JAIST Repository

https://dspace.jaist.ac.jp/

Title	Decision making via End-to-End Lossy Distributed Wireless Cooperative Networks - A Distributed Hypothesis Testing based Formulation -					
Author(s)	Matsumoto, Tad					
Citation	Institute of Electronics, Information and Communication Engineers Technical Committee on Radio Communication Systems (IEICE RCS), RCS2024-23					
Issue Date	2024-05-24					
Туре	Presentation					
Text version	author					
URL	http://hdl.handle.net/10119/19040					
Rights	Copyright (c) Author. Tad Matsumoto, Decision making via End-to-End Lossy Distributed Wireless Cooperative Networks - A Distributed Hypothesis Testing based Formulation -, RCS2024-23, Tutorial Lecture at IEICE Technical Committee on Radio Communication Systems (RCS), May 30–31, 2024, Fukuoka University.					
Description	IEICE Technical Committee on Radio Communication Systems (RCS), May 30–31, 2024, Fukuoka University 電子情報通信学会 無線通信シス テム研究会, 2024年5月30日~31日, 福岡大学					



Decision making via End-to-End Lossy Distributed Wireless Cooperative Networks - A Distributed Hypothesis Testing based Formulation -Full Tutorial: MMMM DD, YYYY @ University of ZZZZ Tad Matsumoto*, **, *** IEEE Life Fellow





*International Chair, Invited Professor, IMT -Atlantic Technopôle Brest-Iroise CS 83818 29238 Brest cedex 03, France

Tadashi.matsumoto@imt-atlantique.fr

** Professor Emeritus, JAIST matumoto@jaist.ac.jp

*** Professor Emeritus, CWC, University of Oulu,

tadashi.matsumoto@oulu.fi









Tree of Tad's SISU: Wireless Communications











Contents of Full Tutorial:

- 0. Revisit to Fundamental Theorems in Network Information Theory
- 1. End-to-End Lossless Relaying: Slepian Wolf Theorem with Source-Channel Separation
- 1.1 EXIT Analysis for Source Bit-Flipped MIMO Transmission with Turbo Equalization
- 1.2 Slepian-Wolf Formulation for Lossless Two-Way Relay Networks
- 2. End-to-End Lossy Distributed Multi-terminal Networks:
 - Rate Distortion Analysis
- 2.1 Wyner-Ziv Formulation for End-to-End Lossy Two-Way Relay Network
- 2.2 Berger-Tung Formulation for Two Source One Helper Network
- 2.3 End-to-End Lossless and Lossy Multiple Access Channels
- 2.4 Two Stage Wyner-Ziv Network: Distortion Transfer Analysis
- 3. Wyner-Ziv Formulation for Decision Making Process
- 3.1 Revisit of Helper-aided Lossy Networks
- 3.2 Distributed Hypothesis Testing (DHT)
- 3.3 Semantic Communications
- 3.4 Learning Process in Machine Learning





Contents of Full Tutorial:

- 0. Revisit to Fundamental Theorems in Network Information Theory
- 1. End-to-End Lossless Relaying: Slepian Wolf Theorem with Source-Channel Separation
- 1.1 EXIT Analysis for Source Bit-Flipped MIMO Transmission with Turbo Equalization
- 1.2 Slepian-Wolf Formulation for Lossless Two-Way Relay Networks
- 2. End-to-End Lossy Distributed Multi-terminal Networks: Rate Distortion Analysis
- 2.1 Wyner-Ziv Formulation for End-to-End Lossy Two-Way Relay Network
- 2.2 Berger-Tung Formulation for Two Source One Helper Network
- 2.3 End-to-End Lossless and Lossy Multiple Access Channels
- 2.4 Two Stage Wyner-Ziv Network: Distortion Transfer Analysis
- 3. Wyner-Ziv Formulation for Decision Making Process
- 3.1 Revisit of Helper-aided Lossy Networks
- 3.2 Distributed Hypothesis Testing (DHT)
- 3.3 Semantic Communications
- 3.4 Learning Process in Machine Learning

Copied from March 27 Seminar Top Page Slide

Lossy Multi-terminal Cooperative Networks, Queueing, and Decision Making:

Erlang, Shannon, and Neyman-Pearson Meet in 6G Networks March 27, 2023 @ Uoulu CWC Research Seminar by Remote Tad Matsumoto*, **, ***

In what fields are they famous as land-make builders? IEEE Life Fellow



JAFAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY







*International Chair, Invited Professor, IMT -Atlantic Technopôle Brest-Iroise CS 83818 29238 Brest cedex 03, France Tadashi.matsumoto@imt-atlantique.fr

- **** Professor Emeritus, JAIST**
 - matumoto@jaist.ac.jp

*** Professor Emeritus, CWC, University of Oulu,

tadashi.matsumoto@oulu.fi











AIST

Erlang: Queueing Theory





Neyman-Pearson: Hypothesis Testing





Why "Neyman Pearson" Involved?



Unclear but still Accident Avoidable Make left turn or right turn?



Clearer but maybe Too Late to Avoid Accident Decision: Make left turn!

Making correct decision is most important than lossless reconstruction of the observation for risk avoidance!





Is Lossless reconstruction Needed?

Risk =Risk(TM(time), CNDM(time))





École Mines-Télécom





Welcome to ITW 2024 (https://IEEE-ITW2024.org)

Register a paper for 2024 IEEE Information Theory Workshop

Topics

- Deep Learning for Communication Networks
- Detection and Estimation
- Distributed Storage and Computing
- Emerging Applications of Information Theory
- Information Theory and Statistics
- Information Theory for Decision and Control
- Information Theory in Biology
- Information Theory in Computer Science



11	٢V	V	'2	0	2	4
				-	_	



Chapter 0. Revisit of Fundamental Theorems in Network

<u>1. Channel Capacity: Point-to-Point Lossless Channel's Maximum Capability</u></u>

Theorem 3.1 (Channel Coding Theorem). The capacity of the discrete memoryless channel p(y|x) is given by the information capacity formula



$$C = \max I(X;Y) \ge I(X;Y) \ge 0$$

 $C \le \log |\mathcal{H}|$ because $C = \max I(X;Y) \le \max H(X) = \log |\mathcal{H}|$ $C \le \log |\mathcal{H}|$ because $C = \max I(X;Y) \le \max H(Y) = \log |\mathcal{H}|$







$$C = \max_{p(x): EX^2 \le P} I(X;Y) = \log(1 + SNR)$$



■ Slepian-Wolf Theorem:

$$\begin{split} \mathsf{R}_{\mathsf{X}} &\geq \mathsf{H}(\mathsf{X} \mid \mathsf{Y}) \\ \mathsf{R}_{\mathsf{Y}} &\geq \mathsf{H}(\mathsf{Y} \mid \mathsf{X}) \\ \mathsf{R}_{\mathsf{X}} &+ \mathsf{R}_{\mathsf{Y}} &\geq \mathsf{H}(\mathsf{X}, \mathsf{Y}) \end{split}$$





Two Sources:



Theorem 10.1 (Slepian–Wolf Theorem). The optimal rate region \mathscr{R}^* for distributed lossless source coding of a 2-DMS (X_1, X_2) is the set of rate pairs (R_1, R_2) such that

$$\begin{split} R_1 &\geq H(X_1 \,|\, X_2), \\ R_2 &\geq H(X_2 \,|\, X_1), \\ R_1 + R_2 &\geq H(X_1, X_2). \end{split}$$







One Source One Helper:



Network

Abbas El Gamal

Information Theory

Theorem 10.2. Let (X, Y) be a 2-DMS. The optimal rate region \mathscr{R}^* for lossless source coding of X with a helper observing Y is the set of rate pairs (R_1, R_2) such that

 $R_1 \ge H(X|U),$ $R_2 \ge I(Y;U)$

for some conditional pmf p(u|y), where $|\mathcal{U}| \le |\mathcal{Y}| + 1$.





3. Lossy Source Coding:

Theorem 3.5 (Lossy Source Coding Theorem). The rate-distortion function for a DMS *X* and a distortion measure $d(x, \hat{x})$ is

 $R(D) = \min_{p(\hat{x}|x): \mathsf{E}(d(X,\hat{X})) \le D} I(X; \hat{X})$

for
$$D \ge D_{\min} = \min_{\hat{x}(x)} \mathsf{E}[d(X, \hat{x}(X))].$$

Source
$$X^n$$

Encoder
 $f_n(X^n) \in \{1, 2, \dots 2^{nR}\}$
 $\hat{X}^n = g_n(f_n(X^n)) \in \{1, 2, \dots 2^n\}$

Example 3.4 (Bernoulli source with Hamming distortion). The rate-distortion function for a Bern(p) source $X, p \in [0, 1/2]$, and Hamming distortion measure is

$$R(D) = \begin{cases} H(p) - H(D) & \text{for } 0 \le D < p, \\ 0 & \text{for } D \ge p. \end{cases}$$

$$I(X; \hat{X}) = H(X) - H(X | \hat{X}) \\ = H(p) - H(X \oplus \hat{X} | \hat{X}) \\ \ge H(p) - H(X \oplus \hat{X}) \\ \stackrel{(a)}{\ge} H(p) - H(D), \\ \text{with } P\{X \ne \hat{X}\} \le D. \end{cases}$$

Network Information Theory Abbas El Gannal Young-Han Kim







<u>4. Source-Channel Separation:</u> A Connecting point between Source Coding and Channel Coding

Theorem 3.7 (Source–Channel Separation Theorem). Given a DMS *U* and a distortion measure $d(u, \hat{u})$ with rate–distortion function R(D) and a DMC p(y|x) with capacity *C*, the following statements hold:

- If rR(D) < C, then (r, D) is achievable.
- If (r, D) is achievable, then $rR(D) \le C$.



Remark 3.14. As a special case of joint source–channel coding, consider the problem of sending *U* over a DMC <u>losslessly</u>, i.e., $\lim_{k\to\infty} P\{\hat{U}^k \neq U^k\} = 0$. The separation theorem holds with the requirement that $rH(U) \leq C$.

<u>Note</u>: **Source-Channel Separation applies to:**

- Orthogonal transmission with multiple Point-to-Point links,
- Both Lossless and Lossy, so far as each link is orthogonal,
- Helper link,
- (Experts say it holds with majority of the cases....)





Multiple Access Relaying with orthogonal links



5. Rate Region: Lossy Source Coding with Side information/a Helper

With Side information:

Theorem 11.3 (Wyner–Ziv Theorem). Let (X, Y) be a 2-DMS and $d(x, \hat{x})$ be a distortion measure. The rate–distortion function for X with side information Y available noncausally at the decoder is

 $R_{\text{SI-D}}(D) = \min \left(I(X; U) - I(Y; U) \right) = \min I(X; U|Y) \quad \text{for } D \ge D_{\min},$

where the minimum is over all conditional pmfs p(u|x) with $|\mathcal{U}| \le |\mathcal{X}| + 1$ and functions $\hat{x}(u, y)$ such that $\mathsf{E}[d(X, \hat{X})] \le D$, and $D_{\min} = \min_{\hat{x}(y)} \mathsf{E}[d(X, \hat{x}(Y))]$.







Notice: $U \rightarrow X \rightarrow Y$ forms a Markov Chain

$$I(X;U) - I(Y;U) = I(XY;U) - I(Y;U|X) - I(Y;U)$$

= $I(XY;U) - I(Y;U) = I(X;U|Y).$

 Q_{n}

18

With a Helper:

Theorem 11.3 (Wyner–Ziv Theorem). Let (X, Y) be a 2-DMS and $d(x, \hat{x})$ be a distortion measure. The rate–distortion function for X with side information Y available noncausally at the decoder is

$$\begin{split} R_{\text{SI-D}}(D) &= \min \left(I(X;U) - I(Y;U) \right) = \min I(X;U|Y) \quad \text{for } D \geq D_{\min}, \\ R_{Helper} &\geq I(Y;U) \end{split}$$

where the minimum is over all conditional pmts p(u|x) with $|\mathcal{U}| \le |\mathcal{X}| + 1$ and functions $\hat{x}(u, y)$ such that $\mathsf{E}[d(X, \hat{X})] \le D$, and $D_{\min} = \min_{\hat{x}(y)} \mathsf{E}[d(X, \hat{x}(Y))]$.



Coded Side Information=Helper



6. Rate Region: Distributed Multipoint-to-Multipoint Lossy Coding

Without Helper:

Theorem 12.1 (Berger–Tung Inner Bound). Let (X_1, X_2) be a 2-DMS and $d_1(x_1, \hat{x}_1)$ and $d_2(x_2, \hat{x}_2)$ be two distortion measures. A rate pair (R_1, R_2) is achievable with distortion pair (D_1, D_2) for distributed lossy source coding if

$$\begin{split} R_1 &> I(X_1; U_1 | U_2, Q), \\ R_2 &> I(X_2; U_2 | U_1, Q), \\ R_1 + R_2 &> I(X_1, X_2; U_1, U_2 | Q) \end{split}$$

for some conditional pmf $p(q)p(u_1|x_1, q)p(u_2|x_2, q)$ with $|\mathcal{U}_j| \le |\mathcal{X}_j| + 4$, j = 1, 2, and functions $\hat{x}_1(u_1, u_2, q)$ and $\hat{x}_2(u_1, u_2, q)$ such that $\mathsf{E}(d_j(X_j, \hat{X}_j)) \le D_j$, j = 1, 2.







With a Helper:

Theorem 12.1 (Berger–Tung Inner Bound). Let (X_1, X_2) be a 2-DMS and $d_1(x_1, \hat{x}_1)$ and $d_2(x_2, \hat{x}_2)$ be two distortion measures. A rate pair (R_1, R_2) is achievable with distortion pair (D_1, D_2) for distributed lossy source coding if

> $R_1 > I(X_1; U_1 | U_2, V, Q),$ $R_2 > I(X_2; U_2 | U_1 | V, Q),$ $R_1 + R_2 > I(X_1, X_2; U_1, U_2 | V, Q),$ $R_{Helper} > I(Y;V)$

for some conditional pmf $p(q)p(u_1|x_1, q)p(u_2|x_2, q)$ with $|\mathcal{U}_i| \le |\mathcal{X}_i| + 4$, j = 1, 2, andfunctions $\hat{x}_1(u_1, u_2, q)$ and $\hat{x}_2(u_1, u_2, q)$ such that $E(d_j(X_j, \hat{X}_j)) \le D_j, j = 1, 2$.





7. Multiple Access Channels (MAC)

Correlated Sources Transmission over MAC:

First consider the following sufficient condition for separate source and channel coding. We know that the capacity region \mathcal{C} of the DM-MAC is the set of rate pairs (R_1, R_2) such that

 $\begin{aligned} R_1 &\leq I(X_1; Y | X_2, Q), \\ R_2 &\leq I(X_2; Y | X_1, Q), \\ R_1 + R_2 &\leq I(X_1, X_2; Y | Q) \end{aligned}$

for some pmf $p(q)p(x_1|q)p(x_2|q)$. We also know from the Slepian–Wolf theorem that the optimal rate region \mathscr{R}^* for distributed lossless source coding is the set of rate pairs (R_1, R_2) such that





There are two regions in this set up: SW and MAC regions.









<u>Revisit to Source-Channel Separation:</u>

- Orthogonal transmission with multiple Point-to-Point links,
- Both Lossless and Lossy, so far as orthogonal,
- Helper link,





Region intersection: The rate-pair plot belongs to the both MAC and SW regions. → Source-channel separation holds!



Separation holds in:

- MAC transmission when the rates-plot is in intersection (Sufficient condition, NOT optimal. Separation vs. Joint),
- (Experts say it holds in many cases....)





A sufficient condition for <u>Lossy</u> Recovery.

- Recovery for one source with one helper

$$\begin{split} R_1 &\leq I(X_1; Y | X_2, Q), \\ R_2 &\leq I(X_2; Y | X_1, Q), \\ R_1 + R_2 &\leq I(X_1, X_2; Y | Q) \end{split}$$

and

 $R_1 \ge H(X_1; U|V)$ $R_2 \ge I(X_2; V)$

Region intersection → Source-channel separation holds!



Properties of Binary Convolution:

Binary Convolution

x * y = x(1 - y) + (1 - x)y = x + y - 2xy

For variables
$$x, t \in [0, 0.5]$$
 and $y \in [0, 0.5)$
 $x * y \le t \Rightarrow x \le \Lambda(t, y) \triangleq \frac{1}{2}(1 - \frac{2t - 1}{2y - 1})$

• $x * y \le t$ inherently involves $y \le t$

• $\Lambda(y, t)$ is a monotonically decreasing function of y, with a maximum

 $\Lambda(y=0,t)=t$

• $\Lambda(y, t)$ is a linearly increasing function of t, with a maximum

$$\Lambda(y, t = 0.5) = 0.5$$

Recursive structure:

$$y * z \le s \Rightarrow y \le \Lambda(s, z)$$
$$x * y * z \le t \Rightarrow x \le \Lambda(t, \Lambda(s, z))$$

$$v * w \le c_1 \Rightarrow v \le \Lambda(c_1, w)$$

 $x * y * z * \cdots v * w \le c_n \Rightarrow x \le \Lambda(c_n, \Lambda(cn_{-1}, \Lambda(\dots, \Lambda(c_1, w))))$



Tad's Book?



Chapter 1. End-to-End Lossless Relaying: Slepian Wolf Theorem with Source-Channel Separation

1.1 EXIT Analysis for Source Bit-Flipped MIMO Transmission with Turbo Equalization



Vertical iteration is expected to improve performance because of space diversity gain

from using antenna 1 and 2, and coding gain. Mariella Sarestoniemi, Tad Matsumoto, Kimmo Kansanen, and Jari Iinatti, "Turbo Diversity Based on SC/MMSE Equalization", *IEEE Transactions on Vehicular Technology*, Vol. 54, No. 2, pp. 749-752, March, 2005

This design is called as Spatial Turbo Code (STC) because coded sequences are multiplexed in the spatial domain, not in the time domain as in the Turbo codes.







- We* developed Frequency Domain Turbo Equalization Algorithms for single carrier signalling: It requires computational complexity of only "high school levelmath"!
- Convergence property analysis made significantly easy!

• Kimmo Kansanen, and Tad Matsumoto, "An Analytical Method for MMSE MIMO Turbo Equalizer EXIT Chart Computation", *IEEE Transactions on Wireless Communications*, Vol. 6, No.1, pp.59-63, January, 2007





K. Anwar and T. Matsumoto, "Spatially Concatenated Codes with Turbo Equalization for Correlated Sources", *IEEE Transaction Signal Processing*, vol. 60, no. 10, Oct. 2012, pp. 5572-5577

EXIT Chart for Source Bit-Flipped MIMO Transmission with Turbo Equalization





Parameters:

Transmitter:
 Encoder: CC
 4(17,15),17
 Interleaver=5000
 (random)
 Correlation Model:
 Bit-flipping

Channel: MIMO 2x2

Equal Power 64path

Receiver: Decoder: BCJR Log-MAP FFT=512



Average BER in Block Frequency-Selective Block Rayleigh Fading Channels: Source Bit-Flipped MIMO Transmission with Turbo Equalization - Bit-flipped sequences are correlated sources!



Parameters: Transmitter: Encoder: CC 4(17,15),17 Interleaver=1024 (random) Correlation Model: Bit-flipping

ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Channel: MIMO 2x2 Equal Power 64-path

Receiver: Decoder: BCJR Log-MAP FFT=512





Source





1.2 Slepian-Wolf Formulation for Lossless Two-Way Relay Networks Observation on Bit-flipped MIMO TEQ: BF Model Works as Correlated Sources!

RD

- Scenario Assumption (Lossy-Forward, LF)
 - 1. Source broadcasts information
 - 2. Errors may occur in S-R link
 - 3. Relay still forwards the lossy information
 - 4. Destination recovers the source information by joint decoding

SD

 $\bullet \bullet \bullet \bullet \bullet$

Relay

5. End-to-End lossless.



Destination



O

Do we need to recover the relay information $b_{\rm R}$?



We do not care about the decoding result (=V) of b_R, but we can use b_R as a helper! → One Source One Helper Slepian Wolf Theorem for Lossless Multi-terminal Source Coding

Zhou, Xiaobo, Meng Cheng, Xin He, and Tad Matsumoto, "Exact and approximated outage probability analyses for decode-and-forward relaying system allowing intra-link errors," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 7062-7071, Dec. 2014.
LF Rate Region Analysis: *Slepian –Wolf Theorem for Lossless Multi-terminal Source Coding with a helper.*

With LF, the S-R link is lossy, the admissible rate region is given by:

$$\begin{cases} R_1 \geq H(X_1 | V) \\ R_2 \geq I(Y; V) = H(Y) - H(Y | V) \end{cases}$$

This is a general expression.

To calculate the rate region using parameters related to the links, we use :

- (1) Shannon's Source-Channel Separation Theorem,
- (2) Test Channel Model of Binary R(D) function to represent each link's threshold, and
- (3) Utilization of Markov Chain.
 - → Binary Convolution



Rate Region Analysis: we need threefold Integral!



LF Rate Region Analysis: SR Link

How can we combine Shannon's Separation Theorem and the Rate Region? By using the lossy Separation theorem $R_{c,1} \cdot R(\mathscr{D}) \leq C(\gamma_0)$ and with Inverse <u>C⁻¹(γ₀)</u> of the <u>Capacity Function C(γ₀)</u>, we calculate the binary distortion (*D*=BER) of the S-R link after decoding, as

$$p_{e} = \begin{cases} H_{b}^{-1}[1 - \Phi_{1}(\gamma_{0})], & \text{for } \Phi^{-1}(0) \leq \gamma_{0} \leq \Phi_{1}^{-1}(1) \\ 0, & \text{for } \gamma_{0} \geq \Phi_{1}^{-1}(1), \end{cases}$$

and $\Phi_{1}(\gamma_{0}) = \frac{C(\gamma_{0})}{R_{c,1}}$ Separation Theorem
with $H_{b}^{-1}(\cdot)$ denotes the inverse function of the binary
entropy function $H_{b}(x) = -x \log_{2} x - (1 - x) \log_{2}(1 - x),$
and $\Phi_{1}^{-1}(\cdot)$ is the inverse function of $\Phi_{1}(\cdot)$.

This means that given the *instantaneous* SNR γ_0 and $R_{c,1}$, we can calculate the binary distortion ($D=BER=p_e$) of S-R link after decoding!

JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY 1990

LF Rate Region Analysis: Test Channel

Binary Source

Consider a Binary source $x \in X$, Prob(x = 1) = p, Prob(x = 0) = 1 - p

Assume that p < 1/2. The rate distortion function is given by:

$$R(D) = \begin{cases} H(p) - H(D) &, \quad 0 \le D \le \min(p, 1-p) \\ 0 &, \quad D > \min(p, 1-p) \end{cases}$$

where Hamming distortion measure is assumed.



LF Rate Region Analysis: RD Link

 We also do not need lossless R-D link. R-D link's error probability α after decoding can be calculated in the same way, as

$$\alpha = \begin{cases} H_b^{-1} [1 - \Phi_1 \gamma_2)], & \text{for } \Phi_1^{-1}(0) \leq \gamma_2 \rangle \leq \Phi_1^{-1}(1) \\ 0, & \text{for } \gamma_2 \rangle \geq \Phi_1^{-1}(1), \end{cases}$$

with $\Phi_1 (\gamma_2) = \begin{bmatrix} C(\gamma_2) \\ R_{c,1} \end{bmatrix}$ Separation Theorem
By combining all, we have

 $\begin{cases} R_1 \geq H(X_1 | \mathsf{V}) = H_b(\alpha * p_e) \text{, because V->Y->X_1 forms Markov Chain.} \\ R_2 \geq I(Y; \mathsf{V}) = H(Y) - H(Y | \mathsf{V}) = 1 - H_b(\alpha) \end{cases}$

with
$$\alpha * p_e = (1 - \alpha) p_e + \alpha (1 - p_e)$$

<u>We do not know γ_0 , γ_1 , γ_2 but we know their distributions.</u>

To Calculate the Outage, we need threefold Integrals

$$\begin{array}{ll} P_{1,a} = \Pr\{p = 0, R_{2} \geq 1, 0 \leq R_{1} \leq H_{b}(p)\} \\ = \Pr\{\gamma_{0} \geq \Phi_{1}^{-1}(1), \gamma_{2} \geq \Phi_{2}^{-1}(1), \\ \Phi_{1}^{-1}(0) \leq \gamma_{1} \leq \Phi_{1}^{-1}(0)\} \\ = \int_{\Phi_{1}^{-1}(0)}^{\Phi_{1}^{-1}(1)} d\gamma_{0} \int_{\Phi_{2}^{-1}(0)}^{\Phi_{2}^{-1}(1)} d\gamma_{2} \\ = \int_{\Phi_{1}^{-1}(0)}^{\Phi_{1}^{-1}(0)} p(\gamma_{0}) \cdot p(\gamma_{1}) \cdot p(\gamma_{2}) d\gamma_{1} \\ = 0, \\ P_{1,b} = \Pr\{p = 0, 0 \leq R_{2} \leq 1, 0 \leq R_{1} \leq H_{b}(\alpha * p)\} \\ = \Phi\{p_{1}^{-1}(0) \quad Dorlog threeform the Details (of the Calculations! \\ \Phi_{1}^{-1}(0) \leq \gamma_{1} \leq \Phi_{1}^{-1}(1) \quad P(\gamma_{2}) d\gamma_{1} \\ = 0, \\ P_{1,b} = \Pr\{p = 0, 0 \leq R_{2} \leq 1, 0 \leq R_{1} \leq H_{b}(\alpha * p)\} \\ = \Phi\{p_{1}^{-1}(0) \quad Dorlog threeform the Details (of the Calculations! \\ \Phi_{1}^{-1}(0) \quad P(\gamma_{0}) \cdot p(\gamma_{1}) \cdot p(\gamma_{2}) d\gamma_{1} \\ = \frac{1}{\Gamma_{0}} \exp\left[-\frac{\Phi_{2}^{-1}(1)}{\Gamma_{0}}\right] \int_{\Phi_{2}^{-1}(0)}^{\Phi_{2}^{-1}(1)} \exp(-\frac{\gamma_{2}}{\Gamma_{2}}) \\ \cdot \int_{\Phi_{1}^{-1}(0)}^{\Phi_{1}^{-1}(1) - \Phi_{2}(\gamma_{2})} p(\gamma_{0}) \cdot p(\gamma_{1}) \cdot p(\gamma_{2}) d\gamma_{1} \\ = \frac{1}{\Gamma_{2}} \exp\left[-\frac{\Phi_{1}^{-1}(1)}{\Gamma_{0}}\right] \int_{\Phi_{2}^{-1}(0)}^{\Phi_{2}^{-1}(1)} \exp(-\frac{\gamma_{2}}{\Gamma_{2}}) \\ \cdot \left[1 - \exp(-\frac{\Phi_{1}^{-1}(1 - \Phi_{2}(\gamma_{2}))}{\Gamma_{1}})\right] d\gamma_{2}, \\ \cdot \left[1 - \exp\left[-\frac{\Phi_{1}^{-1}(1 - \Phi_{2}(\gamma_{2}))}{\Gamma_{1}}\right]\right] d\gamma_{2}, \\ \cdot \left\{1 - \exp\left[-\frac{\Phi_{1}^{-1}(1)}{\Gamma_{0}}\right\right] d\gamma_{0} d\gamma_{2} \\ \cdot \left\{1 - \exp\left[-\frac{\Phi_{1}^{-1}(1)}{\Gamma_{1}}\right\right] d\gamma_{0} d\gamma_{2} \\ \cdot \left\{1 - \exp\left[-\frac{\Phi_{1}^{-1}(1)}{\Gamma_$$

Comparison of exact and approximated SW region with a helper (Orthogonal Case)







Location A, d0=d1=d2 Location B, d0=(1/4)d1, d2=(3/4)d1

[1] X. Zhou, M. Cheng, X. He and T. Matsumoto, "Exact and Approximated Outage Probability Analyses for Decode-and- Forward Relaying System Allowing Intra-Link Errors," in IEEE Transactions on Wireless Communications, vol. 13, no. 12, pp. 7062-7071, Dec. 2014.





Chapter 2. End-to-End Lossy Distributed Multi-terminal Networks: Rate Distortion Analysis

2.1 Wyner-Ziv Formulation for End-to-End Lossy Two-Way Relay Network

• Internet of Things (IoT)

Connect objects to make Right Decisions→ E2E Lossy





Source

E2E Lossy Communications with LF

• Objective of IoT:

- Make a judgement rather to recovering information itself
- The picture exemplifies LF for lossy communications:
 - NOT necessarily be E2E-lossless as long as the system can make **right judgement**.

••••0

Relay



ossy Distributed Multi-terminal Source Coding



- S-R link: point-to-point communication.
- S-D and R-D links: distributed lossy multi-terminal source coding problem.

→ As a whole, Wyner-Ziv Problem



Lin Wensheng, Shen Qian, and Tad Matsumoto, "Lossy-Forward Relaying for Lossy Communications: Rate-Distortion and Outage Probability Analyses", *IEEE Transactions on Wireless Communications*, Vol. 18, No. 8, 05 June 2019, pp. pp. 3974-3986, **Dol:** <u>10.1109/TWC.2019.2919831</u>



Rate-Distortion Region



WZ R(D) function for general sources

 $R_1 > I(X; U|V),$ $R_2 > I(Y; V).$

- For binary sources
 - S-R link

 $R_0 > 1 - H_b(\rho).$

• R-D link

 $R_2 > 1 - H_b(\rho').$

• S-D link

$$R_1 > H_b(\rho' * \rho * D_X) - H_b(D_X).$$

because V->Y->X->U forms a Markov Chain.

 ρ : crossover probability between *X* and *Y* ρ' : crossover probability between *Y* and *V*



Outage Event





• The link rates (R_0, R_1, R_2) supported by channel capacities cannot satisfy the distortion requirement D_X , when they fall outside the achievable rate-distortion region. \rightarrow Outage



Again, we need multifold Integrals

$$P_{0ut} = \Pr \{ \{0 \le R_1 \le H_b(0 * D_X) - H_b(D_X), 0 \le R_2 \\ p = 0 \}$$

$$= \Pr \{ \{p = 0, (R_1, R_2) \in \alpha \cup \beta \} \\ + \Pr \{p = 0, (R_1, R_2) \in \alpha \cup \beta \} \\ + \Pr \{p = 0, (R_1, R_2) \in \alpha \cup \beta \} \\ + \Pr \{p = 0, (R_1, R_2) \in \alpha \} \\ + \Pr \{p = 0, (R_1, R_2) \in \alpha \} \\ + \Pr \{p = 0, (R_1, R_2) \in \alpha \} \\ + \Pr \{p = (0, 0.5], (R_1, R_2) \in \beta \} \\ = \Pr \{0 \le R_1 \le H_b(p * D_X) - H_b(D_X), 0 \le R_2; \\ p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * D_X) - H_b(D_X), 0 \le R_2; \\ p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 \le R_2 \le 1, p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 \le R_2 \le 1, p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0
$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0
$$= \Pr \{A_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0
$$= \Pr \{A_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_2 \le 1, p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_2 \le 1, p = 0 \}$$

$$= \Pr \{0 \le R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * p \times D_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; \\ 0 < R_1 \le H_b(p * P \times P_X) - H_b(D_X), 0 \le R_2; 0 \le R_1 \le H_b(p *$$$$$$$$



Again, we need threefold Integrals!

We calculated, but too boring. Let's skip it!

We evaluate the outage probability also by chain simulations using very simple signaling and joint decoding techniques.





Simulation Results



 Simulation results have the same tendency and the slope decay (=Diversity Order) as the theoretical bound.

2. The gap between the simulation and theoretical results becomes larger as D_X increases. \rightarrow We need *more efficient rate-distortion code.*

SIMULATION PARAMETERS

Parameter	Value
Frame length	10^4 bits
Number of frames	5×10^5
Source coding rate	1
Channel coding rate	1/2
Generator polynomial of CC	$G = ([3, 2]3)_8$
Type of interleaver	random interleaver
Modulation method	BPSK
Maximum iteration time	20







System Model





 $(U_1, U_2) \rightarrow (X_1, X_2) \rightarrow Y \rightarrow V$ and by data processing theorem $I(U_1, U_2; V) \leq I(Y; V) \leq [R_H]^-$)

Lin Wensheng, Qiang Xue, Jiguang He, Markku Juntti, and Tad Matsumoto, "Rate-Distortion and Outage Probability Analyses for Single Helper Assisted Lossy Communications", *IEEE Transactions on Vehicular Technology*, Vol. 68, 2019, No. 11, pp. 10882-10894, DOI:10.1109/TVT.2019.2939622





- What is the achievable rate-distortion region?
 - Condition for reliable communications: link rates can support the transmissions to satisfy the distortion requirements.
- Inner bound on the achievable rate-distortion region, given by Berger Tung Bound

 $\begin{aligned} R_1 &> I(X_1; U_1 | U_2, V, Q), \\ R_2 &> I(X_2; U_2 | U_1, V, Q), \\ R_1 + R_2 &> I(X_1, X_2; U_1, U_2 | V, Q), \\ R_H &> I(Y; V), \end{aligned}$

- U_i : compressed information of X_i
- *V*: compressed information of *Y*
- Q: an auxiliary variable resulting from time-sharing scheme

Use Inner bound \rightarrow Upper bound of the outage probability





- The inner bound for binary sources.
 - (a) for some $0 \le \tilde{d} \le D_2$, $\begin{cases}
 R_1 > H_b(D_1 * \rho * \tilde{d}) - H_b(D_1) - [R_H]^-, \\
 R_2 > 1 - H_b(\tilde{d}),
 \end{cases}$
 - (b) for some $0 \le \tilde{d} \le D_1$, $\begin{cases}
 R_1 > 1 - H_b(\tilde{d}), \\
 R_2 > H_b(\tilde{d} * \rho * D_2) - H_b(D_2) - [R_H]^-,
 \end{cases}$
 - (c) common case, $R_1 + R_2 > 1 + H_b(D_1 * \rho * D_2) - H_b(D_1) - H_b(D_2) - [R_H]^-,$
- $\rho: \text{crossover probability} \\ \text{between } X_1 \text{ and } X_2 \\ \text{(representing correlation)} \\ \tilde{d}: \text{dummy variable} \\ H_b: \text{binary entropy function} \\ a * b = a(1 b) + b(1 a)$
- $[R_{\rm H}]^- = \min\{1, R_{\rm H}\}$
- This is not an inner bound in general. It is only for the case that the following inequality holds with equality.

 $I(U_{\boldsymbol{S}}; V | U_{\boldsymbol{S}^c}) \leq [R_{\mathrm{H}}]^{-},$

where $S \subseteq \{1,2\}$, and S^c represents the complementary set of S.





• The shape of achievable rate-distortion region.

- 1. The achievable rate-distortion region is expanded as $R_{\rm H}$ increases.
- 2. However, the above part of the region for $R_{\rm H} \ge 1$, does not change even if the helper rate continues increasing.



 $\rho = 0.15, D_1 = D_2 = 0.05$





• The achievable rate-distortion region projected on the R_1 - R_2 plane by given R_H .

The derived inner bound perfectly coincides with the Berger-Tung inner bound when $R_{\rm H} = 0$ (equivalent to no helper).





Outage Probability Analysis





Outage event defined as:

• The link rates fall outside the achievable rate-distortion region, i.e., the link rates (R_1, R_2, R_H) supported by channel capacities cannot satisfy the distortion requirements (D_1, D_2) .





• The outage probability is the multiple integral with respect to (R_1, R_2, R_H) .

$$P_{\rm out} = \iiint \cdots dR_1 \, dR_2 \, dR_{\rm H}$$

- The instantaneous rates (R_1, R_2, R_H) are supported by the instantaneous signal-to-noise ratios (SNRs) $(\gamma_1, \gamma_2, \gamma_H)$.
- The outage probability can be calculated by the threefold integral with respect to SNRs $(\gamma_1, \gamma_2, \gamma_H)$.

$$P_{\rm out} = \iiint \cdots d\gamma_1 \, d\gamma_2 \, d\gamma_{\rm H}$$



Again, we need multifold Integrals

$$P_{1} = \Pr\{0 \le R_{1} \le H_{2}(D_{1} * \rho) - H_{2}(D_{1}) - [R_{H}]^{-}, H(X_{2}) \le R_{2}, 0 \le R_{H}\}$$

$$= \Pr\{0 \le \Theta_{1}(\gamma_{1}) \le \lambda_{1}(0), H(X_{2}) \le \Theta_{2}(\gamma_{2}), 0 \le \Theta_{H}(Y_{H})\}$$

$$= \Pr\{\Theta_{1}^{-1}(0) \le \gamma_{1} \le \Theta_{1}^{-1}[\lambda_{1}(0)], 0 \le \gamma_{H}\}$$

$$= \int_{\Theta_{1}^{-1}(0)}^{\Theta_{1}^{-1}(1)} d\gamma_{H} \int_{\Theta_{1}^{-1}(0)}^{\Theta_{1}^{-1}(1)} d\gamma_{1} \int_{\Theta_{1}^{-1}(0)}^{\Theta_{1}^{-1}(1)} d\gamma_{$$





- Numerical Results ($\rho = 0.1$)
- 1. The larger the acceptable distortion , the smaller the outage probability . $(D_j \nearrow \Rightarrow P_{out} \searrow).$
- 2. Without a helper, the curves always exhibit order diversity.
- 3. With a helper, it can achieve second order diversity (*P*_{out} decreases faster).



2. Multiterminal Source Coding with a Helper





Parameter	Value
Frame length	10^4 bits
Number of frames	5×10^5
Source coding rate	1
Channel coding rate	1/2
Generator polynomial of CC	$G = ([3, 2]3)_8$
Type of interleaver	random interleaver
Modulation method	BPSK
Maximum iteration time	20









2.3 End-to-End Lossless and Lossy Multiple Access Channels

Internet of things (IoT) system



Data gathering should not necessarily be lossless.

Shulin Song, Meng Cheng, Jiguang He, Xiaobo Zhou, and TadMatsumoto,"Outage Probability of One-Source-with-One-Helper Sensor Systems in Block Rayleigh Fading Multiple Access Channels", IEEE Sensor Journal, Accepted date: Aug. 24, DOI: <u>10.1109/JSEN.2020.3018787</u>

End-to-End Lossless MAC:



A sufficient condition of successful transmissions is defined as the case Slepian-Wolf region with a helper and MAC rate region intersect, where Source-Channel Separation holds.



Let the transmission rate of the source and the helper to be R_1 and R_2 , respectively.

$$R_{1} \geq \begin{cases} H(pe), & \text{for} \quad R_{2} \geq 1, \\ 1 + H(pe) - R_{2}, & \text{for} \quad H(pe) \leq R_{2} \leq 1, \\ 1, & \text{for} \quad 0 \leq R_{2} \leq H(pe). \end{cases} \quad \begin{cases} R_{1}R_{c} \leq C(\gamma_{1}), \\ R_{2}R_{c} \leq C(\gamma_{2}), \\ R_{1}R_{c} + R_{2}R_{c} \leq C(\gamma_{1} + \gamma_{2}), \\ \text{Source-Channel Separation} \end{cases}$$

Source – Channel separation modeled by Bit–Fipping Model (as in BF MIMO TEQ) with flipping probability p_e

Cases where Outage Happens







Outage Probability Expressions

Outage probability can be calculated by multifold integrals with respect to the instantaneous SNR of each link.

$$\begin{split} P_{out,1} &= \Pr\left\{\frac{1}{R_c}C\left(\gamma_1\right) < H\left(u_1 \mid u_2\right)\right\} & P_{out,2} = \Pr\left\{\frac{H\left(u_1 \mid u_2\right) \le \frac{1}{R_c}C\left(\gamma_1\right) < H\left(u_1\right)\right)}{\frac{1}{R_c}C\left(\gamma_1 + \gamma_2\right) < H\left(u_1, u_2\right)}\right\} \\ &= \int_{\Phi(0)}^{\Phi[H(u_1 \mid u_2)]} \int_{\Phi(0)}^{\Phi(+\infty)} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \int_{0}^{2^{R_cH(P_e)} - 1} \int_{0}^{+\infty} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \int_{0}^{2^{R_cH(P_e)} - 1} \int_{0}^{+\infty} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \int_{2^{R_cH(P_e)} - 1}^{2^{R_cH(P_e)} - 1} \int_{0}^{2^{R_cH(P_e)} - 1}\right] \right\}, \\ &= 1 - \exp\left\{-\frac{1}{\Gamma_1}\left[2^{R_cH(P_e)} - 1\right]\right\}, \\ P_{out} = P_{out,1} + P_{out,2} \\ \end{pmatrix} \\ P_{out} = P_{out,1} + P_{out,2} \\ \end{split}$$



Numerical Result





Lossy Case Helper: WZ System



End-to-End Lossy MAC:



Shulin Song, Jigang He, and Tad Matsumoto, "Rate-Distortion and Outage Probability Analyses of Wyner-Ziv Systems over Multiple Access Channels" IEEE Trans. On Communications. **DOI:** <u>10.1109/TCOMM.2021.3087128</u>



WZ Region Approximation



Intersection Analysis



Admissible WZ rate region

with:
$$\alpha = \frac{I(X;Y)}{I(X;U|Y) - I(X;U)}$$
 $\beta = \frac{I(X;U) - I(X;U|Y) \cdot H(Y|X)}{I(X;U|Y) - I(X;U)}$

68



Two scenarios that outage happens.

 $P_{out,total} = P_{out,1} + P_{out,2}$

69



Twofold Integral Similar to Lossless Case

-

$$\begin{aligned} &P_{out,MAC,1} \\ &= \Pr\left\{ 0 \leq \frac{1}{R_{c,1}} C\left(\gamma_1\right) \leq I\left(X; U | Y\right); \frac{1}{R_{c,2}} C\left(\gamma_2\right) \geq 0 \right\} \\ &= \Pr\left\{ \Phi\left(0\right) \leq \gamma_1 \leq \Phi\left[I\left(X; U | Y\right)\right]; \gamma_2 \geq \Phi\left(0\right) \right\} \\ &= \int_{\Phi(0)}^{\Phi\left[I\left(X; U | Y\right)\right]} \int_{\Phi(0)}^{+\infty} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \Pr\left\{ \begin{array}{c} I\left(X; U | Y\right) \leq \frac{1}{R_{c,1}} C\left(\gamma_1\right) \leq I\left(X; U\right); \\ 0 \leq -\frac{R_{c,1}}{R_{c,2}} R_1 + \frac{1}{R_{c,2}} C\left(\gamma_1 + \gamma_2\right) \leq \alpha R_1 - \beta \end{array} \right\} \\ &= \int_{0}^{2^{R_{c,1}[H_b(D_X * P_c) - H_b(D_X)] - 1}} \int_{0}^{+\infty} p\left(\gamma_1 \mathcal{W} \mathcal{W} \mathcal{W} \mathcal{W} \mathcal{W} \mathcal{D} \mathcal{C} \mathcal{D} \mathcal{C} \mathcal{C} \mathcal{M} \mathcal{P} \mathcal{A} \mathcal{A} \mathcal{A} \mathcal{A} \mathcal{A} \right\} \\ &= \int_{0}^{\Phi[I(X; U]Y)]} \int_{\Phi(0)}^{M} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \int_{2^{R_{c,1}[I - H_b(D_X)] - 1}} \int_{0}^{M} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 \\ &= \int_{2^{R_{c,1}[H_b(D_X * P_c) - H_b(D_X)] - 1}} \int_{0}^{M} p\left(\gamma_1\right) p\left(\gamma_2\right) d\gamma_2 d\gamma_1 , \end{aligned}$$

with
$$M = 2^{-\beta R_{c,2}} (1 + \gamma_1)^{\left(\frac{\alpha R_{c,2}}{R_{c,1}} + 1\right)} - 1 - \gamma_1.$$



WZ MAC Outage Probability






Extension to Networks



2.4 Two Stage Wyner-Ziv Network: Distortion Transfer Analysis

Analyze: How Distortion causing at the previous stage is **forwarded** to the current stage? Distortion Transfer Function (DTF)



Amin Zribi, Lin Wensheng, Reza Asvadi, Elsa Dupraz, Tad Matsumoto, "Two-stage Successive Wyner-Ziv Lossy Forward Relaying for Lossy Communications: Rate-distortion and Outage Probability Analyses", Under Review, IEEE Trans. Vehcular Technology.

Block Diagram



Mathematical assumptions

- i.i.d Bit-Flipping model for the correlation between the observations
- Hamming distortion measure
- Definition of Admissible rate-distortion (RD) region:





Stage-Independent RD Analysis

Stage-by-stage admissible RD region

Stage 1

Stage 2



Smaller rate R_0 is required when D_H is larger or p_h is lower



when D_X is larger or D_H is smaller

75

The Recursive Structure of the Binary Convolution is Referred to as: Distortion Transfer Function (DTF)

DTF Connects the two stages, as

- Stage 1:
 - Assume the required distortion D_H at Helper is given by D_H^*
 - It is found that Bit-Flipping probability between the observations should satisfy

$$p_{h} \leq \Lambda \Big[D_{H}^{*}, H_{b}^{-1} \Big(R_{0} + H_{b} \big(D_{H}^{*} \big) \Big) \Big]$$

to distortion requirement
on Side information From distortion requirement
at Helper

- ▶ When $R_0 \ge 1 H_b(D_H^*)$, using $\Lambda(y, t = 0.5) = 0.5$, we have $p_H \le 0.5$ □ which corresponds to the case no side information is required.
- When R_0 decreases, $H_b^{-1}(R_0 + H_b(D_H^*))$ also decreases, and hence also p_H decreases,
 - \square which corresponds to the case higher correlation is needed to satisfy D_H^*

 $R(D)=H_b(p)-H_b(D)$ \downarrow $p=H_b^{-1}(R+H(D))$



Stage-Dependent RD Analysis: Connecting Stages



Stage 2

- Assume the required distortion at Destination is given by D_X
- It is found that the distortion at Helper should satisfy

$$D_{H}^{*} \leq \Lambda \left[p, \Lambda \left[D_{X}, H_{b}^{-1} \left(R_{1} + H_{b} \left(D_{X} \right) \right) \right] \right]$$
DTF
From disterior

to distortion requirement at Helper

From distortion requirement at Destination

• When R_2 is large enough, $p = H_b^{-1}(1 - R_2) = 0$, then using $\Lambda(0, t) = t$:

$$D_H^* \leq \Lambda \left[D_X, H_b^{-1} \left(R_1 + H_b(D_X) \right) \right] = D_{SI}$$

- In this case, Stage 2 is equivalent to Stage 1 with distortion requirement D_{SI} on Side Information.
- Condition $p \leq D_{SI}$ is required to achieve D_X at Destination



Rate Surface Calculation

Connecting Stage 1 and Stage 2

> 3D admissible RD region



Outage Probability



• Assumptions:

- Static stage 1: D_H is a fixed parameter
- Destination is moving: fading variation on S-D and H-D links
- \rightarrow When rates R_1 and R_2 are in the inadmissible RD region: outage happens



Twofold Integrals for Outage Probability Calculation



Utilizing Lossy Source-Channel Separation theorem,

$$R_1 \le \Phi_S(\gamma_S) \triangleq \frac{C(\gamma_S)}{R_c^{SD}} \qquad R_2 \le \Phi_S(\gamma_H) \triangleq \frac{C(\gamma_H)}{R_c^{HD}}$$

- $C(\gamma) = \log_2(1 + \gamma)$: the channel capacity function with two dimensional signaling
- Case 1 outage probability:

• Case 2 outage probability:

$$P_{out}^{C_1} = \Pr\{0 \le R_1 \le L_S, 0 \le R_2\}$$

= $\Pr\{0 \le \Phi_S(\gamma_S) \le L_S, 0 \le \Phi_H(\gamma_H)\}$
= $\Pr\{\Phi_S^{-1}(0) \le \gamma_S \le \Phi_S^{-1}(L_S), \Phi_H^{-1}(0) \le \gamma_H\}$
= $\int_{\Phi_H^{-1}(0)}^{+\infty} \int_{\Phi_S^{-1}(L_S)}^{\Phi_S^{-1}(L_S)} p(\gamma_S, \gamma_H) d\gamma_S d\gamma_H$

$$P_{out}^{C_2} = \Pr\{L_S \le R_1 \le H_S(\gamma_H), 0 \le R_2 \le 1\}$$

= $\Pr\{L_S \le \Phi_S(\gamma_S) \le H_S(\gamma_H), 0 \le \Phi_H(\gamma_H) \le 1\}$
= $\Pr\{\Phi_S^{-1}(L_S) \le \gamma_S \le \Phi_S^{-1}(H_S(\gamma_H)), \Phi_H^{-1}(0) \le \gamma_H \le \Phi_H^{-1}(1)\}$
= $\int_{\Phi_H^{-1}(0)}^{\Phi_B^{-1}(L_S)} p(\gamma_S, \gamma_H) d\gamma_S d\gamma_H$
 $p = H_b^{-1}(1 - \Phi_H(\gamma_H))$
 $H_S(\gamma_H) = H_b(H_b^{-1}(1 - \Phi_H(\gamma_H)) * D_H * D_X) - H_b(D_X)$

Twofold Integrals for Outage Probability Calculation



For independent fading on S-D and H-D links

$$P_{out}^{C_1} = \frac{1}{\Gamma_S \Gamma_H} \int_{\Phi_H^{-1}(0)}^{+\infty} \int_{\Phi_S^{-1}(L_S)}^{\Phi_S^{-1}(L_S)} \exp\left(-\frac{\gamma_S}{\Gamma_S}\right) \exp\left(-\frac{\gamma_H}{\Gamma_H}\right) d\gamma_S d\gamma_H$$

$$= 1 - \exp\left(\frac{-\Phi_S^{-1}(L_S)}{\Gamma_S}\right),$$

$$\mathbf{TwWKDCpCmPaT!}$$

$$P_{out}^{C_2} = \frac{1}{\Gamma_S \Gamma_H} \int_{\Phi_H^{-1}(0)}^{\Phi_H^{-1}(1)} \int_{\Phi_S^{-1}(L_S)}^{\Phi_S^{-1}(H_S(\gamma_H))} \exp\left(-\frac{\gamma_S}{\Gamma_S}\right) \exp\left(-\frac{\gamma_H}{\Gamma_H}\right) d\gamma_S d\gamma_H$$

$$= \frac{1}{\Gamma_H} \int_{\Phi_H^{-1}(0)}^{\Phi_H^{-1}(1)} \exp\left(-\frac{\gamma_H}{\Gamma_H}\right) \left[\exp\left(-\frac{\Phi_S^{-1}(L_S)}{\Gamma_S}\right) - \exp\left(-\frac{\Phi_S^{-1}(H_S(\gamma_H))}{\Gamma_S}\right)\right] d\gamma_H$$

Outage Probability



For independent fading on the S-D and H-D links



10 **Increasing** the allowed distortion at Destination D_X provides **lower** 10-2 outage probabilities **Outage Probability**O_H=0.0, D_X=0.2 10⁻³**D**....D_H=0.0, D_X=0.3 - D_H=0.01, D_X=0 10⁻⁴ - D_H=0.01, D_x=0.05 However, D_X has no impact on - D_H=0.01, D_x=0.2 the slope of the outage 10⁻⁵ probability (parallel curves) - D_H=0.1, D_X=0 $\nabla - D_{H}^{2}=0.1, D_{X}^{2}=0.05$ $\Theta - D_{H}^{2}=0.1, D_{X}^{2}=0.2$ 10-6 - ⊖ - D_µ=0.1, D_y=0.3

10

For independent fading on the S-D and H-D links





20

15

-10

-5

0

5

Average SNR (dB)

10-7

Impact of Spatial Correlation on Outage Probability



For correlated fading on the S-D and H-D links

• $\rho = \langle h_1, h_2^* \rangle$ the correlation of the complex channel gains h_1 and h_2

The joint PDF of the instantaneous SNRs

$$p(\gamma_S, \gamma_H) = \frac{1}{\Gamma_S \Gamma_H (1 - |\rho|^2)} \exp\left(-\frac{1}{1 - |\rho|^2} \left(\frac{\gamma_S}{\Gamma_S} + \frac{\gamma_H}{\Gamma_H}\right)\right) \times I_0\left(\frac{2|\rho|}{1 - |\rho|^2} \sqrt{\frac{\gamma_S \gamma_H}{\Gamma_S \Gamma_H}}\right)$$

- $I_0(x)$ is the zero-order modified Bessel's function of the first kind **TWWKDE pCmPaT!** $I_0(x) = \sum_{m=0}^{\infty} \frac{p(m!)^2}{(m!)^2} \left(\frac{x}{2}\right)^2$
- The outage probability of cases 1 and 2 can be written as

$$P_{out}^{C_1} \approx \sum_{m=0}^M A_m \int_{\Phi_H^{-1}(0)}^{+\infty} \int_{\Phi_S^{-1}(L_S)}^{\Phi_S^{-1}(L_S)} \left(\frac{\gamma_S \gamma_H}{\Gamma_S \Gamma_H}\right)^m \exp\left(-a\frac{\gamma_S}{\Gamma_S}\right) \exp\left(-a\frac{\gamma_H}{\Gamma_H}\right) \ d\gamma_S \ d\gamma_H,$$
$$P_{out}^{C_2} \approx \sum_{m=0}^M A_m \int_{\Phi_H^{-1}(0)}^{\Phi_H^{-1}(0)} \int_{\Phi_S^{-1}(L_S)}^{\Phi_S^{-1}(H_S(\gamma_H))} \left(\frac{\gamma_S \gamma_H}{\Gamma_S \Gamma_H}\right)^m \exp\left(-a\frac{\gamma_S}{\Gamma_S}\right) \exp\left(-a\frac{\gamma_H}{\Gamma_H}\right) \ d\gamma_S \ d\gamma_H,$$

Impact of Spatial Correlation on Outage Probability



For **correlated fading** on the S-D and H-D links



Fading correlation has no impact on the asymptotic diversity order

Chapter 3 Wyner-Ziv Formulation for Decision Making Process

3.1 Revisit of Helper-aided Lossy Networks

Wyner Ziv Networks:





Notice: $U \rightarrow X \rightarrow Y$ forms a Markov Chain

$$I(X;U) - I(Y;U) = I(XY;U) - I(Y;U| X) - I(Y;U)$$

= $I(XY;U) - I(Y;U) = I(X;U| Y).$

We have used $U \rightarrow X \rightarrow Y$ in the networks of:

- Outage analysis for wireless End-to-End Lossy Communications networks
- A two-stage wireless communications network based on Distortion Transfer Function
- Extension to two-sources one-helper End-to-End Lossy wireless communications network

- ...

Fact: I(X; U) - I(Y; U) = I(X; U|Y) can be understood as:

- Y is training sequence for Machine Learning,
- Y is training sequence, maybe followed by online observation, used for the knowledge updating of 1st and 2nd order statistics, *pdf* and Markov *dynamics*, in Semantic Communications.



Hypothesis Testing HT





3.2 Distributed Hypothesis Testing (DHT) Landmark Builders for Hypothesis Testing (HT): Neyman-Pearson

Information Theoretic formulation of HT: Basically, HT is a Code Design problem: Design $f^{(n)}$ and $g^{(n)}$ such that Minimize Type II Error Probability β_n subject to Type I Error Probability $\alpha_n \leq \varepsilon$

$$\alpha_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right)^{\prime}\right) = H_{1} \mid H_{0} \text{ is true}\right]$$
$$\beta_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right)^{\prime}\right) = H_{0} \mid H_{1} \text{ is true}\right]$$



$$\begin{array}{l} H_0: X \sim \boldsymbol{P}_{0,\mathrm{X}} \\ H_1: X \sim \boldsymbol{P}_{1,\mathrm{X}} \end{array}$$



under constraint the rate *R* being given.

Note: Tradeoff: $\alpha_n \uparrow$, $\beta_n \downarrow$ Note further: Decoder does NOT have to really "decode" to obtain *U*". *e.g.*, by Syndrome check only.



Nyman Pearson Test



With *R*=1.0, the decision problem boils down to traditional Nyman Pearson test using Likelihood:

 $\frac{P_0(X_1, X_2, \dots, X_n)}{P_1(X_1, X_2, \dots, X_n)} > T , \quad [X_1, X_2, \dots, X_n \in X, and H_0: X \sim P_{0,X}] \quad H_1: X \sim P_{1,X}$ Unconstrained $\underbrace{X}_{\text{Encoder}} \xrightarrow{\mathbb{R}}_{\text{Encoder}} \text{Decoder} \xrightarrow{\mathbb{H}_0 \mid \mathbb{H}_1}$ $\alpha_n \text{ and } \beta_n \text{ are given by } \alpha^* = P_1^n(A_n^c(T)), \quad \beta^* = P_2^n(A_n(T)),$

with
$$A_n(T) = \left\{ \frac{P_1(x_1, x_2, \dots, x_n)}{P_2(x_1, x_2, \dots, x_n)} > T \right\}.$$

Note: Tradeoff $\alpha_n \uparrow$, $\beta_n \downarrow$ still holds with the threshold T.



Decoder $g^{(n)}$ does NOT have to really "decode" to obtain U, because the objective is to make a decision under the constraint on rate R.

Objective : find the type-II error exponent θ such that type-I error is imposed and a rate constraint is satisfied

$$\alpha_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right), \mathbf{Y}^{n}\right) = H_{1} \mid H_{0} \text{ is true}\right],\\\beta_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right), \mathbf{Y}^{n}\right) = H_{0} \mid H_{1} \text{ is true}\right].$$

The objective is that minimize β_n , subject to $\alpha_n \leq \varepsilon$. Tradeoff: $\alpha_n \uparrow$, $\beta_n \downarrow$

Ismaila Salihou Adamou, Elsa Dupraz1, Amin Zribi1, and Tad Matsumoto, "Error-Exponent of Distributed Hypothesis Testing for Gilbert-Elliot Source Models", to be published, Proc of IEEE ISTC 2023, Brest



DHT Problems:



(1) In the same way as HT, again, a DHT problem is a Code Design problem: Design $f^{(n)}$ and $g^{(n)}$ such that

Minimize Type II Error Probability β_n subject to Type I Error Probability $\alpha_n \leq \varepsilon$ under constraint on the rate *R* given.

$$\alpha_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right), \mathbf{Y}^{n}\right) = H_{1} \mid H_{0} \text{ is true}\right]$$

$$\beta_{n} = \mathbb{P}\left[g^{(n)}\left(f^{(n)}\left(\mathbf{X}^{n}\right), \mathbf{Y}^{n}\right) = H_{0} \mid H_{1} \text{ is true}\right]$$

$$H_0: (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{0,XY},$$
$$H_1: (\mathbf{X}^n, \mathbf{Y}^n) \sim P_{1,XY}.$$



(2) Find Type-II Error Exponent θ Subject to $\alpha_n \leq \epsilon$ and $\limsup_{n \to \infty} \frac{1}{n} \log M_2 \leq R$



École Mines-Télécom

DHT: Type-II Error Exponent θ

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_n} \ge \theta$$

It has already been known that:

 $\theta = \min \left\{ \frac{G(P_{0,UXY}, R)}{binning\text{-}error}, \frac{D(P_{0,UXY} || P_{1,UXY})}{testing\text{-}error} \right\}$ with the binning-error part being: $G(P_{0,UXY}, R) = R - [I(X; U) - I(U; Y)]$

and

 $D(P_{0,UXY} || P_{1,UXY})$ is the KL divergence

We rewrite the *binning-error* part:

G(R,R(D))=R-R(D)

with

 $R(D) = \min \{ I(X; U) - I(U; Y) \} = \min I(X; U| Y) -.-(1)$

being Wyner Ziv R(D) function! \rightarrow The DHT Problem (1) boils down to WZ Coding Problem, because $U \rightarrow X \rightarrow Y$ forms a Markov Chain. So far up to this point, formulations are generic, and hence distributions are not specified.



Objective: Minimize Type II Error Probability β_n subject to Type I Error Probability $\alpha_n \le \varepsilon$ under constraint on the rate *R* given.



Binary DHT: Recent Results



- Compression of **X** by Short Linear Block Code
- Y is not compressed.
- Nayman-Pearson Test

Consider a binary linear code C defined by a $k \times n$ generator matrix **G** with a rate of $R = \frac{k}{n}$ as the binary quantizer component [17]. According to the standard array concept [18], the minimum Hamming weight vector d_H (.) in each coset C_s associated to the syndrome s is referblack to as the coset leader defined as

$$L(\mathcal{C}_{\mathbf{s}}) \triangleq \arg\min_{\mathbf{z}\in\mathcal{C}_{\mathbf{s}}} d_H(\mathbf{z}),$$

Encoding:
$$\mathbf{u}_q^k = \arg \min_{\mathbf{u}^k \in \{0,1\}^k} d_H \left(\mathbf{x}^n, \mathbf{x}_q^n\right)$$

Neyman Pearson Test: $\sum_{j=1}^n (x_{q,j} \oplus y_j) \leq \gamma_t$, where $\mathbf{x}_q^n = \mathbf{u}^k \mathbf{G}$

 γ_t is the threshold that determines the α_n and the β_n values.



English Letters Appearence Probabilities



ATST

ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Semantic Communications

 \overline{x}



Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document



Copied from:



(a) P(y|x)

(b) $P(x \mid y)$

Conditioned English Letter Appearance

If we can evaluate conditional probabilities $p(x_i|x_{i-1})$, $p(x_i|x_{i-1}, x_{i-2})$, ..., $p(x_i|x_{i-1}, x_{i-2}, ..., x_{i-n})$, empirically or theoretically and create a Markov model of the letter appearances, we can reduce the rate required to encode English.

Shannon's landmark paper presents artificially created English sentences!



School of Information Science



Using empirical – knowledge of $p(x_i|x_{i-1}, x_{i-2}, x_{i-3})$

Using empirical knowledge of the word appearance probability p(w_i)

 $p(w_i|w_{i-1}, w_{i-2})$

Higher-order Conditioning: Letter and Word Levels

Fourth-order approximation. (The frequency of quadruplets of letters matches English text. Each letter depends on the previous three letters. This sentence is from Lucky's book, *Silicon Dreams* [183].)

THE GENERATED JOB PROVIDUAL BETTER TRAND THE

DISPLAYED CODE, ABOVERY UPONDULTS WELL THE

CODERST IN THESTICAL IT DO HOCK BOTHE MERG.

(INSTATES CONS ERATION. NEVER ANY OF PUBLE AND TO

THEORY. EVENTIAL CALLEGAND TO ELAST BENERATED IN

WITH PIES AS IS WITH THE)

Instead of continuing with the letter models, we jump to word models.

First-order word model. (The words are chosen independently but with frequencies as in English.)

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME

CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO

OF TO EXPERT GRAY COME TO FURNISHES THE LINE

MESSAGE HAD BE THESE.

 Second-order word model. (The word transition probabilities match English text.)

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH

WRITER THAT THE CHARACTER OF THIS POINT IS

THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE

TIME OF WHO EVER TOLD THE PROBLEM FOR AN

UNEXPECTED

With the 4th order model, Shannon showed that 2.8 bits are enough to express one English letter!



Adaptive Morse Code Semantic Communications



Morse Code Alphabet

Α	•-	Ν	-•	0
В		0		1 •
С	-•-•	Ρ	••	2 ••
D	-••	Q	•-	3 •••
Е	•	R	•_•	4 ••••-
F	••_•	S	•••	5 ••••
G	•	Т	-	6 -••••
Н	••••	U	••-	7•••
1	••	V	•••	8•
J	•	W	•	9•
K	-•-	Х		• •-•-•-
L	•_••	Y	_•	,••
М		Ζ	••	? ••••

The higher the appearance probability, the shorter the code length, following the Huffman coding rule.

However, the appearance probabilities should change according to the sources, such as Book, Video, File type,, situation, person, ... ← Semantic Dependency.

Joint Source and Channel Coding and Adaptive, to exploit higher order Markov Memory Structure and error correction, depending on "Semantics" → Learning needed to construct Corpus for Natural Language Processing!





Knowledge updating for Semantic Communications: a WZ Problem



Remembering Martin Luther King Jr. | Tory Daily (toryburch.jp)





Knowledge updating for Semantic Communications: a WZ Problem Undeting the L

Updating the knowledge to under the rate R given, so that:

Oulu



Remembering Martin Luther King Jr. | Tory Daily (toryburch.jp)



SEMANTIC COMMUNICATION COVC FOR THE INTERNET OF VEHICLES

A Multiuser Cooperative Approach

Wenjun Xu[©], Yimeng Zhang[©], Fengyu Wang[©], Zhijin Qin[©], Chenyao Liu[©], and Ping Zhang[©]

IEEE VTS Magazine Volume 18, No. 1, pp. 100-109





École Mines-Télécom

Semantic Forward





We know already (see Lossless Relaying) $R_0 \ge I(X;Y)$ $R_1 \ge H(X|U),$ $R_2 \ge I(Y;U).$ With Semantic encoder and decoder, $R_1 \ge H(X|U,V)$

 $R_2 \ge I(Y; U|V).$

V reduces the required rate \rightarrow Compression





(b) $\gamma_1 = -5 \text{ dB}, \rho = 0.1.$

0.31279

0.11853



7

0.31679

0.11361

 $Y = X \oplus E$, where $E \sim \text{Bern}(\rho)$



Original

(a) Example with $\rho = 0$.

0.31285

0.16300



3.4 Training Process in Machine Learning



Updating the knowledge to minimize the rate **R** so that:





IMT Atlantique

Bretagne-Pays de la Loire École Mines-Télécom



Oulu

X corresponding to the Current Observation, *U* to the Lossy Reconstruction, ¹ and *Y* to Data Set for the Learning of Probability Distribution for knowledge updating followed by Codebook generation!







Learning in the WZ framework: Open Questions

(1) *Y* is fully covered by sub-probability space Y_i without overlapping.

(2) *Y* is NOT fully covered. Y_i are overlapping.









(1) is suitable when *decision* is *Ergodic* (time average).

→Learning requires Large size of training data. Suitable for pre-training, such as ML.

(2) is suitable when *decision* is *Instantaneous*. Learning data may require only partial data.

Suitable for online-training by introducing a forgetting factor. A similarity to Information Bottleneck!





Connection to Information Bottleneck





On the Information Bottleneck Problems: Models, Connections, Applications and Information Theoretic Views

Abdellatif Zaidi ^{1,2,*} and Iñaki Estella-Aguerri ² and Shlomo Shamai (Shitz) ³

Specifically, IB formulates the problem of extracting the relevant information that some signal $X \in \mathcal{X}$ provides about another one $Y \in \mathcal{Y}$ that is of interest as that of finding a representation U that is maximally informative about Y (i.e., large mutual information I(U;Y)) while being minimally informative about X (i.e., small mutual information I(U;X)).

$$Y \in \mathcal{Y} \longrightarrow \underbrace{P_{X|Y}} \xrightarrow{X \in \mathcal{X}} \phi \xrightarrow{U = \phi(X)} \psi \longrightarrow \widehat{Y} \in \mathcal{Y}$$



Information Bottleneck as a classification problem

Specifically, IB formulates the problem of extracting the relevant information that some signal $X \in \mathcal{X}$ provides about another one $Y \in \mathcal{Y}$ that is of interest as that of finding a representation U that is maximally informative about Y (i.e., large mutual information I(U;Y)) while being minimally informative about X (i.e., small mutual information I(U;X)). In the IB framework, I(U;Y) is referred

Medical Data Analysis: an Example



This term can not be ZERO because X has some information about Y.

The encoder and decoder need to have some medical factor. The roles can be performed by DNN. They need "training".


Information Bottleneck: Formulation under WZ Framework!

Accordingly, for a given β and source distribution $P_{X,Y}$, the optimal mapping of the data, denoted by $P_{U|X}^{*,\beta}$, is found by solving the IB problem, defined as

 $\mathcal{L}^{\mathrm{IB}}_{\beta}(P_{U|X}) \coloneqq I(U;Y) - \beta I(U;X)$

with $U \rightarrow X \rightarrow Y$ and Lagrange Multiplier β . \longrightarrow We can use some optimization tools.





École Mines-Télécom





Welcome to ITW 2024 (https://IEEE-ITW2024.org)

Register a paper for 2024 IEEE Information Theory Workshop

Topics

- Deep Learning for Communication Networks
- Detection and Estimation
- Distributed Storage and Computing
- Emerging Applications of Information Theory
- Information Theory and Statistics
- Information Theory for Decision and Control
- Information Theory in Biology
- Information Theory in Computer Science



ľ	TV	۷	2	0	2	4
	•••	•	-	~	-	





Any Questions?









Do they meet in 6G Networks?

My SISU Continues. Thank you!

See you soon again somewhere in the world!













