PAPER *Special Section on Information Theory and Its Applications*

Achievable Rate Regions for Source Coding with Delayed Partial Side Information*

Tetsunao MATSUTA[†a)], ***Member and*** **Tomohiko UYEMATSU**[†b)], ***Fellow***

SUMMARY In this paper, we consider a source coding with side information partially used at the decoder through a codeword. We assume that there exists a relative delay (or gap) of the correlation between the source sequence and side information. We also assume that the delay is unknown but the maximum of possible delays is known to two encoders and the decoder, where we allow the maximum of delays to change by the block length. In this source coding, we give an inner bound and an outer bound on the achievable rate region, where the achievable rate region is the set of rate pairs of encoders such that the decoding error probability vanishes as the block length tends to infinity. Furthermore, we clarify that the inner bound coincides with the outer bound when the maximum of delays for the block length converges to a constant.

key words: *achievable rate region, delay, side information, source coding*

1. Introduction

Source coding with partial (or coded) side information is one of important coding systems introduced by Wyner [2] and Ahlswede-Körner [3]. In this coding system, two encoders independently encode source sequences from two correlated sources into codewords, and the decoder reconstructs one of the source sequences from two codewords. The other source sequence is not reconstructed and used partially as side information at the decoder through the codeword. We sometimes refer to the source sequence used as side information as the side information sequence. We also refer to this coding system as the Wyner-Ahlswede-Körner (WAK) coding system for the sake of brevity. For the WAK coding system, Wyner [2] and Ahlswede-Körner [3] characterized the achievable rate region for a discrete stationary memoryless source (DMS), where the achievable rate region is the set of rate pairs of encoders such that the decoding error probability vanishes as the block length tends to infinity.

In the above WAK coding system, it is assumed that two encoders can receive correlated source symbols simultaneously. However, if the encoders are far away from each other, two encoders are not always able to receive correlated source symbols simultaneously. Especially, a situation will occur in which the side information sequence is relatively delayed to the other one. Moreover, the delay time to obtain a correlated symbol at the encoder may be unknown to the coding system. For example, we can consider the following situation previously mentioned in [4]: An observatory (encoder) on an island observes a sequence of wave heights per unit time (source sequence) caused by breeze, an earthquake, a typhoon, and etc. The observatory transmits this sequence to a weather center (decoder) on a coast city distant from there. On the other hand, a sequence of wave heights (side information sequence) can be observed also on the coast of the city and used partially at the center. However, since the wave reaches the coast city later than it reaches the island, these heights at the same time may not be correlated. Furthermore, observatories and the weather center do not know the actual delay of the wave in advance, because there are many uncertainties such as the direction of breeze, the point of the earthquake center, shielding on the sea, etc.

In this paper, we consider the WAK coding system with delayed side information mentioned above. Here, we assume that the delay is unknown but the maximum of possible delays is known to the system. In other words, the system knows the worst case delay which can be roughly setting from the distance between encoders. We allow the maximum of delays to change by the block length. This allows us more detailed analyses such as the case where delays affect half of a source sequence. For this coding system, we give an inner bound and an outer bound on the achievable rate region for a DMS. Furthermore, we clarify that the inner bound coincides with the outer bound when the maximum of delays for the block length converges to a constant. We also clarify that the region does not always coincide with that for the case without delay.

Proof techniques used in this paper are based on our previous study [5] which gives the achievable rate region for a similar coding system of the WAK coding system. In order to obtain the inner bound by using the previous technique, we need to show that there exist encoders and a decoder of which error probability vanishes at a certain desired order of the block length for a certain *mixed* source if the pair of rates is in the inner bound. In the previous study, we used Gallager's random coding technique [6], [7] to show the existence of such encoders and a decoder. Although Gallager's technique provides a detailed analysis for the error probability, it requires the knowledge of special functions and probability distributions. Hence, in order to obtain the inner bound more simply, we use a different technique, i.e., the Chernoff bound (cf. e.g. [8], [9]) in this paper. Specifi-

Manuscript received January 27, 2019.
Manuscript revised July 16, 2019.

[†]The authors are with Dept. of Information and Communications Engineering, Tokyo Institute of Technology, Tokyo, 152-8552 Japan.

[*]A part of this paper was presented at the 2018 International Symposium on Information Theory and Its Applications [1].

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

DOI: 10.1587/transfun.E102.A.1631

cally, to show the existence of encoders and a decoder, we use a known bound [10] on the error probability of encoders and a decoder for the WAK coding system. Then, we show that the bound (and hence the error probability) vanishes at the desired order by exploiting the Chernoff bound if the pair of rates of the encoders is in the inner bound. Since this analysis using the Chernoff bound does not require special functions and probability distributions typically used for Gallager's technique, we can obtain the inner bound more simply.

We note that there are some researches related to coding systems with delayed side information (the Slepian-Wolf coding system [5], [11], [12], and the Wyner-Ziv coding system [4]). Especially, Willems [11] also considered the WAK coding system with delayed side information. He assumes the following three conditions (i)–(iii) that are quite different from our setup: (i) The actual delay is known to the decoder. (ii) Encoders and the decoder continue to carry out coding infinitely on a given block length for infinitely long source sequences (though the block length is finite). (iii) When the decoder reconstructs a source sequence from the codeword, it can always utilize all codewords (actually two adjacent codewords) of side information sequences correlated with the reconstructed source sequence. Under these conditions, for a DMS, he showed a quite different result from our result: The achievable rate region always coincides with that for the case without delay. In other words, delays have no effect on coding. This mainly follows from conditions (ii) and (iii) as explained below: In a conventional manner, suppose that sequences of length n are encoded. Then, due to the delay, there is no correlation between a part of the end (or beginning) of the source sequence and the side information sequence. In particular, in the case where the delay exceeds the block length n, there is no correlation between the sequences. Hence, if we consider a single pair of two codewords, the correlation may not be sufficiently used. This makes the rate large. However, due to the conditions (ii) and (iii), there must exist codewords of side information sequences correlated with a source sequence regardless of the size of the delay, and all these codewords can be used at the decoder. Thus, if we assume (ii) and (iii), the correlation between sequences can be used perfectly. This eliminates the effect of the delay.

There are many controversial problems in the conditions of (i)–(iii). As in the previous example of islands, it is difficult to know the actual delay as in (i). In practice, it is not possible to consider infinitely long sequences as in (ii). If we do not assume such infinitely long sequences, i.e., we stick to source sequences of finite length, we cannot assume (iii) because the sequences of finite length may not be correlated due to the delay. Moreover, if the decoder does not know the delay, it is quite difficult to assume (iii) because the decoder cannot recognize which codewords are of correlated sequences. Even if the delay is known to the decoder, we should note that it must wait a long time until receiving codewords of correlated sequences if the sequences arrive late at the encoders as in the example of islands.

On the other hand, we do not assume (i)–(iii) in this paper. Specifically, we assume that the delay is unknown to the decoder, and we only consider a single pair of source sequences of the length n and encode them at once. Thus, only a single pair of two codewords is used at the decoder. Note that since we do not assume (ii) and (iii), a correlation between the sequences cannot be sufficiently used from a single pair of codewords as described above. Therefore, the rate must be increased compared with that of the case without delay. This is the main reason why there is a difference between achievable regions. Hence, the situation considered in this paper can be regarded as a counterexample to that of Willems.

The rest of this paper is organized as follows. In Sect. 2, we provide some notations and the formal definition of the WAK coding system. In Sect. 3, we show our inner and outer bounds on the achievable rate region. In Sects. 4 and 5, we show proofs for our inner and outer bounds. In Sect. 6, we conclude the paper.

2. Preliminaries

In this section, we provide some notations and the precise definition of the WAK coding system with delayed side information.

We will denote a sequence of symbols $(a_m, a_{m+1}, \cdots, a_{m'})$ by $a_m^{m'}$, where $a_m^{m'} = \emptyset$ if $m > m'$. If $m = 1$, we will denote it by $a^{m'}$ for the sake of simplicity. More generally, we will denote a pair of sequences of symbols $((a_m, b_l), (a_{m+1}, b_{l+1}), \cdots, (a_{m'}, b_{l'}))$ by $(a_m^{m'}, b_l^{l'})$. For any countable sets $\mathcal{X}$ and $\mathcal{Y}$, we will denote the set of all probability mass functions (pmfs) over $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$, and the set of all conditional pmfs from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. We will denote the pmf of a random variable (RV) X on $\mathcal{X}$ by $P_X \in \mathcal{P}(\mathcal{X})$, and the conditional pmf of Y on $\mathcal{Y}$ given X by $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$. We will denote the nth power of a pmf P_X by P_X^n, i.e., $P_X^n(x^n) = \prod_{i=1}^n P_X(x_i)$, and the nth power of a conditional pmf $P_{Y|X}$ by $P_{Y|X}^n$, i.e., $P_{Y|X}^n(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$.

In what follows, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. We will denote a general source $\{(X^n, Y^n)\}_{n=1}^\infty$ (i.e., a sequence of n-length RVs which are not required to satisfy the consistency condition (cf. [13])) by the corresponding boldface letter $(\mathbf{X}, \mathbf{Y})$. Since a DMS is represented by a sequence of independent copies of a pair of RVs (X, Y), we simply express it as (X, Y).

In the WAK coding system, two n-length sequences from a DMS (X, Y) are independently encoded by encoder 1 and encoder 2, respectively. Hence, for positive integers $M_n^{(1)}$ and $M_n^{(2)}$, encoder 1 and encoder 2 are defined by mappings

$$f_n^{(1)} : \mathcal{X}^n \to \mathcal{M}_n^{(1)} = \{1, \cdots, M_n^{(1)}\},$$
$$f_n^{(2)} : \mathcal{Y}^n \to \mathcal{M}_n^{(2)} = \{1, \cdots, M_n^{(2)}\},$$

and rates of these encoders are defined as

$$R_n^{(1)} \triangleq \frac{1}{n} \log M_n^{(1)}, \quad R_n^{(2)} \triangleq \frac{1}{n} \log M_n^{(2)},$$

respectively. Hereafter, log means the natural logarithm.

Since side information may be delayed, encoder 1 encodes a source sequence $X^n = (X_1, X_2, \cdots, X_n)$ while encoder 2 may encode a delayed source sequence $Y_{-2}^{n-3} = (Y_{-2}, Y_{-1}, \cdots, Y_{n-3})$. In general, encoder 1 and encoder 2 encode sequences X^n and $Y_{1-d}^{n-d} = (Y_{1-d}, Y_{2-d}, \cdots, Y_{n-d})$, respectively, where d is a non-negative integer which represents a *relative* delay. We denote Y_{1-d}^{n-d} by $Y_{(d)}^n$ for the sake of brevity.

Without loss of generality, we assume that $d \le n$, because for any $d \ge n$, X^n is independent of $Y_{(d)}^n (= Y_{1-d}^{n-d})$. Thus, we introduce the maximum $d_n \in \{0, 1, 2, \cdots, n\}$ of delays, and denote the sequence $\{d_n\}_{n=1}^{\infty}$ by $\mathbf{d}$. We allow the maximum of delays to change with the block length. We also introduce $\mathcal{D}_n = \{0, 1, 2, \cdots, d_n\}$ that is the set of possible delays. Hence, the delay satisfies $d \in \mathcal{D}_n$ for any block length n. We note that, for $d \in \mathcal{D}_n$, the pmf $P_{X^n Y_{(d)}^n}$ can be written as

$$\begin{aligned} &P_{X^n Y_{(d)}^n}(x^n, y^n) \\ &= P_Y^d(y_1^d) P_{XY}^{n-d}(x_1^{n-d}, y_{d+1}^n) P_X^d(x_{n-d+1}^n), \end{aligned} \tag{1}$$

where P_{XY} is the pmf of the source (X, Y), and P_X and P_Y are marginal pmfs of P_{XY}. We denote the source with delay d by $(\mathbf{X}, \mathbf{Y}_{(d)}) = \{(X^n, Y_{(d)}^n)\}_{n=1}^{\infty}$. By definition, $(\mathbf{X}, \mathbf{Y}_{(d)})$ is a special case of the general source. We note that $(\mathbf{X}, \mathbf{Y}_{(d)})$ denotes the DMS with delay and does not denote the general source with delay. In this paper, we do not consider general sources with delay.

The decoder receives two codewords $f_n^{(1)}(X^n)$ and $f_n^{(2)}(Y_{(d)}^n)$, and outputs an estimate of the source sequence X^n. Hence, the decoder is defined by the mapping

$$\varphi_n : \mathcal{M}_n^{(1)} \times \mathcal{M}_n^{(2)} \to \mathcal{X}^n.$$

Then, for a DMS (X, Y) and a delay d, the error probability is defined as

$$\varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n) \triangleq \Pr\{\varphi_n(f_n^{(1)}(X^n),\ f_n^{(2)}(Y_{(d)}^n)) \ne X^n\}.$$

More generally, we will denote the error probability for a general source $(\mathbf{X}, \mathbf{Y})$ by $\varepsilon_{\mathbf{XY}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$, i.e.,

$$\varepsilon_{\mathbf{XY}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n) \triangleq \Pr\{\varphi_n(f_n^{(1)}(X^n),\ f_n^{(2)}(Y^n)) \ne X^n\}.$$

Thus, by recalling that the source $(\mathbf{X}, \mathbf{Y}_{(d)})$ is a special case of the general source, the error probability $\varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ can be regraded as the error probability $\varepsilon_{\mathbf{XY}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ in the case where $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \mathbf{Y}_{(d)})$. We will sometimes omit the code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ in the notation of $\varepsilon_{\mathbf{XY}}^{(n)}$ when it is clear from the context.

In this coding system, we assume that the delay d is *unknown* but the maximum $\mathbf{d}$ of delays is *known* to the encoders and the decoder. More precisely, the code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ is independent of d, but is allowed to depend on $\mathbf{d}$.

We now define *achievability* and *achievable rate region* for the WAK coding system with delayed side information.

Definition 1 (Achievability)**.** For a DMS (X, Y) and a maximum $\mathbf{d}$ of delays, a pair (R_1, R_2) is called *achievable* if and only if there exists a sequence of codes $\{(f_n^{(1)}, f_n^{(2)}, \varphi_n)\}$ satisfying

$$\limsup_{n\to\infty} R_n^{(1)} \le R_1,\ \limsup_{n\to\infty} R_n^{(2)} \le R_2, \tag{2}$$

and

$$\lim_{n\to\infty} \max_{d\in\mathcal{D}_n} \varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n) = 0. \tag{3}$$

Definition 2 (Achievable rate region)**.** For a DMS (X, Y) and a maximum $\mathbf{d}$ of delays, the achievable rate region $\mathcal{R}_{\mathbf{d}}^{(X,Y)}$ is defined by

$$\begin{aligned} \mathcal{R}_{\mathbf{d}}^{(X,Y)} \triangleq \mathrm{cl}\,(\{&(R_1, R_2) : (R_1, R_2) \text{ is achievable} \\ &\text{for the source } (X, Y) \text{ and the maximum } \mathbf{d}\}), \end{aligned}$$

where $\mathrm{cl}(\cdot)$ denotes the closure operation.

3. Inner and Outer Bounds on the Achievable Rate Region

In this section, we show an inner bound and an outer bound on the achievable rate region. To this end, we introduce some definitions.

In what follows, let $\mathcal{U}$ be a countably infinite set unless otherwise stated. For real numbers $\alpha, \beta \in [0, 1]$, we define

$$\begin{aligned} \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) &\triangleq \{(R_1, R_2) : R_1 \ge H(X|U) + \alpha I(X; U), \\ &\qquad R_2 \ge (1-\beta) I(Y; U)\}, \\ \mathcal{A}_{\alpha,\beta}^{(X,Y)} &\triangleq \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}), \end{aligned}$$

where the triple of RVs (X, Y, U) is drawn according to $P_{XY} \times P_{U|Y}$ (and hence the Markov chain $X - Y - U$ holds). We also define

$$\underline{\Delta}_{\mathbf{d}} \triangleq \liminf_{n\to\infty} \frac{d_n}{n}, \quad \overline{\Delta}_{\mathbf{d}} \triangleq \limsup_{n\to\infty} \frac{d_n}{n}.$$

If $\{d_n/n\}$ converges as $n \to \infty$, we define

$$\Delta_{\mathbf{d}} \triangleq \lim_{n\to\infty} \frac{d_n}{n}.$$

Remark 1. $\mathcal{A}_{0,0}^{(X,Y)}$ is the achievable rate region for the case without delay [2], [3]. By noticing that $\mathcal{A}_{0,0}^{(X,Y)}$ is a closed and convex set (cf. [2]), $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$ is also a closed and convex set. Furthermore, as $\mathcal{A}_{0,0}^{(X,Y)}$ does, $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$ will be unchanged if we only consider $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$ such that $|\mathcal{U}| = |\mathcal{Y}| + 1$ (cf. [14]). We show these properties in Appendix A.

Now we give our bounds. The next theorem shows the outer bound on the achievable rate region.

Theorem 1. For a DMS (X, Y) and a maximum $\mathbf{d}$, we have

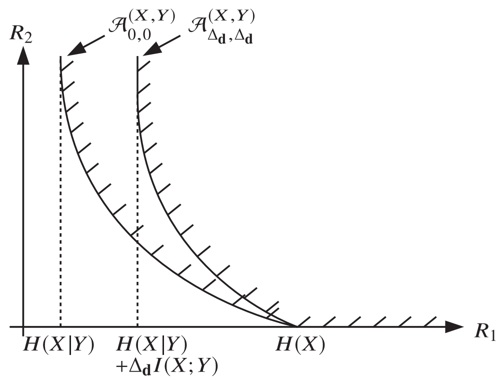

Fig. 1 An image of achievable rate regions.

$$\mathcal{R}^{(X,Y)}_\mathbf{d} \subseteq \mathcal{A}^{(X,Y)}_{\underline{\Delta}_\mathbf{d},\overline{\Delta}_\mathbf{d}}.$$

The next theorem shows the inner bound on the achievable rate region.

Theorem 2. For a DMS (X, Y) and a maximum $\mathbf{d}$ such that $0 < \underline{\Delta}_\mathbf{d} \le \overline{\Delta}_\mathbf{d} < 1$ or $\overline{\Delta}_\mathbf{d} = 0$, we have

$$\mathcal{A}^{(X,Y)}_{\overline{\Delta}_\mathbf{d},\underline{\Delta}_\mathbf{d}} \subseteq \mathcal{R}^{(X,Y)}_\mathbf{d}.$$

Remark 2. If the decoder does not employ side information, the coding system can be regarded as the source coding system without side information. In this case, we can easily show that any pair (R_1, R_2) satisfying $R_1 \ge H(X)$ and $R_2 \ge 0$ is achievable. Thus, it always holds that $\mathcal{A}^{(X,Y)}_{1,1} = \{(R_1, R_1) : R_1 \ge H(X), R_2 \ge 0\} \subseteq \mathcal{R}^{(X,Y)}_\mathbf{d}$.

According to Theorem 1, Theorem 2, and Remark 2, we immediately obtain the following corollary.

Corollary 1. For a DMS (X, Y) and a maximum $\mathbf{d}$ such that $\{d_n/n\}$ converges as $n \to \infty$, we have

$$\mathcal{R}^{(X,Y)}_\mathbf{d} = \mathcal{A}^{(X,Y)}_{\Delta_\mathbf{d},\Delta_\mathbf{d}}.$$

This corollary shows that when $d_n = o(n)$ the achievable rate region coincides with that for the case without delay. However, in general, the achievable rate region does not coincide with it. To show this fact, we consider the minimum rate $R_{1,\Delta_\mathbf{d}} = \inf\{R_1 : \exists (R_1, R_2) \in \mathcal{A}^{(X,Y)}_{\Delta_\mathbf{d},\Delta_\mathbf{d}}\}$ on one side. Then, it holds that $R_{1,\Delta_\mathbf{d}} = H(X|Y) + \Delta_\mathbf{d} I(X; Y)$ (see Appendix B for details). Thus, when $\Delta_\mathbf{d} > 0$, the minimum rate $H(X|Y) + \Delta_\mathbf{d} I(X; Y)$ does not coincide with the minimum rate $H(X|Y)$ for the case without delay (see an image in Fig. 1).

4. Proof of Theorem 1

In this section, we prove Theorem 1 that gives an outer bound on $\mathcal{R}^{(X,Y)}_\mathbf{d}$.

Let $(R_1, R_2) \in \mathcal{R}^{(X,Y)}_\mathbf{d}$. Then there exists a sequence of codes $\{(f^{(1)}_n, f^{(2)}_n, \varphi_n)\}$ satisfying (2) and (3). For these codes and an arbitrarily fixed delay $d \in \mathcal{D}_n$, let $M_1 \triangleq f^{(1)}_n(X^n)$ and $M_{2,d} \triangleq f^{(2)}_n(Y^n_{(d)})$.

By Fano's inequality [15], for any $d \in \mathcal{D}_n$, we have

$$\begin{aligned} H(X^n|M_1, M_{2,d}) &\le H(X^n|\varphi_n(M_1, M_{2,d})) \\ &\le \varepsilon^{(n)}_{\mathbf{XY}_{(d)}} \log|\mathcal{X}|^n + 1 \\ &= n\epsilon_{n,d}, \end{aligned} \tag{4}$$

where $\epsilon_{n,d} = \varepsilon^{(n)}_{\mathbf{XY}_{(d)}} \log|\mathcal{X}| + \frac{1}{n}$. Thus, we have

$$\begin{aligned} nR^{(1)}_n &\ge H(M_1) \\ &\ge H(M_1|M_{2,d}) \\ &\overset{\text{(a)}}{\ge} H(M_1|M_{2,d}) + H(X^n|M_1, M_{2,d}) - n\epsilon_{n,d} \\ &= H(X^n, M_1|M_{2,d}) - n\epsilon_{n,d} \\ &= H(X^n|M_{2,d}) - n\epsilon_{n,d}, \end{aligned} \tag{5}$$

where (a) comes from (4). The first term in the right-hand side is further bounded as follows:

$$\begin{aligned} &H(X^n|M_{2,d}) \\ &= \sum_{i=1}^{n-d} H(X_i|X^{i-1}, M_{2,d}) + \sum_{i=n-d+1}^{n} H(X_i|X^{i-1}, M_{2,d}) \\ &\overset{\text{(a)}}{=} \sum_{i=1}^{n-d} H(X_i|X^{i-1}, M_{2,d}) + \sum_{i=n-d+1}^{n} H(X_i) \\ &\ge \sum_{i=1}^{n-d} H(X_i|X^{i-1}_{1-d}, Y^{i-1}_{1-d}, M_{2,d}) + \sum_{i=n-d+1}^{n} H(X_i) \\ &\overset{\text{(b)}}{=} \sum_{i=1}^{n-d} H(X_i|U_i) + dH(X) \\ &\overset{\text{(c)}}{=} (n-d) \sum_{i=1}^{n-d} P_{Q^{(n)}}(i) H(X|U_{Q^{(n)}}, Q^{(n)} = i) + dH(X) \\ &= (n-d)H(X|U_{Q^{(n)}}, Q^{(n)}) + dH(X) \\ &= nH(X|U_{Q^{(n)}}, Q^{(n)}) + dI(X; U_{Q^{(n)}}, Q^{(n)}) \\ &\overset{\text{(d)}}{=} nH(X|U^{(n)}) + dI(X; U^{(n)}), \end{aligned} \tag{6}$$

where (a) comes from the fact that X_i is independent of $(X^{i-1}, M_{2,d})$ for all $i \ge n - d + 1$, (b) comes from $U_i \triangleq (X^{i-1}_{1-d}, Y^{i-1}_{1-d}, M_{2,d})$, (c) comes from the fact that $X_i - Y_i - U_i$ and the tuple of RVs $(Q^{(n)}, X, Y, U_{Q^{(n)}})$ is defined as

$$P_{Q^{(n)}XYU_{Q^{(n)}}}(i, x, y, u) = \frac{1}{n-d} P_{XY}(x, y) P_{U_i|Y_i}(u|y), \tag{7}$$

and (d) comes from $U^{(n)} \triangleq (U_{Q^{(n)}}, Q^{(n)})$. We note that $P_{Q^{(n)}}$ is the pmf of the RV $Q^{(n)}$, and it holds that $P_{Q^{(n)}}(i) = \frac{1}{n-d}$ for any $i \in \{1, \cdots, n-d\}$.

On the other hand, we have

$$\begin{aligned} nR^{(2)}_n &\ge H(M_{2,d}) \\ &\overset{\text{(a)}}{=} H(M_{2,d}) - H(M_{2,d}|Y^{n-d}_{1-d}) \\ &= I(M_{2,d}; Y^{n-d}_{1-d}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1-d}^{n-d} I(M_{2,d}; Y_i | Y_{1-d}^{i-1}) \\
&\overset{\text{(b)}}{=} \sum_{i=1-d}^{n-d} I(M_{2,d}, Y_{1-d}^{i-1}; Y_i) \\
&\overset{\text{(c)}}{=} \sum_{i=1-d}^{n-d} I(M_{2,d}, Y_{1-d}^{i-1}, X_{1-d}^{i-1}; Y_i) \\
&\geq \sum_{i=1}^{n-d} I(M_{2,d}, Y_{1-d}^{i-1}, X_{1-d}^{i-1}; Y_i) \\
&= \sum_{i=1}^{n-d} I(U_i; Y_i) \\
&\overset{\text{(d)}}{=} (n-d) \sum_{i=1}^{n-d} P_{Q^{(n)}}(i) I(U_{Q^{(n)}}; Y | Q^{(n)} = i) \\
&= (n-d) I(U_{Q^{(n)}}; Y | Q^{(n)}) \\
&\overset{\text{(e)}}{=} (n-d) I(U_{Q^{(n)}}, Q^{(n)}; Y) \\
&= (n-d) I(U^{(n)}; Y), \qquad (8)
\end{aligned}
$$

where (a) follows since $H(M_{2,d}|Y_{1-d}^{n-d}) = 0$ (because $M_{2,d}$ is a function of Y_{1-d}^{n-d}), (b) comes from the fact that Y_i is independent of Y_{1-d}^{i-1}, (c) comes from the fact that $X_{1-d}^{i-1} - (M_{2,d}, Y_{1-d}^{i-1}) - Y_i$, (d) comes from the definition of $(Q^{(n)}, X, Y, U_{Q^{(n)}})$ (see (7)), and (e) comes from the fact that $Q^{(n)}$ is independent of Y.

According to (5), (6), and (8) and setting that $d = d_n$, we have

$$
R_n^{(1)} \geq H(X|U^{(n)}) + \frac{d_n}{n} I(X; U^{(n)}) - \epsilon_{n,d_n}, \qquad (9)
$$

$$
R_n^{(2)} \geq \left(1 - \frac{d_n}{n}\right) I(Y; U^{(n)}). \qquad (10)
$$

On the other hand, for any $\epsilon > 0$ and sufficiently large $n > 0$, we have

$$
R_i \overset{\text{(a)}}{\geq} \limsup_{n\to\infty} R_n^{(i)} \geq R_n^{(i)} - \epsilon, \qquad (11)
$$

$$
\epsilon_{n,d_n} \overset{\text{(b)}}{\leq} \epsilon, \qquad (12)
$$

$$
\overline{\Delta}_{\mathbf{d}} + \epsilon \geq \frac{d_n}{n} \geq \underline{\Delta}_{\mathbf{d}} - \epsilon, \qquad (13)
$$

where (a) comes from the definition of the achievability, and (b) comes from the fact that

$$
\lim_{n\to\infty} \varepsilon_{\mathbf{XY}_{(d_n)}}^{(n)} \leq \lim_{n\to\infty} \max_{d\in\mathcal{D}_n} \varepsilon_{\mathbf{XY}_{(d)}}^{(n)} = 0.
$$

By combining (9)–(13), for sufficiently large $n > 0$, we have

$$
\begin{aligned}
R_1 &\geq H(X|U^{(n)}) + \frac{d_n}{n} I(X; U^{(n)}) - 2\epsilon \\
&\geq H(X|U^{(n)}) + \underline{\Delta}_{\mathbf{d}} I(X; U^{(n)}) - 2\epsilon - \epsilon \log|\mathcal{X}|, \qquad (14)
\end{aligned}
$$

$$
R_2 \geq (1 - \overline{\Delta}_{\mathbf{d}}) I(Y; U^{(n)}) - \epsilon - \epsilon \log|\mathcal{Y}|. \qquad (15)
$$

By noticing that $X - Y - U^{(n)}$, inequalities (14) and (15) show that for any $\epsilon > 0$, there exists $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$ such that

$$
(R_1 + \epsilon, R_2 + \epsilon) \in \hat{\mathcal{A}}_{\underline{\Delta}_{\mathbf{d}}, \overline{\Delta}_{\mathbf{d}}}^{(X,Y)}(P_{U|Y}) \subseteq \mathcal{A}_{\underline{\Delta}_{\mathbf{d}}, \overline{\Delta}_{\mathbf{d}}}^{(X,Y)}.
$$

Since $\epsilon > 0$ is arbitrary and $\mathcal{A}_{\underline{\Delta}_{\mathbf{d}}, \overline{\Delta}_{\mathbf{d}}}^{(X,Y)}$ is a closed set, we have $(R_1, R_2) \in \mathcal{A}_{\underline{\Delta}_{\mathbf{d}}, \overline{\Delta}_{\mathbf{d}}}^{(X,Y)}$. By recalling that $(R_1, R_2) \in \mathcal{R}_{\mathbf{d}}^{(X,Y)}$, this completes the proof of Theorem 1.

5. Proof of Theorem 2

In order to prove Theorem 2, we use a similar proof technique as in [5]. The proof consists of three steps: First, we define a *mixed* source from original sources with delay. Next, we show that if the error probability of a code for the mixed source vanishes at the order $o(n^{-1})$, that of the same code for sources with delay also vanishes. Finally, we show that there exists such a code as long as the pair of rates is in the inner bound $\mathcal{A}_{\underline{\Delta}_{\mathbf{d}}, \overline{\Delta}_{\mathbf{d}}}^{(X,Y)}$. This implies that any rate pair in the inner bound is achievable.

In this final step, as mentioned earlier, we used Gallager's random coding technique [6], [7] in our previous study [5]. However, this is rather difficult to simply show the existence of a code of which error probability vanishes at a desired order. Thus, to simplify the final step, we use a known result [10] and the Chernoff bound (cf. e.g. [8], [9]) in this paper.

First of all, we define a mixed source. For any $n > 0$ and an arbitrarily fixed $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$, let $(\tilde{X}^n, \tilde{Y}^n, \tilde{U}^n)$ be a triple of RVs defined by

$$
\begin{aligned}
P_{\tilde{X}^n\tilde{Y}^n\tilde{U}^n}(x^n, y^n, u^n) &= P_U^{d_n}(u^{d_n}) P_{U|Y}^{n-d_n}(u_{d_n+1}^n | y_{d_n+1}^n) \\
&\times \sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_{X^nY^n_{(d)}}(x^n, y^n). \qquad (16)
\end{aligned}
$$

We note that $\tilde{X}^n - \tilde{Y}^n - \tilde{U}^n$, and

$$
P_{\tilde{Y}^n\tilde{U}^n}(y^n, u^n) = P_U^{d_n}(u^{d_n}) P_{U|Y}^{n-d_n}(u_{d_n+1}^n | y_{d_n+1}^n) P_Y^n(y^n), \qquad (17)
$$

$$
P_{\tilde{X}^n\tilde{U}^n}(x^n, u^n) = \sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_{X^n_{(d)}U^n}(x^n, u^n), \qquad (18)
$$

where $X - Y - U$ and

$$
\begin{aligned}
P_{X^n_{(d)}U^n}(x^n, u^n) &= P_U^n(u^n) P_X^{d_n-d}(x_1^{d_n-d}) \\
&\times P_{X|U}^{n-d_n}(x_{d_n-d+1}^{n-d} | u_{d_n+1}^n) P_X^d(x_{n-d+1}^n). \qquad (19)
\end{aligned}
$$

We give a precise derivation of (18) in Appendix C. By using this pmf, we can define the mixed source $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}, \tilde{\mathbf{U}}) \triangleq \{(\tilde{X}^n, \tilde{Y}^n, \tilde{U}^n)\}_{n=1}^{\infty}$.

For this source, we have the next lemma.

Lemma 1. For any code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$, we have

$$
\max_{d\in\mathcal{D}_n} \varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n) \leq (n+1) \varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n).
$$

Proof. Since the code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ is a set of deterministic functions, for any source $(\mathbf{X},\mathbf{Y})$, the error probability $\varepsilon_{\mathbf{XY}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ is represented as

$$\varepsilon_{\mathbf{XY}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n) = \sum_{(x^n,y^n)\in\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n)} P_{X^nY^n}(x^n,y^n),$$

where $\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ is the set of pairs of sequences which cannot be decoded correctly, i.e.

$$\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n) \triangleq \{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n : \varphi_n(f_n^{(1)}(x^n), f_n^{(2)}(y^n)) \neq x^n)\}.$$

Thus, we have

$$\begin{aligned}
&\max_{d\in\mathcal{D}_n} \varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)\\
&= \max_{d\in\mathcal{D}_n} \sum_{(x^n,y^n)\in\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n)} P_{X^nY^n_{(d)}}(x^n,y^n)\\
&\le \sum_{d\in\mathcal{D}_n} \sum_{(x^n,y^n)\in\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n)} P_{X^nY^n_{(d)}}(x^n,y^n)\\
&= |\mathcal{D}_n| \sum_{(x^n,y^n)\in\mathcal{E}_n(f_n^{(1)}, f_n^{(2)}, \varphi_n)} P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\\
&= |\mathcal{D}_n| \varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)\\
&\le (n+1)\varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n),
\end{aligned}$$

where the last inequality follows since $0 \le d_n \le n$. □

According to this lemma, if the error probability $\varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ vanishes at the order $o(n^{-1})$, the error probability $\max_{d\in\mathcal{D}_n} \varepsilon_{\mathbf{XY}_{(d)}}^{(n)}(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ also vanishes as n increases. We will show the existence of a code of which error probability vanishes exponentially rather than $o(n^{-1})$. To this end, we give a known result [10] for the WAK coding system.

Theorem 3 ([10, Corollary 6]). Let $(\mathbf{X},\mathbf{Y},\mathbf{U}) = \{(X^n,Y^n,U^n)\}$ be a general source such that $X^n - Y^n - U^n$. Then, for arbitrary $\gamma_1,\gamma_2 \ge 0$, $n > 0$, and $M_n^{(1)}, M_n^{(2)} > 0$, there exists a code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ whose error probability satisfies

$$\varepsilon_{\mathbf{XY}}^{(n)} \le \Pr\{(U^n,X^n)\in\mathcal{T}_1^{(n)}(\gamma_1)^c \cup (U^n,Y^n)\in\mathcal{T}_2^{(n)}(\gamma_2)^c\} + e^{-nR_n^{(1)}+\gamma_1} + \frac{1}{2}\sqrt{e^{-nR_n^{(2)}+\gamma_2}},$$

where the superscript c denotes the complement of a set, and

$$\mathcal{T}_1^{(n)}(\gamma_1) \triangleq \left\{(u^n,x^n)\in\mathcal{U}^n\times\mathcal{X}^n : \log\frac{1}{P_{X^n|U^n}(x^n|u^n)} \le \gamma_1\right\},$$

$$\mathcal{T}_2^{(n)}(\gamma_2) \triangleq \left\{(u^n,y^n)\in\mathcal{U}^n\times\mathcal{Y}^n : \log\frac{P_{Y^n|U^n}(y^n|u^n)}{P_{Y^n}(y^n)} \le \gamma_2\right\}.$$

Applying this theorem to our mixed source, we have the following two corollaries. Here, for any RVs (X,Y,U), we use the following notations:

$$\begin{aligned}
i(X) &= -\log P_X(X),\\
i(X|U) &= -\log P_{X|U}(X|U),\\
i(Y;U) &= \log\frac{P_{Y|U}(Y|U)}{P_Y(Y)},\\
\psi_X(\gamma) &= \sup_{\lambda\ge 0}\left(\lambda\gamma - \log \mathrm{E}\left[e^{\lambda X}\right]\right).
\end{aligned}$$

Corollary 2. For arbitrary $\gamma_{1,1},\gamma_{1,2},\gamma_2 \ge 0$, $n > 0$, and $M_n^{(1)}, M_n^{(2)} > 0$, there exists a code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ whose error probability satisfies

$$\begin{aligned}
\varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)} &\le e^{-d_n\psi_{i(X)}\left(\gamma_{1,1}-\frac{\log|\mathcal{D}_n|}{d_n}\right)} + e^{-(n-d_n)\psi_{i(X|U)}(\gamma_{1,2})}\\
&\quad + e^{-(n-d_n)\psi_{i(Y;U)}(\gamma_2)} + e^{-nR_n^{(1)}+d_n\gamma_{1,1}+(n-d_n)\gamma_{1,2}}\\
&\quad + \frac{1}{2}\sqrt{e^{-nR_n^{(2)}+(n-d_n)\gamma_2}},
\end{aligned}$$

where $(X,Y,U) \sim P_{XY}\times P_{U|Y}$.

Proof. In Theorem 3, we substitute $\gamma_1^{(n)} = d_n\gamma_{1,1}+(n-d_n)\gamma_{1,2}$ and $\gamma_2^{(n)} = (n-d_n)\gamma_2$ into γ_1 and γ_2, respectively. Then, we have

$$\begin{aligned}
\varepsilon_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}^{(n)} &\le \Pr\{(\tilde{U}^n,\tilde{X}^n)\in\mathcal{T}_1^{(n)}(\gamma_1^{(n)})^c\}\\
&\quad + \Pr\{(\tilde{U}^n,\tilde{Y}^n)\in\mathcal{T}_2^{(n)}(\gamma_2^{(n)})^c\}\\
&\quad + e^{-nR_n^{(1)}+\gamma_1^{(n)}} + \frac{1}{2}\sqrt{e^{-nR_n^{(2)}+\gamma_2^{(n)}}}.
\end{aligned} \tag{20}$$

The first term in the right-hand side is bounded as follows:

$$\begin{aligned}
&\Pr\{(\tilde{U}^n,\tilde{X}^n)\in\mathcal{T}_1^{(n)}(\gamma_1^{(n)})^c\}\\
&\overset{\text{(a)}}{=} \sum_{\substack{(x^n,u^n)\in\mathcal{X}^n\times\mathcal{U}^n:\\ \log\frac{|\mathcal{D}_n|}{\sum_{d\in\mathcal{D}_n}P_{X^n_{(d)}|U^n}(x^n|u^n)}>\gamma_1^{(n)}}} \sum_{d\in\mathcal{D}_n}\frac{1}{|\mathcal{D}_n|}P_{X^n_{(d)}U^n}(x^n,u^n)\\
&\overset{\text{(b)}}{\le} \sum_{d\in\mathcal{D}_n}\frac{1}{|\mathcal{D}_n|} \sum_{\substack{(x^n,u^n)\in\mathcal{X}^n\times\mathcal{U}^n:\\ \log\frac{|\mathcal{D}_n|}{P_{X^n_{(d)}|U^n}(x^n|u^n)}>\gamma_1^{(n)}}} P_{X^n_{(d)}U^n}(x^n,u^n)\\
&= \sum_{d\in\mathcal{D}_n}\frac{1}{|\mathcal{D}_n|}\Pr\left\{i(X^n_{(d)}|U^n) > \gamma_1^{(n)} - \log|\mathcal{D}_n|\right\},
\end{aligned} \tag{21}$$

where (a) comes from the fact that the pmf of $(\tilde{X}^n,\tilde{U}^n)$ can be expressed as (18), and (b) follows since it holds that

$$\log\frac{|\mathcal{D}_n|}{\sum_{d\in\mathcal{D}_n}P_{X^n_{(d)}|U^n}(x^n|u^n)} \le \log\frac{|\mathcal{D}_n|}{P_{X^n_{(d)}|U^n}(x^n|u^n)}$$

for any $d\in\mathcal{D}_n$. Here, for any $d\in\mathcal{D}_n$, we have

$$\begin{aligned}
&\Pr\left\{i(X^n_{(d)}|U^n) > \gamma_1^{(n)} - \log|\mathcal{D}_n|\right\}\\
&= \Pr\left\{-\log P_X^{d_n-d}(X^{d_n-d}_{(d),1})P_X^d(X^n_{(d),n-d+1})\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \log P^{n-d_n}_{X|U}(X^{n-d}_{(d),d_n-d+1}|U^n_{d_n+1}) > \gamma_1^{(n)} - \log|\mathcal{D}_n|\Big\} \\
&\overset{(a)}{=} \Pr\left\{i(X^{d_n}) + i(X^n_{d_n+1}|U^n_{d_n+1}) > \gamma_1^{(n)} - \log|\mathcal{D}_n|\right\} \qquad (22)\\
&\overset{(b)}{\le} \Pr\left\{i(X^{d_n}) > d_n\gamma_{1,1} - \log|\mathcal{D}_n|\right\} \\
&\quad + \Pr\left\{i(X^n_{d_n+1}|U^n_{d_n+1}) > (n-d_n)\gamma_{1,2}\right\} \\
&\overset{(c)}{\le} e^{-\psi_{i(X^{d_n})}\left(d_n\gamma_{1,1} - \log|\mathcal{D}_n|\right)} + e^{-\psi_{i(X^n_{d_n+1}|U^n_{d_n+1})}((n-d_n)\gamma_{1,2})} \\
&= e^{-d_n\psi_{i(X)}\left(\gamma_{1,1} - \frac{\log|\mathcal{D}_n|}{d_n}\right)} + e^{-(n-d_n)\psi_{i(X|U)}(\gamma_{1,2})}, \qquad (23)
\end{aligned}
$$

where the sequence of RVs $\{(X_i, U_i)\}_{i=1}^n$ is i.i.d. drawn according to P_{XU}, (a) comes from the fact that $(X^{d_n-d}_{(d),1}, X^n_{(d),n-d+1}) = X^{d_n}$ and $(X^{n-d}_{(d),d_n-d+1}, U^n_{d_n+1}) = (X^n_{d_n+1}, U^n_{d_n+1})$ according to (19), (b) comes from the fact that

$$\Pr\{X + Y > \alpha + \beta\} \le \Pr\{X > \alpha\} + \Pr\{Y > \beta\},$$

and (c) follows the Chernoff bound:

$$\Pr\{X \ge \gamma\} \le e^{-\psi_X(\gamma)}.$$

On the other hand, the second term of (20) is bounded as

$$
\begin{aligned}
&\Pr\{(\tilde{U}^n, \tilde{Y}^n) \in \mathcal{T}_2^{(n)}(\gamma_2^{(n)})^c\} \\
&= \Pr\left\{\log\frac{P_{\tilde{Y}^n|\tilde{U}^n}(\tilde{Y}^n|\tilde{U}^n)}{P_{\tilde{Y}^n}(\tilde{Y}^n)} > \gamma_2^{(n)}\right\} \\
&\overset{(a)}{=} \Pr\left\{\log\frac{P^{n-d_n}_{Y|U}(\tilde{Y}^n_{d_n+1}|\tilde{U}^n_{d_n+1})}{P^{n-d_n}_{Y}(\tilde{Y}^n_{d_n+1})} > \gamma_2^{(n)}\right\} \\
&\overset{(b)}{=} \Pr\left\{\log\frac{P^{n-d_n}_{Y|U}(Y^{n-d_n}|U^{n-d_n})}{P^{n-d_n}_{Y}(Y^{n-d_n})} > \gamma_2^{(n)}\right\} \\
&= \Pr\left\{i(Y^{n-d_n}; U^{n-d_n}) > \gamma_2^{(n)}\right\} \\
&\overset{(c)}{\le} e^{-\psi_{i(Y^{n-d_n};U^{n-d_n})}(\gamma_2^{(n)})} \qquad (24)\\
&= e^{-(n-d_n)\psi_{i(Y;U)}(\gamma_2)}, \qquad (25)
\end{aligned}
$$

where the sequence of RVs $\{(Y_i, U_i)\}_{i=1}^n$ is i.i.d. drawn according to P_{YU}, (a) comes from (17), (b) comes from $(\tilde{Y}^{n-d_n}, \tilde{U}^n_{d_n+1}) = (Y^{n-d_n}, U^{n-d_n})$, and (c) comes from the Chernoff bound.

Combining (20), (21), (23), and (25), we have the desired bound. □

Corollary 3. Let $\overline{\Delta}_{\mathbf{d}} = 0$. Then, for arbitrary $\gamma_1, \gamma_2 \ge 0$, $\delta > 0$, $M_n^{(1)}, M_n^{(2)} > 0$, and sufficiently large $n > 0$, there exists a code $(f_n^{(1)}, f_n^{(2)}, \varphi_n)$ whose error probability satisfies

$$
\begin{aligned}
\varepsilon^{(n)}_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}} &\le e^{-n(\psi_{i(X|U)}(\gamma_1)-\delta)} + e^{-n(\psi_{i(Y;U)}(\gamma_2)-\delta)} \\
&\quad + e^{-nR_n^{(1)}+n\gamma_1} + \frac{1}{2}\sqrt{e^{-nR_n^{(2)}+n\gamma_2}}.
\end{aligned}
$$

Proof. In Theorem 3, we substitute $\gamma_1^{(n)} = n\gamma_1$ and $\gamma_2^{(n)} = n\gamma_2$ into γ_1 and γ_2, respectively. Then, by following the same way as the proof of Corollary 2, we have (20), (21), (22), and (24). In what follows, let $p_1^{(n)}$ denote the right-hand side of (22), and $p_2^{(n)}$ denote the right-hand side of (24) for the sake of brevity.

For any $\lambda \ge 0$, $p_1^{(n)}$ can be bounded as follows:

$$
\begin{aligned}
p_1^{(n)} &\overset{(a)}{\le} e^{-\psi_{i(X^{d_n})+i(X^n_{d_n+1}|U^n_{d_n+1})}(\gamma_1^{(n)} - \log|\mathcal{D}_n|)} \\
&\le e^{-\lambda(\gamma_1^{(n)} - \log|\mathcal{D}_n|) + \log \mathrm{E}\left[\exp\left(\lambda i(X^{d_n}) + \lambda i(X^n_{d_n+1}|U^n_{d_n+1})\right)\right]} \\
&= e^{-n\left(\lambda\gamma_1 - \log \mathrm{E}\left[\exp(\lambda i(X|U))\right] - \delta_n\right)},
\end{aligned}
$$

where (a) comes from the Chernoff bound, and $\delta_n = \lambda\frac{\log|\mathcal{D}_n|}{n} - \frac{d_n}{n}\log \mathrm{E}\left[\exp\left(\lambda i(X|U)\right)\right] + \frac{d_n}{n}\log \mathrm{E}\left[\exp\left(\lambda i(X)\right)\right]$.

Since $\mathrm{E}\left[\exp\left(\lambda i(X|U)\right)\right] \ge 1$, δ_n is bounded as

$$\delta_n \le \lambda\frac{\log|\mathcal{D}_n|}{n} + \frac{d_n}{n}\log \mathrm{E}\left[\exp\left(\lambda i(X)\right)\right].$$

Since $\overline{\Delta}_{\mathbf{d}} = 0$ and $\mathrm{E}[\exp(\lambda i(X))] < \infty$, we have $\limsup_{n\to\infty} \delta_n \le 0$ for any $\lambda \ge 0$. Thus, we have

$$\limsup_{n\to\infty} \frac{1}{n}\log p_1^{(n)} \le -\lambda\gamma_1 + \log \mathrm{E}\left[\exp\left(\lambda i(X|U)\right)\right].$$

Since this holds for any $\lambda \ge 0$, we have

$$\limsup_{n\to\infty} \frac{1}{n}\log p_1^{(n)} \le -\psi_{i(X|U)}(\gamma_1).$$

Now, for any $\delta > 0$ and sufficiently large $n > 0$, we have

$$\frac{1}{n}\log p_1^{(n)} \le \limsup_{n\to\infty} \frac{1}{n}\log p_1^{(n)} + \delta \le -\psi_{i(X|U)}(\gamma_1) + \delta.$$

That is

$$p_1^{(n)} \le e^{-n(\psi_{i(X|U)}(\gamma_1)-\delta)}. \qquad (26)$$

On the other hand, $p_2^{(n)}$ can be bounded as follows:

$$
\begin{aligned}
p_2^{(n)} &\le e^{-\lambda\gamma_2^{(n)} + (n-d_n)\log\left(\mathrm{E}\left[\exp(\lambda i(Y;U))\right]\right)} \\
&= e^{-n\left(\lambda\gamma_2 - \log\left(\mathrm{E}\left[\exp(\lambda i(Y;U))\right]\right) + \delta_n\right)}
\end{aligned}
$$

where

$$\delta_n = \frac{d_n}{n}\log\left(\mathrm{E}\left[\exp\left(\lambda i(Y;U)\right)\right]\right).$$

Since $\underline{\Delta}_{\mathbf{d}} = \overline{\Delta}_{\mathbf{d}} = 0$ and $\mathrm{E}[\exp(\lambda i(Y;U))] > 0$, we have

$$\limsup_{n\to\infty} \frac{1}{n}\log p_2^{(n)} \le -\lambda\gamma_2 + \log\left(\mathrm{E}\left[\exp\left(\lambda i(Y;U)\right)\right]\right).$$

Since this holds for any $\lambda \ge 0$, we have

$$\limsup_{n\to\infty} \frac{1}{n}\log p_2^{(n)} \le -\psi_{i(Y;U)}(\gamma_2).$$

Now, for any $\delta > 0$ and sufficiently large $n > 0$, we have

$$\frac{1}{n}\log p_2^{(n)} \le \limsup_{n\to\infty} \frac{1}{n}\log p_2^{(n)} + \delta \le -\psi_{i(Y;U)}(\gamma_2) + \delta.$$

That is

$$p_2^{(n)} \le e^{-n(\psi_{i(Y;U)}(\gamma_2)-\delta)}. \quad (27)$$

Combining (20), (21), (26), and (27), we have the desired bound. □

The function $\psi_X(\gamma)$ has the following basic property (cf. e.g. [9, Theorem 14.3]).

Lemma 2. $\psi_X(\gamma) = 0$ for any $\gamma \le \mathrm{E}[X]$ and $\psi_X(\gamma) > 0$ for any $\gamma > \mathrm{E}[X]$.

Now we prove Theorem 2.

Proof of Theorem 2. For $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$, $\epsilon > 0$, and (R_1, R_2) such that

$$R_1 = H(X|U) + \overline{\Delta}_{\mathbf{d}} I(X;U) + 2\epsilon + \epsilon \log|\mathcal{X}|,$$
$$R_2 = (1-\underline{\Delta}_{\mathbf{d}})I(Y;U) + 2\epsilon + \epsilon \log|\mathcal{Y}|,$$

we set $M_n^{(i)} = \lceil e^{nR_i} \rceil$ for all $i \in \{1,2\}$.

We will show that (R_1, R_2) is achievable for any $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$ and $\epsilon > 0$. This implies that

$$\left(H(X|U) + \overline{\Delta}_{\mathbf{d}} I(X;U), (1-\underline{\Delta}_{\mathbf{d}})I(Y;U)\right) \in \mathcal{R}_{\mathbf{d}}^{(X,Y)}$$

for any $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$. Thus we have the inner bound

$$\mathcal{A}_{\overline{\Delta}_{\mathbf{d}},\underline{\Delta}_{\mathbf{d}}}^{(X,Y)} \subseteq \mathcal{R}_{\mathbf{d}}^{(X,Y)}.$$

First, we consider the case where $0 < \underline{\Delta}_{\mathbf{d}} \le \overline{\Delta}_{\mathbf{d}} < 1$. From this assumption, for sufficiently small $\delta \in (0,1)$ and large $n > 0$, it holds that $\delta n < d_n < (1-\delta)n$. Thus, we have $d_n \to \infty$ and $n - d_n \to \infty$. By setting $\gamma_{1,1} = H(X) + \epsilon$, $\gamma_{1,2} = H(X|U) + \epsilon$, and $\gamma_2 = I(Y;U) + \epsilon$ in Corollary 2, we have for sufficiently large $n > 0$,

$$\begin{aligned}
\varepsilon_{\hat{\mathbf{X}}\hat{\mathbf{Y}}}^{(n)} &\overset{\text{(a)}}{\le} e^{-d_n\psi_{i(X)}(H(X)+\epsilon/2)} + e^{-(n-d_n)\psi_{i(X|U)}(H(X|U)+\epsilon)} \\
&\quad + e^{-(n-d_n)\psi_{i(Y;U)}(I(Y;U)+\epsilon)} \\
&\quad + e^{-nR_n^{(1)}+d_nH(X)+(n-d_n)H(X|U)+n\epsilon} \\
&\quad + \frac{1}{2}\sqrt{e^{-nR_n^{(2)}+(n-d_n)(I(Y;U)+\epsilon)}} \\
&\overset{\text{(b)}}{\le} e^{-d_n\psi_{i(X)}(H(X)+\epsilon/2)} + e^{-(n-d_n)\psi_{i(X|U)}(H(X|U)+\epsilon)} \\
&\quad + e^{-(n-d_n)\psi_{i(Y;U)}(I(Y;U)+\epsilon)} \\
&\quad + e^{-n(R_1-H(X|U)-\frac{d_n}{n}I(X;U)-\epsilon)} \\
&\quad + \frac{1}{2}\sqrt{e^{-n(R_2-\frac{n-d_n}{n}I(Y;U)-\epsilon)}} \\
&\overset{\text{(c)}}{\le} e^{-d_n\psi_{i(X)}(H(X)+\epsilon/2)} + e^{-(n-d_n)\psi_{i(X|U)}(H(X|U)+\epsilon)} \\
&\quad + e^{-(n-d_n)\psi_{i(Y;U)}(I(Y;U)+\epsilon)} \\
&\quad + e^{-n(R_1-H(X|U)-\overline{\Delta}_{\mathbf{d}}I(X;U)-\epsilon-\epsilon\log|\mathcal{X}|)} \\
&\quad + \frac{1}{2}\sqrt{e^{-n(R_2-(1-\underline{\Delta}_{\mathbf{d}})I(Y;U)-\epsilon-\epsilon\log|\mathcal{Y}|)}} \\
&= e^{-d_n\psi_{i(X)}(H(X)+\epsilon/2)} + e^{-(n-d_n)\psi_{i(X|U)}(H(X|U)+\epsilon)} \\
&\quad + e^{-(n-d_n)\psi_{i(Y;U)}(I(Y;U)+\epsilon)} + e^{-n\epsilon} + \frac{1}{2}\sqrt{e^{-n\epsilon}},
\end{aligned}$$

where (a) comes from the fact that $\lim_{n\to\infty}\frac{\log|\mathcal{D}_n|}{d_n} = 0$, (b) follows since $R_n^{(i)} \ge R_i$, and (c) comes from (13). Since $\mathrm{E}[i(X)] = H(X)$, $\mathrm{E}[i(X|U)] = H(X|U)$, and $\mathrm{E}[i(Y;U)] = I(Y;U)$, we have $\psi_{i(X)}(H(X)+\epsilon/2) > 0$, $\psi_{i(X|U)}(H(X|U)+\epsilon) > 0$, and $\psi_{i(Y;U)}(I(Y;U)+\epsilon) > 0$ according to Lemma 2. Thus, the error probability $\varepsilon_{\hat{\mathbf{X}}\hat{\mathbf{Y}}}^{(n)}$ vanishes exponentially. According to Lemma 1 and noticing that $\lim_{n\to\infty}\frac{\log(n+1)}{d_n} = 0$ and $\lim_{n\to\infty}\frac{\log(n+1)}{n-d_n} = 0$, the error probability $\max_{d\in\mathcal{D}_n}\varepsilon_{\mathbf{X}\mathbf{Y}_{(d)}}^{(n)}$ with the same code also vanishes. Thus by noticing that $\limsup_{n\to\infty} R_n^{(i)} = R_i$ for all $i \in \{1,2\}$, (R_1, R_2) is achievable.

Finally, we consider the case where $\overline{\Delta}_{\mathbf{d}} = 0$. By setting that $\gamma_1 = H(X|U) + \epsilon$ and $\gamma_2 = I(Y;U) + \epsilon$, and choosing $\delta > 0$ so that $\psi_{i(X|U)}(\gamma_1) - \delta > 0$ and $\psi_{i(Y;U)}(\gamma_2) - \delta > 0$ in Corollary 3, we have for sufficiently large $n > 0$,

$$\begin{aligned}
\varepsilon_{\hat{\mathbf{X}}\hat{\mathbf{Y}}}^{(n)} &\le e^{-n(\psi_{i(X|U)}(H(X|U)+\epsilon)-\delta)} + e^{-n(\psi_{i(Y;U)}(I(Y;U)+\epsilon)-\delta)} \\
&\quad + e^{-n(R_n^{(1)}-H(X|U)-\epsilon)} + \frac{1}{2}\sqrt{e^{-n(R_n^{(2)}-I(Y;U)-\epsilon)}} \\
&\le e^{-n(\psi_{i(X|U)}(H(X|U)+\epsilon)-\delta)} + e^{-n(\psi_{i(Y;U)}(I(Y;U)+\epsilon)-\delta)} \\
&\quad + e^{-n(\epsilon+\epsilon\log|\mathcal{X}|)} + \frac{1}{2}\sqrt{e^{-n(\epsilon+\epsilon\log|\mathcal{Y}|)}}.
\end{aligned}$$

Thus, the error probability vanishes exponentially. According to Lemma 1 and noticing that $\limsup_{n\to\infty} R_n^{(i)} = R_i$ for all $i \in \{1,2\}$, this implies that (R_1, R_2) is also achievable in this case. □

6. Conclusion

In this paper, we have considered the WAK coding system with delayed side information. We have given inner and outer bounds on the achievable rate region for a DMS. These bounds coincide with each other when the maximum of delays for the block length converges to a constant. We have clarified that the achievable rate region does not always coincide with that for the case without delay.

Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Numbers JP15K15935 and JP19K04368, and by JSPS and NUS under the Japan-Singapore Research Cooperative Program.

References

[1] T. Matsuta and T. Uyematsu, "Achievable rate regions for source coding with delayed partial side information," Proc. Int. Symp. on Inf. Theory and its Apps., pp.369–373, Oct. 2018.

[2] 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.

[3] R. 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, Nov. 1975.
[4] T. Matsuta and T. Uyematsu, "On the Wyner-Ziv source coding problem with unknown delay," IEICE Trans. Fundamentals, vol.E97-A, no.12, pp.2288–2299, Dec. 2014.
[5] T. Matsuta and T. Uyematsu, "Achievable rate regions for asynchronous Slepian-Wolf coding systems," Proc. 2015 IEEE Inf. Theory Workshop, pp.312–316, Oct. 2015.
[6] R.G. Gallager, Information Theory and Reliable Communication, Wiley, New York, 1968.
[7] R.G. Gallager, "Source coding with side information and universal coding," unpublished manuscript, Sept. 1976.
[8] S. Boucheron, G. Lugosi, and P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence, Oxford University Press, 2013.
[9] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," 2012–2017, Lecture Notes for 6.441 (MIT), ECE563 (UIUC), and STAT664 (Yale).
[10] 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.
[11] F.M.J. Willems, "Totally asynchronous Slepian-Wolf data compression," IEEE Trans. Inf. Theory, vol.34, no.1, pp.35–44, Jan. 1988.
[12] N. Oki and Y. Oohama, "Coding for the asynchronous Slepian-Wolf data compression," Proc. 20th Symp. on Inf. Theory and its Apps. (SITA97), pp.89–92, Dec. 1997 (in Japanese).
[13] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.
[14] A. El Gamal and Y.H. Kim, Network Information Theory, Cambridge University Press, 2011.
[15] T.M. Cover and J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.

Appendix A: Some Properties of $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$

First of all, we give the next lemma.

Lemma 3. For any $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$, we have

$$\hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) = \hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}),$$

where

$$\hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) = \left\{(R_1, R_2) : f_{\alpha,\beta}(R_1, R_2) \in \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\},$$

and $f_{\alpha,\beta}(R_1, R_2) = \left(\frac{R_1-\alpha H(X)}{1-\alpha}, \frac{R_2}{1-\beta}\right)$.

Proof. Let $f_\alpha(R_1) = \frac{R_1-\alpha H(X)}{1-\alpha}$ and $f_\beta(R_2) = \frac{R_2}{1-\beta}$. We note that $f_{\alpha,\beta}(R_1, R_2) = (f_\alpha(R_1), f_\beta(R_2))$.

For any $(R_1, R_2) \in \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y})$, it holds that

$$R_1 \geq H(X|U) + \alpha I(X; U),$$
$$R_2 \geq (1-\beta)I(Y; U).$$

Thus, we have

$$f_\alpha(R_1) \geq \frac{H(X|U) + \alpha I(X; U) - \alpha H(X)}{1-\alpha} = H(X|U),$$
$$f_\beta(R_2) \geq \frac{(1-\beta)I(Y; U)}{1-\beta} = I(Y; U).$$

This means that

$$f_{\alpha,\beta}(R_1, R_2) = (f_\alpha(R_1), f_\beta(R_2)) \in \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y}),$$

and

$$\hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) \subseteq \hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}). \quad \text{(A· 1)}$$

On the other hand, for any $(R_1, R_2) \in \hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y})$, it holds that

$$H(X|U) \leq f_\alpha(R_1) = \frac{R_1 - \alpha H(X)}{1-\alpha},$$
$$I(Y; U) \leq f_\beta(R_2) = \frac{R_2}{1-\beta}.$$

Thus, we have

$$R_1 \geq (1-\alpha)H(X|U) + \alpha H(X) = H(X|U) + \alpha I(X; U),$$
$$R_2 \geq (1-\beta)I(Y; U).$$

This means that

$$\hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) \subseteq \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}). \quad \text{(A· 2)}$$

Due to (A· 1) and (A· 2), we have the lemma. □

According to this lemma, we have

$$\begin{aligned}
&\mathcal{A}_{\alpha,\beta}^{(X,Y)} \\
&= \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) \\
&= \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) \\
&= \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \left\{(R_1, R_2) : f_{\alpha,\beta}(R_1, R_2) \in \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\} \\
&= \left\{(R_1, R_2) : f_{\alpha,\beta}(R_1, R_2) \in \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\} \quad \text{(A· 3)} \\
&= \left\{(R_1, R_2) : f_{\alpha,\beta}(R_1, R_2) \in \mathcal{A}_{0,0}^{(X,Y)}\right\}. \quad \text{(A· 4)}
\end{aligned}$$

We consider an arbitrary convergent sequence $(R_1^{(k)}, R_2^{(k)}) \to (R_1, R_2)$ $(k \to \infty)$, where $(R_1^{(k)}, R_2^{(k)}) \in \mathcal{A}_{\alpha,\beta}^{(X,Y)}$. Here, according to (A· 4), it holds that $f_{\alpha,\beta}(R_1^{(k)}, R_2^{(k)}) \in \mathcal{A}_{0,0}^{(X,Y)}$. Since $f_{\alpha,\beta}$ is a continuous function, we have $f_{\alpha,\beta}(R_1^{(k)}, R_2^{(k)}) \to f_{\alpha,\beta}(R_1, R_2)$. Since $\mathcal{A}_{0,0}^{(X,Y)}$ is a closed set, this implies that $f_{\alpha,\beta}(R_1, R_2) \in \mathcal{A}_{0,0}^{(X,Y)}$. Thus, we have $(R_1, R_2) \in \mathcal{A}_{\alpha,\beta}^{(X,Y)}$ due to (A· 4). This implies that $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$ is a closed set.

We assume that $(R_1, R_2) \in \mathcal{A}_{\alpha,\beta}^{(X,Y)}$ and $(r_1, r_2) \in \mathcal{A}_{\alpha,\beta}^{(X,Y)}$. By recalling that $\mathcal{A}_{0,0}^{(X,Y)}$ is a convex set and noticing that $f_{\alpha,\beta}(R_1, R_2) \in \mathcal{A}_{0,0}^{(X,Y)}$ and $f_{\alpha,\beta}(r_1, r_2) \in \mathcal{A}_{0,0}^{(X,Y)}$, we have $\lambda f_{\alpha,\beta}(R_1, R_2) + (1-\lambda) f_{\alpha,\beta}(r_1, r_2) \in \mathcal{A}_{0,0}^{(X,Y)}$. Since $f_\alpha(R_1)$ and $f_\beta(R_2)$ are linear functions, we have, for any $\lambda \in [0, 1]$

$$\lambda f_{\alpha,\beta}(R_1,R_2)+(1-\lambda)f_{\alpha,\beta}(r_1,r_2) = f_{\alpha,\beta}(\lambda R_1+(1-\lambda)r_1, \lambda R_2+(1-\lambda)r_2).$$

Thus, we have $f_{\alpha,\beta}(\lambda R_1+(1-\lambda)r_1, \lambda R_2+(1-\lambda)r_2) \in \mathcal{A}_{0,0}^{(X,Y)}$. This means that $(\lambda R_1+(1-\lambda)r_1, \lambda R_2+(1-\lambda)r_2) \in \mathcal{A}_{\alpha,\beta}^{(X,Y)}$. Thus, $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$ is a convex set.

On the other hand, again according to Lemma 3, we have

$$\begin{aligned}
&\mathcal{A}_{\alpha,\beta}^{(X,Y)} \\
&= \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}) \qquad \text{(A·5)}\\
&\overset{\text{(a)}}{=} \left\{(R_1,R_2): f_{\alpha,\beta}(R_1,R_2) \in \bigcup_{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y})} \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\}\\
&\overset{\text{(b)}}{=} \left\{(R_1,R_2): f_{\alpha,\beta}(R_1,R_2) \in \bigcup_{\substack{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y}):\\ |\mathcal{U}|=|\mathcal{Y}|+1}} \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\}\\
&= \bigcup_{\substack{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y}):\\ |\mathcal{U}|=|\mathcal{Y}|+1}} \left\{(R_1,R_2): f_{\alpha,\beta}(R_1,R_2) \in \hat{\mathcal{A}}_{0,0}^{(X,Y)}(P_{U|Y})\right\}\\
&= \bigcup_{\substack{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y}):\\ |\mathcal{U}|=|\mathcal{Y}|+1}} \hat{\mathcal{B}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y})\\
&= \bigcup_{\substack{P_{U|Y}\in\mathcal{P}(\mathcal{U}|\mathcal{Y}):\\ |\mathcal{U}|=|\mathcal{Y}|+1}} \hat{\mathcal{A}}_{\alpha,\beta}^{(X,Y)}(P_{U|Y}), \qquad \text{(A·6)}
\end{aligned}$$

where (a) comes from the right-hand side of (A· 3) and (b) comes from the fact that $\mathcal{A}_{0,0}^{(X,Y)}$ will be unchanged if we only consider $P_{U|Y} \in \mathcal{P}(\mathcal{U}|\mathcal{Y})$ such that $|\mathcal{U}| = |\mathcal{Y}| + 1$ (cf. [14, Theorem 10.2]). Thus, $\mathcal{A}_{\alpha,\beta}^{(X,Y)}$ is also unchanged, which means that the right-hand side of (A· 5) is equal to the right-hand side of (A· 6).

Appendix B: The Minimum Rate on One Side

In this appendix, we show that $R_{1,\Delta_{\mathbf{d}}} = H(X|Y) + \Delta_{\mathbf{d}} I(X;Y)$.

By the definition of $\mathcal{A}_{\Delta_{\mathbf{d}},\Delta_{\mathbf{d}}}^{(X,Y)}$, for any $R_1 \in \{R_1 : \exists (R_1,R_2) \in \mathcal{A}_{\Delta_{\mathbf{d}},\Delta_{\mathbf{d}}}^{(X,Y)}\}$, there exists an RV U such that $X-Y-U$ and

$$\begin{aligned}
R_1 &\geq H(X|U) + \Delta_{\mathbf{d}} I(X;U)\\
&= (1-\Delta_{\mathbf{d}})H(X|U) + \Delta_{\mathbf{d}} H(X)\\
&\geq (1-\Delta_{\mathbf{d}})H(X|Y) + \Delta_{\mathbf{d}} H(X)\\
&= H(X|Y) + \Delta_{\mathbf{d}} I(X;Y).
\end{aligned}$$

Thus, we have

$$R_{1,\Delta_{\mathbf{d}}} \geq H(X|Y) + \Delta_{\mathbf{d}} I(X;Y). \qquad \text{(A·7)}$$

On the other hand, by noticing that $X-Y-Y$, we have

$$(H(X|Y) + \Delta_{\mathbf{d}} I(X;Y), (1-\Delta_{\mathbf{d}})I(Y;Y)) \in \hat{\mathcal{A}}_{\Delta_{\mathbf{d}},\Delta_{\mathbf{d}}}^{(X,Y)}(P_{Y|Y}) \subseteq \mathcal{A}_{\Delta_{\mathbf{d}},\Delta_{\mathbf{d}}}^{(X,Y)}.$$

This implies that

$$R_{1,\Delta_{\mathbf{d}}} \leq H(X|Y) + \Delta_{\mathbf{d}} I(X;Y). \qquad \text{(A·8)}$$

The inequalities (A· 7) and (A· 8) show that $R_{1,\Delta_{\mathbf{d}}} = H(X|Y) + \Delta_{\mathbf{d}} I(X;Y)$.

Appendix C: Precise Derivation of (18)

We have

$$\begin{aligned}
&P_{\tilde{X}^n\tilde{U}^n}(x^n,u^n)\\
&= \sum_{y^n} P_{\tilde{X}^n\tilde{Y}^n\tilde{U}^n}(x^n,y^n,u^n)\\
&= \sum_{y^n}\sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^{d_n}(u^{d_n}) P_{U|Y}^{n-d_n}(u_{d_n+1}^n|y_{d_n+1}^n)\\
&\quad \times P_Y^d(y_1^d) P_{XY}^{n-d}(x_1^{n-d}, y_{d+1}^n) P_X^d(x_{n-d+1}^n)\\
&= \sum_{y_{d+1}^n}\sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^{d_n}(u^{d_n}) P_{U|Y}^{n-d_n}(u_{d_n+1}^n|y_{d_n+1}^n)\\
&\quad \times P_{XY}^{n-d}(x_1^{n-d}, y_{d+1}^n) P_X^d(x_{n-d+1}^n)\\
&= \sum_{y_{d+1}^n}\sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^{d_n}(u^{d_n}) P_{U|Y}^{n-d_n}(u_{d_n+1}^n|y_{d_n+1}^n)\\
&\quad \times P_{XY}^{d_n-d}(x_1^{d_n-d}, y_{d+1}^{d_n}) P_{XY}^{n-d_n}(x_{d_n-d+1}^{n-d}, y_{d_n+1}^n)\\
&\quad \times P_X^d(x_{n-d+1}^n)\\
&\overset{\text{(a)}}{=} \sum_{y_{d+1}^n}\sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^{d_n}(u^{d_n}) P_{XY}^{d_n-d}(x_1^{d_n-d}, y_{d+1}^{d_n})\\
&\quad \times P_{XYU}^{n-d_n}(x_{d_n-d+1}^{n-d}, y_{d_n+1}^n, u_{d_n+1}^n) P_X^d(x_{n-d+1}^n)\\
&= \sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^{d_n}(u^{d_n}) P_X^{d_n-d}(x_1^{d_n-d})\\
&\quad \times P_{XU}^{n-d_n}(x_{d_n-d+1}^{n-d}, u_{d_n+1}^n) P_X^d(x_{n-d+1}^n)\\
&= \sum_{d\in\mathcal{D}_n} \frac{1}{|\mathcal{D}_n|} P_U^n(u^n) P_X^{d_n-d}(x_1^{d_n-d})\\
&\quad \times P_{X|U}^{n-d_n}(x_{d_n-d+1}^{n-d}|u_{d_n+1}^n) P_X^d(x_{n-d+1}^n),
\end{aligned}$$

where (a) follows since $X-Y-U$.

Tetsunao Matsuta was born in Fukui, Japan, on March 17, 1985. He received the B.E. degree from Utsunomiya University in 2007. He received the M.E. degree, and the Dr.Eng. degree from Tokyo Institute of Technology, Japan, in 2009, 2012, respectively. Since 2012, he has been an assistant professor at Tokyo Institute of Technology, and currently he is in the Department of Information and Communications Engineering of Tokyo Institute of Technology. He received the Best Paper Award in 2016 from IEICE. His current research interests are in the area 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 Supelec (Ecole superieure d'electricite), 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.