\
&\quad\Big(\Big|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\Big|^+ + D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\Big)\Big\}\Big]
\end{aligned}$$

where Eq. (5.32) comes from Lemma 2.2.1. This completes the proof of Theorem 5.4.6. □

We can see that for any real value $R\ge R_F(\boldsymbol{X})$ and sufficiently large n we have

$$\begin{aligned}
&\min_{Q_{\boldsymbol{X}}\in\mathcal{M}(\mathcal{X}^{\mathcal{I}_{N_s}})}\big(\big|\max_{j\in\mathcal{I}_{N_d}}H(V_j|Q_j)-(R+\epsilon_n(1))\big|^+ + D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\big)\big\}\\
&= \left|\max_{j\in\mathcal{I}_{N_d}}H\left(P_{\boldsymbol{X}^{\mathcal{S}_j}|\boldsymbol{X}^{\mathcal{S}_j^c}}|P_{\boldsymbol{X}^{\mathcal{S}_j^c}}\right)-(R+\epsilon_n(1))\right|^+
\end{aligned}$$

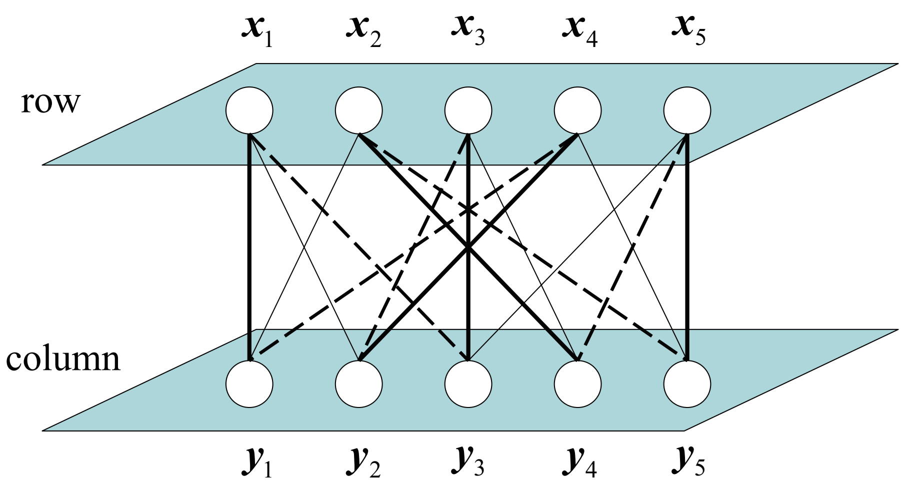

Figure 5.7: Bipartite graph equivalent to the coding graph shown in Fig. 5.6

$$= \quad 0.$$

On the other hand, for any real value $R < R_F(\boldsymbol{X})$ we have

$$\begin{aligned}
&\min_{Q_{\boldsymbol{X}} \in \mathcal{T}(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}}) \\
&\geq \min_{Q_{\boldsymbol{X}} \in \mathcal{M}(\mathcal{X}^{(\mathcal{I}_{N_d})})} \left(\left| \max_{j \in \mathcal{I}_{N_d}} H(V_j|Q_j) - (R + \epsilon_n(1)) \right|^+ + D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}}) \right\} \\
&\geq 0.
\end{aligned}$$

This implies that the error exponent obtained in Theorem 5.4.5 might not be tight.

5.4.4 Comparison with results shown in Section 5.2

The chromatic number of a coding graph is explicitly determined when $N_d = 2$. A representative special case is the situation considered in Section 5.2. In the case of $N_d = 2$, the coding graph $G(Q)$ can translated into an equivalent bipartite graph $\widetilde{G}(Q)$ such that

- each vertex in one set corresponds to a sequence set $\boldsymbol{x}^{(\mathcal{S}_1^c)} \in T_{Q_1}^n$, and each vertex in the other set corresponds to a sequence set $\boldsymbol{x}^{(\mathcal{S}_2^c)} \in T_{Q_2}^n$.
- each edge corresponds to a sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T_Q^n$, and the edge links between two vertices, each of which corresponds to the sequence subset $\boldsymbol{x}^{(\mathcal{S}_j^c)} \in T_{Q_j}^n$ $(j = 1, 2)$ of the sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})}$.

Fig. 5.3 shows an example of such bipartite graphs, which is equivalent to the coding graph shown in Fig. 5.6. From the nature of the equivalent bipartite graph $\widetilde{G}(Q)$, we can easily obtain

$$\chi(G) \quad = \quad \chi'(\widetilde{G}).$$

Therefore, the coding rate of the proposed coding scheme is determined by the edge chromatic number of the equivalent bipartite graph $\widetilde{G}(Q)$. To this end, we introduce the following lemmas.

Lemma 5.4.3. *If the number of decoders equals $N_d = 2$, then the degree of the bipartite graph $\widetilde{G}(Q)$ equivalent to the coding graph $G(Q)$ is constant for a given joint type $Q \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})$, obtained as follows:*

$$\Delta(\widetilde{G}(Q)) \quad = \quad \max_{j=1,2} |T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})|,$$

where $\boldsymbol{x}^{(\mathcal{S}_j^c)} \in T^n_{Q_j}$. This equals the clique number $\omega(G(Q))$ of the coding graph $G(Q)$.

Proof. Note that the number of edges connected to the node $\boldsymbol{x}^{(\mathcal{S}_j^c)}$ equals $|T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})|$. □

Lemma 5.4.4. *If the number of decoders equals $N_d = 2$, then the edge chromatic number of the bipartite graph $\widetilde{G}(Q)$ equivalent to the coding graph $G(Q)$ of a joint type $Q \in \mathcal{T}_n(R)$ is bounded as*

$$\chi'(\widetilde{G}(Q)) \leq \exp(nR).$$

Proof. This property is directly derived from Lemmas 2.2.2, 2.4.3 and 5.4.4 as follows:

$$\begin{aligned}
\chi'(\widetilde{G}(Q)) &= \Delta(\widetilde{G}(Q)) && (5.33)\\
&= \max_{j=1,2} |T^n_{V_j}(\boldsymbol{x}^{(\mathcal{S}_j^c)})| && (5.34)\\
&\leq \max_{j=1,2} \exp\{nH(V_j|Q_j)\} && (5.35)\\
&\leq \exp(nR), && (5.36)
\end{aligned}$$

where Eq. (5.33) comes from Lemma 2.4.3, Eq. (5.34) from Lemma 5.4.4, Eq. (5.35) from Lemma 2.2.2, and Eq. (5.36) from the definition of $\mathcal{T}_n(R)$. This concludes the proof of Lemma 5.4.4. □

To summarize the above discussions, we obtain

$$\chi(G(Q)) = \omega(G(Q)) \leq \exp(nR).$$

From the above discussions, we can obtain the following direct theorems for universal lossless f-FGCD codes of $N_d = 2$, which cannot be derived as corollaries of the theorems shown in the previous section.

Theorem 5.4.7. *If the number of decoders equals $N_d = 2$, then for a given real number $R > 0$ there exists a universal lossless f-FGCD code*

$$\{(\varphi_n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)},)\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any integer $n \geq 1$ and any DMS $\boldsymbol{X}$ with a generic distribution $P_{\boldsymbol{X}} \in \mathcal{M}(\mathcal{X}^{(\mathcal{I}_{N_s})})$

$$\begin{aligned}
\frac{1}{n}\log M_n &\leq R + \epsilon_n(1),\\
e_n^{(1)} + e_n^{(2)} &\leq \exp\left\{-n\left(-\epsilon_n(2) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\}.
\end{aligned}$$

Theorem 5.4.8. *For a given real number $R > 0$, there exists a universal lossless f-FGCD code*

$$\{(\varphi_n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any integer $n \geq 1$ and any DMS $\boldsymbol{X}$

$$\begin{aligned} \frac{1}{n}\log M_n &\leq R + \epsilon_n(1), \\ 1 - (e_n^{(1)} + e_n^{(2)}) &\geq \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n(R)} D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\right\}, \end{aligned}$$

5.5 Variable-length coding for multiple sources

This section discusses variable-length coding for the generalized complementary delivery coding system, and shows an explicit construction of universal variable-length codes. The coding scheme is similar to that of fixed-length codes, and also utilizes the coding graphs defined in Section 5.4.2.

5.5.1 Formulation

Definition 5.5.1. (Lossless fixed-to-variable fully-informed generalized complementary delivery (lossless v-FGCD) code)
A sequence $\left\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\right\}_{n=1}^{\infty}$ of codes $(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})$ is a lossless v-FGCD code for the system (i.e. indices of side information sequences) $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_j\}_{j=1}^{N_d}$ and the source $\boldsymbol{X}$ if and only if

$$\begin{aligned} &\varphi_n : \{\mathcal{X}^{(\mathcal{I}_{N_s})}\}^n \to \mathcal{B}^* \\ &\widehat{\varphi}_n^{(j)} : \varphi_n(\{\mathcal{X}^{(\mathcal{I}_{N_s})}\}^n) \times \{\mathcal{X}^{(\mathcal{S}_j^c)}\}^n \to \{\mathcal{X}^{(\mathcal{S}_j)}\}^n, \\ &e_n^{(j)} = \Pr\left\{\boldsymbol{X}^{(\mathcal{S}_j)n} \neq \widehat{\boldsymbol{X}}^{(\mathcal{S}_j)n}\right\} = 0, \quad \forall j \in \mathcal{I}_{N_d}, \\ &S_{j_1} \neq S_{j_2} \quad \forall j_1, j_2 \in \mathcal{I}_{N_d}, \end{aligned}$$

where

$$\widehat{\boldsymbol{X}}^{(\mathcal{S}_j)n} \stackrel{\text{def.}}{=} \widehat{\varphi}_n^{(j)}(\varphi_n(\boldsymbol{X}^n), \boldsymbol{X}^{(\mathcal{S}_j^c)n}).$$

and the image of φ_n is a prefix set.

Definition 5.5.2. (Lossless f-FGCD achievable rate)
R is an lossless v-FGCD achievable rate of the source $\boldsymbol{X}$ for the system $\boldsymbol{\mathcal{S}}$ if and only if there exists a lossless v-FGCD code

$$\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ and the source $\boldsymbol{X}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} E\left[l(\varphi_n(\boldsymbol{X}^n))\right] \leq R.$$

Definition 5.5.3. (Inf lossless v-FGCD achievable rate)

$$R_v(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) = \inf\{R : R \text{ is an lossless v-FGCD achievable rate of } \boldsymbol{X} \text{ for } \boldsymbol{\mathcal{S}}\}.$$

5.5.2 Code construction

We can construct universal lossless v-FGCD codes in a similar manner to universal lossless f-FGCD codes (fixed-length codes). Note that the coding rate depends on the type of sequence set to be encoded when constructing variable-length codes, whereas the coding rate is fixed beforehand for fixed-length coding. The coding scheme is described as follows:
[Encoding]

1. Create a coding graph for each joint type $Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})$ and assign a symbol to each vertex of the coding graph $G(Q_{\boldsymbol{X}})$ in the same way as Step 2 and 3 of Section 5.4.2. Note that a coding graph is created for every joint type $Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})$.

2. For an input sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in T^n_{Q_{\boldsymbol{X}}}$, the index assigned to the joint type $Q_{\boldsymbol{X}}$ is the first part of the codeword, and the symbol assigned to the corresponding vertex of the coding graph is determined as the second part of the codeword. Note that a codeword is assigned to every input sequence set $\boldsymbol{x}^{(\mathcal{I}_{N_s})} \in \mathcal{X}^{(\mathcal{I}_{N_s})n}$.

[Decoding]
Decoding can be accomplished in almost the same way as the fixed-length coding. Note that the decoder can always find the coding table used in the encoding scheme, and therefore it can always reconstruct the original sequence.

5.5.3 Coding theorems

We begin by showing a theorem for (non-universal) variable-length coding, which indicates that the inf coding rate of variable-length coding is the same as that of fixed-length coding. The proof is almost the same as Theorem 5.3.1.

Theorem 5.5.1. (Coding theorem of lossless v-FGCD code)

$$\begin{aligned} R_v(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) &= R_F(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \\ &= \max_{j \in \mathcal{I}_{N_d}} H(\boldsymbol{X}^{(\mathcal{S}_j)}|\boldsymbol{X}^{(\mathcal{I}_{N_s}-\mathcal{S}_j)}) \end{aligned}$$

The following direct theorem for universal coding indicates that the coding scheme presented in Section 5.5.2 can achieve the inf achievable rate.

Theorem 5.5.2. *There exists a universal lossless v-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ such that for any integer $n \geq 1$ and any DMS $\boldsymbol{X}$, the overflow probability $\overline{\rho}_n(R)$ is bounded as

$$\begin{aligned} \overline{\rho}_n(R) &\overset{\text{def.}}{=} \Pr\{l(\varphi_n(\boldsymbol{X}^n)) > nR\} \\ &\leq \exp\left\{-n\left(-\epsilon_n(N_d) + \min_{Q_{\boldsymbol{X}} \in \mathcal{T}_n^c(R-\epsilon_n(N_d))} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right)\right\}, \end{aligned}$$

where $\epsilon_n(N)$ is defined in Eq. (5.4). This implies that there exists a universal lossless v-FGCD code

$$\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *that satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} l(\varphi_n(\boldsymbol{X}^n)) \quad \le \quad R_v(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \quad a.s. \tag{5.37}$$

The converse theorem for variable-length coding can be easily obtained in the same way as Theorem 5.4.4.

Theorem 5.5.3. *Any lossless v-FGCD code*

$$\{(\varphi_n, \{\widehat{\varphi}_n^{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *and the DMS* $\boldsymbol{X}$ *must satisfy*

$$\overline{\rho}_n(R) \quad \ge \quad \exp\left\{-n\left(\epsilon_n(2) + \min_{Q_{\boldsymbol{X}}\in\mathcal{T}_n^c(R+\epsilon_n(2))} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right)\right\}$$

for a given real number $R > 0$ *and any integer* $n \ge 1$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

The following corollary is directly derived from Theorems 5.5.2 and 5.5.3.

Corollary 5.5.1. *There exists a universal lossless v-FGCD code*

$$\{(\varphi^n, \{\widehat{\varphi}_{(j)}^n\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *such that for any DMS* $\boldsymbol{X}$

$$\limsup_{n\to\infty} \frac{1}{n} l(\varphi^n(\boldsymbol{X}^n)) \quad \le \quad R_v(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \quad a.s.$$
$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\rho}_n(R) \quad = \quad \min_{Q_{\boldsymbol{X}}\in\mathcal{T}^c(R)} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})$$

Next, we investigate the underflow probability. For this purpose, we present the following two theorems. The proofs are almost the same as those of Theorems 5.4.5 and 5.4.6.

Theorem 5.5.4. *There exists a universal lossless f-FGCD code*

$$\{(\varphi^n, \{\widehat{\varphi}_{(j)}^n\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *such that for any integer* $n \ge 1$ *and any DMS* $\boldsymbol{X}$*, the underflow probability* $\underline{\rho}_n(R)$ *is bounded as*

$$\begin{aligned}\underline{\rho}_n(R) &\stackrel{\text{def.}}{=} \Pr\{l(\varphi^n(\boldsymbol{X}^n)) < nR\} \\ &\le \exp\left\{-n\left(\epsilon_n(1) + \min_{Q_{\boldsymbol{X}}\in\mathcal{T}_n^c(R-\epsilon_n(N_d))} D(Q_{\boldsymbol{X}}\|P_{\boldsymbol{X}})\right)\right\},\end{aligned}$$

where $\epsilon_n(N)$ *is defined in Eq. (5.4). This implies that there exists a universal lossless v-FGCD code*

$$\{(\varphi^n, \{\widehat{\varphi}_{(j)}^n\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *that satisfies*

$$\liminf_{n\to\infty} \frac{1}{n} l(\varphi^n(\boldsymbol{X}^n)) \quad \ge \quad R_v(\boldsymbol{X}|\boldsymbol{\mathcal{S}}) \quad a.s.$$

Theorem 5.5.5. *Any lossless f-FGCD code*

$$\{(\varphi^n, \{\widehat{\varphi}^n_{(j)}\}_{j=1}^{N_d})\}_{n=1}^{\infty}$$

for the system $\boldsymbol{\mathcal{S}}$ *and the DMS* $\boldsymbol{X}$ *must satisfy*

$$\underline{\rho}_n(R) \le \exp\Big[-n\Big\{-\epsilon_n(1) + \min_{Q_{\boldsymbol{X}} \in \mathcal{P}_n(\mathcal{X}^{(\mathcal{I}_{N_s})})} \left(\left|\max_{j \in \mathcal{I}_{N_d}} H(V_j|Q_j) - (R + \epsilon_n(1))\right|^+ + D(Q_{\boldsymbol{X}} \| P_{\boldsymbol{X}})\right)\Big\}\Big]$$

for a given real number $R > 0$ *and any integer* $n \ge 1$*, where* $\epsilon_n(N)$ *is defined in Eq. (5.4).*

Chapter 6

Conclusions

This thesis investigated multiterminal source coding for a general framework that involve the SW system, called coding systems with cooperation. Especially, two types of coding systems that incorporate cooperation with encoders were focused on.

First, a coding system that extends the SW system is presented, where there are some mutual linkages between two encoders. Coding problems for this coding system called the SW system with linkage (SWL system) were considered, where the coding rate of the mutual linkage between two encoders is negligible. It is shown that this mutual linkage is enough to effectively reduce the coding rate in several cases even though the rate of the mutual linkage is zero.

Next, a coding system that contrasts with the SW system in the sense of cooperation, called the complementary delivery coding system. In this coding system, the encoder delivers messages emitted from two correlated sources to decoders, and each decoder has access to one of two messages to reproduce the other message. Lossy coding problems and universal coding problems were investigated for this coding system. First, lossy coding problems for the complementary delivery coding system were considered, and the minimum achievable rate for given distortion criteria was clarified. Next, universal coding schemes were proposed for the complementary delivery coding system and its extensions called generalized complementary delivery coding systems. Explicit constructions of universal lossless codes and the bounds of the error probabilities are clarified via methods of types and graph-theoretical analyses.

The above results clarified several parts of fundamental limits for multiterminal source coding systems with cooperation. However, a lot of open problems remain in a general framework of coding systems with cooperation.

First, the achievable rate region has not been clarified for some communication systems in the presence of *cascading* and/or *feedback* information channels from a decoder so as to refine reproduced messages, which correspond to the lower half part of Figure 1.3. This framework provides another type of refinement structures which is different from the successive refinement coding system [43, 19], and is known as an information-theoretical model of information retrieval with index structures [41]. Kimura and Uyematsu [42] first considered this framework, and clarified inner and outer bounds of the achievable rate regions for given distortion criteria in several special cases. However, the achievable rate regions have not fully settled yet.

Two important problems still remain in universal coding for the complementary delivery coding system: First, the coding scheme proposed in Chapter 5 must be off from practical use due to the difficulty in finding codewords from the coding table (the coding graph) and a huge amount of storage space needed for storing the coding table (the coding graph). Practical coding schemes with less computations should be addressed. Recently, Kuzuoka et al. [44] proposed a simple coding scheme for the complementary delivery coding system with 2 sources which universally attains the minimum achievable rate. However, universal coding schemes for generalized complementary delivery coding systems still remains as an open problem. Second, this paper dealt with only lossless coding. Recent results by Kuzuoka et al. [44] also presented a simple lossy coding scheme for the complementary delivery coding system, however, this coding scheme does not attain the minimum achievable rate for a given distortion criterion. Therefore constructing universal lossy codes for the complementary delivery coding system that attain the optimal coding rate still remains as an open problem.

Acknowledgement

I would like to express my sincere gratitude for all the supports and encouragements I have received during my student days.

I am grateful to Prof. Tomohiko Uyematsu of Tokyo Institute of Technology for his constructive discussions and valuable comments, which leds to great improvements of this work. I would like to thank Prof. Te Sun Han of the University of Electo-Communications, Prof. Yoshinori Sakai, Prof. Koichi Sakaniwa, Prof. Isao Yamada and Prof. Ryutaroh Matsumoto of Tokyo Institute of Technology for their helpful comments on this thesis. Also, I wish to thank professors and researchers who made useful comments on this work: most notably, Prof. Fumio Kanaya, Prof. Yoshinao Shiraki of Shonan Institute of Technology, Prof. Kingo Kobayashi, Prof. Hiroshi Nagaoka, Prof. Hiroyoshi Morita, Prof. Mamoru Hoshi of the University of Electro-Communications, Prof. Hirosuke Yamamoto of the University of Tokyo, Prof. Raymond Yeung of Chinese University of Hong Kong, Prof. Yasutada Oohama of Tokushima University, Prof. Hiroki Koga of University of Tsukuba, Dr. Jun Muramatsu of NTT Communication Science Laboratoties, NTT Corporation, Prof. Shigeaki Kuzuoka of Wakayama University, and Mr. Shun Watanabe of Tokyo Institute of Technology. I thank to colleagues, friends and family for their encouragements: Mr. Tomomichi Kimura, Ms. Ritsuko Kimura and Ms. Michiko Kimura, all members of Chair of Communication Networks (Uyematsu-Matsumoto Laboratory) of Tokyo Institute of Technology, Dr. Shoji Makino, Dr. Junji Yamato, Dr. Kunio Kashino of NTT Communication Science Laboratories, NTT Corporation, and all members of Media Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories, NTT Corporation.

Lastly, I would like to give special thanks to my dear wife, Ms. Junko Kimura, and my precious child Mr. Masaki Kimura for their untiring dedication, supports and encouragements.

List of publications

(Irrelevant publications are excluded from the list. Please see the following URL for the excluded publications if necessary.)

http://www.brl.ntt.co.jp/people/akisato/publications.html (in English)
http://www.brl.ntt.co.jp/people/akisato/publications-j.html (in Japanese)

Journal papers

1. Akisato Kimura and Tomohiko Uyematsu, "Weak variable-length Slepian-Wolf coding with linked encoders for mixed sources," IEEE Transactions on Information Theory, Vol.50, No.1, pp.183–193, January 2004.
   (The material in this paper is included in Chapter 3.)

2. Akisato Kimura and Tomohiko Uyematsu, "Multiterminal source coding with complementary delivery," submitted, IEICE Transactions on Fundamentals.
   (The material in this paper is included in Chapter 4.)

3. Akisato Kimura, Tomohiko Uyematsu and Shigeaki Kuzuoka, "Universal coding for correlated sources with complementary delivery," to appear, IEICE Transactions on Fundamentals, September 2007.
   (The material in this paper is included in Chapter 5.)

4. Akisato Kimura, Tomohiko Uyematsu and Shigeaki Kuzuoka, "Universal coding for correlated sources with generalized complementary delivery," to be submitted, IEEE Transactions on Information Theory.
   (The material in this paper is included in Chapter 5.)

Peer-reviewed conference papers

1. Akisato Kimura and Tomohiko Uyematsu, "Weak variable-lenth Slepian-Wolf coding with linked encoders for mixed source," Proceedings of IEEE Information Theory Workshop (ITW2001), pp.82–84, Cairns, Australia, September 2001.
   (Parts of the material in this paper is included in Chapter 3.)

2. Akisato Kimura and Tomohiko Uyematsu, "Multiterminal source coding with complementary delivery," Proceedings of International Symposium

on Information Theory and its Applications (ISITA2006), pp.189–194, Seoul, South Korea, Octover 2006.
(Parts of the material in this paper is included in Chapter 4.)

3. Akisato Kimura, Tomohiko Uyematsu and Shigeaki Kuzuoka, "Universal coding for correlated sources with complementary delivery," Proceedings of International Symposium on Information Theory (ISIT2007), pp. 1756–1760, Nice, France, June 2007.
(Parts of the material in this paper is included in Chapter 5.)

Miscellaneous

1. Akisato Kimura and Tomohiko Uyematsu, "Weak variable-length Slepian-Wolf coding with linked encoders for mixed sources," IEICE Technical Report, IT99-59, pp.7-12, Okayama, Japan, January 2000.
(Parts of the material in this paper is included in Chapter 3.)

2. Akisato Kimura, "Weak variable-length Slepian-Wolf coding with linked encoders for mixed sources," Master thesis, Tokyo Institute of Technology, March 2000.
(Parts of the material in this paper is included in Chapter 3.)

3. Akisato Kimura and Tomohiko Uyematsu, "Multiterminal source coding with complementary delivering," Presented at 2006 Hawaii, IEICE and SITA Joint Conference on Information Theory (HISC2006), IEICE Technical Report, IT2006-8, pp.7-12, Nara, Japan, May 2006.
(Parts of the material in this paper is included in Chapter 4.)

4. Akisato Kimura, Tomohiko Uyematsu and Shigeaki Kuzuoka, "Universal source coding for complementary delivery," Proceedings of Symposium on Information Theory and its Applications (SITA2006), pp.803–806, Hokkaido, Japan, December 2006.
(Parts of the material in this paper is included in Chapter 5.)

5. Akisato Kimura, Tomohiko Uyematsu and Shigeaki Kuzuoka, "Universal coding for correlated sources with generalized complementary delivery," presented at a recent result session, International Symposium on Information Theory (ISIT2007), Nice, France, June 2007.
(Parts of the material in this paper is included in Chapter 5.)

Bibliography

[1] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 721–726, November 1985.

[2] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permitations," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 430–443, May 1982.

[3] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, November 1975.

[4] J. Barros and S. D. Servetto, "On the rate-distortion region for separate encoding of correlated sources," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2003.

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

[6] ——, "Multiterminal source coding," in *The information theory approach to communications*, ser. CISM Courses and Lectures, G. Longo, Ed. Springer-Verlag, 1978, vol. 229, pp. 171–231.

[7] T. Berger and R. W. Yeung, "Multiterminal source coding encoding with one distortion criterion," *IEEE Trans. Inf. Theory*, vol. 35, pp. 228–236, 1989.

[8] R. L. Brooks, "On coloring the nodes of a network," in *Proc. Cambridge Phios. Soc.*, vol. 37, 1941, pp. 194–197.

[9] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 21, pp. 226–228, March 1975.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 1991.

[11] I. Csiszár, "Linear codes for source and source networks: Error exponents, universal coding," *IEEE Trans. Inf. Theory*, vol. 28, no. 4, pp. 585–592, July 1982.

[12] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, March 1980.

[13] ——, *Information theory: Coding theorems for discrete memoryless systems*. New York: Academic Press, 1981.

[14] R. Diestel, *Graph Theory*, 3rd ed. Springer-Verlag, 2005.

[15] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source coding," *IEEE Trans. Inf. Theory*, vol. 45, no. 9, pp. 1887–1910, September 1999.

[16] ——, "Universal multiresolution source codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2113–2129, September 2001.

[17] A. A. ElGamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, November 1982.

[18] W. H. R. Equitz, "Successive refinement of information," Ph.D. dissertation, Stanford University, 1989.

[19] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–274, March 1991.

[20] T. Ericson and J. Körner, "Successive encoding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 390–395, May 1983.

[21] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, 1996.

[22] F. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012–2021, July 2002.

[23] T. Fukatani, R. Matsumoto, and T. Uyematsu, "Two methods for decreasing the computational complexity of the MIMO ML decoder," *IEICE Trans. Fundamentals*, vol. E87-A, no. 10, pp. 2571–2576, October 2004.

[24] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, November 2004.

[25] R. M. Gray, *Probability, Random Processes, Ergodic Properties*. New York: Springer-Verlag, 1988.

[26] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Systems Technical Journal*, vol. 53, no. 9, pp. 1681–1721, November 1974.

[27] W. Gu and M. Effros, "Source coding for a multiple network," in *Proc. Data Compression Conference (DCC)*, 2005, p. 461.

[28] ——, "Source coding for a simple multiple network," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2005, pp. 2335–2339.

[29] T. S. Han, "Weak variable-length source coding," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1217–1226, July 2000.

[30] ——, *Information-spectrum methods in information theory.* New York: Springer-Verlag, 2003.

[31] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 277–288, 1980.

[32] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, pp. 752–772, May 1993.

[33] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, November 1985.

[34] K. Iwata and J. Muramatsu, "An information-spectrum approach to rate-distortion function with side information," *IEICE Trans. Fundamentals*, vol. E85-A, no. 6, pp. 1387–1395, June 2002.

[35] A. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 11, pp. 2031–2034, November 1994.

[36] A. H. Kaspi, "Rate-distortion for correlated sources with partially separated encoders," Ph.D. dissertation, Cornell University, January 1979.

[37] ——, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 735–740, November 1985.

[38] A. H. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 828–840, November 1982.

[39] D. König, "Graphok és alkalmazásuk a determinánsok és a halmazok elméletére," *Mathematikai és Természettudományi Értesitö*, vol. 34, pp. 104–119, 1916, (in Hungarian).

[40] J. Körner and K. Marton, "Images of a set via two channels and their role in multi-user communication," *IEEE Trans. Inf. Theory*, vol. 23, no. 6, pp. 751–761, November 1975.

[41] A. Kimura and T. Uyematsu, "Information-theoretical analysis of index searching: revised," in *Proc. Symposium on Information Theory and its Applications (SITA)*, November 2006, pp. 73–76.

[42] ——, "Multiterminal source coding for cascading and feedback refinement systems," in *Proc. Shannon Theory Workshop (STW)*, September 2006, pp. 25–31.

[43] V. N. Koshelev, "Hierarchical coding of discrete sources," *Problems of Information Transmission*, vol. 16, no. 3, pp. 186–203, July 1980.

[44] S. Kuzuoka, A. Kimura, and T. Uyematsu, "Simple coding schemes for lossless and lossy complementary delivery problems," in *Proc. Shannon Theory Workshowp (STW)*, September 2007, to appear.

[45] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. Fundamentals*, vol. E78-A, no. 9, pp. 1063–1070, September 1995.

[46] J. Muramatsu, "Universal data compression algorithms for stationary ergodic sources based on the complexity of sequences," Ph.D. dissertation, Nagoya University, March 1998.

[47] Y. Oohama, "Universal coding for correlated sources with linked encoders," *IEEE Trans. Inf. Theory*, vol. 42, pp. 837–847, May 1996.

[48] ——, "The rate-distortion function for the quadratic gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[49] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.

[50] T. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Systems Technical Journal*, vol. 59, no. 10, pp. 84–91, December 1980.

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

[52] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 179–182, March 1977.

[53] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, July 1948.

[54] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[55] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, August 2004.

[56] E. Telatar, "capacity of multi-antenna Gaussian channels," *Europian Transactions on Telecommunications*, vol. 10, pp. 585–595, November 1999.

[57] E. Tuncel, P. Koulgi, and K. Rose, "Rate-distortion approach to databases: storage and content-based retrieval," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 953–967, June 2004.

[58] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, May 1978.

[59] T. Uyematsu, *Contemporary Shannon theory.* Tokyo, Japan: Baifu-kan, 1998, (In Japanese).

[60] ——, "An algebraic construction of codes for Slepian-Wolf source networks," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3082–3088, November 2001.

[61] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 236–246, January 2000.

[62] V. G. Vizing, "On an estimate of the chromatic class of a p-graph," *Diskret. Analiz.*, vol. 3, pp. 23–30, 1964, (in Russian).

[63] A. B. Wagner and V. Anantharam, "An improved outer bound for the multiterminal source-coding problem," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2005, pp. 1406–1410.

[64] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, 2005, submitted.

[65] ——, "Achievable rates for pattern recognition: binary and Gaussian cases," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, September 2005, pp. 28–32.

[66] F. M. J. Willems, J. K. Wolf, and A. D. Wyner, "Communicating via a processing broadcast satellite," in *Proc. of the 1989 IEEE/CAM Information Theory Workshop*, June 1989.

[67] H. S. Witsenhausen, "On source networks with minimal breakdown degradation," *Bell Systems Technical Journal*, vol. 59, no. 6, pp. 1083–1087, July 1980.

[68] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions II: A binary source," *Bell Systems Technical Journal*, vol. 60, pp. 2281–2292, November 1981.

[69] J. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Systems Technical Journal*, vol. 59, no. 8, pp. 1417–1426, October 1980.

[70] J. Wolfowitz, *Coding theorems of information theory.* Berlin-Heidelberg: Springer-Verlag, 1961.

[71] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[72] A. D. Wyner, J. K. Wolf, and F. M. J. Willems, "Communicating via a processing broadcast satellite," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1243–1249, June 2002.

[73] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[74] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 299–308, May 1981.

[75] ——, "Source coding theory for a triangular communication system," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 848–853, May 1996.

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

[77] R. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1111–1120, May 1999.

[78] Z. Zhang and T. Berger, "New results for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 502–521, July 1987.