

# An Equivalent Expression for the Wyner-Ziv Source Coding Problem\*

Tetsunao MATSUTA<sup>†a)</sup>, Member and Tomohiko UYEMATSU<sup>††b)</sup>, Fellow

**SUMMARY** We consider the coding problem for lossy source coding with side information at the decoder, which is known as the Wyner-Ziv source coding problem. The goal of the coding problem is to find the minimum rate such that the probability of exceeding a given distortion threshold is less than the desired level. We give an equivalent expression of the minimum rate by using the chromatic number and notions of covering of a set. This allows us to analyze the coding problem in terms of graph coloring and covering.

**key words:** chromatic number, covering, finite blocklength, lossy source coding, side information, zero-error coding

## 1. Introduction

Lossy source coding with side information at the decoder [2] is fundamental and important in information theory. This source coding deals with sequences generated by two correlated sources. A sequence generated by one of two sources is encoded into a codeword. Then, the sequence is reproduced allowing some distortion from the codeword and a source sequence generated by the other source that is often referred to as side information. This source coding is depicted in Fig. 1, where  $X$ ,  $Y$ , and  $\hat{X}$  denote a source sequence, side information, and the reproduced sequence, respectively.

The coding problem for the source coding is to find the minimum rate (or equivalently the minimum number of codewords) such that the probability of exceeding a given distortion threshold is less than the desired level. This coding problem is known as the Wyner-Ziv source coding problem. The limit of the minimum rate for the coding problem has already been clarified [2], [3], where the blocklength (i.e., the length of the sequence to be encoded) goes to infinity.

On the other hand, to the best of our knowledge, the minimum rate in finite blocklengths has not been fully clarified except for some special cases of the coding problem: lossy source coding problem [4], the lossless source coding problem with side information [4], and their zero-error cases [5], [6]. Only some upper bounds, lower bounds, and

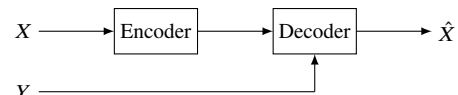


Fig. 1 Wyner-Ziv source coding.

asymptotic behavior [7]–[9] of the minimum rate have been reported so far. We note that the Wyner-Ziv source coding problem in the zero-error case has not been clarified, where the zero-error means that the probability of exceeding a given distortion threshold is zero.

In this paper, to clarify the minimum rate, we give an equivalent expression of the minimum rate. We note that an equivalent expression [4] for the lossy source coding problem is given by using  $\epsilon$ -covering which is a collection of centers of balls that cover a given set. Moreover, equivalent expressions [4], [6] for the lossless source coding problem with side information is given by using the *chromatic number* which is the minimum number of colors needed to color all vertices of a graph, where no two adjacent vertices have the same color. Thus, an equivalent expression for the Wyner-Ziv source coding problem seems to be given by combining these two notions. However, in order to give an equivalent expression, we need an additional notion of  $\epsilon$ -partition which is a collection of disjoint subsets of balls of  $\epsilon$ -covering. This implies that using  $\epsilon$ -covering and graph coloring may not be enough to construct an optimal Wyner-Ziv source coding, and noticing this is important in the practical code construction.

More precisely, we show that the minimum rate is equivalent to the chromatic number of a graph induced by  $\epsilon$ -covering and  $\epsilon$ -partition of a subset of source sequences. Due to this, every idea for the chromatic number, covering, and partition can be used to analyze the minimum rate. In other words, our equivalent expression gives a new perspective to the coding problem in the finite-blocklength regime. We note that the given equivalent expression is also valid for the zero-error case. We also show that the given equivalent expression recovers the previous results in [4]–[6]. Furthermore, for stationary memoryless sources, we show that the equivalent expression converges to the well-known single-letter characterization [2] of the minimum rate as the block length tends to infinity.

The rest of this paper is organized as follows. In Sect. 2, we provide notations and the formal definitions of the coding problem,  $\epsilon$ -covering,  $\epsilon$ -partition, and the chromatic number.

Manuscript received February 15, 2021.

Manuscript revised June 17, 2021.

Manuscript publicized September 9, 2021.

<sup>†</sup>The author is with the Graduate School of Science and Engineering, Saitama University, Saitama-shi, 338-8570 Japan.

<sup>††</sup>The author is with Dept. of Information and Communications Engineering, Tokyo Institute of Technology, Tokyo, 152-8552 Japan.

\*A part of this paper was presented at the 2020 International Symposium on Information Theory and Its Applications [1].

a) E-mail: matsuta@mail.saitama-u.ac.jp

b) E-mail: uematsu@ict.e.titech.ac.jp

DOI: 10.1587/transfun.2021TAP0005

In Sect. 3, we give our equivalent expression and some special cases involving stationary memoryless cases and show that it recovers the previous results in [4]–[6]. In Sect. 4, we show proofs for our equivalent expressions. In Sect. 5, we conclude the paper.

## 2. Preliminaries

In this section, we provide notations and the formal definitions of the Wyner-Ziv source coding problem,  $\epsilon$ -covering,  $\epsilon$ -partition, and the chromatic number.

### 2.1 Notations

For a pair of integers  $i \leq j$ , the set of integers  $\{i, i+1, \dots, j\}$  is denoted by  $[i : j]$ . For a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , the image of  $f$  is denoted by  $f(\mathcal{X}) = \{y \in \mathcal{Y} : \exists x \in \mathcal{X}, f(x) = y\}$ . The probability distribution of a random variable (RV)  $X$  is denoted by the subscript notation  $P_X$ , and the conditional probability distribution of  $X$  given  $Y$  is denoted by  $P_{X|Y}$ . For an RV  $X \in \mathcal{X}$  and a subset  $\mathcal{A} \subseteq \mathcal{X}$ , the probability  $\Pr\{X \in \mathcal{A}\}$  is denoted by  $P_X(\mathcal{A})$ . The  $n$ th Cartesian power of a set  $\mathcal{X}$  is denoted by  $\mathcal{X}^n$ , and an  $n$ -length sequence of symbols  $(a_1, a_2, \dots, a_n)$  is denoted by  $a^n$ . In what follows, all logarithms and exponentials are taken to the base 2.

### 2.2 Wyner-Ziv Source Coding Problem

In this section, we define the Wyner-Ziv source coding problem (see Fig. 1).

Let  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\hat{\mathcal{X}}$  be finite sets. Let  $X \in \mathcal{X}$  and  $Y \in \mathcal{Y}$  be RVs that represent single source symbols, where  $Y$  denotes the side information at the decoder. If  $\mathcal{X}$  is the  $n$ th Cartesian power of a finite set, we can regard the source symbol  $X$  as an  $n$ -length source sequence. Hence, for the sake of brevity, we deal with single source symbols unless otherwise stated.

The encoder  $f$  is defined as  $f : \mathcal{X} \rightarrow [1 : M]$ , where  $M$  is a positive integer that represents the number of codewords. The decoder  $\varphi$  is defined as  $\varphi : [1 : M] \times \mathcal{Y} \rightarrow \hat{\mathcal{X}}$ . In order to measure the distortion between the source symbol and the reproduction symbol in  $\hat{\mathcal{X}}$ , we introduce the distortion measure defined by a map  $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, +\infty)$ .

We define the minimum number of codewords as follows:

**Definition 1.** For  $\epsilon \geq 0$  and  $M \geq 1$ , we say  $(f, \varphi)$  is an  $(M, D, \epsilon)$ -code if and only if  $\Pr\{d(X, \varphi(f(X), Y)) > D\} \leq \epsilon$  and the number of codewords is  $M$ .

**Definition 2** (Minimum number of codewords). For  $D, \epsilon \geq 0$ ,

$$M^*(D, \epsilon) \triangleq \min\{M : \exists(M, D, \epsilon)\text{-code}\}.$$

The goal of the Wyner-Ziv source coding problem is to find  $M^*(D, \epsilon)$ .

For given  $D, \epsilon \geq 0$ , there may not exist an  $(M, D, \epsilon)$ -code for any  $M \geq 1$ , and  $M^*(D, \epsilon)$  may not be defined.

In this case, the coding problem itself becomes meaningless. Whether an  $(M, D, \epsilon)$ -code exists or not needs to be discussed individually for each distortion measure  $d$  and source. In this paper, in order to avoid such complications, we assume that, for given  $D, \epsilon \geq 0$ , an  $(M, D, \epsilon)$ -code exists<sup>†</sup> for some  $M \geq 1$ , and deal only with the case where  $M^*(D, \epsilon)$  can be defined. Note, however, that in Remark 4, we give a necessary and sufficient condition for the existence of  $(M, D, \epsilon)$ -codes.

In addition, since it obviously holds that  $M^*(D, 1) = 1$ , we do not consider this obvious case and assume that  $\epsilon \in [0, 1)$ .

### 2.3 Covering and Partition

In order to show the equivalent expression for the coding problem, we introduce  $\epsilon$ -covering,  $\epsilon$ -partition, and also  $\epsilon$ -entropy [10].

We define a ball of radius  $\epsilon$  with center at  $\hat{x} \in \hat{\mathcal{X}}$  as

$$\mathcal{B}_\epsilon(\hat{x}) \triangleq \{x \in \mathcal{X} : d(x, \hat{x}) \leq \epsilon\}.$$

For a given set  $\mathcal{A} \subseteq \mathcal{X}$ , we say a subset  $C \subseteq \hat{\mathcal{X}}$  is an  $\epsilon$ -covering of  $\mathcal{A}$  if and only if  $\mathcal{B}_\epsilon(\hat{x}) \neq \emptyset$  for all  $\hat{x} \in C$  and

$$\mathcal{A} \subseteq \bigcup_{\hat{x} \in C} \mathcal{B}_\epsilon(\hat{x}).$$

This means that the set  $\mathcal{B}_\epsilon(C) = \{\mathcal{B}_\epsilon(\hat{x}) : \hat{x} \in C\}$  of balls induced by  $C$  covers the set  $\mathcal{A}$ .

The  $\epsilon$ -entropy  $H_\epsilon(\mathcal{A})$  of a set  $\mathcal{A}$  is defined as

$$H_\epsilon(\mathcal{A}) \triangleq \log \min\{|C| : C \text{ is an } \epsilon\text{-covering of } \mathcal{A}\}.$$

We say a family  $\bar{\mathcal{B}}_\epsilon(C) = \{\bar{\mathcal{B}}_\epsilon(\hat{x}) \subseteq \mathcal{B}_\epsilon(\hat{x}) : \hat{x} \in C\}$  of subsets of balls is an  $\epsilon$ -partition of  $\mathcal{A}$  given by  $C$  if and only if  $\bar{\mathcal{B}}_\epsilon(\hat{x}) \neq \emptyset$  for all  $\hat{x} \in C$ ,  $\bar{\mathcal{B}}_\epsilon(\hat{x}) \cap \bar{\mathcal{B}}_\epsilon(\hat{x}') = \emptyset$  if  $\hat{x} \neq \hat{x}' \in C$ , and

$$\mathcal{A} \subseteq \bigcup_{\hat{x} \in C} \bar{\mathcal{B}}_\epsilon(\hat{x}).$$

This means that a family of these non-empty disjoint subsets gives a partition of the set  $\mathcal{A}$ .

As in the notation used here, a set of balls and an  $\epsilon$ -partition may be written with an  $\epsilon$ -covering  $C$ . However, for simplicity, if it is clear from the context, we omit  $C$  and, for example, write  $\bar{\mathcal{B}}_\epsilon$  instead of  $\bar{\mathcal{B}}_\epsilon(C)$ .

### 2.4 Graph Coloring and Chromatic Number

In this section, we introduce the chromatic number (cf. e.g. [11, Section 6.1]).

Since a graph  $G$  consists of the set  $V$  of vertices and the set  $E$  of edges, we denote the graph as  $G = (V, E)$ . A vertex coloring is an assignment  $c : V \rightarrow \mathbb{N}$  of integer numbers

<sup>†</sup>For example, we consider a very important case where  $\mathcal{X} = \hat{\mathcal{X}}$  and  $d(x, x) = 0 \forall x \in \mathcal{X}$ . In this case, for any  $D, \epsilon \geq 0$ , there exists an  $(M, D, \epsilon)$ -code with some  $M \geq 1$ .

to vertices such that no two adjacent vertices have the same integer number, i.e., color. Then, the minimum number of colors needed to color a given graph is called its chromatic number, and defined as

$$\chi(G) \triangleq \min\{|c(V)| : c \text{ is a vertex coloring of } G\}.$$

### 3. Main Results

In this section, we give equivalent expressions for the Wyner-Ziv source coding problem and some special cases.

For a subset  $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$ , we define

$$\mathcal{X}_{\mathcal{A}} \triangleq \{x \in \mathcal{X} : \exists y \in \mathcal{Y}, (x, y) \in \mathcal{A}\},$$

$$\mathcal{X}_{\mathcal{A}|y} \triangleq \{x \in \mathcal{X} : (x, y) \in \mathcal{A}\}.$$

$\mathcal{X}_{\mathcal{A}}$  denotes the  $\mathcal{X}$  side of  $\mathcal{A}$ , and  $\mathcal{X}_{\mathcal{A}|y}$  denotes the  $\mathcal{X}$  side of  $\mathcal{A}$  when the  $\mathcal{Y}$  side is fixed to  $y$ . We define  $\mathcal{Y}_{\mathcal{A}}$  and  $\mathcal{Y}_{\mathcal{A}|x}$  similarly.

In order to give our equivalent expression, for  $D \geq 0$  and  $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$ , we introduce  $(D, \mathcal{A})$ -covering family and  $(D, \mathcal{A})$ -partition family. For  $y \in \mathcal{Y}_{\mathcal{A}}$ , let  $C_{y, \mathcal{A}} \subseteq \hat{\mathcal{X}}$  be a  $D$ -covering of  $\mathcal{X}_{\mathcal{A}|y}$ , and  $\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(C_{y, \mathcal{A}}) = \{\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(\hat{x}) \subseteq \mathcal{B}_D(\hat{x}) : \hat{x} \in C_{y, \mathcal{A}}\}$  be a  $D$ -partition of  $\mathcal{X}_{\mathcal{A}|y}$  given by  $C_{y, \mathcal{A}}$ . Hence, it holds that

$$\mathcal{X}_{\mathcal{A}|y} \subseteq \bigcup_{\hat{x} \in C_{y, \mathcal{A}}} \bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(\hat{x}) \subseteq \bigcup_{\hat{x} \in C_{y, \mathcal{A}}} \mathcal{B}_D(\hat{x}).$$

Then, we refer to families  $\mathcal{C} = \{C_{y, \mathcal{A}} : y \in \mathcal{Y}_{\mathcal{A}}\}$  and  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(C_{y, \mathcal{A}}) : y \in \mathcal{Y}_{\mathcal{A}}\}$  as a  $(D, \mathcal{A})$ -covering family and a  $(D, \mathcal{A})$ -partition family, respectively.

For simplicity, we often omit  $\mathcal{A}$  from  $C_{y, \mathcal{A}}$ ,  $\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(\hat{x})$ , and  $\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(C_{y, \mathcal{A}})$  and write them as  $C_y$ ,  $\bar{\mathcal{B}}_D^{(y)}(\hat{x})$ , and  $\bar{\mathcal{B}}_D^{(y)}(C_y)$ , respectively. Furthermore, as mentioned in the previous section, we may omit  $C_y$  from  $\bar{\mathcal{B}}_D^{(y)}(C_y)$  and write it as  $\bar{\mathcal{B}}_D^{(y)}$ .

We note that for any  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $x \in \mathcal{X}_{\mathcal{A}|y}$ , there is the unique element  $\hat{x} \in C_y$  such that  $x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x})$ . We denote this element  $\hat{x}$  as  $\hat{x}_y(x)$ .

For a  $(D, \mathcal{A})$ -covering family  $\mathcal{C}$  and a  $(D, \mathcal{A})$ -partition family  $\bar{\mathcal{B}}$ , let  $G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})$  be a graph such that the set of vertices corresponds to  $\mathcal{X}_{\mathcal{A}}$  and two vertices  $x, x' \in \mathcal{X}_{\mathcal{A}}$  are connected by an edge if and only if the following two conditions are satisfied for some  $y \in \mathcal{Y}_{\mathcal{A}}$ ,

- (i)  $(x, y) \in \mathcal{A}$  and  $(x', y) \in \mathcal{A}$ ,
- (ii)  $\hat{x}_y(x) \neq \hat{x}_y(x')$ .

Then, we have the next theorem which gives an equivalent expression for the Wyner-Ziv source coding problem.

**Theorem 1.** For any  $D \geq 0$  and  $\epsilon \in [0, 1)$ , we have

$$\log M^*(D, \epsilon) = \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})), \quad (1)$$

where  $\mathcal{F}_{D, \epsilon}$  is triples of  $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$  satisfying  $P_{XY}(\mathcal{A}) \geq$

$1 - \epsilon$ , a  $(D, \mathcal{A})$ -covering family  $\mathcal{C} = \{C_y : y \in \mathcal{Y}_{\mathcal{A}}\}$ , and a  $(D, \mathcal{A})$ -partition family  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_D^{(y)}(C_y) : y \in \mathcal{Y}_{\mathcal{A}}\}$ .

**Remark 1.** As we will show in Remark 4, there exists  $(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$  if and only if there exists an  $(M, D, \epsilon)$ -code. Thus, the both sides of the equality (1) are well defined as long as there exists an  $(M, D, \epsilon)$ -code or  $(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ .

We note that, according to Remark 1 and the assumption that there exists an  $(M, D, \epsilon)$ -code, there also exists  $(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ .

Theorem 1 shows that the Wyner-Ziv source coding problem is equivalent to the chromatic number of a graph given by  $\epsilon$ -covering and  $\epsilon$ -partition. Thus, we can use many known results in the chromatic number such as known bounds and algorithms (cf. e.g. [11, Section 7]) to approximate the minimum number of codewords.

As we will show later, equivalent expressions for the lossy source coding problem and the lossless source coding problem with side information are given by using the chromatic number and  $\epsilon$ -covering. On the other hand, for the Wyner-Ziv source coding problem, we need to use  $\epsilon$ -partition. This is due to the fact that  $\epsilon$ -partition gives the unique index  $\hat{x}_y(x)$  of the ball including the given symbol  $x \in \mathcal{X}$ . If we only use  $\epsilon$ -covering, there are many choices of indices of balls for a given symbol  $x \in \mathcal{X}$ . Thus, the graph depends on what index is chosen for  $\hat{x}_y(x)$  and is not defined uniquely from  $\epsilon$ -covering. Since the proof of Theorem 1 largely depends on the graph  $G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})$ , we need  $\epsilon$ -partition to define the graph uniquely.

We show two special cases of the right-hand side of (1), which give minimum numbers of codewords of the lossy source coding problem [4, Theorem 1] and the lossless source coding problem with side information [4, Theorem 2].

**Theorem 2.** Let  $|\mathcal{Y}| = 1$ , i.e.,  $Y$  be constant. Then, for any  $D \geq 0$  and  $\epsilon \in [0, 1)$ , we have

$$\min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})) = \min_{\substack{\mathcal{A} \subseteq \mathcal{X}: \\ P_X(\mathcal{A}) \geq 1 - \epsilon}} H_D(\mathcal{A}).$$

**Theorem 3.** Let  $D = 0$ ,  $\hat{\mathcal{X}} = \mathcal{X}$ , and  $d(x, x') = 0$  if and only if  $x = x'$ . Then, for any  $\epsilon \in [0, 1)$ , we have

$$\begin{aligned} & \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})) \\ &= \min_{\substack{\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}: \\ P_{XY}(\mathcal{A}) \geq 1 - \epsilon}} \log \chi(G(\mathcal{A})), \end{aligned}$$

where  $G(\mathcal{A})$  denotes the graph such that the set of vertices corresponds to  $\mathcal{X}_{\mathcal{A}}$  and two vertices  $x, x' \in \mathcal{X}_{\mathcal{A}}$  are connected by an edge if and only if  $(x, y) \in \mathcal{A}$  and  $(x', y) \in \mathcal{A}$  for some  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $x \neq x'$ .

Tuncel et al. [5] give an equivalent expression for the lossy source coding problem in the zero-error case. This is given by a dominating set  $\mathcal{D}$  of a graph  $G = (V, E)$  and its

domination number  $\gamma(G)$ . The dominating set  $\mathcal{D}$  (cf. e.g. [11, Section 15.1]) is a subset of  $V$  such that every vertex of the graph either belongs to  $\mathcal{D}$  or is adjacent to some vertex in  $\mathcal{D}$ , and the domination number is the minimum cardinality of a dominating set, i.e.,

$$\gamma(G) = \min\{|\mathcal{D}| : \mathcal{D} \text{ is a dominating set of } G\}.$$

This is a similar notion of the  $\epsilon$ -entropy. In fact, we have the next lemma.

**Lemma 1.** We assume that there is an  $\epsilon$ -covering of  $\mathcal{A} \subseteq \mathcal{X}$ . Let  $\mathcal{A}^{(1)} = \mathcal{A} \times \{1\}$  and  $\hat{\mathcal{X}}^{(2)} = \hat{\mathcal{X}} \times \{2\}$ . We denote  $(x, 1) \in \mathcal{A}^{(1)}$  and  $(\hat{x}, 2) \in \hat{\mathcal{X}}^{(2)}$  as  $x^{(1)}$  and  $\hat{x}^{(2)}$ , respectively. Let  $G_\epsilon(\mathcal{A})$  be a graph<sup>†</sup> such that the set of vertices corresponds to  $\mathcal{A}^{(1)} \cup \hat{\mathcal{X}}^{(2)}$ , where any  $\hat{x}^{(2)}, \hat{x}'^{(2)} (\neq \hat{x}^{(2)}) \in \hat{\mathcal{X}}^{(2)}$  are connected and  $x^{(1)} \in \mathcal{A}^{(1)}$  and  $\hat{x}^{(2)} \in \hat{\mathcal{X}}^{(2)}$  are connected if and only if  $d(x, \hat{x}) \leq \epsilon$ . Then, we have

$$H_\epsilon(\mathcal{A}) = \log \gamma(G_\epsilon(\mathcal{A})). \quad (2)$$

We prove this lemma in Appendix A. Due to this lemma and Theorems 1 and 2, we obtain an equivalent expression for the lossy source coding problem in the zero-error case (see [5, Section 2]). This is shown in the next corollary.

**Corollary 1.** Let  $|\mathcal{Y}| = 1$ , i.e.,  $Y$  be constant, and  $\mathcal{A}' = \{x \in \mathcal{X} : P_X(x) > 0\}$ . Then, we have

$$\begin{aligned} \log M^*(D, 0) &= H_D(\mathcal{A}') \\ &= \log \gamma(G_D(\mathcal{A}')). \end{aligned} \quad (3)$$

*Proof.* For any  $\mathcal{A} \subseteq \mathcal{X}$  such that  $P_X(\mathcal{A}) = 1$ , it holds that  $\mathcal{A}' \subseteq \mathcal{A}$ , and hence any  $D$ -covering of  $\mathcal{A}$  is also a  $D$ -covering of  $\mathcal{A}'$ . Thus, we have  $H_D(\mathcal{A}) \geq H_D(\mathcal{A}')$ . This yields

$$\log M^*(D, 0) \stackrel{(a)}{=} \min_{\substack{\mathcal{A} \subseteq \mathcal{X}: \\ P_X(\mathcal{A})=1}} H_D(\mathcal{A}) \geq H_D(\mathcal{A}'),$$

where (a) comes from Theorems 1 and 2. Since the opposite direction is obvious, we have

$$\log M^*(D, 0) = H_D(\mathcal{A}').$$

We immediately obtain the corollary from this equality and Lemma 1.  $\square$

**Remark 2.** Tuncel et al. [5] only deal with the case where  $\mathcal{X} = \hat{\mathcal{X}}$  and  $d(x, x) = 0$  for any  $x \in \mathcal{X}$ . If we assume this and also  $\mathcal{X} = \mathcal{A}'$ , the equivalent expression (3) holds by using the graph  $\tilde{G}_D(\mathcal{A}')$  (instead of  $G_D(\mathcal{A}')$ ) such that the set of vertices corresponds to  $\mathcal{A}'$ , where  $x, x' (\neq x) \in \mathcal{A}'$  are connected if and only if  $d(x, x') \leq D$ . Then, the equivalent expression (3) is exactly the same as that of Tuncel et al.

On the other hand, due to Theorems 1 and 3, we obtain

<sup>†</sup>We introduce  $\mathcal{A}^{(1)}$  and  $\hat{\mathcal{X}}^{(2)}$  to make  $\mathcal{A}^{(1)} \cup \hat{\mathcal{X}}^{(2)}$  a disjoint union. If  $\mathcal{A} \cap \hat{\mathcal{X}} = \emptyset$ , we do not need to introduce these sets.

an equivalent expression for the lossless source coding problem with side information in the zero-error case [6]. This is shown in the next corollary.

**Corollary 2** ([6, Proposition 2]). Let  $\hat{\mathcal{X}} = \mathcal{X}$ ,  $d(x, x') = 0$  if and only if  $x = x'$ , and  $\mathcal{A}' = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(x, y) > 0\}$ . Then, we have

$$\log M^*(0, 0) = \log \chi(G(\mathcal{A}')).$$

*Proof.* For any  $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$  such that  $P_{XY}(\mathcal{A}) = 1$ , it holds that  $G(\mathcal{A}')$  is a subgraph of  $G(\mathcal{A})$ . Thus, we have

$$\begin{aligned} \log M^*(0, 0) &\stackrel{(a)}{=} \min_{\substack{\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}: \\ P_{XY}(\mathcal{A})=1}} \log \chi(G(\mathcal{A})) \\ &\stackrel{(b)}{\geq} \log \chi(G(\mathcal{A}')), \end{aligned}$$

where (a) comes from Theorems 1 and 3, and (b) follows since the chromatic number of a graph is greater than or equal to that of its subgraph (cf. e.g. [11, Theorem 6.1]). Since the opposite direction is obvious, this completes the proof.  $\square$

As with the above zero-error cases, due to Theorem 1, we obtain a bit simpler expression for the Wyner-Ziv source coding problem in the zero-error case. This is shown in the next corollary.

**Corollary 3.** Let  $\mathcal{A}' = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(x, y) > 0\}$ . Then, we have

$$\log M^*(D, 0) = \min_{(C, \bar{\mathcal{B}}) \in \mathcal{F}_D(\mathcal{A}')} \log \chi(G(\mathcal{A}', C, \bar{\mathcal{B}})), \quad (4)$$

where  $\mathcal{F}_D(\mathcal{A}')$  is pairs of a  $(D, \mathcal{A}')$ -covering family  $\mathcal{C} = \{C_{y, \mathcal{A}'} : y \in \mathcal{Y}_{\mathcal{A}'}\}$  and a  $(D, \mathcal{A}')$ -partition family  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_{D, \mathcal{A}'}^{(y)}(C_{y, \mathcal{A}'}) : y \in \mathcal{Y}_{\mathcal{A}'}\}$  (i.e.,  $\mathcal{F}_D(\mathcal{A}') = \{(C, \bar{\mathcal{B}}) : (\mathcal{A}', C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,0}\}$ ).

*Proof.* For  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,0}$ , let  $C' = \{C_{y, \mathcal{A}} : y \in \mathcal{Y}_{\mathcal{A}'}\}$  and  $\bar{\mathcal{B}}' = \{\bar{\mathcal{B}}_{D, \mathcal{A}}^{(y)}(C_{y, \mathcal{A}}) : y \in \mathcal{Y}_{\mathcal{A}'}\}$ . Then, we have  $(C', \bar{\mathcal{B}}') \in \mathcal{F}_D(\mathcal{A}')$  and  $G(\mathcal{A}', C', \bar{\mathcal{B}}')$  is a subgraph of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$ . Thus, we have

$$\begin{aligned} \log M^*(D, 0) &\stackrel{(a)}{=} \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,0}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \\ &\stackrel{(b)}{\geq} \min_{(C, \bar{\mathcal{B}}) \in \mathcal{F}_D(\mathcal{A}')} \log \chi(G(\mathcal{A}', C, \bar{\mathcal{B}})), \end{aligned}$$

where (a) comes from Theorem 1, and (b) follows since the chromatic number of a graph is greater than or equal to that of its subgraph. Since the opposite direction is obvious, this completes the proof.  $\square$

We regard the  $n$ th Cartesian power  $\mathcal{X}^n$  (resp.  $\hat{\mathcal{X}}^n, \mathcal{Y}^n$ ) as the alphabet  $\mathcal{X}$  (resp.  $\hat{\mathcal{X}}, \mathcal{Y}$ ). We also regard the  $n$ -length sequence  $X^n$  (resp.  $\hat{X}^n, Y^n$ ) as the symbol  $X$  (resp.  $\hat{X}, Y$ ), and the distortion measure  $d_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \rightarrow [0, +\infty)$  for each blocklength  $n$  as the distortion measure  $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, +\infty)$ .



Then, the above theorems and corollaries can be directly applied to  $n$ -length source sequences. In particular, we consider the case where sources are stationary memoryless, i.e., the pair  $(X^n, Y^n)$  is a sequence of independent copies of a pair of RVs  $(X, Y)$  over  $\mathcal{X} \times \mathcal{Y}$ . For these sources, the equivalent expression converges to the well-known single-letter characterization [2] of the minimum rate as the block length tends to infinity, which is shown in the next theorem.

**Theorem 4.** Let the distortion measure  $d_n$  be additive, i.e., for a function  $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, +\infty)$ , the distortion measure is represented as  $d_n(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i)$ . Then, for any  $D > D_{\min}$ , we have

$$\lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})) = R_{\text{wz}}(D),$$

where  $R_{\text{wz}}(D) = \min I(X; U|Y)$  of which minimum is taken over all RVs  $U \in \mathcal{U}$  and functions  $\psi : \mathcal{U} \times \mathcal{Y} \rightarrow \hat{\mathcal{X}}$  such that the Markov chain  $Y \leftrightarrow X \leftrightarrow U$  holds and  $E[d(X, \psi(U, Y))] \leq D$ , and  $D_{\min} = \inf_{\psi, U: Y \leftrightarrow X \leftrightarrow U} E[d(X, \psi(U, Y))]$ .

As shown in the following corollaries, by using this theorem, we can immediately show that two special cases in Theorems 2 and 3 converge to single-letter characterizations (cf. e.g. [12]) for the lossy source coding problem and the lossless source coding problem with side information. This is also shown in [4].

**Corollary 4** ([4, Theorem 3]). Let the distortion measure  $d_n$  be additive. Then, for any  $D > D_{\min}$ , we have

$$\lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \min_{\substack{\mathcal{A} \subseteq \mathcal{X}^n: \\ P_{X^n}(\mathcal{A}) \geq 1 - \epsilon}} H_D(\mathcal{A}) = R(D),$$

where  $R(D) = \min_{\hat{X}: E[d(X, \hat{X})] \leq D} I(X; \hat{X})$  and  $D_{\min} = \inf_{\hat{X}} E[d(X, \hat{X})]$ .

**Corollary 5** ([4, Theorem 4]). We have

$$\lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \min_{\substack{\mathcal{A} \subseteq \mathcal{X}^n \times \mathcal{Y}^n: \\ P_{X^n Y^n}(\mathcal{A}) \geq 1 - \epsilon}} \log \chi(G(\mathcal{A})) = H(X|Y).$$

#### 4. Proofs of Theorems

In this section, we prove Theorems 1–4.

##### 4.1 Proof of Theorem 1

First, we show that

$$\log M^*(D, \epsilon) \leq \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})). \quad (5)$$

Let  $(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$  and  $c$  be a vertex coloring of  $G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})$ . For  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $i \in c(\mathcal{X}_{\mathcal{A}|y})$ , we consider the set

$$\mathcal{S}_{i,y} = \{x \in \mathcal{X}_{\mathcal{A}|y} : c(x) = i\}.$$

For any  $x, x' \in \mathcal{S}_{i,y}$ , we have  $c(x) = i = c(x')$ . Thus, vertices  $x$  and  $x'$  are not adjacent in the graph  $G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})$ . Moreover, according to the definition of the graph, it holds that  $\hat{x}_y(x) = \hat{x}_y(x')$  because  $(x, y) \in \mathcal{A}$  and  $(x', y) \in \mathcal{A}$ . This implies that there is  $\hat{x}_{i,y} \in \hat{\mathcal{X}}$  such that  $\hat{x}_y(x) = \hat{x}_{i,y}$  for any  $x \in \mathcal{S}_{i,y}$ . Hence, for any  $x \in \mathcal{S}_{i,y}$ , we have

$$x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x}_y(x)) = \bar{\mathcal{B}}_D^{(y)}(\hat{x}_{i,y}). \quad (6)$$

By using the vertex coloring  $c$  and  $\hat{x}_{i,y}$ , we define an encoder and a decoder as follows:

- The encoder is defined as

$$f(x) \triangleq \begin{cases} c(x) & \text{if } x \in \mathcal{X}_{\mathcal{A}}, \\ e & \text{otherwise,} \end{cases}$$

where  $e$  is any fixed integer in  $c(\mathcal{X}_{\mathcal{A}})$ .

- The decoder is defined as

$$\varphi(i, y) \triangleq \begin{cases} \hat{x}_{i,y} & \text{if } y \in \mathcal{Y}_{\mathcal{A}}, i \in c(\mathcal{X}_{\mathcal{A}|y}), \\ \hat{e} & \text{otherwise,} \end{cases}$$

where  $\hat{e}$  is any fixed symbol in  $\hat{\mathcal{X}}$ .

For these encoder and decoder, we show that the distortion is less than  $D$  if  $(x, y) \in \mathcal{A}$ . For any  $(x, y) \in \mathcal{A}$ , we have  $x \in \mathcal{S}_{c(x),y}$  because  $y \in \mathcal{Y}_{\mathcal{A}}$ ,  $x \in \mathcal{X}_{\mathcal{A}|y}$ , and  $c(x) \in c(\mathcal{X}_{\mathcal{A}|y})$ . Hence, according to (6), we have

$$x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x}_{c(x),y}) \subseteq \mathcal{B}_D(\hat{x}_{c(x),y}). \quad (7)$$

On the other hand, we have  $f(x) = c(x)$  and  $\varphi(f(x), y) = \varphi(c(x), y) = \hat{x}_{c(x),y}$  because  $x \in \mathcal{X}_{\mathcal{A}|y} \subseteq \mathcal{X}_{\mathcal{A}}$ ,  $y \in \mathcal{Y}_{\mathcal{A}}$ , and  $c(x) \in c(\mathcal{X}_{\mathcal{A}|y})$ . Hence, we have

$$d(x, \varphi(f(x), y)) = d(x, \hat{x}_{c(x),y}) \leq D,$$

where the last inequality comes from (7). Since the distortion exceeds  $D$  only if  $(x, y) \notin \mathcal{A}$ , we have

$$\Pr\{d(X, \varphi(f(X), Y)) > D\} \leq P_{XY}(\mathcal{A}^c) \leq \epsilon, \quad (8)$$

where  $\mathcal{A}^c$  denotes the complement of the set  $\mathcal{A}$  and the last inequality comes from the definition of  $\mathcal{A}$ . This means that  $(f, \varphi)$  is a  $(|\mathcal{C}(\mathcal{X}_{\mathcal{A}})|, D, \epsilon)$ -code, and hence we have

$$\log M^*(D, \epsilon) \leq \log |\mathcal{C}(\mathcal{X}_{\mathcal{A}})|.$$

Since this holds for any  $(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$  and coloring  $c$ , we have

$$\begin{aligned} \log M^*(D, \epsilon) &\leq \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \min_c \log |\mathcal{C}(\mathcal{X}_{\mathcal{A}})| \\ &= \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})). \end{aligned}$$

Next, we show the opposite direction, i.e.,

$$\log M^*(D, \epsilon) \geq \min_{(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, \mathcal{C}, \bar{\mathcal{B}})). \quad (9)$$

For any given  $(M, D, \epsilon)$ -code, let

$$\begin{aligned} \mathcal{A} &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} : d(x, \varphi(f(x), y)) \leq D\}, \\ C_y &= \{\hat{x} \in \hat{\mathcal{X}} : \exists (x, y) \in \mathcal{A}, \hat{x} = \varphi(f(x), y)\}, \\ \bar{\mathcal{B}}_D^{(y)}(\hat{x}) &= \{x \in \mathcal{X} : \hat{x} = \varphi(f(x), y), d(x, \hat{x}) \leq D\}. \end{aligned} \quad (10)$$

(11)

Note that  $\mathcal{A} \neq \emptyset, C_y \neq \emptyset$  for any  $y \in \mathcal{Y}_{\mathcal{A}}$ , and  $\bar{\mathcal{B}}_D^{(y)}(\hat{x}) \neq \emptyset$  for any  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $\hat{x} \in C_y$ .

For any  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $\hat{x} \in C_y$ , we have  $\bar{\mathcal{B}}_D^{(y)}(\hat{x}) \subseteq \mathcal{B}_D(\hat{x})$  and  $\bar{\mathcal{B}}_D^{(y)}(\hat{x}) \cap \bar{\mathcal{B}}_D^{(y)}(\hat{x}') = \emptyset$  if  $\hat{x} \neq \hat{x}' \in C_y$ . This is because if there is  $x$  such that  $x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x})$  and  $x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x}')$ , we have  $\hat{x} = \varphi(f(x), y) = \hat{x}'$ . This contradicts the assumption that  $\hat{x} \neq \hat{x}'$ . We also have

$$\begin{aligned} \mathcal{X}_{\mathcal{A}|y} &= \{x \in \mathcal{X} : (x, y) \in \mathcal{A}\} \\ &= \{x \in \mathcal{X} : (x, y) \in \mathcal{A}, d(x, \varphi(f(x), y)) \leq D\} \\ &= \{x \in \mathcal{X} : (x, y) \in \mathcal{A}, \exists \hat{x} \in C_y, \hat{x} = \varphi(f(x), y), \\ &\quad d(x, \hat{x}) \leq D\} \\ &\subseteq \{x \in \mathcal{X} : \exists \hat{x} \in C_y, \hat{x} = \varphi(f(x), y), d(x, \hat{x}) \leq D\} \\ &= \bigcup_{\hat{x} \in C_y} \bar{\mathcal{B}}_D^{(y)}(\hat{x}) \subseteq \bigcup_{\hat{x} \in C_y} \mathcal{B}_D(\hat{x}). \end{aligned}$$

Thus,  $C_y \subseteq \hat{\mathcal{X}}$  is a  $D$ -covering of  $\mathcal{X}_{\mathcal{A}|y}$ , and  $\bar{\mathcal{B}}_D^{(y)}$  is a  $D$ -partition of  $\mathcal{X}_{\mathcal{A}|y}$  given by  $C_y$ . Furthermore, we have

$$P_{XY}(\mathcal{A}) = \Pr\{d(X, \varphi(f(X), Y)) \leq D\} \geq 1 - \epsilon.$$

Hence, we have  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ , where  $C = \{C_y : y \in \mathcal{Y}_{\mathcal{A}}\}$  and  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_D^{(y)} : y \in \mathcal{Y}_{\mathcal{A}}\}$ .

We note that for any  $x \in \mathcal{X}_{\mathcal{A}|y}$ , it holds that  $x \in \bar{\mathcal{B}}_D^{(y)}(\hat{x}_y(x))$ . Hence, by the definition of  $\bar{\mathcal{B}}_D^{(y)}$ , we have

$$\hat{x}_y(x) = \varphi(f(x), y).$$

Since adjacent vertices  $x, x' \in \mathcal{X}_{\mathcal{A}}$  ( $x \neq x'$ ) of the graph  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  satisfy conditions (i) and (ii) for some  $y \in \mathcal{Y}_{\mathcal{A}}$ , we have  $x, x' \in \mathcal{X}_{\mathcal{A}|y}$  and  $\hat{x}_y(x) \neq \hat{x}_y(x')$ . Hence, we have

$$\varphi(f(x), y) = \hat{x}_y(x) \neq \hat{x}_y(x') = \varphi(f(x'), y).$$

This means that  $f(x) \neq f(x')$ , i.e., no adjacent vertices of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  have the same codeword. In other words,  $f$  is a vertex coloring of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$ . Hence we have

$$\begin{aligned} \log M &\geq \log |f(\mathcal{X}_{\mathcal{A}})| \\ &\geq \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \\ &\geq \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})), \end{aligned}$$

where the last inequality comes from the fact that  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ . Since this inequality holds for any  $(M, D, \epsilon)$ -code, we have (9). This completes the proof.

**Remark 3.** According to the proof, the code using the vertex coloring  $c$  and  $\hat{x}_{i,y}$  can exactly attain the minimum rate for the coding problem.

**Remark 4.** We assume that there exists  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ . Then, according to the first half of the proof, there exists an  $(M, D, \epsilon)$ -code. On the other hand, we assume that there exists an  $(M, D, \epsilon)$ -code. Then, according to the second half of the proof, there exists  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ . Thus, there exists  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$  if and only if there exists an  $(M, D, \epsilon)$ -code.

## 4.2 Proof of Theorem 2

Let  $\mathcal{Y} = \{b\}$ . For any  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$ , we have  $C = \{C_b\}$  and  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_D^{(b)}\}$ . Let  $\bar{C}_b = \{\hat{x} \in C_b : \bar{\mathcal{B}}_D^{(b)}(\hat{x}) \cap \mathcal{X}_{\mathcal{A}|b} \neq \emptyset\}$ . We note that  $\bar{C}_b$  is also a  $D$ -covering of  $\mathcal{X}_{\mathcal{A}|b}$ . This is because

$$\begin{aligned} \mathcal{X}_{\mathcal{A}|b} &\subseteq \bigcup_{\hat{x} \in C_b} \bar{\mathcal{B}}_D^{(b)}(\hat{x}) \cap \mathcal{X}_{\mathcal{A}|b} \\ &= \bigcup_{\hat{x} \in \bar{C}_b} \bar{\mathcal{B}}_D^{(b)}(\hat{x}) \cap \mathcal{X}_{\mathcal{A}|b} \\ &\subseteq \bigcup_{\hat{x} \in \bar{C}_b} \bar{\mathcal{B}}_D^{(b)}(\hat{x}) \subseteq \bigcup_{\hat{x} \in \bar{C}_b} \mathcal{B}_D(\hat{x}). \end{aligned}$$

For each  $\hat{x} \in \bar{C}_b$ , we choose an  $x \in \bar{\mathcal{B}}_D^{(b)}(\hat{x}) \cap \mathcal{X}_{\mathcal{A}|b}$  and denote it as  $x(\hat{x})$ . It obviously holds that  $\hat{x}_b(x(\hat{x})) = \hat{x}$ . We consider a subset  $\mathcal{S} \subseteq \mathcal{X}_{\mathcal{A}}$  of vertices of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  such that

$$\mathcal{S} = \bigcup_{\hat{x} \in \bar{C}_b} \{x(\hat{x})\}.$$

For any  $\hat{x}, \hat{x}' \in \bar{C}_b$  such that  $\hat{x} \neq \hat{x}'$ , we have  $(x(\hat{x}), b) \in \mathcal{A}$ ,  $(x(\hat{x}'), b) \in \mathcal{A}$ , and  $\hat{x}_b(x(\hat{x})) = \hat{x} \neq \hat{x}' = \hat{x}_b(x(\hat{x}'))$ . Thus, any vertices in  $\mathcal{S} \subseteq \mathcal{X}_{\mathcal{A}}$  are connected with each other. This means that the subgraph induced by  $\mathcal{S}$  of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  is a clique. Hence, we have

$$\begin{aligned} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) &\stackrel{(a)}{\geq} \log |\mathcal{S}| \\ &\stackrel{(b)}{\geq} H_D(\mathcal{X}_{\mathcal{A}|b}) \\ &\stackrel{(c)}{\geq} \min_{\substack{\bar{\mathcal{A}} \subseteq \mathcal{X}: \\ P_X(\bar{\mathcal{A}}) \geq 1 - \epsilon}} H_D(\bar{\mathcal{A}}), \end{aligned}$$

where (a) follows since the chromatic number is greater than or equal to the number of vertices of a clique (cf. e.g. [11, Corollary 6.2]), (b) follows since  $|\mathcal{S}| = |\bar{C}_b|$  and  $\bar{C}_b$  is a  $D$ -covering of  $\mathcal{X}_{\mathcal{A}|b}$ , and (c) comes from the fact that

$$\begin{aligned} 1 - \epsilon &\leq \sum_{(x,y) \in \mathcal{A}} P_{XY}(x, y) \\ &= \sum_{x:(x,b) \in \mathcal{A}} P_{XY}(x, b) \end{aligned}$$

$$= P_X(\mathcal{X}_{\mathcal{A}|b}).$$

Since this holds for any  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$ , we have

$$\min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \geq \min_{\substack{\bar{\mathcal{A}} \subseteq \mathcal{X}: \\ P_X(\bar{\mathcal{A}}) \geq 1-\epsilon}} H_D(\bar{\mathcal{A}}). \quad (12)$$

Next, we show the inequality in the opposite direction. Let  $\bar{\mathcal{A}}^* \subseteq \mathcal{X}$  and  $D$ -covering  $C_b^*$  of  $\bar{\mathcal{A}}^*$  attain the right-hand side of (12), i.e.,

$$\min_{\substack{\bar{\mathcal{A}} \subseteq \mathcal{X}: \\ P_X(\bar{\mathcal{A}}) \geq 1-\epsilon}} H_D(\bar{\mathcal{A}}) = H_D(\bar{\mathcal{A}}^*) = \log |C_b^*|.$$

For the sake of brevity, let  $C_b^* = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{|C_b^*|}\}$ . We define  $\bar{\mathcal{B}}_D^{(b)}$  as

$$\bar{\mathcal{B}}_D^{(b)}(\hat{x}_i) \triangleq \mathcal{B}_D(\hat{x}_i) \setminus \bigcup_{k=1}^{i-1} \mathcal{B}_D(\hat{x}_k), \forall i \in \{1, \dots, |C_b^*|\}.$$

As shown in Appendix B,  $\bar{\mathcal{B}}_D^{(b)}$  is a  $D$ -partition of  $\bar{\mathcal{A}}^*$  given by  $C_b^*$ .

Let  $\mathcal{A} = \bar{\mathcal{A}}^* \times \{b\}$ . Since  $\mathcal{X}_{\mathcal{A}|b} = \bar{\mathcal{A}}^*$ , we have  $P_{XY}(\mathcal{A}) = P_X(\mathcal{X}_{\mathcal{A}|b}) = P_X(\bar{\mathcal{A}}^*) \geq 1 - \epsilon$ . Hence, it holds that  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$ , where  $C = \{C_b^*\}$  and  $\bar{\mathcal{B}} = \{\bar{\mathcal{B}}_D^{(b)}\}$ .

Since  $\bar{\mathcal{A}}^* = \mathcal{X}_{\mathcal{A}|b} = \mathcal{X}_{\mathcal{A}}$ , we can define  $c : \mathcal{X}_{\mathcal{A}} \rightarrow \mathbb{N}$  as

$$c(x) \triangleq i \text{ if } x \in \bar{\mathcal{B}}_D^{(b)}(\hat{x}_i).$$

Then, for any adjacent vertices  $x, x' \in \mathcal{X}_{\mathcal{A}} (= \mathcal{X}_{\mathcal{A}|b})$  in the graph  $G(\mathcal{A}, C, \bar{\mathcal{B}})$ , we have  $\hat{x}_{c(x)} = \hat{x}_b(x) \neq \hat{x}_b(x') = \hat{x}_{c(x')}$ . Since this means that  $c(x) \neq c(x')$ ,  $c$  is a vertex coloring of the graph. Hence, we have

$$\begin{aligned} \min_{\substack{\bar{\mathcal{A}} \subseteq \mathcal{X}: \\ P_X(\bar{\mathcal{A}}) \geq 1-\epsilon}} H_D(\bar{\mathcal{A}}) &= \log |C_b^*| \\ &\geq \log |c(\mathcal{X}_{\mathcal{A}})| \\ &\geq \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \\ &\geq \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})). \end{aligned}$$

This completes the proof.

### 4.3 Proof of Theorem 3

Since  $D = 0$ ,  $\hat{\mathcal{X}} = \mathcal{X}$ , and  $d(x, x') = 0$  if and only if  $x = x'$ , we have  $\mathcal{B}_D(x) = \{x\}$  for any  $x \in \mathcal{X}$ . Hence, for any  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$ , any  $y \in \mathcal{Y}_{\mathcal{A}}$ , and any  $x \in C_y$ , we have  $\bar{\mathcal{B}}_D^{(y)}(x) = \{x\}$ . Moreover, for any  $x \in \mathcal{X}_{\mathcal{A}|y}$ , we have  $\hat{x}_y(x) = x$ . Hence, two vertices  $x, x' \in \mathcal{X}_{\mathcal{A}}$  in the graph  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  are connected by an edge if and only if  $(x, y) \in \mathcal{A}$  and  $(x', y) \in \mathcal{A}$  for some  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $x \neq x'$ . This is because it always holds that  $\hat{x}_y(x) = x \neq x' = \hat{x}_y(x')$ . Thus, for any  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$ , we have  $G(\mathcal{A}, C, \bar{\mathcal{B}}) = G(\mathcal{A})$ .

Now, we have

$$\begin{aligned} &\min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \\ &= \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A})) \\ &\geq \min_{\substack{\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}: \\ P_{XY}(\mathcal{A}) \geq 1-\epsilon}} \log \chi(G(\mathcal{A})), \end{aligned}$$

where the last inequality follows since for any  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$ , it holds that  $P_{XY}(\mathcal{A}) \geq 1 - \epsilon$ .

On the other hand, for any  $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$  such that  $P_{XY}(\mathcal{A}) \geq 1 - \epsilon$ , let  $C' = \{C'_y : y \in \mathcal{Y}_{\mathcal{A}}\}$  and  $\bar{\mathcal{B}}' = \{\bar{\mathcal{B}}_D'^{(y)} : y \in \mathcal{Y}_{\mathcal{A}}\}$ , where  $C'_y = \mathcal{X}_{\mathcal{A}|y}$  and  $\bar{\mathcal{B}}_D'^{(y)}(x) = \{x\}$  for any  $y \in \mathcal{Y}_{\mathcal{A}}$  and  $x \in C'_y$ . Then, we have  $(\mathcal{A}, C', \bar{\mathcal{B}}') \in \mathcal{F}_{D,\epsilon}$ , and hence  $G(\mathcal{A}) = G(\mathcal{A}, C', \bar{\mathcal{B}}')$ . Thus, we have

$$\begin{aligned} &\min_{\substack{\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}: \\ P_{XY}(\mathcal{A}) \geq 1-\epsilon}} \log \chi(G(\mathcal{A})) \\ &= \min_{\substack{\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}: \\ P_{XY}(\mathcal{A}) \geq 1-\epsilon}} \log \chi(G(\mathcal{A}, C', \bar{\mathcal{B}}')) \\ &\geq \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})). \end{aligned}$$

This completes the proof.

### 4.4 Proof of Theorem 4

Let  $(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}$  and a coloring  $c$  of  $G(\mathcal{A}, C, \bar{\mathcal{B}})$  satisfy

$$\log |c(\mathcal{X}_{\mathcal{A}}^n)| = \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})). \quad (13)$$

For these sets  $(\mathcal{A}, C, \bar{\mathcal{B}})$  and the coloring  $c$ , we define a code  $(f, \varphi)$  in the same way as that in the proof of Theorem 1 in Sect. 4.1. Since  $(f, \varphi)$  is a  $(|c(\mathcal{X}_{\mathcal{A}}^n)|, D, \epsilon)$ -code (see (8)), it holds that  $\Pr\{d_n(X^n, \hat{X}^n) > D\} \leq \epsilon$ , where  $\hat{X}^n = \varphi(f(X^n), Y^n)$ . Hence, we have

$$E[d_n(X^n, \hat{X}^n)] \leq D + \epsilon d_{\max}, \quad (14)$$

where  $d_{\max} = \max_{(x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}} d(x, \hat{x})$ . By using this code and the equality (13), we have

$$\begin{aligned} &\min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}})) \\ &= \log |c(\mathcal{X}_{\mathcal{A}}^n)| \\ &= \log |f(\mathcal{X}^n)| \\ &\geq H(F) \\ &\stackrel{(a)}{\geq} nR_{\text{wz}}(E[d_n(X^n, \hat{X}^n)]) \\ &\stackrel{(b)}{\geq} nR_{\text{wz}}(D + \epsilon d_{\max}), \end{aligned}$$

where  $F = f(X^n)$ , (a) comes from the converse proof of the Wyner-Ziv theorem in [13, Section 11.3.2], and (b) comes from (14) and the fact that  $R_{\text{wz}}$  is a nonincreasing function. Since  $R_{\text{wz}}$  is a continuous function, we have

$$\lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \min_{(\mathcal{A}, C, \bar{\mathcal{B}}) \in \mathcal{F}_{D,\epsilon}} \log \chi(G(\mathcal{A}, C, \bar{\mathcal{B}}))$$

$$\geq R_{\text{wz}}(D).$$

Next, we show the inequality in the opposite direction. We define the set of typical sequences [13, Section 2.4] as

$$\mathcal{T}_\epsilon^{(n)}(X) \triangleq \{x^n \in \mathcal{X}^n : \forall x \in \mathcal{X}, |\pi(x|x^n) - P_X(x)| \leq \epsilon P_X(x)\},$$

where  $\pi(x|x^n) = |\{i : x_i = x\}|/n$ . We will omit the RV  $X$  when it is clear from the context. For  $D > D_{\min}$  and sufficiently small  $\epsilon > \epsilon' > 0$ , let an RV  $U$  and a function  $\psi$  attain the function  $R_{\text{wz}}(D/(1+\epsilon))$ . For two constants  $\tilde{R} \geq R \geq 0$ , according to the achievability proof of the Wyner-Ziv theorem in [13, Section 11.3.1], we can show the existence of a set  $\{u^n(l) \in \mathcal{U}^n : l \in [1 : 2^{\lceil n\tilde{R} \rceil}]\}$  and a code  $(f, \varphi)$  that performs the following encoding and decoding:

- Encoding: For a given  $x^n \in \mathcal{X}^n$ , the encoder finds  $l \in [1 : 2^{\lceil n\tilde{R} \rceil}]$  such that  $(u^n(l), x^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ . If there is more than one such index, it selects a specific one of them. If there is no such index, it selects a specific index from  $[1 : 2^{\lceil n\tilde{R} \rceil}]$ . The encoder outputs  $m \in [1 : 2^{\lceil nR \rceil}]$  such that  $l \in \mathcal{B}(m)$ , where  $\mathcal{B}(m) = [(m-1)2^{\lceil n\tilde{R} \rceil - \lceil nR \rceil} + 1 : m2^{\lceil n\tilde{R} \rceil - \lceil nR \rceil}]$ .
- Decoding: For given  $y^n \in \mathcal{Y}^n$  and  $m \in [1 : 2^{\lceil nR \rceil}]$ , the decoder finds the unique index  $\hat{l} \in \mathcal{B}(m)$  such that  $(u^n(\hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}$ . If there is more than one or no such index, it sets  $\hat{l} = 1$ . The decoder outputs  $\hat{x}^n$  such that

$$\hat{x}_i = \psi(u_i(\hat{l}), y_i),$$

where  $u_i(\hat{l})$  denotes the  $i$ th component of  $u^n(\hat{l})$ .

Here, we define

$$\mathcal{S}_n \triangleq \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : (u^n(\hat{l}), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\},$$

where  $\hat{l} \in \mathcal{B}(m)$  is the index selected by the decoder based on  $(x^n, y^n)$ . Then, for this code, we have (see also [13, Section 11.3.1])

$$\lim_{n \rightarrow \infty} \Pr\{(X^n, Y^n) \in \mathcal{S}_n\} = 1$$

by setting that  $\tilde{R} = I(X; U) + \delta(\epsilon') + \epsilon$  and  $R = R_{\text{wz}}(D/(1+\epsilon)) + \delta(\epsilon) + \delta(\epsilon') + 2\epsilon$ , where  $\delta(\epsilon)$  and  $\delta(\epsilon')$  are some constants that go to zero as  $\epsilon$  goes to zero.

On the other hand, for any  $(x^n, y^n) \in \mathcal{S}_n$ , we have

$$\begin{aligned} d_n(x^n, \varphi(f(x^n), y^n)) &= \frac{1}{n} \sum_{i=1}^n d(x_i, \psi(u_i(\hat{l}), y_i)) \\ &= \sum_{(u, x, y) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}} \pi(u, x, y | u^n(\hat{l}), x^n, y^n) d(x, \psi(u, y)) \\ &\stackrel{(a)}{\leq} (1+\epsilon) \mathbb{E}[d(X, \psi(U, Y))] \stackrel{(b)}{\leq} D, \end{aligned}$$

where (a) follows since  $(u^n(\hat{l}), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$  and (b) follows since  $U$  and  $\psi$  attain the function  $R_{\text{wz}}(D/(1+\epsilon))$ . Thus,

for sufficiently large  $n > 0$ , we have

$$1 - \epsilon \leq \Pr\{(X^n, Y^n) \in \mathcal{S}_n\} \leq \Pr\{(X^n, Y^n) \in \mathcal{A}\},$$

where

$$\mathcal{A} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : d(x^n, \varphi(f(x^n), y^n)) \leq D\}.$$

Now, we define  $C_{y^n}$  and  $\tilde{\mathcal{B}}_D^{(y^n)}(\hat{x}^n)$  exactly the same as (10) and (11), respectively. Then, according to the proof of Theorem 1 in Sect. 4.1,  $(\mathcal{A}, C, \tilde{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}$  and  $f$  is a vertex coloring of  $G(\mathcal{A}, C, \tilde{\mathcal{B}})$ , where  $C = \{C_{y^n} : y^n \in \mathcal{Y}_{\mathcal{A}}^n\}$  and  $\tilde{\mathcal{B}} = \{\tilde{\mathcal{B}}_D^{(y^n)} : y^n \in \mathcal{Y}_{\mathcal{A}}^n\}$ . Thus, for sufficiently large  $n > 0$ , we have

$$\begin{aligned} \frac{1}{n} \min_{(\mathcal{A}, C, \tilde{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, C, \tilde{\mathcal{B}})) &\leq \frac{1}{n} \log |f(\mathcal{X}_{\mathcal{A}}^n)| \\ &\leq R_{\text{wz}}(D/(1+\epsilon)) + \delta(\epsilon) + \delta(\epsilon') + 2\epsilon + \frac{1}{n}. \end{aligned}$$

Since  $\epsilon > 0$  can be arbitrarily small, we have

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \min_{(\mathcal{A}, C, \tilde{\mathcal{B}}) \in \mathcal{F}_{D, \epsilon}} \log \chi(G(\mathcal{A}, C, \tilde{\mathcal{B}})) &\leq R_{\text{wz}}(D). \end{aligned}$$

This completes the proof.

## 5. Conclusion

In this paper, we have considered the Wyner-Ziv source coding problem and given an equivalent expression of the minimum rate. Our equivalent expression is given by using the chromatic number,  $\epsilon$ -covering, and  $\epsilon$ -partition. As special cases of our expression, we have given equivalent expressions for the lossy source coding problem, the lossless source coding problem with side information, and zero-error cases. Furthermore, for stationary memoryless sources, we have shown that our equivalent expression converges to the single-letter characterization of the minimum rate as the block length tends to infinity.

It is an interesting future direction to give tight upper and lower bounds on the minimum rate in the finite-blocklength regime using our equivalent expression and known results in the chromatic number. Another direction is to find an equivalent expression for the coding problem when other criteria are used, such as the average distortion criterion, i.e.,  $\mathbb{E}[d(X, \varphi(f(X), Y))] \leq D$  instead of the excess distortion probability criterion, i.e.,  $\Pr\{d(X, \varphi(f(X), Y)) > D\} \leq \epsilon$ .

## Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Numbers JP19K04368.



## References

- [1] T. Matsuta and T. Uyematsu, "An equivalent expression for the Wyner-Ziv source coding problem," Proc. Int. Symp. on Inf. Theory and its Apps., pp.21–25, Oct. 2020.
- [2] A. 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, Jan. 1976.
- [3] 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.1063–1070, June 2002.
- [4] T. Matsuta and T. Uyematsu, "Equivalent representations for two source coding problems," Proc. 42th Symp. on Inf. Theory and its Apps. (SITA2018), pp.143–148, Dec. 2018.
- [5] E. Tuncel, P. Koulgi, S. Regunathan, and K. Rose, "Zero-error source coding with maximum distortion criterion," Proc. DCC 2002. Data Compression Conference, pp.92–101, April 2002.
- [6] H. Witsenhausen, "The zero-error side information problem and chromatic numbers (corresp.)," IEEE Trans. Inf. Theory, vol.22, no.5, pp.592–593, Sept. 1976.
- [7] S. Watanabe, S. Kuzuoka, and V.Y.F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," IEEE Trans. Inf. Theory, vol.61, no.4, pp.1574–1605, April 2015.
- [8] Y. Oohama, "An inequality useful for proofs of strong converse theorems in network information theory," Proc. IEEE Int. Symp. on Inf. Theory, pp.2883–2887, 2019.
- [9] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," IEEE Trans. Inf. Theory, vol.66, no.2, pp.689–703, Feb. 2020.
- [10] A.N. Kolmogorov, "On certain asymptotic characteristics of completely bounded metric spaces," Dokl. Akad. Nauk SSSR, vol.108, no.3, pp.385–388, 1956.
- [11] G. Chartrand and P. Zhang, Chromatic Graph Theory, 2nd ed., Chapman and Hall/CRC, 2019.
- [12] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.
- [13] A. El Gamal and Y.H. Kim, Network Information Theory, Cambridge University Press, 2011.
- [14] V. Kann, "On the approximability of NP-complete optimization problems," PhD dissertation, Royal Institute of Technology, 1992.

## Appendix A: Proof of Lemma 1

In this appendix, we prove Lemma 1. To this end, we use the next lemma.

**Lemma 2.** We assume that there is an  $\epsilon$ -covering of  $\mathcal{A} \subseteq \mathcal{X}$ . Let  $\mathcal{D}$  be a dominating set of  $G_\epsilon(\mathcal{A})$ . Then, for  $x_e^{(1)} \in \mathcal{A}^{(1)} \cap \mathcal{D}$  and its adjacent vertex  $\hat{x}_a^{(2)} \in \hat{\mathcal{X}}^{(2)}$ ,  $\mathcal{D}_{ea} = \mathcal{D} \setminus \{x_e^{(1)}\} \cup \{\hat{x}_a^{(2)}\}$  is still a dominating set of  $G_\epsilon(\mathcal{A})$ .

*Proof.* Let  $N[v]$  be a closed neighborhood of  $v$ , i.e.,  $N[v] = \{v\} \cup \{v' : v' \text{ is adjacent to } v\}$ . Then,  $\mathcal{D}$  is a dominating set if and only if

$$\mathcal{A}^{(1)} \cup \hat{\mathcal{X}}^{(2)} \subseteq \bigcup_{v \in \mathcal{D}} N[v]. \quad (\text{A.1})$$

We will show that  $\mathcal{D}_{ea} = \mathcal{D} \setminus \{x_e^{(1)}\} \cup \{\hat{x}_a^{(2)}\}$  satisfies (A.1).

Since there is an  $\epsilon$ -covering of  $\mathcal{A}$ , for any vertex  $x^{(1)} \in \mathcal{A}^{(1)}$ , there is some vertex  $\hat{x}^{(2)} \in \hat{\mathcal{X}}^{(2)}$  such that  $d(x, \hat{x}) \leq \epsilon$ . Hence, by the definition of  $G_\epsilon(\mathcal{A})$ , the vertex  $x_e^{(1)} \in \mathcal{A}^{(1)} \cap$

$\mathcal{D}$  must be adjacent to the vertex  $\hat{x}_a^{(2)} \in \hat{\mathcal{X}}^{(2)}$  and is not isolated. Then, we have

$$\{x_e^{(1)}\} \subseteq \{v' : v' \text{ is adjacent to } \hat{x}_a^{(2)}\}. \quad (\text{A.2})$$

Moreover, since any adjacent vertex of  $x_e^{(1)}$  is in  $\hat{\mathcal{X}}^{(2)}$  by the definition of  $G_\epsilon(\mathcal{A})$ , we have

$$\{v' : v' \text{ is adjacent to } x_e^{(1)}\} \subseteq \hat{\mathcal{X}}^{(2)}. \quad (\text{A.3})$$

Now, we have

$$\begin{aligned} N[x_e^{(1)}] &= \{x_e^{(1)}\} \cup \{v' : v' \text{ is adjacent to } x_e^{(1)}\} \\ &\stackrel{(a)}{\subseteq} \{v' : v' \text{ is adjacent to } \hat{x}_a^{(2)}\} \cup \hat{\mathcal{X}}^{(2)} \\ &\stackrel{(b)}{=} \{\hat{x}_a^{(2)}\} \cup \{v' : v' \text{ is adjacent to } \hat{x}_a^{(2)}\} \\ &= N[\hat{x}_a^{(2)}], \end{aligned}$$

where (a) comes from (A.2) and (A.3), and (b) comes from the fact that the vertex  $\hat{x}_a^{(2)}$  is adjacent to any other vertex in  $\hat{\mathcal{X}}^{(2)}$ . Due to this, we have

$$\begin{aligned} \mathcal{A}^{(1)} \cup \hat{\mathcal{X}}^{(2)} &\stackrel{(a)}{\subseteq} N[x_e^{(1)}] \cup \bigcup_{v \in \mathcal{D} \setminus \{x_e^{(1)}\}} N[v] \\ &\subseteq N[\hat{x}_a^{(2)}] \cup \bigcup_{v \in \mathcal{D} \setminus \{x_e^{(1)}\}} N[v] \\ &= \bigcup_{v \in \mathcal{D}_{ea}} N[v], \end{aligned}$$

where (a) follows since  $\mathcal{D}$  is a dominating set. This means that  $\mathcal{D}_{ea}$  is a dominating set.  $\square$

Now, we prove the lemma.

*Proof of Lemma 1.* The proof is based on the proof of [14, Theorem A.1].

Let  $C \subseteq \hat{\mathcal{X}}$  be an  $\epsilon$ -covering of  $\mathcal{A}$ . Then, for any  $x \in \mathcal{A}$ , there is  $\hat{x} \in C$  such that  $d(x, \hat{x}) \leq \epsilon$ . Hence, by letting  $C^{(2)} = C \times \{2\} (\subseteq \hat{\mathcal{X}}^{(2)})$ ,  $x^{(1)} \in \mathcal{A}^{(1)}$  is adjacent to  $\hat{x}^{(2)} \in C^{(2)}$  in the graph  $G_\epsilon(\mathcal{A})$ . Furthermore, any  $\hat{x}^{(2)} \in \hat{\mathcal{X}}^{(2)}$  belongs to  $C^{(2)}$  or is adjacent to any  $\hat{x}^{(2)} \in C^{(2)}$ . Hence,  $C^{(2)}$  is a dominating set of the graph. This gives

$$\begin{aligned} H_\epsilon(\mathcal{A}) &= \log |C^*| \\ &= \log |C^{*(2)}| \\ &\geq \log \min\{|\mathcal{D}| : \mathcal{D} \text{ is a dominating set of } G_\epsilon(\mathcal{A})\} \\ &= \log \gamma(G_\epsilon(\mathcal{A})), \end{aligned} \quad (\text{A.4})$$

where  $C^*$  is an  $\epsilon$ -covering that attains the  $\epsilon$ -entropy and  $C^{*(2)} = C^* \times \{2\}$ .

Let  $\mathcal{D}$  be a dominating set of  $G_\epsilon(\mathcal{A})$ . According to Lemma 2, the set obtained by eliminating  $x^{(1)} \in \mathcal{A}^{(1)} \cap \mathcal{D}$  from  $\mathcal{D}$  and adding an adjacent vertex  $\hat{x}^{(2)} \in \hat{\mathcal{X}}^{(2)}$  of  $x^{(1)}$  is still a dominating set. Repeating this elimination and addition until all vertex in  $\mathcal{A}^{(1)}$  are eliminated from  $\mathcal{D}$ ,

we obtain a dominating set  $\mathcal{D}_{\text{ea}}$  such that  $\mathcal{D}_{\text{ea}} \subseteq \hat{\mathcal{X}}^{(2)}$  and  $|\mathcal{D}_{\text{ea}}| \leq |\mathcal{D}|$ . From the definition of the dominating set, any  $x^{(1)} \in \mathcal{A}^{(1)}$  is adjacent to some  $\hat{x}^{(2)} \in \mathcal{D}_{\text{ea}} (\subseteq \hat{\mathcal{X}}^{(2)})$  and hence  $d(x, \hat{x}) \leq \epsilon$ . This means that  $\mathcal{D}_{\text{ea} \setminus 2} = \{\hat{x} : (\hat{x}, 2) \in \mathcal{D}_{\text{ea}}\}$  is an  $\epsilon$ -covering of  $\mathcal{A}$ . Hence, we have

$$\begin{aligned} \log \gamma(G_\epsilon(\mathcal{A})) &= \log |\mathcal{D}^*| \\ &\geq \log |\mathcal{D}_{\text{ea}}^*| \\ &= \log |\mathcal{D}_{\text{ea} \setminus 2}^*| \\ &\geq \log \min\{|\mathcal{C}| : \mathcal{C} \text{ is an } \epsilon\text{-covering of } \mathcal{A}\} \\ &= H_\epsilon(\mathcal{A}), \end{aligned} \quad (\text{A.5})$$

where  $\mathcal{D}^*$  is a dominating set that attains the domination number, and  $\mathcal{D}_{\text{ea}}^* (\subseteq \hat{\mathcal{X}}^{(2)})$  is a dominating set obtained from  $\mathcal{D}^*$  by repeating the above elimination and addition process.

Due to (A.4) and (A.5), we have (2).  $\square$

## Appendix B: Existence of Partition

Let  $\bar{\mathcal{A}}^* \subseteq \mathcal{X}$  and  $D$ -covering  $C_b^*$  of  $\bar{\mathcal{A}}^*$  attain the right-hand side of (12). We note that, for any  $\hat{x} \in C_b^*$ ,

$$\mathcal{B}_D(\hat{x}) \setminus \bigcup_{\hat{x}' \in C_b^* \setminus \{\hat{x}\}} \mathcal{B}_D(\hat{x}') \neq \emptyset. \quad (\text{A.6})$$

This comes from the following fact: If there exists  $\hat{x} \in C_b^*$  such that

$$\mathcal{B}_D(\hat{x}) \setminus \bigcup_{\hat{x}' \in C_b^* \setminus \{\hat{x}\}} \mathcal{B}_D(\hat{x}') = \emptyset,$$

we have

$$\mathcal{B}_D(\hat{x}) \subseteq \bigcup_{\hat{x}' \in C_b^* \setminus \{\hat{x}\}} \mathcal{B}_D(\hat{x}').$$

Thus,  $C_b^* \setminus \{\hat{x}\}$  is a  $D$ -covering of  $\bar{\mathcal{A}}^*$ , and this yields a contradiction that

$$H_D(\bar{\mathcal{A}}^*) \leq \log |C_b^* \setminus \{\hat{x}\}| < \log |C_b^*| = H_D(\bar{\mathcal{A}}^*).$$

For the sake of brevity, let  $C_b^* = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{|C_b^*|}\}$ . We define  $\bar{\mathcal{B}}_D^{(b)}$  as in the proof of Theorem 2, i.e.,

$$\bar{\mathcal{B}}_D^{(b)}(\hat{x}_i) \triangleq \mathcal{B}_D(\hat{x}_i) \setminus \bigcup_{k=1}^{i-1} \mathcal{B}_D(\hat{x}_k), \forall i \in \{1, \dots, |C_b^*|\}.$$

Then, for any  $i \in \{1, \dots, |C_b^*|\}$ , we have  $\bar{\mathcal{B}}_D^{(b)}(\hat{x}_i) \subseteq \mathcal{B}_D(\hat{x}_i)$  and  $\bar{\mathcal{B}}_D^{(b)}(\hat{x}_i) \neq \emptyset$  due to (A.6). Furthermore, according to this definition,  $\bar{\mathcal{B}}_D^{(b)}(\hat{x}_i)$  does not include any ball  $\mathcal{B}_D(\hat{x}_k)$  such that  $k < i$ . Thus, for any  $k, i \in \{1, \dots, |C_b^*|\}$  such that  $k < i$ , we have  $\bar{\mathcal{B}}_D^{(b)}(\hat{x}_k) \cap \bar{\mathcal{B}}_D^{(b)}(\hat{x}_i) = \emptyset$ . We also have

$$\bar{\mathcal{A}}^* \subseteq \bigcup_{i=1}^{|C_b^*|} \mathcal{B}_D(\hat{x}_i)$$

$$\begin{aligned} &\stackrel{(a)}{=} \bigcup_{i=1}^{|C_b^*|} \left( \mathcal{B}_D(\hat{x}_i) \setminus \bigcup_{k=1}^{i-1} \mathcal{B}_D(\hat{x}_k) \right) \\ &= \bigcup_{i=1}^{|C_b^*|} \bar{\mathcal{B}}_D^{(b)}(\hat{x}_i), \end{aligned}$$

where (a) comes from the fact that

$$\bigcup_{i=1}^M A_i = \bigcup_{i=1}^M \left( A_i \setminus \bigcup_{k=1}^{i-1} A_k \right).$$

Therefore,  $\bar{\mathcal{B}}_D^{(b)}$  is a  $D$ -partition of  $\bar{\mathcal{A}}^*$  given by  $C_b^*$ .



**Tetsunao Matsuta** was born in Fukui, Japan, in 1985. He received the B.E. degree from Utsunomiya University in 2007, and the M.E. and Dr.Eng. degrees from Tokyo Institute of Technology, Japan, in 2009 and 2012, respectively. From 2012 to 2021, he was an Assistant Professor with Tokyo Institute of Technology. Since 2021, he has been an Associate Professor with the Graduate School of Science and Engineering, Saitama University. He received the Best Paper Award from IEICE in 2016. His current

research interests include the areas of multi-terminal information theory and non-asymptotic analysis of coding problems.



**Tomohiko Uyematsu** received the B.E., M.E. and Dr.Eng. degrees from Tokyo Institute of Technology in 1982, 1984 and 1988, respectively. From 1984 to 1992, he was with the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology, first as research associate, next as lecturer, and lastly as associate professor. From 1992 to 1997, he was with School of Information Science of Japan Advanced Institute of Science and Technology as associate professor. Since 1997, he returned to

Tokyo Institute of Technology as associate professor, and currently he is with the Department of Information and Communications Engineering as professor. In 1992 and 1996, he was a visiting researcher at the Supélec (Ecole supérieure d'électricité), France and Delft University of Technology, Netherlands, respectively. He was an associate editor of IEEE Trans. Information Theory from 2010 to 2013. He received the Achievement Award in 2008, and the Best Paper Award in 1993, 1996, 2002, 2007, 2011, 2014 and 2016 both from IEICE. His current research interests are in the areas of information theory, especially Shannon theory and multi-terminal information theory. Dr. Uyematsu is a senior member of IEEE.