# Universal coding for correlated sources with complementary delivery

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

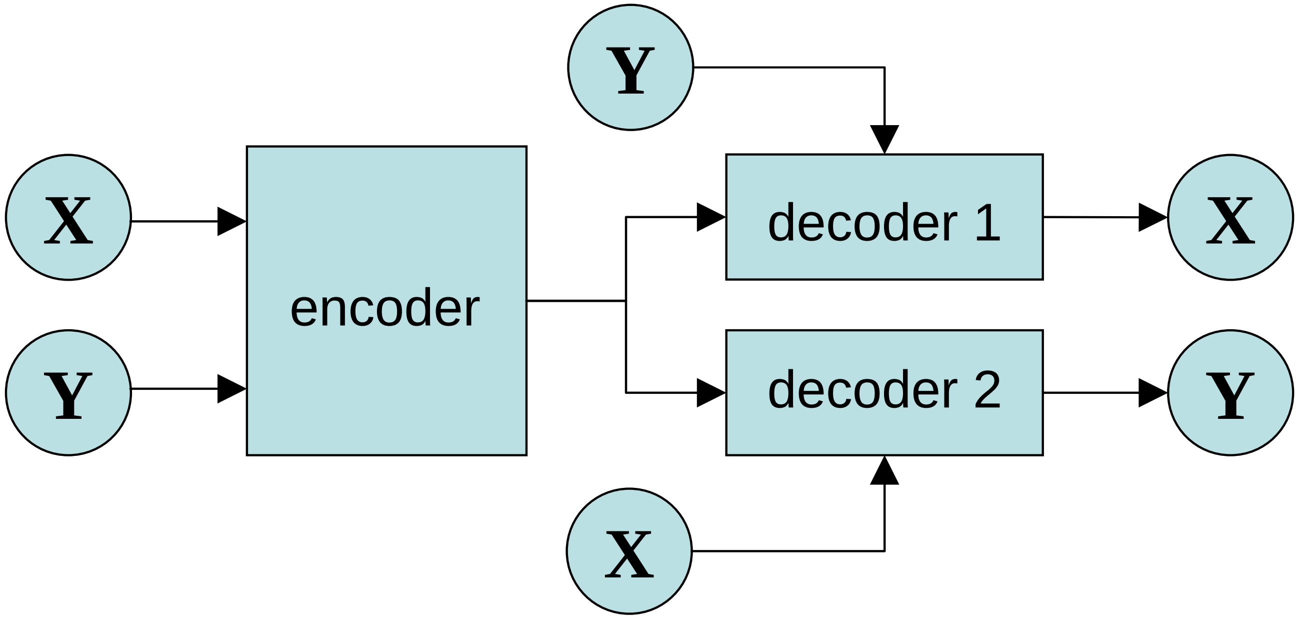
[†] Tokyo Institute of Technology

[*] NTT Communication Science Laboratories, NTT Corporation

[‡] Wakayama University

# Abstract

- Universal source coding problem for multiterminal coding system shown below
- Explicit constructions of universal codes
  - Fixed-length lossless codes
  - Variable-length lossless codes



# Outline

1. Introduction
2. Preliminaries
3. Code construction
4. Coding theorems
5. Summary

# Introduction

1. Introduction
2. Preliminaries
3. Code construction
4. Coding theorems
5. Summary

# Introduction

- Slepian-Wolf (1973)
  - beginning of multiterminal source coding
- Csiszar-Korner (1980)
  - first investigated the problem of universal coding for the Slepian-Wolf coding system
  - showed the existence of universal codes
- Subsequent research mainly focused on the Slepian-Wolf coding system
  - Csiszar (1982), Ahlswede-Dueck (1982),Oohama (1996), Uyematsu (2001)

# Main contribution

- Universal source coding problem for another multiterminal coding system
  - Complementary delivery coding system ("broadcast network" in the plenary talk by Prof.Effros)
  - Coding theorem was clarified by Csiszar-Korner (1980)
- Explicit constructions of universal codes
  - Fixed-length codes & variable-length codes
  - Utilize method of types and graph-theoretical techniques.

# Preliminaries

# Basic definitions

- Alphabets:

  $\mathcal{X}, \mathcal{Y}$ : finite sets, ($\mathcal{B}$ : binary set)

  $\mathcal{B}^*$ : set of all finite sequences in $\mathcal{B}$,

  $\mathcal{I}_M = \{1, 2, \cdots, M\}$

- Sequences

  Member of $\mathcal{X}^n$: $x^n = (x_1, x_2, \cdots, x_n)$,

  Sequences: boldface letters, i.e., $\boldsymbol{x} \in \mathcal{X}^n$

# Basic definitions

- Probability distributions

  $\mathcal{P}(\mathcal{X})$: set of all probability distributions on $\mathcal{X}$,

  $\mathcal{P}(\mathcal{X}|P_Y)$:

  set of all probability distributions on $\mathcal{X}$

  given distribution $P_Y \in \mathcal{P}(\mathcal{Y})$.

- Sources

  $X$: discrete memoryless source (DMS) taking values

  in $\mathcal{X}$ with generic distribution $P_X \in \mathcal{P}(\mathcal{X})$.

- Entropy

  $H(X), H(X, Y), H(H|Y)$

# Basic definitions

- Types of sequences

$Q_{\boldsymbol{x}}$: type of sequence $\boldsymbol{x}$

$\mathcal{P}_n(\mathcal{X})$: set of types of sequences in $\mathcal{X}^n$.

$\mathcal{V}_n(\mathcal{Y}|Q)$:

set of all stochastic matrices $V : \mathcal{X} \rightarrow \mathcal{Y}$

s.t. for a given $Q \in \mathcal{P}_n(\mathcal{X})$ there exists

$(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ that satisfy $Q_{\boldsymbol{x},\boldsymbol{y}}(x, y) = Q(x)V(y|x)$.

$T_Q^n$: set of sequences with type $Q \in \mathcal{P}_n(\mathcal{X})$

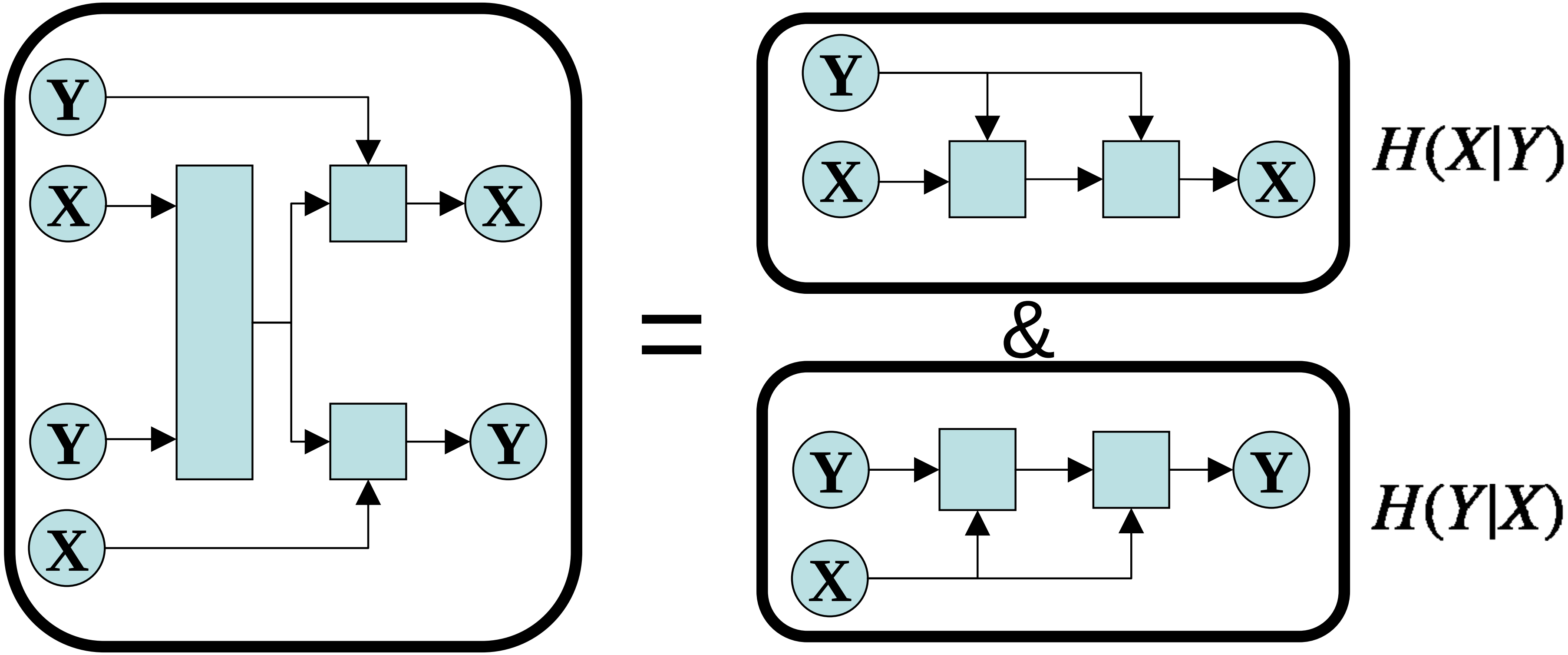
$T_V^n(\boldsymbol{x})$: set of sequences with stochastic matrix

$V \in \mathcal{V}_n(\mathcal{Y}|Q_{\boldsymbol{x}})$ for a given $\boldsymbol{x} \in \mathcal{X}^n$

# Previous result

**Theorem 1** (Csiszár-Körner 1980).

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



# Code construction

1. Introduction
2. Preliminaries
3. Code construction
4. Coding theorems
5. Summary

# Code construction

| | $y_1$ | $y_2$ | $y_3$ | $y_4$ | $y_5$ | $(\in T^n_{Q_Y})$ |
|---|---|---|---|---|---|---|
| $x_1$ | ○ | ○ | ○ | | | |
| $x_2$ | ○ | | | ○ | ○ | |
| $x_3$ | | ○ | ○ | ○ | | $(\in T^n_{Q_{XY}})$ |
| $x_4$ | ○ | ○ | | | ○ | |
| $x_5$ | | | ○ | ○ | ○ | |
| $(\in T^n_{Q_X})$ | | | | | | |

$$\left(\begin{array}{l} Q_{XY} = Q_X V = Q_Y W, \\ Q_X \in \mathcal{P}_n(\mathcal{X}), \\ Q_Y \in \mathcal{P}_n(\mathcal{Y}), \\ V \in \mathcal{V}_n(\mathcal{Y}|Q_X), \\ W \in \mathcal{V}_n(\mathcal{X}|Q_Y) \end{array}\right)$$

Create a coding table for each joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$

- each row corresponds to a sequence $\boldsymbol{x} \in T^n_{Q_X}$
- each column corresponds to a sequence $\boldsymbol{y} \in T^n_{Q_Y}$
- mark cells corresponding to sequences $(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{Q_{XY}}$

# Code construction

- To reproduce all sequence pairs at decoders…

| | $y_1$ | $y_2$ | $y_3$ | $y_4$ | $y_5$ |
|---|---|---|---|---|---|
| $x_1$ | ① | ② | ③ | | |
| $x_2$ | ② | | | ○ | ○ |
| $x_3$ | | ○ | ○ | ○ | |
| $x_4$ | ③ | ○ | | | ○ |
| $x_5$ | | | ○ | ○ | ○ |

For a given $\boldsymbol{x}_i$, we must distinguish each $\boldsymbol{y}_j$ that satisfy $(\boldsymbol{x}_i, \boldsymbol{y}_j) \in T^n_{Q_{XY}}$. (about $\exp(nH(V|Q_X))$ sequences)

For a given $\boldsymbol{y}_j$, we must distinguish each $\boldsymbol{x}_i$ such that $(\boldsymbol{x}_i, \boldsymbol{y}_j) \in T^n_{Q_{XY}}$. (about $\exp(nH(W|Q_Y))$ sequences)

$$\left(\begin{array}{l} Q_{XY} = Q_X V = Q_Y W, \\ Q_X \in \mathcal{P}_n(\mathcal{X}), \\ Q_Y \in \mathcal{P}_n(\mathcal{Y}), \\ V \in \mathcal{V}_n(\mathcal{Y}|Q_X), \\ W \in \mathcal{V}_n(\mathcal{X}|Q_Y) \end{array}\right)$$

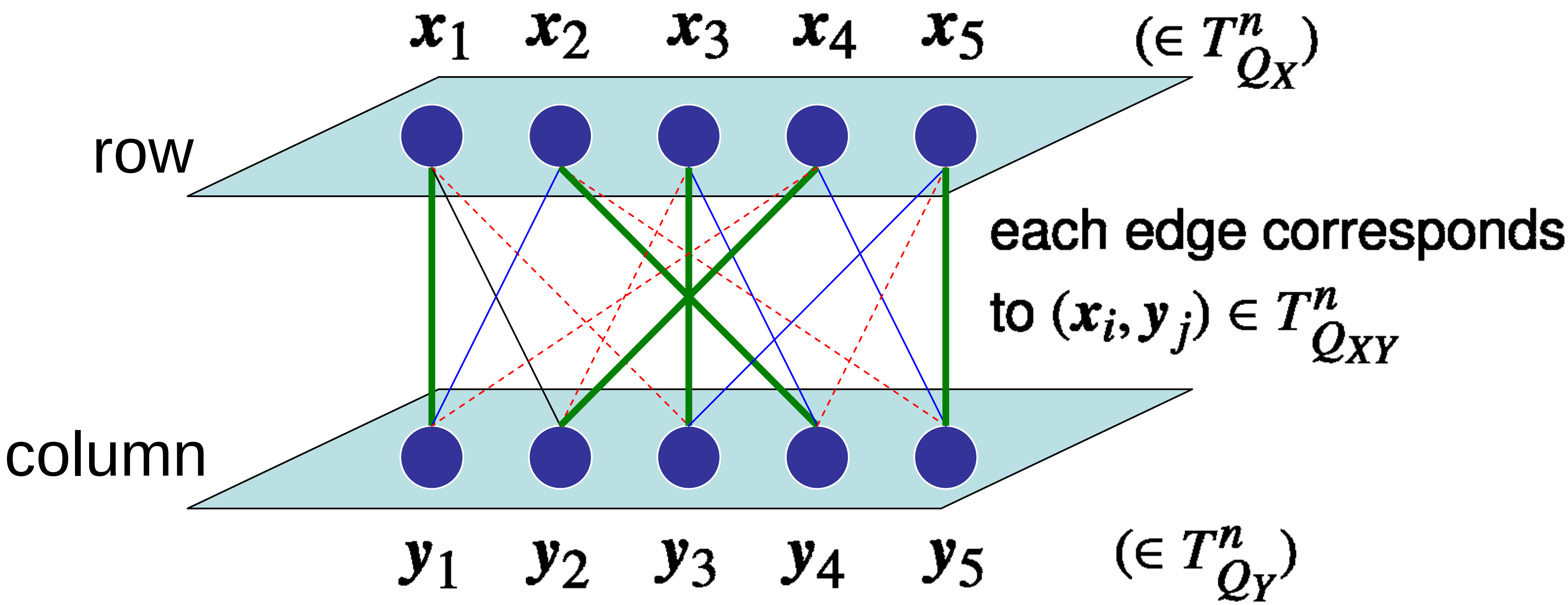
# Code construction

| | $y_1$ | $y_2$ | $y_3$ | $y_4$ | $y_5$ |
|---|---|---|---|---|---|
| $x_1$ | ① | ② | ③ | | |
| $x_2$ | ② | | | ① | ③ |
| $x_3$ | | ③ | ① | ② | |
| $x_4$ | ③ | ① | | | ② |
| $x_5$ | | | ② | ③ | ① |

About $\exp\{n \max(H(W|Q_Y), H(W|Q_Y))\}$ symbols are needed to distinguish every $(x_i, y_j) \in Q_{XY}$.

# Equivalence to graph coloring

- The table can be converted to a bipartite graph
- Encoding is equivalent to edge coloring of the bipartite graph



# Necessary number of symbols

**Lemma 3** (König). *If a graph* $G$ *is bipartite, then the edge chromatic number (i.e. the number of symbols needed for edge coloring)* $\chi'(G)$ *of* $G$ *is given by*

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

*where* $\Delta(G)$ *is the degree of the graph* $G$*.*

**Lemma .** *For a given joint type* $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$*, its coding graph* $G(Q_{XY})$ *has a degree*

$$\Delta(G(Q_{XY})) \quad \leq \quad \exp[n \max\{H(V|Q_X), H(W|Q_Y)\}].$$

# Coding scheme

- Fixed-length coding:

Create a table for each joint type $Q_{XY} \in \mathcal{T}_n(R)$.

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

- Variable-length coding

Create a table for every joint type $Q_{XY}$. The codeword length of an input sequence pair depends on its type.

# Coding theorems

1. Introduction
2. Preliminaries
3. Code construction
4. <u>Coding theorems</u>
5. Summary

# Fixed-length coding theorem

**Theorem** (Coding theorem for universal f-CD codes). *For any real number $R > R_f(X, Y)$, there exists a universal f-CD code $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ such that for any source $(X, Y)$*

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R,$$

$$\lim_{n\to\infty} -\frac{1}{n} \log(e_n^{(1)} + e_n^{(2)}) = \min_{Q_{XY}\in\mathcal{T}^c(R)} D(Q_{XY}\|P_{XY}) > 0,$$

$$\limsup_{n\to\infty} -\frac{1}{n} \log(1 - e_n^{(1)} - e_n^{(2)}) \le \min_{Q_{XY}\in\mathcal{T}(R)} D(Q_{XY}\|P_{XY}),$$

$$e_n^{(1)} = \Pr\{X^n \ne \widehat{X}^n\}, \quad e_n^{(2)} = \Pr\{Y^n \ne \widehat{Y}^n\}.$$

# Variable-length coding theorem

**Theorem** (Coding theorem for universal v-CD codes).

*There exists a universal v-CD code* $\{(\varphi_n, \widehat{\varphi}_n^{(1)}, \widehat{\varphi}_n^{(2)})\}_{n=1}^{\infty}$ *such that for any source* $(X, Y)$

$$\lim_{n\to\infty} \frac{1}{n} l(\varphi_n(X^n, Y^n)) = R_v(X, Y) = R_f(X, Y) \quad a.s.$$

$$\lim_{n\to\infty} -\frac{1}{n} \log \overline{\rho}_n(R) = \min_{Q_{XY}\in\mathcal{T}^c(R)} D(Q_{XY}\|P_{XY}) > 0$$

$$\limsup_{n\to\infty} -\frac{1}{n} \log \underline{\rho}_n(R) \leq \min_{Q_{XY}\in\mathcal{T}(R)} D(Q_{XY}\|P_{XY})$$

$$\overline{\rho}_n(R) = \Pr\{l(\varphi_n(X^n, Y^n)) > R\},$$

$$\underline{\rho}_n(R) = \Pr\{l(\varphi_n(X^n, Y^n)) < R\}.$$

# Summary

# Summary

- Investigated universal coding problem for complementary delivery coding system
  - Presented explicit constructions of universal codes
  - Established a coding theorem for each problem

- Future work
  - More than 3 messages and/or more than 3 decoders
    → will be presented in the recent result session
  - Practical universal coding schemes
  - Lossy universal coding schemes
    → will be presented in another conference

# Thank you very much