# *Universal source coding for complementary delivery*

Akisato Kimura[*1,*2], Tomohiko Uyematsu[*2], Shigeaki Kuzuoka[*2]

[*1] Media Information Laboratory,

NTT Communication Science Laboratories,

NTT Corporation

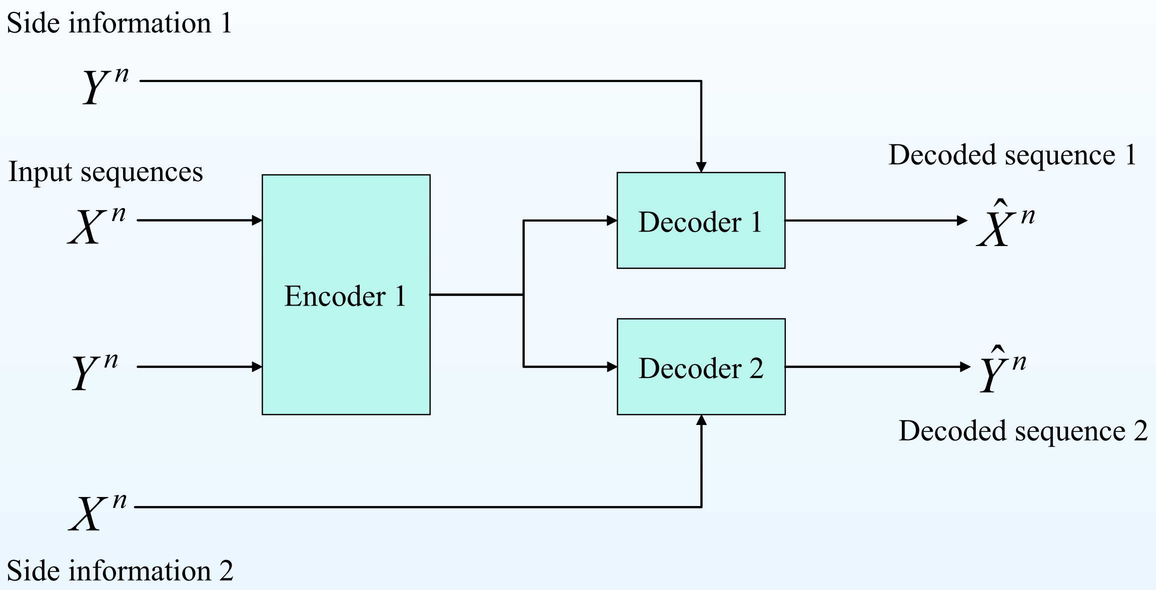
[*2] Department of Communications and Integrated Systems,

Graduate School of Science and Engineering,

Tokyo Institute of Technology

# Abstract

Present explicit constructions of universal codes for the following multiterminal source coding system



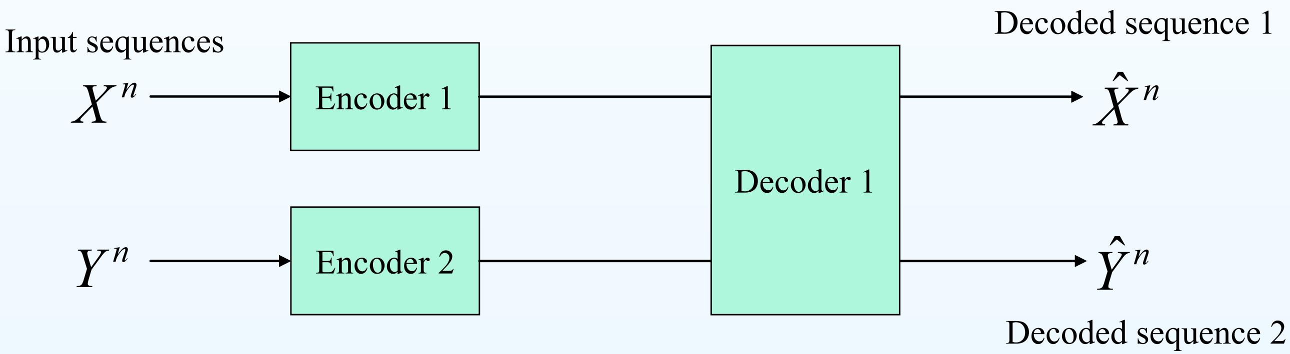

- Fixed-to-fixed length (FF) and fixed-to-variable length (FV) lossless coding schemes are considered.
- Key techniques: type theory, graph theory

## Contents

- Introduction
- FF universal coding
  - Coding scheme
  - Lemmas that support the coding scheme
- FF coding theorems
- FV unversal coding (See the proceedings.)
  - Coding scheme
  - Coding theorems
- Conclusion

# Introduction

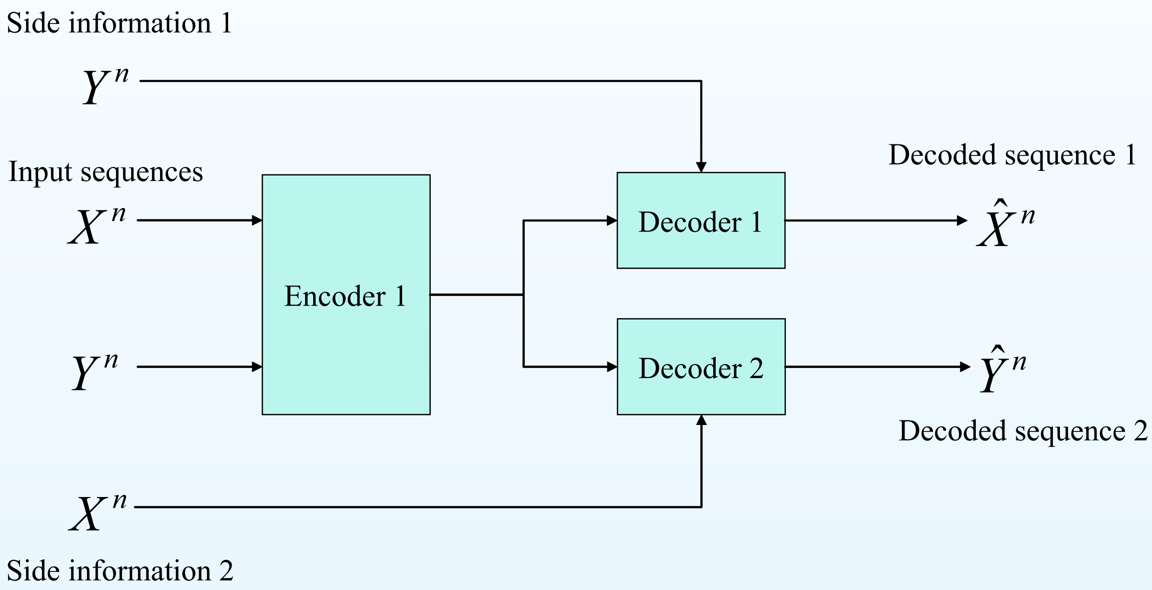
- Coding problems for correlated information sources were originally investigated by Slepian and Wolf [SW73]




- The problem of universal coding for this systems was first investigated by Csiszár and Körner [CK80].
- Subsequent work has mainly focused on the Slepian-Wolf coding system
  - Since it appears to be difficult to construct universal codes for most of the other coding systems.

# Main contributions

Construct universal codes for the complementary delivery coding system




- Our previous report [KU06]: Considered the lossy configuration and clarified the rate-distortion function.
- This report: Consider a lossless configuration and construct universal codes.

# Preliminaries (1/2)

- $\mathcal{X}, \mathcal{Y}$: finite alphabets, $\mathcal{B}$: a binary set,
  $\mathcal{B}^*$: a set of all finite sequences in $\mathcal{B}$,
  $|\mathcal{X}|$: cardinality of $\mathcal{X}$.
- $x^n = (x_1, x_2, \cdots, x_n)$: member of $\mathcal{X}^n$.
  - When the dimension is clear from the content, vectors will be denoted by boldface letters, i.e., $\boldsymbol{x} \in \mathcal{X}^n$
- $\mathcal{M}(\mathcal{X})$: the set of all probability distributions on $\mathcal{X}$,
  $\mathcal{M}(\mathcal{X}|P_Y)$: the set of all probability distributions on $\mathcal{X}$ given a distribution $P_Y \in \mathcal{M}(\mathcal{Y})$
- $X$: a discrete memoryless source taking values in $\mathcal{X}$ with a generic distribution $P_X \in \mathcal{M}(\mathcal{X})$

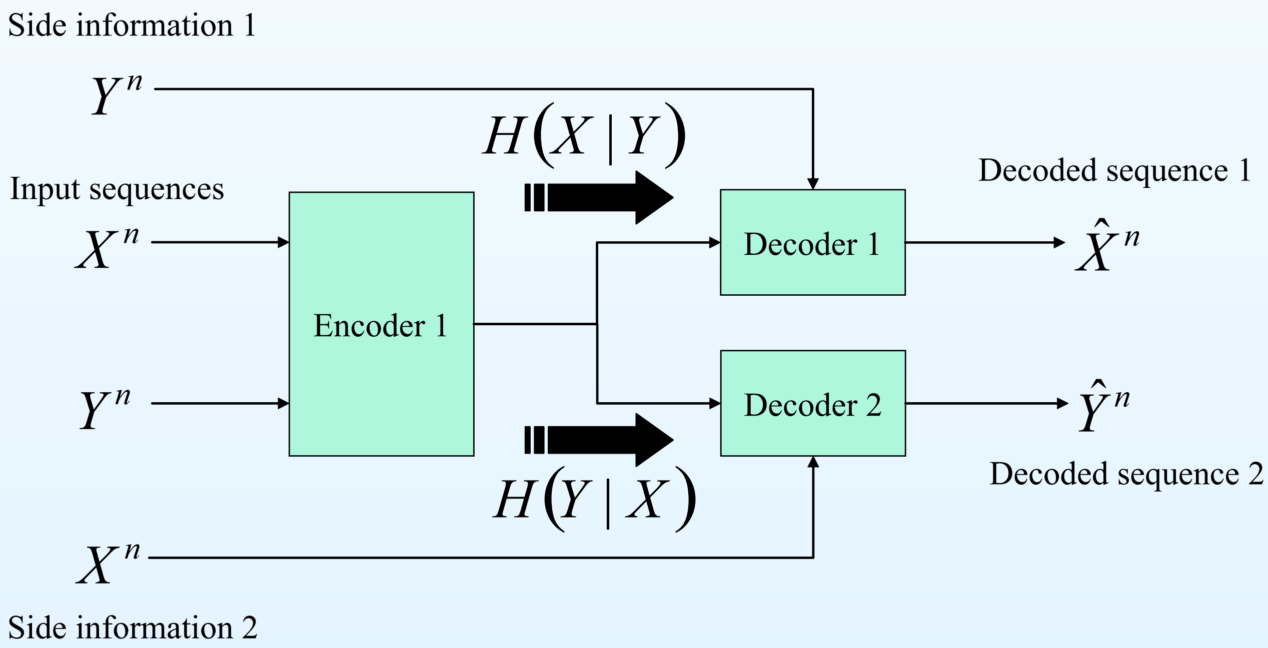
## Preliminaries (2/2)

- $H(X)$, $H(P_X)$: entropy of $X$,
  $H(X|Y)$, $H(P_{X|Y}|P_Y)$: conditional entropy of $X$ given $Y$,
  $D(P_X \| P_Y)$: divergence between $P_X$ and $P_Y$.
- $T_Q^n$: the set with sequences of type $Q$,
  $T_V^n(\boldsymbol{x})$: the set of sequences with conditional type $V : \mathcal{X} \to \mathcal{Y}$ for a given $\boldsymbol{x} \in \mathcal{X}^n$.
- $\mathcal{P}_n(\mathcal{X})$: the set of types of sequences in $\mathcal{X}^n$,
  $\mathcal{V}_n(\mathcal{Y}|Q)$: the set of conditional types of sequences in $\mathcal{Y}^n$ for a given $Q \in \mathcal{P}_n(\mathcal{X})$ s.t. $T_V^n(\boldsymbol{x}) \neq \emptyset$ for any $\boldsymbol{x} \in T_Q^n$.

# Minimum achievable rate

**Theorem 1.** *[WWW02]*

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



# Contents

# FF Coding scheme (1/4)

1. *(Type selection)*
   For a given coding rate $R > 0$, select joint types $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ that satisfy

   $$\max\{H(W|Q_Y), H(V|Q_X)\} \leq R,$$

   where $Q_X \in \mathcal{P}_n(\mathcal{X})$, $Q_Y \in \mathcal{P}_n(\mathcal{Y})$, $V \in \mathcal{V}_n(\mathcal{Y}|Q_X)$ and $W \in \mathcal{V}_n(\mathcal{X}|Q_Y)$ that satisfy $Q_X V = Q_Y W = Q_{XY}$.

   If $\max\{H(W|Q_Y), H(V|Q_X)\} > R$
   $\longrightarrow$ codewords will not be assigned.

# FF Coding scheme (2/4)

2. *(Table creation)*
   Create a coding table for each joint type $Q_{XY}$ selected in Step 1. Each row of the coding table corresponds to a sequence $\boldsymbol{x} \in T^n_{Q_X}$, and each column corresponds to a sequence $\boldsymbol{y} \in T^n_{Q_Y}$.

| $Q_{XY}$ | $\boldsymbol{y}_1$ | $\boldsymbol{y}_2$ | $\boldsymbol{y}_3$ | $\boldsymbol{y}_4$ | $\boldsymbol{y}_5$ |
|---|---|---|---|---|---|
| $\boldsymbol{x}_1$ | | | | | |
| $\boldsymbol{x}_2$ | | | | | |
| $\boldsymbol{x}_3$ | | | | | |
| $\boldsymbol{x}_4$ | | | | | |
| $\boldsymbol{x}_5$ | | | | | |

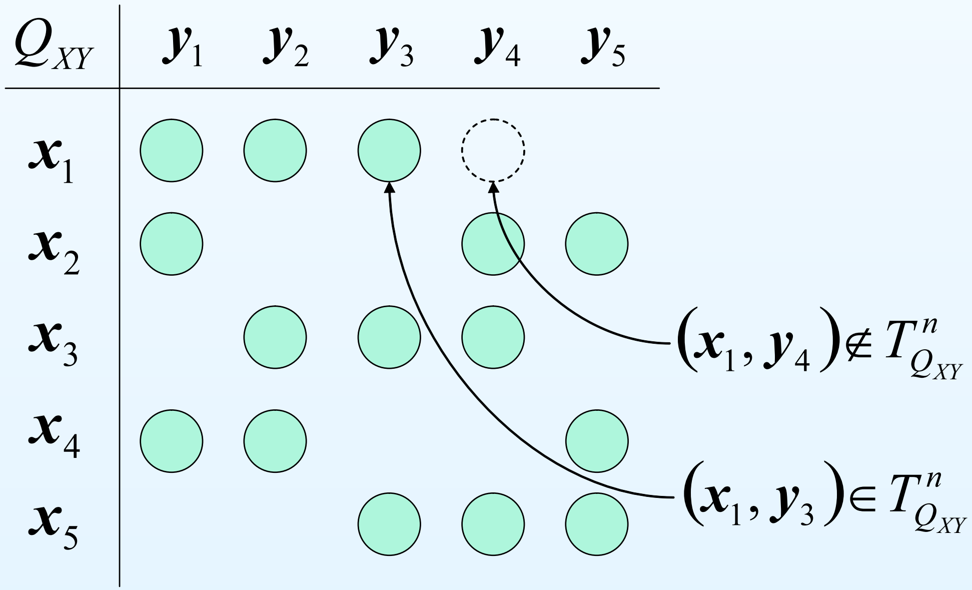
$\Leftarrow T^n_{Q_Y}$

$\Uparrow$

$T^n_{Q_X}$

# FF Coding scheme (3/4)

3. *(Cell marking)*
   Mark cells that correspond to sequence pairs $(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{Q_{XY}}$
   Codewords will be given only to sequence pairs that correspond to marked cells.



# Coding scheme (4/4)

4. *(Codeword assignment)*
   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.
5. For a given $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$, the codeword is

index assigned to the type of the sequence pair +

| | $\boldsymbol{y}_1$ | $\boldsymbol{y}_2$ | $\boldsymbol{y}_3$ | $\boldsymbol{y}_4$ | $\boldsymbol{y}_5$ |
|---|---|---|---|---|---|
| $\boldsymbol{x}_1$ | 1 | 2 | 3 | | |
| $\boldsymbol{x}_2$ | 2 | | | 1 | 3 |
| $\boldsymbol{x}_3$ | | 3 | 1 | 2 | |
| $\boldsymbol{x}_4$ | 3 | 1 | | | 2 |
| $\boldsymbol{x}_5$ | | | 2 | 3 | 1 |

## Lemmas that support the coding scheme (1/3)

**Lemma 1.** *For a given coding table of a joint type* $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$*,*

- *the number of marked cells in every row of the coding table is a constant* $N_x$ s.t. $N_x \leq \exp(nR)$*,*
- *the number of marked cells in every column of the coding table is also a constant* $N_y$ s.t. $N_y \leq \exp(nR)$

*both of which depend solely on the joint type* $Q_{XY}$*.*

**Sketch of the proof.**
For a given type $Q_{XY}$ and $(\boldsymbol{x}, \boldsymbol{y}) \in T^n_{Q_{XY}}$, both $N_x = |T^n_V(\boldsymbol{x})|$ and $N_y = |T^n_W(\boldsymbol{y})|$ are constant. □

# Lemmas that support the coding scheme (2/3)

**Lemma 2.** *For given positive 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 a $m_x \times m_y$ table filled with $\max(n_x, n_y)$ different symbols such that*

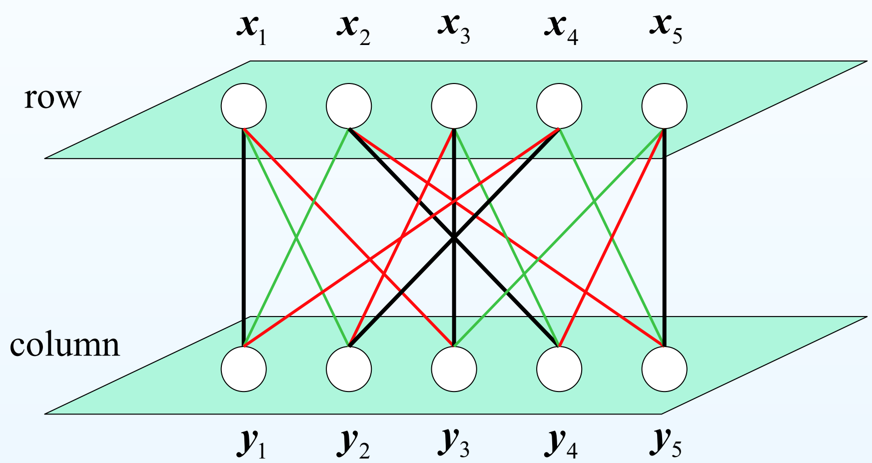
- *at most $n_y$ cells are filled with a certain symbol for each row (blank cells are possible),*
- *at most $n_x$ cells are filled with a certain symbol for each column (blank cells are possible),*
- *each symbol occurs at most once in each row and at most once in each column.*

# Lemmas that support the coding scheme (3/3)

**Sketch of the proof.**

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

$\Longleftrightarrow$




The following lemma ensures the existence of the above bipartite graph.

**Lemma 3.** (König [K16])
*If a graph $G$ is bipartite, the minimum number of colors necessary for edge coloring of the graph $G$ equals the maximum degree of $G$.*

□

# Contents

- Introduction
- FF coding scheme
  - Coding scheme
  - Lemmas that support the coding scheme
- FF coding theorems
- Conclusion

# FF coding theorems (1/2)

**Theorem 2.** *(Direct part)*
*For a given real number* $R > 0$*, there exists a sequence of universal FF codes* $\{(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ *such that for any integer* $n \geq 1$ *and any source* $(X, Y)$

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

*where*

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

# FF coding theorems (2/2)

**Theorem 3.** *(Converse part)*
*Any sequence of FF codes* $\{(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ *for the source* $(X, Y)$ *must satisfy*

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

*for any integer* $n \geq 1$ *and a given coding rate* $R = 1/n \log M_n > 0$*, where* $\epsilon_n \to 0$ $(n \to \infty)$*.*
*This theorem implies that the error exponent obtained in Theorem 2 is asymptotically optimal.*

# Conclusions

Investigated a universal coding problem for the complementary delivery coding system.

- Presented an explicit construction of universal FF codes
  - The proposed coding scheme can achieve the optimal error exponent.
- Can apply the FF coding scheme to construction of FV universal codes
  - Codeword lengths converge to the minimum achievable rate almost surely.

Future work

- Extend the results to other classes of information sources
- Construct universal codes for the lossy configuration

# Thank you

## References


[CK80] I. Csiszár and J. Körner. Towards a general theory of source networks. *IEEE Trans. Inform. Theory*, 26(2):155–165, March 1980.

[K1̣6] D. König. Graphok és alkalmazásuk a determinánsok és a halmazok elméletére. *Mathematikai és Természettudományi Értesitö*, 34:104–119, 1916. (in Hungarian).

[KU06] A. Kimura and T. Uyematsu. Multiterminal source coding with complementary delivery. In *Proc. International Symposium on Information Theory and its Applications (ISITA)*, pages 189–194, October 2006.

[SW73] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, 19(4):471–480, July 1973.

[WWW02] A. D. Wyner, J. K. Wolf, and F. M. J. Willems. Communicating via a processing broadcast satellite. *IEEE Trans. Inform. Theory*, 48(6):1243–1249, June 2002.


Some materials will be available at *http://www.brl.ntt.co.jp/people/akisato/*

# Appendix

# Variable-length coding

Basically the same as the FF coding scheme. The codeword is

index assigned to the type +

| | $\boldsymbol{y}_1$ | $\boldsymbol{y}_2$ | $\boldsymbol{y}_3$ | $\boldsymbol{y}_4$ | $\boldsymbol{y}_5$ |
|---|---|---|---|---|---|
| $\boldsymbol{x}_1$ | 1 | 2 | 3 | | |
| $\boldsymbol{x}_2$ | 2 | | | 1 | 3 |
| $\boldsymbol{x}_3$ | | 3 | 1 | 2 | |
| $\boldsymbol{x}_4$ | 3 | 1 | | | 2 |
| $\boldsymbol{x}_5$ | | | 2 | 3 | 1 |

- Coding tables are created for every type $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, and therefore codewords are assigned to every sequence pair.
- $\max\{\exp(nH(V|Q_X)), \exp(nH(W|Q_Y))\}$ different symbols are necessary to fill the coding table with.

# FV coding theorems (1/2)

**Theorem 4.** *(Direct part)*
*There exists a sequence of universal FV codes* $\{(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ *such that for any integer* $n \geq 1$ *and any source* $(X, Y)$ *the overflow probability is bounded as*

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

*where* $\epsilon_n \to 0$ $(n \to \infty)$*. This implies that there exists a sequence of universal FV codes* $\{(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ *that satisfies*

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

# FV coding theorems (2/2)

**Theorem 5.** *(Converse part)*
*Any sequence of FV codes* $\{(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)})\}_{n=1}^{\infty}$ *for the source* $(X, Y)$ *must satisfy*

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

*for a given real number* $R > 0$ *and any integer* $n \geq 1$*, where* $\epsilon_n \to 0$ $(n \to \infty)$*.*