論文 / 著書情報
Article / Book Information

Title	On the Wyner-Ziv Source Coding Problem with Unknown Delay
Authors	Tetsunao Matsuta, Tomohiko Uyematsu
Citation	IEICE Trans. Fundamentals, Vol. E97-A, No. 12, pp. 2288-2299
Pub. date	2014, 12
URL	http://search.ieice.org/
Copyright	(c) 2014 Institute of Electronics, Information and Communication Engineers

IEICE TRANS. FUNDAMENTALS, VOL.E97–A, NO.12 DECEMBER 2014

PAPER *Special Section on Information Theory and Its Applications*

On the Wyner-Ziv Source Coding Problem with Unknown Delay*

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

SUMMARY In this paper, we consider the lossy source coding problem with delayed side information at the decoder. We assume that delay is unknown but the maximum of delay is known to the encoder and the decoder, where we allow the maximum of delay to change with the block length. In this coding problem, we show an upper bound and a lower bound of the rate-distortion (RD) function, where the RD function is the infimum of rates of codes in which the distortion between the source sequence and the reproduction sequence satisfies a certain distortion level. We also clarify that the upper bound coincides with the lower bound when maximums of delay per block length converge to a constant. Then, we give a necessary and sufficient condition in which the RD function is equal to that for the case without delay. Furthermore, we give an example of a source which does not satisfy this necessary and sufficient condition.
key words: *delay, rate-distortion function, side information, source coding*

1. Introduction

In multi-terminal information theory, various coding problems have been considered and analyzed by many researchers (cf., e.g., [1], [2]). One of the most famous and important coding problems is the *Wyner-Ziv source coding problem* introduced by Wyner and Ziv [3]. This is a lossy source coding problem with side information at the decoder. For this coding problem, Wyner and Ziv [3] clarified the rate-distortion (RD) function for stationary memoryless sources, where the RD function is the infimum of rates of codes in which the distortion between the source sequence and the reproduction sequence satisfies a certain distortion level.

In the Wyner-Ziv source coding problem, it is assumed that the decoder can receive a side information symbol correlated with the source symbol simultaneously. However, in many practical situations (e.g., the case where the decoder is far away from the encoder, the case where it takes a little while to generate a sequence of side information, and the case where the path connecting side information and the decoder has some delay, etc.), the decoder can not receive correlated symbols in the beginning of the decoding. Moreover, the delay time to get a correlated symbol at the decoder may be unknown to the coding system. For example, we can consider the following situation: an observatory (encoder) on an island observes a sequence of wave heights per unit time (source sequence) caused by an earthquake near there. The observatory sends this sequence to a weather center (decoder) on a coast city distant from there. On the other hand, the center also can observe a sequence of wave heights (side information sequence) on the coast of the city. However, since the wave reaches the coast city later than it reaches the island, these heights at the same time may not be correlated. Further, the observatory and the weather center do not know the actual delay of the wave in advance, because there are many uncertainties such as the point of the earthquake center, sea breeze, shielding on the sea, etc.

In this paper, we consider the RD function for the above lossy source coding problem with delayed (noncausal) side information. Then, we show an upper bound and a lower bound of the RD function when the delay is unknown but the maximum of delay is known to the encoder and the decoder, where we allow the maximum of delay to change with the block length. In the above example, the maximum of delay depends on the distance between the island and the city, and is known to the observatory and the weather center because the distance is usually known to them. Since the wave moves during the encoding process, the maximum delay may be changed with the block length. We also clarify that the upper bound coincides with the lower bound when maximums of delay per block length converges to a constant. Then, we give a necessary and sufficient condition in which the RD function is equal to that for the case without delay. Furthermore, we give an example of a source which does not satisfy this necessary and sufficient condition.

There are some related works [4], [5] to our coding problem. In [4], a lossy source coding problem with delayed *causal* side information is considered, while delayed *noncausal* side information is considered in our setting. In [5], the lossy source coding problem with *feedforward* is considered, in which the delayed source sequence is available at the decoder. On the other hand, in our coding problem, we do not restrict delayed side information to the delayed source sequence.

The rest of this paper is organized as follows. In Sect. 2, we provide a precise definition of our coding problem. In Sect. 3, we show both upper and lower bounds of the RD function, and clarify the case where the upper bound coincides with the lower bound. We also give the necessary and sufficient condition in which the RD function is equal to that for the case without delay in this section. In Sect. 4, we

Manuscript received February 8, 2014.
Manuscript revised June 23, 2014.
[†]The authors are with the Dept. of Communications and Computer Engineering, Tokyo Institute of Technology, Tokyo, 152-8550 Japan.
*This work was presented in part at the 36th Symposium on Information Theory and Its Applications (SITA2013).
a) E-mail: tetsu@it.ce.titech.ac.jp
b) E-mail: uyematsu@ieee.org
DOI: 10.1587/transfun.E97.A.2288

deal with a source which does not satisfy the necessary and sufficient condition, and give a numerical result for the RD function of this source. In Sect. 5, we show the proof for the upper and lower bounds. In Sect. 6, we conclude the paper.

2. Preliminaries

In this section, we provide notations and a precise definition of the coding problem dealt in this paper.

We will denote an n-length sequence of symbols $(a_1, a_2, \cdots, a_n)$ by a^n, and a sequence of symbols $(a_l, a_{l+1}, \cdots, a_m)$ by a_l^m. For 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 $\mathcal{W}(\mathcal{Y}|\mathcal{X})$. The pmf of the random variable (RV) X taking a value of $\mathcal{X}$ will be denoted by $P_X \in \mathcal{P}(\mathcal{X})$, and the conditional pmf of the RV Y taking a value of $\mathcal{Y}$ given X will be denoted by $P_{Y|X} \in \mathcal{W}(\mathcal{Y}|\mathcal{X})$. For a pair of integers $i \leq j$, we will denote the set $\{i, i+1, \cdots, j\}$ as $[i : j]$. In what follows, all logarithms are taken to the base 2.

Let $\mathcal{X}$, $\mathcal{Y}$, and $\hat{\mathcal{X}}$ be arbitrary finite sets. A discrete stationary memoryless source (DMS) $(\mathbf{X}, \mathbf{Y})$ is a sequence $\{(X_i, Y_i)\}_{i=-\infty}^{\infty}$ of independent copies of a pair of correlated RVs $(X, Y) \in \mathcal{X} \times \mathcal{Y}$. For a DMS $(\mathbf{X}, \mathbf{Y})$, we consider the following coding problem with an encoder and a decoder: For an integer $M_n > 0$, the encoder $f_n : \mathcal{X}^n \to [1 : M_n]$ encodes a given source sequence $X^n \in \mathcal{X}^n$, and sends a codeword $f_n(X^n)$ to the decoder. Then, we define the rate R_n of the encoder as $R_n \triangleq \frac{1}{n} \log M_n$. The decoder $\varphi_n : \mathcal{Y}^n \times [1 : M_n] \to \hat{\mathcal{X}}^n$ outputs the reproduction sequence $\hat{X}^n = \varphi_n(Y_{1-t}^{n-t}, f_n(X^n))$ from the codeword and a delayed side information sequence $Y_{1-t}^{n-t} \in \mathcal{Y}^n$, where $t \in [0 : u_n]$ is a nonnegative integer which represents a delay, and u_n is the maximum of delay. We allow the maximum to change with the block length n, and denote the sequence $\{u_n\}_{n=1}^{\infty}$ of maximums by $\mathbf{u}$. Without loss of generality, $u_n \leq n$ because for any $t \geq n$, X^n is independent of Y_{1-t}^{n-t}. We assume that a delay t is *unknown* but the sequence $\mathbf{u}$ of maximums of delay is *known* to the encoder and the decoder.

The distortion between the source sequence X^n and the reproduction sequence $\hat{X}^n$ is measured by a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{\max}]$ as

$$d^n(X^n, \hat{X}^n) \triangleq \sum_{i=1}^{n} d(X_i, \hat{X}_i).$$

We define an (n, M_n, u_n)-code as a pair of an encoder f_n and a decoder φ_n. Now, we define the RD function.

Definition 1. For a DMS $(\mathbf{X}, \mathbf{Y})$ and a sequence $\mathbf{u}$ of maximums of delay, we call R is D-achievable when there exists a sequence of (n, M_n, u_n)-codes such that

$$\limsup_{n\to\infty} R_n \leq R,$$

and

$$\limsup_{n\to\infty} \max_{t\in[0:u_n]} \frac{1}{n} E[d^n(X^n, \hat{X}^n)|t] \leq D, \tag{1}$$

where $E[\cdot|t]$ represents the expectation when the delay is $t \in [0 : u_n]$, i.e.,

$$E[d^n(X^n, \hat{X}^n)|t] = E[d^n(X^n, \varphi_n(Y_{1-t}^{n-t}, f_n(X^n)))].$$

Then, for a nonnegative constant $D \geq 0$, the RD function $R_{\mathbf{u}}(D)$ is defined as

$$R_{\mathbf{u}}(D) \triangleq \inf\{R : R \text{ is } D\text{-achievable}\}.$$

3. Upper and Lower Bounds of the RD Function

In this section, we show upper and lower bounds of the RD function. We also clarify the case where the upper bound coincides with the lower bound, and give a necessary and sufficient condition in which the RD function is equal to that for the case without delay.

When $u_n = 0$ for all $n > 0$, i.e., a delay does not occur, it is known [3] that the RD function can be represented as

$$R_{\mathbf{u}}(D) = \min I(X; V|Y) \triangleq R_{\rm wz}(D),$$

where the minimum is taken over all conditional pmfs $P_{V|X} \in \mathcal{W}(\mathcal{V}|\mathcal{X})$ and functions $g : \mathcal{Y} \times \mathcal{V} \to \hat{\mathcal{X}}$ such that

$$|\mathcal{V}| \leq |\mathcal{X}| + 1,$$
$$Y \leftrightarrow X \leftrightarrow V,$$
$$E[d(X, g(Y, V))] \leq D,$$

where $|\cdot|$ denotes the cardinality of the set, and $Y \leftrightarrow X \leftrightarrow V$ represent that the RVs (Y, X, V) form a Markov chain in this order.

When for a DMS $(\mathbf{X}, \mathbf{Y})$, X is independent of Y, it is known [6] that the RD function can be represented as

$$R_{\mathbf{u}}(D) = \min I(X; \hat{X}) \triangleq R(D),$$

where the minimum is taken over all conditional pmfs $P_{\hat{X}|X} \in \mathcal{W}(\hat{\mathcal{X}}|\mathcal{X})$ such that $E[d(X, \hat{X})] \leq D$.

$R_{\rm wz}(D)$ and $R(D)$ have the following properties (cf., e.g., [2]):

Property 1. $R_{\rm wz}(D) \leq R(D)$ for any $D \geq 0$.

Property 2. $R_{\rm wz}(D)$ and $R(D)$ are monotone nonincreasing, convex, and continuous functions.

For a given sequence $\mathbf{u}$, we define $\underline{\Delta}_{\mathbf{u}}$ and $\overline{\Delta}_{\mathbf{u}}$ as

$$\underline{\Delta}_{\mathbf{u}} \triangleq \liminf_{n\to\infty} \frac{u_n}{n}$$

and

$$\overline{\Delta}_{\mathbf{u}} \triangleq \limsup_{n\to\infty} \frac{u_n}{n},$$

respectively. Then, we have the next two theorems which show lower and upper bounds of the RD function.

Theorem 1. For a DMS $(\mathbf{X}, \mathbf{Y})$ and a sequence $\mathbf{u}$,

$$R_{\mathbf{u}}(D) \ge \min_{(D_1,D_2)\in\underline{\mathcal{D}}_{\mathbf{u}}} \{(1-\overline{\Delta}_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \underline{\Delta}_{\mathbf{u}}R(D_2)\},$$

where

$$\underline{\mathcal{D}}_{\mathbf{u}} \triangleq \{(D_1,D_2)\in[0,d_{\max}]^2 : D \ge (1-\overline{\Delta}_{\mathbf{u}})D_1 + \underline{\Delta}_{\mathbf{u}}D_2\}.$$

Remark 1. The lower bound of the RD function in the above theorem is the minimum of a convex function over a simple convex set. Thus, if closed-forms of $R_{\mathrm{wz}}(D)$ and $R(D)$ are given, one can easily compute the bound. On the other hand, as shown in Remark 5, it also can be bounded by the following rather complex formula involving an optimization over a set of infinite sequences.

$$R_{\mathbf{u}}(D) \ge \inf_{\substack{\{(D_1^{(n)},D_2^{(n)}\}_{n=1}^{\infty}:\\ D\ge \limsup_{n\to\infty}\{(1-\Delta_n)D_1^{(n)}+\Delta_n D_2^{(n)}\}}} \limsup_{n\to\infty}\{(1-\Delta_n)R_{\mathrm{wz}}(D_1^{(n)}) + \Delta_n R(D_2^{(n)})\}, \tag{2}$$

where $\Delta_n \triangleq \frac{u_n}{n}$. According to this bound, we have

$$\begin{aligned}R_{\mathbf{u}}(D) &\overset{(a)}{\ge} \inf_{\substack{\{(D_1^{(n)},D_2^{(n)}\}_{n=1}^{\infty}:\\ D\ge \limsup_{n\to\infty}\{(1-\Delta_n)D_1^{(n)}+\Delta_n D_2^{(n)}\}}} \limsup_{n\to\infty}\{(1-\Delta_n)R_{\mathrm{wz}}(D_1^{(n)}) + \Delta_n R_{\mathrm{wz}}(D_2^{(n)})\}\\ &\overset{(b)}{\ge} \inf_{\substack{\{(D_1^{(n)},D_2^{(n)}\}_{n=1}^{\infty}:\\ D\ge \limsup_{n\to\infty}\{(1-\Delta_n)D_1^{(n)}+\Delta_n D_2^{(n)}\}}} \times \limsup_{n\to\infty}\{R_{\mathrm{wz}}((1-\Delta_n)D_1^{(n)} + \Delta_n D_2^{(n)})\}\\ &\overset{(c)}{\ge} R_{\mathrm{wz}}(D).\end{aligned} \tag{3}$$

where (a) comes from Property 1, (b) comes from Property 2, and (c) comes from the fact that $\limsup_{n\to\infty} R_{\mathrm{wz}}(a_n) \ge R_{\mathrm{wz}}(\limsup_{n\to\infty} a_n)$. This bound is tighter than the bound in Theorem 1. In fact, for $D \in [0, d_{\max}]$, $(\mathbf{X},\mathbf{Y})$ and $\mathbf{u}$ such that $R_{\mathrm{wz}}(D) > 0$, $\overline{\Delta}_{\mathbf{u}} = 1$ and $\underline{\Delta}_{\mathbf{u}} = 0$, we have $R_{\mathbf{u}}(D) \ge 0$ from the bound in Theorem 1, while we have $R_{\mathbf{u}}(D) \ge R_{\mathrm{wz}}(D) (> 0)$ from the bound (2).

Theorem 2. If a sequence $\mathbf{u}$ satisfies $\overline{\Delta}_{\mathbf{u}} = 0$ or $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$, then for a DMS $(\mathbf{X},\mathbf{Y})$,

$$R_{\mathbf{u}}(D) \le \min_{(D_1,D_2)\in\overline{\mathcal{D}}_{\mathbf{u}}} \{(1-\underline{\Delta}_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \overline{\Delta}_{\mathbf{u}}R(D_2)\},$$

where

$$\overline{\mathcal{D}}_{\mathbf{u}} \triangleq \{(D_1,D_2)\in[0,d_{\max}]^2 : D \ge (1-\underline{\Delta}_{\mathbf{u}})D_1 + \overline{\Delta}_{\mathbf{u}}D_2\}.$$

If the sequence $\mathbf{u}$ does not satisfy the above condition, it holds that $R_{\mathbf{u}}(D) \le R(D)$.

Remark 2. Just like the lower bound, the upper bound of the RD function in the above theorem is also the minimum of a convex function over a simple convex set. On the other hand, if a sequence $\mathbf{u}$ satisfies $\overline{\Delta}_{\mathbf{u}} = 0$ or $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$, it also can be bounded by the following rather complex formula as shown in Remark 7.

$$R_{\mathbf{u}}(D) \le \inf_{\substack{(D_1,D_2)\in[0,d_{\max}]^2:\\ D\ge \limsup_{n\to\infty}\{(1-\Delta_n)D_1+\Delta_n D_2\}}} \limsup_{n\to\infty}\{(1-\Delta_n)R_{\mathrm{wz}}(D_1) + \Delta_n R(D_2)\}.$$

According to this bound, we have

$$\begin{aligned}R_{\mathbf{u}}(D) &\le \limsup_{n\to\infty}\{(1-\Delta_n)R_{\mathrm{wz}}(D) + \Delta_n R(D)\}\\ &\le \limsup_{n\to\infty}\{(1-\Delta_n)R(D) + \Delta_n R(D)\}\\ &= R(D),\end{aligned} \tag{4}$$

where the second inequality comes from Property 1. This bound is also tighter than the bound in Theorem 2. In fact, for a sufficiently small $\delta > 0$, $D \in [0, d_{\max}]$, $(\mathbf{X},\mathbf{Y})$ and $\mathbf{u}$ such that $R_{\mathrm{wz}}(D) > 0$, $\overline{\Delta}_{\mathbf{u}} = 1-\delta$ and $\underline{\Delta}_{\mathbf{u}} = \delta$, the upper bound in Theorem 2 is bounded as

$$\begin{aligned}&\min_{(D_1,D_2)\in\overline{\mathcal{D}}_{\mathbf{u}}} \{(1-\underline{\Delta}_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \overline{\Delta}_{\mathbf{u}}R(D_2)\}\\ &= \min_{\substack{(D_1,D_2)\in[0,d_{\max}]^2:\\ D/(1-\delta)-D_1\ge D_2}} \{(1-\delta)R_{\mathrm{wz}}(D_1) + (1-\delta)R(D_2)\}\\ &\ge \min_{D_1\in[0,d_{\max}]} \{(1-\delta)R_{\mathrm{wz}}(D_1) + (1-\delta)R(D/(1-\delta)-D_1)\}\\ &> R(D),\end{aligned}$$

where the last inequality follows from the fact that

$$\min_{D_1\in[0,d_{\max}]} \{R_{\mathrm{wz}}(D_1) + R(D-D_1)\} > R(D),$$

since δ is sufficiently small.

We postpone the proof of Theorem 1 and Theorem 2 to Sect. 5.

Especially, when the sequence $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$, we have the next corollary.

Corollary 1. For a DMS $(\mathbf{X},\mathbf{Y})$ and a sequence $\mathbf{u}$ such that $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$, it holds that

$$R_{\mathbf{u}}(D) = \min_{(D_1,D_2)\in\mathcal{D}_{\mathbf{u}}} \{(1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}}R(D_2)\},$$

where

$$\Delta_{\mathbf{u}} \triangleq \lim_{n\to\infty}\frac{u_n}{n},$$

and

$$\mathcal{D}_{\mathbf{u}} \triangleq \{(D_1,D_2)\in[0,d_{\max}]^2 : D \ge (1-\Delta_{\mathbf{u}})D_1 + \Delta_{\mathbf{u}}D_2\}.$$

According to Corollary 1, we have the next theorem.

Theorem 3. For a constant $D \in [0, d_{\max}]$, a DMS $(\mathbf{X},\mathbf{Y})$, and a sequence $\mathbf{u}$ such that $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$, let

$$(D_1^*, D_2^*) = \operatorname*{argmin}_{(D_1,D_2)\in\mathcal{D}_{\mathbf{u}}} \{(1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}}R(D_2)\}.$$

Then,

$$R_{\mathbf{u}}(D) = R_{\mathrm{wz}}(D),$$

if and only if $(\mathbf{X}, \mathbf{Y})$ and $\mathbf{u}$ satisfy the following two conditions:

1). $\Delta_{\mathbf{u}} R(D_2^*) = \Delta_{\mathbf{u}} R_{\mathrm{wz}}(D_2^*)$.
2). $(1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*)+\Delta_{\mathbf{u}} R_{\mathrm{wz}}(D_2^*) = R_{\mathrm{wz}}((1-\Delta_{\mathbf{u}})D_1^*+\Delta_{\mathbf{u}} D_2^*)$.

To prove this theorem, we use the next lemma which shows an alternative formula of the RD function.

Lemma 1. For a constant $D \in [0, d_{\max}]$, a DMS $(\mathbf{X}, \mathbf{Y})$, and a sequence $\mathbf{u}$ such that $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$, it holds that

$$R_{\mathbf{u}}(D) = \min_{(D_1, D_2) \in \mathcal{D}'_{\mathbf{u}}} \{(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}} R(D_2)\},$$

where

$$\mathcal{D}'_{\mathbf{u}} \triangleq \{(D_1, D_2) \in [0, d_{\max}]^2 : D = (1-\Delta_{\mathbf{u}})D_1 + \Delta_{\mathbf{u}} D_2\}.$$

Proof. For any $D \in [0, d_{\max}]$ and $(D_1, D_2) \in \mathcal{D}_{\mathbf{u}}$, there exists a pair of constants $(D'_1, D'_2) \in [0, d_{\max}]^2$ such that $D'_1 \geq D_1$, $D'_2 \geq D_2$ and

$$D = (1 - \Delta_{\mathbf{u}})D'_1 + \Delta_{\mathbf{u}} D'_2.$$

Then, according to Property 2, we have

$$\begin{aligned}(1 - \Delta_{\mathbf{u}})&R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}} R(D_2)\\ &\geq (1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D'_1) + \Delta_{\mathbf{u}} R(D'_2).\end{aligned}$$

Hence, by using Corollary 1, we have

$$\begin{aligned}R_{\mathbf{u}}(D) &= \min_{(D_1, D_2) \in \mathcal{D}_{\mathbf{u}}} \{(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}} R(D_2)\}\\ &\geq \min_{\substack{(D_1, D_2) \in [0, d_{\max}]^2:\\ D = (1-\Delta_{\mathbf{u}})D_1 + \Delta_{\mathbf{u}} D_2}} \{(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}} R(D_2)\}.\end{aligned}$$

Since the inequality in the opposite direction is trivial, this completes the proof. □

Remark 3. By using this lemma, one can easily delete the variable D_2 as

$$\begin{aligned}R_{\mathbf{u}}(D) = &\min_{\max\{0, \frac{D - d_{\max}\Delta_{\mathbf{u}}}{1-\Delta_{\mathbf{u}}}\} \leq D_1 \leq \min\{d_{\max}, \frac{D}{1-\Delta_{\mathbf{u}}}\}} \{(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1)\\ &+ \Delta_{\mathbf{u}} R((D - (1 - \Delta_{\mathbf{u}})D_1)/\Delta_{\mathbf{u}})\}.\end{aligned}$$

This alternative formula is little bit complicated, but it is convenient for the numerical calculation.

Proof of Theorem 3. For an optimal pair (D_1^*, D_2^*), we have

$$\begin{aligned}R_{\mathbf{u}}(D) &= (1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}} R(D_2^*)\\ &\overset{(a)}{\geq} (1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}} R_{\mathrm{wz}}(D_2^*)\\ &\overset{(b)}{\geq} R_{\mathrm{wz}}((1 - \Delta_{\mathbf{u}})D_1^* + \Delta_{\mathbf{u}} D_2^*)\\ &\overset{(c)}{=} R_{\mathrm{wz}}(D), \qquad (5)\end{aligned}$$

where (a) comes from Property 1, (b) comes from Property 2, and (c) comes from Lemma 1. According to the above inequality,

$$R_{\mathbf{u}}(D) = R_{\mathrm{wz}}(D),$$

if and only if inequalities at (a) and (b) are equality. Since the inequality at (a) is equality if and only if $\Delta_{\mathbf{u}} R_{\mathrm{wz}}(D_2^*) = \Delta_{\mathbf{u}} R(D_2^*)$, and the inequality at (b) is equality if and only if $(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}} R_{\mathrm{wz}}(D_2^*) = R_{\mathrm{wz}}((1 - \Delta_{\mathbf{u}})D_1 + \Delta_{\mathbf{u}} D_2)$, we have the conditions in the theorem. □

A sequence $\mathbf{u}$ satisfying $u_n = o(n)$ satisfies conditions 1) and 2) because $\Delta_{\mathbf{u}} = 0$. Hence, according to Theorem 3, we have

$$R_{\mathbf{u}}(D) = R_{\mathrm{wz}}(D).$$

On the other hand, according to Theorem 3, when a DMS $(\mathbf{X}, \mathbf{Y})$ and a sequence $\mathbf{u}$ do not satisfy one or two of conditions 1) and 2), we have

$$R_{\mathbf{u}}(D) > R_{\mathrm{wz}}(D),$$

i.e., the RD function is strictly larger than that for the case without delay. In the next section, we give an example of this case.

According to Corollary 1, we also have some properties of the RD function as shown in the next theorem.

Theorem 4. For a DMS $(\mathbf{X}, \mathbf{Y})$ and a sequence $\mathbf{u}$ such that $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$, it holds that

$$R_{\mathrm{wz}}(D) \leq R_{\mathbf{u}}(D) \leq R(D), \qquad (6)$$

and for any $\mathbf{u}$ and $\mathbf{u}'$ such that $\Delta_{\mathbf{u}'} \leq \Delta_{\mathbf{u}}$, it holds that

$$R_{\mathbf{u}'}(D) \leq R_{\mathbf{u}}(D). \qquad (7)$$

Proof. The first inequality in (6) comes from (5), and the second inequality in (6) follows since

$$\begin{aligned}R_{\mathbf{u}}(D) &\leq (1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D) + \Delta_{\mathbf{u}} R(D)\\ &\leq (1 - \Delta_{\mathbf{u}})R(D) + \Delta_{\mathbf{u}} R(D)\\ &= R(D),\end{aligned}$$

where the second inequality comes from Property 1.

To prove the inequality (7), let

$$(D_1^*, D_2^*) = \operatorname*{argmin}_{(D_1, D_2) \in \mathcal{D}_{\mathbf{u}}} \{(1 - \Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \Delta_{\mathbf{u}} R(D_2)\},$$

and we consider the case where $D_1^* \leq D_2^*$. According to Lemma 1, we have

$$\begin{aligned}D &= (1 - \Delta_{\mathbf{u}})D_1^* + \Delta_{\mathbf{u}} D_2^*\\ &= (1 - \Delta_{\mathbf{u}'})D_1^* + \Delta_{\mathbf{u}'} D_2^* + (\Delta_{\mathbf{u}} - \Delta_{\mathbf{u}'})(D_2^* - D_1^*).\end{aligned}$$

Since $(\Delta_{\mathbf{u}} - \Delta_{\mathbf{u}'})(D_2^* - D_1^*) \geq 0$, there exists a constant $D'_1 \geq D_1^*$ such that

$$(1 - \Delta_{\mathbf{u}})D_1^* + \Delta_{\mathbf{u}} D_2^* = (1 - \Delta_{\mathbf{u}'})D'_1 + \Delta_{\mathbf{u}'} D_2^*.$$

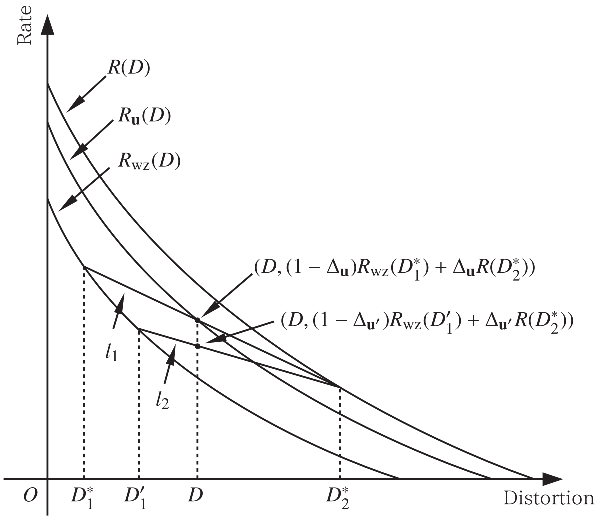

Fig. 1 Lines l_1 and l_2.

Then, by noting that $R_{\mathrm{wz}}(D)$ is a convex function, the line l_1 between the point $(D_1^*, R_{\mathrm{wz}}(D_1^*))$ and the point $(D_2^*, R(D_2^*))$ is above the line l_2 between the point $(D_1', R_{\mathrm{wz}}(D_1'))$ and the point $(D_2^*, R(D_2^*))$ on the interval $[0, D_2^*]$ (see Fig. 1). On the other hand, the point $(D, (1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}}R(D_2^*))$ is on the line l_1, and the point $(D, (1-\Delta_{\mathbf{u}'})R_{\mathrm{wz}}(D_1') + \Delta_{\mathbf{u}'}R(D_2^*))$ is on the line l_2. Hence, we have

$$\begin{aligned} R_{\mathbf{u}}(D) &= (1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}}R(D_2^*) \\ &\geq (1-\Delta_{\mathbf{u}'})R_{\mathrm{wz}}(D_1') + \Delta_{\mathbf{u}'}R(D_2^*) \\ &\geq R_{\mathbf{u}'}(D), \end{aligned}$$

where the last inequality comes from the fact that $D = (1-\Delta_{\mathbf{u}'})D_1' + \Delta_{\mathbf{u}'}D_2^*$. Since the case where $D_2^* \leq D_1^*$ can be proved in a similar way to the case where $D_1^* \leq D_2^*$, we omit the proof. □

Remark 4. Inequalities (6) hold even if $\{u_n/n\}_{n=1}^{\infty}$ does not converge to a constant as $n \to \infty$. Indeed, these inequalities are straightforward from (3) and (4).

4. Example: Doubly Symmetric Binary Source

In this section, we give an example of a source which does not satisfy the necessary and sufficient condition in Theorem 3. In other words, we give an example of the RD function which is strictly larger than that for the case without delay.

Let the distortion measure d be the Hamming distortion measure, i.e.,

$$d(x, \hat{x}) = \begin{cases} 1 & \text{if } x \neq \hat{x}, \\ 0 & \text{if } x = \hat{x}. \end{cases}$$

For $p \in [0, 1/2]$, let $(X, Y) \in \{0,1\} \times \{0,1\}$ be a pair of binary RVs such that

$$P_{XY}(0,0) = P_{XY}(1,1) = \frac{1-p}{2},$$

$$P_{XY}(0,1) = P_{XY}(1,0) = \frac{p}{2},$$

and $(\mathbf{X}, \mathbf{Y})$ be a DMS characterized by (X, Y). This type of source is called the *doubly symmetric binary source* (DSBS). Wyner and Ziv [3] evaluated the RD function for the DSBS, and showed that

$$R_{\mathrm{wz}}(D) = \begin{cases} g(D) & \text{for } 0 \leq D \leq D_c, \\ (D-p)g'(D_c) & \text{for } D_c \leq D \leq p, \\ 0 & \text{for } p \leq D, \end{cases} \tag{8}$$

where

$$\begin{aligned} g(D) &\triangleq h(p * D) - h(D), \\ h(x) &\triangleq -x \log x - (1-x)\log(1-x), \\ x * y &\triangleq (1-x)y + x(1-y), \end{aligned}$$

$g'(D)$ is the derivative of $g(D)$ with respect to D, and D_c is the solution of the equation

$$\frac{g(D_c)}{(D_c - p)} = g'(D_c).$$

On the other hand, since $P_X(0) = 1/2$, we have (see [1, Theorem 10.3.1])

$$R(D) = h(1/2) - h(D) = 1 - h(D). \tag{9}$$

In this case, for any $p \in [0, 1/2)$ and any $D \in [0, 1/2)$, we have

$$R(D) > R_{\mathrm{wz}}(D).$$

Thus, for any $p \in [0, 1/2)$, any $D \in [0, (1-\Delta_{\mathbf{u}})p + \Delta_{\mathbf{u}}/2)$, and any sequence $\mathbf{u}$ such that $\Delta_{\mathbf{u}} > 0$ and $D_2^* \neq 1/2$, the condition 1) in Theorem 3 is not satisfied.

On the other hand, if $D_2^* = 1/2$, D_1^* must satisfy that $D_1^* < p$ and $D_1^* < D$. Since $p < 1/2$, $D < 1/2$, and $R_{\mathrm{wz}}(D) = 0$ for all $D \geq p$, the line l_1 between the point $(D_1^*, R_{\mathrm{wz}}(D_1^*))$ and the point $(D_2^*, R_{\mathrm{wz}}(D_2^*)) = (1/2, 0)$ is strictly above the line l_2 between the point $(D_1^*, R_{\mathrm{wz}}(D_1^*))$ and the point $(D, R_{\mathrm{wz}}(D))$ on the interval $(D_1^*, 1/2]$. On the other hand, the point $(D, (1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}}R_{\mathrm{wz}}(D_2^*))$ is on the line l_1, and the point $(D, R_{\mathrm{wz}}(D))$ is on the line l_2. Thus, we have

$$\begin{aligned} &(1-\Delta_{\mathbf{u}})R_{\mathrm{wz}}(D_1^*) + \Delta_{\mathbf{u}}R_{\mathrm{wz}}(D_2^*) \\ &\qquad > R_{\mathrm{wz}}(D) \\ &\qquad = R_{\mathrm{wz}}((1-\Delta_{\mathbf{u}})D_1^* + \Delta_{\mathbf{u}}D_2^*). \end{aligned}$$

Thus, for any $p \in [0, 1/2)$, any $D \in [0, (1-\Delta_{\mathbf{u}})p + \Delta_{\mathbf{u}}/2)$, and any sequence $\mathbf{u}$ such that $\Delta_{\mathbf{u}} > 0$ and $D_2^* = 1/2$, the condition 2) in Theorem 3 is not satisfied.

Consequently, for any $p \in [0, 1/2)$, any $D \in [0, (1-\Delta_{\mathbf{u}})p + \Delta_{\mathbf{u}}/2)$, and any sequence $\mathbf{u}$ such that $\Delta_{\mathbf{u}} > 0$, we have

$$R_{\mathbf{u}}(D) > R_{\mathrm{wz}}(D).$$

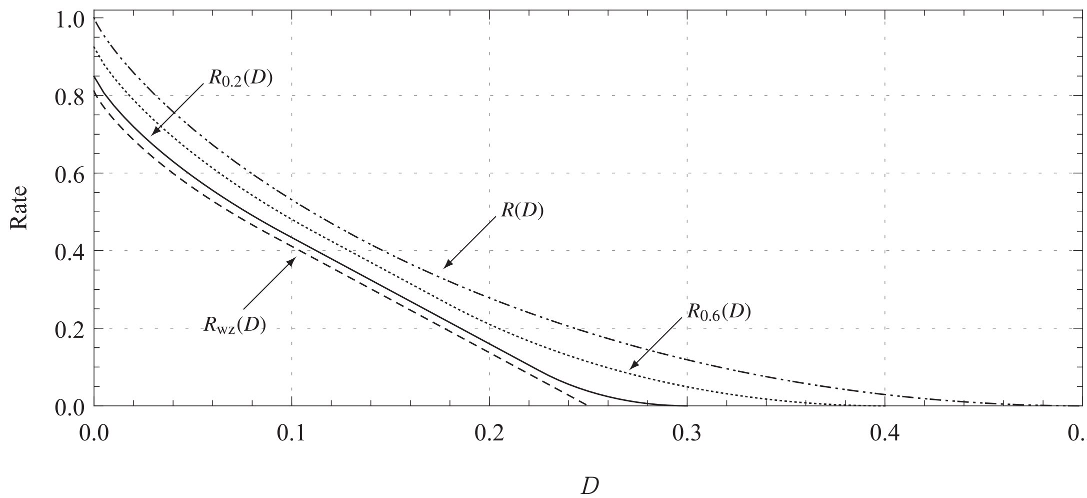

Fig. 2 RD functions for the DSBS.

On the other hand, according to (8), (9), and Corollary 1, for any $D \geq (1-\Delta_{\mathbf{u}})p + \Delta_{\mathbf{u}}/2$, we have

$$R_{\mathbf{u}}(D) = 0.$$

These facts also be confirmed by the numerical calculation as shown in Fig. 2. In Fig. 2, we set $p = 0.25$, and plot RD functions $R(D)$, $R_{\mathrm{wz}}(D)$, $R_{0.2}(D)$, and $R_{0.6}(D)$, where $R_{0.2}(D)$ and $R_{0.6}(D)$ are the RD functions $R_{\mathbf{u}}(D)$ in the case where $\Delta_{\mathbf{u}} = 0.2$ and $\Delta_{\mathbf{u}} = 0.6$, respectively. It can be seen that $R_{\mathrm{wz}}(D) < R_{0.6}(D)$ on the interval $[0, 0.4)$, and $R_{\mathrm{wz}}(D) < R_{0.2}(D)$ on the interval $[0, 0.3)$ as shown in the above. It also can be seen that $R_{0.2}(D) \leq R_{0.6}(D) \leq R(D)$ as shown in Theorem 4.

5. Proof of Theorems

In this section, we prove Theorem 1 and Theorem 2.

Proof of Theorem 1. For any (n, M_n, u_n)-code and any $t \in [0 : u_n]$, by letting $F_n = f_n(X^n)$, we have

$$\begin{aligned}\log M_n \geq& H(F_n)\\ \geq& H(F_n|Y_{1-t}^{n-t})\\ =& I(X^n; F_n|Y_{1-t}^{n-t})\\ =& \sum_{i=1}^{n-t} I(X_i; F_n|Y_{1-t}^{n-t}, X^{i-1})\\ &+ \sum_{i=n-t+1}^{n} I(X_i; F_n|Y_{1-t}^{n-t}, X^{i-1})\\ \stackrel{(a)}{=}& \sum_{i=1}^{n-t} I(X_i; F_n, Y_{1-t}^{i-1}, Y_{i+1}^{n-t}, X^{i-1}|Y_i)\\ &+ \sum_{i=n-t+1}^{n} I(X_i; F_n|Y_{1-t}^{n-t}, X^{i-1}),\end{aligned} \tag{10}$$

where (a) follows since (X_i, Y_i) is independent of $(Y_{1-t}^{i-1}, Y_{i+1}^{n-t}, X^{i-1})$. The first term in the right-hand side (RHS) of (10) can be lower bounded as

$$\begin{aligned}&\sum_{i=1}^{n-t} I(X_i; F_n, Y_{1-t}^{i-1}, Y_{i+1}^{n-t}, X^{i-1}|Y_i)\\ &\geq \sum_{i=1}^{n-t} I(X_i; V_i|Y_i)\\ &\stackrel{(a)}{\geq} \sum_{i=1}^{n-t} R_{\mathrm{wz}}(E[d(X_i, \hat{X}_i)|t])\\ &\stackrel{(b)}{\geq} (n-t)R_{\mathrm{wz}}\Big(\frac{1}{n-t}\sum_{i=1}^{n-t} E[d(X_i, \hat{X}_i)|t]\Big),\end{aligned} \tag{11}$$

where $V_i = (F_n, Y_{1-t}^{i-1}, Y_{i+1}^{n-t})$, $\hat{X}_i$ is the i-th element of the sequence $\hat{X}^n = \varphi_n(Y_{1-t}^{n-t}, F_n)$, (a) comes from the definition of $R_{\mathrm{wz}}(D)$ and the fact that $Y_i \leftrightarrow X_i \leftrightarrow V_i$ (because F_n is a function of X^n and not Y^n), and (b) follows from the convexity of the function $R_{\mathrm{wz}}(D)$.

The second term in the RHS of (10) can be lower bounded as

$$\begin{aligned}&\sum_{i=n-t+1}^{n} I(X_i; F_n|Y_{1-t}^{n-t}, X^{i-1})\\ &\stackrel{(a)}{=} \sum_{i=n-t+1}^{n} I(X_i; F_n, Y_{1-t}^{n-t}, X^{i-1})\\ &\geq \sum_{i=n-t+1}^{n} I(X_i; F_n, Y_{1-t}^{n-t})\\ &\stackrel{(b)}{\geq} \sum_{i=n-t+1}^{n} I(X_i; \hat{X}_i)\\ &\stackrel{(c)}{\geq} \sum_{i=n-t+1}^{n} R(E[d(X_i, \hat{X}_i)|t])\\ &\stackrel{(d)}{\geq} tR\Big(\frac{1}{t}\sum_{i=n-t+1}^{n} E[d(X_i, \hat{X}_i)|t]\Big),\end{aligned} \tag{12}$$

where (a) follows since X_i is independent of (Y_{1-t}^{n-t}, X^{i-1}) for any $i \geq n-t+1$, (b) follows from $X_i \leftrightarrow (F_n, Y_{1-t}^{n-t}) \leftrightarrow \hat{X}_i$ and the data processing inequality [7, Lemma 3.11], (c) comes from the definition of $R(D)$, and (d) follows from the convexity of the function $R(D)$.

On the other hand, if a sequence of (n, M_n, u_n)-codes satisfies (1), then for any $\gamma > 0$ and all sufficiently large $n > 0$, we have

$$\begin{aligned} D+\gamma &\geq \max_{t\in[0:u_n]} \frac{1}{n} E[d_n(X^n, \hat{X}^n)|t] \\ &\geq \frac{1}{n} E[d_n(X^n, \hat{X}^n)|u_n] \\ &= \frac{n-u_n}{n}\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\quad + \frac{u_n}{n}\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big) \qquad (13) \\ &\geq \Big(\liminf_{n\to\infty}\frac{n-u_n}{n} - \gamma\Big)\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\quad + \Big(\liminf_{n\to\infty}\frac{u_n}{n} - \gamma\Big)\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\geq (1-\overline{\Delta}_{\mathbf{u}})\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\quad + \underline{\Delta}_{\mathbf{u}}\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big) - 2\gamma d_{\max}. \qquad (14) \end{aligned}$$

Hence from (10)–(12), by letting $t = u_n$, for any D-achievable rate R, any $\gamma > 0$, and all sufficiently large $n > 0$, we have

$$\begin{aligned} R+\gamma &\geq \frac{1}{n}\log M_n \\ &\geq \frac{n-u_n}{n} R_{\mathrm{wz}}\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\quad + \frac{u_n}{n} R\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big) \qquad (15) \\ &\geq (1-\overline{\Delta}_{\mathbf{u}}) R_{\mathrm{wz}}\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &\quad + \underline{\Delta}_{\mathbf{u}} R\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big) - 2\gamma\log|\mathcal{X}| \\ &\overset{(a)}{\geq} \min_{\{(D_1,D_2)\in[0,d_{\max}]^2 : D+\gamma\geq(1-\overline{\Delta}_{\mathbf{u}})D_1+\underline{\Delta}_{\mathbf{u}}D_2-2\gamma d_{\max}\}} \\ &\quad \times\{(1-\overline{\Delta}_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \underline{\Delta}_{\mathbf{u}} R(D_2)\} - 2\gamma\log|\mathcal{X}|, \qquad (16) \end{aligned}$$

where (a) comes from (14), and assuming that

$$D_1 = \frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n],$$

$$D_2 = \frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n].$$

Since (16) holds for any D-achievable rate R and any $\gamma > 0$, by noting that $R_{\mathrm{wz}}(D)$ and $R(D)$ are continuous functions, we have Theorem 1. □

Remark 5. According to (1) and (13), we have

$$\begin{aligned} D \geq \limsup_{n\to\infty}\Big\{&(1-\Delta_n)\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &+ \Delta_n\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big)\Big\}. \end{aligned}$$

Further, according to (15), we have

$$\begin{aligned} R \geq \limsup_{n\to\infty}\Big\{&(1-\Delta_n)R_{\mathrm{wz}}\Big(\frac{1}{n-u_n}\sum_{i=1}^{n-u_n} E[d(X_i,\hat{X}_i)|u_n]\Big) \\ &+ \Delta_n R\Big(\frac{1}{u_n}\sum_{i=n-u_n+1}^{n} E[d(X_i,\hat{X}_i)|u_n]\Big)\Big\}. \end{aligned}$$

Thus, the same argument used to derive the inequality (16), we have the following bound stated in Remark 1,

$$\begin{aligned} R_{\mathbf{u}}(D) \geq &\inf_{\substack{\{(D_1^{(n)},D_2^{(n)})\}_{n=1}^{\infty}: \\ D\geq\limsup_{n\to\infty}\{(1-\Delta_n)D_1^{(n)}+\Delta_n D_2^{(n)}\}}} \limsup_{n\to\infty}\{(1-\Delta_n)R_{\mathrm{wz}}(D_1^{(n)}) \\ &+ \Delta_n R(D_2^{(n)})\}. \end{aligned}$$

In order to prove Theorem 2, we introduce the *typical set* and the *conditionally typical set* as defined below.

Definition 2. For a pair of RVs $(X, Y) \in \mathcal{X}\times\mathcal{Y}$, a constant $\epsilon > 0$, and a sequence $x^k \in \mathcal{X}^k$, we define

$$\begin{aligned} \mathcal{T}_{\epsilon}^{(k)}(X) &\triangleq \{x^k \in \mathcal{X}^k : |\pi(x|x^k) - P_X(x)| \leq \epsilon P_X(x),\ \forall x\in\mathcal{X}\}, \\ \mathcal{T}_{\epsilon}^{(k)}(Y|x^k) &\triangleq \{y^k \in \mathcal{Y}^k : |\pi(x,y|x^k,y^k) - \pi(x|x^k)P_{Y|X}(y|x)| \\ &\qquad \leq \epsilon P_{Y|X}(y|x),\ \forall (x,y)\in\mathcal{X}\times\mathcal{Y}\}, \end{aligned}$$

where $\pi(a|a^k) \triangleq |\{i\in[1:k] : a_i = a\}|/k$.

These sets have some well-known properties shown as the following lemmas.

Lemma 2 ([7, Lemma 2.12])**.** For any pair of RVs $(X, Y) \in \mathcal{X}\times\mathcal{Y}$, any $\epsilon > 0$, any $x^k \in \mathcal{X}^k$, and the sequence of i.i.d. RVs $(X^k, Y^k) \in \mathcal{X}^k\times\mathcal{Y}^k$ such that $(X_i, Y_i) \sim P_{XY}$, there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \to 0$ such that

$$\Pr\{X^k \notin \mathcal{T}_{\epsilon}^{(k)}(X)\} \leq 2^{-k\delta(\epsilon)}, \qquad (17)$$

$$\Pr\{Y^k \notin \mathcal{T}_{\epsilon}^{(k)}(Y|x^k)|X^k = x^k\} \leq 2^{-k\delta(\epsilon)}. \qquad (18)$$

Lemma 3 ([2, Lemma 24.2])**.** For a sufficiently small $\epsilon > 0$, any pair of RVs $(X, Y) \in \mathcal{X}\times\mathcal{Y}$, and the sequence of i.i.d. RVs $\{(X_i, Y_i)\}_{i=-\infty}^{\infty}$ such that $(X_i, Y_i) \sim P_{XY}$, there exists a constant $\gamma(\epsilon, P_{XY}) > 0$ such that

$$\Pr\{(X^k, Y_{1+t}^{k+t}) \in \mathcal{T}_{\epsilon}^{(k)}(X, Y)\} \leq 2^{-k\gamma(\epsilon, P_{XY})},\ \forall t \neq 0.$$

Lemma 4 ([2, Typical Average Lemma]). For any $x^k \in \mathcal{T}_\epsilon^{(k)}(X)$ and any nonnegative function $g : \mathcal{X} \to \mathbb{R}$,

$$\frac{1}{k}\sum_{i=1}^{k} g(x_i) \le (1+\epsilon)E[g(X)].$$

Lemma 5 (Packing and Covering Lemma). Let $(X, Y) \in \mathcal{X}\times\mathcal{Y}$ be a pair of RVs, and for each $m \in [1 : M]$, $Y^k(m) \in \mathcal{Y}^k$ be a sequence of RVs distributed according to $\prod_{i=1}^{k} P_Y(y_i)$. Then, for any $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \to 0$ such that

$$\begin{aligned}&\Pr\{(X^k, Y^k(m)) \in \mathcal{T}_\epsilon^{(k)}(X, Y), \exists m \in [1 : M]\}\\ &\quad\le 2^{-k(I(X;Y)-\frac{1}{k}\log M-\delta(\epsilon))}.\end{aligned} \tag{19}$$

Furthermore, for sufficiently large $k > 0$,

$$\begin{aligned}&\Pr\{(X^k, Y^k(m)) \notin \mathcal{T}_\epsilon^{(k)}(X, Y), \forall m \in [1 : M]\}\\ &\quad\le 2^{-k\delta(\epsilon)} + \exp\{-2^{k(\frac{1}{k}\log M-I(X;Y)-\delta(\epsilon))}\}.\end{aligned} \tag{20}$$

Proof of Lemma 5. This lemma is immediately obtained from the proof of [2, Lemma 3.1], [2, Lemma 3.3], and (17) in Lemma 2. □

Now, we prove the theorem.

Proof of Theorem 2. If the sequence $\mathbf{u}$ does not satisfy that $\overline{\Delta}_{\mathbf{u}} = 0$ or $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$, i.e., the sequence $\mathbf{u}$ satisfies that $\overline{\Delta}_{\mathbf{u}} = 1$ or $0 = \underline{\Delta}_{\mathbf{u}} < \overline{\Delta}_{\mathbf{u}} < 1$, the advantage of the correlation of the source might not be used in the coding system. Hence, in this case, we employ (n, M_n, u_n)-codes that do not use side information. Then, we have $R_{\mathbf{u}}(D) \le R(D)$. Thus, in what follows, we assume that $\overline{\Delta}_{\mathbf{u}} = 0$ or $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$.

For an arbitrarily fixed $(D_1, D_2) \in \overline{\underline{\mathcal{D}}}_{\mathbf{u}}$, fix a conditional pmf $P_{V|X} \in \mathcal{W}(\mathcal{V}|\mathcal{X})$ and a function $g : \mathcal{Y}\times\mathcal{V} \to \hat{\mathcal{X}}$ that attain $R_{\mathrm{wz}}(D_1/(1+\epsilon))$, and fix a conditional pmf $P_{\hat{X}|X} \in \mathcal{W}(\hat{\mathcal{X}}|\mathcal{X})$ that attains $R(D_2/(1+\epsilon))$. Then, we consider the following random coding scheme:

Codebook generation: Randomly and independently generate $\tilde{M}_n$ sequences $v^{n-\lfloor\sqrt{n}\rfloor-u_n}(l) \in \mathcal{V}^{n-\lfloor\sqrt{n}\rfloor-u_n}$ each according to $\prod_{i=1}^{n-\lfloor\sqrt{n}\rfloor-u_n} P_V(v_i(l))$, $l \in [1 : \tilde{M}_n]$, where

$$P_V(v) = \sum_{x\in\mathcal{X}} P_X(x)P_{V|X}(v|x),\ \forall v \in \mathcal{V}.$$

Then, partition the set of indices $l \in [1 : \tilde{M}_n]$ into equal-size $M_n^{(1)}$ bins $\mathcal{B}(m_1)$, $m_1 \in [1 : M_n^{(1)}]$. If the indices cannot be partitioned into equal-size bins, assign the indices to each bin $\mathcal{B}(m_1)$ in ascending order such that the size of the bin satisfies $|\mathcal{B}(m_1)| = \lceil\tilde{M}_n/M_n^{(1)}\rceil$ whenever it is possible. In this case, if some bins are left over, these bins are not used. We note that the size of a bin is at most $\lceil\tilde{M}_n/M_n^{(1)}\rceil$ in each case. In what follows, for the sake of simplicity, we will omit the notation $\lfloor\cdot\rfloor$ for $\sqrt{n}$, and denote $n - \sqrt{n} - u_n$ by n_1. On the other hand, randomly and independently generate $M_n^{(2)}$ sequences $\hat{x}^{u_n}(m_2) \in \hat{\mathcal{X}}^{u_n}$, each according to $\prod_{i=1}^{u_n} P_{\hat{X}}(\hat{x}_i(m_2))$, $m_2 \in [1 : M_n^{(2)}]$, where

$$P_{\hat{X}}(\hat{x}) = \sum_{x\in\mathcal{X}} P_X(x)P_{\hat{X}|X}(\hat{x}|x),\ \forall \hat{x} \in \hat{\mathcal{X}}.$$

Encoding: Let $\epsilon > 0$ be a sufficiently small constant, and ϵ_1 be a constant satisfying $0 < \epsilon_1 < \epsilon$. For a source sequence $x^n \in \mathcal{X}^n$, the encoding procedure is described as follows:

1. For the sequence $x_{\sqrt{n}+1}^{n-u_n}$, the encoder finds an index $l \in [1 : \tilde{M}_n]$ such that
$$(x_{\sqrt{n}+1}^{n-u_n}, v^{n_1}(l)) \in \mathcal{T}_{\epsilon_1}^{(n_1)}(X, V).$$
If there is more than one such index, it selects one of them uniformly at random. If there is no such index, it selects an index from $[1 : \tilde{M}_n]$ uniformly at random.
2. For the sequence $x_{n-u_n+1}^{n}$, the encoder finds an index $m_2 \in [1 : M_n^{(2)}]$ such that
$$(x_{n-u_n+1}^{n}, \hat{x}^{u_n}(m_2)) \in \mathcal{T}_{\epsilon}^{(u_n)}(X, \hat{X})$$
If there is more than one such index, choose the smallest one among them. If there is no such index, it sets $m_2 = 1$.
3. The encoder sends the triple $(x^{\sqrt{n}}, m_1, m_2)$ to the decoder, where m_1 is the bin index such that $l \in \mathcal{B}(m_1)$.

Thus, the rate of this encoder is

$$R_n = \frac{\sqrt{n}}{n}\log|\mathcal{X}| + \frac{n_1}{n}R_n^{(1)} + \frac{u_n}{n}R_n^{(2)}, \tag{21}$$

where $R_n^{(1)} = \frac{1}{n_1}\log M_n^{(1)}$, and $R_n^{(2)} = \frac{1}{u_n}\log M_n^{(2)}$.

Decoding: For the triple $(x^{\sqrt{n}}, m_1, m_2)$, the decoding procedure is described as follows:

1. The decoder finds the unique estimate $\hat{t} \in [0 : u_n]$ of the delay such that
$$(x^{\sqrt{n}}, y_{1+\hat{t}}^{\sqrt{n}+\hat{t}}) \in \mathcal{T}_\epsilon^{(\sqrt{n})}(X, Y),$$
otherwise it sets $\hat{t} = 0$.
2. For the estimate $\hat{t}$, the decoder finds the unique index $\hat{l} \in \mathcal{B}(m_1)$ such that
$$(v^{n_1}(\hat{l}), y_{\sqrt{n}+1+\hat{t}}^{n-u_n+\hat{t}}) \in \mathcal{T}_\epsilon^{(n_1)}(V, Y),$$
otherwise it sets $\hat{l} = 1$.
3. The decoder outputs the reconstruction sequence as
$$\hat{x}^n = (\hat{x}^{\sqrt{n}}, \hat{x}^{n_1}(m_1), \hat{x}^{u_n}(m_2))$$
where $\hat{x}^{\sqrt{n}} \in \hat{\mathcal{X}}^{\sqrt{n}}$ is an arbitrarily fixed sequence, and
$$\hat{x}_i(m_1) = g(y_{\sqrt{n}+i+\hat{t}}, v_i(\hat{l})),\ \forall i \in \{1, 2, \cdots, n_1\}.$$

We now analyze the expected distortion for this random coding. Let (L, M_1, M_2) denote the indices found at the encoder, $\hat{T}$ be the estimate of the delay, and $\hat{L}$ be the index chosen at the decoder. Then, we consider the following error events:

$$\mathcal{E}_1 = \left\{(V^{n_1}(\hat{L}), X^{n-u_n}_{\sqrt{n}+1}, Y^{n-u_n-t+\hat{T}}_{\sqrt{n}+1-t+\hat{T}}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)\right\},$$
$$\mathcal{E}_2 = \left\{(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2)) \notin \mathcal{T}^{(u_n)}_{\epsilon}(X,\hat{X})\right\}.$$

Briefly speaking, if one of these events $\mathcal{E}_1$ and $\mathcal{E}_2$ occurs, it can not be guaranteed that the distortion of the random code is less than the given distortion level D.

We also consider the following events:

$$\mathcal{E}_0 = \{\hat{T} \neq t\},$$
$$\mathcal{E}_{1,1} = \left\{(V^{n_1}(l), X^{n-u_n}_{\sqrt{n}+1}) \notin \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X), \forall l \in [1:\tilde{M}_n]\right\},$$
$$\mathcal{E}_{1,2} = \left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}, Y^{n-u_n}_{\sqrt{n}+1}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)\right\},$$
$$\mathcal{E}_{1,3} = \left\{(V^{n_1}(\tilde{l}), Y^{n-u_n-t+\hat{T}}_{\sqrt{n}+1-t+\hat{T}}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\ \exists \tilde{l} \in \mathcal{B}(M_1) \text{ s.t. } \tilde{l} \neq L\right\}.$$

Whenever the event $\mathcal{E}_0^c \cap \mathcal{E}_{1,1}^c \cap \mathcal{E}_{1,2}^c \cap \mathcal{E}_{1,3}^c$ occurs, the event $\mathcal{E}_1^c$ occurs. Hence, for a delay $t \in [0:u_n]$, we have

$$\Pr\{\mathcal{E}_1^c|t\} \geq \Pr\{\mathcal{E}_0^c \cap \mathcal{E}_{1,1}^c \cap \mathcal{E}_{1,2}^c \cap \mathcal{E}_{1,3}^c|t\},$$

and we have

$$\begin{aligned}\Pr\{\mathcal{E}_1|t\} &\leq \Pr\{\mathcal{E}_0 \cup \mathcal{E}_{1,1} \cup \mathcal{E}_{1,2} \cup \mathcal{E}_{1,3}|t\} \\ &\leq \Pr\{\mathcal{E}_0|t\} + \Pr\{\mathcal{E}_{1,1}|t\} + \Pr\{\mathcal{E}_{1,1}^c \cap \mathcal{E}_{1,2}|t\} \\ &\quad + \Pr\{\mathcal{E}_0^c \cap \mathcal{E}_{1,3}|t\}. \end{aligned} \tag{22}$$

The first term in the RHS of (22) is upper bounded as

$$\begin{aligned}\Pr\{\mathcal{E}_0|t\} &\leq \Pr\left\{(X^{\sqrt{n}}, Y^{\sqrt{n}}) \notin \mathcal{T}^{(\sqrt{n})}_{\epsilon}(X,Y)\middle|t\right\} \\ &\quad + \Pr\left\{(X^{\sqrt{n}}, Y^{\sqrt{n}-t+\hat{t}}_{1-t+\hat{t}}) \in \mathcal{T}^{(\sqrt{n})}_{\epsilon}(X,Y),\right. \\ &\quad \left.\exists \hat{t} \in [0:u_n] \text{ s.t. } \hat{t} \neq t\middle|t\right\} \\ &\leq 2^{-\sqrt{n}\delta(\epsilon)} + u_n 2^{-\sqrt{n}\gamma(\epsilon,P_{XY})},\end{aligned}$$

where the last inequality comes from the inequality (17) in Lemma 2, and Lemma 3. Since $\overline{\Delta}_{\mathbf{u}} < 1$, we have $n_1 \to \infty$ $(n \to \infty)$. Thus, by using (20) in Lemma 5, the second term in the RHS of (22) is upper bounded for sufficiently large $n > 0$ as

$$\Pr\{\mathcal{E}_{1,1}|t\} \leq 2^{-n_1\delta(\epsilon_1)} + \exp\{-2^{n_1(\tilde{R}_n - I(X;V) - \delta(\epsilon_1))}\},$$

where $\tilde{R}_n = \frac{1}{n_1}\log \tilde{M}_n$. The third term in the RHS of (22) is upper bounded as

$$\begin{aligned}&\Pr\{\mathcal{E}_{1,1}^c \cap \mathcal{E}_{1,2}|t\} \\ &\leq \Pr\left\{\left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X)\right\}\right. \\ &\quad \left.\cap \left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}, Y^{n-u_n}_{\sqrt{n}+1}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)\right\}\middle|t\right\} \\ &= \sum_{(v^{n_1},x^{n_1}) \in \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X)} \sum_{\substack{y^{n_1}: \\ (v^{n_1},x^{n_1},y^{n_1}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)}} \\ &\quad \times \Pr\left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}, Y^{n-u_n}_{\sqrt{n}+1}) = (v^{n_1}, x^{n_1}, y^{n_1})\right\} \\ &= \sum_{(v^{n_1},x^{n_1}) \in \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X)} \Pr\left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}) = (v^{n_1}, x^{n_1})\right\} \\ &\quad \times \sum_{\substack{y^{n_1}: \\ (v^{n_1},x^{n_1},y^{n_1}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)}} \Pr\left\{Y^{n-u_n}_{\sqrt{n}+1} = y^{n_1}\middle|X^{n-u_n}_{\sqrt{n}+1} = x^{n_1}\right\} \\ &\overset{(a)}{=} \sum_{(v^{n_1},x^{n_1}) \in \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X)} \Pr\left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}) = (v^{n_1}, x^{n_1})\right\} \\ &\quad \times \sum_{\substack{y^{n_1}: \\ (v^{n_1},x^{n_1},y^{n_1}) \notin \mathcal{T}^{(n_1)}_{\epsilon}(V,X,Y)}} \prod_{i=1}^{n_1} P_{Y|VX}(y_i|v_i,x_i) \\ &\overset{(b)}{\leq} \sum_{(v^{n_1},x^{n_1}) \in \mathcal{T}^{(n_1)}_{\epsilon_1}(V,X)} \Pr\left\{(V^{n_1}(L), X^{n-u_n}_{\sqrt{n}+1}) = (v^{n_1}, x^{n_1})\right\} \\ &\quad \times \sum_{y^{n_1} \notin \mathcal{T}^{(n_1)}_{\epsilon'}(Y|v^{n_1},x^{n_1})} \prod_{i=1}^{n_1} P_{Y|VX}(y_i|v_i,x_i) \\ &\overset{(c)}{\leq} 2^{-n_1\delta(\epsilon')},\end{aligned}$$

where (a) follows since $Y \leftrightarrow X \leftrightarrow V$ and

$$\Pr\left\{Y^{n-u_n}_{\sqrt{n}+1} = y^{n_1}\middle|X^{n-u_n}_{\sqrt{n}+1} = x^{n_1}\right\} = \prod_{i=1}^{n_1} P_{Y|X}(y_i|x_i),$$

(b) comes from the fact that for $(v^k, x^k) \in \mathcal{T}^{(k)}_{\epsilon_1}(V,X)$, there exists $\epsilon' \in (0,\epsilon)$ such that

$$\{y^k \in \mathcal{Y}^k : (v^k, x^k, y^k) \in \mathcal{T}^{(k)}_{\epsilon}(V,X,Y)\} \supseteq \mathcal{T}^{(k)}_{\epsilon'}(Y|v^k,x^k) \tag{23}$$

(see Appendix), and (c) comes from (18) in Lemma 2.

The last term in the RHS of (22) is upper bounded as

$$\begin{aligned}&\Pr\{\mathcal{E}_0^c \cap \mathcal{E}_{1,3}|t\} \\ &\leq \Pr\left\{(V^{n_1}(\tilde{l}), Y^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\right. \\ &\quad \left.\exists \tilde{l} \in \mathcal{B}(M_1) \text{ s.t. } \tilde{l} \neq L\right\} \\ &= \sum_{(m_1,l)} \Pr\{(M_1,L) = (m_1,l)\} \\ &\quad \times \Pr\left\{(V^{n_1}(\tilde{l}), Y^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\right. \\ &\quad \left.\exists \tilde{l} \in \mathcal{B}(m_1) \text{ s.t. } \tilde{l} \neq l | M_1 = m_1, L = l\right\} \\ &\overset{(a)}{=} \sum_{(m_1,l)} \Pr\{(M_1,L) = (m_1,l)\} \\ &\quad \times \Pr\left\{(V^{n_1}(\tilde{l}), Y^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\right. \\ &\quad \left.\exists \tilde{l} \in \mathcal{B}(m_1) \text{ s.t. } \tilde{l} \neq l | L = l\right\} \\ &\overset{(b)}{\leq} \sum_{(m_1,l)} \Pr\{(M_1,L) = (m_1,l)\} \\ &\quad \times \Pr\left\{(V^{n_1}(\tilde{l}), Y^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\right. \\ &\quad \left.\exists \tilde{l} \in [1:\lceil \tilde{M}_n/M_n^{(1)} \rceil - 1] | L = l\right\}\end{aligned}$$

$$= \Pr\left\{(V^{n_1}(\tilde{l}), Y^{n-u_n}_{\sqrt{n}+1}) \in \mathcal{T}^{(n_1)}_{\epsilon}(V,Y),\right.$$
$$\left.\exists \tilde{l} \in [1 : \lceil \tilde{M}_n/M^{(1)}_n \rceil - 1]\right\}$$
$$\overset{(c)}{\le} 2^{-n_1(I(Y;V)-\tilde{R}_n+R^{(1)}_n-\delta(\epsilon))},$$

where (a) comes from the fact that M_1 is a function of L, (b) follows since $V^{n_1}(\tilde{l})$ has the same distribution for $\tilde{l} \neq l$, and (c) comes from (19) in Lemma 5. Hence, for sufficiently large $n > 0$, we set

$$I(X;V) + 2\delta(\epsilon_1) \le \tilde{R}_n \le I(X;V) + 3\delta(\epsilon_1), \quad (24)$$
$$\tilde{R}_n - I(Y;V) + 2\delta(\epsilon) \le R^{(1)}_n \le \tilde{R}_n - I(Y;V) + 3\delta(\epsilon). \quad (25)$$

Then, there exists constants $\gamma_{1,1}, \gamma_{1,2} > 0$ that do not depend on a delay t such that

$$\Pr\{\mathcal{E}_1|t\} \le 2^{-\gamma_{1,1}\sqrt{n}} + 2^{-\gamma_{1,2}n_1}. \quad (26)$$

On the other hand, when $\underline{\Delta}_{\mathbf{u}} > 0$, we have $u_n \to \infty$ $(n \to \infty)$. Thus, by using (20) in Lemma 5, the probability of the event $\mathcal{E}_2$ is upper bounded for sufficiently large $n > 0$ as

$$\Pr\{\mathcal{E}_2|t\} \le 2^{-u_n\delta(\epsilon)} + \exp\{-2^{u_n(R^{(2)}_n - I(X;\hat{X}) - \delta(\epsilon))}\}.$$

Thus, for sufficiently large $n > 0$, we set

$$I(X;\hat{X}) + 2\delta(\epsilon) \le R^{(2)}_n \le I(X;\hat{X}) + 3\delta(\epsilon). \quad (27)$$

Then, there exists a constant $\gamma_2 > 0$ that does not depend on a delay t such that

$$\Pr\{\mathcal{E}_2|t\} \le 2^{-\gamma_2 u_n}. \quad (28)$$

Let C be the RV that denotes the above random coding, and c be a realization of C, i.e., c denotes an (n, M_n, u_n)-code. Then, for a realization c, and a delay $t \in [0 : u_n]$, we have

$$E[d^n(X^n, \hat{X}^n)|t, c]$$
$$= E[d^{\sqrt{n}}(X^{\sqrt{n}}, \hat{X}^{\sqrt{n}})|t, c] + E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c]$$
$$+ E[d^{u_n}(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2))|t, c]. \quad (29)$$

The second term in the RHS of (29) is upper bounded as

$$E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c]$$
$$= \Pr\{\mathcal{E}^c_1|t, c\}E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c, \mathcal{E}^c_1]$$
$$+ \Pr\{\mathcal{E}_1|t, c\}E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c, \mathcal{E}_1]$$
$$= \Pr\{\mathcal{E}^c_1|t, c\}$$
$$\times E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, g(Y^{n-u_n-t+\hat{T}}_{\sqrt{n}+1-t+\hat{T}}, V^{n_1}(\hat{L})))|t, c, \mathcal{E}^c_1]$$
$$+ \Pr\{\mathcal{E}_1|t, c\}E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c, \mathcal{E}_1]$$
$$\overset{(a)}{\le} n_1(1+\epsilon)E[d(X, g(Y,V))]$$
$$+ \Pr\{\mathcal{E}_1|t, c\}E[d^{n_1}(X^{n-u_n}_{\sqrt{n}+1}, \hat{X}^{n_1}(M_1))|t, c, \mathcal{E}_1]$$
$$\overset{(b)}{\le} n_1 D_1 + \Pr\{\mathcal{E}_1|t, c\}n_1 d_{\max}, \quad (30)$$

where (a) comes from Lemma 4, and (b) follows since the conditional pmf $P_{V|X}$ and the function g attain the RD function $R_{\mathrm{wz}}(D_1/(1+\epsilon))$. The third term in the RHS of (29) is upper bounded as

$$E[d^{u_n}(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2))|t, c]$$
$$= \Pr\{\mathcal{E}^c_2|t, c\}E[d^{u_n}(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2))|t, c, \mathcal{E}^c_2]$$
$$+ \Pr\{\mathcal{E}_2|t, c\}E[d^{u_n}(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2))|t, c, \mathcal{E}_2]$$
$$\overset{(a)}{\le} u_n(1+\epsilon)E[d(X, \hat{X})]$$
$$+ \Pr\{\mathcal{E}_2|t, c\}E[d^{u_n}(X^n_{n-u_n+1}, \hat{X}^{u_n}(M_2))|t, c, \mathcal{E}_2]$$
$$\overset{(b)}{\le} u_n D_2 + \Pr\{\mathcal{E}_2|t, c\}u_n d_{\max}, \quad (31)$$

where (a) comes from Lemma 4, and (b) follows since the conditional pmf $P_{\hat{X}|X}$ attains the RD function $R(D_2/(1+\epsilon))$.

We now show the existence of a sequence of (n, M_n, u_n)-codes with the desired expected distortion for the following two cases:

The case where $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$**:** Since C does not depend on a delay t, we have

$$\sum_c \Pr\{C = c\} \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_1 \cup \mathcal{E}_2|t, c\}$$
$$\le \sum_c \Pr\{C = c\} \sum_{t\in[0:u_n]} \Pr\{\mathcal{E}_1 \cup \mathcal{E}_2|t, c\}$$
$$= \sum_{t\in[0:u_n]} \Pr\{\mathcal{E}_1 \cup \mathcal{E}_2|t\}$$
$$\le (u_n + 1)(2^{-\gamma_{1,1}\sqrt{n}} + 2^{-\gamma_{1,2}n_1} + 2^{-\gamma_2 u_n}),$$

where the last inequality comes from (26) and (28). Hence, there exists a sequence $\{\tilde{c}_n\}^{\infty}_{n=1}$ of (n, M_n, u_n)-codes such that

$$\lim_{n\to\infty} \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_1 \cup \mathcal{E}_2|t, \tilde{c}_n\} = 0. \quad (32)$$

Thus for this sequence $\{\tilde{c}_n\}^{\infty}_{n=1}$, according to (29)–(31), we have

$$\limsup_{n\to\infty} \max_{t\in[0:u_n]} \frac{1}{n}E[d^n(X^n, \hat{X}^n)|t, \tilde{c}_n]$$
$$\le \limsup_{n\to\infty} \frac{n_1}{n}\Big(D_1 + \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_1|t, \tilde{c}_n\}d_{\max}\Big)$$
$$+ \limsup_{n\to\infty} \frac{u_n}{n}\Big(D_2 + \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_2|t, \tilde{c}_n\}d_{\max}\Big)$$
$$= (1 - \underline{\Delta}_{\mathbf{u}})D_1 + \overline{\Delta}_{\mathbf{u}}D_2$$
$$\le D,$$

where the last inequality comes from the fact that $(D_1, D_2) \in \overline{\mathcal{D}}_{\mathbf{u}}$.

The case where $\overline{\Delta}_{\mathbf{u}} = 0$**:** In the similar way to the above argument, we can show the existence of a sequence $\{\hat{c}_n\}^{\infty}_{n=1}$ of (n, M_n, u_n)-codes such that

$$\lim_{n\to\infty} \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_1|t, \hat{c}_n\} = 0.$$

Thus for this sequence $\{\hat{c}_n\}_{n=1}^{\infty}$, according to (29)–(31), we have

$$\begin{aligned}
&\limsup_{n\to\infty} \max_{t\in[0:u_n]} \frac{1}{n} E[d^n(X^n, \hat{X}^n)|t, \hat{c}_n] \\
&\quad \le \limsup_{n\to\infty} \frac{n_1}{n}\Big(D_1 + \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_1|t, \hat{c}_n\} d_{\max}\Big) \\
&\qquad + \limsup_{n\to\infty} \frac{u_n}{n}\Big(D_2 + \max_{t\in[0:u_n]} \Pr\{\mathcal{E}_2|t, \hat{c}_n\} d_{\max}\Big) \\
&\quad =(1-\underline{\Delta}_{\mathbf{u}})D_1 \\
&\quad \le D.
\end{aligned}$$

On the other hand, according to (21), (24), (25), and (27), rates of these two sequences $\{\tilde{c}_n\}_{n=1}^{\infty}$ and $\{\hat{c}_n\}_{n=1}^{\infty}$ of (n, M_n, u_n)-codes satisfy

$$\begin{aligned}
&\limsup_{n\to\infty} R_n \\
&\quad \le \limsup_{n\to\infty} \frac{n_1}{n}\Big(I(X;V) - I(Y;V) + 3\delta(\epsilon) + 3\delta(\epsilon_1)\Big) \\
&\qquad + \limsup_{n\to\infty} \frac{u_n}{n}\Big(I(X;\hat{X}) + 3\delta(\epsilon)\Big) \\
&\quad =(1-\underline{\Delta}_{\mathbf{u}})\Big(R_{\mathrm{wz}}(D_1/(1+\epsilon)) + 3\delta(\epsilon) + 3\delta(\epsilon_1)\Big) \\
&\qquad + \overline{\Delta}_{\mathbf{u}}\Big(R(D_2/(1+\epsilon)) + 3\delta(\epsilon)\Big).
\end{aligned}$$

Thus, by recalling that $(D_1, D_2) \in \overline{\mathcal{D}}_{\mathbf{u}}$ is arbitrary, and noting that $R_{\mathrm{wz}}(D)$ and $R(D)$ are continuous functions, we have

$$R_{\mathbf{u}}(D) \le \min_{(D_1,D_2)\in\overline{\mathcal{D}}_{\mathbf{u}}} \{(1-\underline{\Delta}_{\mathbf{u}})R_{\mathrm{wz}}(D_1) + \overline{\Delta}_{\mathbf{u}} R(D_2)\}.$$

This completes the proof. □

Remark 6. If we use the following three types of coding for sufficiently small $\epsilon > 0$, we might be able to remove the condition $\overline{\Delta}_{\mathbf{u}} = 0$ or $0 < \underline{\Delta}_{\mathbf{u}} \le \overline{\Delta}_{\mathbf{u}} < 1$.

- For $n > 0$ such that $1-\epsilon \le u_n/n$, we use the ordinary lossy source coding without side information.
- For $n > 0$ such that $u_n/n \le \epsilon$, we use the Wyner-Ziv source coding after estimating the delay by using the above method.
- For $n > 0$ such that $\epsilon < u_n/n < 1-\epsilon$, we use the source coding described in the above proof.

However, for the sake of brevity, we do not employ such method in this paper.

Remark 7. For the sequence $\{\tilde{c}_n\}_{n=1}^{\infty}$ (and also $\{\hat{c}_n\}_{n=1}^{\infty}$) of codes, according to (29)–(32), we have

$$\begin{aligned}
&\limsup_{n\to\infty} \max_{t\in[0:u_n]} \frac{1}{n} E[d^n(X^n, \hat{X}^n)|t, \tilde{c}_n] \\
&\quad \le \limsup_{n\to\infty} \{(1-\Delta_n)D_1 + \Delta_n D_2\}.
\end{aligned}$$

Further, according to (21), (24), (25), and (27), for any $\delta > 0$, we have

$$\limsup_{n\to\infty} R_n \le \limsup_{n\to\infty} \{(1-\Delta_n)R_{\mathrm{wz}}(D_1) + \Delta_n R(D_2)\} + \delta.$$

Thus, by using an arbitrary fixed $(D_1, D_2) \in [0, d_{\max}]^2$ satisfying $D \ge \limsup_{n\to\infty}\{(1-\Delta_n)D_1 + \Delta_n D_2\}$, and repeating the same augment of the proof, we have

$$\begin{aligned}
R_{\mathbf{u}}(D) \le \inf_{\substack{(D_1,D_2)\in[0,d_{\max}]^2: \\ D\ge\limsup_{n\to\infty}\{(1-\Delta_n)D_1+\Delta_n D_2\}}} &\limsup_{n\to\infty} \{(1-\Delta_n)R_{\mathrm{wz}}(D_1) \\
&+ \Delta_n R(D_2)\}.
\end{aligned}$$

6. Conclusion

This paper has dealt with the lossy source coding problem with delayed side information at the decoder, assuming that a delay is unknown but the sequence $\mathbf{u}$ of maximums of delay is known to both the encoder and the decoder. We have shown upper and lower bounds of the RD function $R_{\mathbf{u}}(D)$. We also have clarified that the upper bound coincides with the lower bound when $\{u_n/n\}_{n=1}^{\infty}$ converges to a constant as $n \to \infty$. Further, we have given a necessary and sufficient condition in which $R_{\mathbf{u}}(D) = R_{\mathrm{wz}}(D)$.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

This work was partially supported by Grant-in-Aid for Young Scientists (B) No. 25820154.

References

[1] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley-Interscience New York, 2006.
[2] A. El Gamal and Y.H. Kim, Network Information Theory, Cambridge University Press, 2012.
[3] 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.
[4] O. Simeone and H. Permuter, "Source coding when the side information may be delayed," IEEE Trans. Inf. Theory, vol.59, no.6, pp.3607–3618, June 2013.
[5] R. Venkataramanan and S. Pradhan, "Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," IEEE Trans. Inf. Theory, vol.53, no.6, pp.2154–2179, June 2007.
[6] C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," IRE Nat. Conv. Rec., vol.4, no.142-163, 1959.
[7] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed., Cambridge University Press, 2011.

Appendix

In this appendix we show (23).

For any $(v^k, x^k) \in \mathcal{T}_{\epsilon_1}^{(k)}(V, X)$ and any $(v, x, y) \in \mathcal{V}\times\mathcal{X}\times\mathcal{Y}$, we have

$$\begin{aligned}
&\{y^k \in \mathcal{Y}^k : |\pi(v, x, y|v^k, x^k, y^k) - P_{VXY}(v, x, y)| \\
&\quad \le \epsilon P_{VXY}(v, x, y)\}
\end{aligned}$$

$$\begin{aligned}
&\overset{\text{(a)}}{\supseteq} \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) \\
&\quad - \pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x)| \\
&\quad + |\pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x) - P_{VXY}(v,x,y)| \\
&\quad \le \epsilon P_{VXY}(v,x,y)\} \\
&= \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) \\
&\quad - \pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x)| \\
&\quad + |\pi(v,x|v^k,x^k) - P_{VX}(v,x)|P_{Y|VX}(y|v,x) \\
&\quad \le \epsilon P_{VXY}(v,x,y)\} \\
&\overset{\text{(b)}}{\supseteq} \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) \\
&\quad - \pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x)| \\
&\quad + \epsilon_1 P_{VX}(v,x)P_{Y|VX}(y|v,x) \\
&\quad \le \epsilon P_{VXY}(v,x,y)\} \\
&\supseteq \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) \\
&\quad - \pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x)| \\
&\quad \le (\epsilon - \epsilon_1)p_{VX}P_{Y|VX}(y|v,x)\},
\end{aligned}$$

where $p_{VX} = \min\{P_{VX}(v,x) : P_{VX}(v,x) > 0\}$, (a) comes from the triangle inequality, and (b) comes from the fact that

$$|\pi(v,x|v^k,x^k) - P_{VX}(v,x)| \le \epsilon_1 P_{VX}(v,x), \quad \forall (v,x) \in \mathcal{V} \times \mathcal{X}.$$

Hence, for any $(v^k, x^k) \in \mathcal{T}^{(k)}_{\epsilon_1}(V,X)$, we have

$$\begin{aligned}
&\{y^k \in \mathcal{Y}^k : (v^k,x^k,y^k) \in \mathcal{T}^{(k)}_{\epsilon}(V,X,Y)\} \\
&= \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) - P_{VXY}(v,x,y)| \\
&\quad \le \epsilon P_{VXY}(v,x,y), \forall (v,x,y) \in \mathcal{V} \times \mathcal{X} \times \mathcal{Y}\} \\
&\supseteq \{y^k \in \mathcal{Y}^k : |\pi(v,x,y|v^k,x^k,y^k) \\
&\quad - \pi(v,x|v^k,x^k)P_{Y|VX}(y|v,x)| \\
&\quad \le (\epsilon - \epsilon_1)p_{VX}P_{Y|VX}(y|v,x), \forall (v,x,y) \in \mathcal{V} \times \mathcal{X} \times \mathcal{Y}\} \\
&= \mathcal{T}^{(k)}_{(\epsilon-\epsilon_1)p_{VX}}(Y|v^k,x^k).
\end{aligned}$$

Now, by setting $\epsilon' = (\epsilon - \epsilon_1)p_{VX}$, we have (23).

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. From 2012 to 2013, he was an assistant professor in the Department of Communications and Integrated Systems of Tokyo Institute of Technology. He is currently an assistant professor in the Department of Communications and Computer Engineering of Tokyo Institute of Technology. His current research interests are in the area of multi-terminal information theory.

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 Communications and Computer Engineering as professor. In 1992 and 1996, he was a visiting researcher at the Centre National de la Recherche Scientifique, France and Delft University of Technology, Netherlands, respectively. He received the Achievement Award in 2008, and the Best Paper Award in 1993, 1996, 2002, 2007 and 2011 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.