Multiterminal source coding with complementary delivering

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

akisato@eye.brl.ntt.co.jp, uematsu@it.ss.titech.ac.jp

[*1] Department of Communications and Integrated Systems,

Graduate School of Science and Engineering,

Tokyo Institute of Technology

[*2] Media Information Laboratory,

NTT Communication Science Laboratories,

NTT Corporation

Abstract

- Investigate a coding problem for correlated information sources, where
  - messages from two correlated sources are jointly encoded
  - each decoder has access to one of two messages to reproduce the other message
- Clarify the rate-distortion function for this problem

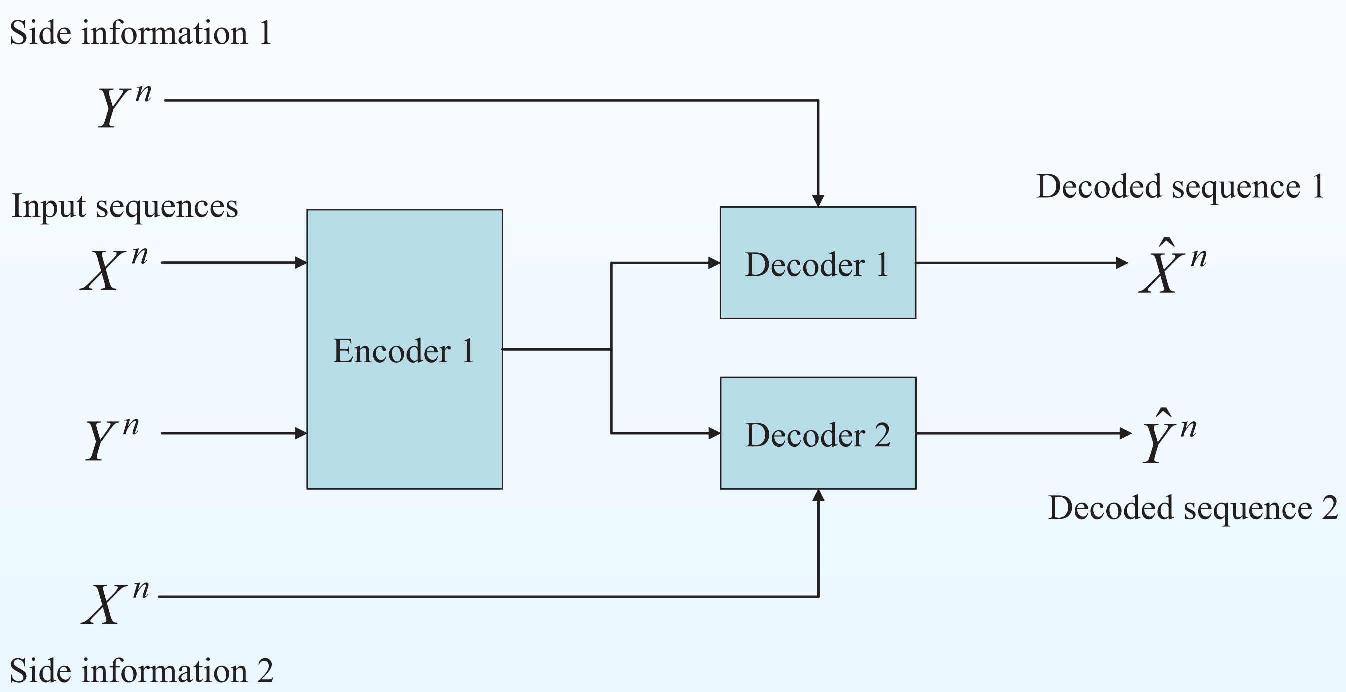
Introduction (1/2)

- Coding problems for correlated information sources were originally investigated by Slepian and Wolf [SW73]
- A huge number of coding problems have been considered.
  - Corresponding rate-distortion coding problems (e.g. Wyner-Ziv [WZ76], Berger-Tung [Tun78])
  - Various coding problems derived from the Slepian-Wolf coding problem (e.g. Wyner [Wyn75], Sgarro [Sga77], and Körner-Marton [KM75])
- Including the above studies, the main focus has been on coding problems with
  - Separate encoding : each message is separately encoded
  - Joint decoding : some of codewords are sent to a decoder and decoded at once

Introduction (2/2)

- The situation we investigate here involves
  - Joint encoding : messages from several sources are encoded at once
  - Separate decoding : each message is separately decoded
- Separate decoding processes have already been considered (e.g. multiple description [EC82])
- Joint encoding processes have rarely been studied
  - Gray-Wyner [GW74], Willems-Wolf-Wyner [WWW02] and Yamamoto [Yam96] are almost the only studies.

Main contributions: complementary delivering



- We clarify the rate-distortion function and some interesting properties for the above coding problem.
  - The minimum achievable rate for lossless configuration have been clarified. (Willems-Wolf-Wyner [WWW02])

Preliminaries (1/2)

- $\mathcal{X}$: finite alphabet, $|\mathcal{X}|$: cardinality of $\mathcal{X}$, $\mathcal{I}_M = \{1, 2, \cdots, M\}$
  - In what follows, only finite alphabets will be considered.
- $x^n = (x_1, x_2, \cdots, x_n)$: member of $\mathcal{X}^n$,
  $x_i^j = (x_i, x_{i+1}, \cdots, x_j)$: substring of x^n ($i \leq j$)
  - 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)

- $H(X)$: entropy of X,
  $H(X|Y)$: conditional entropy of X given Y,
  $I(X;Y)$: mutual information between X and Y,
  $I(X;Y|Z)$: conditional mutual information between X and Y given Z
- $\widehat{\mathcal{X}}$: reconstruction alphabet,
  $\Delta_X : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \infty)$: single-letter distortion function,
  $\Delta_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}})$: vector distortion function
  $$\Delta_X^n(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{n}\sum_{k=1}^{n} \Delta_X(x_k, \widehat{x}_k)$$
- Similar conventions are used for other alphabets, random variables and vectors.

Problem formulation (1/2)

Definition 1. (Lossy CD (Complementary Delivering) code)
A set $\left(\varphi^n, \widehat{\varphi}^n_{(1)}, \widehat{\varphi}^n_{(2)}\right)$ *of an encoder and decoders is a lossy CD code* $\left(n, M_n, \rho_n^{(X)}, \rho_n^{(Y)}\right)$ *for the source* (X, Y) *if and only if*

$$
\begin{aligned}
\varphi^n &: \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{I}_{M_n} \\
\widehat{\varphi}^n_{(1)} &: \varphi^n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{Y}^n \to \widehat{\mathcal{X}}^n, \\
\widehat{\varphi}^n_{(2)} &: \varphi^n(\mathcal{X}^n, \mathcal{Y}^n) \times \mathcal{X}^n \to \widehat{\mathcal{Y}}^n, \\
\rho_n^{(X)} &= E\left[\Delta_X^n(X^n, \widehat{X}^n)\right],\ \rho_n^{(Y)} = E\left[\Delta_Y^n(Y^n, \widehat{Y}^n)\right],
\end{aligned}
$$

where

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

Problem formulation (2/2)

Definition 2. (Lossy CD-achievable rate)
R is a lossy CD-achievable rate of the source (X, Y) for a given distortion pair (D_1, D_2) if and only if there exists a sequence of lossy CD codes $\left\{\left(n, M_n, \rho_n^{(X)}, \rho_n^{(Y)}\right)\right\}_{n=1}^{\infty}$ for the source (X, Y) such that

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

$$\limsup_{n\to\infty} \rho_n^{(X)} \le D_1, \quad \limsup_{n\to\infty} \rho_n^{(Y)} \le D_2.$$

Definition 3. (Inf lossy CD-achievable rate)

$$\begin{aligned} & R_1(X, Y | D_1, D_2) \\ &= \inf\left\{R : R \text{ is a lossy CD-achievable rate of } (X, Y) \text{ for } (D_1, D_2)\right\}. \end{aligned}$$

Main results

Theorem 1. (Coding theorem of lossy CD code)

$$R_1(X, Y|D_1, D_2) = \min\Big[\max\{I(X;U|Y), I(Y;U|X)\}\Big],$$

where the auxiliary random variable U *takes a value over the alphabet* $\mathcal{U}$ *satisfying* $|\mathcal{U}| \leq |\mathcal{X} \times \mathcal{Y}| + 2$*, and the minimum is taken over all* $P_{U|XY} \in \mathcal{M}(\mathcal{U}|P_{XY})$ *such that there exist functions* $\phi_{(1)} : \mathcal{U} \times \mathcal{Y} \to \widehat{\mathcal{X}}$ *and* $\phi_{(2)} : \mathcal{U} \times \mathcal{X} \to \widehat{\mathcal{Y}}$ *which satisfy*

$$D_1 \geq E\left[\Delta_X(X, \phi_{(1)}(U, Y))\right],$$
$$D_2 \geq E\left[\Delta_Y(Y, \phi_{(2)}(U, X))\right].$$

- This theorem can be easily extended to any finite number of correlated sources.

Properties (1/4)

Corollary 1. (Consistency with the known results for lossless configuration)
If $\Delta_X(x, \widehat{x}) = 0 \Leftrightarrow x = \widehat{x}$ *and* $\Delta_Y(y, \widehat{y}) = 0 \Leftrightarrow y = \widehat{y}$*, then*

$$\begin{aligned} R_2(X, Y) &\stackrel{\text{def.}}{=} R_1(X, Y | D_1 = 0, D_2 = 0) \\ &= \max\{H(X|Y), H(Y|X)\}, \end{aligned}$$

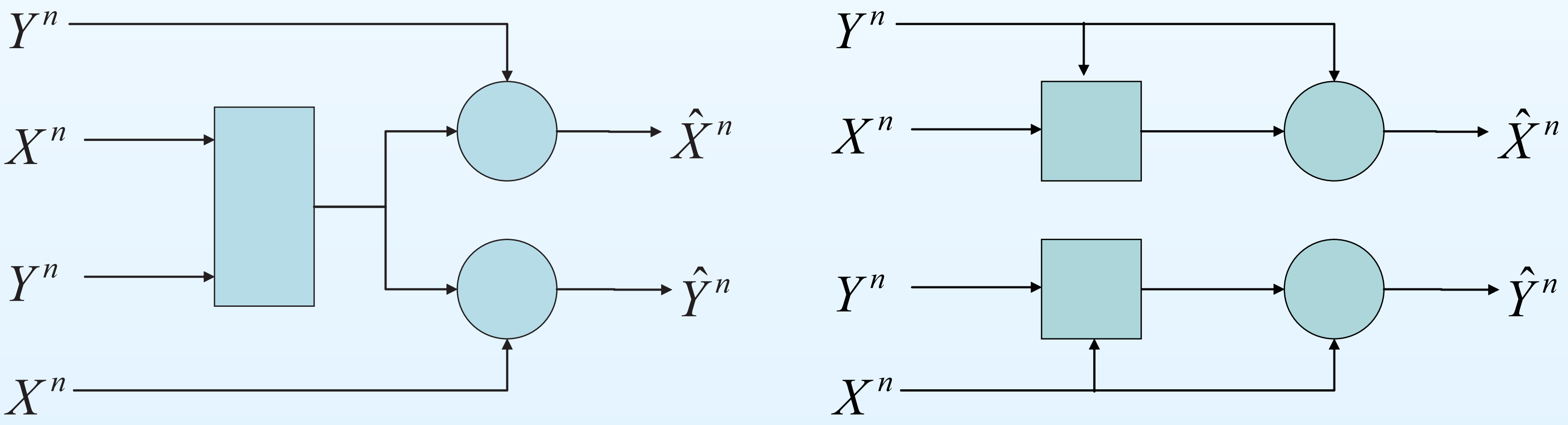
which coincides with the results reported by Han-Kobayashi [HK80], Sgarro [Sga77] and Willems-Wolf-Wyner [WWW02].

Properties (2/4)

Corollary 2. (Relationships to the known results for lossy configuration)

$$
\begin{aligned}
\max\{R_{C1}(X|Y,D_1), R_{C1}(Y|X,D_2)\}& \\
\le\ & R_1(X,Y|D_1,D_2) \\
\le\ & R_{C1}(X|Y,D_1) + R_{C1}(Y|X,D_2),
\end{aligned}
$$

where $R_{C1}(X|Y,D)$ *is the conditional rate-distortion function [Ber71].*

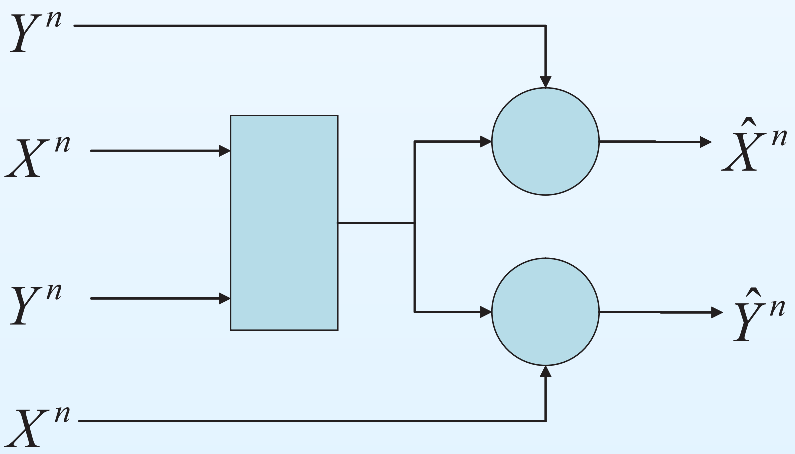


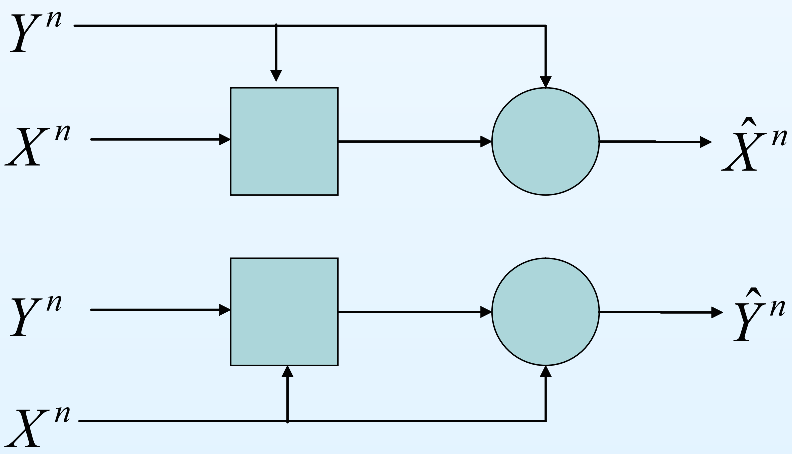
Properties (3/4)

Corollary 2. Contd.
On the other hand, for lossless configuration

$$\begin{aligned}\max\{R_{C2}(X|Y), R_{C2}(Y|X)\} \\ = \quad R_2(X,Y) \le R_{C2}(X|Y) + R_{C2}(Y|X),\end{aligned}$$

where $R_{C2}(X|Y)$ *is the minimum achievable rate when* X *is encoded and reproduced, both with side information* Y *(see e.g. [CT91]).*





Properties (4/4)

Corollary 3. (Consistency with the known results for special cases)

$$\begin{aligned} R_1(X,Y|D_1=d_1,D_2) &= R_{C1}(Y|X,D_2), \\ R_1(X,Y|D_1,D_2=d_2) &= R_{C1}(X|Y,D_1) \end{aligned}$$

if $d_1 \geq \overline{\Delta}_X$ *and* $d_2 \geq \overline{\Delta}_Y$, *where*

$$\begin{aligned} \overline{\Delta}_X &\stackrel{\text{def.}}{=} \max_{(x,\hat{x})\in\mathcal{X}\times\widehat{\mathcal{X}}} \Delta_X(x,\widehat{x}) < \infty, \\ \overline{\Delta}_Y &\stackrel{\text{def.}}{=} \max_{(y,\hat{y})\in\mathcal{Y}\times\widehat{\mathcal{Y}}} \Delta_Y(y,\widehat{y}) < \infty. \end{aligned}$$

Sketch of proof (1/4)

[Converse]
Consist of 3 parts:

1. Derive a lower bound of the coding rate R
2. Evaluate average distortions, e.g. $E[\Delta_X^n(X^n, \widehat{X}^n)]$
3. limit the size of the alphabet $\mathcal{U}$

1. $U_k = A_n X^{k-1} Y^{k-1}$, where $A_n = \varphi^n(X^n, Y^n)$

$$\longrightarrow \begin{cases} R+\delta \geq \frac{1}{n}\sum_{k=1}^{n} I(X_k; U_k | Y_k) \\ R+\delta \geq \frac{1}{n}\sum_{k=1}^{n} I(Y_k; U_k | X_k) \end{cases}$$

$$U = (J, U_J) \quad \longrightarrow \quad R+\delta \geq I(X;U|Y), \quad R+\delta \geq I(Y;U|X)$$

Sketch of proof (2/4)

2. $Y_{k+1}^n(U_k, Y_k)$: r.v. selected to minimize $E[\Delta_X(X_k, \widehat{X}_k)]$

$$\longrightarrow \quad \phi_{(1)k}(U_k, Y_k) \stackrel{\text{def.}}{=} \widehat{\varphi}_{(1)k}(A_n, Y^k * Y_{k+1}^n(U_k, Y_k))$$

$$\begin{aligned} D_1 + \gamma &\geq \frac{1}{n}\sum_{k=1}^n E\left[\Delta_X(X_k, \widehat{\varphi}_{(1)k}(A_n, Y^n))\right] \\ &\geq \frac{1}{n}\sum_{k=1}^n E\left[\Delta_X(X_k, \phi_{(1)k}(U_k, Y_k))\right] \\ &= E\left[\Delta_X(X, \phi_{(1)}(U, Y))\right] \end{aligned}$$

3. Introduce the support lemma.
   Need $|\mathcal{X} \times \mathcal{Y}| + 2$ constraints to preserve
   $P_{XY}(x, y)\ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, $\quad \max\{I(X; U|Y), I(Y; U|X)\}$,
   $E[\Delta_X(X, \phi_{(1)}(U, Y))]$, $\quad$ and $E[\Delta_Y(Y, \phi_{(2)}(U, X))]$.

sketch of Proof (3/4)

[Achievability]

Codeword selection

- Generate $\mathcal{A}_U = \{\boldsymbol{u}_i\}_{i=1}^{M_U}$ according to P_U.
- Divide $\mathcal{A}_U$ into N_U bins, each containing $L_U = M_U/N_U$ members.

Encoding

- $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{u}_i$ s.t. $(\boldsymbol{u}_i, \boldsymbol{x}, \boldsymbol{y}) \in T_{UXY}(k_1\delta)$ (Strong typicality)
  No $\boldsymbol{u}_i$ we can find : encoding error.
  $\longrightarrow$ $M_U \geq \exp\{n(I(XY;U) + \gamma_1)\}$ is necessary to prevent encoding errors.
- $\varphi^n(\boldsymbol{x}, \boldsymbol{y}) = j$ s.t. $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y})$ belongs to j-th bin.

Decoding

- $\widehat{\boldsymbol{u}}(\boldsymbol{y}) = \boldsymbol{u}$ s.t. $(\boldsymbol{u}, \boldsymbol{y}) \in T_{UY}(k_2\delta)$, $\boldsymbol{u}$ belongs to j-th bin.
  No or more than one $\boldsymbol{u}$ we can find : decoding error.
  $\longrightarrow$ $L_U \leq \exp\{n(I(Y;U) - \gamma_2)\}$ is necessary to avoid decoding errors.
- $\widehat{\boldsymbol{x}} = (x_1, \cdots, x_n)$, $\widehat{x}_k = \phi_{(1)}(\widehat{u}_k(\boldsymbol{y}), y_k)$ $k \in \mathcal{I}_n$.

Proof of theorems (4/4)

Distortion

- $(\boldsymbol{u}(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{x},\boldsymbol{y}) \in T_{UXY}(k_1\delta)$ from the encoding scheme.
- If no error occures in the encoding/decoding processes,

$$\begin{aligned}
\Delta_X(\boldsymbol{x},\widehat{\boldsymbol{x}}) &\le \sum_{(u,x,y)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} \{P(u,x,y)+k_1\delta\}\Delta_X(x,\phi_{(1)}(u,y)) \\
&\le E[\Delta_X(x,\phi_{(1)}(u,y))] + k_1\delta\overline{\Delta}_X|\mathcal{U}\times\mathcal{V}\times\mathcal{W}| \\
&\le D_1 + k_1\delta\overline{\Delta}_X|\mathcal{U}\times\mathcal{V}\times\mathcal{W}|.
\end{aligned}$$

Coding rate

$$\begin{aligned}
\frac{1}{n}\log N_U &= \frac{1}{n}\log\frac{M_U}{L_U} \\
&\ge I(XY;U) - \min\{I(Y;U), I(X;U)\} \\
&= \max\{I(X;U|Y), I(Y;U|X)\}.
\end{aligned}$$

Conclusions

- Investigated a coding problem for correlated information sources, where
  - messages from two correlated sources are jointly encoded
  - each decoder has access to one of two messages to reproduce the other message
- Clarified the rate-distortion function for the problem

<u>Future work</u>

- Extend the results to other classes of information sources
- Investigate some extended models
  (e.g. the case that only an encoded sequence of a message is available as side information at each decoder)
  $\longrightarrow$ to be presented at ISITA2006

Thank you

References

[Ber71] T. Berger. *Rate distortion theory: A mathematical basis for data compression*. Prentice Hall, Englewood Cliffs, NJ, 1971.

[CT91] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 1991.

[EC82] A. A. ElGamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inform. Theory*, 28(6):851–857, November 1982.

[GW74] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell Syst. Tech. J.*, 53(9):1681–1721, November 1974.

[HK80] T. S. Han and K. Kobayashi. A unified achievable rate region for a general class of multiterminal source coding systems. *IEEE Trans. Inform. Theory*, 26(3):277–288, 1980.

[KM75] J. Körner and K. Marton. Images of a set via two channels and thier role in multi-user communication. *IEEE Trans. Inform. Theory*, 23(6):751–761, November 1975.

[Sga77] A. Sgarro. Source coding with side information at several decoders. *IEEE Trans. Inform. Theory*, 23(2):179–182, March 1977.

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

[Tun78] S. Y. Tung. *Multiterminal source coding*. PhD thesis, School of Electrical Engineering, Cornell University, May 1978.

[WWW89] F. M. J. Willems, J. K. Wolf, and A. D. Wyner. Communicating via a processing broadcast satellite. In *Proc. the 1989 IEEE/CAM Information Theory Workshop*, June 1989.

[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.

[Wyn75] A. D. Wyner. On source coding with side information at the decoder. *IEEE Trans. Inform. Theory*, 21(3):294–300, May 1975.

[WZ76] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inform. Theory*, 22(1):1–10, January 1976.

[Yam96] H. Yamamoto. Source coding theory for a triangular communication system. *IEEE Trans. Inform. Theory*, 42(3):848–853, May 1996.

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