

Universal coding for correlated sources with complementary delivery

Akisato Kimura^{*†‡}, Tomohiko Uyematsu^{*}, Shigeaki Kuzuoka[†]

^{*} Department of Communications and Integrated Systems, Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan.

[†] Department of Computer and Communication Sciences, Wakayama University,
930 Sakaedani, Wakayama, Wakayama, 640-8510 Japan.

[‡] NTT Communication Science Laboratories, NTT Corporation,
3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan.

URL: <http://www.brl.ntt.co.jp/people/akisato/>

Abstract—This report deals with a universal coding problem for a certain kind of multiterminal source coding system that we call the complementary delivery coding system. Both fixed-to-fixed length and fixed-to-variable length lossless coding schemes are considered. Explicit constructions of universal codes and the bounds of the error probabilities are clarified via type-theoretical and graph-theoretical analyses.

I. INTRODUCTION

A coding problem for correlated information sources was first described and investigated by Slepian and Wolf [11], and later, various coding problems derived from that work were considered (e.g. Wyner [14], Körner and Marton [7], Sgarro [10]). Meanwhile, the problem of universal coding for these systems was first investigated by Csiszár and Körner [3]. Universal coding problems are not only interesting in their own right but also very important in terms of practical applications. Subsequent work has mainly focused on the Slepian-Wolf coding system [2], [9], [12] since it appears to be difficult to construct universal codes for most of the other coding systems.

This report deals with a universal coding problem for a certain kind of multiterminal source coding system that we call a complementary delivery coding system [8]. Figure 1 shows a block diagram of the complementary delivery coding system. The encoder observes messages emitted from two correlated sources, and delivers these messages to other locations (i.e. decoders). Each decoder has access to one of two messages, and therefore wants to reproduce the other message. Although the previous articles [8] considered lossy configurations, this report considers a lossless configuration. We show an explicit construction of fixed-to-fixed length universal codes. We also clarify the upper and lower bounds of the error probabilities via type-theoretical and graph-theoretical analyses. Fixed-to-variable universal codes can also be constructed in a similar manner.

II. PRELIMINARIES

A. Basic definitions

Let \mathcal{X} be a finite set, \mathcal{B} be a binary set, and \mathcal{B}^* be a set of all finite sequences in the alphabet \mathcal{B} . Let $|\mathcal{X}|$ be the cardinality of \mathcal{X} and $\mathcal{I}_M = \{1, 2, \dots, M\}$. A member of \mathcal{X}^n

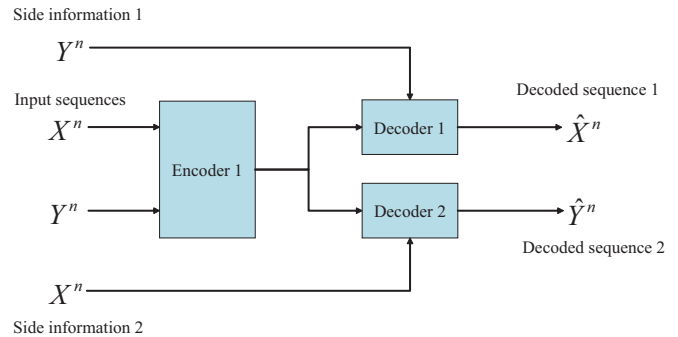


Fig. 1. Complementary delivery coding system

is written as $x^n = (x_1, x_2, \dots, x_n)$, and substrings of x^n are written as $x_i^j = (x_i, x_{i+1}, \dots, x_j)$ for $i \leq j$. When the dimension is clear from the context, vectors will be denoted by boldface letters, i.e., $\mathbf{x} \in \mathcal{X}^n$. $\mathcal{M}(\mathcal{X})$ denotes the set of all probability distributions on \mathcal{X} . Also, $\mathcal{M}(\mathcal{X}|P_Y)$ denotes the set of all probability distributions on \mathcal{X} given a distribution $P_Y \in \mathcal{M}(\mathcal{Y})$, namely each member $P_{X|Y}$ of $\mathcal{M}(\mathcal{X}|P_Y)$ is characterized by $P_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ as $P_{XY} = P_{X|Y}P_Y$. A discrete memoryless source (\mathcal{X}, P_X) is an infinite sequence of independent copies of a random variable X taking values in \mathcal{X} with a generic distribution $P_X \in \mathcal{M}(\mathcal{X})$. We will denote a source (\mathcal{X}, P_X) by referring to its generic distribution P_X or random variable X . For a correlated source (X, Y) , $H(X|Y)$ denotes the conditional entropy of X given Y . For a generic distribution $P_Y \in \mathcal{M}(\mathcal{Y})$ and a conditional distribution $P_{X|Y} \in \mathcal{M}(\mathcal{X}|P_Y)$, $H(P_{X|Y}|P_Y)$ also denotes the conditional entropy of X given Y . $D(P||Q)$ denotes the Kullback-Leibler divergence between two distributions P and Q . In the following, all bases of exponentials and logarithms are set at 2.

B. Types of sequences

Let us define the *type* of a sequence $\mathbf{x} \in \mathcal{X}^n$ as the empirical distribution $Q_{\mathbf{x}} \in \mathcal{M}(\mathcal{X})$ of the sequence \mathbf{x} , i.e. $Q_{\mathbf{x}}(a) = \frac{1}{n}N(a|\mathbf{x}) \forall a \in \mathcal{X}$, where $N(a|\mathbf{x})$ represents the number of occurrences of the letter a in the sequence \mathbf{x} . Similarly, the joint

type $Q_{\mathbf{x}, \mathbf{y}} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is defined in a similar manner. Let $\mathcal{P}_n(\mathcal{X})$ be the set of types of sequences in \mathcal{X}^n . Similarly, for every type $Q \in \mathcal{P}_n(\mathcal{X})$, let $\mathcal{V}_n(\mathcal{Y}|Q)$ be the set of all stochastic matrices $V: \mathcal{X} \rightarrow \mathcal{Y}$ such that for some pairs $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ of sequences we have $Q_{\mathbf{x}, \mathbf{y}}(\mathbf{x}, \mathbf{y}) = Q(\mathbf{x})V(\mathbf{y}|\mathbf{x})$. For every type $Q \in \mathcal{P}_n(\mathcal{X})$ we denote

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

Similarly, for every $\mathbf{x} \in T_Q^n$ and $V \in \mathcal{V}_n(\mathcal{Y}|Q)$ we define

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

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

III. PREVIOUS RESULTS

This section formulates the coding problem investigated in this report, and shows the fundamental bound of the coding rate. First, we formulate the coding problem of the complementary delivery coding system.

Definition 1: (Fixed-to-fixed complementary delivery (FF-CD) code)

A set $(\varphi^n, \hat{\varphi}_{(1)}^n, \hat{\varphi}_{(2)}^n)$ of an encoder and two decoders is an FF-CD code with parameters $(n, M_n, e_n^{(X)}, e_n^{(Y)})$ for the source (X, Y) if and only if

$$\begin{aligned} \varphi^n &: \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \mathcal{I}_{M_n} \\ \hat{\varphi}_{(1)}^n &: \mathcal{I}_{M_n} \times \mathcal{Y}^n \rightarrow \mathcal{X}^n, \quad \hat{\varphi}_{(2)}^n : \mathcal{I}_{M_n} \times \mathcal{X}^n \rightarrow \mathcal{Y}^n, \\ e_n^{(X)} &= \Pr\{X^n \neq \hat{X}^n\}, \quad e_n^{(Y)} = \Pr\{Y^n \neq \hat{Y}^n\}, \end{aligned}$$

where

$$\begin{aligned} \hat{X}^n &\stackrel{\text{def.}}{=} \hat{\varphi}_{(1)}^n(\varphi^n(X^n, Y^n), Y^n), \\ \hat{Y}^n &\stackrel{\text{def.}}{=} \hat{\varphi}_{(2)}^n(\varphi^n(X^n, Y^n), X^n). \end{aligned}$$

Definition 2: (FF-CD-achievable rate)

R is an FF-CD-achievable rate of the source (X, Y) if and only if there exists a sequence $\{(\varphi^n, \hat{\varphi}_{(1)}^n, \hat{\varphi}_{(2)}^n)\}_{n=1}^{\infty}$ of FF-CD codes with parameters $\{(n, M_n, e_n^{(X)}, e_n^{(Y)})\}_{n=1}^{\infty}$ for the source (X, Y) such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log M_n \leq R, \quad \lim_{n \rightarrow \infty} e_n^{(X)} = \lim_{n \rightarrow \infty} e_n^{(Y)} = 0.$$

Definition 3: (Inf FF-CD-achievable rate)

$$R_f(X, Y) = \inf\{R :$$

R is an FF-CD-achievable rate of $(X, Y)\}$.

Willems, Wolf and Wyner [13], [15] investigated a coding problem where several users are physically separated but communicate with each other via a satellite, and determined the minimum coding rate for the three users when transmitting to and from the satellite. The complementary delivery coding system is a special case of the system described by Willems et al., which considers the case of two users. Therefore, we can immediately obtain the closed form of $R_f(X, Y)$ from the result obtained by Willems et al.

	\mathcal{Y}_1	\mathcal{Y}_2	\mathcal{Y}_3	\mathcal{Y}_4	\mathcal{Y}_5		\mathcal{Y}_1	\mathcal{Y}_2	\mathcal{Y}_3	\mathcal{Y}_4	\mathcal{Y}_5		\mathcal{Y}_1	\mathcal{Y}_2	\mathcal{Y}_3	\mathcal{Y}_4	\mathcal{Y}_5
x_1						x_1	1	2	3			x_1	1	2	3		
x_2						x_2						x_2	2			1	3
x_3						x_3						x_3		3	1	2	
x_4						x_4						x_4	3	1			2
x_5						x_5						x_5			2	3	1

Fig. 2. Example of coding scheme (left) Coding table (middle) Positions where codewords will be provided (right) Provided codewords

Theorem 1: (Coding theorem of FF-CD codes [8])

$$\begin{aligned} R_f(X, Y) &= \max\{H(X|Y), H(Y|X)\} \\ &= \max\{H(P_{X|Y}|P_Y), H(P_{Y|X}|P_X)\} \end{aligned}$$

IV. CODE CONSTRUCTION

This section shows an explicit construction of universal codes for the complementary delivery coding system defined by Definition 1. The coding scheme is described as follows: [Encoding]

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

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

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

- 2) Create a table (henceforth we call this a *coding table*, see the left side of Figure 2) for each joint type $Q_{XY} \in \mathcal{S}_n(R)$. Each row of the coding table corresponds to a sequence $\mathbf{x} \in T_{Q_X}^n$, and each column corresponds to a sequence $\mathbf{y} \in T_{Q_Y}^n$.
- 3) Mark cells that correspond to sequence pairs $(\mathbf{x}, \mathbf{y}) \in T_{Q_{XY}}^n$ (see the middle of Figure 2). Codewords will be given only to sequence pairs that correspond to marked cells.
- 4) Fill the marked cells with $\exp(nR)$ different symbols such that each symbol occurs at most once in each row and at most once in each column. An example of symbol filling is shown on the right side of Figure 2.
- 5) For a given pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$ with the joint type Q_{XY} , if $Q_{XY} \in \mathcal{S}_n(R)$, the index assigned to the joint type Q_{XY} of (\mathbf{x}, \mathbf{y}) is the first part of the codeword, and the symbol filling the cell of (\mathbf{x}, \mathbf{y}) in the coding table of Q_{XY} is determined as the second part of the codeword. For the sequence pairs (\mathbf{x}, \mathbf{y}) whose joint type Q_{XY} does not belong to $\mathcal{S}_n(R)$, the corresponding codeword is determined arbitrarily and an encoding error is declared.

[Decoding: $\hat{\varphi}_{(1)}^n$] (Almost the same as for $\hat{\varphi}_{(2)}^n$)

- 1) Find the coding table of the type \hat{Q}_{XY} that corresponds to the first part of the received codeword. The decoder

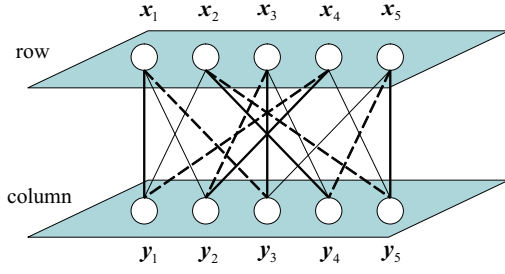


Fig. 3. Example of a bipartite graph ($m_x = m_y = 5$, $n_x = n_y = 3$, equivalent to the table in Fig. 2 right)

can find the coding table used in the encoding scheme if no encoding error occurs. In this case, \hat{Q}_{XY} should be Q_{XY} .

- 2) Find the cell filled with the second part of the received codeword from the column of the side information sequence $\mathbf{y} \in T_{Q_Y}^n$. The sequence $\hat{\mathbf{x}} \in T_{Q_X}^n$ that corresponds to the row of the cell found in this step is reproduced.

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

Lemma 1: For a given coding table of a joint type $Q_{XY} \in \mathcal{S}_n(R)$, the number of marked cells in every row of the coding table $N_y(Q_{XY}) = |T_Y^n(\mathbf{x})|$ is a constant value that is less than $\exp(nR)$, and the number of marked cells in every column of the coding table $N_x(Q_{XY}) = |T_W^n(\mathbf{y})|$ is also a constant value that is less than $\exp(nR)$, both of which depend solely on the joint type Q_{XY} .

Proof: This lemma can be directly derived from [4, Lemma 2.5] and the definition of the set $\mathcal{S}_n(R)$ used when creating a coding table. ■

Lemma 2: For given integers m_x , m_y , n_x and n_y that satisfy $m_x \geq n_x$ and $m_y \geq n_y$, there exists an $m_x \times m_y$ table filled with $\max(n_x, n_y)$ different symbols such that

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

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

- each node in one set corresponds to a row in the table, and each node in the other set corresponds to a column in the table,
- each edge corresponds to a cell in the table, to which a certain symbol is assigned,
- $\max(n_x, n_y)$ different colors are given to edges, each of which corresponds to a symbol in the table,
- no two edges with the same color share a common node.

Figure 3 shows an example of such a graph. Here, let us introduce the following lemma for bipartite graphs:

Lemma 3: (König [6], [1])

If a graph G is bipartite, the minimum number of colors necessary for edge coloring of the graph G equals the maximum degree of G .

Lemma 3 ensures the existence of the above mentioned bipartite graph. ■

From Lemmas 1 and 2, we can easily show the existence of coding tables by setting $m_x = |T_{Q_X}^n|$, $m_y = |T_{Q_Y}^n|$, $n_x = |T_W^n(\mathbf{y})|$ and $n_y = |T_V^n(\mathbf{x})|$ in Lemma 2.

V. CODING THEOREMS

We can obtain the following theorem for the universal FF-CD codes constructed in Section IV.

Theorem 2: For a given real number $R > 0$, there exists a sequence of universal FF-CD codes with parameters $\{(n, M_n, e_n^{(X)}, e_n^{(Y)})\}_{n=1}^{\infty}$ such that for any integer $n \geq 1$ and any source (X, Y) with a generic distribution $P_{XY} \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$

$$\begin{aligned} \frac{1}{n} \log M_n &\leq R + \frac{1}{n} |\mathcal{X} \times \mathcal{Y}| \log(n+1), \\ e_n^{(X)} + e_n^{(Y)} &\leq 2(n+1)^{|\mathcal{X} \times \mathcal{Y}|} \\ &\quad \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} D(Q_{XY} \| P_{XY}) \right\}, \end{aligned}$$

where $\overline{\mathcal{S}}_n(R) = \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) - \mathcal{S}_n(R)$.

Proof: Lemmas 1 and 2 ensure the existence of a coding table for every joint type $Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. From the coding scheme, the size of the codeword set is bounded as

$$\begin{aligned} M_n &\leq |\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \exp(nR) \\ &\leq (n+1)^{|\mathcal{X} \times \mathcal{Y}|} \exp(nR), \quad (\because [4, \text{Lemma 2.2}]) \end{aligned}$$

which implies the first inequality of Theorem 2. Next, we evaluate decoding error probabilities. Since every sequence pair $(\mathbf{x}, \mathbf{y}) \in T_{Q_{XY}}^n$ that satisfies $Q_{XY} \in \mathcal{S}_n(R)$ is reproduced correctly at the decoder, the sum of error probabilities is bounded as

$$\begin{aligned} e_n^{(X)} + e_n^{(Y)} &\leq 2 \Pr \{ \exists Q_{XY} \in \overline{\mathcal{S}}_n(R), (X^n, Y^n) \in T_{Q_{XY}}^n \} \\ &\leq 2 \sum_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} \exp\{-nD(Q_{XY} \| P_{XY})\} \\ &\quad (\because [4, \text{Lemma 2.6}]) \\ &\leq 2 \sum_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} D(Q_{XY} \| P_{XY}) \right\} \\ &\leq 2(n+1)^{|\mathcal{X} \times \mathcal{Y}|} \\ &\quad \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} D(Q_{XY} \| P_{XY}) \right\} \\ &\quad (\because [4, \text{Lemma 2.2}]) \end{aligned}$$

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

$$\min_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} D(Q_{XY} \| P_{XY}) > 0.$$

This implies that any real value $R \geq R_f(X, Y)$ is a *universal FF-CD achievable rate* of (X, Y) , namely, there exists a sequence of universal FF-CD codes with parameters $\{(n, M_n, e_n^{(X)}, e_n^{(Y)})\}_{n=1}^{\infty}$ that satisfies the conditions shown in Definition 2.

The following converse theorem indicates that the error exponent obtained in Theorem 2 is tight.

Theorem 3: Any sequence of FF-CD codes with parameters $\{(n, M_n, e_n^{(X)}, e_n^{(Y)})\}_{n=1}^{\infty}$ for the source (X, Y) must satisfy

$$e_n^{(X)} + e_n^{(Y)} \geq \frac{1}{2}(n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R+\epsilon_n)} D(Q_{XY} \| P_{XY}) \right\}$$

for any integer $n \geq 1$ and a given coding rate $R = 1/n \log M_n > 0$, where

$$\epsilon_n \stackrel{\text{def.}}{=} \frac{1}{n} \{ |\mathcal{X} \times \mathcal{Y}| \log(n+1) + 1 \}.$$

Proof: Note that the number of sequences to be decoded correctly for each decoder are at most $\exp(nR)$. Here, let us consider a joint type $Q_{XY} \in \overline{\mathcal{S}}_n(R + \epsilon_n)$. The definition of $\overline{\mathcal{S}}_n(R + \epsilon_n)$ and [4, Lemma 2.5] imply that for $(\mathbf{x}, \mathbf{y}) \in T_{Q_{XY}}^n$ we have

$$\begin{aligned} & \max\{|T_V^n(\mathbf{x})|, |T_W^n(\mathbf{y})|\} \\ & \geq (n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \\ & \quad \times \max\{\exp\{nH(V|Q_X)\}, \exp\{nH(W|Q_Y)\}\} \\ & \geq (n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \exp\{n(R + \epsilon_n)\} \\ & = 2 \exp(nR). \end{aligned}$$

Therefore, at least half of the sequences in the V-shell $T_V^n(\mathbf{x})$ will not be decoded correctly at the decoder $\varphi_{(2)}^n$, or at least half of sequences in the V-shell $T_W^n(\mathbf{y})$ will not be decoded correctly at the decoder $\varphi_{(1)}^n$. Thus, the sum of the error probabilities is bounded as

$$\begin{aligned} & e_n^{(X)} + e_n^{(Y)} \\ & \geq \frac{1}{2} \sum_{Q_{XY} \in \overline{\mathcal{S}}_n(R+\epsilon_n)} \Pr\{(X^n, Y^n) \in T_{Q_{XY}}\} \\ & \geq \frac{1}{2}(n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \\ & \quad \times \sum_{Q_{XY} \in \overline{\mathcal{S}}_n(R+\epsilon_n)} \exp\{-nD(Q_{XY} \| P_{XY})\} \\ & \quad (\because [4, Lemma 2.6]) \\ & \geq \frac{1}{2}(n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \\ & \quad \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R+\epsilon_n)} D(Q_{XY} \| P_{XY}) \right\} \end{aligned}$$

■

VI. VARIABLE-LENGTH CODING

This section discusses variable-length coding for the complementary delivery coding system, and shows an explicit construction of universal variable-length codes. The coding

scheme is similar to that of fixed-length codes, and also utilizes the coding tables defined in Section IV.

A. Formulation

Definition 4: (Fixed-to-variable complementary delivery (FV-CD) code)

A set $(\varphi^n, \widehat{\varphi}_{(1)}^n, \widehat{\varphi}_{(2)}^n)$ of an encoder and two decoders is an FV-CD code for the source (X, Y) if and only if

$$\begin{aligned} & \varphi^n : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow \mathcal{B}^* \\ & \widehat{\varphi}_{(1)}^n : \varphi^n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{Y}^n \rightarrow \mathcal{X}^n, \\ & \widehat{\varphi}_{(2)}^n : \varphi^n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{X}^n \rightarrow \mathcal{Y}^n, \\ & \Pr\{X^n \neq \widehat{X}^n\} = \Pr\{Y^n \neq \widehat{Y}^n\} = 0, \end{aligned}$$

where the image of φ^n is a prefix set.

Definition 5: (FV-CD-achievable rate)

R is an FV-CD-achievable rate of the source (X, Y) if and only if there exists a sequence of FV-CD codes $\{(\varphi^n, \widehat{\varphi}_{(1)}^n, \widehat{\varphi}_{(2)}^n)\}_{n=1}^{\infty}$ for the source (X, Y) such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} E[l(\varphi^n(X^n, Y^n))] \leq R,$$

where $l(\cdot) : \mathcal{B}^* \rightarrow \mathcal{R}$ is a length function.

Definition 6: (Inf FV-CD-achievable rate)

$$R_v(X, Y) = \inf\{R : R \text{ is an FV-CD-achievable rate of } (X, Y)\}.$$

B. Code construction

We can construct universal FV-CD codes in a similar manner to universal FF-CD codes. Note that the coding rate depends on the type of sequence pair to be encoded, whereas the coding rate is fixed beforehand for fixed-length coding. The coding scheme is described as follows:

[Encoding]

- 1) Create a coding table for each joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ in the same way as Step 2 of Section IV.
- 2) Mark cells that correspond to sequence pairs $(\mathbf{x}, \mathbf{y}) \in T_{Q_{XY}}$.
- 3) Fill the marked cells on the coding table with different $\max\{|T_V^n(\mathbf{x})|, |T_W^n(\mathbf{y})|\}$ symbols such that each symbol occurs at most once in each row and at most once in each column, where $\mathbf{x} \in T_{Q_X}^n$, $\mathbf{y} \in T_{Q_Y}^n$.
- 4) For a given pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, the number (index) assigned to the joint type Q_{XY} of (\mathbf{x}, \mathbf{y}) is the first part of the codeword, and the symbol filling the cell of (\mathbf{x}, \mathbf{y}) in the coding table of Q_{XY} is determined as the second part of the codeword.

[Decoding]

Decoding can be accomplished in almost the same way as the fixed-length coding. Note that the decoder can always find the coding table used in the encoding scheme.

C. Coding theorems

We begin by showing a theorem for (non-universal) variable-length coding, which indicates that the inf coding rate of variable-length coding is the same as that of fixed-length coding.

Theorem 4: (Coding theorem of FV-CD code)

$$R_v(X, Y) = R_f(X, Y) = \max\{H(X|Y), H(Y|X)\}.$$

Proof: (Converse part) We can prove it in a similar manner to that for fixed-length coding.

(Direct part) We can apply a sequence of *achievable* FF-CD codes. The encoder φ^n assigns the same codeword as that of the fixed-length code to a sequence pair $(x, y) \in \mathcal{X}^n \times \mathcal{Y}^n$ that is correctly reproduced by the fixed-length code. Otherwise, the encoder sends the sequence pair itself as a codeword. ■

The following direct theorem for universal coding indicates that the coding scheme presented in the previous subsection can achieve the inf achievable rate clarified in Theorem 4.

Theorem 5: There exists a sequence of universal FV-CD codes $\{(\varphi^n, \widehat{\varphi}_{(1)}^n, \widehat{\varphi}_{(2)}^n)\}_{n=1}^{\infty}$ such that for any integer $n \geq 1$ and any source (X, Y) , the overflow probability $\bar{p}_n(R)$ (the probability that the length of a codeword exceeds a given real number $R > 0$) is bounded as

$$\begin{aligned} \bar{p}_n(R) &\stackrel{\text{def.}}{=} \Pr \{l(\varphi^n(X^n, Y^n)) > n(R + \epsilon_n)\} \\ &\leq (n+1)^{|\mathcal{X} \times \mathcal{Y}|} \\ &\quad \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R)} D(Q_{XY} \| P_{XY}) \right\}, \end{aligned}$$

where ϵ_n is defined in Theorem 3. This implies that there exists a sequence of universal FV-CD codes $\{(\varphi^n, \widehat{\varphi}_{(1)}^n, \widehat{\varphi}_{(2)}^n)\}_{n=1}^{\infty}$ that satisfies

$$\limsup_{n \rightarrow \infty} \frac{1}{n} l(\varphi^n(X^n, Y^n)) \leq R_v(X, Y) \quad \text{a.s.} \quad (1)$$

Proof: The overflow probability can be obtained in the same way as an upperbound of the error probability of the FF-CD code, which has been shown in the proof of Theorem 2. Thus, we have

$$\sum_{n=1}^{\infty} \Pr \left\{ \frac{1}{n} l(\varphi^n(X^n, Y^n)) > R_v(X, Y) + \delta \right\} < \infty$$

for a given $\delta > 0$. From Borel-Cantelli's lemma [5, Lemma 4.6.3], we immediately obtain Eq. (1). ■

The following converse theorem for variable-length coding can be easily obtained in the same way as Theorem 3.

Theorem 6: Any sequence of FV-CD codes $\{(\varphi^n, \widehat{\varphi}_{(1)}^n, \widehat{\varphi}_{(2)}^n)\}_{n=1}^{\infty}$ for the source (X, Y) must satisfy

$$\begin{aligned} \bar{p}_n(R) &\geq (n+1)^{-|\mathcal{X} \times \mathcal{Y}|} \\ &\quad \times \exp \left\{ -n \min_{Q_{XY} \in \overline{\mathcal{S}}_n(R+\epsilon_n)} D(Q_{XY} \| P_{XY}) \right\}. \end{aligned}$$

for a given real number $R > 0$ and any integer $n \geq 1$, where ϵ_n is defined in Theorem 3.

VII. CONCLUDING REMARKS

We presented an explicit construction of universal fixed-length codes for the complementary delivery coding system. We clarified that the error exponent achieved by the proposed coding scheme is optimal. Next, we applied the coding scheme to construction of universal variable-length codes. We clarified that there exists a universal code such that the codeword length converges to the minimum achievable rate almost surely, and that the exponent of the overflow probability attained by the proposed coding scheme is optimal. This paper dealt with the case where the number of decoders was two, and therefore constructing universal codes for cases where the number of decoders is more than three still remains as an open problem.

ACKNOWLEDGEMENTS

The authors thank Prof. Ryutaroh Matsumoto of Tokyo Institute of Technology, Dr. Yoshinobu Tonomura, Dr. Hiromi Nakaiwa, Dr. Tatsuto Takeuchi, Dr. Shoji Makino and Dr. Junji Yamato of NTT Communication Science Laboratories for their help. The first author contributed to this work during his doctoral program at Tokyo Institute of Technology.

REFERENCES

- [1] N. L. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory*. Oxford University Press, 1976.
- [2] I. Csiszár, "Linear codes for source and source networks: Error exponents, universal coding," *IEEE Trans. Inf. Theory*, vol. 28, no. 4, pp. 585–592, July 1982.
- [3] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, March 1980.
- [4] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. New York: Academic Press, 1981.
- [5] R. M. Gray, *Probability, Random Processes, Ergodic Properties*. New York: Springer-Verlag, 1988.
- [6] D. König, "Graphok és alkalmazásuk a determinánsok és a halmazok elméletére," *Mathematikai és Természettudományi Értesítő*, vol. 34, pp. 104–119, 1916, (in Hungarian).
- [7] J. Körner and K. Marton, "Images of a set via two channels and their role in multi-user communication," *IEEE Trans. Inf. Theory*, vol. 23, no. 6, pp. 751–761, November 1975.
- [8] A. Kimura and T. Uyematsu, "Multiterminal source coding with complementary delivery," in *Proc. International Symposium on Information Theory and its Applications (ISITA)*, October 2006, pp. 189–194.
- [9] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, November 1994.
- [10] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 179–182, March 1977.
- [11] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [12] T. Uyematsu, "An algebraic construction of codes for Slepian-Wolf source networks," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3082–3088, November 2001.
- [13] F. M. J. Willems, J. K. Wolf, and A. D. Wyner, "Communicating via a processing broadcast satellite," in *Proc. of the 1989 IEEE/CAM Information Theory Workshop*, June 1989.
- [14] 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.
- [15] A. D. Wyner, J. K. Wolf, and F. M. J. Willems, "Communicating via a processing broadcast satellite," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1243–1249, June 2002.