2011

Correlated Sources In Distributed Networks - Data Transmission, Common Information Characterization and Inferencing

Wei Liu
Syracuse University

Abstract

Correlation is often present among observations in a distributed system. This thesis deals with various design issues when correlated data are observed at distributed terminals, including: communicating correlated sources over interference channels, characterizing the common information among dependent random variables, and testing the presence of dependence among observations.

It is well known that separated source and channel coding is optimal for point-to-point communication. However, this is not the case for multi-terminal communications. In this thesis, we study the problem of communicating correlated sources over interference channels (IC), for both the lossless and the lossy case. For lossless case, a sufficient condition is found using the technique of random source partition and correlation preserving codeword generation. The sufficient condition reduces to the Han-Kobayashi achievable rate region for IC with independent observations. Moreover, the proposed coding scheme is optimal for transmitting a special correlated sources over a class of deterministic interference channels. We then study the general case of lossy transmission of two correlated sources over a two-user discrete memoryless interference channel (DMIC). An achievable distortion region is obtained and Gaussian examples are studied.

The second topic is the generalization of Wyner's definition of common information of a pair of random variables to that of N random variables. Coding theorems are obtained to show that the same operational meanings for the common information of two random variables apply to that of N random variables. We establish a monotone property of Wyner's common information which is in contrast to other notions of the common information, specifically Shannon's mutual information and Gács and Körner's common randomness. Later, we extend Wyner's common information to that of continuous random variables and provide an operational meaning using the Gray-Wyner network with *lossy* source coding. We show that Wyner's common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding.

Finally, we consider the problem of *distributed* test of statistical independence under communication constraints. Focusing on the Gaussian case because of its tractability, we study in this thesis the characteristics of optimal scalar quantizers for distributed test of independence where the optimality is both in the finite sample regime and in the asymptotic regime.

Correlated Sources In Distributed Networks - Data Transmission, Common Information Characterization and Inferencing

by

Wei Liu
M.S. Syracuse University, Syracuse, NY, U.S.A, 2009
B.S. University of Science & Technology of China, Hefei, China, 2006

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering in the Graduate School of Syracuse University

May 2011

Approved: Dr. Biao Chen

Acknowledgement

I am truly privileged to have the opportunity to work with my advisor, Dr. Biao Chen. I am heartily thankful for his guidance, encouragement and support throughout my Ph.D study. Dr. Chen is a knowledgeable scholar, and a great advisor. Without him, I would not have finished this degree, nor would I have such a strong motivation and interest in research. I benefit tremendously from his wisdom, vision and insight into research topics. The weekly meeting with him will be one of my best memories in my graduate study. I will always remember the day we discussed on a white board for hours about the problems of sending correlated Gaussian signals over Gaussian interference channels. I am deeply indebted to his insightful suggestions on my research problems. His enthusiasm, responsibility and hardworking make him a role model for my future career. His understanding, encouragement and patience are the basis for this thesis.

I want to thank Dr. Lixin Shen for chairing my Ph.D thesis defense. I also want to thank Dr. Pramod Varshney for his kind encouragement during my first academic presentation at Asilomar conference, 2007. I am also very grateful to him for chairing my Ph.D. oral qualifying exam and my Master thesis defense.

Thanks to my officemates and friends: Bin Liu, Ying Lin, Xiaohu Shang, Yi Cao, Jin Xu, Minna Chen, Kapil Borle, Fangfang Zhu, Ge Xu, Fangrong Peng, Pengfei Yang and Yu Zhao with whom I have shared a great time in my study. The discussions with them benefit me a lot. Special thanks to Bin Liu, Xiaohu Shang, Yi Cao and Jin Xu who helped me a lot during the first several months after I arrived at U.S.

I thank all my defense committee members for carefully reading my thesis and giving me helpful suggestions. They are (in alphabetic order) Dr. Makan Fardad, Dr. Yingbin Liang, Dr. Ruixin Niu, Dr. Lixin Shen and Dr. Pramod Varshney.

Special thanks to Ge Xu for coauthoring the common information chapter in this thesis and Minna Chen for coauthoring the testing of independence chapter.

Last but not the least, I want to thank my family for their encouragement, understanding and support for my studying towards my Ph.D. degree.

Contents

Abstract **i**

Acknowledgement **iv**

List of Figures **vii**

1 Introduction **1**

1.1 Discrete memoryless interference channels 3

1.2 Communicating correlated sources over MAC 4

1.3 Communicating correlated sources over BC 6

1.4 Wyner's common information for two random variables 8

1.5 Hypothesis testing with communication constraints 10

1.6 Outline of the thesis 13

2 Interference Channels With Arbitrarily Correlated Sources: The Lossless Case **16**

2.1 Problem statement 17

2.2 A sufficient condition 18

2.2.1 Reduce to Han and Kobayashi's region on interference channels 21

2.2.2 Reduce to Jiang *et al* and Cao *et al*'s region on interference channels with common information 22

2.2.3 A necessary and sufficient condition for a class of deterministic interference channels 23

2.3 Conclusion 25

3 Communicating Correlated Gaussian Sources over Gaussian Z Interference Channels **26**

3.1 Gaussian multiple access channels with single distortion constraint . . 29

3.1.1 System model 29
3.1.2 Lower and upper bound on the distortion constraint 29
3.1.3 Comparison 35
3.2 Gaussian Z Interference Channel With Correlated Gaussian sources . 36
3.2.1 System Model 36
3.2.2 Inner and upper bounds on the achievable distortion region . . 37
3.3 Conclusion and Discussions 40

4 Communicating Correlated Sources Over Interference Channels: The Lossy Case 42
4.1 Arbitrarily correlated sources: the general case 44
4.1.1 Sufficient conditions: a joint source-channel coding approach . 44
4.2 The Gaussian special case 47
4.2.1 Lower and upper bounds on the achievable distortion region . 48
4.2.2 The strong interference regime 50
4.2.3 The noisy interference regime 52
4.2.4 Numerical examples 52
4.3 Summary 53

5 The Common Information of N Dependent Random Variables 56
5.1 Common information of N random variables 57
5.1.1 A Gray-Wyner source coding network interpretation 57
5.1.2 A random variable generator interpretation 58
5.1.3 Main results: $C_1 = C_2 = C$ 59
5.2 Examples and discussions 59
5.2.1 The monotone property of common information C 59
5.2.2 Circularly symmetric binary sources 61
5.2.3 Asymptotic results for N binary random variables 63
5.3 Conclusions 64

6 Common Information For Continuous Random Variables 66
6.1 A Gray-Wyner lossy source coding interpretation 67
6.2 Examples 70
6.2.1 Doubly symmetric binary source (DSBS) 70
6.2.2 Gaussian source 71
6.3 Conclusion 72

7 **Quantization for Distributed Testing of Independence** **73**
7.1 Problem statement and main results 75
7.1.1 Large sample regime 76
7.1.2 Finite sample regime 77
7.2 Numerical examples 78
7.3 Conclusion 81

8 Conclusion and future work **82**

A Proof of the sufficient condition: Theorem 1 **85**

B Proof of the necessary and sufficient condition: Theorem 2 **93**

C Proof of the sufficient condition: Theorem 8 **100**

D Proof of the main results: Theorem 10 **104**

E Proof of $C_1 = C$ **106**

F Proof of $C_2 = C$ **113**

G Proof of the sufficient condition of $C_3 = C$: Theorem 11 **114**

H Proof of the optimality of the distortion threshold: Theorem 12 **119**

I Proof of the optimality of quantization threshold: Theorem 13 **121**

J Proof of the optimal local decision rule : Proposition 15 and 16 **125**

K Proof of Proposition 13 **127**

L Proof of Corollary 7 **129**

Bibliography **129**

List of Figures

1.1 A two user DMIC. 4

1.2 Source coding over a simple network. 10

1.3 Random variable generators. 10

2.1 Interference channels with correlated sources. 17

2.2 Coding structure for IC with correlated sources 19

2.3 A class of deterministic interference channel with correlated sources. . 24

3.1 Gaussian ZIC with correlated Gaussian sources. 27

3.2 GMAC with a single distortion constraint. 27

3.3 Comparison of D_{l1} and D_{l2} as a function of ρ for $P_1 = P_2 = 10, \sigma = 1, N = 1$. 31

3.4 HDA scheme and its CEO representation. 34

3.5 Comparison of D_u, D_s and D_h together with two lower bounds D_{l1} and D_{l2} as a function of $\frac{P}{N}$ for $\sigma = 1, \rho = 0.5$, for simplicity, we assume $P_1 = P_2$. 35

3.6 Comparison of D_u, D_s and D_h together with two lower bounds D_{l1} and D_{l2} as a function of ρ for $\sigma = 1$ and $P_1 = 20, P_2 = 20, N = 1$. 36

4.1 Lossy transmission of correlated sources over a DMIC. 43

4.2 Comparison of different achievable schemes for a=1, where lower bound 1 is by Proposition 11 and lower bound 2 is from Theorem 9. 54

4.3 Comparison of different achievable schemes for a=2, where the lower bound is from Theorem 9. 54

4.4 Comparison of different achievable schemes for a=0.1. 55

5.1 A simple Bayesian graph model. 62

5.2 Comparison of common information. 63

6.1 Gray-Wyner source coding network. 67

7.1 Hypothesis testing with communication constraints. 76
7.2 Plot of $I(U;V)$ as a function of thresholds t_1 and t_2 for $\rho = 0.65$. . . 79
7.3 Probability of error for spectrum sensing. 80

Chapter 1

Introduction

Shannon, in his landmark work [1] in 1948, showed that for point-to-point communication, separated source and channel coding incurs no loss of optimality in terms of reliable transmission. This result has important practical applications. Traditional communication systems follow exactly such an approach by separately and independently designing source codes and channel codes. The combination of the two will achieve the same optimality as designing the system considering the source and channel coding together.

However, today's communication systems are usually distributed, i.e., they consist of multiple transmitters and multiple receivers. Typical multiuser channels includes multiple access channels (MAC), broadcast channels (BC) and interference channels (IC). A MAC models the situation where several transmitters compete for a common communication medium and send individual information to one central node, while in a BC, one central transmitter node broadcast independent information to different users. An IC, on the other hand, models the situation where all transmitters and receivers are distributed.

It has been shown in the literature that separated source and channel coding fails to achieve optimality for multiuser communication systems. Cover, El-Gamal, and Salehi first studied the problem of communicating discrete correlated sources over a MAC [2], and gave a concrete example showing that separated source and channel coding scheme is strictly suboptimal. Han and Costa [3] studied the problem of communicating arbitrarily correlated sources over a discrete memoryless BC and also provided an example showing that the separated scheme is strictly suboptimal for broadcast channels. Recent results on sending correlated Gaussian sources over a Gaussian MAC in [4] proposed joint source and channel coding schemes that uniformly outperform separated source and channel coding. All of these works imply that a new communication framework is needed for communication over multi-terminal networks, that is, the joint design of source and channel coding. This motivates the study of the first topic of this thesis, namely communicating correlated sources over multiuser channels.

Our primary focus is on interference channels. Due to the decentralized transceiver structure, communication between one sender and its corresponding receiver will cause interference to the communication between other senders and their corresponding receivers. The capacity region of a two user interference channel, where two independent sources are transmitted through a two user interference channel, has been a long standing problem. The first information-theoretic study of this problem dates back to Shannon [5], and extensive research has since been carried out by many authors [6,7,8,9,10,11,12,13,14,16,17,18]. Unfortunately, the problem of characterizing the capacity region of interference channels remains open except for some special cases, such as ICs with strong and very strong interference [7, 11, 13], and a class of deterministic interference channel [14]. So far, the best known achievable region is due to Han and Kobayashi [12], which includes all the cases where the capacity region is known as special cases.

In this thesis, we study the transmission of two arbitrarily correlated sources over a two-user interference channel. We aim to provide a better understanding on the fundamental performance limits and what is the optimal way in designing source channel matching code utilizing the dependence structure of the sources. Both lossless and lossy cases are studied, i.e., the receivers need to recover their sources with arbitrarily small probability of errors or within certain distortion constraints. The special case of Gaussian sources over Gaussian ICs is also studied.

For arbitrarily correlated sources, the notion of common information arises naturally. In the case that two correlated random variables X and Y can be decomposed as $X = (X^{'}, K)$ and $Y = (Y^{'}, K)$ such that $X^{'}, Y^{'}, K$ are mutually independent, the common information between X and Y is clearly $H(K)$, where H is the Shannon entropy function. For general sources, measuring the common information is not an easy task. Indeed, characterizing the common information of dependent random variables has been a topic of research interest in the past decades [56,57,58,59,60] and our focus is on Wyner's common information. However, Wyner's common information was originally proposed for two random variables with finite alphabet. Our focus in this part of the thesis is to generalize this classical notion of common information. In particular, we provide a generalized Wyner's common information for N arbitrarily distributed random variables, as well as its extension to continuous random variables.

Test of statistical independence between random variables X and Y has been a classical inference problem [74] and has found a wide range of applications, e.g., in image processing [75], economics [76]. The emerging wireless sensor networks bring new dimensions and challenges to this classical problem as the data are no longer centrally available. Dependence detection in distributed systems is often the first and crucial step in event detection/identification; thus its relevance in various sensor network applications is quite evident. In the centralized case where X and Y sequences are available, this statistical inference problem can be solved straightforwardly by applying some standard statistical inference frameworks [77]. The problem becomes much more interesting and complicated when X and Y are not directly available; instead, compressed versions of X and Y subject to some rate constraints are used for the test of independence. Focusing on the Gaussian case because of its tractability, we study in this thesis the characteristics of optimal scalar quantizers for distributed test of independence where the optimality is in the sense of optimizing the error exponent. We also discuss the optimal quantizer properties for the finite sample regime, i.e., that of directly minimizing the error probability.

In the following, we provide some background knowledge related to this thesis. We start with some primer on interference channels. This is followed by a comprehensive overview of the state of art on communicating correlated sources over multiuser channels. The notion of common information as well as the problem of hypothesis testing with communication constraint are introduced later. We conclude this section by giving an outline of this thesis.

1.1 Discrete memoryless interference channels

The interference channel consists of several transmitters and receivers, where each transmitter sends information to its intended receiver while causing interference to all other receivers. A two user discrete memoryless interference channel (DMIC) can be denoted as $\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2, p(y_1y_2|x_1x_2)\}$, where X_1, X_2 are the channel inputs, Y_1, Y_2 are the channel outputs and $p(y_1y_2|x_1x_2)$ is the channel transition probability. The model is shown in Fig. 1.1.

The capacity region for the general DMIC is still unknown. The best known achievable region is due to Han and Kobayashi [12] by using superposition coding

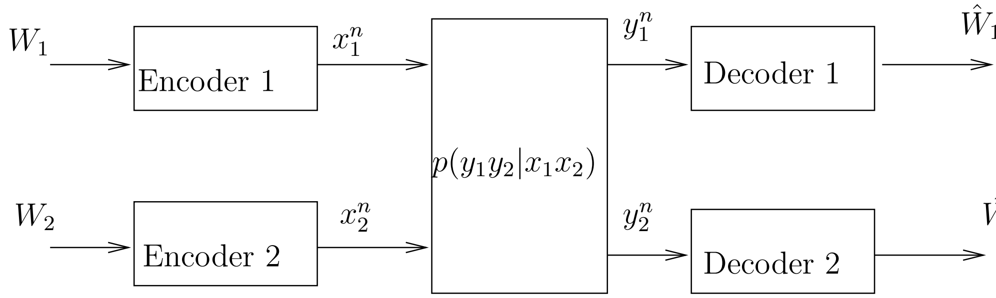

Figure 1.1: A two user DMIC.

and joint decoding. A simplified version of the Han and Kobayashi (HK) region was given by Chong *et al* [18], repeated below.

Proposition 1 *(Chong* et al *[18, Theorem 2]) Let $\mathcal{P}$ be the set of probability distributions that factor as $P(q, w_1, w_2, x_1, x_2) = p(q)p(w_1|q)p(w_2|q)p(x_1|w_1q)p(x_2|w_2q)$. Then the rate pair (R_1, R_2) is achievable for a discrete memoryless interference channel $p(y_1y_2|x_1x_2)$, if the following conditions are satisfied:*

$$
\begin{aligned}
R_1 &< I(X_1;Y_1|QW_2), && (1.1)\\
R_2 &< I(X_2;Y_2|QW_1), && (1.2)\\
R_1 + R_2 &< I(X_1;Y_1|QW_1W_2) + I(W_1X_2;Y_2|Q), && (1.3)\\
R_1 + R_2 &< I(X_2;Y_2|QW_1W_2) + I(W_2X_1;Y_1|Q), && (1.4)\\
R_1 + R_2 &< I(W_2X_1;Y_1|QW_1) + I(W_1X_2;Y_2|QW_2), && (1.5)\\
2R_1 + R_2 &< I(X_1;Y_1|QW_1W_2) + \\
&\quad I(W_2X_1;Y_1|Q) + I(W_1X_2;Y_2|QW_2), && (1.6)\\
R_1 + 2R_2 &< I(X_2;Y_2|QW_1W_2) + \\
&\quad I(W_1X_2;Y_2|Q) + I(W_2X_1;Y_1|QW_1). && (1.7)
\end{aligned}
$$

1.2 Communicating correlated sources over MAC

Communicating correlated sources over multi-terminal networks have been a topic of research interest in the past decades. Slepian and Wolf [19] studied the problem of communicating correlated information over a two-user multiple access channel where the correlation is of a special structure in the form of three independent sources, with one of them observed by both encoders while each of the other two observed only at individual encoders. Later, Cover, El Gamal, and Salehi studied the problem of communicating discrete correlated sources over a multiple access channel (MAC) [2], where the correlation structure can be arbitrary. A sufficient condition was obtained for the lossless transmission of such correlated source pair over a multiple access channel.

Proposition 2 *([2, Theorem 1]) A source pair* $(S^n, T^n) \sim \prod_{i=1}^{n} p(s_i, t_i)$ *can be sent with arbitrarily small probability of error over a discrete memoryless multiple access channel* $p(y|x_1, x_2)$ *if*

$$\begin{aligned} H(S|T) &< I(X_1; Y|X_2TW), && (1.8)\\ H(T|S) &< I(X_2; Y|X_1SW), && (1.9)\\ H(ST|K) &< I(X_1X_2; Y|KW), && (1.10)\\ H(ST) &< I(X_1X_2; Y), && (1.11) \end{aligned}$$

where

$$p(s, t, w, x_1, x_2, y) = p(w)p(s, t)p(x_1|sw)p(x_2|tw)p(y|x_1x_2), \tag{1.12}$$

and $K = f(S) = g(T)$ *is the common part of two variables* (S, T)*, in the sense of Gács, Körner [23] and Witsenhausen [58].*

The key technique in deriving the sufficient condition is the correlation preserving codeword generation. For fixed distribution $p(w), p(x_1|w, s), p(x_2|w, t)$, independently generate one codeword $w^n(k^n) \sim \prod_{i=1}^{n} p(w_i)$ for each $k^n \in \mathcal{K}^n$ that carries the information of the common part. Next, for each source sequence $s^n \in \mathcal{S}^n$, find the corresponding $k^n = f(s^n) = (f(s_1), f(s_2), \cdots, f(s_n))$ and independently generate one codeword $x_1^n \sim \prod_{i=1}^{n} p(x_{1i}|s_i, w_i)$. The codeword x_2^n is similarly generated. Therefore the correlation between the source pair induces correlation in the generated codewords, the so-called correlation preserving codeword generation. To transmit s^n, encoder 1 sends the corresponding codeword x_1^n. Similarly encoder 2 sends the corresponding codeword x_2^n for the given source sequence $t^n \in \mathcal{T}^n$. The decoder uses joint typicality decoding: upon receiving y^n, the decoder finds a unique pair of (s^n, t^n) such that $(s^n, t^n, k^n, w^n, x_1^n, x_2^n, y^n) \in T_\epsilon^n(STKWX_1X_2Y)$.

This sufficient condition includes various known capacity results as its special cases. These include the capacity region for a MAC [20, 21]; distributed lossless source coding, i.e., the Slepian-Wolf coding [22]; cooperative multiple access channel capacity; and the correlated source multiple access channel capacity region of Slepian and Wolf [19]. The key technique used in [2], aside from making use of the so-called common part of correlated random variables (in the sense of Gács, Körner [23] and Witsenhausen [58]), is the correlation preserving codeword generation. By generating codewords that depend, probabilistically, on the source sequences, the correlation between the sources induces correlation in the generated codewords. In addition, a simple example was given to show that the separation approach which concatenates a Slepian-Wolf code [22] and the optimal channel code for MAC [20, 21] is strictly suboptimal.

1.3 Communicating correlated sources over BC

Han and Costa [3] studied the problem of communicating arbitrarily correlated sources over a discrete memoryless BC and obtained the following sufficient condition for the lossless transmission of such correlated sources over a BC.

Proposition 3 *(Han and Costa [3], with correction by Kramer and Nair [25]) A source pair* $(S^n, T^n) \sim \prod_{i=1}^n p(s_i, t_i)$ *can be sent with arbitrarily small probability of error over a discrete memoryless broadcast channel* $p(y_1 y_2|x)$ *if there exist auxiliary random variables* W, U, V *satisfying the Markov chain property* $ST \to WUV \to X \to Y_1 Y_2$ *such that*

$$\begin{aligned} H(S) &< I(SWU;Y_1) - I(T;WU|S), & (1.13)\\ H(T) &< I(TWV;Y_2) - I(S;WV|T), & (1.14)\\ H(ST) &< \min\{I(W;Y_1), I(W;Y_2)\} + I(SU;Y_1|W), & (1.15)\\ & \quad +I(TV;Y_2|W) - I(SU;TV|W), & (1.16)\\ H(ST) &< I(SWU;Y_1) + I(TWV;Y_2) - I(SU;TV|W) - I(ST;W). & (1.17) \end{aligned}$$

The sufficient condition derived in [3] (with correction by Kramer and Nair [25]) recovers the Marton region for broadcast channels with independent messages [26, Theorem 2]. In [27], Minero and Kim proposed an alternative coding scheme and the obtained region was shown to be equivalent to that of Han and Costa. It was pointed out in [27] that the common part does not play a role for the broadcast channel case which is consistent with engineering intuition because of the centralized transmitter. We comment here that the same coding scheme proposed by Han and Costa can also be easily modified to obtain the same region without the use of the common part, as to be elaborated in chapter 2.

Notice that the above conditions in Proposition 3, due to Minero and Kim [27], are slightly different from the original expressions in [3]. The original expressions in [3] involve the common part K although the region was shown to be equivalent to that specified in Proposition 2 [27].

The key technique in Han and Coast's coding scheme that is of particular use to our problem is random source partition, which reminisces superposition coding for the channel coding problem. Specifically, source sequences $s^n \in \mathcal{S}^n$, $t^n \in \mathcal{T}^n$ are randomly placed into 2^{nr_1} and 2^{nr_2} cells, respectively. The cell indices for s^n and t_n, denoted by α and β, respectively, play the role as the common information to be decoded by both receivers. The coding scheme is sketched as follows: fix distribution $p(w), p(u|w,s)$, and $p(v|w,t)$. For each α, β and k^n, independently generate $2^{n\rho_0}$ codewords $w^n(\alpha, \beta, k^n) \sim \prod_{i=1}^n p(w_i)$. Next, for each pair of (s^n, w^n), independently generate $2^{n\rho_1}$ codewords $u^n(s^n, w^n) \sim \prod_{i=1}^n p(u_i|s_i, w_i)$, and $2^{n\rho_2}$ codewords $v^n(t^n, w^n) \sim \prod_{i=1}^n p(v_i|t_i, w_i)$. For each pair of source sequences (s^n, t^n), the encoder will choose a triple (w^n, u^n, v^n) such that $(s^n, t^n, k^n, w^n, u^n, v^n) \in T_\epsilon^n(STKWUV)$, which is ensured with high probability by properly chosen ρ_1, ρ_2 and ρ_3. The two decoders use joint typicality decoding, that is, decoder Y_1 finds a unique sequence s^n such that $(s^n, k^n, w^n, u^n, y_1^n) \in T_\epsilon^n(SKWUY_1)$. Similarly, decoder Y_2 finds a unique sequence t^n such that $(t^n, k^n, w^n, v^n, y_2^n) \in T_\epsilon^n(TKWVY_2)$.

In [27], Minero and Kim proposed an alternative, and conceptually simple, coding scheme. The obtained region does not involve the common part K of the two variables (S, T), but was shown to be equivalent to that of Han and Costa, thus yielding the intuitive explanation that the common part does not play a role for the broadcast channel case because of the centralized transmitter. Indeed, the same coding scheme proposed by Han and Costa can also be easily modified to obtain the same region without the use of the common part. For the encoding scheme in [3], sketched above, if we remove the part related to the common variable K, in both the encoding and decoding processes, straightforward error probability analysis leads to the same sufficient condition as in Proposition 3.

1.4 Wyner's common information for two random variables

Consider a pair of dependent random variables X and Y with joint distribution $P(x, y)$. Characterizing the common information between X and Y has been a topic of research interest in the past decades [56, 57, 58, 59, 60]. There have been three classical notions reported in the literature.

Shannon's [1] mutual information $I(X; Y)$

Shannon's mutual information measures how much uncertainty can be reduced with respect to one random variable by observation the other random variable. In the case that X and Y are independent, mutual information $I(X; Y) = 0$, indicating that observing one variable X does not give any information about Y and vice versa. Shannon's mutual information carries operational meanings that are instrumental in laying the foundation for information theory.

Gács and Körner's [56] common randomness $K(X, Y)$

Consider a pair of independent and identically distributed random sequences X^n, Y^n with each pair $(X_i, Y_i) \sim P(x, y)$. These two sequences are observed respectively by two nodes, which attempt to map the sequences to a common message set $\mathcal{W}$. Specifically, let f_n and g_n be such mappings, i.e.,

$$\begin{aligned} f_n : &\quad \mathcal{X}^n \to \mathcal{W}, \\ g_n : &\quad \mathcal{Y}^n \to \mathcal{W}. \end{aligned}$$

Define $\epsilon_n = Pr(W_1 \neq W_2)$ where $W_1 = f_n(X^n)$ and $W_2 = g_n(Y^n)$. Gács and Körner's common randomness is defined as

$$K(X, Y) = \lim_{n\to\infty, \epsilon_n \to 0} \sup \frac{1}{n} H(W_1).$$

Gács and Körner's common randomness has found extensive applications in cryptography, i.e., for key generation [63,61,62]. On the other hand, the common randomness notion is rather restrictive as it equals 0 in most cases except for the following special case (or random variable pairs that can be converted to such distributions through relabeling of realizations, i.e., permutation of joint distribution matrix). Let X and Y be $X = (X', V)$ and $Y = (Y', V)$, respectively, where X', Y', V are independent. Clearly, the common part between X and Y is V and it follows that $K(X; Y) = H(V)$. Note that for this example $I(X; Y) = K(X; Y) = H(V)$.

Wyner's [59] common information $C(X, Y)$

Wyner's common information is defined as

$$C(X, Y) = \min_{X \to W \to Y} I(XY; W). \tag{1.18}$$

Thus the hidden (or auxiliary) variable W induces a Markov chain $X - W - Y$, or, equivalently, a conditional independence structure of X, Y being independent given W. Wyner gave two operational meanings for the above definition. The first approach is shown in Fig. 1.2. The encoder observes a pair of sequences (X^n, Y^n), and maps them to three messages W_0, W_1, W_2, taking values in alphabets of respective sizes $2^{nR_0}, 2^{nR_1}, 2^{nR_2}$. Decoder 1, upon receiving (W_0, W_1), needs to reproduce X^n reliably while decoder 2, upon receiving (W_0, W_2), needs to reproduce Y^n reliably. Let C_1 be the infimum of all admissible R_0 for the system in Fig. 1 such that the total rate $R_0 + R_1 + R_2 \approx H(X, Y)$.

The second approach is shown in Fig. 1.3. A common input W, uniformly distributed on $\mathcal{W} = \{1, \cdots, 2^{nR_0}\}$ is given to two separate processors which are otherwise independent of each other. These processors (random variable generators) generating independent and identically distributed sequences according to $q_1(X|W)$ and $q_2(Y|W)$ respectively. The output sequences of the two processors are denoted by $\tilde{X}^n$ and $\tilde{Y}^n$ respectively. Thus the joint distribution of the output sequences is,

$$Q(\tilde{X}^n, \tilde{Y}^n) = \sum_{w \in \mathcal{W}} \frac{1}{\mathcal{W}} q_1(X^n|W) q_2(Y^n|W). \tag{1.19}$$

Define C_2 of (X, Y) to be infimum of rate R_0 for the common input such that $q(\tilde{X}^n, \tilde{Y}^n)$ close to $p(X^n, Y^n)$, where the closeness is defined using the average divergence of the two distributions

$$D_n(P, Q) = \frac{1}{n} \sum_{x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n} P(x^n, y^n) \log \frac{P(x^n, y^n)}{Q(x^n, y^n)}. \tag{1.20}$$

Wyner proved that

$$C_1 = C_2 = C(X, Y). \tag{1.21}$$

It was observed in [59] that

$$K(X, Y) \leq I(X; Y) \leq C(X, Y). \tag{1.22}$$

Wyner [59] and Witsenhausen [60] also provide several examples on how to calculate the common information $C(X, Y)$. For the example of $X = (X', V)$ and $Y = (Y', V)$ with (X', Y', V) mutually independent, $C(X, Y) = I(X; Y) = K(X, Y) = H(V)$.

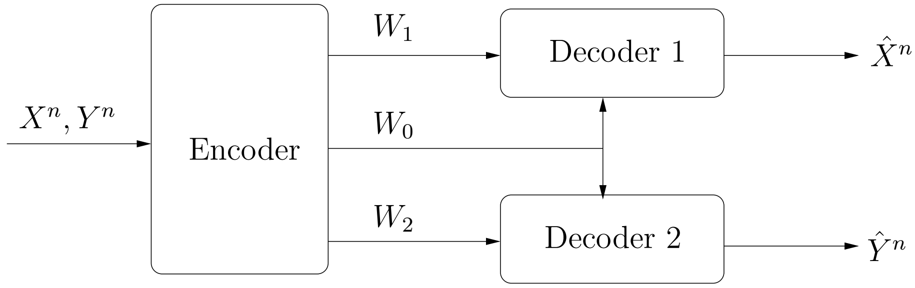

Figure 1.2: Source coding over a simple network.

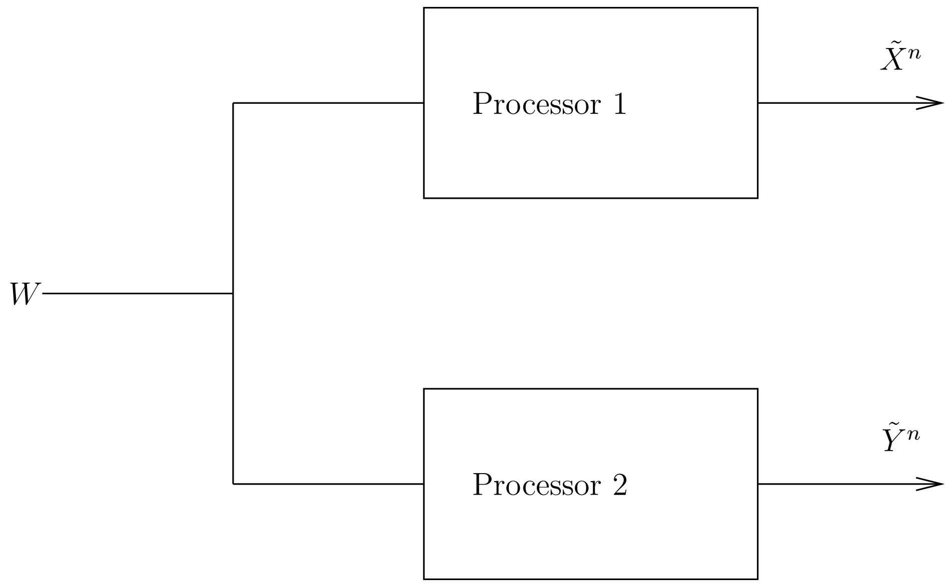

Figure 1.3: Random variable generators.

1.5 Hypothesis testing with communication constraints

Consider the classical hypothesis testing problem, $H_0 : X \sim P(X)$ against $H_1 : X \sim Q(X)$. Suppose that a length n sequence $x^n = (x_1, x_2, \cdots, x_n)$ is emitted from a source that is assumed to have a probability distribution of $\prod_{i=1}^n P(x_i)$ under H_0 and $\prod_{i=1}^n Q(x_i)$ under H_1 . By fully observing the data x^n, the statistician attempts to decide, which hypothesis of H_0 or H_1 is true. It is easy to show that any reasonable test will lead to a diminishing error probability as n grows to infinity. Thus a sensible criterion is the speed with which the error probability approaches zero, i.e., the error exponent characterization. Under the Neyman-Pearson criterion, Stein's lemma [35, Theorem 12.8.1] gives us the best error exponent of second kind of error β_n, under the condition that the first kind of error probability $\alpha_n \leq \epsilon$ for a prescribed $0 < \epsilon < 1$.

On the other hand, in some cases, such as in resource constrained sensor network, it is expensive to transmit all the sampling data to the central processor. Instead, local sensors encode the data x^n and transmit its encoded version $f(x^n)$ with rate constraint R (data compression). It is trivial to see that the introduction of such a rate constraint adds no new aspects in such a simple hypothesis testing problem, because even one bit (local sensor sends the binary decision to the central processor) is sufficient for the statistician to attain the same optimal decision as that attained with the full knowledge of x^n.

Things will change dramatically, however, if we consider the distributed bivariate hypothesis testing $H_0 : XY$ against $H_1 : \bar{X}\bar{Y}$, i.e., x^n from $X^n(\bar{X}^n)$ and y^n from $Y^n(\bar{Y}^n)$ are communicated with rate constraints R_1 and R_2, separately, at remote sites. The distributed behavior in this framework makes the problem extremely involved. Characterizing optimal error exponents for dependence test with communication constraints was first considered by Ahlswede and Csiszár [78]. In particular, for the special case of test of independence problem with one sided data compression, i.e., $R_2 = \infty$, a single letter characterization of the optimal error exponent was obtained in [78]. An overview of related work can be found in [80] and the references therein.

We now formally state the problem as follows.

Consider the bivariate hypothesis testing problem:

$$H_0 \ : \ XY \sim P(XY), \tag{1.23}$$

$$H_1 \ : \ XY \sim Q(XY). \tag{1.24}$$

We will denote $H_0 : XY$ and $H_1 : \bar{X}\bar{Y}$ for continence.

The data sequences $x^n = (x_1, x_2, \cdots, x_n)$ and $y^n = (y_1, y_2, \cdots, y_n)$ are communicated by two separate encoders:

$$\phi_1 \ : \ \mathcal{X}^n \to \mu_1 = \{1, 2, \cdots, M_1\}, \tag{1.25}$$

$$\phi_2 \ : \ \mathcal{Y}^n \to \mu_2 = \{1, 2, \cdots, M_2\}, \tag{1.26}$$

with rates R_1 and R_2 defined in the sense that, for any fixed $\eta > 0$,

$$\frac{1}{n} \log M_1 \ \leq \ R_1 + \eta, \tag{1.27}$$

$$\frac{1}{n}\log M_2 \quad \leq \quad R_2+\eta. \tag{1.28}$$

The decoder $\psi : \mu_1 \times \mu_2 \rightarrow \{H_0, H_1\}$, observes outputs $(\phi_1(x^n), \phi_2(y^n))$ from two encoders ϕ_1 and ϕ_2 and decide which of the two hypothesis H_0 and H_1 is true. That is, the decoder declares that H_0 is true if $\psi = H_0$, and that H_1 is true if $\psi = H_1$.

The acceptance region $\mathcal{A}_n$ is defined as

$$\mathcal{A}_n = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \psi(\phi_1(x^n), \phi_2(y^n)) = H_0\} \tag{1.29}$$

The first kind of error is defined as

$$\alpha_n = Pr(X^nY^n \in \mathcal{A}_n^c), \tag{1.30}$$

and the probability of second kind of error is defined as

$$\beta_n = Pr(\bar{X}^n\bar{Y}^n \in \mathcal{A}_n). \tag{1.31}$$

With these notations, the hypothesis testing problem is formulated as follows. For any fixed $0 < \epsilon < 1$, impose the condition $\alpha_n \leq \epsilon$, and define

$$\beta_n(R_1, R_2, \eta, \epsilon) = \min_{\phi_1, \phi_2, \psi} \beta_n \tag{1.32}$$

Further, define

$$\theta(R_1, R_2, \eta, \epsilon) \quad = \quad \liminf_{n\rightarrow\infty}(\frac{1}{n}\log \beta_n(R_1, R_2, \eta, \epsilon)) \tag{1.33}$$

$$\theta(R_1, R_2, \epsilon) \quad = \quad \lim_{\eta\rightarrow 0}\theta(R_1, R_2, \eta, \epsilon) \tag{1.34}$$

where, $\theta(R_1, R_2, \epsilon)$ is the error exponent for the hypothesis testing of H_0 against H_1. The goal is to characterize $\theta(R_1, R_2, \epsilon)$ for different R_1 and R_2.

One important special case is when y^n is fully available at the decoder side, i.e., $R_2 \geq \log|\mathcal{Y}|$. In this case, we set

$$\theta(R, \epsilon) = \theta(R_1, \log|\mathcal{Y}|, \epsilon) \tag{1.35}$$

For this general two sided hypothesis testing problem, we have the following proposition due to Han [79].

Proposition 4 *([79, Theorem 6]) Define two sets of finite-value taking auxiliary random variables:*

$$L_1(R_1, R_2) \quad = \quad \{UV : I(U;X) \leq R_1, I(V;Y) \leq R_2 \quad \textit{and} \quad U \rightarrow X \rightarrow Y \rightarrow V\}$$

$$(1.36)$$

$$L_2(UV) = \{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} : p(\tilde{U}\tilde{X}) = p(UX), p(\tilde{V}\tilde{Y}) = p(VY), p(\tilde{U}\tilde{V}) = p(UV)\} \tag{1.37}$$

and let $\bar{U}, \bar{V}$ be the random variables such that $p(\bar{U}|\bar{X}) = p(U|X)$ and $p(\bar{V}|\bar{Y}) = p(V|Y)$. Further more, define

$$\theta_L(R_1, R_2) = \sup_{UV \in L_1(R_1,R_2)} \min_{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} \in L_2(UV)} D(\tilde{U}\tilde{X}\tilde{Y}\tilde{V} || \bar{U}\bar{X}\bar{Y}\bar{V}) \tag{1.38}$$

Then we have the following result. Let $\theta(R_1, R_2, \epsilon)$ be the error exponent for the hypothesis testing of H_0 against H_1. Then, for any $R_1 \geq 0$, $R_2 \geq 0$ and $0 < \epsilon < 1$.

$$\theta(R_1, R_2, \epsilon) \geq \theta_L(R_1, R_2). \tag{1.39}$$

In the case of test of independent, we have the following corollary.

Corollary 1 *([79, Corollary 6]) Consider the case where $p(\bar{X}) = p(X), p(\bar{Y}) = p(Y)$, and $\bar{X}, \bar{Y}$ are independent. Then, for $R_1 \geq 0, R_2 \geq 0$ and $0 < \epsilon < 1$,*

$$\theta(R_1, R_2, \epsilon) \geq \max_{U \in L_1(R_1,R_2)} I(U;V) \tag{1.40}$$

1.6 Outline of the thesis

This thesis attempts to make progress toward a better understanding of various issues related to correlated sources in distributed systems, including: how to exploit correlation structure to facilitate transmission of correlated sources over interference channels, how to characterize the common information of random variables that carries meaningful operational interpretation, and how to detect the presence of data dependence in a distributed network.

This thesis intends to make progress to have a better understanding of communicating correlated sources over multi-terminal channels, as well as to have a better understanding of the dependence structure in correlated sources.

In Chapter 2, communicating arbitrarily correlated sources over interference channels is considered. A sufficient condition is found for the lossless transmission of a pair of correlated sources over a discrete memoryless interference channel. With independent sources, the sufficient condition reduces to the Han-Kobayashi achievable rate region for the interference channel. For a special correlation structure (in the sense of Slepian-Wolf, 1973), the proposed region reduces to the known achievable region for interference channels with common information. A simple example is given to show that the separation approach, with Slepian-Wolf encoding followed by optimal channel coding, is strictly suboptimal.

In Chapter 3, we consider the transmission of two correlated Gaussian sources over a two-user Gaussian ZIC. To facilitate our study, we first investigate Gaussian multiple access channels with correlated Gaussian sources and single distortion constraint at the receiver. Lower bounds on the distortion as well as several achievable schemes are proposed. In particular, hybrid digital analog transmission is proposed whose achievable region is resolved through coupling it with the quadratic Gaussian CEO problem. These lower bounds and achievable schemes are then applied to the source channel communication over Gaussian ZIC.

In Chapter 4, we consider lossy transmission of two correlated sources over a two-user discrete memoryless interference channel (DMIC). An achievable distortion region is obtained and it is shown that it includes Salehi and Kurtas's result [28] on lossless transmission of correlated sources over a DMIC as a special case. Sending correlated Gaussian sources over a Gaussian interference channel is also studied, and several achievable schemes as well as lower bounds are proposed.

Chapter 5 generalizes Wyner's definition of common information of a pair of random variables to that of N random variables. We prove coding theorems that show the same operational meanings for the common information of two random variables generalize to that of N random variables. As a byproduct of our proof, we show that the Gray-Wyner source coding network can be generalized to N source squences with N decoders. We also establish a monotone property of Wyner's common information which is in contrast to other notions of the common information, specifically Shannon's mutual information and Gács and Körner's common randomness. Examples about the computation of Wyner's common information of N random variables are also given.

In Chapter 6, Wyner's common information is generalized for continuous random variables. We provide an operational meaning for such generalization using the Gray-Wyner network with *lossy* source coding. Specifically, a Gray-Wyner network consists of one encoder and two decoders. A sequence of independent copies of a pair of random variables $(X, Y) \sim p(x, y)$ is encoded into three messages, one of them is a common input to both decoders. The two decoders attempt to reconstruct the two sequences respectively subject to individual distortion constraints. We show that Wyner's common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds. An interpretation for such thresholds is given for the symmetric case.

In Chapter 7, we consider the problem of *distributed* test of statistical independence under communication constraints. While independence test is frequently encountered in various applications, *distributed* independence test is particularly useful for events detection in sensor networks: data correlation often occurs among sensor observations in the presence of a target. Focusing on the Gaussian case because of its tractability, we study in this chapter the characteristics of optimal scalar quantizers for distributed test of independence where the optimality is in the sense of optimizing the error exponent. We also discuss the optimal quantizer properties for the finite sample regime, i.e., that of directly minimizing the error probability.

We conclude the thesis in Chapter 8 by summarizing the major contributions as well as possible directions for future work.

Chapter 2

Interference Channels With Arbitrarily Correlated Sources: The Lossless Case

Communicating correlated sources over interference channels has previously been studied by Salehi and Kurtas [28]. However, the obtained rate region, derived by largely following the coding scheme for MAC [2,29] does not reduce to the well known Han and Kobayashi (HK) region for interference channels [12] when the sources are independent. The HK region, originally proposed in 1981 [12] and recently simplified by Chong *et al* [18], remains to be the largest achievable rate region for interference channels with independent messages. In addition, there is no definitive answer to the question of whether the separation approach is strictly suboptimal, even though intuition suggests that this is likely the case.

In this chapter, we derive a sufficient condition for lossless transmission of a pair of arbitrarily correlated sources over a discrete memoryless interference channel (DMIC). The coding scheme takes advantage of the common part of the random source pair, if it exists. Moreover, it utilizes the correlation preserving technique for the multiple access channel [2] and the random source partition for the broadcast channel [3]. We show that the proposed region includes the HK region as its special case. In addition, for a special correlation structure (in the sense of Slepian-Wolf, 1973 [19]), the proposed region coincides with the known achievable region for interference channels with common information [30,31,32]. Finally, the proposed coding scheme is shown to be optimal for transmitting such set of correlated sources over a class of deterministic interference channels studied by El Gamal and Costa [14].

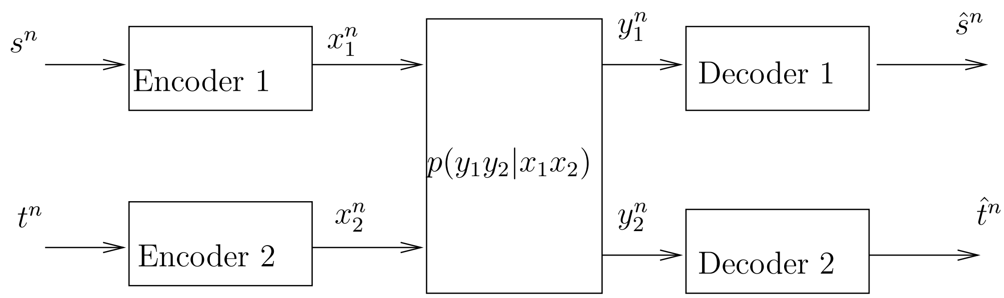

Figure 2.1: Interference channels with correlated sources.

The rest of this chapter is organized as follows. Section 2.1 gives the problem formulation and introduces some previous results related to this work. The main result as well as its implications are presented in Section 2.2. Section 2.3 concludes this chapter.

2.1 Problem statement

The model studied in this chapter is shown in Fig. 2.1. The source sequences (S^n, T^n) are arbitrarily correlated discrete memoryless sources, generated independently according to:

$$p(s^n, t^n) = \prod_{i=1}^{n} p(s_i, t_i). \tag{2.1}$$

This pair of source sequences S^n and T^n are to be transmitted losslessly over a two user DMIC defined by the transition probability $p(y_1y_2|x_1x_2)$, where X_1, X_2 are the channel inputs and Y_1, Y_2 are the channel outputs.

A length n source channel block code for the channel consists of two encoder mappings:

$$\psi_1 \; : \; \mathcal{S}^n \to \mathcal{X}_1^n, \tag{2.2}$$

$$\psi_2 \; : \; \mathcal{T}^n \to \mathcal{X}_2^n, \tag{2.3}$$

and two decoder mappings:

$$\phi_1 \; : \; \mathcal{Y}_1^n \to \mathcal{S}^n, \tag{2.4}$$

$$\phi_2 \; : \; \mathcal{Y}_2^n \to \mathcal{T}^n. \tag{2.5}$$

The probability of error at decoders 1 and 2 are defined as

$$P_{e1} = \sum_{s^n \in \mathcal{S}^n} p(s^n) Pr\{s^n \neq \phi_1(y_1^n) | S^n = s^n\}, \tag{2.6}$$

$$P_{e2} = \sum_{t^n \in \mathcal{T}^n} p(t^n) Pr\{t^n \neq \phi_2(y_2^n) | T^n = t^n\}. \tag{2.7}$$

Definition 1 *The source* $(S, T) \sim \prod_{i=1}^n p(s_i, t_i)$ *is said to be* admissible *for the interference channel* $p(y_1 y_2 | x_1 x_2)$ *if for any* $\epsilon \in (0, 1)$ *and sufficiently large* n*, there exist a sequence of block codes* $(\psi_1, \psi_2, \phi_1, \phi_2)$ *such that*

$$\max\{P_{e1}, P_{e2}\} \leq \epsilon. \tag{2.8}$$

The goal is to find a sufficient condition for a source pair (S, T) to be admissible for a given DMIC.

2.2 A sufficient condition

We begin with a quick review of the HK achievable rate region for interference channels with independent messages. The major ingredients in the coding scheme for the HK region are rate splitting and joint decoding. Specifically, user i, $i = 1, 2$, splits the message M_i into two parts, common message M_{i0} and private message M_{i1}. Therefore, $|\mathcal{M}_i| = |\mathcal{M}_{i0}| \times |\mathcal{M}_{i1}|$ where $|\cdot|$ denotes the cardinality of a set. The common message needs to be decoded by both decoders and the private message is only intended for its own receiver. This rate splitting can be implemented using sequential superposition encoding as described in [18]. Let $R_{ij} = \frac{1}{n} \log |\mathcal{M}_{ij}|$, $i = 1, 2$ and $j = 0, 1$. First generate $2^{nR_{i0}}$ auxiliary codewords w_i^n, which carry the information of common message M_{i0}. Next, for each W_i^n, generate $2^{nR_{i1}}$ codewords x_i^n superimposed on top of w_i^n, which carry the information of the private message M_{i1}. Each decoder jointly decodes both common messages and its own private message, i.e., decoder 1 finds unique codewords w_1^n, w_2^n and x_1^n such that $(w_1^n, w_2^n, x_1^n, y_1^n) \in T_\epsilon^n(W_1 W_2 X_1 Y_1)$, and decoder 2 finds unique codewords w_1^n, w_2^n and x_2^n such that $(w_1^n, w_2^n, x_2^n, y_2^n) \in T_\epsilon^n(W_1 W_2 X_2 Y_2)$.

Consider now the model of interest in the present chapter, i.e., DMIC with correlated sources. Let us first disregard the common part K between the source variables S and T. We start with Han and Costa's random source partition: the sequences $s^n \in \mathcal{S}^n$ and $t^n \in \mathcal{T}^n$ are randomly placed respectively into 2^{nr_1} and 2^{nr_2} cells. This source partition is tantamount to rate splitting in the channel coding problem: the

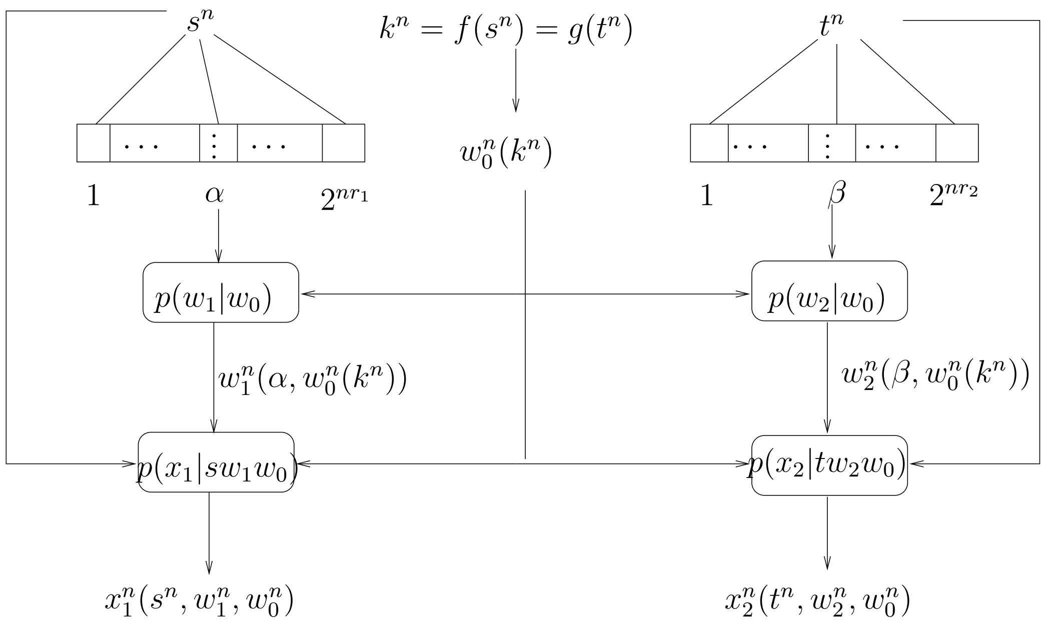

Figure 2.2: Coding structure for IC with correlated sources

cell index associated with a given sequence plays the role of common information and the index of the source within the cell the private information. This is then followed by superposition coding [18]. First generate an auxiliary codeword w_i^n for each cell index. The codeword x_i^n is then generated to be superimposed on top of w_i^n that also carries the source index within the cell. Different from [18] is that the codeword x_i^n is statistically dependent on the input source, thereby preserving the correlation contained in the original source pair. The common part K, if it exists, is then put back in the encoding process by generating an auxiliary codeword w_0^n. This codeword, known to both encoders, will be used in generating all the other codewords through a superposition code structure. This encoding process is illustrated in Fig. 2.2.

For decoders, joint typicality decoding is used at both decoders. That is, decoder 1 finds a unique s^n such that $(s^n, k^n, w_0^n, w_1^n, x_1^n, w_2^n, y_1^n) \in T_n^\epsilon(SKW_0W_1X_1W_2Y_1)$, and decoder 2 finds a unique t^n such that $(t^n, k^n, w_0^n, w_2^n, x_2^n, w_1^n, y_2^n) \in T_n^\epsilon(TKW_0W_2X_2W_1Y_2)$. The above coding scheme leads to the following sufficient condition for lossless transmission of a correlated source pair over a DMIC.

Theorem 1 *A source pair $(S,T) \sim p(s,t)$ is admissible for a discrete memoryless interference channel $p(y_1y_2|x_1x_2)$ if there exist auxiliary random variables W_0, W_1, W_2*

with joint distribution of all the variables factoring as:

$$p(s,t,w_0,w_1,w_2,x_1,x_2) = p(s,t)p(w_0)p(w_1|w_0)p(w_2|w_0)p(x_1|sw_1w_0)p(x_2|tw_2w_0), \tag{2.9}$$

and that the following conditions are satisfied:

$$\begin{aligned}
H(S|K) &< I(SX_1;Y_1|W_0W_2K), && (2.10)\\
H(T|K) &< I(TX_2;Y_2|W_0W_1K), && (2.11)\\
H(S) &< I(W_0W_2SX_1;Y_1), && (2.12)\\
H(T) &< I(W_0W_1TX_2;Y_2), && (2.13)\\
H(S|K)+H(T|K) &< I(SX_1;Y_1|W_0W_1W_2K)+I(W_1TX_2;Y_2|W_0K), && (2.14)\\
H(S|K)+H(T|K) &< I(TX_2;Y_2|W_0W_1W_2K)+I(W_2SX_1;Y_1|W_0K), && (2.15)\\
H(S|K)+H(T|K) &< I(SW_2X_1;Y_1|W_0W_1K)+I(TW_1X_2;Y_2|W_0W_2K), && (2.16)\\
H(S|K)+H(T) &< I(W_0W_1TX_2;Y_2)+I(SX_1;Y_1|W_0W_1W_2K), && (2.17)\\
H(S)+H(T|K) &< I(W_0W_2SX_1;Y_1)+I(TX_2;Y_2|W_0W_1W_2K), && (2.18)\\
2H(S|K)+H(T|K) &< I(SX_1;Y_1|W_0W_1W_2K)+I(SW_2X_1;Y_1|W_0K)\\
&\quad +I(TW_1X_2;Y_2|W_0W_2K), && (2.19)\\
H(S|K)+2H(T|K) &< I(TX_2;Y_2|W_0W_1W_2K)+I(TW_1X_2;Y_2|W_0K)\\
&\quad +I(SW_2X_1;Y_1|W_0W_1K), && (2.20)\\
H(S)+H(S|K)+H(T|K) &< I(SX_1;Y_1|W_0W_1W_2K)+I(W_0W_2SX_1;Y_1)\\
&\quad +I(TW_1X_2;Y_2|W_0W_2K), && (2.21)\\
H(T)+H(S|K)+H(T|K) &< I(TX_2;Y_2|W_0W_1W_2K)+I(W_0W_1TX_2;Y_2)\\
&\quad +I(SW_2X_1;Y_1|W_0W_1K), && (2.22)
\end{aligned}$$

where $K = f(S) = g(T)$ is the common part of S and T in the sense of Gács, Körner [23] and Witsenhausen [58].

The proof of Theorem 1 is given in Appendix A.

Remark 1: Separate source and channel coding is known to be *strictly suboptimal* for transmitting correlated sources over multiple access channels [2] and broadcast channels [3]. The same statement can be made for transmitting correlated sources over interference channels. Consider the special case of transmitting the triangular source (S, T) as in [2] with the joint distribution $p(s, t)$ given by

$$p(s = 0, t = 0) = p(s = 1, t = 0) = p(s = 1, t = 1) = \frac{1}{3}, \tag{2.23}$$

over an interference channel defined by $Y_1 = X_1$ and $Y_2 = X_1 + X_2$, where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \{0, 1\}$ and $\mathcal{Y}_2 = \{0, 1, 2\}$. Notice that this is a special case of one sided deterministic interference channels studied in [14], where the sum rate capacity is given by

$$R_1 + R_2 \leq \max_{p(x_1)p(x_2)} H(Y_2) = 1.5. \tag{2.24}$$

For the source pair, $H(S, T) = \log_2 3 = 1.58$ bits, therefore, lossless transmission is not attainable by a simple concatenation of the Slepian-Wolf code followed by an optimal channel code. However, it can be easily checked that a trivial way to reliably transmit this source is to choose $X_1 = S$ and $X_2 = T$, which results in zero error probability at both receivers. This example shows that separate source and channel coding is strictly suboptimal. In fact, the source and channel in this example are perfectly matched when choosing $X_1 = S$ and $X_2 = T$ in the sense that $H(S|Y_1) = 0$ and $H(S, T|Y_2) = 0$. More specifically, given $Y_1 = y_1$ there is no uncertainty of S and given $Y_2 = y_2$, the only ambiguity of T when $y_2 = 1$ is removed by considering the structure of the sources. We comment here that one can easily check that this special case is included in Theorem 1 by letting $K = W_0 = W_1 = W_2 = \phi$, $X_1 = S$, and $X_2 = T$.

The detailed proof of Theorem 1 is given in Section A. We now discuss some implications of Theorem 1.

2.2.1 Reduce to Han and Kobayashi's region on interference channels

Corollary 2 *If there is no common part for the source pair (S, T), i.e., $K = \emptyset$, let $W_0 = Q$ be the time sharing variable. Theorem 1 yields the following sufficient condition for lossless transmission of (S, T) over a discrete memoryless interference channel.*

$$H(S) < I(SX_1; Y_1|QW_2), \tag{2.25}$$
$$H(T) < I(TX_2; Y_2|QW_1), \tag{2.26}$$

$$
\begin{aligned}
H(S)+H(T) &< I(SX_1;Y_1|QW_1W_2)+I(W_1TX_2;Y_2|Q), && (2.27)\\
H(S)+H(T) &< I(TX_2;Y_2|QW_1W_2)+I(W_2SX_1;Y_1|Q), && (2.28)\\
H(S)+H(T) &< I(SW_2X_1;Y_1|QW_1)+I(TW_1X_2;Y_2|QW_2), && (2.29)\\
2H(S)+H(T) &< I(SX_1;Y_1|QW_1W_2)+I(SW_2X_1;Y_1|Q) && \\
&\quad +I(TW_1X_2;Y_2|QW_2), && (2.30)\\
H(S)+2H(T) &< I(TX_2;Y_2|QW_1W_2)+I(TW_1X_2;Y_2|Q) && \\
&\quad +I(SW_2X_1;Y_1|QW_1), && (2.31)
\end{aligned}
$$

where W_1, W_2 are auxiliary random variables such that the joint distribution of all variables can be factored as

$$p(s,t,q,w_1,w_2,x_1,x_2) = p(s,t)p(q)p(w_1|q)p(w_2|q)p(x_1|sw_1q)p(x_2|tw_2q). \tag{2.32}$$

The fact that Theorem 1 includes the HK region as its special case comes directly from Corollary 2. If S and T are independent, choose the joint distribution as

$$p(s,t,q,w_1,w_2,x_1,x_2) = p(s)p(t)p(q)p(w_1|q)p(w_2|q)p(x_1|w_1q)p(x_2|w_2q), \tag{2.33}$$

and let $R_1 = H(S)$ and $R_2 = H(T)$. Corollary 2 yields an achievable region for the interference channel which coincides with that described in Proposition 1.

2.2.2 Reduce to Jiang *et al* and Cao *et al*'s region on interference channels with common information

Consider now another special case where the source has a special correlation structure similar to that of [19].

Corollary 3 *Suppose that the source (S, T) can be decomposed into three parts: $S = (S', K)$ and $T = (T', K)$ where S', T', K are independent random variables. Choose the joint distribution*

$$p(s,t,w_0,w_1,w_2,x_1,x_2) = p(s')p(t')p(k)p(w_0)p(w_1|w_0)p(w_2|w_0)p(x_1|w_0w_1)p(x_2|w_0w_2), \tag{2.34}$$

where $s = (s', k)$ and $t = (t', k)$. Theorem 1 gives the following sufficient condition for lossless transmission of the source pair (S, T).

$$H(S') < I(X_1;Y_1|W_0W_2), \tag{2.35}$$

$$\begin{aligned}
H(T') &\leq I(X_2;Y_2|W_0W_1), && (2.36)\\
H(K)+H(S') &< I(W_0W_2X_1;Y_1), && (2.37)\\
H(K)+H(T') &< I(W_0W_1X_2;Y_2), && (2.38)\\
H(S')+H(T') &< I(X_1;Y_1|W_0W_1W_2)+I(W_1X_2;Y_2|W_0), && (2.39)\\
H(S')+H(T') &< I(X_2;Y_2|W_0W_1W_2)+I(W_2X_1;Y_1|W_0), && (2.40)\\
H(S')+H(T') &< I(W_2X_1;Y_1|W_0W_1)+I(W_1X_2;Y_2|W_0W_2), && (2.41)\\
H(K)+H(S')+H(T') &< I(W_0W_1X_2;Y_2)+I(X_1;Y_1|W_0W_1W_2), && (2.42)\\
H(K)+H(S')+H(T') &< I(W_0W_2X_1;Y_1)+I(X_2;Y_2|W_0W_1W_2), && (2.43)\\
2H(S')+H(T') &< I(X_1;Y_1|W_0W_1W_2)+I(W_2X_1;Y_1|W_0)\\
&\quad +I(W_1X_2;Y_2|W_0W_2), && (2.44)\\
H(S')+2H(T') &< I(X_2;Y_2|W_0W_1W_2)+I(W_1X_2;Y_2|W_0)\\
&\quad +I(W_2X_1;Y_1|W_0W_1), && (2.45)\\
H(K)+2H(S')+H(T') &< I(X_1;Y_1|W_0W_1W_2)+I(W_0W_2X_1;Y_1)\\
&\quad +I(W_1X_2;Y_2|W_0W_2), && (2.46)\\
H(K)+H(S')+2H(T') &< I(X_2;Y_2|W_0W_1W_2)+I(W_0W_1X_2;Y_2)\\
&\quad +I(W_2X_1;Y_1|W_0W_1). && (2.47)
\end{aligned}$$

Corollary 3 can be used to establish that the sufficient condition includes that of [30, 31, 32] as its special case. Specifically, define $R_0 = H(K), R_1 = H(S')$ and $R_2 = H(T')$, the sufficient condition reduces to the rate region of interference channels with common information obtained in [30,31,32]. Moreover, Corollary 3 can be used to establish a necessary and sufficient condition for transmitting the above class of correlated sources over a class of deterministic interference channels.

2.2.3 A necessary and sufficient condition for a class of deterministic interference channels

Consider the class of deterministic interference channel defined in [14] (see figure 2.3). The definitions of length n block code $(\psi_1, \psi_2, \phi_1, \phi_2)$ and admissible sources remain the same as in (2.2)-(6). The difference lies in the channel model, which is given by the following deterministic functions:

$$\begin{aligned}
V_1 &= g_1(X_1), && (2.48)\\
V_2 &= g_2(X_2), && (2.49)
\end{aligned}$$

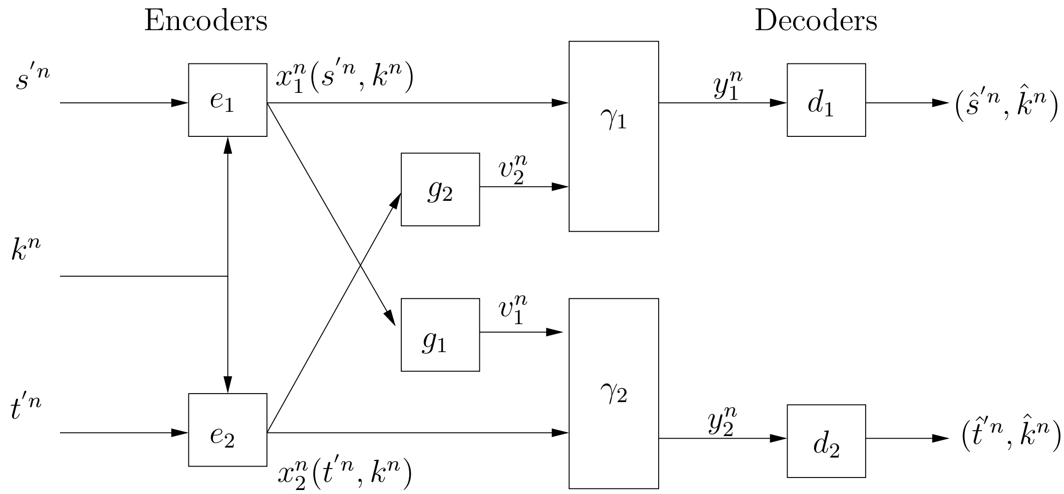

Figure 2.3: A class of deterministic interference channel with correlated sources.

$$
\begin{aligned}
Y_1 &= \gamma_1(X_1, V_2), && (2.50)\\
Y_2 &= \gamma_2(X_2, V_1), && (2.51)
\end{aligned}
$$

where V_1 and V_2 represent the interference signals caused by transmitter 1 and 2, and $g_1, g_2, \gamma_1, \gamma_2$ are deterministic functions. Furthermore, we assume that there exist two deterministic functions,

$$
\begin{aligned}
V_2 &= h_1(Y_1, X_1), && (2.52)\\
V_1 &= h_2(Y_2, X_2). && (2.53)
\end{aligned}
$$

As pointed out in [32], conditions (2.52) and (2.53) are more general than conditions $H(Y_1|X_1) = H(V_2)$ and $H(Y_2|X_2) = H(V_1)$ as used in [14]. We denote this class of deterministic interference channel as $\mathcal{C}_d$. Let $\mathcal{P}_d$ denote the set of joint distributions $p(\cdot)$ that factor as

$$p(q, x_1, x_2) = p(q)p(x_1|q)(x_2|q), \tag{2.54}$$

where q is the realization of an auxiliary random variable Q defined over a finite set $\mathcal{Q}$. We have the following theorem.

Theorem 2 *The necessary and sufficient condition for transmitting the correlated sources $S = (S', K)$ and $T = (T', K)$ with S', T', K being independent through the channel $\mathcal{C}_d$ is the following,*

$$H(S') \leq H(Y_1|V_2Q), \tag{2.55}$$

$$
\begin{aligned}
H(T^{'}) &\leq H(Y_2|V_1Q), & (2.56)\\
H(K) + H(S^{'}) &\leq H(Y_1), & (2.57)\\
H(K) + H(T^{'}) &\leq H(Y_2), & (2.58)\\
H(S^{'}) + H(T^{'}) &\leq H(Y_1|V_1Q) + H(Y_2|V_2Q), & (2.59)\\
H(S^{'}) + H(T^{'}) &\leq H(Y_1|Q) + H(Y_2|V_1V_2Q), & (2.60)\\
H(S^{'}) + H(T^{'}) &\leq H(Y_2|Q) + H(Y_1|V_1V_2Q), & (2.61)\\
H(K) + H(S^{'}) + H(T^{'}) &\leq H(Y_1) + H(Y_2|V_1V_2Q), & (2.62)\\
H(K) + H(S^{'}) + H(T^{'}) &\leq H(Y_2) + H(Y_1|V_1V_2Q), & (2.63)\\
2H(S^{'}) + H(T^{'}) &\leq H(Y_1|Q) + H(Y_1|V_1V_2Q) + H(Y_2|V_2Q), & (2.64)\\
2H(T^{'}) + H(S^{'}) &\leq H(Y_2|Q) + H(Y_2|V_1V_2Q) + H(Y_1|V_1Q), & (2.65)\\
H(K) + 2H(S^{'}) + H(T^{'}) &\leq H(Y_1) + H(Y_1|V_1V_2Q) + H(Y_2|V_2Q), & (2.66)\\
H(K) + 2H(S^{'}) + H(T^{'}) &\leq H(Y_2) + H(Y_2|V_1V_2Q) + H(Y_1|V_1Q) & (2.67)
\end{aligned}
$$

for some fixed joint distribution $p(\cdot) \in \mathcal{P}_d$.

The proof of Theorem 2 is given in Appendix B.

Remark 3: Theorem 2 includes [32, Theorem 4] as its special case. Specifically, define $R_0 = H(K), R_1 = H(S')$ and $R_2 = H(T')$, the necessary and sufficient condition reduces to the capacity region of a class of deterministic interference channels with common information obtained in [32, Theorem 4].

2.3 Conclusion

In this chapter, we studied the problem of communicating arbitrarily correlated sources over a discrete memoryless interference channel. Using the techniques of correlation preserving coding and random source partition, a sufficient condition was derived for lossless transmission of correlated sources over interference channels. The proposed region includes the Han and Kobayashi achievable rate region for interference channels with independent messages as its special case. Furthermore, it includes the known rate region for interference channels with common information as its special case when the source correlation is in the sense of [19]. Finally, the proposed coding scheme is shown to be optimal for transmitting the set of correlated sources as in [19] over a class of deterministic interference channels [14].

Chapter 3

Communicating Correlated Gaussian Sources over Gaussian Z Interference Channels

In this chapter, we consider primarily communicating a bivariate Gaussian source over a two-user Gaussian interference channel. Each receiver is interested in one of the two Gaussian sources and we study the admissible distortion region of this source channel communication problem. We focus on the simple case of a Gaussian Z interference channel (ZIC) where interference is only one-sided. As illustrated in Fig. 3.1, Two source sequences, S_1^n and S_2^n, are available respectively at the two transmitters, and each receiver is interested in recovering the source at its own transmitter. The source sequences are assumed to be independent and identically distributed (i.i.d.) zero mean bivariate Gaussian random variables with covariance matrix

$$K_{S_1S_2} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{3.1}$$

where $\rho \in [0, 1]$, and $0 < \sigma_i < \infty, i = 1, 2$. Notice that restricting ρ to be non-negative does not lose any generality as the transmitter can always multiply the source by -1 if $\rho < 0$.

The results for the source channel communication over a ZIC in this chapter are built on detailed studies of a simpler model: communicating bivariate Gaussian sources over a Gaussian multiple access channel (GMAC) with a single distortion constraint at the receiver. The model is illustrated in Fig. 3.2 where *the receiver is only interested in recovering S_1^n from the received sequence Y^n with a distortion*

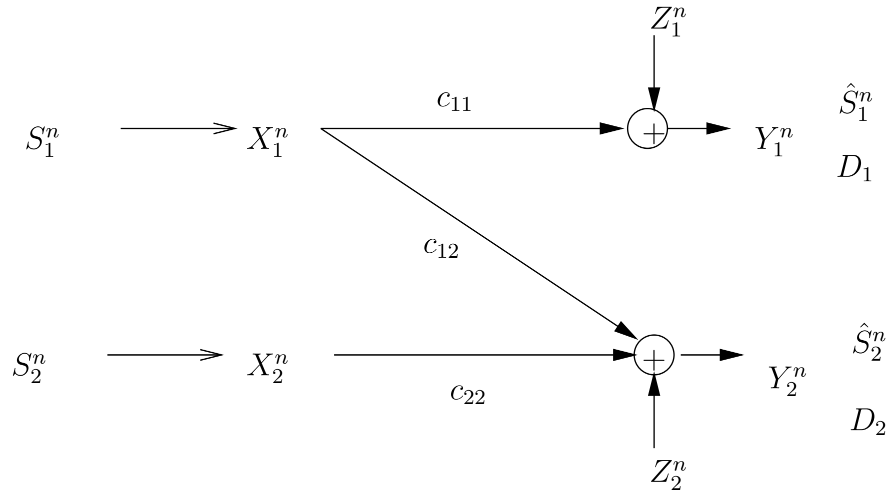

Figure 3.1: Gaussian ZIC with correlated Gaussian sources.

constraint D. Clearly, transmitter 2, in encoding S_2^n, should facilitate the estimate of S_1^n at the receiver. For example, if $\rho = 0$, it can be easily shown that the optimal scheme is to keep transmitter 2 silent while transmitter 1 uses optimal point-to-point source channel coding. In this case, transmitter 2 plays the role of "unwanted interference" and transmitter 1 can use either analog (since source and channel match) or digital (separation) transmission to achieve the optimum distortion. On the other hand, if $\rho = 1$, it can be shown that the optimal transmission is that both transmitters use uncoded transmission, i.e., amplify their source sequences with full power. For ρ between 0 and 1, the problem is interesting and remains largely open. This particular model can be viewed as a lossy extension of the model studied by Ahlswede and Han [29], where they studied the problem of lossless transmission of correlated source through a multiple access channel with a single distortion constraint.

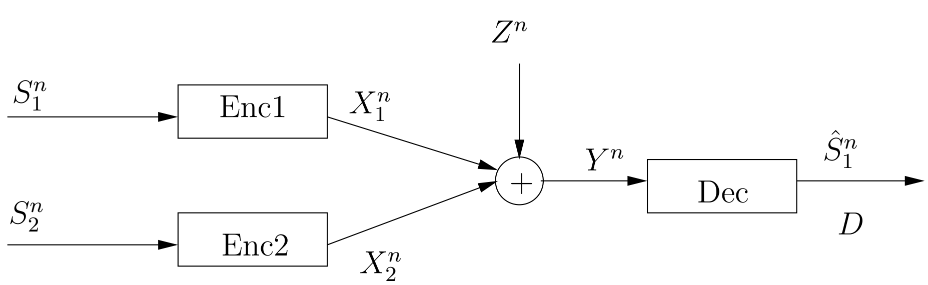

Figure 3.2: GMAC with a single distortion constraint.

Sending correlated source over a multiple access channel was first studied in [2], which lossless transmission of an i.i.d. correlated discrete memoryless source over a discrete memoryless multiple access channel. The lossy extension of this model was studied in [28], where an inner bound for the set of achievable distortions is given. Sending a correlated Gaussian source over a GMAC was recently studied in [33, 4], where the authors proposed several necessary conditions and sufficient conditions and showed that below certain signal to noise ratio (SNR) threshold, uncoded transmission is optimal for the symmetric case. Lossless transmission of i.i.d. correlated sources over some multi-user channels was also considered in [34], in which receivers have access to correlated source side information and various separation results were obtained.

Our GMAC channel with a single distortion constraint can be viewed as a special case of [33]. However, directly extracting the result out of [33,4] is not sufficient as the model in [33,4] is more general and does not take into account the special structure of having only the single constraint. For example, the lower bound obtained by removing the distortion constraint on source 2 is quite loose in general. More importantly, the optimality of uncoded transmission for the GMAC under certain SNR regime does not directly apply to our model because of the difference in the distortion constraint.

In this chapter, for communicating correlated Gaussian sources over a GMAC with a single distortion constraint, we compare three different transmission schemes, namely uncoded, digital (i.e., separation), and hybrid digital analog (HDA). For the last case, the achievable distortion is obtained via coupling the problem with the quadratic Gaussian CEO problem. The obtained achievable distortions are compared against lower bounds for the distortion. The results naturally extend to the source channel communication over a ZIC and we derive inner and outer bounds for the distortion region.

The rest of the chapter is organized as follows. In section 3.1, we consider the problem of sending a bivariate Gaussian source over a GMAC channel with a single distortion constraint. In section 3.2, we extend our model to Gaussian ZIC channels. We conclude in section 3.3, where we give a sufficient condition for sending a correlated source over a general Z interference channel.

3.1 Gaussian multiple access channels with single distortion constraint

3.1.1 System model

Consider a length-n sequence of i.i.d. zero mean bivariate Gaussian random source (S_{1j}, S_{2j}), $j = 1, \cdots, n$, with covariance matrix (3.1), with S_i^n available at transmitter i, $i = 1, 2$, as illustrated in Fig. 3.2. The codeword length is also n, hence, there is no bandwidth mismatch in our model. The encoders are a set of mappings $f_i : S_i^n \to X_i^n(S_i^n) = f(S_i^n)$, $i = 1, 2$, and are subject to individual power constraints:

$$\frac{1}{n}\sum_{j=1}^{n} EX_{ij}^2(S_i^n) \leq P_i, \quad j = 1, 2 \tag{3.2}$$

where $X_i^n = [X_{i1}, \cdots, X_{in}]$ is the codeword for the ith transmitter, and E denotes the expectation. The channel output of the GMAC is

$$Y^n = X_1^n + X_2^n + Z^n \tag{3.3}$$

where $Z^n \sim \mathcal{N}(0, NI_n)$ is a Gaussian noise sequence and I_n is the $n \times n$ identity matrix. The decoder map is $\phi(Y^n) = \hat{S}_1^n$ which is an estimate of the sequence S_1^n. As usual, we use squared distortion hence the average distortion is the mean squared error (MSE),

$$D_{1n} \triangleq \frac{1}{n}\sum_{j=1}^{n} E(S_{1j} - \hat{S}_{1j})^2.$$

Definition 2 *A distortion D is said to be achievable if there exists a sequence of (X_1^n, X_2^n, n) codes and a decoding map $\phi(Y^n)$ such that ,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} E(S_{1j} - \hat{S}_{1j})^2 \leq D. \tag{3.4}$$

We are interested in finding the minimum achievable D.

3.1.2 Lower and upper bound on the distortion constraint

We first give a lower bound on the distortion.

Theorem 3 *Any achievable distortion D should satisfy*

$$D \geq \max\{D_{l1}, D_{l2}\} \tag{3.5}$$

where

$$D_{l1} = \sigma^2 \frac{N}{P_1 + P_2 + 2\rho\sqrt{P_1 P_2} + N} \tag{3.6}$$

$$D_{l2} = \sigma^2 \frac{(1-\rho^2)N}{P_1 + N} \tag{3.7}$$

The bound D_{l1} is the cut-set bound and can be proved using Witsenhausen's Lemma [58]. It can also be directly obtained from [33, 4] by specializing $D_2 = \sigma^2$ in their result.

To prove D_{2l}, by the rate distortion theory [35], we have,

$$\frac{n}{2} \log \frac{\sigma^2}{D_1} \leq I(S_1^n; \hat{S}_1^n) \tag{3.8}$$

We next upper bound the RHS of (3.8) by giving S_2^n as the side information to the receiver.

$$\begin{aligned}
& I(S_1^n; \hat{S}_1^n) \\
\leq \; & I(S_1^n; Y^n) \\
\leq \; & I(S_1^n; S_2^n Y^n) \\
= \; & I(S_1^n; S_2^n) + I(S_1^n; Y^n | S_2^n) \\
= \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + h(Y^n | S_2^n) - h(Y^n | S_1^n S_2^n) \\
= \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + h(Y^n | S_2^n X_2^n) - h(Y^n | X_1^n X_2^n) \\
= \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + h(X_1^n + Z^n | S_2^n X_2^n) - h(Z^n) \\
\leq \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + h(X_1^n + Z^n) - h(Z^n) \\
\leq \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + \frac{n}{2} \log Var(X_{1i} + Z_i) - h(Z^n) \\
\leq \; & \frac{n}{2} \log \frac{1}{1-\rho^2} + \frac{n}{2} \log \frac{P_1 + N}{N} \\
= \; & \frac{n}{2} \log \frac{P_1 + N}{(1-\rho^2)N}
\end{aligned} \tag{3.9}$$

Combining (3.9) and (3.8), we complete the proof. □

Remark 1: The bound D_{l1} is tight when $\rho = 1$. In this case, the lower bound is trivially achieved by the uncoded transmission, i.e., choosing $X_i^n = \sqrt{\frac{P_i}{\sigma^2}} S_i^n, i = 1, 2,$ which gives the received signal $Y^n = \sqrt{\frac{P_1}{\sigma^2}} S_1^n + \sqrt{\frac{P_2}{\sigma^2}} S_2^n + Z^n$. The minimum MSE (MMSE) at the receiver can be trivially shown to be $D_{1n} = \sigma^2 \frac{N}{P_1+P_2+2\sqrt{P_1 P_2}+N}$, using standard estimation theory.

Remark 2: The bound D_{l2} is tight when $\rho = 0$. In this case, the lower bound is trivially achieved by keeping transmitter 2 quiet and using an optimum source channel code for transmitter 1. Notice that transmitter 1 can either implement a separation approach (i.e., a vector quantization followed by channel coding) or uncoded transmission. For the latter choice, we set, $X_1^n = \sqrt{\frac{P_1}{\sigma^2}} S_1^n$ and $X_2^n = 0$; the receiver can thus achieve the MMSE $D = \sigma^2 \frac{N}{P_1+N}$, which coincides with the lower bound D_{l2} when $\rho = 0$.

Remark 3: As shown in Fig. 3.3, D_{l1} is tighter than D_{l2} for large ρ while D_{l2} dominates D_{l1} for small ρ and for high signal to noise ratio (SNR). In addition to the

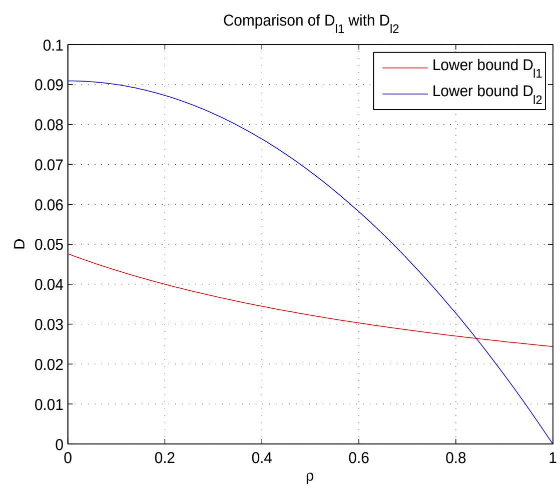

Figure 3.3: Comparison of D_{l1} and D_{l2} as a function of ρ for $P_1 = P_2 = 10, \sigma = 1, N = 1$.

above lower bound, we have the following achievable distortion.

Theorem 4 *For a GMAC with a single distortion constraint, the following distortion is achievable.*

$$D = \min\{D_u, D_s\}$$

where

$$D_u = \min_{\beta\in[0,1]} \sigma^2 \frac{\beta P_2(1-\rho^2)+N}{P_1+\beta P_2+2\sqrt{\beta P_1 P_2}\rho+N} \tag{3.10}$$

$$D_s = \sigma^2 \frac{P_1+P_2(1-\rho^2)+N}{P_1+P_2+N}\frac{N}{P_1+N} \tag{3.11}$$

The two achievable distortions correspond to respectively the uncoded transmission and separation approach, which we describe in details below.

Uncoded transmission

We have already shown that uncoded transmission is optimal when $\rho = 0$ and $\rho = 1$. One nice property of the uncoded transmission is that the codewords preserve the correlation of the source components. It is well known that uncoded transmission is optimal for a Gaussian source transmitted through an additive white Gaussian noise (AWGN) channel [36]. Gastpar in [37] generalized this optimality to a symmetric Gaussian sensor network, where noisy versions of a Gaussian source are sent to a fusion center through a GMAC with symmetric parameters. It has also been shown recently that the uncoded transmission is optimal for Gaussian broadcast channel with correlated Gaussian sources [38] and for GMAC channel with correlated Gaussian sources [33, 4] under certain SNR constraints.

For the uncoded transmission, the transmitted signals at encoders 1 and 2 are respectively

$$X_{1j}(\alpha) = \sqrt{\frac{\alpha P_1}{\sigma^2}} S_{1,j} \tag{3.12}$$

$$X_{2j}(\beta) = \sqrt{\frac{\beta P_2}{\sigma^2}} S_{2,j} \tag{3.13}$$

where $j = 1, 2, \cdots, n$ and $\alpha, \beta \in [0, 1]$. The receiver implements a MMSE estimator using the received signal:

$$Y^n = \sqrt{\frac{\alpha P_1}{\sigma^2}} S_1^n + \sqrt{\frac{\beta P_2}{\sigma^2}} S_2^n + Z^n. \tag{3.14}$$

The MMSE corresponding to a given (α, β) can be shown to be, using standard estimation theory,

$$D_u(\alpha,\beta) = \sigma^2 \frac{\beta P_2(1-\rho^2)+N}{\alpha P_1+\beta P_2+2\sqrt{\alpha\beta P_1 P_2}\rho+N} \tag{3.15}$$

Clearly, $D_u(\alpha, \beta)$ is monotone decreasing in α, i.e., transmitter 1 should always use full power. This leads to the achievable distortion D_u in (3.10).

Separation approach

For the separation approach, the encoders construct codewords whose rates are subject to the GMAC capacity constraint. Here, we can directly use Oohama's result for the Gaussian multi-terminal source coding problem with a single distortion constraint [41], i.e., $D_2 = \sigma^2$. The rate distortion region is characterized in the following lemma.

Lemma 1 *For the two terminal source coding problem with a single distortion constraint, the optimal rate distortion region is given as follows:*

$$R_1 \geq \frac{1}{2} \log \left[\frac{\sigma^2}{D_1} (1 - \rho^2 + \rho^2 2^{-2R_2}) \right] \tag{3.16}$$

The remaining problem is to find suitable (R_1, R_2) pair that falls into the capacity region of the GMAC. We note that the capacity region for GMAC with correlated message is still an open problem. We will hence choose an achievable, possibly sub-optimal, rate pair:

$$(R_1, R_2) = \left(\frac{1}{2} \log{(1 + \frac{P_1}{N})}, \frac{1}{2} \log{(1 + \frac{P_2}{P_1 + N})} \right).$$

Together with (3.16), we can get the desired distortion D_s as in (3.11) for the separation scheme.

Remark 4: Both uncoded and separation schemes are tight at $\rho = 0$. The uncoded scheme is also tight for the other extreme $\rho = 1$. In general, the uncoded scheme is better when ρ is close to 1 while the separation approach has a smaller distortion for small ρ values.

Remark 5: A close examination of the uncoded transmission also motivates a hybrid approach. From (3.10), one can easily arrive at the conclusion that $\beta = 0$ for ρ less than a certain threshold $\rho*$ which is a function of SNR. For example, if $P_1 = P_2 = 10$, and $N = 1$, then $\rho* \approx 0.42$. As such, even if there is still significant correlation between S_1 and S_2, transmitter 2 remains silent, reducing the system to a simple point-to-point source channel communication system. However, it is conceivable that if one resorts to digital transmission at transmitter 2, as long as the rate is below a certain rate constraint, the receiver can always decode the codeword reliably and subsequently subtract the codeword to recover the signal of the corresponding point-to-point channel model. Thus the decoded codeword from transmitter 2 serves as side information which may strictly improve upon the uncoded transmission for those $\rho < \rho^*$. We note that the HDA approach has been used when there is bandwidth mismatch in the source channel transmission system [42,43,44,45].

Specifically, by restricting transmitter 1 to analog transmission, transmitter 2 may encode at a rate $R_2 \leq \frac{1}{2}\log(1 + \frac{P_2}{P_1+N})$ to ensure its reliable decoding at the receiver. The receiver will estimate S_1^n based on both the analog signal $\sqrt{\frac{P_1}{\sigma^2}}S_1^n + Z^n$ and the digital message from encoder 2. To explicitly compute the achievable distortion of the hybrid scheme, we couple it with the quadratic Gaussian CEO problem through the following decomposition. For correlated Gaussian sources (S_1, S_2) with a correlation coefficient ρ, S_2 is statistically equivalent to

$$S_2 = \rho S_1 + W \tag{3.17}$$

where $W \sim \mathcal{N}(0, (1-\rho^2)\sigma^2)$ is independent of S_1. Hence, our original problem can now be reduced to a quadratic Gaussian CEO problem, whose rate distortion region was resolved in [46,47]. The equivalent CEO problem is illustrated in Fig. 3.4, where $\tilde{Z}^n \sim \mathcal{N}(0, \frac{N}{P_1/\sigma^2}I_n)$ and $\tilde{W}^n \sim \mathcal{N}(0, \frac{1-\rho^2}{\rho^2}\sigma^2 I_n)$, and I_n is the $n \times n$ identity matrix. For the present problem, since $S_1^n + \tilde{Z}^n$ is available at the receiver, this corresponds to $R_1 = \infty$ and $R_2 = \frac{1}{2}\log(1 + \frac{P_2}{P_1+N})$.

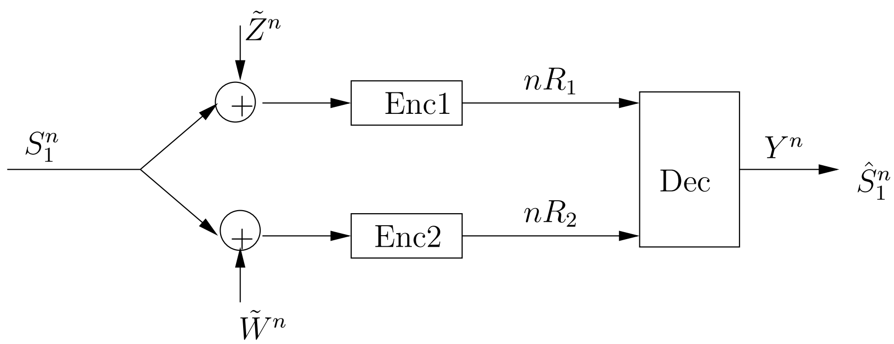

Figure 3.4: HDA scheme and its CEO representation.

The distortion D achieved by the HDA scheme is

$$D_h = \sigma^2 N \frac{(P_1+P_2)(1-\rho^2)+N}{(P_1+P_2+N)(P_1(1-\rho^2)+N)} \tag{3.18}$$

Note that by coupling the problem to a quadratic Gaussian CEO problem, the HDA is in fact a special case of the separation approach. Indeed, it is straightforward to show that D_h is no smaller than that achieved by the separation approach, c.f. (3.11). We also want to point out that the Gaussian CEO problem via a GMAC with encoder cooperation was studied recently in [48], where separation and uncoded transmission are considered. Our hybrid scheme can be viewed as a combination of the two coding schemes.

3.1.3 Comparison

We can compare the performance of the above achievable schemes together with the proposed lower bounds. Fig. 3.5 is the plot of various distortion bounds as a function of SNR. We choose $\sigma = 1, \rho = 0.5$ and $P_1 = P_2$. Fig. 3.6 plots the achievable distortions and lower bounds as a function of ρ for $\sigma = 1$ and $P_1 = 20, P_2 = 20, N = 1$. As we can see from Fig. 3.5 and Fig. 3.6, uncoded transmission is better than the others in the low SNR regime. For the high SNR regime, separation is the best.

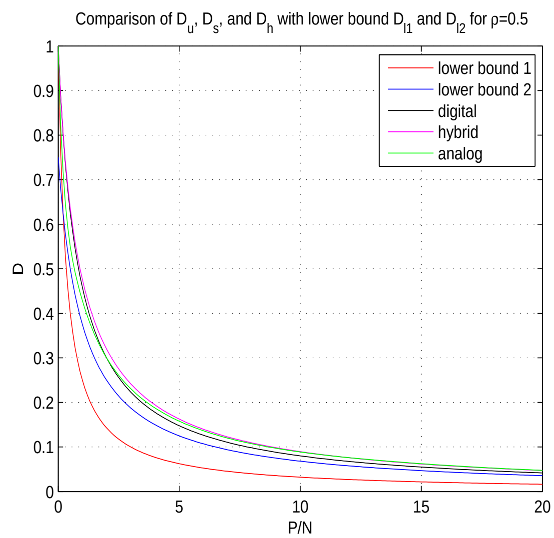

Figure 3.5: Comparison of D_u, D_s and D_h together with two lower bounds D_{l1} and D_{l2} as a function of $\frac{P}{N}$ for $\sigma = 1, \rho = 0.5$, for simplicity, we assume $P_1 = P_2$.

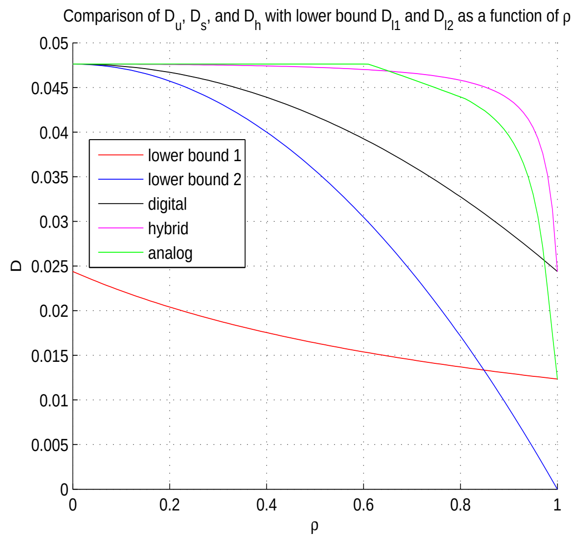

Figure 3.6: Comparison of D_u, D_s and D_h together with two lower bounds D_{l1} and D_{l2} as a function of ρ for $\sigma = 1$ and $P_1 = 20, P_2 = 20, N = 1$.

3.2 Gaussian Z Interference Channel With Correlated Gaussian sources

3.2.1 System Model

We now consider the problem of sending correlated Gaussian sources through a Gaussian Z interference channel. The problem is illustrated in Fig. 3.1 where the source statistics are the same as that defined in section II. Receivers 1 and 2 attempt to recover S_1^n and S_2^n respectively with as small distortions as possible. The Gaussian Z interference channel is defined as follows:

$$Y_1^n = c_{11}X_1^n + Z_1^n; \tag{3.19}$$

$$Y_2^n = c_{12}X_1^n + c_{22}X_2^n + Z_2^n; \tag{3.20}$$

where $Z_i^n \sim \mathcal{N}(0, I_n N_i)$, $i = 1, 2$ and I_n is the $n \times n$ identity matrix, and c_{11}, c_{12}, c_{22} are nonnegative channel parameters. We assume without loss of any generality $N_1 = N_2 = N$, since one can always scale the channel output. On the other hand, we do not assume Gaussian interference channels in standard form due to the power constraint: while standardization preserves capacity region by changing the power constraint, it does not necessarily preserve the distortion region for the source channel communication problem.

The encoder mappings are $f_i : S_i^n \to X_i^n(S_i^n) = f(S_i^n)$, $i = 1, 2$, subject to an individual power constraint as in (3.2). The decoder maps at the two receivers are $\phi_1(Y_1^n) = \hat{S}_1^n$ and $\phi_2(Y_2^n) = \hat{S}_2^n$. Again, squared error distortion is used in each receiver.

Definition 3 *A distortion pair (D_1, D_2) is said to be achievable if there exists a sequence of (X_1^n, X_2^n, n) codes and two decoding mapping functions $\phi_1(Y_1^n)$ and $\phi_2(Y_2^n)$ such that the MSEs for the source estimation satisfy*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} E(S_{ij} - \hat{S}_{ij})^2 \leq D_i \quad i = 1, 2 \tag{3.21}$$

The distortion region $\mathbb{D}$ is the convex closure of all achievable distortion pairs.

In addition, we define

$$\begin{aligned} D_{1min} = f(D_1) &= \min\{D_1 : (D_1, D_2) \in \mathbb{D}\} \\ D_{2min} = g(D_2) &= \min\{D_2 : (D_1, D_2) \in \mathbb{D}\} \end{aligned}$$

and we then define $D_1^* = \min\{D_1 : (D_1, D_{2min}) \in \mathbb{D}\}$ and $D_2^* = \min\{D_1 : (D_{1min}, D_2) \in \mathbb{D}\}$. Therefore, (D_{1min}, D_2^*) and (D_1^*, D_{2min}) are the two corner points of the distortion region.

3.2.2 Inner and upper bounds on the achievable distortion region

An inner bound for the distortion region $\mathbb{D}$ can be obtained by directly borrowing the result from Theorem 3.

Theorem 5 *For the Gaussian ZIC with correlated Gaussian source, the distortion region should always satisfy*

$$D_1 \geq \sigma^2 \frac{N}{c_{11}^2 P_1 + N} \tag{3.22}$$

$$D_2 \geq \sigma^2 \frac{N}{c_{12}^2 P_1 + c_{22}^2 P_2 + 2c_{12}c_{22}\sqrt{P_1 P_2}\rho + N} \tag{3.23}$$

$$D_2 \geq \sigma^2 \frac{(1-\rho^2)N}{c_{22}^2 P_2 + N} \tag{3.24}$$

For the achievable region, we consider first the uncoded transmission.

Theorem 6 *For the Gaussian ZIC with correlated Gaussian source, the uncoded transmission can achieve the following region:*

$$\begin{aligned} D_1 &\geq \sigma^2 \frac{N}{c_{11}^2 \alpha P_1 + N} && (3.25) \\ D_2 &\geq \sigma^2 \frac{c_{12}^2 \alpha P_1 (1-\rho^2) + N}{c_{12}^2 \alpha P_1 + c_{22}^2 \beta P_2 + 2 c_{12} c_{22} \rho \sqrt{\alpha \beta P_1 P_2} + N} && (3.26) \end{aligned}$$

for all $\alpha, \beta \in [0, 1]$.

We next consider two extreme cases: $\rho = 1$ and $\rho = 0$, and we have the following two propositions.

Proposition 5 *For the ZIC with correlated Gaussian source, if the correlation coefficient* $\rho = 1$*, then uncoded transmission is optimal and the optimal distortion region is given as follows.*

$$\begin{aligned} D_1 &\geq \sigma^2 \frac{N}{c_{11}^2 P_1 + N} && (3.27) \\ D_2 &\geq \sigma^2 \frac{N}{c_{12}^2 P_1 + c_{22}^2 P_2 + 2 c_{12} c_{22} \sqrt{P_1 P_2} + N} && (3.28) \end{aligned}$$

This can be established from the fact that the inner bound in Theorem 5 and the outer bound in Theorem 6 match when $\rho = 1$.

Proposition 6 *For the Gaussian ZIC with correlated Gaussian source, if the correlation coefficient* $\rho = 0$ *and the channel parameters* $c_{11} \geq c_{12}$*, then uncoded transmission can achieve the following corner point of the distortion region:*

$$\begin{aligned} D_{1min} &= \sigma^2 \frac{N}{c_{11}^2 P_1 + N} && (3.29) \\ D_2^* &= \sigma^2 \frac{c_{12}^2 P_1 + N}{c_{12}^2 P_1 + c_{22}^2 P_2 + N} && (3.30) \end{aligned}$$

Proof: This corner point can be achieved by uncoded transmission from Theorem 6. For the lower bound, we note that D_{1min} is obvious. To prove a lower bound for D_2^*, we need the following bound.

$$\begin{aligned} & \frac{n}{2} \log \frac{\sigma^2}{D_{1n}} + \frac{n}{2} \log \frac{\sigma^2}{D_{2n}} \\ \leq\ & I(S_1^n; Y_1^n) + I(S_2^n; Y_2^n) && (3.31) \\ \overset{(a)}{\leq}\ & I(X_1^n; Y_1^n) + I(X_2^n; Y_2^n) && (3.32) \end{aligned}$$

$$\stackrel{(b)}{\leq} \quad \frac{n}{2}\log\frac{c_{11}^2P_1+N}{N}+\frac{n}{2}\log\frac{c_{12}^2P_1+c_{22}^2P_2+N}{c_{12}^2P_1+N} \tag{3.33}$$

where (a) follows from the data processing inequality, i.e., $S_2^n \to X_2^n \to Y_2^n$ is a Markov chain when $\rho = 0$, (b) is from the sum rate capacity of the Gaussian ZIC with weak interference [17, 49]. Hence,

$$\frac{\sigma^4}{D_{1n}D_{2n}} \leq \frac{c_{11}^2P_1+N}{N}\cdot\frac{c_{12}^2P_1+c_{22}^2P_2+N}{c_{12}^2P_1+N} \tag{3.34}$$

Substitute $D_{1min} = \sigma^2\frac{N}{c_{11}^2P_1+N}$ into (3.34), we get

$$D_2^* \geq \sigma^2\frac{c_{12}^2P_1+N}{c_{12}^2P_1+c_{22}^2P_2+N} \tag{3.35}$$

Remark 6: In the case of $\rho = 0$, and by constraining $D_{2n} = D_{2min}$, uncoded scheme can only achieve a corner point (σ^2, D_{2min}), where $D_{2min} = \sigma^2\frac{N_2}{c_{22}^2P_2+N_2}$. This is because in order to achieve D_{2min} at decoder 2, transmitter 1 needs to shut down its transmission as discussed in the previous section, i.e., the uncoded transmission for the GMAC. This, however, can be improved by a hybrid transmission scheme. In this case, transmitter 1 can always transmit a coded message with a rate $R_1 = \frac{1}{2}\log(1+\frac{c_{12}^2P_1}{c_{22}^2P_2+N_2})$, so that decoder 2 can always decode it first. Transmitter 2 can use a simple uncoded transmission for S_2^n. Hence, we can achieve the following distortion pair:

$$D_{1h} = 2^{-2R_1} = \sigma^2\frac{c_{22}^2P_2+N_2}{c_{12}^2P_1+c_{22}^2P_2+N_2} \tag{3.36}$$

$$D_{2min} = \sigma^2\frac{N_2}{c_{22}^2P_2+N_2} \tag{3.37}$$

We note that one can also use the separation approach to attain the same distortion pair.

We can also obtain a lower bound for D_1^* for the case $c_{11} \geq c_{12}$, i.e., ZIC with weak interference. Notice that (3.34) still holds in this case, thus,

$$D_1^* \geq \sigma^2\Big(\frac{c_{22}^2P_2+N}{c_{12}^2P_1+c_{22}^2P_2+N}\Big)\frac{(c_{12}^2P_1+N)}{(c_{11}^2P_1+N)} \tag{3.38}$$

This lower bound on D_1^* matches the achievable D_{1h} if and only if $c_{12} = c_{11}$, which corresponds to the transition point to Gaussian ZIC with strong interference.

3.3 Conclusion and Discussions

In this chapter, we studied the problem of sending correlated Gaussian sources over Gaussian Z interference channels. We first considered communicating Gaussian courses over Gaussian multiple access channels with a single distortion constraint. Lower bounds and several achievable schemes were proposed. These bounds were applied to the Gaussian Z interference channel model for which we provided an inner bound and an achievable region obtained by uncoded transmission. For the extreme case of $\rho = 1$, uncoded scheme is optimal. For $\rho = 0$, uncoded transmission can achieve a corner point of the distortion region if the channel parameter $c_{11} \geq c_{12}$. We also give a sufficient condition for sending a correlated source over a general Z interference channel.

One can also consider the problem of communicating discrete correlated sources over discrete memoryless Z interference channels, defined as

$$\omega(y_1^n y_2^n | x_1^n x_2^n) = \prod_{i=1}^{n} p(y_{1i}|x_{1i}) p(y_{2i}|x_{1i}x_{2i}). \tag{3.39}$$

Following similar coding scheme as that of [29], we can obtain the following sufficient condition for lossy transmission with a given distortion constraint. Extension and evaluation of this sufficient condition for the Gaussian case will be reported in the next chapter.

Theorem 7 *For any correlated source* $(s_1^n, s_2^n) \sim \prod_{i=1}^n p(s_{1i}, s_{2i})$ *and a discrete memoryless Z interference channel* $\omega(y_1 y_2 | x_1 x_2)$ *as defined above, if there exists auxiliary random variables* Q, W_1, W_2, U*, such that the joint distribution of* $(s_1, s_2, q, w_1, w_2, u, x_1, x_2)$ *can be expanded as*

$$p(s_1 s_2) p(q) p(w_1|s_1 q) p(w_2|s_2 q) p(u|w_1 q) p(x_1|u w_1 q) p(x_2|w_2 q) \tag{3.40}$$

and the following conditions are satisfied:

$$\begin{aligned} I(W_1; S_1|Q) - I(W_1; U|Q) &< I(W_1 X_1; Y_1|UQ) && (3.41) \\ I(W_1; S_1|Q) &< I(W_1 X_1; Y|Q) && (3.42) \\ I(W_2; S_2|UQ) &< I(W_2 X_2; Y_2|UQ) && (3.43) \\ I(W_2; S_2|UQ) + I(W_1; U|Q) &< I(W_2 U_1 X_2; Y_2|Q) && (3.44) \end{aligned}$$

and

$$E[(S_1 - E(S_1|W_1, U))^2] \leq D_1, \tag{3.45}$$

$$E[(S_2 - E(S_2|W_2, U))^2] \quad \leq \quad D_2, \tag{3.46}$$

then $(D_1, D_2) \in \mathbb{D}$.

Chapter 4

Communicating Correlated Sources Over Interference Channels: The Lossy Case

We consider in this chapter the problem of lossy transmission of a pair of correlated sources over a two-user discrete memoryless interference channel (DMIC). The model is shown in Fig. 4.1. The source sequences $(s_1^n, s_2^n) \in (\mathcal{S}_1^n, \mathcal{S}_2^n)$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are finite, generated according to

$$p(s_1^n, s_2^n) = \prod_{i=1}^{n} p(s_{1i}, s_{2i}). \tag{4.1}$$

This pair of arbitrarily correlated source sequences are to be transmitted through a two user DMIC defined by

$$p(y_1^n y_2^n | x_1^n x_2^n) = \prod_{i=1}^{n} p(y_1 y_2 | x_1 x_2), \tag{4.2}$$

where X_1, X_2 are the channel inputs and Y_1, Y_2 are the channel outputs. Receivers 1 and 2 try to recover the source sequences S_1^n and S_2^n, respectively, subject to individual distortion constraints.

We are interested in characterizing the achievable distortion region for such lossy transmission.

Communicating correlated sources over multi-terminal channels have been a topic of research interest in the past decades. Cover, El-Gamal, and Salehi [2] first studied lossless transmission of arbitrarily correlated sources over a multiple access channel

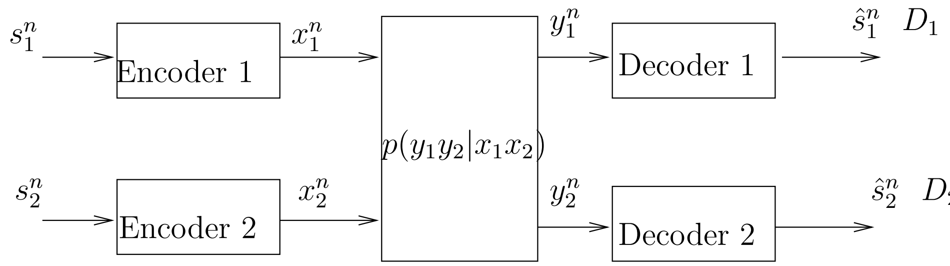

Figure 4.1: Lossy transmission of correlated sources over a DMIC.

(MAC). A sufficient condition was given for reliable transmission of a correlated source pair that includes various known capacity results as its special cases. The lossy counterpart of this problem was first studied by Salehi in [50], and the problem of sending a correlated Gaussian sources over a Gaussian MAC has been studied recently in [33, 51, 4]. Lossless transmission of arbitrarily correlated sources over a broadcast channel (BC) was studied by Han and Costa [3](with correction by Kramer and Nair [25]) and Minero and Kim [27]. Bross *et al* [38]studied the problem of sending correlated Gaussian sources over a Gaussian BC, and this problem was further investigated in [52, 53].

Lossless transmission of correlated sources over interference channels has previously been studied in [28] and chapter 2, where different coding schemes were proposed. The sufficient condition in chapter 2 includes the Han and Kobayashi region [12] when the sources are independent. In addition, it was shown in chapter 2 that separate source and channel coding is strictly sub-optimal. The lossy counterpart was initially studied in chapter 3 by considering the special case of transmitting correlated Gaussian sources over a Gaussian Z interference (GZIC).

In this Chapter, we derive a sufficient condition for lossy transmission with a prescribed distortion constraint of a pair of arbitrarily correlated discrete memoryless sources over a DMIC. The special case of sending correlated Gaussian sources over a Gaussian interference channel (GIC) is also studied. Several achievable schemes as well as lower bounds for the distortion regions are proposed.

The rest of this Chapter is organized as follows. Section gives the problem formulation and the main results. Section considers the Gaussian case. Finally, we conclude this chapter in section .

4.1 Arbitrarily correlated sources: the general case

For lossy transmission of correlated sources (S_1^n, S_2^n) generated according to (4.1) over a DMIC defined in (4.2), a length n joint source channel code consists of two encoder mappings, for $i = 1, 2$,

$$f_i : \mathcal{S}_i^n \to \mathcal{X}_i^n, \tag{4.3}$$

two decoder mappings:

$$\phi_i : \mathcal{Y}_i^n \to \hat{\mathcal{S}}_i^n, \tag{4.4}$$

and distortion measure functions:

$$d_i : \mathcal{S}_i \times \mathcal{S}_i \to R^+. \tag{4.5}$$

Note that the codeword length is assumed to be the same as the source sequence length, hence, there is no bandwidth mismatch issue in the present study.

Definition 4 *A distortion pair (D_1, D_2) is said to be* achievable *for transmitting the source (S_1^n, S_2^n) through the DMIC $p(y_1y_2|x_1x_2)$, if for any $\epsilon > 0$ and sufficiently large n, there exist a sequence of block codes $(f_1, f_2, \phi_1, \phi_2)$, such that, for $i = 1, 2$*

$$\frac{1}{n}\sum_{j=1}^{n} E[d_i(S_i, \hat{S}_i)] \;\; \leq \;\; D_i + \epsilon. \tag{4.6}$$

The achievable distortion region $\mathbb{D}$ is the union of all achievable distortion pairs.

4.1.1 Sufficient conditions: a joint source-channel coding approach

Theorem 8 *For transmitting sources $(s_1^n, s_2^n) \sim \prod_{i=1}^{n} p(s_{1i}, s_{2i})$ over the DMIC $p(y_1y_2|x_1x_2)$, a distortion pair (D_1, D_2) is achievable if there exist auxiliary random variables W_1, W_2, U_1, U_2 with joint distribution of all the variables factoring as:*

$$\begin{aligned} p(s_1, s_2, q, w_1, w_2, u_1, u_2, x_1, x_2) \;\; = \;\; & p(s_1, s_2)p(q)p(w_1|s_1q)p(u_1|s_1q)p(w_2|s_2q)p(u_2|s_2q) \\ & \cdot p(x_1|w_1u_1q)p(x_2|w_2u_2q), \end{aligned} \tag{4.7}$$

and functions, for $i = 1, 2$,

$$\phi_i \;\; : \;\; W_i \times U_i \times U_{(3-i)} \to \hat{S}_i, \tag{4.8}$$

such that,

$$
\begin{aligned}
I(W_1; S_1|U_1U_2Q) &< I(W_1X_1; Y_1|U_1U_2Q), && (4.9)\\
I(W_1U_1; S_1|U_2Q) &< I(W_1U_1X_1; Y_1|U_2Q), && (4.10)\\
I(W_1; S_1|U_1U_2Q) &< I(W_1U_2X_1; Y_1|U_1Q) \\
&\quad -I(U_2; S_2|U_1Q), && (4.11)\\
I(W_1U_1; S_1|U_2Q) &< I(W_1U_1U_2X_1; Y_1|Q) \\
&\quad -I(U_2; S_2|Q), && (4.12)\\
I(W_2; S_2|U_1U_2Q) &< I(W_2X_2; Y_2|U_1U_2Q), && (4.13)\\
I(W_2U_2; S_2|U_1Q) &< I(W_2U_2X_2; Y_2|U_1Q), && (4.14)\\
I(W_2; S_2|U_1U_2Q) &< I(W_2U_1X_2; Y_2|U_2Q) \\
&\quad -I(U_1; S_1|U_2Q), && (4.15)\\
I(W_2U_2; S_2|U_1Q) &< I(W_2U_1U_2X_2; Y_2|Q) \\
&\quad -I(U_1; S_1|Q), && (4.16)
\end{aligned}
$$

and D_1 and D_2 are given by

$$D_i = E[d_i(S_i, \phi_i(W_iU_iU_{(3-i)}))]. \qquad (4.17)$$

Remark: Theorem 8 gives a sufficient condition for lossy transmission of correlated sources (S_1^n, S_2^n) over the DMIC $p(y_1y_2|x_1x_2)$ with a distortion constraint (D_1, D_2). The coding scheme is a natural extension of the lossless case studied by Salehi and Kurtas [28]. The difference is that, before transmitting the source through the channel, the encoders first vector quantize the source S_1^n and S_2^n into W_1^n and W_2^n, respectively, and then implement a lossless joint source channel coding through the channel by treating (W_1^n, W_2^n) as the new source.

For the joint source channel coding part, we choose Salehi and Kurtas's [28] coding scheme for its simplicity. That is, for each source sequence s_1^n, encoder 1 chooses a covering u_1^n to represent the partial information embedded in s_1^n, and then choose a codeword x_1^n according to $p(x_1^n|w_1^n, u_1^n) = \prod_{j=1}^n p(x_{1j}|w_{1j}u_{1j})$, where w_1^n is the quantized version of the source sequence s_1^n. Similarly, for each source sequence s_2^n, encoder 2 chooses a covering u_2^n to represent the partial information embedded in s_2^n, and then choose a codeword x_2^n according to $p(x_2^n|w_2^n, u_2^n) = \prod_{j=1}^n p(x_{2j}|w_{2j}u_{2j})$. Here, u_1^n and u_2^n can be viewed as the partial information that needs to be decoded at both decoders for the purpose of cooperation. We also remark that, here u_1^n and u_2^n represent information directly from the source s_1^n and s_2^n, respectively, other than by representing information embedded in their quantized versions w_1^n and w_2^n. At the decoder side, decoder i needs to recover the codewords $(w_i^n, u_i^n, u_{(3-i)}^n)$ for $i = 1, 2$. For the distortion part, upon correct decoding, decoder i estimates the original s_i^n based on $(w_i^n, u_i^n, u_{(3-i)}^n)$, that is, $\hat{s}_{ij} = \phi_i(w_{ij}, u_{ij}, u_{(3-i)j})$.

One may choose our coding scheme proposed in Chapter 2 for the joint source channel coding part, that is, random source partition followed by a correlation preserving coding. While this coding scheme reduces to the Han and Kobayashi [12] region for DMIC with independent messages, the use of random source partition makes the analysis much more complex for lossy transmission case.

The detailed proof of Theorem 8 is given in Appendix C. In the following, we discuss two special cases of Theorem 8.

When specializing our result to the lossless case, i.e., $D_1 = D_2 = 0$, we get the following sufficient conditions for the lossless transmission of correlated sources (s_1^n, s_2^n) over the DMIC $p(y_1 y_2 | x_1 x_2)$.

Proposition 7 *A source pair* $(s_1^n, s_2^n) \sim \prod_{j=1}^n p(s_{1j}, s_{2j})$ *can be reliably transmitted through a DMIC* $p(y_1 y_2 | x_1 x_2)$, *if*

$$\begin{aligned}
H(S_1|U_1U_2Q) &< I(S_1X_1;Y_1|U_1U_2Q),\\
H(S_1|U_2Q) &< I(S_1X_1;Y_1|U_2Q),\\
H(S_1|U_1U_2Q) + I(S_2;U_2|U_1Q) &< I(S_1U_2X_1;Y_1|U_1Q),\\
H(S_1|U_2Q) + I(S_2;U_2|Q) &< I(S_1U_2X_1;Y_1|Q),\\
H(S_2|U_1U_2Q) &< I(S_2X_2;Y_2|U_1U_2Q),\\
H(S_2|U_1Q) &< I(S_2X_2;Y_2|U_1Q),\\
H(S_2|U_1U_2Q) + I(S_1;U_1|U_2Q) &< I(S_2U_1X_2;Y_2|U_2Q),\\
H(S_2|U_1Q) + I(S_1;U_1|Q) &< I(S_2U_1X_2;Y_2|Q).
\end{aligned}$$

Proposition 7 follows by setting $W_1 = S_1$ and $W_2 = S_2$ in Theorem 8. It is not surprising that the conditions in Proposition 7 coincide with Salehi and Kurtas's result [28] for the lossless transmission of correlated sources over DMIC.

When specializing our result to the discrete memoryless Z interference channel (DMZIC) case, defined by $p(y_1 y_2 | x_1 x_2) = p(y_1|x_1) p(y_2|x_1 x_2)$, i.e., there is no communication link between encoder 2 and receiver 1, we get the following achievable distortion region.

Proposition 8 *For any correlated sources* $(s_1^n, s_2^n) \sim \prod_{j=1}^n p(s_{1j}, s_{2j})$*, and a DMZIC* $p(y_1y_2|x_1x_2) = p(y_1|x_1)p(y_2|x_1x_2)$*, a distortion pair* (D_1, D_2) *is achievable if*

$$\begin{aligned} I(W_1; S_1|U_1Q) &< I(W_1X_1; Y_1|U_1Q), \\ I(W_1U_1; S_1|Q) &< I(W_1U_1X_1; Y|Q), \\ I(W_2; S_2|U_1Q) &< I(W_2X_2; Y_2|U_1Q), \\ I(W_2; S_2|U_1Q) + I(S_1; U_1|Q) &< I(W_2U_1X_2; Y_2|Q), \end{aligned}$$

and D_i *for* $i = 1, 2$*, are given by*

$$D_i = E[d_i(S_i, \phi_i(W_iU_1))]. \tag{4.18}$$

Proposition 8 is obtained by setting $U_2 = \emptyset$ in Theorem 8. The conditions in Proposition 8 is slightly different from Theorem 7. The difference is that, in Theorem 7, the auxiliary random variables W_1, W_2, U_1 satisfy the Markov chain $W_2 \to S_2 \to S_1 \to W_1 \to U_1$, where in this chapter, they satisfy the Markov chain $W_2 \to S_2 \to S_1 \to (W_1, U_1)$ and $W_1 \to S_1 \to U_1$. This also implies a slight difference between the two coding schemes. In Theorem 7, U_1^n is generated based on W_1^n only, i.e., U_1^n carries the partial information embedded in W_1^n. In the present chapter, as remarked before, U_1^n represents the information embedded in the original source S_1^n. We believe that the latter one is a more natural scheme.

4.2 The Gaussian special case

In this section, we examine the problem of sending correlated Gaussian sources over a GIC.

Consider a pair of Gaussian random sources (S_1^n, S_2^n), independent and identically distributed (i.i.d.) generated according to a zero mean bivariate Gaussian distribution. The covariance matrix of (S_1, S_2) is given by

$$K_{S_1S_2} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{4.19}$$

where $\rho \in [-1, 1]$, and $0 < \sigma_i < \infty, i = 1, 2$. The source sequences are to be transmitted over a Gaussian interference channel defined by

$$Y_1^n = c_{11}X_1^n + c_{12}X_2^n + Z_1^n, \tag{4.20}$$

$$Y_2^n = c_{21}X_1^n + c_{22}X_2^n + Z_2^n, \tag{4.21}$$

where X_1, X_2 are the channel input and Y_1, Y_2 are the output, and $Z_i^n \sim \mathcal{N}(0, N_i I_n)$, $i = 1, 2$, are i.i.d. Gaussian noise sequences and I_n is the $n \times n$ identity matrix.

The realization S_1^n is available at encoder 1 and S_2^n available at encoder 2. The encoders choose a sequence of codewords $X_i^n, i = 1, 2$, based on S_1^n and S_2^n, respectively. The encoding is subject to individual power constraints, for $i = 1, 2$,

$$\frac{1}{n}\sum_{j=1}^{n} EX_{ij}^2(S_i^n) \leq P_i. \tag{4.22}$$

The decoding maps are the same as defined before, and the corresponding distortion is measured by mean squared error (MSE), defined as, for $i = 1, 2$,

$$D_{in} \triangleq \frac{1}{n}\sum_{j=1}^{n} E(S_{ij} - \hat{S}_{ij})^2. \tag{4.23}$$

Without loss of generality [33], we assume that two source components have equal variance ($\sigma_1 = \sigma_2 = \sigma$) and have a non-negative correlation coefficient ($\rho \in [0, 1]$). We also assume that $c_{11} = c_{22} = 1$, since decoders can always scale their channel output. For simplicity, we only consider symmetric GIC with equal power constraints, i.e., $c_{12} = c_{21} = a, P_1 = P_2 = P$ and $N_1 = N_2 = N$, and we are interested in characterizing the achievable distortion region, i.e., $\mathbb{D} = \bigcup(D_1, D_2)$, where the union is taken over all achievable distortion pairs.

4.2.1 Lower and upper bounds on the achievable distortion region

Before discussing the achievable schemes, we introduce a simple lower bound (cut set bound) for the distortion region.

Theorem 9 *A necessary condition for* $(D_1, D_2) \in \mathbb{D}$ *is that*

$$D_1 \geq \sigma^2 \frac{N}{(1 + 2a\rho + a^2)P + N}, \tag{4.24}$$

$$D_2 \geq \sigma^2 \frac{N}{(1 + 2a\rho + a^2)P + N}. \tag{4.25}$$

Proof: By the rate distortion theory [35, Theorem 13.3.2], we have

$$I(S^n, \hat{S}^n) \geq \frac{n}{2} \log \frac{\sigma^2}{D_1} \tag{4.26}$$

We next want to upper bound $I(S^n, \hat{S}^n)$. Notice that $S^n \to Y_1^n \to \hat{S}^n$ forms a Markov chain, we get,

$$\begin{aligned} I(S^n; \hat{S}^n) &\leq I(S^n; Y_1^n) & (4.27)\\ &\leq I(S^n X_1^n X_2^n; Y_1^n) & (4.28)\\ &= H(Y_1^n) - H(Y_1^n | X_1^n X_2^n) & (4.29)\\ &\leq \frac{n}{2} \log \frac{(1 + 2a\rho + a^2)P + N}{N} & (4.30) \end{aligned}$$

where the last step uses the fact that $\frac{EX_{1i}X_{2i}}{\sqrt{EX_{1i}^2}\sqrt{EX_{2i}^2}} \leq \rho$, $i = 1, 2, \cdots, n$, by Witsenhausen's lemma [58]. Combining (4.26) and (4.30), we get lower bound (4.24), lower bound (4.25) can be obtained similarly.□

For the achievability part, the achievable distortion region obtained in Theorem 8 still applies to the Gaussian case. However, the evaluation of this region is hard in general, since we do not know the optimal choices of the auxiliary random variables (W_1, W_2, U_1, U_2). In the following, we give a sufficient condition which can be viewed as a special case of Theorem 8.

Proposition 9 *For transmitting correlated Gaussian source over Gaussian interference channels, a distortion pair (D_1, D_2) is achievable if*

$$\begin{aligned} I(W_1; S|W_2) &< I(W_1 X_1; Y_1 | W_2)\\ I(W_2; T|W_1) &< I(W_2 X_1; Y_1 | W_1)\\ I(W_1; S|W_2) + I(W_2; T) &< I(W_1 W_2 X_1; Y_1)\\ I(W_2; T|W_1) &< I(W_2 X_2; Y_2 | W_1)\\ I(W_1; S|W_2) &< I(W_1 X_2; Y_2 | W_2)\\ I(W_2; T|W_1) + I(W_1; S) &< I(W_1 W_2 X_2; Y_2) & (4.31) \end{aligned}$$

and D_i for $i = 1, 2$, are given by

$$D_i = E[d_i(S_i, \phi_i(W_i))]. \tag{4.32}$$

The Proposition is proved by setting $U_1 = W_1$, $U_2 = W_2$ and $Q = \emptyset$ in Theorem 8. To further calculate the bound, we study two special regimes of the GIC, the strong interference regime ($a \geq 1$) and the noisy interference regime ($2\sqrt{a}(aP + 1) \leq 1$).

4.2.2 The strong interference regime

In this regime, the capacity region of GIC with independent messages is that of a compound MAC [16], and for our symmetric GIC case, is given by

$$
\begin{aligned}
R_1 &\leq \frac{1}{2}\log(1+\frac{P}{N}), &(4.33)\\
R_2 &\leq \frac{1}{2}\log(1+\frac{P}{N}), &(4.34)\\
R_1+R_2 &\leq \frac{1}{2}\log(1+\frac{P+a^2P}{N}). &(4.35)
\end{aligned}
$$

For sending correlated Gaussian sources over this GIC, by directly evaluating the sufficient conditions in Theorem 8, we get the following achievable distortion region.

Corollary 4 *For sending correlated Gaussian sources over a symmetric GIC with $a \geq 1$, a distortion pair (D_1, D_2) is achievable if there exist rates $R_1 > 0$ and $R_2 > 0$ such that the following conditions hold*

$$
\begin{aligned}
R_i &< \frac{1}{2}\log\frac{(1-\tilde{\rho}^2)P+N}{(1-\tilde{\rho}^2)N}, &(4.36)\\
R_1+R_2 &< \frac{1}{2}\log\frac{(1+2a\tilde{\rho}+a^2)P+N}{(1-\tilde{\rho}^2)N}, &(4.37)\\
D_i &> \sigma^2 2^{-2R_i}\frac{1-(1-2^{-2R_{(3-i)}})\rho^2}{1-\tilde{\rho}^2}, &(4.38)
\end{aligned}
$$

where $i = 1, 2$ and $\tilde{\rho} = \sqrt{1-2^{-2R_1}}\sqrt{1-2^{-2R_2}}\rho$.

Sketch of proof: Corollary 4 follows directly from Proposition 9. To evaluate the bound in Proposition 9, we further choose the test channels $W_1 = c_1S + Z_{W_1}$ and $W_2 = c_2T + Z_{W_2}$, where $c_i = \sqrt{1-2^{-2R_i}}$, $Z_{W_i} \sim \mathcal{N}(0, 2^{-2R_i})$, $i = 1, 2$, and R_1 and R_2 are the quantization rates. By choosing the encoding function as $X_{ij} = \sqrt{\frac{P}{\sigma^2}}W_{ij}$, $i = 1, 2$ and $j = 1, 2, \cdots, n$, and implement a MMSE estimator at both decoder (hence, $D_1 = \mathrm{Var}(S|W_1W_2)$ and $D_2 = \mathrm{Var}(T|W_1W_2)$), we get the conditions as in Corollary 4. □

Remark: In the evaluation part, the choices of test channels and encoding schemes are exactly the same as the vector quantizer scheme introduced by Lapidoth and Tinguely [33,4] for the problem of sending correlated Gaussian sources over a Gaussian MAC. It is not surprising that the conditions obtained here are similar to theirs as in the regime $a \geq 1$, we use the IC as a compound MAC by setting $U_i = W_i$, $i = 1, 2$ in the evaluation.

If only symmetric distortion is considered, i.e., $D_1 = D_2 = D$, we get the following proposition.

Proposition 10 *A symmetric distortion D is achievable if there exists some $R > 0$ such that*

$$R < \frac{1}{2}\log\frac{(1-\tilde{\rho}^2)P+N}{(1-\tilde{\rho}^2)N}, \tag{4.39}$$

$$2R < \frac{1}{2}\log\frac{(1+2a\tilde{\rho}+a^2)P+N}{(1-\tilde{\rho}^2)N}, \tag{4.40}$$

$$D = \sigma^2 2^{-2R}\frac{1-(1-2^{-2R})\rho^2}{1-\tilde{\rho}^2}, \tag{4.41}$$

where $\tilde{\rho} = (1-2^{-2R})\rho$.

To compare the performance, we consider two alternative coding schemes: uncoded transmission and the separation approach.

Uncoded transmission: by choosing $X_{ij} = \sqrt{\frac{P}{\sigma^2}}S_{ij}$, for $i = 1, 2$ and $j = 1, 2, \cdots, n$, and implementing a minimum mean squared error (MMSE) estimation at both decoders, we get the following achievable distortion for the uncoded scheme

$$D_1 = D_2 = D_u = \sigma^2\frac{a^2(1-\rho^2)P+N}{(1+2a\rho+a^2)P+N}. \tag{4.42}$$

Separation: For the channel coding part, the capacity region is given by (4.33-4.35). For the symmetric case, we choose[1]

$$R_1 = R_2 = R = \frac{1}{4}\log(1+\frac{(1+a^2)P}{N}). \tag{4.43}$$

For the source coding part, since in the strong interference channel regime, the two decoders can decode both users' messages, it is reasonable to implement an optimal quadratic Gaussian two-terminal source coding scheme given by [54]. In the case of $R_1 = R_2 = R$, the distortion region in [54] reduces to

$$D = \sqrt{2^{-4R}(1-\rho^2)+\rho^2 2^{-8R}}. \tag{4.44}$$

Together with (4.43), we get the following achievable distortion for the separation scheme

$$D_s = \frac{\sqrt{N[(1+a^2)P(1-\rho^2)+N]}}{(1+a^2)P+N}. \tag{4.45}$$

[1] Assuming without loss of generality, constraint (4.35) is active, i.e., $a^2 \leq P+N$.

We remark here that, in the case of $a = 1$, the received signals Y_1^n and Y_2^n are statistically the same, hence, the lower bound [33, Corollary 1] developed by Lapidoth and Tinguely for the Gaussian MAC case is still valid for our problem. We have the following proposition.

Proposition 11 *For $a = 1$, a necessary condition for $(D, D) \in \mathbb{D}$ is that*

$$D \geq \sigma^2 \frac{P(1-\rho^2)+N}{2P(1+\rho)+N} \quad \text{if} \quad \frac{P}{N} \leq \frac{\rho}{1-\rho^2}, \tag{4.46}$$

$$D \geq \sigma^2 \sqrt{\frac{(1-\rho^2)N}{2P(1+\rho)+N}} \quad \text{if} \quad \frac{P}{N} > \frac{\rho}{1-\rho^2}. \tag{4.47}$$

4.2.3 The noisy interference regime

In the noisy interference regime, the analysis follows similarly as above, and we only summarize the results below.

Proposition 12 *In the symmetric setting with $2\sqrt{a}(aP+1) \leq 1$, a symmetric distortion D is achievable if there exists rate $R > 0$ such that*

$$R < \frac{1}{2}\log\left(1 + \frac{(1+a\tilde{\rho})^2 P}{a^2(1-\tilde{\rho}^2)P+N}\right), \tag{4.48}$$

$$D = \sigma^2 2^{-2R}, \tag{4.49}$$

where $\tilde{\rho} = (1-2^{-2R})\rho$.

Corollary 5 *In the symmetric setting with $2\sqrt{a}(aP+1) \leq 1$, uncoded transmission and separation scheme can achieve respectively*

$$D_u = \sigma^2 \frac{a^2(1-\rho^2)P+N}{(1+2a\rho+a^2)P+N} \tag{4.50}$$

$$D_s = \sigma^2 2^{-2R} = \sigma^2 \frac{a^2P+N}{(1+a^2)P+N} \tag{4.51}$$

4.2.4 Numerical examples

Fig.4.2 shows the performance of different achievable schemes for $a = 1$. From the figure, uncoded scheme is optimal when $\frac{P}{N}$ is less than a certain threshold, which coincides with the observation in [33]. In general, uncoded scheme is the best for low SNR, while joint source and channel coding scheme outperforms others for high SNR.

Another observation is that, joint source and channel coding uniformly outperforms separation scheme for all SNR. We also observe that the distortion attained with the uncoded scheme will not converge to 0 as SNR increases, while that of the joint source channel coding and separation schemes converge to 0. Similar observations can be made for a=2, as shown in Fig.4.3. For the noisy interference regime, the performance is shown in Fig.4.4. In this case, uncoded transmission uniformly beats all others for all SNR. The joint source channel coding described above, although utilizing the rate gain from the correlation between codewords, does not make use of the correlation at the final estimation stage. This also implies that the choice of auxiliary random variables is not optimal.

4.3 Summary

In this chapter, we studied lossy transmission of two correlated sources over a two-user DMIC. An achievable distortion region was proposed and it was shown to include Salehi and Kurtas's result on lossless transmission of correlated sources over DMIC as a special case. The Gaussian case was also studied. A lower bound and several achievable schemes including uncoded transmission, the separation approach, and joint source channel coding were introduced. It was shown that in the strong interference regime, uncoded transmission is the best at low SNR while the joint source channel coding outperforms others at high SNR. In addition, the joint source channel coding scheme uniformly outperforms the separation scheme for all SNR. In the noisy interference regime, uncoded transmission outperforms others for all SNR range. The joint source channel coding scheme introduced in section 4.2 is only a special case of our main result. The optimal choice of auxiliary random variables is still unknown. Possible improvement can be made by considering the superposition approach proposed in [55, 4]. Moreover, the lower bound in Theorem 9 is loose in general.

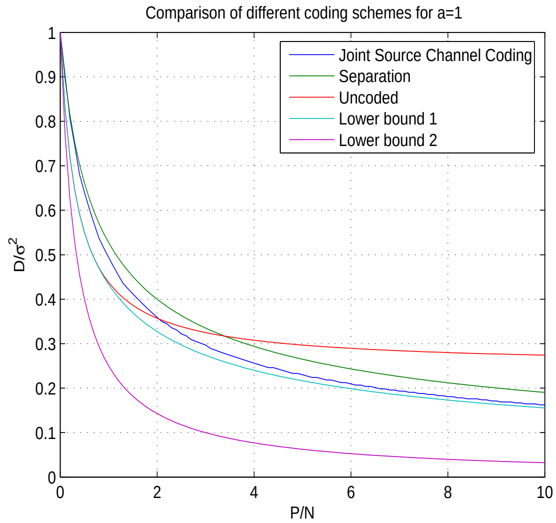

Figure 4.2: Comparison of different achievable schemes for a=1, where lower bound 1 is by Proposition 11 and lower bound 2 is from Theorem 9.

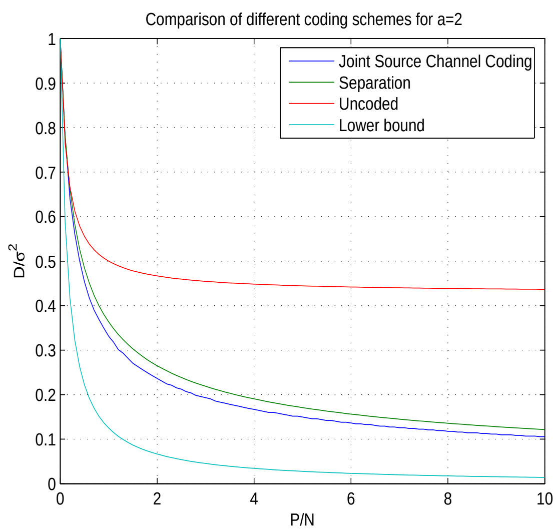

Figure 4.3: Comparison of different achievable schemes for a=2, where the lower bound is from Theorem 9.

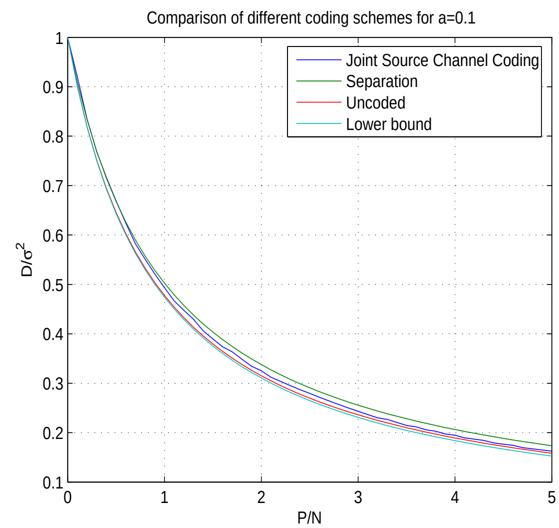

Figure 4.4: Comparison of different achievable schemes for a=0.1.

Chapter 5

The Common Information of N Dependent Random Variables

We briefly introduced the three classic measure of common information in the introduction part of this thesis. They are Shannon's mutual information, Gács and Körner's common randomness and Wyner's common information. However, all these three notions of common information were originally proposed for two random variables case.

Generalizing of mutual information to N random variables was first reported in [64]. The generalization comes from the observation that for a pair of random variables, Shannon's information measures is consistent with the Venn diagram for set operation and a comprehensive treatment was available in [65, 66]. Gács and Körner's common randomness was recently generalized to multiple random variables by Tyagi, Narayan and Gupta in [67], which extends the encoding process in the definition of common randomness to that of N terminals.

In this chapter, we generalize Wyner's common information of a pair of random variables to that of N dependent variables. We show that the operational meaning defined in both approaches are still valid. Moreover, we establish some monotone property of such generalization which contrast to the notion of 'common' information. Specifically, we show that the common information does *not* decrease as the number of variables increases while keeping the same marginal distribution. This is different from the other two notions of common information. Examples on evaluating $C(X_1, X_2, \cdots, X_N)$ are given for circularly symmetric binary sources and the asymptotic results are also studied.

The rest of this chapter is organized as follows. Section 5.1 gives the problem formulation and main results. Section 5.2 gives some examples and discussions. Section 5.3 concludes the chapter.

5.1 Common information of N random variables

Let $X_1, X_2, \cdots, X_N$ be random variables that take values on the finite alphabet sets $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_N$ with joint distribution $P(x_1, x_2, \cdots, x_N)$. A length-n source sequence $(x_1^n, x_2^n, \cdots, x_N^n)$ is distributed according to

$$P^{(n)}(x_1^n, x_2^n, \cdots, x_N^n) = \prod_{i=1}^{n} P(x_{1i}, x_{2i}, \cdots, x_{Ni}). \tag{5.1}$$

Our generalization of Wyner's common information is to define a similar measure for N random variables by preserving the conditional independence structure through the introduction of an auxiliary random variable. Specifically, we define

$$C(X_1, X_2, \cdots, X_N) \triangleq \inf I(X_1, X_2, \cdots, X_N; W), \tag{5.2}$$

where the infimum is taken over all the joint distributions of $(X_1, X_2, \cdots, X_N, W)$ such that

$$\sum_w P(x_1, x_2, \cdots, x_n, w) \quad = \quad P(x_1, x_2, \cdots, x_N), \tag{5.3}$$

$$P(x_1, ..., x_n|w) \quad = \quad \prod_{i=1}^{n} P(x_i|w). \tag{5.4}$$

Thus the marginal distribution of $(X_1, X_2, \cdots, X_N)$ is $P(x_1, x_2, \cdots, x_N)$ and $(X_1, \cdots, X_N)$ are conditionally independent given W.

We now give two interpretation for the common information C of N dependent random variables.

5.1.1 A Gray-Wyner source coding network interpretation

For the Gray-Wyner source coding network, we start with the definition of encoder-decoders.

Definition 5 *A* $(n, \mathcal{M}_0, \mathcal{M}_1, \cdots, \mathcal{M}_N)$ *code consists of the following:*

- *An encoder mapping*

$$f : \mathcal{X}_1^n \times \mathcal{X}_2^n \times \cdots \times \mathcal{X}_N^n \to \mathcal{M}_0 \times \mathcal{M}_1 \times \cdots \times \mathcal{M}_N,$$

where $\mathcal{M}_i = \{1, 2, \cdots, 2^{nR_i}\}$.

- N *decoders* g_i, *for* $i = 1, 2, \cdots, N$,

$$g_i : \mathcal{M}_i \times \mathcal{M}_0 \to \mathcal{X}_i^n. \tag{5.5}$$

The probability of error is defined as

$$P_e^{(n)} = Pr\{(\hat{X}_1^n \hat{X}_2^n \cdots \hat{X}_N^n) \neq (X_1^n, X_2^n, \cdots, X_N^n)\}, \tag{5.6}$$

where $\hat{X}_i^n = g_i(M_i, M_0)$ for $i = 1, \cdots, N$.

Definition 6 *A number* R_0 *is said to be* achievable *if for any* $\epsilon > 0$, *we can find an* n *sufficiently large such that there exists a* $(n, \mathcal{M}_0, \mathcal{M}_1, \cdots, \mathcal{M}_N)$ *code with*

$$\mathcal{M}_0 \leq 2^{nR_0} \tag{5.7}$$

$$P_e^{(n)} \leq \epsilon, \tag{5.8}$$

$$\frac{1}{n} \sum_{i=0}^{N} \log \mathcal{M}_i \leq H(X_1, X_2, \cdots, X_N) + \epsilon. \tag{5.9}$$

As with the case for two random variables, C_1 is defined as the infimum of all achievable R_0.

5.1.2 A random variable generator interpretation

For the second approach of approximating joint distribution, we again start with the following definition.

Definition 7 *An* $(n, \mathcal{M}, \Delta)$ generator *consists of the following:*

- *a message set* $\mathcal{W} \in \{1, 2, \cdots, 2^{nR}\}$;
- *for all* $w \in \mathcal{W}$ *and* N *conditional probability distributions* $q_i^{(n)}(x_i^n|w)$, *for* $i = 1, 2, \cdots, N$, *define the probability distribution on* $\mathcal{X}_1^n \times \mathcal{X}_2^n \times \cdots \times \mathcal{X}_N^n$

$$Q^{(n)}(X_1^n, X_2^n, \cdots, X_N^n) = \sum_{w \in \mathcal{W}} \frac{1}{\mathcal{M}} \prod_{i=1}^{N} q_i^{(n)}(x_i^n|w). \tag{5.10}$$

Thus the N processors serve as random number generators each generating independent and identically distributed (i.i.d.) sequence $\hat{X}_i^n$ according to $q(x_i|w)$ and the output of the processors follow joint distribution defined in (5.1). Let

$$\Delta = D_n(P^{(n)};Q^{(n)}) = \frac{1}{n} \sum_{x_i^n \in X_i^n, i=1,2,\cdots,N} P^{(n)} \log \frac{P^{(n)}}{Q^{(n)}}, \tag{5.11}$$

where $P^{(n)}$ and $Q^{(n)}$ are defined as in (5.1) and (5.10) respectively.

Definition 8 *A number R is said to be achievable if for all $\epsilon > 0$, we can find an n sufficiently large such that there exists a $(n, \mathcal{M}, \Delta)$ generator with $\mathcal{M} \leq 2^{nR}$ and $\Delta \leq \epsilon$.*

We define C_2 as the infimum of all achievable R.

5.1.3 Main results: $C_1 = C_2 = C$

The main result of this chapter is the following Theorem.

Theorem 10

$$C_1 = C_2 = C(X_1, X_2, \cdots, X_N). \tag{5.12}$$

Thus both C_1 and C_2 admit single letter characterization which coincides with $C(X_1, \cdots, X_N)$.

The proof of Theorem 10 is given in Appendix D

5.2 Examples and discussions

5.2.1 The monotone property of common information C

We start with the following example. Let $X = (X^{'}, U, V)$, $Y = (Y^{'}, V, W)$ and $Z = (Z^{'}, W, U)$ where the random variables $X^{'}, Y^{'}, Z^{'}, U, V, W$ are mutually independent. It is easy to show that for this example

$$I(X;Y;Z) = K(X,Y,Z) = 0,$$

whereas

$$C(X,Y,Z) = H(UVW).$$

On the other hand,

$$C(X,Y) \;\; = \;\; H(V),$$

$$
\begin{aligned}
C(X,Z) &= H(U), \\
C(Y,Z) &= H(W).
\end{aligned}
$$

What is interesting is that the inclusion of an additional variable increases the common information. This is somewhat surprising: if the information is *common* it ought to be non-increasing when more random variables are included. Indeed, we can prove the following general result:

Lemma 2 *Let $(X_1, \cdots, X_N) \sim p(x_1, \cdots, x_N)$. For any two sets A, B that satisfy $A \subseteq B \subseteq \mathcal{N} = \{1, 2, \cdots, N\}$,*

$$
C(\mathbf{X}_A) \leq C(\mathbf{X}_B), \tag{5.13}
$$

where $\mathbf{X}_A = \{X_i, i \in A\}$ and $\mathbf{X}_B = \{X_i, i \in B\}$.

Proof: Let W' be the W that achieves $C(\mathbf{X}_B)$, i.e., $I(W'; \mathbf{X}_B) = \inf I(W; \mathbf{X}_B)$. But $A \subseteq B$, thus $\mathbf{X}_B$ conditionally independent given W' implies that $\mathbf{X}_A$ is conditionally independent given W'. Thus

$$
\begin{aligned}
I(\mathbf{X}_B; W') &\geq I(\mathbf{X}_A; W') \\
&\geq \inf I(\mathbf{X}_A; W)
\end{aligned}
$$

where the infimum is taken over all W such that $\mathbf{X}_A$ is independent given W.

This monotone property perhaps suggests that the name common information, while meaningful for pair of variables, no longer suits the generalization to N variables. We comment here that Gács and Körner's common randomness follows a different monotone property

$$
K(\mathbf{X}_A) \geq K(\mathbf{X}_B)
$$

while there is no definitive inequality relationship for mutual information.

As a consequence, we have for any N random variables

$$
C(X_1, X_2, \cdots, X_N) \geq K(X_1, X_2, \cdots, X_N).
$$

We now examine another example in which Wyner's common information increases as the number of the observations increases. Moreover the common information eventually converges and the asymptote suggests that the notion of common information may have potential application in certain inference problem.

5.2.2 Circularly symmetric binary sources

We now examine another example in which Wyner's common information increases as the number of the observations increases. Moreover the common information eventually converges and the asymptote suggests that the notion of common information may have potential application in certain inference problem.

Consider first the example of three binary random variables X_1, X_2, X_3 with joint distribution

$$P(x_1, x_2, x_3) = \begin{cases} \frac{1}{2} - \frac{3}{4}a_0 & \text{if } x_1 = x_2 = x_3 \\ \frac{1}{4}a_0 & \text{otherwise} \end{cases} \tag{5.14}$$

where the parameter a_0 satisfies $0 \leq a_0 \leq \frac{1}{2}$.

It can be easily verified that

$$Pr\{X_i = 0\} = \frac{1}{2}, \tag{5.15}$$

for $i = 1, 2, 3$ and that for $1 \leq i, j \leq 3, i \neq j$,

$$Pr(X_i = x_i, X_j = x_j) = \frac{1}{2}(1 - a_0)\delta_{x_i,x_j} + \frac{1}{2}a_0(1 - \delta_{x_i,x_j}), \tag{5.16}$$

where $\delta_{a,b} = 1$ if $a = b$ and 0 otherwise.

Thus, each pair of (X_i, X_j), $i \neq j$, can be viewed as a doubly symmetric binary source as defined in [59]. We refer to this set of exchangeable binary sources circularly symmetric binary source. For such circularly symmetric binary source (X_1, X_2, X_3) with joint distribution given in (5.14) and random variables (X_1, X_2, X_3, W) that satisfy (5.3) and (5.4), we have the following lemma.

Lemma 3

$$H(X_1|W) + H(X_2|W) + H(X_3|W) \leq 3h(a_1), \tag{5.17}$$

where $a_1 = \frac{1}{2} - \frac{1}{2}(1 - 2a_0)^{\frac{1}{2}}$.

This lemma is a direct consequence of Wyner's result on doubly symmetric binary source [59]. Therefore, we have,

$$\begin{aligned} & I(X_1X_2X_3; W) \\ = \; & H(X_1X_2X_3) - H(X_1X_2X_3|W), \end{aligned}$$

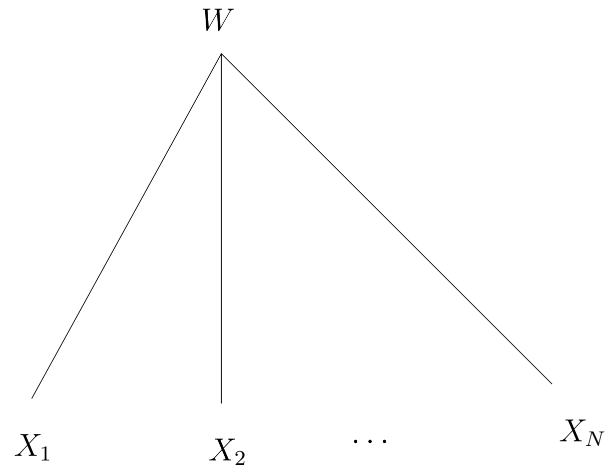

Figure 5.1: A simple Bayesian graph model.

$$
\begin{aligned}
&= H(X_1X_2X_3) - \sum_{i=1}^{3} H(X_i|W), \\
&\geq H(X_1X_2X_3) - 3h(a_1), \\
&= 1 + h(a_0) + a_0 + (1-a_0)h\left(\frac{a_0}{2(1-a_0)}\right) - 2h(a_1),
\end{aligned}
\tag{5.18}
$$

This lower bound can indeed be achieved by choosing the following random variables. Let W be a random variable with $p_W(0) = p_W(1) = 1/2$, i.e., a Bernoulli(1/2) random variable. Let each X_i be the output of a binary symmetric channel (BSC) with crossover probability a_1 with W as input. The channels share the common input W but are otherwise independent of each other. This is illustrated in the simple Bayesian graph model in Fig. 5.1 with $N = 3$ where each link represents a BSC with crossover probability a_1.

Thus, the common information of this circularly symmetric binary source is,

$$
\begin{aligned}
C(X_1, X_2, X_3) &= 1 + a_0 + h(a_0) + \\
&\quad (1-a_0)h\left(\frac{a_0}{2(1-a_0)}\right) - 3h(a_1),
\end{aligned}
\tag{5.19}
$$

Notice that any pair of (X_i, X_j) is a doubly symmetric binary source [59], therefore,

$$C(X, Y) = 1 + h(a_0) - 2h(a_1).$$

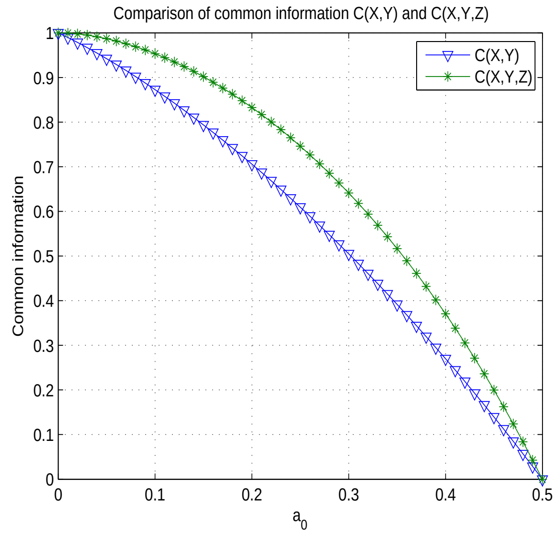

Figure 5.2: Comparison of common information.

It is straightforward to check that

$$C(X, Y, Z) > C(X, Y)$$

when $0 < a_0 < \frac{1}{2}$. This is also shown numerically in Fig. 5.2.

5.2.3 Asymptotic results for N binary random variables

We now study the generalization of above example to arbitrary N and in particular the asymptotic value of the common information for the circularly symmetric binary sources.

Consider N binary random variables $X_1, X_2, \cdots, X_N$ with joint distribution $p(X_1, X_2, \cdots, X_N)$ generated by an underlying Bayesian graph model as in Fig. 5.1, where W is a Bernoulli(1/2) random variable and each X_i, $i = 1, 2, \cdots, N$, is the output of a BSC with crossover probability $a_1 (0 \leq a_1 \leq \frac{1}{2})$ with a common input W. Hence, for $x_1, x_2, \cdots, x_N \in \{0, 1\}$,

$$P(x_1, x_2, \cdots, x_n) = \sum_{w \in \{0,1\}} \frac{1}{2} \prod_{i=1}^{N} P_i(x_i|w), \tag{5.20}$$

where for each $i = 1, 2, \cdots, N$, $p_i(x_i|w) = (1 - a_1)$ if $x_i = w$ and a_1 otherwise.

Similarly, we have,

$$\sum_{i=1}^{N} H(X_i|W) \leq N h(a_1), \tag{5.21}$$

for any random variable W that satisfies (5.3) and (5.4).

Therefore, $C(X_1, X_2, \cdots, X_N)$ can be lower bounded by

$$C(X_1, X_2, \cdots, X_N) \geq H(X_1, X_2, \cdots, X_N) - Nh(a_1). \tag{5.22}$$

On the other hand, the above lower bound is achievable by exactly the same W in the above Bayesian model. Hence, we have,

$$C(X_1, X_2, \cdots, X_N) = H(X_1, X_2, \cdots, X_N) - Nh(a_1), \tag{5.23}$$

where $H(X_1, X_2, \cdots, X_N)$ can be calculated from (5.20).

Now consider the above model but with increasing N. For any ϵ and $a_1 < 1/2$, it is clear that

$$H(W|X_1, X_2, \cdots, X_N) < \epsilon$$

for N sufficiently large. This can be established by the Fano's inequality as one can estimate W with arbitrary reliability given $X_1, \cdots, X_N$ for sufficiently large N. Therefore,

$$\begin{aligned}
& C(X_1, X_2, \cdots, X_N) \\
= \; & H(X_1, X_2, \cdots, X_N) - Nh(a_1), \\
= \; & H(X_1, X_2, \cdots, X_N, W) - Nh(a_1) \\
& -H(W|X_1, X_2, \cdots, X_N), \\
\geq \; & H(W) - \epsilon,
\end{aligned} \tag{5.24}$$

where the last step is from the fact that $H(X_1, X_2, \cdots, X_N|W) = Nh(a_1)$. On the other hand,

$$C(X_1, \cdots, X_N) \leq H(W)$$

for any N. Thus, for $a_1 < 1/2$,

$$\lim_{N \to \infty} C(X_1, X_2, \cdots, X_N) = H(W) = 1$$

If $a_1 = 1/2$, then $X_1, \cdots, X_N$ are mutually independent hence $C(X_1, \cdots, X_N) = 0$.

5.3 Conclusions

This chapter generalized Wyner's common information, defined for a pair of random variables, to that of N dependent random variables. We showed that it is the minimum common information rate R_0 needed for N separate decoders to recover their intended sources losslessly while keeping the total rate close to the entropy bound. It is also equivalently to the smallest rate of the common input to N independent processors (random number generators), such that the output distribution is approximately the same as the given joint distribution. It was shown that such generalization leads to the phenomenon of 'common' information non-decreasing as the number of sources increases.

For the example of circularly symmetric binary sources, we show that common information not only increases as N grows, but eventually converges to the entropy of W that achieves $C(X_1, \cdots, X_N)$.

Chapter 6

Common Information For Continuous Random Variables

In previous chapter, we generalized Wyner's common information to N dependent random variables. The interpretations there only apply to random variables with finite alphabet sets. Intuitively, the definition of common information itself applies to random variables with continuous alphabet. However, it is not clear what physical interpretation such quantity carries for continuous random variables.

In this chapter, we provide such an interpretation using the rate distortion result for the Gray-Wyner network as described in Fig. 6.1. That is, instead of requiring the sources to be reproduced losslessly at the two decoders, we allow certain distortions subject to given distortion constraints [68]. It turns out that Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. A surprising result is that, as Wyner's common information is only a function of the joint distribution, this smallest common rate remains constant even if the distortion constraints vary, as long as they are less than certain thresholds.

We limit ourself to two random variable case in this chapter. The extension to N continuous random variables will be reported in the future works.

The rest of the chapter is organized as follows. Section 6.1 gives the problem formulation and the main results. In section 6.2, two examples, the doubly symmetric binary source and the bivariate Gaussian source, are given. We conclude in Section 6.3.

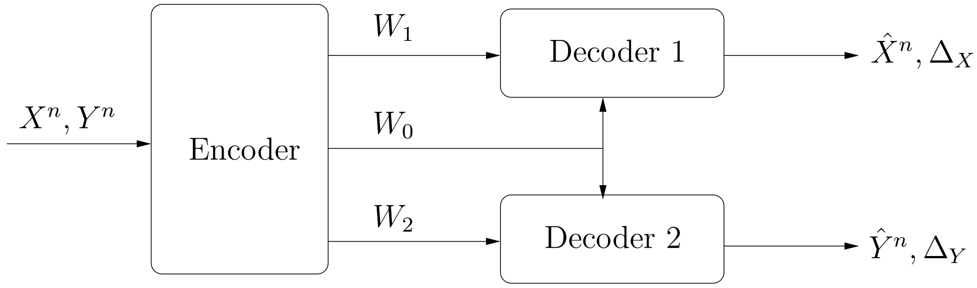

Figure 6.1: Gray-Wyner source coding network.

6.1 A Gray-Wyner lossy source coding interpretation

Let $\{(X_k, Y_k)\}_{k=1}^{\infty}$ be independent copies of a pair of dependent random variables $(X, Y) \sim Q(x, y)$ which take values in some arbitrary (finite, countable, or continuous) spaces $\mathcal{X} \times \mathcal{Y}$. Here, we use $Q(x, y)$ to denote the joint distribution of (X, Y), i.e., probability mass function if (X, Y) are discrete and probability density function if (X, Y) are continuous. Thus the joint distribution of length n vectors (X^n, Y^n) is

$$Q^n(x^n, y^n) = \prod_{i=1}^{n} Q(x_i, y_i). \tag{6.1}$$

The common information of the pair (X, Y) is a functional of Q and is defined as

$$C(X, Y) \triangleq \inf I(X, Y; W), \tag{6.2}$$

where the infimum is taken over all random variable triples X, Y, W satisfying

- (C1) The marginal distribution for X, Y is $Q(x, y)$,
- (C2) X and Y are conditionally independent given W.

Let us consider the lossy source coding problem described in Fig. 6.1. The encoder observes a pair of sequences (X^n, Y^n), and map them to three messages W_0, W_1, W_2 with

$$W_i \in \{1, \cdots, 2^{nR_i}\},$$

for $i = 0, 1, 2$. Let $d_1(x, \hat{x})$ and $d_2(x, \hat{y})$ be bounded single letter distortion functions defined on $\mathcal{X} \times \mathcal{X}$ and $\mathcal{Y} \times \mathcal{Y}$ respectively. Decoder 1 reproduces X^n from (W_0, W_1) subject to an average distortion constraint Δ_1; decoder 2 reproduces Y^n

from (W_0, W_2) subject to an average distortion constraint Δ_2. We now give a precise definition of the quantity $C_3(\Delta_1, \Delta_2)$, which is the smallest common rate R_0 such that the total rate meets the rate distortion bound with joint decoding.

Definition 9 *An (n, M_0, M_1, M_2) rate distortion code consists of the following:*

- *One encoder mapping f_E*

$$f_E \;:\; \mathcal{X}^n \times \mathcal{Y}^n \to I_{M_0} \times I_{M_1} \times I_{M_2}, \tag{6.3}$$

where $I_{M_i} = \{0, 1, 2, \cdots, M_i - 1\}$ for $i = 0, 1, 2$.

- *Two decoder mappings $f_D^{(X)}, f_D^{(Y)}$*

$$f_D^{(X)} \;:\; I_{M_0} \times I_{M_1} \to \mathcal{X}^n, \tag{6.4}$$

$$f_D^{(Y)} \;:\; I_{M_0} \times I_{M_2} \to \mathcal{Y}^n. \tag{6.5}$$

Let $f_E(X^n, Y^n) = (W_0, W_1, W_2)$, $1 \le W_i \le M_i$ and

$$\hat{X}^n = f_D^{(X)}(W_0, W_1), \tag{6.6}$$

$$\hat{Y}^n = f_D^{(Y)}(W_0, W_2). \tag{6.7}$$

Denote by (Δ_X, Δ_Y) the average distortion between encoder inputs and decoder outputs:

$$\Delta_X = Ed_1(X^n, \hat{X}^n), \tag{6.8}$$

$$\Delta_Y = Ed_2(Y^n, \hat{Y}^n), \tag{6.9}$$

where

$$d_1(x^n, \hat{x}^n) = \frac{1}{n}\sum_{k=1}^{n} d_1(x_k, \hat{x}_k), \tag{6.10}$$

$$d_2(y^n, \hat{y}^n) = \frac{1}{n}\sum_{k=1}^{n} d_2(y_k, \hat{y}_k). \tag{6.11}$$

An (n, M_0, M_1, M_2) code with distortion (Δ_X, Δ_Y) is referred to as an $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ rate distortion code.

Definition 10 *For any $\Delta_1, \Delta_2 \geq 0$, a number R_0 is said to be (Δ_1, Δ_2)-achievable if for any $\epsilon > 0$ we can find n sufficiently large such that there exists a $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ rate distortion code with*

$$\begin{aligned} M_0 &\leq 2^{nR_0}, &(6.12)\\ \sum_{i=0}^{2} \frac{1}{n} \log M_i &\leq R_{XY}(\Delta_1, \Delta_2) + \epsilon, &(6.13)\\ \Delta_X \leq \Delta_1 + \epsilon \quad , &\quad \Delta_Y \leq \Delta_2 + \epsilon. &(6.14)\end{aligned}$$

where $R_{XY}(\Delta_1, \Delta_2)$ is the rate distortion function for (X, Y) with joint encoding and decoding, i.e.,

$$R_{XY}(\Delta_1, \Delta_2) = \min I(X, Y; \hat{X}, \hat{Y}), \tag{6.15}$$

where the minimum is taken over all the test channels $q_t(\hat{x}, \hat{y}|x, y)$ such that $Ed_1(X, \hat{X}) \leq \Delta_1,\ Ed_2(Y, \hat{Y}) \leq \Delta_2$.

Definition 11 *$C_3(\Delta_1, \Delta_2)$ is defined as the infimum of all R_0 that is (Δ_1, Δ_2)-achievable.*

We now state the main results.

Theorem 11 *The common information $C(X;Y) = C_3(\Delta_1, \Delta_2)$ in some neighborhood of the origin $\{(\Delta_1, \Delta_2) : 0 \leq \Delta_1, \Delta_2 \leq \gamma\}$ provided that*

$$Q(x, y) > 0 \quad \textit{all } x \in \mathcal{X}, y \in \mathcal{Y}, \tag{6.16}$$

and d_1, d_2 satisfy

$$\begin{aligned} d_1(x, \hat{x}) &> d_1(x, x) = 0, x \neq \hat{x}, &(6.17)\\ d_2(y, \hat{y}) &> d_2(y, y) = 0, y \neq \hat{y}. &(6.18)\end{aligned}$$

A proof of Theorem 11 is given in Appendix G.

The condition on d_1 and d_2 set forth in the theorem amounts to requiring the distortion function be normal, as defined in [66].

If (X, Y) are discrete random variables with finite alphabet, and $d_1 = d_2 = d_H$ are the Hamming distortion, defined as

$$d_H(u, \hat{u}) = \begin{cases} 0, & u = \hat{u} \\ 1, & u \neq \hat{u}, \end{cases} \tag{6.19}$$

then for $\Delta_1 = \Delta_2 = 0$, $C_3(\Delta_1, \Delta_2) = C_1 = C(X;Y)$. Therefore, approach 1 in [59] is a special case of Theorem 1.

Theorem 12 *For the symmetric case $\Delta_1 = \Delta_2 = \Delta$, $C_3(\Delta) = C(X,Y)$ if and only if $\Delta \le R_{XY}^{-1}(C(X,Y))$, where $R_{XY}^{-1}(\cdot)$ denotes the inverse function of $R_{XY}(\Delta,\Delta)$, i.e, the distortion rate function.*

A proof of Theorem 12 is given in Appendix H.

6.2 Examples

6.2.1 Doubly symmetric binary source (DSBS)

Consider a DSBS as in [59, 68]. That is, a binary source where $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and for $x, y = 0, 1$,

$$Q(x,y) = \frac{1}{2}(1-a_0)\delta_{x,y} + \frac{1}{2}a_0(1-\delta_{x,y}), \tag{6.20}$$

$0 \le a_0 \le \frac{1}{2}$ and $\delta_{x,y}$ is an indicator function of $x = y$. X can be considered as an unbiased binary input to a binary symmetric channel (BSC) with crossover probability a_0 and Y as the corresponding output, or vice versa.

It was shown in [59] that for the DSBS

$$C(X;Y) = 1 + h(a_0) - 2h(a_1), \tag{6.21}$$

where $h(a_0)$ is the binary entropy function for $0 \le a_0 \le 1$ and $a_1 = \frac{1}{2} - \frac{1}{2}(1-2a_0)^{\frac{1}{2}}$.

For a DSBS with Hamming distortion $d_1 = d_2 = d_H$ and symmetric distortion constraint $\Delta_1 = \Delta_2 = \Delta$, the joint rate distortion function [71] is given by,

$$R_{XY}(\beta,\beta) = \begin{cases} 1 + h(a_0) - 2h(\beta) & \text{if } 0 \le \beta \le a_1 \\ L(1-a_0) - \frac{1}{2}\{L(2\beta - a_0) + L[2(1-\beta) - a_0]\} & a_1 \le \beta \le \frac{1}{2} \end{cases}, \tag{6.22}$$

where $L(x) = -x \log x$.

It can be seen that

$$R_{XY}(a_1, a_1) = 1 + h(a_0) - 2h(a_1) = C(X,Y).$$

Therefore, by Theorem 12 we have $\gamma = a_1$ and $C_3(\Delta,\Delta) = 1 + h(a_0) - 2h(a_1)$ for any $0 \le \Delta \le a_1$.

Remark: $C_3(\Delta,\Delta)$ for any $0 \le \Delta \le a_1$ is achieved by $R_0 = R_{XY}(a_1,a_1) = 1 + h(a_0) - 2h(a_1)$, $R_1 = R_{X|\tilde{X}\tilde{Y}}(\Delta)$, and $R_2 = R_{Y|\tilde{X}\tilde{Y}}(\Delta)$, where $(\tilde{X},\tilde{Y})$ are the random variables achieving $R_{XY}(a_1,a_1)$. The test channels are

$$Pr\{X = x|\tilde{x}\tilde{y}\} = (1-a_1)\delta_{x,\tilde{x}} + a_1(1-\delta_{x,\tilde{x}}), \tag{6.23}$$

$$Pr\{Y = y|\tilde{x}\tilde{y}\} \quad = \quad (1-a_1)\delta_{y,\tilde{y}} + a_1(1-\delta_{y,\tilde{y}}). \tag{6.24}$$

Hence, $R_1 = R_2 = h(a_1) - h(\Delta)$.

6.2.2 Gaussian source

In this section we consider the case when X, Y are bivariate Gaussian with zero mean and covariance matrix

$$K = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{6.25}$$

Proposition 13 *For the Gaussian random variable (X, Y) described above, the common information is*

$$C(X;Y) \quad = \quad \frac{1}{2}\log\frac{1+\rho}{1-\rho}. \tag{6.26}$$

The proof is given in Appendix K. Proposition 13 can be extended to multivariate Gaussian distributions.

Corollary 6 *For N joint Gaussian random variables $X_1, X_2, \cdots, X_N$ with covariance matrix*

$$K_N = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{6.27}$$

the common information is

$$C(X_1, X_2, \cdots, X_N) = \frac{1}{2}\log(1 + \frac{N\rho}{1-\rho}). \tag{6.28}$$

Proposition 14 *For bivariate Gaussian random variables X, Y with zero mean and covariance matrix in (6.25) and squared error distortion $d_1(u, \hat{u}) = d_2(u, \hat{u}) = (u - \hat{u})^2$, we have*

$$C_3(\Delta, \Delta) = C(X;Y), \tag{6.29}$$

for any $\Delta \leq 1 - \rho$.

Proof: The joint rate distortion function for Gaussian random variables with symmetric squared error distortion [71] is

$$R_{XY}(\beta, \beta) = \begin{cases} \frac{1}{2}\log\frac{1-\rho^2}{\beta^2} & 0 \le \beta \le 1-\rho \\ \frac{1}{2}\log\frac{1+\rho}{2\beta-(1-\rho)} & 1-\rho \le \beta \le 1 \\ 0 & \beta \ge 1 \end{cases}. \tag{6.30}$$

Thus we have

$$R_{XY}(1-\rho, 1-\rho) = \frac{1}{2}\log\frac{1+\rho}{1-\rho} = C(X,Y). \tag{6.31}$$

By Theorem 12, $\gamma = 1-\rho$. This means that $C_3(\Delta, \Delta) = C(X;Y)$ for any $\Delta \le 1-\rho$.

Remark: $C_3(\Delta, \Delta)$ for any $0 \le \Delta \le 1-\rho$ is achieved by $R_0 = R_{XY}(1-\rho, 1-\rho)$, $R_1 = R_{X|\tilde{X}\tilde{Y}}(\Delta) = \frac{1}{2}\log\frac{1-\rho}{\Delta}$, $R_2 = R_{Y|\tilde{X}\tilde{Y}}(\Delta) = \frac{1}{2}\log\frac{1-\rho}{\Delta}$, where $(\tilde{X}, \tilde{Y})$ are the random variables achieving $R_{XY}(1-\rho, 1-\rho)$.

6.3 Conclusion

In this chapter, we generalized Wyner's common information to that of continuous random variables and provided a lossy source coding interpretation using the Gray-Wyner network. A surprising observation is that the the minimum common rate for lossy source coding is invariant to the distortion constraint as long as it is less than a certain threshold.

Chapter 7

Quantization for Distributed Testing of Independence

In this chapter, we consider the problem of *distributed* test of statistical independence under communication constraints. While independence test is frequently encountered in various applications, *distributed* independence test is particularly useful for events detection in sensor networks: data correlation often occurs among sensor observations in the presence of a target. Focusing on the Gaussian case because of its tractability, we study in this chapter the characteristics of optimal quantization rule for distributed test of independence.

Consider the following hypothesis testing problem: a pair of random sequences (X_i, Y_i), $i = 1, \cdots, n$, with (X_i, Y_i) independent and identically distributed (i.i.d.) according to the joint probability density function

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\Big).$$

The two hypotheses under test are

$$\begin{cases} H_0: & \rho \neq 0, \\ H_1: & \rho = 0. \end{cases} \tag{7.1}$$

i.e., (X, Y) is bivariate Gaussian and they are independent under H_1 and dependent under H_0. Notice that assuming zero mean and unit variance does not lose any generality as long as the mean values and variances are known, since we can always transform the random variables to standard bivariate Gaussian by proper scaling and shifting.

In the centralized case where X and Y sequences are available, this statistical inference problem can be solved straightforwardly by applying some standard statistical inference frameworks depending on the situations (e.g., whether or not ρ is known under H_0) [77].

The problem becomes much more interesting and challenging when $\{X_i\}$ and $\{Y_i\}$ are not directly available; instead, compressed versions of X and Y subject to some rate constraints are used for the test of independence. This distributed test of independence is the focus of the present work. To be more specific, we assume that $\{X_i\}$ and $\{Y_i\}$ are available respectively at two distributed sensor nodes. The sensor nodes communicate their data to the fusion center under a communication constraint of R_1 and R_2 bits per observation. The fusion center, upon receiving the sensor data, makes a final decision on whether $\{X_i\}$ and $\{Y_i\}$ are correlated or not. Our attempt is to understand properties of optimal quantizers at distributed nodes where the optimality is associated with the performance at the fusion center with regard to the dependence test.

Consider first the large sample regime, i.e., n is large. Given that (X_i, Y_i) form an i.i.d. sequence, it is easy to show that any reasonable quantizers will lead to a test with diminishing error probability as n grows for $R_1 > 0$ and $R_2 > 0$. Thus a sensible criterion is the speed with which the error probability approaches zero, i.e., the error exponent characterization. This is indeed the underlying reason for the problem setting where the null hypothesis H_0 represents dependence while independence occurs under H_1. Applying Stein's lemma [35] to the hypothesis testing problem (7.12), for a given type I error constraint, the error exponent for the type II error (i.e., the Kullback Leibler distance between the distributions under H_0 and H_1) reduces to the mutual information between suitable random variables. For example, with centralized test, the optimal error exponent becomes $I(X;Y)$. Our focus in the large sample regime is to study quantizer properties in the context of distributed test against independence with Gaussian sources. Motivated by practical constraints that often require simple sensor processing, we consider only scalar quantizers at local sensors with 1 bit per observation. That is, $R_1 = R_2 = 1$ and each sensor quantizer is 'memoryless'. Our objective will be therefore to determine the optimal scalar quantizer structure that maximizes $I(U, V)$ where U and V are the one bit quantizer output for the two sensors.

Characterizing optimal error exponents for dependence test with communication constraints was first considered by Ahlswede and Csiszár [78]. In particular, for the special case of test of independence problem with one sided data compression, i.e., $R_2 = \infty$, a single letter characterization of the optimal error exponent was obtained in [78]. An overview of related work can be found in [80] and the references therein. We note here that the majority of the reported work are largely restricted to (X, Y) being discrete memoryless sources. Distributed test of independence with continuous alphabet sources (e.g., Gaussian sources) have been much less investigated.

We will also study distributed test of independence in the finite sample regime, where asymptotic error exponent result can not apply. We consider a Bayesian approach where the priors for the two hypotheses are assumed to be π_0 and π_1 respectively. We derive quantizer properties for minimum error probability with both two-sided and one-sided tests, with the latter referring to the situation of $H_0 : \rho > 0$ and $H_1 : \rho = 0$.

The rest of the chapter is organized as follows. In Section 7.1, we give the problem statement and our main results. Section 7.2 are numerical examples. At last, we conclude in section 7.3.

7.1 Problem statement and main results

Consider the following hypothesis testing of independence problem for standard zero mean bivariate Gaussian source.

$$\begin{aligned} H_0 :& \quad (X_1, X_2) \sim P_{X_1 X_2}, \\ H_1 :& \quad (X_1, X_2) \sim Q_{X_1 X_2}, \end{aligned} \tag{7.2}$$

where $P_{X_1X_2}$ and $Q_{X_1X_2}$ are given by

$$\begin{aligned} H_0 : \quad & P_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\Big), \quad \rho \neq 0, \\ H_1 : \quad & Q_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi} \exp(-(x_1^2 + x_2^2)/2). \end{aligned} \tag{7.3}$$

We assume that the source sequences (x_1^n, x_2^n) are i.i.d. according to the corresponding hypothesis. The fusion center does not have direct access to the source sequence (x_1^n, x_2^n), but can be informed about the source only at limited rates. Precisely, the local sensors encode the source sequence (x_1^n, x_2^n) into $(u_1^{k_1}, u_2^{k_2})$, respectively, where

$$\frac{1}{n} \log \|U_1^{k_1}\| = \frac{k_1}{n} \quad \leq \quad R_1 \tag{7.4}$$

$$\frac{1}{n}\log\|U_2^{k_2}\| = \frac{k_2}{n} \quad \leq \quad R_2 \tag{7.5}$$

We assume that there are no errors for the transmission of $u_1^{k_1}$ and $u_2^{k_2}$. The fusion center will make the final decision U_0 based on received $u_1^{k_1}$ and $u_2^{k_2}$. The model is shown in Fig.7.1.

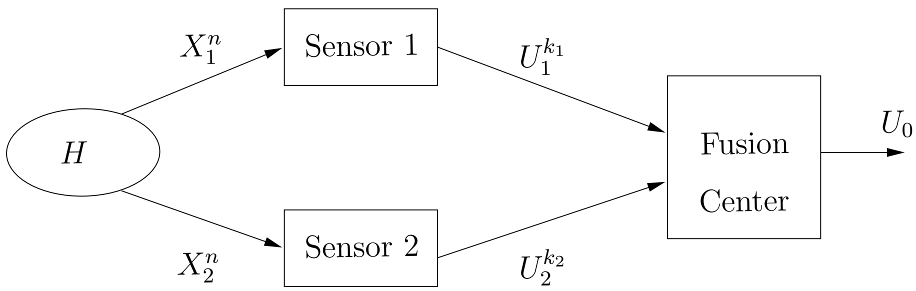

Figure 7.1: Hypothesis testing with communication constraints.

7.1.1 Large sample regime

Consider the hypothesis test described in (7.2). The fusion center does not have direct access to the source sequence (X_i, Y_i), $i = 1, 2, \cdots, n$, but can be informed about the sources only at limited rates. Precisely, the local sensors apply scalar quantizers to their respective observations:

$$\begin{aligned} U_i &= \gamma_1(X_i), \\ V_i &= \gamma_2(Y_i), \end{aligned}$$

where U_i and $V_i \in \{0, 1\}$.

For the large sample regime, the fusion center will decide H_0 or H_1 given the sequence (U_i, V_i) $i = 1, \cdots, n$ and we are to characterize the optimal quantizers that maximize the error exponent. Using the Neyman-Pearson criterion, we assume that the rejection region is the set $B \subset \mathcal{X}^n$ whose complement of B is $\bar{B}$. The minimum probability of type II error for a prescribed arbitrary small probability of type I error, denoted by $\beta_{R_1,R_2}(n, \epsilon)$, is defined as

$$\beta_{R_1,R_2}(n, \epsilon) = \min_B \{Q^n(\bar{B}) | B \subset \mathcal{X}^n, P^n(B) \leq \epsilon\}. \tag{7.6}$$

The error exponent associated with $\beta_{R_1,R_2}(n,\epsilon)$ is, under the problem setup, the mutual information between U and V, $I(U,V)$. Our problem becomes finding a pair of binary quantizers such that $I(U,V)$ is maximized.

By restricting each sensor to an one bit scalar quantizer, we have the following result.

Theorem 13 *For the distributed test of independence problem described in (7.2) where each local quantizer is restricted to be one bit scalar quantizer with a single threshold, the optimal quantizers that maximize the error exponent are a sign detector, i.e., a binary quantizer with threshold*

$$t_1 = t_2 = 0. \tag{7.7}$$

Remark: while the result is rather intuitive with the symmetric problem setting, the proof is rather lengthy and is sketched in Appendix I. Notice that the result relies on the assumption of a single threshold quantizer: it is not known if such restriction may be relaxed though it appears to be the case from extensive numerical examples.

7.1.2 Finite sample regime

For the finite sample regime, we consider a Bayesian approach where the priors for the two hypotheses are assumed to be π_0 and π_1 respectively. We derive quantizer properties for minimum error probability with both two-sided and one-sided compression, with the latter referring to the situation in which the fusion center has full data from one sensor while compressed data from another. This situation arises naturally in the case where one of the sensors is tasked with the final decision making.

For the finite sample regime, we adopt the person-by-person optimal approach and obtain the following result for two-sided compression, following standard approach described in [81].

Proposition 15 *For the distributed testing of independence problem with one bit quantization defined above. If we further assume the fusion rule satisfies,*

$$\begin{aligned} P(U_0=1|U=1,V=j) &\geq P(U_0=1|U=0,V=j), \\ P(U_0=0|U=0,V=j) &\geq P(U_0=0|U=1,V=j), \end{aligned}$$

for all $j=\{0,1\}$, then the optimal local decision rule at ith sensor is given by:

$$P(U_i=1|x_i) = \begin{cases} 1 & if\ \frac{\int_{x_{\bar{i}}} B_i P(x_{\bar{i}}|x_i H_1) dx_{\bar{i}}}{\int_{x_{\bar{i}}} A_i P(x_{\bar{i}}|x_i H_0) dx_{\bar{i}}} \geq \frac{\pi_0}{\pi_1} \\ 0 & otherwise \end{cases} \tag{7.8}$$

where $\bar{i} = 3 - i$, *for* $i = 1, 2$, *(hence,* $x_{\bar{1}} = Y$*)*, $\pi_0 = P(H_0)$, $\pi_1 = P(H_1)$, *and* A_i, B_i, $i = 1, 2$, *are given by*

$$A_i = \sum_{j=0}^{1}[P(U_0 = 1|U_i = 1, U_{\bar{i}} = j) - P(U_0 = 1|U_i = 0, U_{\bar{i}} = j)]P(U_{\bar{i}} = j|x_{\bar{i}}), \quad (7.9)$$

$$B_i = \sum_{j=0}^{1}[P(U_0 = 0|U_i = 0, U_{\bar{i}} = j) - P(U_0 = 0|U_i = 1, U_{\bar{i}} = j)]P(U_{\bar{i}} = j|x_{\bar{i}}). \quad (7.10)$$

If furthermore the fusion center uses the AND rule, we have

Proposition 16 *For the distributed test of independence problem with one bit quantization defined above, if we assume further that AND rule is used at the fusion center, i.e.,* $U_0 = 1$ *if and only if* $U = V = 1$, *then the optimal local decision rule is given by:*

$$P(U_i = 1|x_i) = \begin{cases} 1 & if \ \frac{\int_{D_{\bar{i}}} P(x_{\bar{i}}|x_i H_1) dx_{\bar{i}}}{\int_{D_{\bar{i}}} P(x_{\bar{i}}|x_i H_0) dx_{\bar{i}}} \geq \frac{\pi_0}{\pi_1} \\ 0 & otherwise \end{cases} \quad (7.11)$$

where $D_i = \{x_i : P(U_i = 1|x_i) = 1\}$ *is the rejection region for hypothesis* H_0 *at ith local sensor.*

For the case of one sided hypothesis testing of independence, e.g., $H_0 : \rho > 0$ versus $H_1 : \rho = 0$, we have the following corollary.

Corollary 7 *For the distributed one sided hypothesis testing of independence problem with one bit quantization defined above, single semi-infinite intervals for* D_1 *and* D_2 *form a PBPO solution for minimum probability of error.*

The fact that optimal quantizer has semi-infinite quantization intervals is rather appealing as it allows efficient search of a single threshold for quantizer design. Proofs of Propositions 15 and 16 are sketched in Appendix J. Corollary 7 is proved in Appendix L.

7.2 Numerical examples

Fig. 7.2 plots $I(U; V)$ as a function of thresholds t_1 and t_2 for $\rho = 0.65$. Apparently $I(U; V)$ achieves its maximum (≈ 0.15) when $(t_1, t_2) = (0, 0)$. We further conjecture that, this point is actually a global maximum which is corroborated by extensive

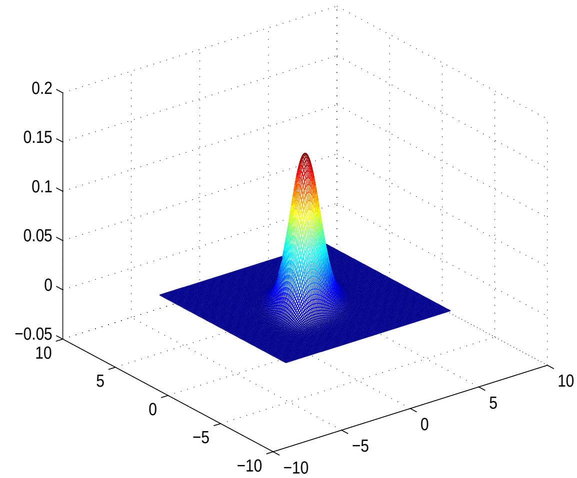

Figure 7.2: Plot of $I(U;V)$ as a function of thresholds t_1 and t_2 for $\rho = 0.65$.

numerical results. The difficulty in proving it's global maximum is that we do not have an analytical expression of bivariate normal cumulative function in general.

An interesting example of applying our main result is the spectrum sensing problem in cognitive radio network, where a secondary user tries to detect whether the primary user is present or not, i.e.,

$$\begin{cases} H_0: & X \neq 0, \\ H_1: & X = 0. \end{cases} \tag{7.12}$$

While the problem is well understood when the primary user's signal is fully observed (maybe corrupted by noise), it is more challenging when only a finite bits of information can be received. In this example, we study this spectrum sensing problem in a distributed fashion. Consider the following model, local sensor Y_1 and Y_2 receive noisy version of the original signal through independent additive Gaussian channels.

$$Y_1^n = X^n + N_1^n, \tag{7.13}$$

$$Y_2^n = X^n + N_2^n, \tag{7.14}$$

where X^n is a n length sample of the original random process, either present or not, i.e., $X^n = [x_1, x_2, \cdots, x_n]$ or $X^n = [0, 0, \cdots, 0]$. In this example, we assume that X is a zero mean independent Gaussian random process with variance P. The noise N_1 and N_2 are independent standard Gaussian random variables.

After receiving Y_i^n, sensor i will send a binary decision vector u_i^n to the fusion center, the fusion center will then decide wether the original signal is present or not. Clearly, if X is present, with high probability, the received signals at local sensors Y_1 and Y_2 are correlated (In the simulation, we choose $P = 2.857$ to make sure that the correlation of Y_1 and Y_2 under H_0 is 0.65). Based on this property, we use the following decision rules, for $k = 1, 2$ and $i = 1, 2, \cdots, n$

$$u_{ki} = 1 \quad \text{if} \quad \frac{y_{ki}}{\sqrt{P+1}} > t. \tag{7.15}$$

The fusion center will first calculate $u_{0i} = 1$ if and only if $u_{1i} = u_{2i}$ for $i = 1, 2, \cdots, n$, and then make a final decision using the following rule,

$$u = 1 \quad \text{if} \quad \sum_{i=1}^{n} u_{0i} \geq t_0(n), \tag{7.16}$$

where $t_0(n)$ is chosen so that the probability of type I error $P_{e1} = 0.1$. Since, under H_0, $\sum_{i=1}^{n} u_{0i}$ is a binomial distribution with probability of success $p = Pr_0(u_{1i} = u_{2i})$, which can be easily calculated, hence, $t_0(n)$ can be easily calculated numerically for each n.

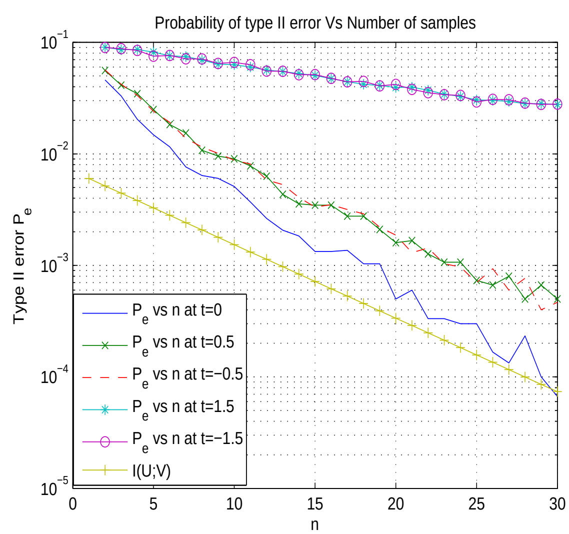

Figure 7.3: Probability of error for spectrum sensing.

Fig. 7.3 shows the performances of the above algorithm. In the simulation, we assume that $Pr(H_0) = 0.8$, and we choose five different local decision thresholds

($t = -1.5, -0.5, 0, 0.5, 1.5$) in (7.15). To compare the performance, we also plot the optimal error exponent $I(U;V)$ as plotted in Fig. 7.2. As we can see from the figure, even with 1 bit of information, we can still achieve a relatively low probability of error. We also observe that as number of samples increases, the probability of error decreases, and the threshold $t = 0$ performs the best among others.

7.3 Conclusion

In this chapter, we studied distributed test of independence of bivariate Gaussian sources with communication constraints. In particular, with one bit quantization, we derived quantization rules for single threshold quantizer at the local sensors that optimize the error exponent. For distributed one sided independence testwe proved that semi-infinite interval quantizers form a person by person optimal (PBPO) solution for minimum probability of error.

Chapter 8

Conclusion and future work

This dissertation deals with various aspects of correlated observations in distributed systems, including transmitting correlated sources over interference channels, characterizing the common information of random variables that carries meaningful operational interpretation, and testing the presence of data dependence in a distributed network.

First, a sufficient condition for the lossless transmission of a pair of correlated sources over a DMIC was given. By exploring the correlated source structure, a coding scheme that uses the random source partition as well as the correlation preserving codeword generation was introduced. The proposed coding scheme was proved to be optimal in a class of deterministic interference channels. The lossy counterpart was first studied by considering the special case of transmitting correlated Gaussian sources over a Gaussian Z interference channel (GZIC). Lower bounds on the distortion as well as several achievable schemes including uncoded transmission, separated source channel coding, and hybrid digital analog coding were proposed. We then studied the general case of lossy transmission of two arbitrarily correlated sources over a DMIC, as well as sending correlated Gaussian source over a Gaussian interference channel (GIC).

The measure of common information of correlated random variables was also studied in this thesis. A generalization of Wyner's definition of common information to N random variables was given, and an operational meaning was provided. Moreover, a monotone property of Wyner's common information was given, which is in contrast to other notions of the common information, specifically Shannon's mutual information and Gács and Körner's common randomness. Later, a generalization of Wyner's common information for continuous random variables is provided. An operational meaning for such generalization using the Gray-Wyner network with *lossy* source coding was provided. It was shown that Wyner's common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds.

Finally, the problem of distributed testing of independence was studied. A necessary condition for the optimal scalar quantizer is derived where the optimality is in the sense of optimizing the error exponent. The optimal quantizer properties for the finite sample regime was also provided.

For future work, some possible directions are summarized as follows.

- The proposed coding scheme in Chapter 2 is shown to be optimal in a class of deterministic interference channels. It will be interesting to see whether such optimality is still valid for more general cases. A possible direction is to provide a necessary condition to quantify the gap between the proposed sufficient condition and the necessary condition, and to see under which conditions there two coincide with each other.

- The hybrid digital analog coding scheme proposed in Chapter 3, though combining both the uncoded scheme and digital transmission, does not seems to dominate those two by simulation. The reason is that, the hybrid scheme in Chapter 3 restricts sender 1 to implementing an uncoded transmission with full power and sender 2 to a digital coding scheme to facilitate the transmission. A possible extension is hence to let both transmitters do a hybrid digital analog transmission. Intuitively, this scheme will include the uncoded transmission and separated transmission as special cases.

- For the general case of lossy transmission of correlated sources over DMIC studied in Chapter 4. The proposed scheme does reduce to Salehi and Kurtas's result for the lossless transmission. However, it can not reduce to the result studied in Chapter 2. The sufficient conditions in Chapter 2 includes Han and Kobayashi achievable region for ICs when the sources are independent. The technical reason behind this observation is that the proposed coding scheme uses the idea of random source partition that allows the other receiver to decode part of the interference, hence mimic the idea of partial interference cancelation as proposed in Han an Kobayashi's coding scheme. The coding scheme proposed in Chapter 4 does not have such source partition. A possible direction is thus to propose a more general coding scheme that will include the sufficient condition in Chapter 2 as a special case.

- In Chapter 5, we generalized Wyner's common information to N random variables. It is interesting to observe that Wyner's common information is a non-decreasing function of the number of random variables. Moreover, the common information eventually converges and the asymptote suggests that the notion of common information may have potential application and in certain inference problems. The circular symmetric binary source example in Chapter 5 as well as the Gaussian source example in Chapter 6 both indicate such relationship between the common information and inferencing about the hidden source. What is common between the two examples is that the sources can be decomposed into a simple Bayesian graph model that shares a common source, either through a binary symmetric channel or an additive Gaussian channel. Such source decomposition may provide a way to quantify the common information as well as a structure for source inferencing based on correlated observations. A quantitive measure of such relationship is hence worthwhile to pursue for this kind of sources or even more general classes of sources.

- For the distributed testing of independence studied in Chapter7, a locally optimal quantization rule is provided, although simulation results showed that such optimality might be global. A proof for the global optimality is worthwhile to pursue in the future. Moreover, the model considered in Chapter7 is a two sided test in the sense that, both transmitters need to quantize its observation. Another situation that one of the observation is fully available at the fusion center is also important both in theory and practical. Theoretically, the optimal error exponent for this case is known, and practically, this models the situation where the fusion center has its own measurement about the phenomenon. The optimal scale quantization rule for this model is another possible direction for future research.

Appendix A

Proof of the sufficient condition: Theorem 1

Theorem 1 can be obtained via Fourier-Motzkin elimination from the following constraints.

$$
\begin{aligned}
H(S|K) - r_1 &< I(SX_1; Y_1|W_0W_1W_2K), && \text{(A.1)}\\
H(S|K) &< I(SX_1; Y_1|W_0W_2K), && \text{(A.2)}\\
H(S|K) - r_1 + r_2 &< I(SW_2X_1; Y_1|W_0W_1K), && \text{(A.3)}\\
H(S|K) + r_2 &< I(W_2SX_1; Y_1|W_0K), && \text{(A.4)}\\
H(S) + r_2 &< I(W_0W_2SX_1; Y_1), && \text{(A.5)}\\
H(T|K) - r_2 &< I(TX_2; Y_2|W_0W_1W_2K), && \text{(A.6)}\\
H(T|K) &< I(TX_2; Y_2|W_0W_1K), && \text{(A.7)}\\
H(T|K) - r_2 + r_1 &< I(TW_1X_2; Y_2|W_0W_2K), && \text{(A.8)}\\
H(T|K) + r_1 &< I(W_1TX_2; Y_2|W_0K), && \text{(A.9)}\\
H(T) + r_1 &< I(W_0W_1TX_2; Y_2), && \text{(A.10)}\\
r_1 &\geq 0, && \text{(A.11)}\\
r_2 &\geq 0. && \text{(A.12)}
\end{aligned}
$$

Therefore, it suffices to prove, for decoder 1, that equations (A.1-A.5) constitute a sufficient condition. As sketched in Section III, the coding scheme utilizes Cover-El Gamal-Salehi's [2] correlation preserving coding and also Han and Costa's [3] random source partition.

a) *Random partition of the source sequences:* Let $r_1 \geq 0, r_2 \geq 0$ be any nonnegative real numbers. Randomly place source sequences $S^n \in \mathcal{S}^n$ into 2^{nr_1} cells and denote the cell index for a given s^n by $\alpha = l_1(s^n) \in I_1 = \{1, 2, \cdots, 2^{nr_1}\}$. Similarly, randomly place each $T^n \in \mathcal{T}^n$ into 2^{nr_2} cells and denote the cell index for a given t^n by $\beta = l_2(t^n) \in I_2 = \{1, 2, \cdots, 2^{nr_2}\}$. For this random source partition, we have the following lemma as in [3].

Lemma 4 *Let $\mathcal{S}_0, \mathcal{T}_0$ be any subset of $\mathcal{S}^n$ and $\mathcal{T}^n$, respectively. Then for any $\alpha \in I_1$ and $\beta \in I_2$, we have,*

$$\mathcal{E}(|\{s^n \in \mathcal{S}_0\} : l_1(s^n) = \alpha|) = |\mathcal{S}_0| \times 2^{-nr_1}, \tag{A.13}$$

$$\mathcal{E}(|\{t^n \in \mathcal{T}_0\} : l_2(t^n) = \beta|) = |\mathcal{T}_0| \times 2^{-nr_2}, \tag{A.14}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation.

b) *Codebook generation:* For any given joint distribution defined in (2.9), we first calculate the following distributions: $p(w_0), p(w_1|w_0)$, $p(w_2|w_0)$, $p(x_1|w_0w_1s)$ and $p(x_2|w_0w_2t)$.

For each $k^n \in \mathcal{K}^n$, independently generate one w_0^n sequence according to $\prod_{i=1}^n p(w_{0i})$. Index them by $w_0^n(k^n)$. For each source sequence s^n, find its cell index α and the corresponding auxiliary sequence $w_0^n(f(s^n))$, and independently generate *one* codeword w_1^n according to $\prod_{i=1}^n p(w_{1i}|w_{0i})$. Index them by $w_1^n(\alpha, w_0^n)$. Next, for each s^n, find the corresponding $w_0^n(f(s^n))$ and $w_1^n(\alpha, w_0^n)$, independently generate one codeword x_1^n according to $\prod_{i=1}^n p(x_{1i}|w_{0i}w_{1i}s_i)$. Index them by $x_1^n(s^n, w_0^n, w_1^n)$. Similarly, generate codewords $w_2^n(\beta, w_0^n)$ and $x_2^n(t^n, w_0^n, w_2^n)$ for user 2.

Notice that, for user 1, there are three sets of codewords: $w_0^n(k^n), w_1^n(\alpha, w_0^n)$ and $x_1^n(s^n, w_0^n, w_1^n)$. Here, $w_0^n(k^n)$ carries the information corresponding to the common part of the sources; $w_1^n(\alpha, w_0^n)$ carries the information of the cell index of the source s^n which is superimposed on top of w_0^n; $x_1^n(s^n, w_0^n, w_1^n)$ carries the private information of the source s^n and is superimposed on top of both w_0^n and w_1^n. The codebook structure for user 2 is similar.

c) *Encoding:* Upon observing s^n, encoder 1 finds the cell index $\alpha = l_1(s^n)$, codewords $w_0^n(f(s^n))$ and $w_1^n(\alpha, w_0^n)$, and then sends the corresponding $x_1^n(s^n, w_0^n, w_1^n)$. Similarly, encoder 2 sends $x_2^n(t^n, w_0^n, w_2^n)$.

d) *Decoding:* Upon receiving y_1^n, decoder 1 declares $\hat{s}^n = s^n$ to be the transmitted source sequence if s^n is the unique sequence such that

$$(s^n, k^n, w_0^n, w_1^n, x_1^n, w_2^n, y_1^n) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1). \tag{A.15}$$

Similarly, decoder 2 finds the unique t^n such that

$$(t^n, k^n, w_0^n, w_2^n, x_2^n, w_1^n, y_2^n) \in T_\epsilon^n(TKW_0W_2X_2W_1Y_2). \tag{A.16}$$

e) *Error analysis:* Encoding error occurs only if the source sequence pairs $(s^n, t^n) \notin T_\epsilon^n(ST)$ whose probability is bounded by ϵ from the asymptotic equipartition property (AEP) [35]. By symmetry, we only need to consider decoding errors at receiver 1.

Suppose $(s_0^n, t_0^n) \in T_\epsilon^n$ are the source outputs, with $k_0^n = f(s_0^n) = g(t_0^n)$. Without loss of generality, we assume that $\alpha = 1$ and $\beta = 1$, i.e., $l_1(s_0^n) = 1$ and $l_2(t_0^n) = 1$, and also $w_0^n(k_0^n) = w_{00}^n$. Denote $\mathcal{L}_1(1) = \{s^n : l_1(s^n) = 1\}$, $\mathcal{L}_2(1) = \{t^n : l_2(t^n) = 1\}$. Then an error occurs if any one of the following events happens:

1. $E_{11} : \Big(s_0^n, k_0^n, w_0^n(k_0^n), w_1^n(1, w_0^n), x_1^n(s_0^n, w_0^n, w_1^n), w_2^n(1, w_0^n), y_1^n\Big) \notin T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

2. E_{12}: there exists some $s^n \neq s_0^n$ in the cell $\alpha = 1$, i.e., $l_1(s^n) = 1$, and $\beta = 1$, such that, $\Big(s^n, k_0^n, w_0^n(k_0^n), w_1^n(1, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(1, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

3. E_{13}: there exists some $s^n \neq s_0^n$ in the cell $\alpha = 1$, i.e., $l_1(s^n) = 1$, and $\beta \neq 1$, such that,
$\Big(s^n, k_0^n, w_0^n(k_0^n), w_1^n(1, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(\beta, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

4. E_{14}: there exists some $s^n \neq s_0^n$ in the cell $\alpha \neq 1$, i.e., $l_1(s^n) = \alpha$, and $\beta = 1$, such that,
$\Big(s^n, k_0^n, w_0^n(k_0^n), w_1^n(\alpha, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(1, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

5. E_{15}: there exists some $s^n \neq s_0^n$ in the cell $\alpha \neq 1$, i.e., $l_1(s^n) = \alpha$, and $\beta \neq 1$, such that,
$\Big(s^n, k_0^n, w_0^n(k_0^n), w_1^n(\alpha, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(\beta, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

6. E_{16}: there exists some $s^n \neq s_0^n$ in the cell $\alpha \neq 1$, i.e., $l_1(s^n) = \alpha$, and $\beta \neq 1$, such that,
$k^n = f(s^n) \neq k_0^n$, $w_0^n(k^n) \neq w_0^n(k_0^n)$
and $\Big(s^n, k^n, w_0^n(k^n), w_1^n(\alpha, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(\beta, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

7. E_{17}: there exists some $s^n \neq s_0^n$ in the cell $\alpha \neq 1$, i.e., $l_1(s^n) = \alpha$, and $\beta \neq 1$, such that,
$k^n = f(s^n) \neq k_0^n$, $w_0^n(k^n) = w_0^n(k_0^n)$
and $\Big(s^n, k^n, w_0^n(k^n), w_1^n(\alpha, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(\beta, w_0^n), y_1^n\Big) \in T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$.

Hence, the probability of error at decoder 1 is

$$P_{e1} = Pr\{\cup_{i=1}^{7} E_{1i}\} \leq \sum_{i=1}^{7} Pr\{E_{1i}\}. \tag{A.17}$$

We evaluate the probabilities of the seven error events individually. First, by the AEP,

$$Pr\{E_{11}\} \leq \epsilon, \tag{A.18}$$

for sufficiently large n. For the second event, we have,[1]

$$\begin{aligned}
& Pr\{E_{12}\} \\
= \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} Pr\Big((s^n, k_0^n, w_0^n(k_0^n), w_1^n(1, w_0^n), x_1^n(s^n, w_0^n, w_1^n), w_2^n(1, w_0^n), y_1^n) \in T_\epsilon^n\Big), \\
= \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} \sum_{(s^n, k^n, w_0^n, w_1^n, x_1^n, w_2^n, y_1^n) \in T_\epsilon^n} p(k^n w_0^n w_1^n w_2^n y_1^n) p(s^n) p(x_1^n | s^n w_1^n w_0^n)\Big), \\
\leq \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} |T_\epsilon^n(SKW_0W_1X_1W_2Y_1)| 2^{-n(H(KW_0W_1W_2Y_1)} 2^{-n(H(S)+H(X_1|SW_1W_0)-3\epsilon)}\Big), \\
\leq \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} 2^{n(H(SKW_0W_1X_1W_2Y_1)-H(KW_0W_1W_2Y_1))} 2^{-n(H(S)+H(X_1|SW_1W_0)-4\epsilon)}\Big), \\
= \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} 2^{n(H(S|KW_0W_1W_2)+H(X_1Y_1|SKW_0W_1W_2))}, \\
& \cdot 2^{-n(H(Y_1|KW_0W_1W_2)+H(S)+H(X_1|SW_1W_0)-4\epsilon)}\Big), \\
\stackrel{(a)}{=} \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} 2^{n(H(S|K)+H(Y_1|SKW_0W_1W_2X_1))} 2^{-n(H(S)+H(Y_1|KW_0W_1W_2)-4\epsilon)}\Big), \\
\leq \ & 2^{n(H(S)+\epsilon-r_1)} 2^{-n(I(SX_1;Y_1|KW_0W_1W_2)+I(S;K))+4\epsilon)}, \\
= \ & 2^{n(H(S|K)-r_1-I(SX_1;Y_1|KW_0W_1W_2)+5\epsilon)},
\end{aligned} \tag{A.19}$$

where (a) is because $W_0W_1W_2 \to K \to S$ and $W_2 \to SW_0W_1 \to X_1$ form Markov chains and K is a deterministic function of S.

For the third event, we have,

$$\begin{aligned}
& Pr\{E_{13}\} \\
= \ & \mathcal{E}\Big(\sum_{s^n \in \mathcal{L}_1(1) \cap T_\epsilon^n(S)} \sum_{\beta \neq 1} Pr((s^n, k_0^n, w_0^n, w_1^n(1, w_0^n), x_1^n(s^n, w_1^n, w_0^n), w_2^n(\beta, w_0^n), y_1^n) \in T_\epsilon^n)\Big),
\end{aligned}$$

[1] For notational ease, we use T_ϵ^n to denote the typical set $T_\epsilon^n(SKW_0W_1X_1W_2Y_1)$ below.

$$
\begin{aligned}
&= \mathcal{E}\Big(\sum_{s^n\in\mathcal{L}_1(1)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{(s^n,k^n,w_0^n,w_1^n,x_1^n,w_2^n,y_1^n)\in T_\epsilon^n} p(s^n)p(w_2^n|w_0^n)p(x_1^n|s^nw_0^nw_1^n)p(k^nw_0^nw_1^ny_1^n)\Big),\\
&\leq \mathcal{E}\Big(\sum_{s^n\in\mathcal{L}_1(1)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1} 2^{n(H(SKW_0W_1W_2X_1Y_1)}2^{-n(H(S)+H(W_2|W_0)+H(X_1|W_1W_0S)+H(KW_0W_1Y_1)-5\epsilon)},\\
&= \mathcal{E}\Big(\sum_{s^n\in\mathcal{L}_1(1)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1} 2^{n(H(S|KW_0W_1)+H(W_2X_1Y_1|SKW_0W_1))},\\
&\quad \cdot 2^{-n(H(S)+H(W_2|W_0)+H(X_1|W_1W_0S)+H(Y_1|KW_0W_1)-5\epsilon)}\Big),\\
&\overset{(a)}{=} \mathcal{E}\Big(\sum_{s^n\in\mathcal{L}_1(1)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1} 2^{-n(I(S;K)+I(SW_2X_1;Y_1|KW_0W_1)-5\epsilon)}\Big)\\
&\leq 2^{n(H(S)+\epsilon-r_1+r_2)}2^{-n(I(S;K)+I(SW_2X_1;Y_1|KW_0W_1)-5\epsilon)},\\
&= 2^{n(H(S|K)-r_1+r_2-I(SW_2X_1;Y_1|KW_0W_1)+6\epsilon)},
\end{aligned}
\tag{A.20}
$$

where (a) is because $SW_1 \to W_0 \to W_2$ and $W_2 \to SW_0W_1 \to X_1$ form Markov chains and K is a deterministic function of S.

For the fourth event, we have,

$$
\begin{aligned}
&Pr\{E_{14}\}\\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)} Pr((s^n,k_0^n,w_0^n,w_1^n(\alpha,w_0^n),x_1^n(s^n,w_1^n,w_0^n),w_2^n(1,w_0^n),y_1^n)\in T_\epsilon^n)\Big),\\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{(s^n,k^n,w_0^n,w_1^n,x_1^n,w_2^n,y_1^n)\in T_\epsilon^n} p(s^n)p(w_1^n|w_0^n)p(x_1^n|s^nw_0^nw_1^n)p(w_0^n)\\
&\quad \cdot p(k^nw_2^ny_1^n|w_0^n)\Big),\\
&\leq 2^{n(H(S)+\epsilon)}2^{n(H(SKW_0W_1W_2X_1Y_1)}2^{-n(H(SW_0W_1X_1)+H(KW_2Y_1|W_0)+6\epsilon)},\\
&= 2^{n(H(S)-I(SW_1X_1;KW_2Y_1|W_0)+7\epsilon)},\\
&= 2^{n(H(S)-I(SW_1X_1;W_2|W_0)-I(SX_1;Y_1K|W_0W_2))}2^{-n(I(W_1;Y_1|W_0W_2X_1S)-7\epsilon)},\\
&\overset{(a)}{=} 2^{n(H(S)-I(SX_1;Y_1K|W_0W_2)+7\epsilon)},\\
&= 2^{n(H(S)-I(S;K|W_0W_2)-I(X_1;K|W_0W_2S))}2^{-n(I(SX_1;Y_1|KW_0W_2)-7\epsilon)},\\
&\overset{(b)}{=} 2^{n(H(S|K)-I(SX_1;Y_1|KW_0W_2)+7\epsilon)},
\end{aligned}
\tag{A.21}
$$

where (a) is because $W_2 \to W_0 \to SW_1X_1$ and $W_1 \to SW_0W_2X_1 \to Y_1$ form Markov chains; (b) is because S is independent of W_0W_2 and $W_0W_2 \to K \to S$ forms a Markov chain and also K is a deterministic function of S.

For the fifth event, we have,

$$
\begin{aligned}
& Pr\{E_{15}\} \\
= \;& \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1} Pr((s^n,k_0^n,w_0^n,w_1^n(\alpha,w_0^n),x_1^n(s^n,w_1^n,w_0^n),w_2^n(\beta,w_0^n),y_1^n)\in T_\epsilon^n)\Big), \\
= \;& \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{(s^n,k^n,w_0^n,w_1^n,x_1^n,w_2^n,y_1^n)\in T_\epsilon^n} p(s^n)p(w_1^n|w_0^n) \\
& \cdot p(w_2^n|w_0^n)p(x_1^n|s^nw_0^nw_1^n)p(k^nw_0^ny_1^n)\Big), \\
\leq \;& 2^{n(H(S)+r_2+\epsilon)}2^{n(H(SKW_0W_1W_2X_1Y_1)-H(KW_0Y_1))}2^{-n(H(S)+H(W_1W_2|W_0)+H(X_1|SW_0W_1)-6\epsilon)}, \\
= \;& 2^{n(H(S)+r_2+H(SKX_1Y_1|W_0W_1W_2))}2^{-n(H(KY_1|W_0)+H(S)+H(X_1|SW_0W_1)-7\epsilon)}, \\
\stackrel{(a)}{=} \;& 2^{n(H(S)+r_2+H(SX_1Y_1|KW_0W_1W_2))}2^{-n(H(Y_1|KW_0)+H(S)+H(X_1|SW_0W_1)-7\epsilon)}, \\
\stackrel{(b)}{=} \;& 2^{n(H(S)+r_2+H(S|K)+H(X_1Y_1|SKW_0W_1W_2))}2^{-n(H(S)+H(Y_1|KW_0)+H(X_1|SW_0W_1)+7\epsilon)}, \\
\stackrel{(c)}{=} \;& 2^{n(H(S|K)+r_2-I(SW_1W_2X_1;Y_1|KW_0)+7\epsilon)}, \\
= \;& 2^{n(H(S|K)+r_2-I(W_2SX_1;Y_1|KW_0))}2^{-n(I(W_1;Y_1|SKW_0W_2X_1)-7\epsilon)}, \\
\stackrel{(d)}{=} \;& 2^{n(H(S|K)+r_2-I(W_2SX_1;Y_1|KW_0)+7\epsilon)},
\end{aligned}
\tag{A.22}
$$

where (a) is because K is independent of $W_0W_1W_2$; (b) is because $W_0W_1W_2 \to K \to S$ forms a Markov chain; (c) is because $W_2 \to SW_0W_1 \to X_1$ forms a Markov chain; (d) is because $W_1 \to SW_0W_2X_1 \to Y_1$ forms a Markov chain.

For the sixth event, we have,

$$
\begin{aligned}
& Pr\{E_{16}\} \\
= \;& \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}, \\
& Pr(s^n,k^n,w_0^n(k^n),w_1^n(\alpha,w_0^n),x_1^n(s^n,w_1^n,w_0^n),w_2^n(\beta,w_0^n),y_1^n)\in T_\epsilon^n);w_0^n(k^n)\neq w_{00}^n)\Big), \\
= \;& \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{w_0^n\in\mathcal{W}_0^n} p(w_0^n=w_0'^n) \\
& \cdot Pr(s^n,k^n,w_0'^n(k^n),w_1^n(\alpha,w_0'^n),x_1^n(s^n,w_1^n,w_0'^n),w_2^n(\beta,w_0'^n),y_1^n)\in T_\epsilon^n)|w_0'^n\neq w_{00}^n)\Big), \\
= \;& \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{(s^n,k^n,w_0^n,w_1^n,x_1^n,w_2^n,y_1^n)\in T_\epsilon^n}\sum_{w_0^n\in T_\epsilon^n(W_0)} p(w_0'^n) \\
& p(s^nk^nw_0'^nw_1^nw_2^nx_1^n)p(y_1^n)\Big), \\
\leq \;& 2^{n(H(S)+r_2+\epsilon)}2^{n(H(W_0)+\epsilon)}2^{-n(H(W_0)+\epsilon)}2^{n(H(SKW_0W_1W_2X_1Y_1)-H(SKW_0W_1W_2X_1)-H(Y_1)+3\epsilon)} \\
= \;& 2^{n(H(S)+r_2-I(SKW_0W_1W_2X_1;Y_1)+4\epsilon)},
\end{aligned}
$$

$$
\begin{aligned}
&= 2^{n(H(S)+r_2-I(SW_0W_2X_1;Y_1)-I(W_1;Y_1|SW_0W_2X_1)+4\epsilon)}, \\
&\overset{(a)}{=} 2^{n(H(S)+r_2-I(SW_0W_2X_1;Y_1)+4\epsilon)},
\end{aligned} \tag{A.23}
$$

where (a) is because $W_1 \to SW_0W_2X_1 \to Y_1$ forms a Markov chain.

For the last event, we have,

$$
\begin{aligned}
& Pr\{E_{17}\} \\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1} Pr(s^n,k^n,w_0^n(k^n),w_1^n(\alpha,w_0^n),x_1^n(s^n,w_1^n,w_0^n), \\
&\quad w_2^n(\beta,w_0^n),y_1^n)\in T_\epsilon^n; w_0^n(k^n)=w_{00}^n)\Big), \\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{w_0'^n\in\mathcal{W}_0^n} p(w_0'^n)p(w_0'^n=w_{00}^n)Pr(s^n,k^n,w_0'^n(k^n),w_1^n(\alpha,w_0'^n), \\
&\quad x_1^n(s^n,w_1^n,w_0'^n),w_2^n(\beta,w_0'^n),y_1^n)\in T_\epsilon^n|w_0'^n=w_{00}^n)\Big), \\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{w_0'^n\in T_n^\epsilon(W_0)} p(w_0'^n)p(w_0'^n=w_{00}^n)Pr(s^n,k^n,w_0'^n(k^n),w_1^n(\alpha,w_0'^n), \\
&\quad x_1^n(s^n,w_1^n,w_0'^n),w_2^n(\beta,w_0'^n),y_1^n)\in T_\epsilon^n|w_0'^n=w_{00}^n)\Big), \\
&= \mathcal{E}\Big(\sum_{\alpha\neq 1}\sum_{s^n\in\mathcal{L}_1(\alpha)\cap T_\epsilon^n(S)}\sum_{\beta\neq 1}\sum_{(s^n,k^n,w_0^n,w_1^n,x_1^n,w_2^n,y_1^n)\in T_\epsilon^n}\sum_{w_0'^n\in T_\epsilon^n(W_0)} p(w_0'^n)p(w_0'^n=w_{00}^n) \\
&\quad p(s^nk^nw_1^nw_2^nx_1^n|w_0^n)p(w_0^ny_1^n)\Big), \\
&\leq 2^{n(H(S)+r_2+\epsilon)}2^{n(H(W_0)+\epsilon)}2^{-n(H(W_0-\epsilon))}2^{-n(H(W_0)-\epsilon)} \\
&\quad \cdot 2^{n(H(SKW_0W_1W_2X_1Y_1)-H(SKW_0W_1W_2X_1)-H(Y_1|W_0)+3\epsilon)}, \\
&= 2^{n(H(S)+r_2-H(W_0)-I(SKW_1W_2X_1;Y_1|W_0)+7\epsilon)}, \\
&\overset{(a)}{=} 2^{n(H(S)+r_2-H(W_0)-I(SW_2X_1;Y_1|W_0)+7\epsilon)},
\end{aligned} \tag{A.24}
$$

where (a) is because $W_1 \to SW_0W_2X_1 \to Y_1$ forms a Markov chain and K is a deterministic function of S.

From (A.19)-(A.24), if the following conditions are satisfied, then the probability of error at decoder 1 will vanish as n goes to infinity.

$$
H(S|K) - r_1 < I(SX_1;Y_1|W_0W_1W_2K), \tag{A.25}
$$

$$
H(S|K) < I(SX_1;Y_1|W_0W_2K), \tag{A.26}
$$

$$
H(S|K) - r_1 + r_2 < I(SW_2X_1;Y_1|W_0W_1K), \tag{A.27}
$$

$$
\begin{aligned}
H(S|K)+r_2 &< I(W_2SX_1;Y_1|W_0K), && \text{(A.28)}\\
H(S)+r_2 &< I(W_0W_2SX_1;Y_1), && \text{(A.29)}\\
H(S)+r_2 &< I(W_2SX_1;Y_1|W_0)+H(W_0). && \text{(A.30)}
\end{aligned}
$$

One can easily check that (A.30) is dominated by (A.29), since, $I(W_0;Y_1) \leq H(W_0)$.

This establishes (A.1-A.5). Similarly, (A.6-A.10) can be established symmetrically for decoder 2. The proof of Theorem 1 is complete by applying the Fourier-Motzkin elimination to (A.1-A.12).

Appendix B

Proof of the necessary and sufficient condition: Theorem 2

1. **Achievability**: follows from Corollary 2 by setting $W_0 = Q, W_1 = V_1, W_2 = V_2$ and choosing the joint distribution as

$$p(q, v_1, v_2, x_1, x_2) = p(q)p(v_1|q)p(v_2|q)p(x_1|q, v_1)p(x_1|q, v_2). \tag{B.1}$$

2. **Converse**: We first prove that for any non-deterministic encoders that achieve a decoding error P_e^* for transmitting the source through the channel, there exists a deterministic encoder which achieves a decoding error $P_e \leq P_e^*$. This can be proved following the lines of [32, Appendix D]. Next, we want to show that the inequalities (2.55-2.67) should be satisfied.

   Consider a deterministic encoder with $P_e \leq \epsilon$. Notice that this implies $P_{e1} \leq \epsilon$ and $P_{e2} \leq \epsilon$. By Fano's inequality, we have (for notation convenience, we denote $S^{'}$ as S_1 and $T^{'}$ as T_1 from now on),

$$H(K^n, S_1^n|Y_1^n) \leq n\epsilon. \tag{B.2}$$

   Hence, we get,

$$\begin{aligned} H(S_1^n|K_1^n, Y_1^n) &\leq n\epsilon, && \text{(B.3)}\\ H(T_1^n|K_1^n, Y_1^n) &\leq n\epsilon. && \text{(B.4)} \end{aligned}$$

   Notice that

$$H(Y_1^n, V_2^n|K^n, S_1^n) = H(Y_1^n, V_2^n|K^n, S_1^n, X_1^n), \tag{B.5}$$

and by expanding $H(Y_1^n, V_2^n | K^n, S_1^n, X_1^n)$ in two different ways, it can be easily shown that

$$H(V_2^n | X_1^n S_1^n K^n) = H(Y_1^n | X_1^n S_1^n K^n). \tag{B.6}$$

Hence,

$$\begin{aligned} H(V_2^n | S_1^n K^n) &= H(Y_1^n | S_1^n K^n), && \text{(B.7)} \\ H(V_2^n | K^n) &= H(Y_1^n | S_1^n K^n). && \text{(B.8)} \end{aligned}$$

Similarly, we have

$$H(V_1^n | K^n) = H(Y_2^n | T_1^n K^n). \tag{B.9}$$

Next, we want to show that

$$\begin{aligned} I(S_1^n; Y_1^n | K^n) &\leq I(S_1^n; Y_1^n V_1^n | V_2^n K^n), && \text{(B.10)} \\ I(T_1^n; Y_2^n | K^n) &\leq I(T_1^n; Y_2^n V_2^n | V_1^n K^n). && \text{(B.11)} \end{aligned}$$

Inequality (B.10) can be shown as follows,

$$\begin{aligned} & I(S_1^n; Y_1^n | K^n) \\ = \;& H(S_1^n | K^n) - H(S_1^n | Y_1^n K^n), && \text{(B.12)} \\ \leq \;& H(S_1^n | V_2^n K^n) - H(S_1^n | Y_1^n V_2^n K^n), && \text{(B.13)} \\ \leq \;& H(S_1^n | V_2^n K^n) - H(S_1^n | Y_1^n V_1^n V_2^n K^n), && \text{(B.14)} \\ = \;& I(S_1^n; Y_1^n V_1^n | V_2^n K^n). && \text{(B.15)} \end{aligned}$$

Similarly, we can obtain (B.11).

We now proceed to prove each of (2.55-2.67).

For inequality (2.55), we have,

$$\begin{aligned} H(S_1^n) &= H(S_1^n | K^n), && \text{(B.16)} \\ &= H(S_1^n | K^n V_2^n), && \text{(B.17)} \\ &= I(S_1^n; Y_1^n | K^n V_2^n) + H(S_1^n | Y_1^n K^n V_2^n), && \text{(B.18)} \\ &\leq I(S_1^n; Y_1^n | K^n V_2^n) + n\epsilon, && \text{(B.19)} \\ &= H(Y_1^n | K^n V_2^n) + n\epsilon, && \text{(B.20)} \\ &\leq \sum_{i=1}^{n} H(Y_{1i} | K^n V_{2i}) + n\epsilon. && \text{(B.21)} \end{aligned}$$

Similarly, for (2.56), we get

$$H(T_1^n) \le \sum_{i=1}^{n} H(Y_{2i}|K^n V_{1i}) + n\epsilon. \tag{B.22}$$

As for (2.57), we have,

$$\begin{aligned}
H(K^n) + H(S_1^n) &= H(K^n S_1^n), && \text{(B.23)}\\
&= I(K^n S_1^n; Y_1^n) + H(K^n S_1^n | Y_1^n), && \text{(B.24)}\\
&\le I(K^n S_1^n; Y_1^n + n\epsilon, && \text{(B.25)}\\
&\le H(Y_1^n) + n\epsilon, && \text{(B.26)}\\
&\le \sum_{i=1}^{n} H(Y_{1i}) + n\epsilon. && \text{(B.27)}
\end{aligned}$$

Similarly, we get

$$H(K^n) + H(T_1^n) \le \sum_{i=1}^{n} H(Y_{2i}) + n\epsilon. \tag{B.28}$$

For inequality (2.59), we have,

$$\begin{aligned}
& H(S_1^n) + H(T_1^n) \\
=\;& H(S_1^n|K^n) + H(T_1^n|K^n), && \text{(B.29)}\\
\le\;& I(S_1^n; Y_1^n|K^n) + I(T_1^n; Y_2^n|K^n) + 2n\epsilon, && \text{(B.30)}\\
=\;& H(Y_1^n|K^n) - H(Y_1^n|K^n S_1^n) + \\
& H(Y_2^n|K^n) - H(Y_2^n|K^n T_1^n) + 2n\epsilon, && \text{(B.31)}\\
=\;& H(Y_1^n|K^n) - H(V_2^n|K^n) + \\
& H(Y_2^n|K^n) - H(V_1^n|K^n) + 2n\epsilon, && \text{(B.32)}\\
\le\;& H(Y_1^n V_1^n|K^n) - H(V_2^n|K^n) + \\
& H(Y_2^n V_2^n|K^n) - H(V_1^n|K^n) + 2n\epsilon, && \text{(B.33)}\\
=\;& H(Y_1^n|V_1^n K^n) + H(Y_2^n|V_2^n K^n) + 2n\epsilon, && \text{(B.34)}\\
\le\;& \sum_{i=1}^{n} H(Y_{1i}|V_{1i}K^n) + H(Y_{2i}|V_{2i}K^n) + 2n\epsilon. && \text{(B.35)}
\end{aligned}$$

Regarding (2.60), we get,

$$H(S_1^n) + H(T_1^n)$$

$$
\begin{aligned}
&= H(S_1^n|K^n) + H(T_1^n|K^n), && \text{(B.36)}\\
&\leq I(S_1^n;Y_1^n|K^n) + I(T_1^n;Y_2^n|K^n) + 2n\epsilon, && \text{(B.37)}\\
&\stackrel{(a)}{\leq} I(S_1^n;Y_1^n|K^n) + I(T_1^n;Y_2^nV_2^n|K^nV_1^n) + 2n\epsilon, && \text{(B.38)}\\
&= H(Y_1^n|K^n) - H(Y_1^n|S_1^nK^n) + H(V_2^n|K^nV_1^n)\\
&\quad -H(V_2^n|K^nV_1^nT_1^n) + H(Y_2^n|K^nV_1^nV_2^n)\\
&\quad -H(Y_2^n|K^nV_1^nV_2^nT_1^n), && \text{(B.39)}\\
&\stackrel{(b)}{=} H(Y_1^n|K^n) + H(Y_2^n|V_1^nV_2^nK^n) + 2n\epsilon, && \text{(B.40)}\\
&\leq \sum_{i=1}^{n} H(Y_{1i}|K^n) + H(Y_{2i}|V_{1i}V_{2i}K^n) + 2n\epsilon, && \text{(B.41)}
\end{aligned}
$$

where (a) is because of (B.11) and (b) is because that $H(Y_1^n|S_1^nK^n) = H(V_2^n|V_1^nK^n)$, $H(V_2^n|V_1^nT_1^nK^n) = 0$ and $H(Y_2^n|K^nV_1^nV_2^nT_1^n) = 0$.

Similarly, we get,

$$
H(S_1^n) + H(T_1^n) \leq \sum_{i=1}^{n} H(Y_{2i}|K^n) + H(Y_{1i}|V_{1i}V_{2i}K^n) + 2n\epsilon. \qquad \text{(B.42)}
$$

For (2.62), we have,

$$
\begin{aligned}
& H(K^n) + H(S_1^n) + H(T_1^n)\\
&= H(K^nS_1^n) + H(T_1^n|K^n), && \text{(B.43)}\\
&\leq I(K^nS_1^n;Y_1^n) + I(T_1^n;Y_2^n|K^n) + 2n\epsilon, && \text{(B.44)}\\
&\stackrel{(a)}{\leq} I(K^nS_1^n;Y_1^n) + I(T_1^n;Y_2^nV_2^n|K^nV_1^n) + 2n\epsilon, && \text{(B.45)}\\
&= I(K^nS_1^n;Y_1^n) + I(T_1^n;V_2^n|K^nV_1^n)\\
&\quad +I(T_1^n;Y_2^n|K^nV_1^nV_2^n) + 2n\epsilon, && \text{(B.46)}\\
&= H(Y_1^n) - H(Y_1^n|K^nS_1^n) + H(V_2^n|V_1^nK^n)\\
&\quad -H(V_2^n|V_1^nK^nT_1^n) + H(Y_2^n|V_1^nV_2^nK^n)\\
&\quad -H(Y_2^n|V_1^nV_2^nK^nT_1^n) + 2n\epsilon, && \text{(B.47)}\\
&\stackrel{(b)}{=} H(Y_1^n) - H(Y_2^n|V_1^nV_2^nK^n) + 2n\epsilon, && \text{(B.48)}\\
&\leq \sum_{i=1}^{n} H(Y_{1i}) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}K^n) + 2n\epsilon, && \text{(B.49)}
\end{aligned}
$$

where (a) is because of (B.11) and (b) is because that $H(Y_1^n|S_1^nK^n) = H(V_2^n|V_1^nK^n)$, $H(V_2^n|V_1^nT_1^nK^n) = 0$ and $H(Y_2^n|K^nV_1^nV_2^nT_1^n) = 0$.

Similarly, we get,

$$\begin{aligned}
& H(K^n) + H(S_1^n) + H(T_1^n) \\
\leq \; & \sum_{i=1}^{n} [H(Y_{2i}) + H(Y_{1i}|V_{1i}V_{2i}K^n)] + 2n\epsilon. \qquad \text{(B.50)}
\end{aligned}$$

For (2.64), we get,

$$\begin{aligned}
& 2H(S_1^n) + H(T_1^n) \\
= \; & H(S_1^n|K^n) + H(S_1^n|K^n) + H(T_1^n|K^n), && \text{(B.51)} \\
\leq \; & I(S_1^n; Y_1^n|K^n) + I(S_1^n; Y_1^n|K^n) \\
& + I(T_1^n; Y_2^n|K^n) + 3n\epsilon, && \text{(B.52)} \\
\leq \; & I(S_1^n; Y_1^n|K^n) + I(S_1^n; Y_1^n V_1^n|V_2^n K^n) \\
& + I(T_1^n; Y_2^n|K^n) + 3n\epsilon, && \text{(B.53)} \\
= \; & I(S_1^n; Y_1^n|K^n) + I(S_1^n; V_1^n|V_2^n K^n) + \\
& I(S_1^n; Y_1^n|V_1^n V_2^n K^n) + I(T_1^n; Y_2^n|K^n) + 3n\epsilon, && \text{(B.54)} \\
= \; & H(Y_1^n|K^n) - H(Y_1^n|S_1^n K^n) + H(V_1^n|V_2^n K^n) \\
& - H(V_1^n|V_2^n K^n S_1^n) + H(Y_1^n|V_1^n V_2^n K^n) \\
& - H(Y_1^n|S_1^n V_1^n V_2^n K^n) + H(Y_2^n|K^n) \\
& - H(Y_2^n|K^n T_1^n) + 2n\epsilon, && \text{(B.55)} \\
\stackrel{(a)}{=} \; & H(Y_1^n|K^n) - H(Y_1^n|S_1^n K^n) + H(Y_1^n|V_1^n V_2^n K^n) \\
& + H(Y_2^n|K^n) + 3n\epsilon, && \text{(B.56)} \\
\stackrel{(b)}{=} \; & H(Y_1^n|K^n) - H(V_2^n|K^n) + H(Y_1^n|V_1^n V_2^n K^n) \\
& + H(Y_2^n|K^n) + 3n\epsilon, && \text{(B.57)} \\
\leq \; & H(Y_1^n|K^n) - H(V_2^n|K^n) + H(Y_1^n|V_1^n V_2^n K^n) \\
& + H(Y_2^n V_2^n|K^n) + 3n\epsilon, && \text{(B.58)} \\
= \; & H(Y_1^n|K^n) + H(Y_1^n|V_1^n V_2^n K^n) + H(Y_2^n|K^n V_2^n) \\
& + 3n\epsilon, && \text{(B.59)} \\
\leq \; & \sum_{i=1}^{n} H(Y_{1i}|K^n) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}V_{2i}K^n) \\
& + \sum_{i=1}^{n} H(Y_{2i}|V_{2i}K^n) + 3n\epsilon, && \text{(B.60)}
\end{aligned}$$

where (a) is because $H(V_1^n|V_2^nK^n) = H(V_1^n|K^n) = H(Y_2^n|K^nT_1^n)$, $H(V_1^n|V_2^nK^nS_1^n) = H(V_1^n|X_1^nV_2^nK^nS_1^n) = 0$ and $H(Y_1^n|V_1^nV_2^nS_1^nK^n) = H(Y_1^n|X_1^nV_1^nV_2^nS_1^nK^n) = 0$; (b) is from the fact that $H(V_2^n|K^n) = H(Y_1^n|K^nS_1^n)$.

Similarly,we have,

$$
\begin{aligned}
& H(S_1^n) + 2H(T_1^n) \\
\leq\ & \sum_{i=1}^{n} H(Y_{2i}|K^n) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}K^n) \\
& + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}K^n) + 3n\epsilon. \qquad \text{(B.61)}
\end{aligned}
$$

As for (2.66), we have,

$$
\begin{aligned}
& H(K^n) + 2H(S_1^n) + H(T_1^n) \\
=\ & H(S_1^nK^n) + H(S_1^n|K^n) + H(T_1^n|K^n), \qquad \text{(B.62)} \\
\leq\ & I(S_1^nK^n;Y_1^n) + I(S_1^n;Y_1^n|K^n) \\
\leq\ & I(S_1^nK^n;Y_1^n) + I(S_1^n;Y_1^nV_1^n|V_2^nK^n) \\
& +I(T_1^n;Y_2^n|K^n) + 3n\epsilon, \qquad \text{(B.63)} \\
=\ & I(S_1^nK^n;Y_1^n) + I(S_1^n;V_1^n|V_2^nK^n) + \\
& I(S_1^n;Y_1^n|V_1^nV_2^nK^n) + I(T_1^n;Y_2^n|K^n) + 3n\epsilon, \qquad \text{(B.64)} \\
=\ & H(Y_1^n) - H(Y_1^n|S_1^nK^n) + H(V_1^n|V_2^nK^n) \\
& -H(V_1^n|V_2^nK^nS_1^n) + H(Y_1^n|V_1^nV_2^nK^n) \\
& -H(Y_1^n|S_1^nV_1^nV_2^nK^n) + H(Y_2^n|K^n) \\
& -H(Y_2^n|K^nT_1^n) + 2n\epsilon, \qquad \text{(B.65)} \\
\overset{(a)}{\leq}\ & H(Y_1^n) + H(Y_1^n|V_1^nV_2^nK^n) + H(Y_2^n|K^nV_2^n) \\
& +3n\epsilon, \qquad \text{(B.66)} \\
\leq\ & \sum_{i=1}^{n} H(Y^n) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}V_{2i}K^n) \\
& + \sum_{i=1}^{n} H(Y_{2i}|V_{2i}K^n) + 3n\epsilon, \qquad \text{(B.67)}
\end{aligned}
$$

where (a) follows similar procedure as in proving (2.64).

Similarly, we have,

$$\begin{aligned}
& H(K^n) + 2H(S_1^n) + H(T_1^n) \\
\leq \; & \sum_{i=1}^{n} H(Y_{2i}) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}K^n) \\
& + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}K^n) + 3n\epsilon. \qquad \text{(B.68)}
\end{aligned}$$

Now introduce a random variable Q with $p(q_i) = p(k^n)$, i.e., q_i is an auxiliary random variable uniformly distributed over the set $\mathcal{K}^n = \{1, 2, \cdots, |\mathcal{K}|^n\}$. Noticing that $p(x_{1i}x_{2i}|q) = p(x_{1i}|k^n)p(x_{2i}|k^n)$, hence, we have, $p(x_{1i}x_{2i}|k^n) = p(x_{1i}|q)p(x_{2i}|q)$. By standard convexity argument, we completes the proof of converse.

Appendix C

Proof of the sufficient condition: Theorem 8

In this section, we give the proof of Theorem 8.

For any fixed joint distribution as in (4.7), we first calculate the distributions $p(w_1|q)$, $p(w_2|q)$, $p(u_1|q)$, $p(u_2|q)$, $p(x_1|w_1u_1q)$, and $p(x_2|w_2u_2q)$, and fix the decoding functions ϕ_1, ϕ_2 satisfying the distortion constraints as in (4.17).

Codebook generation: For each fixed q^n, let $R_1 = I(W_1;S|Q) + \epsilon$, and $R_2 = I(W_2;T|Q) + \epsilon$, first randomly generate 2^{nR_i} codewords W_i^n, $i = 1, 2$, independent and identically distributed (i.i.d.) according to $p(w_i|q)$. Next, randomly generate $2^{n(I(S;U_1|Q)+\epsilon)}$ codewords U_1^n, i.i.d. according to $p(u_1|q)$ and $2^{n(I(T;U_2|Q)+\epsilon)}$ codewords U_2^n, i.i.d. according to $p(u_2|q)$. For each source sequence s^n, note that (4.7) implies W_1, S, U_1 (given Q) forms a Markov chain in this order. By the Markov lemma [35], for n sufficiently large, there exist at least one pair of codewords (w_1^{n*}, u_1^{n*}) such that $(s^n, w_1^{n*}, u_1^{n*}) \in T_\epsilon^n(SW_1U_1)$. Similarly, for each source sequence t^n, there exist at least one pair of codewords (w_2^{n*}, u_2^{n*}) such that $(t^n, w_2^{n*}, u_2^{n*}) \in T_\epsilon^n(TW_2U_2)$. For each pair of (w_i^{n*}, u_i^{n*}), generate one codeword $x_i^n(w_i^{n*}, u_i^{n*}, q^n)$ according to distribution $p(x_i^n|w_i^n, u_i^n, q^n) = \prod_{j=1}^n p(x_{ij}|w_{ij}u_{ij}q_{ij})$, $i = 1, 2$.

Encoding: Upon observing s^n, encoder 1 transmits the corresponding codeword $x_1^n(w_1^{n*}, u_1^{n*}, q^n)$. Similarly, encoder 2 transmits the codeword $x_2^n(w_2^{n*}, u_2^{n*}, q^n)$.

Decoding: Upon receiving y_1^n, decoder 1 finds a unique set of codewords (w_1^n, u_1^n, u_2^n) such that $(w_1^n, u_1^n, u_2^n, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$. Similarly, decoder 2 finds a unique set of codewords (w_2^n, u_2^n, u_1^n) such that $(w_2^n, u_2^n, u_1^n, x_2^n, y_2^n, q^n) \in T_\epsilon^n(W_2U_2U_1X_2Y_2Q)$. If it fails, the decoder i, $i = 1, 2$, declares an error and chooses the a set of fixed codewords $(w_{i0}^n, u_{10}^n, u_{20}^n)$ as the decoder output, which leads a fixed bounded distortion

$d_0 = \max\{E[d_1(S, \phi_1(W_1U_1U_2))], E[d_2(T, \phi_2(W_2U_2U_1))]\}$ at each decoder.

Error analysis: By symmetry, we only consider the error events at decoder 1. An error occurs if one of the following events happens:

1. E_{11}: the true codewords transmitted satisfy that $(w_1^{n*}, u_1^{n*}, u_2^{n*}, x_1^n, y_1^n, q^n) \notin T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$.

2. E_{12}: there exists some codeword $w_1^n \neq w_1^{n*}$ such that $(w_1^n, u_1^{n*}, u_2^{n*}, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$.

3. E_{13}: there exist $(w_1^n, u_1^n) \neq (w_1^{n*}, u_1^{n*})$ such that $(w_1^n, u_1^n, u_2^{n*}, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$.

4. E_{14}: there exist $(w_1^n, u_2^n) \neq (w_1^{n*}, u_2^{n*})$ such that $(w_1^n, u_1^{n*}, u_2^n, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$.

5. E_{15}: there exist $(w_1^n, u_1^n, u_2^n) \neq (w_1^{n*}, u_1^{n*}, u_2^{n*})$ such that $(w_1^n, u_1^n, u_2^n, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)$.

Hence, the probability of error at decoder 1 is

$$P_{e1} = Pr\{\cup_{i=1}^5 E_{1i}\} \leq \sum_{i=1}^5 Pr\{E_{1i}\} \tag{C.1}$$

Next, we will evaluate the seven probability of errors individually. First, by the extended Markov Lemma [39,40], as $n \to \infty$, $Pr\{E_{11}\} \leq \epsilon$. For the second event, we have,

$$\begin{aligned}
& Pr\{E_{12}\} \\
= \;& Pr\Big((w_1^n, u_1^{n*}, u_2^{n*}, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)\Big) \\
= \;& \sum_{w_1^n} \sum_{(w_1^n, u_1^{n*}, u_2^{n*}, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)} \\
& p(w_1^n|q^n)p(x_1^n|w_1^n u_1^n q^n)p(u_1^n u_2^n y_1^n q^n) \\
\leq \;& \sum_{w_1^n} 2^{n(H(W_1U_1U_2X_1Y_1Q) - H(W_1|Q))} \\
& \cdot 2^{-n(H(X_1|W_1U_1Q) + H(U_1U_2Y_1Q) - 4\epsilon)} \\
= \;& 2^{n(R_1 - I(U_1U_2;W_1|Q) - I(W_1X_1;Y_1|U_1U_2Q) + 4\epsilon)} \\
= \;& 2^{n(I(W_1;S|U_1U_2Q) - I(W_1X_1;Y_1|U_1U_2Q) + 5\epsilon)}
\end{aligned} \tag{C.2}$$

where the last equality is because that given Q, $W_1 \to S \to (U_1, U_2)$ forms a Markov chain.

For the third event, we have,

$$\begin{aligned}
& Pr\{E_{13}\} \\
= \;& Pr\Big((w_1^n, u_1^n, u_2^{n*}, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)\Big) \\
= \;& \sum_{w_1^n}\sum_{u_1^n} \sum_{(w_1^n,u_1^n,u_2^{n*},x_1^n,y_1^n,q^n)\in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)} \\
& p(w_1^n|q^n)p(u_1^n|q^n)p(x_1^n|w_1^nu_1^nq^n)p(u_2^ny_1^nq^n) \\
\leq \;& \sum_{w_1^n}\sum_{u_1^n} 2^{n(H(W_1U_1U_2X_1Y_1Q)-H(W_1|Q))} \\
& \cdot 2^{-n(H(U_1|Q)+H(X_1|W_1U_1Q)+H(U_2Y_1Q)-5\epsilon)} \\
= \;& 2^{n(R_1+I(U_1;S|Q)-I(W_1;U_2|Q)-I(W_1U_2;U_1|Q))} \\
& \cdot 2^{-n(I(W_1U_1X_1;Y_1|U_2Q)-6\epsilon)} \\
= \;& 2^{n(I(W_1U_1;S|U_2Q)-I(W_1U_1X_1;Y_1|U_2Q)+6\epsilon)}
\end{aligned} \tag{C.3}$$

where the last equality is because that given Q, $W_1 \to S \to U_2$ and $U_1 \to S \to (W_1, U_2)$ form Markov chains.

For the fourth event, we have,

$$\begin{aligned}
& Pr\{E_{14}\} \\
= \;& Pr\Big((w_1^n, u_1^{n*}, u_2^n, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)\Big) \\
= \;& \sum_{w_1^n}\sum_{u_2^n} \sum_{(w_1^n,u_1^{n*},u_2^n,x_1^n,y_1^n,q^n)\in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)} \\
& p(w_1^n|q^n)p(u_2^n|q^n)p(x_1^n|w_1^nu_1^nq^n)p(u_1^ny_1^nq^n) \\
\leq \;& \sum_{w_1^n}\sum_{u_2^n} 2^{n(H(W_1U_1U_2X_1Y_1Q)-H(W_1|Q))} \\
& \cdot 2^{-n(H(U_2|Q)+H(X_1|W_1U_1Q)+H(U_1Y_1Q)-5\epsilon)} \\
= \;& 2^{n(R_1+I(U_2;T|Q)-I(U_1;U_2|Q))} \\
& \cdot 2^{-n(I(U_1U_2;W_1|Q)+I(W_1U_2X_1;Y_1|U_1Q)-6\epsilon)} \\
= \;& 2^{n(I(W_1;S|U_1U_2Q)+I(U_2;T|U_1Q)-I(W_1U_2X_1;Y_1|U_1Q)+7\epsilon)}
\end{aligned} \tag{C.4}$$

where the last equality is because that given Q, $W_1 \to S \to (U_1, U_2)$ and $U_2 \to T \to U_1$ form Markov chains.

For the last event, we have,

$$Pr\{E_{15}\}$$

$$
\begin{aligned}
&= Pr\Big((w_1^n, u_1^n, u_2^n, x_1^n, y_1^n, q^n) \in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)\Big) \\
&= \sum_{w_1^n}\sum_{u_1^n}\sum_{u_2^n}\sum_{(w_1^n,u_1^n,u_2^n,x_1^n,y_1^n,q^n)\in T_\epsilon^n(W_1U_1U_2X_1Y_1Q)} \\
&\quad p(w_1^n|q^n)p(u_1^n|q^n)p(u_2^n|q^n)p(x_1^n|w_1^nu_1^nq^n)p(y_1^nq^n) \\
&\leq \sum_{w_1^n}\sum_{u_1^n}\sum_{u_2^n} 2^{n(H(W_1U_1U_2X_1Y_1Q)-H(W_1|Q))} \\
&\quad \cdot 2^{-n(H(U_1|Q)+H(U_2|Q)+H(X_1|W_1U_1Q)+H(Y_1Q)-6\epsilon)} \\
&= 2^{n(R_1+I(U_1;S|Q)+I(U_2;T|Q)-I(W_1;U_2|Q))} \\
&\quad \cdot 2^{-n(I(W_1U_2;U_1|Q)+I(W_1U_2X_1;Y_1|Q)-8\epsilon)} \\
&= 2^{n(I(W_1U_1;S|U_2Q)+I(U_2;T|Q)-I(W_1U_1U_2X_1;Y_1|Q)+8\epsilon)} \qquad \text{(C.5)}
\end{aligned}
$$

From (C.2-C.5), if the following conditions are satisfied, then the probability of error at decoder 1 will vanish as n goes to infinity.

$$
\begin{aligned}
I(W_1;S|U_1U_2Q) &< I(W_1X_1;Y_1|U_1U_2Q), \\
I(W_1U_1;S|U_2Q) &< I(W_1U_1X_1;Y_1|U_2Q), \\
I(W_1;S|U_1U_2Q) &< I(W_1U_2X_1;Y_1|U_1Q) \\
&\quad -I(U_2;T|U_1Q), \\
I(W_1U_1;S|U_2Q) &< I(W_1U_1U_2X_1;Y_1|Q) \\
&\quad -I(U_2;T|Q). \qquad \text{(C.6)}
\end{aligned}
$$

Similarly, the following conditions can be obtained for decoder 2.

$$
\begin{aligned}
I(W_2;T|U_1U_2Q) &< I(W_2X_2;Y_2|U_1U_2Q), \qquad \text{(C.7)} \\
I(W_2U_2;T|U_1Q) &< I(W_2U_2X_2;Y_2|U_1Q) \qquad \text{(C.8)} \\
I(W_2;T|U_1U_2Q) &< I(W_2U_1X_2;Y_2|U_2Q) \\
&\quad -I(U_1;S|U_2Q) \qquad \text{(C.9)} \\
I(W_2U_2;T|U_1Q) &< I(W_2U_1U_2X_2;Y_2|Q) \\
&\quad -I(U_1;S|Q) \qquad \text{(C.10)}
\end{aligned}
$$

For the distortion part, upon correctly decoding as $n \to \infty$, decoder 1 finds the correct codewords $(w_1^{n*}, u_1^{n*}, u_2^{n*})$ which are jointly typical with s^n. Similarly, decoder 2 finds the correct codewords $(w_2^{n*}, u_1^{n*}, u_2^{n*})$ which are jointly typical with t^n. By the standard argument as in the rate distortion theory [65], we get the desired distortion (D_1, D_2), as $n \to \infty$. This completes the proof of Theorem 8. □

Appendix D

Proof of the main results: Theorem 10

First, as with [59], we define a quantity $\Gamma(\delta_1, \delta_2)$ which plays an important role in the proof.

Let $(X_1, X_2, \cdots, X_N) \sim P(x_1, x_2, \cdots, x_N)$ where $X_1, \cdots, X_N$ take values in finite alphabet $\mathcal{X}_1, \cdots, \mathcal{X}_N$. Let $(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N, W)$ be a $(N+1)$tuple of random variables where $\hat{X}_1 \in \mathcal{X}_1, \hat{X}_2 \in \mathcal{X}_2, \cdots, \hat{X}_N \in \mathcal{X}_N$ and $W \in \mathcal{W}$, a finite set. Denote the marginal distribution of $(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N)$ by

$$Q(x_1, x_2, \cdots, x_N) = Pr(\hat{X}_1 = x_1, \hat{X}_2 = x_2, \cdots, \hat{X}_N = x_n), \tag{D.1}$$

for $x_i \in \mathcal{X}_i$, $i = 1, 2, \cdots, N$.

For any $\delta_1, \delta_2 \geq 0$, define

$$\Gamma(\delta_1, \delta_2) = \sup H(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N | W), \tag{D.2}$$

where the sumpremum is taken over all $(N+1)$-tuples $(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N, W)$ that satisfy

$$D(P; Q) = \sum_{x,y} P(x_1, x_2, \cdots, x_N) \log \frac{P(x_1, x_2, \cdots, x_N)}{Q(x_1, x_2, \cdots, x_N)} \leq \delta_1, \tag{D.3}$$

and

$$\sum_{i=1}^{N} H(\hat{X}_i | W) - H(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N | W) \leq \delta_2. \tag{D.4}$$

It follows that $C(X_1, X_2, \cdots, X_N)$ as defined in Theorem 1, is equivalent to

$$C(X_1, X_2, \cdots, X_N) = H(X_1, X_2, \cdots, X_N) - \Gamma(0, 0). \tag{D.5}$$

The following lemma gives some properties of $\Gamma(\delta_1, \delta_2)$.

Lemma 5 *1) For all $\delta_1, \delta_2 \geq 0$, there exists a $(N+1)$-tuple $(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N, W)$ such that (D.3) and (D.4) are satisfied and*

$$\Gamma(\delta_1, \delta_2) = H(\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_N | W). \tag{D.6}$$

Moreover, for $\delta_1, \delta_2 = 0$,

$$|\mathcal{W}| \leq \prod_{i=1}^{N} |\mathcal{X}_i|. \tag{D.7}$$

2) $\Gamma(\delta_1, \delta_2)$ is a concave function of (δ_1, δ_2) and it is continuous for all $\delta_1, \delta_2 \geq 0$.

3) For $\delta \geq 0$, define $\Gamma_1(\delta) = \Gamma(0, \delta)$ and $\Gamma_2(\delta) = \Gamma(\delta, 0)$, then $\Gamma_1(\delta)$ and $\Gamma_2(\delta)$ are concave and continuous for $\delta \geq 0$.

The proof of Lemma 1 follows similarly as the proof of Theorem 4.4 in [59].

Appendix E

Proof of $C_1 = C$

In this section, we prove the first part of Theorem 1, that is $C_1 = C(X_1, X_2, \cdots, X_N)$. We first prove the converse part, that is for any R_0 that is achievable for the Gray-Wyner source coding network, we have,

Theorem 14 (Converse)

$$C_1 \geq C(X_1, X_2, \cdots, X_N). \tag{E.1}$$

To prove the converse, first let (f, g_i), $i = 1, 2, \cdots, N$ be any $(n, \mathcal{M}_0, \mathcal{M}_1, \cdots, \mathcal{M}_N)$ code that satisfies (5.7), (5.8) and (5.9).

Then, we have,

$$\begin{aligned}
& \log \mathcal{M}_0 \\
\geq\ & H(M_0), && \text{(E.2)}\\
\geq\ & I(X_1^n X_2^n \cdots X_N^n; M_0), && \text{(E.3)}\\
=\ & H(X_1^n X_2^n \cdots X_N^n) - H(X_1^n X_2^n \cdots X_N^n | M_0), && \text{(E.4)}\\
=\ & nH(X_1 X_2 \cdots X_N) - \sum_{j=1}^{n} H(X_{1j} X_{2j} \cdots X_{Nj} | W_j), && \text{(E.5)}
\end{aligned}$$

where $W_j \triangleq (M_0, X_1^{j-1}, X_2^{j-1}, \cdots, X_N^{j-1})$ and $X_i^{j-1} = (X_{i1}, X_{i2}, \cdots, X_{i,j-1})$ for $i = 1, 2, \cdots, N$.

Notice that, the $(N+1)$-tuple $(X_{1j}, X_{2j}, \cdots, X_{Nj}, W_j)$ satisfies condition (D.3) and (D.4) with $\delta_1 = 0$ and

$$\delta_2^{(j)} = \sum_{i=1}^{N} H(X_{i,j} | W_j) - H(X_{1j}, X_{2j}, \cdots, X_{Nj} | W_j). \tag{E.6}$$

Hence, by the definition of $\Gamma(\delta_1, \delta_2)$, we have

$$H(X_{1j}X_{2j}\cdots X_{Nj}|W_j) \leq \Gamma_1(\delta_2^{(j)}). \tag{E.7}$$

Substitute (E.7) into (E.5), we get,

$$\begin{aligned}
\log \mathcal{M}_0 &\geq nH(X_1X_2\cdots X_N) - \sum_{j=1}^{n} \Gamma_1(\delta_2^{(j)}), & \text{(E.8)}\\
&\geq nH(X_1X_2\cdots X_N) - n\Gamma_1(\frac{1}{n}\sum_{j=1}^{n}\delta_2^{(j)}). & \text{(E.9)}
\end{aligned}$$

where the last step is from the concavity of $\Gamma_1(\cdot)$ function. Now define

$$\eta = \frac{1}{n}\sum_{j=1}^{n}\delta_2^{(j)}. \tag{E.10}$$

The following lemma gives an upper bound on η.

Lemma 6 *For any $(n, \mathcal{M}_0, \mathcal{M}_1, \cdots, \mathcal{M}_N)$ code that satisfies (5.7), (5.8) and (5.9), we have*

$$\eta \leq (N+1)\epsilon. \tag{E.11}$$

Proof :

By Fano's inequality, we have, for $i = 1, 2, \cdots, N$,

$$H(X_i^n|M_0M_i) \leq n\epsilon. \tag{E.12}$$

Hence, we have, for $i = 1, 2, \cdots, N$,

$$\begin{aligned}
\log \mathcal{M}_i &\geq H(M_i), & \text{(E.13)}\\
&\geq H(M_i|M_0), & \text{(E.14)}\\
&= H(X_i^n M_i|M_0) - H(X_i^n|M_iM_0), & \text{(E.15)}\\
&\geq H(X_i^n M_i|M_0) - n\epsilon, & \text{(E.16)}\\
&= H(X^n|M_0) - n\epsilon. & \text{(E.17)}
\end{aligned}$$

Then, we get,

$$\sum_{i=1}^{N} \log \mathcal{M}_i \geq \sum_{i=1}^{N} H(X_i^n|M_0) - n\epsilon'. \tag{E.18}$$

where $\epsilon' = N\epsilon$. Together with (E.5), we get,

$$\begin{aligned}\sum_{i=0}^{N} \log \mathcal{M}_i \quad \geq \quad & nH(X_1 X_2 \cdots X_N) \\ & - \sum_{j=1}^{n} H(X_{1j} X_{2j} \cdots X_{Nj} | W_j) \\ & + \sum_{i=1}^{N} H(X_i^n | M_0) - n\epsilon'. \end{aligned} \tag{E.19}$$

Together with (5.9), we get,

$$\sum_{i=1}^{N} H(X_i^n | M_0) - \sum_{j=1}^{n} H(X_{1j} X_{2j} \cdots X_{Nj} | W_j) \leq n\epsilon''. \tag{E.20}$$

where $\epsilon'' = (N+1)\epsilon$. On the other hand, we have,

$$\begin{aligned} & \sum_{i=1}^{N} H(X_i^n | M_0) \\ = \quad & \sum_{i=1}^{N} \sum_{j=1}^{n} H(X_{ij} | X_i^{j-1} M_0), && \text{(E.21)} \\ \geq \quad & \sum_{i=1}^{N} \sum_{j=1}^{n} H(X_{ij} | X_1^{j-1}, X_2^{j-1}, \cdots, X_N^{j-1}, M_0), && \text{(E.22)} \\ = \quad & \sum_{i=1}^{N} \sum_{j=1}^{n} H(X_{ij} | W_j). && \text{(E.23)} \end{aligned}$$

Combine (E.20) and (E.23), we have,

$$\sum_{j=1}^{n} \Big[\sum_{i=1}^{N} H(X_{ij} | W_j) - H(X_{1j} X_{2j} \cdots X_{Nj} | W_j) \Big] \leq n\epsilon''. \tag{E.24}$$

Hence, we have,

$$\frac{1}{n} \sum_{j=1}^{n} \delta_2^{(j)} \leq \epsilon''. \tag{E.25}$$

This completes the proof of Lemma 6.□

Now, from Lemma 6 and (E.9), we get,

$$R_0 \geq \frac{1}{n} \log \mathcal{M}_0 \geq H(X_1, X_2, \cdots, X_N) - \Gamma_1(\eta). \tag{E.26}$$

Together with the continuity of $\Gamma_1(\cdot)$, we have, as $n \to \infty$,

$$\begin{aligned} R_0 &\geq H(X_1, X_2, \cdots, X_N) - \Gamma_1(0), && \text{(E.27)} \\ &= C(X_1, X_2, \cdots, X_N). && \text{(E.28)} \end{aligned}$$

This completes the proof of converse part.$\square$

We now prove the achievability part, that is, let the joint distribution $P(x_1, x_2, \cdots, x_N)$ be given, we have,

Theorem 15 (Achievability)

$$C_1 \leq C(X_1, X_2, \cdots, X_N). \quad \text{(E.29)}$$

Our proof mainly involves generalizing Gray-Wyner source coding network [68] to that of N sources. The system model we considered here is the same as Fig. **??** described in section II except that definition 6 is replaced by,

Definition 12 *A rate tuple* $(R_0, R_1, \cdots, R_N)$ *is said to be* achievable *if for all* $\epsilon > 0$*, we can find an* n *sufficiently large such that there exists a* $(n, 2^{nR_0}, 2^{nR_1}, \cdots, 2^{nR_N})$ *code with*

$$P_e^{(n)} \leq \epsilon. \quad \text{(E.30)}$$

Our purpose is to find all achievable rate tuples $(R_0, R_1, \cdots, R_N)$. The rate region of this source coding problem is summarized in the following theorem.

Theorem 16 *For the source coding model described above, a rate tuple* $(R_0, R_1, \cdots, R_N)$ *is achievable if and only if the following conditions are satisfied,*

$$\begin{aligned} R_0 &\geq I(X_1, X_2, \cdots, X_N; W), && \text{(E.31)} \\ R_i &\geq H(X_i|W), && \text{(E.32)} \end{aligned}$$

for $i = 1, 2, \cdots, N$*, and for some* $W \sim P(w|x_1, x_2, \cdots, x_N)$*, where* $W \in \mathcal{W}$ *and* $|\mathcal{W}| \leq \prod_{i=1}^{N} |\mathcal{X}_i| + 2$.

Proof of Theorem 16 (Sketch):

For the achievability part, we want to show that for any rate tuple $(R_0, R_1, \cdots, R_N)$ that satisfies above conditions, we can construct a $(n, 2^{nR_0}, 2^{nR_1}, \cdots, 2^{nR_N})$ code such that the decoding error $P_e^{(n)} \to 0$ as codeword length $n \to \infty$.

Codeword Generation: for any given distributions $P(x_1, x_2, \cdots, x_N)$ and $P(w|x_1, x_2, \cdots, x_N)$, we calculate the marginal distribution $P(w)$.

1. Codebook C_0: we first randomly generate 2^{nR_0} sequences w^n i.i.d. $\sim P(w)$, and index them by $m_0 \in \{1, 2, \cdots, 2^{nR_0}\}$.
2. Codebook $C(X_i)$: for each $i = 1, 2 \cdots, N$, for each $x_i^n \in \mathcal{X}_i^n$, randomly put them into 2^{nR_i} bins and index them bins by $m_i \in \{1, 2, \cdots, 2^{nR_i}\}$.

Encoding:

1. for each source sequences $(x_1^n, x_2^n, \cdots, x_N^n)$, encoder f_0 finds a $w^n(m_0) \in C_0$ such that $(x_1^n, x_2^n, \cdots, x_N^n, w^n(m_0)) \in T_\epsilon^n$, where T_ϵ^n is the jointly typical set as defined in [35], and send the index m_0 to the decoder. If there is no more than one w^n, choose the sequence w^n with the smallest index; if there exist no such sequence, choose sequence $w^n(1)$,
2. for $i = 1, 2, \cdots, N$, encoder f_i sends the bin index m_i of sequence x_i^n.

Decoding: for $i = 1, 2, \cdots, N$, decoder i looks at bin m_i for codebook $C(X_i)$ and finds the sequence $\hat{x}_i^n$ such that $(\hat{x}_i^n, w^n(m_0)) \in T_\epsilon^n$. If there is more than one or none such sequence, declare an error.

Error analysis: Assuming m_i, $i = 0, 1, \cdots, N$ are the chosen indices for encoding $(x_1^n, x_2^n, \cdots, x_N^n)$. There are three error events.

1. E_1: $(x_1^n, x_2^n, \cdots, x_N^n, w^n(m_0)) \notin T_\epsilon^n$ for all $m_0 \in \{1, 2, \cdots, 2^{nR_0}\}$.
2. E_2: $(x_i^n, w^n(m_0)) \notin T_\epsilon^n$ for each i.
3. E_3: for some i, there exists $\tilde{x}_i^n \neq x_i^n$ in bin m_i of codebook $C(X_i)$ such that $(\tilde{x}_i^n, w^n(m_0)) \in T_\epsilon^n$.

Hence,

$$P_e^{(n)} \leq P(E_1) + P(E_2|E_1^c) + P(E_3|E_1^c, E_2^c). \tag{E.33}$$

By some standard argument, we can get, as $n \to \infty$,

1. $P(E_1) \to 0$ if
$$R_0 \geq I(X_1, X_2, \cdots, X_N; W) + \epsilon, \tag{E.34}$$
2. $P(E_2|E_1^c) \to 0$,
3. $P(E_3|E_1^c, E_2^c) \to 0$ if for each $i = 1, 2, \cdots, N$,
$$R_i \geq H(X_i|W) + \epsilon. \tag{E.35}$$

This completes the achievability proof.

For the converse part, we want to show that for any achievable rate tuple $(R_0, R_1, \cdots, R_N)$, it should satisfy (E.31) and (E.32).

By Fano's inequality, we have

$$H(X_i^n|M_iM_0) \leq n\epsilon. \tag{E.36}$$

Hence, we have, for $i = 1, 2, \cdots, N$

$$\begin{aligned}
& nR_i \\
\geq\ & H(M_i), && \text{(E.37)}\\
\geq\ & H(M_i|M_0), && \text{(E.38)}\\
\geq\ & H(M_i|M_0) + H(X_i^n|M_iM_0) - n\epsilon, && \text{(E.39)}\\
=\ & H(X_i^nM_i|M_0) - n\epsilon, && \text{(E.40)}\\
=\ & H(X_i^n|M_0) - n\epsilon, && \text{(E.41)}\\
=\ & \sum_{j=1}^{n} H(X_{ij}|M_0X_i^{j-1}) - n\epsilon, && \text{(E.42)}\\
\geq\ & \sum_{j=1}^{n} H(X_{ij}|M_0, X_1^{j-1}, X_2^{j-1}, \cdots, X_N^{j-1}) - n\epsilon. && \text{(E.43)}
\end{aligned}$$

and

$$\begin{aligned}
& nR_0 \\
\geq\ & H(M_0), && \text{(E.44)}\\
\geq\ & I(M_0; X_1^n, X_2^n, \cdots, X_N^n), && \text{(E.45)}\\
=\ & \sum_{j=1}^{n} I(M_0; X_{1j}X_{2j}\cdots X_{Nj}|X_1^{j-1}X_2^{j-1}\cdots X_N^{j-1}), && \text{(E.46)}\\
=\ & \sum_{j=1}^{n} I(M_0X_1^{j-1}X_2^{j-1}\cdots X_N^{j-1}; X_{1j}X_{2j}\cdots X_{Nj}). && \text{(E.47)}
\end{aligned}$$

Define $W_j = (M_0, X_1^{j-1}, X_2^{j-1}, \cdots, X_N^{j-1})$, and using a standard time sharing argument, we can get, for $i = 1, 2, \cdots, N$,

$$\begin{aligned}
R_i &\geq H(X_i|W) - \epsilon, && \text{(E.48)}\\
R_0 &\geq I(X_1X_2\cdots X_N; W). && \text{(E.49)}
\end{aligned}$$

Let $n \to \infty$, then $\epsilon \to 0$, and this completes the proof of converse. The cardinality bound can be obtained using the technique introduced in [69, Appdendix C]. We skip the details. This completes the proof of Theorem 16.□

Now we proceed to prove Theorem 15. We will show that if $R_0 > C(X_1, X_2, \cdots, X_N)$, it is achievable for Model I.

Let $R_0 > C(X_1, X_2, \cdots, X_N)$ and any $\epsilon > 0$ be given and let random variables $(X_1, X_2, \cdots, X_N, W)$ satisfy (5.3) and (5.4), such that

$$C(X_1, X_2, \cdots, X_N) = I(X_1 X_2 \cdots X_N; W). \tag{E.50}$$

Notice that, the existence of such random variables is guaranteed by Lemma 5. Now define

$$\epsilon_1 = \min\{\frac{\epsilon}{N+1}, R_0 - C(X_1, X_2, \cdots, X_N)\}, \tag{E.51}$$

and hence $\epsilon_1 > 0$. By Theorem 16, there exists a $(n, \mathcal{M}_0, \mathcal{M}_1, \cdots, \mathcal{M}_N)$ code with $P_e^{(n)} \leq \epsilon'$ and $\epsilon' \leq \epsilon_1$. Hence,

$$\frac{1}{n} \log \mathcal{M}_0 \leq C(X_1, X_2, \cdots, X_N) + \epsilon_1 \leq R_0, \tag{E.52}$$

$$\frac{1}{n} \log \mathcal{M}_i \leq H(X_i|W) + \epsilon_1. \tag{E.53}$$

Hence, we have,

$$\begin{aligned} & \sum_{i=0}^{N} \frac{1}{n} \log \mathcal{M}_i \\ \leq \; & C(X_1, X_2, \cdots, X_N) + \sum_{i=1}^{N} H(X_i|W) + \epsilon, & \text{(E.54)} \\ \stackrel{(a)}{=} \; & H(X_1, X_2, \cdots, X_N) + \epsilon. & \text{(E.55)} \end{aligned}$$

where (a) is from condition (5.4). Thus, condition (5.9) is also satisfied. This implies that R_0 is achievable in Model I, which completes the proof of achievability part. This completes the proof of Theorem 15.□

Appendix F

Proof of $C_2 = C$

In this section, we prove the second part of theorem 1, that is $C_2 = C(X_1, X_2, \cdots, X_N)$. We have the following theorem.

Theorem 17

$$C_2 \geq C(X_1, X_2, \cdots, X_N), \tag{F.1}$$

$$C_2 \leq C(X_1, X_2, \cdots, X_N). \tag{F.2}$$

For the converse part , that is (F.1), the proof follows almost the same line as in [59, Section 5.2]. For the achievability part, that is (F.2), the proof follows similarly as in [59, Seciton 6.2] by applying $\mathcal{U} = \mathcal{X}_1 \times \mathcal{X}_2, \cdots \times \mathcal{X}_N$ in [59, Theorem 6.3]. We omit the details here.

Appendix G

Proof of the sufficient condition of $C_3 = C$: Theorem 11

We first introduce the following two lemmas. The first one is given by Gray [70].

Lemma 7 *Given a two-dimensional source X, Y and a compound distortion measure, we have the following inequalities*

$$R_{XY}(\Delta_1, \Delta_2) \geq R_{X|Y}(\Delta_1) + R_Y(\Delta_2), \tag{G.1}$$

$$R_{X|Y}(\Delta_1) \geq R_X(\Delta_1) - I(X;Y), \tag{G.2}$$

and equalities hold in some neighborhood of the origin $\{(\Delta_1, \Delta_2) : 0 \leq \Delta_1, \Delta_2 \leq \gamma\}$, provided that

$$Q(x,y) > 0 \quad \text{all } x \in \mathcal{X}, y \in \mathcal{Y}, \tag{G.3}$$

and d_1, d_2 satisfy

$$d_1(x,\hat{x}) > d_1(x,x) = 0, x \neq \hat{x}, \tag{G.4}$$

$$d_2(y,\hat{y}) > d_2(y,y) = 0, y \neq \hat{y}. \tag{G.5}$$

Here $R_{X|Y}(\Delta)$ is the conditional rate distortion function which is defined as

$$R_{X|Y}(\Delta) = \min I(X;\hat{X}|Y), \tag{G.6}$$

where the minimum is taken with respect to all test channels $q_t(\hat{x}|x,y)$ such that $Ed(X,\hat{X}) \leq \Delta$.

The second lemma is given by Gray and Wyner [68].

Lemma 8 *For the lossy source coding problem described in the previous section, for* $\Delta_1, \Delta_2 \geq 0$*, the rate distortion region is given by*

$$\begin{aligned}\mathcal{R}(\Delta_1, \Delta_2) = \{(R_0, R_1, R_2) : R_0 \geq I(X, Y; W), \\ R_1 \geq R_{X|W}(\Delta_1), R_2 \geq R_{Y|W}(\Delta_2)\},\end{aligned} \tag{G.7}$$

for some distributions $p(w|x, y)Q(x, y)$*.*

Note that Lemma 8 is valid for both the discrete case and the continuous case. Although Gray and Wyner only treated the discrete case in [68], the result can be generalized to the continuous case [72].

We now prove Theorem 11.

1. Achievability:

   For a given $Q(x, y) > 0$ $x \in \mathcal{X}, y \in \mathcal{Y}$, let $C(X; Y) = I(XY; W)$ where (X, Y, W) satisfies $X - W - Y$ and $\sum_w p(x, y, w) = Q(x, y)$, i.e., W is the auxiliary variable that achieves $C(X, Y)$. Let (Δ_1, Δ_2) be in the range $\{0 \leq \Delta_1, \Delta_2 \leq \gamma\}$ where γ is chosen such that the following equalities hold

   $$R_X(\Delta_1) = R_{X|W}(\Delta_1) + I(X; W), \tag{G.8}$$

   $$R_Y(\Delta_2) = R_{Y|W}(\Delta_2) + I(Y; W), \tag{G.9}$$

   $$R_{XY}(\Delta_1, \Delta_2) = R_X(\Delta_1) + R_Y(\Delta_2) - I(X; Y). \tag{G.10}$$

   We now prove that $C_3(\Delta_1, \Delta_2) \leq C(X, Y)$ in the range $\{0 \leq \Delta_1, \Delta_2 \leq \gamma\}$. For any $R_0 > C(X, Y)$ and $\epsilon > 0$ let

   $$\epsilon_1 = \min(\epsilon/3, R_0 - C(X; Y)). \tag{G.11}$$

   Since $\epsilon_1 > 0$, we know from Lemma 8 that there exists a code $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ with $\Delta_X \leq \Delta_1 + \epsilon_1$, $\Delta_Y \leq \Delta_2 + \epsilon_1$ and

   $$\begin{aligned}\frac{1}{n} \log M_0 &\leq I(X, Y; W) + \epsilon_1 \\ &= C(X; Y) + \epsilon_1 \leq R_0,\end{aligned} \tag{G.12}$$

   $$\frac{1}{n} \log M_1 \leq R_{X|W}(\Delta_1) + \epsilon_1, \tag{G.13}$$

   $$\frac{1}{n} \log M_2 \leq R_{Y|W}(\Delta_2) + \epsilon_1. \tag{G.14}$$

From (G.12-G.14), we have that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=0}^{2}\log M_i \\
&\le I(X,Y;W) + R_{X|W}(\Delta_1) + R_{Y|W}(\Delta_2) + 3\epsilon_1, && \text{(G.15)}\\
&= I(X;W) + R_{X|W}(\Delta_1) + I(Y;W) + R_{Y|W}(\Delta_2) \\
&\quad -I(X;Y) + 3\epsilon_1, && \text{(G.16)}\\
&= R_X(\Delta_1) + R_Y(\Delta_2) - I(X;Y) + 3\epsilon_1, && \text{(G.17)}\\
&\le R_{XY}(\Delta_1,\Delta_2) + \epsilon. && \text{(G.18)}
\end{aligned}
$$

where (G.16) follows from the chain rule and the Markov Chain $X - W - Y$, (G.17) and (G.18) follow from (G.8-G.11).

This proves that the code satisfies (6.12)-(6.14), i.e., R_0 is (Δ_1,Δ_2)-achievable. This completes the proof of $C_3(\Delta_1,\Delta_2) \le C(X;Y)$.

2. Converse:

Let Δ_1,Δ_2 be in the region $\{0 \le \Delta_1,\Delta_2 \le \gamma\}$ such that

$$R_X(\Delta_1) + R_Y(\Delta_2) - I(X;Y) = R_{XY}(\Delta_1,\Delta_2). \tag{G.19}$$

Let R_0 be (Δ_1,Δ_2)-achievable. We will show that $R_0 \ge C(X;Y)$. The proof follows similar procedures as the proof of Theorem 5.1 in [59].

Since R_0 is (Δ_1,Δ_2)-achievable, there exists an $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ code satisfying (6.12)-(6.14). Let $f_E(X^n, Y^n) = (W_0, W_1, W_2)$, we have that

$$
\begin{aligned}
R_0 &\ge \frac{1}{n}\log M_0 \ge \frac{1}{n}H(W_0), && \text{(G.20)}\\
&\ge \frac{1}{n}I(X^n,Y^n;W_0), && \text{(G.21)}\\
&= \frac{1}{n}H(X^n,Y^n) - \frac{1}{n}H(X^n,Y^n|W_0), && \text{(G.22)}\\
&= H(X,Y) - \frac{1}{n}\sum_{k=1}^{n} H(X_k,Y_k|X^{k-1},Y^{k-1},W_0), && \text{(G.23)}\\
&\ge H(X,Y) - \frac{1}{n}\sum_{k=1}^{n}\Gamma_1(\delta^{(k)}), && \text{(G.24)}\\
&\ge H(X,Y) - \Gamma_1(\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}), && \text{(G.25)}
\end{aligned}
$$

where (G.24) comes from the definition of $\Gamma_1(\cdot)$ (c.f. Corollary 4.5, [59]) and the definition of $\delta^{(k)}$, where

$$\delta^{(k)} = I(X_k; Y_k | X^{k-1}, Y^{k-1}, W_0).$$

Inequality (G.25) follows from the concavity of $\Gamma_1(\delta)$.

Therefore, since $C(X;Y) = H(X,Y) - \Gamma_1(0)$ (c.f. equation (4.4) in [59]) and (G.25), to establish $R_0 \geq C(X;Y)$ we only need to prove that, for arbitrary $\epsilon > 0$,

$$\frac{1}{n}\sum_{k=1}^{n} \delta^{(k)} \leq v(\epsilon), \tag{G.26}$$

$$\lim_{\epsilon \to 0} v(\epsilon) = 0. \tag{G.27}$$

From (G.22), we have that

$$\begin{aligned}
\frac{1}{n}\log M_0 &\geq \frac{1}{n}H(X^n, Y^n) - \frac{1}{n}H(X^n, Y^n | W_0), && \text{(G.28)}\\
&= \frac{1}{n}H(X^n, Y^n) + \frac{1}{n}I(X^n; Y^n | W_0) \\
&\quad -\frac{1}{n}H(X^n|W_0) - \frac{1}{n}H(Y^n|W_0). && \text{(G.29)}
\end{aligned}$$

Consider again the $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ code that satisfies (6.12)-(6.14) for arbitrary $\epsilon > 0$. Set $\hat{X}^n = f_D^{(X)}(W_0, W_1)$ and $\hat{Y}^n = f_D^{(Y)}(W_0, W_2)$, we have

$$\begin{aligned}
\frac{1}{n}\log M_1 &\geq \frac{1}{n}H(W_1), && \text{(G.30)}\\
&\geq \frac{1}{n}H(W_1|W_0), && \text{(G.31)}\\
&\geq \frac{1}{n}I(X^n; W_1|W_0), && \text{(G.32)}\\
&\geq \frac{1}{n}I(X^n; \hat{X}^n|W_0). && \text{(G.33)}
\end{aligned}$$

where inequality (G.33) follows from the Markov chain $X^n - W_0, W_1 - \hat{X}^n$. Similarly, we have

$$\frac{1}{n}\log M_2 \geq \frac{1}{n}I(Y^n; \hat{Y}^n|W_0). \tag{G.34}$$

Adding (G.29), (G.33) and (G.34), we obtain

$$\sum_{i=0}^{2}\frac{1}{n}\log M_i$$

$$\geq \frac{1}{n}(H(X^n,Y^n) + I(X^n;Y^n|W_0) - H(X^n|W_0)$$
$$-H(Y^n|W_0) + I(X^n;\hat{X}^n|W_0) + I(Y^n;\hat{Y}^n|W_0)), \tag{G.35}$$
$$= \frac{1}{n}(I(X^n;W_0) + I(Y^n;W_0) - I(X^n;Y^n)$$
$$+I(X^n;Y^n|W_0)+I(X^n;\hat{X}^n|W_0)+I(Y^n;\hat{Y}^n|W_0)), \tag{G.36}$$
$$= \frac{1}{n}(I(X^n;\hat{X}^n,W_0) + I(Y^n;\hat{Y}^n,W_0) - I(X^n;Y^n)$$
$$+I(X^n;Y^n|W_0)), \tag{G.37}$$
$$\geq \frac{1}{n}(I(X^n;\hat{X}^n)+I(Y^n;\hat{Y}^n)-nI(X;Y)$$
$$+I(X^n;Y^n|W_0)), \tag{G.38}$$

$$\geq R_X(\Delta_1)+R_Y(\Delta_2)-I(X;Y)+\frac{1}{n}I(X^n;Y^n|W_0), \tag{G.39}$$
$$= R_{XY}(\Delta_1,\Delta_2) + \frac{1}{n}I(X^n;Y^n|W_0). \tag{G.40}$$

where (G.36), (G.37) follow from the chain rule, (G.38) follows from the fact that conditioning does not increase entropy, (G.39) follows from the definition of rate distortion function and (G.40) is from (G.19).

On the other hand, the code satisfies (6.13), so we have

$$\sum_{i=0}^{2}\frac{1}{n}\log M_i \leq R_{XY}(\Delta_1,\Delta_2) + \epsilon. \tag{G.41}$$

Combining (G.40) and (G.41) we will have that

$$\frac{1}{n}I(X^n;Y^n|W_0) \leq \epsilon. \tag{G.42}$$

Also, it is easy to check that the following inequality is true.

$$\frac{1}{n}I(X^n;Y^n|W_0) \geq \frac{1}{n}\sum_{k=1}^{n}\delta^{(k)}. \tag{G.43}$$

Combining (G.42) and (G.43), we obtain

$$\frac{1}{n}\sum_{k=1}^{n}\delta^{(k)} \leq \epsilon, \tag{G.44}$$

which completes the proof.

Appendix H

Proof of the optimality of the distortion threshold: Theorem 12

Before proving Theorem 12, we first introduce two lemmas.

Lemma 9 *For any* Δ_1, Δ_2,

$$C_3(\Delta_1, \Delta_2) \leq R_{XY}(\Delta_1, \Delta_2). \tag{H.1}$$

Proof: The lemma follows from the fact that $R_0 = R_{XY}(\Delta_1, \Delta_2)$ is (Δ_1, Δ_2)-achievable.

Lemma 10 *Let* $\tau = R_{XY}^{-1}(C(X,Y))$, $\Delta \leq \tau$, *if* R_0 *is* Δ*-achievable, then there exists a* W *such that* $X - W - Y$, $R_0 \geq I(X,Y;W)$, *and*

$$I(X,Y;W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta,\Delta). \tag{H.2}$$

Proof: For $\Delta \leq \tau$, if R_0 is Δ-achievable, we have that for any $\epsilon > 0$, there exists a code $(n, M_0, M_1, M_2, \Delta, \Delta)$ that satisfies (6.12)-(6.14). Let $R_i' = \frac{1}{n}\log M_i$ for $i = 0, 1, 2$, we have that

$$\sum_{i=0}^{2} R_i' \leq R_{XY}(\Delta,\Delta) + \epsilon. \tag{H.3}$$

From the definition of rate distortion region [68], we know that $(R_0', R_1' - \epsilon/2, R_2' - \epsilon/2)$ is in the rate distortion region $\mathcal{R}$. By Lemma 8, there exists a W jointly distributed with X, Y as $p(w|x,y)Q(x,y)$ and satisfies

$$R_0' + R_1' + R_2' - \epsilon \geq I(X,Y;W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta), \tag{H.4}$$

$$\geq I(X,Y;W) + R_{XY|W}(\Delta,\Delta), \tag{H.5}$$
$$\geq R_{XY}(\Delta,\Delta), \tag{H.6}$$

where inequalities (H.5) and (H.6) are from Theorem 3.1 in [70]. The equality in (H.5) holds only when X is conditionally independent of Y given W and equality in (H.6) holds only when $0 \leq \Delta \leq \gamma$. For $\Delta = \tau$, combined with (H.3), we have that $I(X,Y;W) = R_{XY}(\tau,\tau)$. Hence for any $\Delta \leq \tau$,

$$I(X,Y;W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta,\Delta). \tag{H.7}$$

This completes the proof.

We now prove Theorem 12.

First we show that for any Δ such that $C_3(\Delta) = C(X,Y)$, we have $\Delta \leq \tau$. From Lemma 9, $R_{XY}(\Delta,\Delta) \geq C(X,Y)$. $R_{XY}(\Delta,\Delta)$ is a non increasing function of Δ, therefore, $\Delta \leq R_{XY}^{-1}(C(X,Y)) = \tau$.

Next we will show that for any distortion $\Delta \leq \tau$, $C_3(\Delta) = C(X,Y)$.

For any R_0 that is Δ-achievable, from Lemma 10, there exists a W such that $X - W - Y$ and $R_0 \geq I(X,Y;W)$. Hence, $R_0 \geq I(X,Y;W) \geq C(X,Y)$, which implies $C_3(\Delta) \geq C(X,Y)$.

From Lemma 3, $C_3(\tau) \leq C(X,Y)$. Hence, $C_3(\tau) = C(X,Y)$. Thus, any rate $R_0 > C(X,Y)$ is τ-achievable. By Lemma 10, we have

$$C(X,Y) + R_{X|W}(\tau) + R_{Y|W}(\tau) = R_{XY}(\tau,\tau),$$

where W is the random variable such that $C(X,Y) = I(X,Y;W)$. Thus, by Lemma 7, for any $\Delta \leq \tau$,

$$C(X,Y) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta,\Delta).$$

Then use the same proof as the achievability part of Theorem 11, we can prove that when the distortion $\Delta \leq \tau$, any rate $R_0 > C(X,Y)$ is Δ-achievable. Hence, $C_3(\Delta) \leq C(X,Y)$, completing the proof.

Appendix I

Proof of the optimality of quantization threshold: Theorem 13

With one bit scalar quantization, optimizing error exponent is equivalent to maximize the mutual information $I(U;V)$. Define, under H_0, $P_{ij} = Pr(U = i; V = j), \quad i, j = \{0, 1\}$, which can be expressed in terms of integration of (7.2) given the single threshold quantizer assumption. We have:

$$\begin{aligned} P_{00} &= \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} f_{x,y}(x,y)dxdy, && \text{(I.1)} \\ P_{01} &= \int_{-\infty}^{t_1} \int_{t_2}^{\infty} f_{X,Y}(x,y)dxdy, && \text{(I.2)} \\ P_{10} &= \int_{t_1}^{\infty} \int_{\infty}^{t_2} f_{X,Y}(x,y)dxdy, && \text{(I.3)} \\ P_{11} &= \int_{t_1}^{\infty} \int_{t_2}^{\infty} f_{X,Y}(x,y)dxdy. && \text{(I.4)} \end{aligned}$$

By definition,

$$Pr(U = 1) = Pr(X \geq t_1) = Q(t_1), \tag{I.5}$$

$$Pr(V = 1) = Pr(Y \geq t_2) = Q(t_2), \tag{I.6}$$

where the Q function is complementary cumulative distribution function for standard Gaussian distribution, defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-t^2/2)dt. \tag{I.7}$$

We want to maximize $I(U;V)$, where

$$\begin{aligned} I(U;V) &= H(U) + H(V) - H(U;V), && \text{(I.8)} \\ &= H(Q(t_1)) + H(Q(t_2)) - H(P_{00}, P_{01}, P_{10}, P_{11}), && \text{(I.9)} \end{aligned}$$

where $H(\cdot)$ is the Shannon entropy function, i.e.,

$$H(Q(t)) = -Q(t)\log Q(t) - (1 - Q(t))\log(1 - Q(t)), \tag{I.10}$$

and

$$H(P_{00}, P_{01}, P_{10}, P_{11}) = -\sum_{i=1}^{2}\sum_{j=1}^{2} P_{ij}\log P_{ij}. \tag{I.11}$$

We now compute the first partial derivative of $I(U;V)$ with respect to t_1 and t_2, receptively. We get, with tedious but straightforward computation,

$$\begin{aligned} \frac{\partial I(U;V)}{\partial t_1} &= \frac{1}{\sqrt{2\pi}}\exp\frac{-t_1^2}{2}\Big\{\log\frac{Q(t_1)}{1-Q(t_1)} + [1 - Q(\frac{t_2 - \rho t_1}{\sqrt{1-\rho^2}})]\log\frac{P_{00}}{P_{10}} \\ &\quad + Q(\frac{t_2 - \rho t_1}{\sqrt{1-\rho^2}})\log\frac{P_{01}}{P_{11}}\Big\}, && \text{(I.12)} \\ \frac{\partial I(U;V)}{\partial t_2} &= \frac{1}{\sqrt{2\pi}}\exp\frac{-t_2^2}{2}\Big\{\log\frac{Q(t_2)}{1-Q(t_2)} + [1 - Q(\frac{t_1 - \rho t_2}{\sqrt{1-\rho^2}})]\log\frac{P_{00}}{P_{01}} \\ &\quad + Q(\frac{t_1 - \rho t_2}{\sqrt{1-\rho^2}})\log\frac{P_{10}}{P_{11}}\Big\}. && \text{(I.13)} \end{aligned}$$

One can easily check that $(t_1, t_2) = (0, 0)$ is a critical point, i.e., the first partial derivatives equal 0. We next check its Hessian matrix:

$$M = \begin{pmatrix} a(\rho) & b(\rho) \\ b(\rho) & c(\rho) \end{pmatrix} \tag{I.14}$$

where

$$\begin{aligned} a(\rho) &= \frac{\partial^2 I(U;V)}{\partial t_1^2}\Big|_{(t_1,t_2)=(0,0)}, \\ b(\rho) &= \frac{\partial^2 I(U;V)}{\partial t_1 t_2}\Big|_{(t_1,t_2)=(0,0)}, \\ c(\rho) &= \frac{\partial^2 I(U;V)}{\partial t_2^2}\Big|_{(t_1,t_2)=(0,0)}. \end{aligned}$$

We want to show that $a(\rho) < 0$ and $\det M = b(\rho)^2 - a(\rho)c(\rho) > 0$ for all $\rho \in [-1, 0) \cup (0, 1]$.

We can easily calculate that,

$$\begin{aligned} a(\rho) &= c(\rho), && \text{(I.15)} \\ &= \frac{1}{2\pi}[-4 + \frac{2\rho}{\sqrt{1-\rho^2}} \log \frac{P_{10}}{P_{11}} + \frac{1}{4P_{10}P_{11}}]|_{(0,0)}, && \text{(I.16)} \\ b(\rho) &= \frac{1}{2\pi}[\frac{2}{\sqrt{1-\rho^2}} \log \frac{P_{11}}{P_{10}} + \frac{P_{10}-P_{11}}{2P_{10}P_{11}}]|_{(0,0)}. && \text{(I.17)} \end{aligned}$$

Next, we introduce a lemma concerning evaluating the cumulative distribution function of a standard bivariate Gaussian distribution at point $(0,0)$.

Lemma 11 *[82, page 290]*

$$P_{00}(t_1 = t_2 = 0) = P_{11}(t_1 = t_2 = 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho), \quad \text{(I.18)}$$

$$P_{01}(t_1 = t_2 = 0) = P_{10}(t_1 = t_2 = 0) = \frac{1}{4} - \frac{1}{2\pi}\arcsin(\rho). \quad \text{(I.19)}$$

Using (I.18) and (I.19), we can further get

$$\begin{aligned} a(\rho) = c(\rho) &= \frac{1}{2\pi}[-4 + \frac{2\rho}{\sqrt{1-\rho^2}} \log \frac{\pi - 2\arcsin\rho}{\pi + 2\arcsin\rho} \\ &\quad + \frac{4\pi^2}{\pi^2 - 4\arcsin^2\rho}], && \text{(I.20)} \\ b(\rho) &= \frac{1}{2\pi}[\frac{2}{\sqrt{1-\rho^2}} \log \frac{\pi + 2\arcsin\rho}{\pi - 2\arcsin\rho} \\ &\quad - \frac{8\pi\arcsin\rho}{\pi^2 - 4\arcsin^2\rho}]. && \text{(I.21)} \end{aligned}$$

Next, we want to evaluate functions $a(\rho), b(\rho)$ and $c(\rho)$ with the help of the following two lemmas.

Lemma 12 *For $a(\rho)$ and $c(\rho)$ defined above, we have:*

$$a(\rho) = c(\rho) \leq 0, \quad \text{(I.22)}$$

for all $\rho \in [-1,1]$ and the maximum is achieved when $\rho = 0$.

Lemma 13 *For the function $b(\rho)$ defined above, we have*

$$\begin{aligned} b(\rho) &> 0, \quad \textit{if} \quad \rho \in (0,1], && \text{(I.23)} \\ b(\rho) &< 0, \quad \textit{if} \quad \rho \in [-1,0), && \text{(I.24)} \\ b(\rho) &= 0, \quad \textit{if} \quad \rho = 0. && \text{(I.25)} \end{aligned}$$

Form Lemma 12, we can see that $a(\rho) < 0$ for all $\rho \in [-1, 0) \cup (0, 1]$ is satisfied. Next, we want to prove that $b^2(\rho) - a(\rho)c(\rho) < 0$ for all $\rho \neq 0$ is also true. Notice that, form Lemmas 12 and I.25, we only need to prove that

$$\begin{aligned} -a(\rho) &> b(\rho) \quad \text{if} \quad \rho \in (0, 1], && \text{(I.26)} \\ a(\rho) &< b(\rho) \quad \text{if} \quad \rho \in [-1, 0). && \text{(I.27)} \end{aligned}$$

Define, $d(\rho) = -a(\rho) - b(\rho)$ and $e(\rho) = a(\rho) - b(\rho)$. We want to show that

$$\begin{aligned} d(\rho) &> 0 \quad \text{if} \quad \rho \in (0, 1], && \text{(I.28)} \\ e(\rho) &< 0 \quad \text{if} \quad \rho \in [-1, 0). && \text{(I.29)} \end{aligned}$$

This can be verified by noting that

$$\begin{aligned} d(\rho) &= \frac{1}{2\pi}[-2\sqrt{\frac{1-\rho}{1+\rho}} \log \frac{\pi + 2 \arcsin \rho}{\pi - 2 \arcsin \rho} + \frac{8(\pi - 2 \arcsin \rho) \arcsin \rho}{\pi^2 - 4 \arcsin^2 \rho}], && \text{(I.30)} \\ e(\rho) &= \frac{1}{2\pi}[2\sqrt{\frac{1+\rho}{1-\rho}} \log \frac{\pi - 2 \arcsin \rho}{\pi + 2 \arcsin \rho} + \frac{8(\pi + 2 \arcsin \rho) \arcsin \rho}{\pi^2 - 4 \arcsin^2 \rho}]. && \text{(I.31)} \end{aligned}$$

Appendix J

Proof of the optimal local decision rule : Proposition 15 and 16

We begin with the probability of error for the hypothesis testing problem. We first expand the probability of error with respect to sensor 1. The result is given by

$$\begin{aligned}
P_e &= \pi_0 P(U=1|H_0) + \pi_1 P(U=0|H_1), \\
&= \pi_0 \int_{x_1}\int_{x_2}\sum_{U_1}\sum_{U_2} P(U=1,U_1,U_2,x_1,x_2|H_0)dx_1dx_2 \\
&\quad +\pi_1 \int_{x_1}\int_{x_2}\sum_{u_1}\sum_{u_2} P(U=0,U_1,U_2,x_1,x_2|H_1)dx_1dx_2, \\
&= \pi_0 \int_{x_1}\int_{x_2}\sum_{U_2}\big[P(U_0=1|U_1=1,U_2)P(x_1x_2|H_0)P(U_1=1|x_1)P(U_2|x_2) \\
&\quad +P(U_0=1|U_1=0,U_2)P(x_1x_2|H_0)P(U_1=0|x_1)P(U_2|x_2)\big]dx_1dx_2 \\
&\quad +\pi_1 \int_{x_1}\int_{x_2}\sum_{U_2}\big[P(U_0=0|U_1=1,U_2)P(x_1x_2|H_1)P(U_1=1|x_1)P(U_2|x_2) \\
&\quad +P(U_0=0|U_1=0,U_2)P(x_1x_2|H_1)P(U_1=0|x_1)P(U_2|x_2)\big]dx_1dx_2, \\
&= \int_{x_1}\int_{x_2}\sum_{U_2}\Big\{\big[\pi_0 P(U_0=1|U_1=1,U_2)P(x_1x_2|H_0)+\pi_1 P(U_0=0|U_1=1,U_2) \\
&\quad \cdot P(x_1x_2|H_1)\big]P(U_1=1|x_1)P(U_2|x_2)\Big\}dx_1dx_2 \\
&\quad + \int_{x_1}\int_{x_2}\sum_{U_2}\Big\{\big[\pi_0 P(U_0=1|U_1=0,U_2)P(x_1x_2|H_0)+\pi_1 P(U_0=0|U_1=0,U_2) \\
&\quad \cdot P(x_1x_2|H_1)\big]P(U_1=0|x_1)P(U_2|x_2)\Big\}dx_1dx_2
\end{aligned}$$

$$
\begin{aligned}
&= \int_{x_1}\int_{x_2} \big\{\pi_0 A_1 P(x_1x_2|H_0) - \pi_1 B_1 P(x_1x_2|H_1)\big\} P(U_1=1|x_1)dx_1dx_2 + C_1, \\
&= \int_{x_1} P(U_1=1|x_1) \int_{x_2} \big\{\pi_0 A_1 P(x_1x_2|H_0) - \pi_1 B_1 P(x_1x_2|H_1)\big\} dx_2dx_1 + C_1,
\end{aligned} \tag{J.1}
$$

where A_1 and B_1 are defined as in (7.9) and (7.10), and C_i is defined as

$$
\begin{aligned}
C_i \;=\; & \int_{x_1}\int_{x_2}\sum_{U_{\bar{i}}} \big[\pi_0 P(U_0=1|U_i=0U_{\bar{i}})P(x_1x_2|H_0) \\
& + \pi_1 P(U_0=0|U_i=0U_{\bar{i}})P(x_1x_2|H_1)\big] P(U_{\bar{i}}|x_{\bar{i}})dx_1dx_2.
\end{aligned}
$$

Note that, for fixed decision rule at sensor $\bar{i}$, the term C_i is a constant. Then, minimize P_e is equivalent to minimize the first term in (J.1). If we further assume that the fusion rule satisfies

$$
\begin{aligned}
P(U_0=1|U_1=1,U_2=j) &\geq P(U_0=1|U_1=0,U_2=j), \\
P(U_0=0|U_1=0,U_2=j) &\geq P(U_0=0|U_1=1,U_2=j),
\end{aligned}
$$

for all $j=\{0,1\}$, then the optimal decision rule at sensor 1 is given as follows:

$$
P(U_1=1|x_1) = \begin{cases} 1 & \text{if } \frac{\int_{x_2} B_1 P(x_1x_2|H_1)dx_2}{\int_{x_2} A_1 P(x_1x_2|H_0)dx_2} \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{otherwise} \end{cases} \tag{J.2}
$$

The optimal decision rule at sensor 2 can be proved similarly if we expand the error probability with respect to sensor 1. This completes the proof of Proposition 15.

If we assume AND rule is used at the fusion center. Then

$$
A_i = B_i = I_{D_{\bar{i}}}, \tag{J.3}
$$

Where I is the indicator function, i.e.,

$$
I_{D_i} = \begin{cases} 1 & \text{if } x_i \in D_i \\ 0 & \text{otherwise} \end{cases} \tag{J.4}
$$

Hence,

$$
\int_{x_{\bar{i}}} B_i P(x_1x_2|H_1)dx_i = \int_{D_{\bar{i}}} P(x_1x_2|H_1)dx_{\bar{i}}, \tag{J.5}
$$

$$
\int_{x_{\bar{i}}} A_i P(x_1x_2|H_0)dx_i = \int_{D_{\bar{i}}} P(x_1x_2|H_0)dx_i. \tag{J.6}
$$

Together with Proposition 15, we complete the proof of Proposition 16.

Appendix K

Proof of Proposition 13

Let W, N_1 and N_2 be standard Gaussian random variables independent of each other and express X, Y as

$$X = \sqrt{\rho}W + \sqrt{1-\rho}N_1, \tag{K.1}$$
$$Y = \sqrt{\rho}W + \sqrt{1-\rho}N_2. \tag{K.2}$$

It is easy to verify that conditions (C1) and (C2) are satisfied. Straightforward calculation yields $I(X,Y;W) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}$.

The proof is thus complete if one can prove $I(X,Y;W) > \frac{1}{2}\log\frac{1+\rho}{1-\rho}$ for all W satisfying the conditions (C1) and (C2).

Let $P_{X,W,Y}$ be any joint distribution satisfying the conditions (C1) and (C2) and let K denote the corresponding covariance matrix. Let $\tilde{P}_{X,W,Y}$ be joint Gaussian satisfying the conditions (C1) and (C2) with zero mean and the same covariance matrix K. From the fact that conditional differential entropy is maximized under Gaussian distribution for a given covariance matrix [73], we have

$$h(X,Y|W) \leq h_{\tilde{P}}(X,Y|W). \tag{K.3}$$

Therefore $I(X,Y;W) \geq I_{\tilde{P}}(X,Y;W)$. Hence we only need to consider (X,W,Y) that are jointly Gaussian distributed.

Without loss of generality, let W be a Gaussian random variable with zero mean and variance σ^2, and

$$X = \rho_1 W + \sqrt{1-\rho_1^2\sigma^2}N_1, \tag{K.4}$$
$$Y = \rho_2 W + \sqrt{1-\rho_2^2\sigma^2}N_2, \tag{K.5}$$

where N_1 and N_2 are standard Gaussian random variables and W, N_1, N_2 are mutually independent with each other.

Since $EXY = \rho$, we have

$$\rho = \rho_1 \rho_2 \sigma^2, \tag{K.6}$$

and due to the Markov chain $X - W - Y$, we have $H(X|W) = H(X|W, Y)$, i.e.,

$$1 - \rho_1^2 = \frac{1 + 2\rho\rho_1\rho_2 - \rho^2 - \rho_1^2 - \rho_2^2}{1 - \rho_2^2}. \tag{K.7}$$

Combining (K.6) and (K.7), we get $\sigma^2 = 1$. Therefore, we can lower bound $I(X, Y; W)$ by

$$\begin{aligned}
I(X, Y; W) &= h(X, Y) - h(X|W) - h(Y|W), && \text{(K.8)} \\
&= \frac{1}{2} \log \frac{1 - \rho^2}{(1 - \rho_1^2)(1 - \rho_2^2)}, && \text{(K.9)} \\
&= \frac{1}{2} \log \frac{1 - \rho^2}{1 + \rho^2 - \rho_1^2 - \rho_2^2}, && \text{(K.10)} \\
&\geq \frac{1}{2} \log \frac{1 - \rho^2}{1 + \rho^2 - 2\rho}, && \text{(K.11)} \\
&= \frac{1}{2} \log \frac{1 + \rho}{1 - \rho}, && \text{(K.12)}
\end{aligned}$$

where we use the facts that $\rho_1 \rho_2 = \rho$ and $\rho_1^2 + \rho_2^2 \geq 2\rho_1\rho_2$.

Appendix L

Proof of Corollary 7

Without any loss of generality, we only consider the case $H_0 : \rho > 0$ Vs $H_1 : \rho = 0$. We assume that the rejection region of sensor 1 is a single semi-infinite interval, that is,

$$D_1 = (l_1, +\infty). \tag{L.1}$$

From Corollary 1, we have that, the rejection region D_2 for sensor 2 is characterized by

$$D_2 = \left\{ x_2 : \frac{Q(l_1)}{Q(\frac{l_1 - \rho x_2}{\sqrt{1-\rho^2}})} \geq \frac{\pi_0}{\pi_1} \right\}, \tag{L.2}$$

Where $Q(\cdot)$ is the Q function defined in (I.7).

Since $Q(\frac{l_1 - \rho x_2}{\sqrt{1-\rho^2}})$ is monotone increasing as a function of x_2 from 0 to 1 given $\rho > 0$ (monotone decreasing from 1 to 0 given $\rho < 0$), hence D_2 is a single semi-infinite interval of the form $(-\infty, l_2)$ given $\rho > 0$ ((l_2, ∞) given $\rho < 0$). This completes the proof of Corollary 7.

Bibliography

[1] C.E. Shannon, “A mathematical theory of communication”, *Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.

[2] T. M. Cover, A. El-Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources”, *IEEE. Trans. Inform. Theory*, vol. 26, no.6, pp. 648-657, Nov. 1980.

[3] T. S. Han and M. H. M. Costa, “Broadcast channels with arbitrarily correlated sources”, *IEEE Trans. Inform. Theory*, vol. 33, no. 5, pp. 641-650, Sep. 1987.

[4] A. Lapidoth and S. Tinguely, “Sending a bi-variate Gaussian over a Gaussian MAC”, *IEEE. Trans. Inform. Theory*, vol. 56, no. 6, pp. 2714-2752, June 2010.

[5] C. E. Shannon, “Two-way communication channels”, In *Proc. 4th Berkeley Symp. Math. Sta. Prob.*, Berkeley, CA, vol. 1, pp. 611-644, 1961.

[6] R. Ahlswede, “The capacity region of a channel with two senders and two receivers”, *Ann. Prob.*, vol. 2, no. 5, pp. 805-814, 1974.

[7] A. B. Carleial, “A case where interference does not reduce capacity”, *IEEE Trans. Inf. Theory*, vol. 21, pp. 569-570, Sep. 1975.

[8] H. Sato, “Two-user communication channels”, *IEEE Trans. Inf. Theory*, vol. 23, pp. 295-304, May 1977.

[9] A. B. Carleial, “Interference Channels”, *IEEE Trans. Inf. Theory*, vol. 24, pp. 60-70, Jan. 1978.

[10] H. Sato, “On degraded Gaussian two-user channels”, *IEEE Trans. Inform. Theory*, vol. 24, pp. 634-640, Sep. 1978.

[11] H. Sato, "On the capacity region of a discrete two-user channel for strong interference", *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 377-379, May 1978.

[12] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel", *IEEE Trans. Inf. Theory*, vol. 27, pp. 49-60, Jan. 1981.

[13] H. Sato, "The capacity of the Gaussian interference channel under strong interference", *IEEE Trans. Inf. Theory*, vol. 27, pp. 786-788, Nov. 1981.

[14] A. El Gamal and M. H. M. Costa, "The capacity region of a class of deterministic interference channels", *IEEE Trans. Inf. Theory*, vol. 28, pp. 343-346, Mar. 1982.

[15] M. H. M. Costa, "On the Gaussian interference channel", *IEEE Trans. Inf. Theory*, vol. 31, pp. 607-615, Sep. 1985.

[16] M. H. M. Costa and A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference", *IEEE Trans. Inform. Theory*, vol. 33, pp. 710-711, Sep. 1987.

[17] I. Sason, "On achievable rate regions for the Gaussian interference channels", *IEEE Trans. Inform. Theory*, vol. 50, pp. 1345-1356, June 2004.

[18] H. F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel", *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 3188-3195, July 2008.

[19] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources", *Bell Syst. Tech. J.*, vol. 52, pp. 1037-1076, Sep. 1973.

[20] R. Ahlswede, "Multi-way communication channels", In *Proc. 2nd IEEE Int. Symp. on Information Theory*, Tsahkadsor, USSR, 1971.

[21] H. Liao, "*Multiple access channels*", PhD thesis, University of Hawaii, Honolulu, 1972.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471-480, July 1973.

[23] P. Gács and J. Körner, "Common information is much less than mutual information", *Problems Contr. Inform. Theory*, vol. 2, pp. 149-162, 1973.

[24] H. S. Witsenhausen, "On sequences of pairs of dependent random variables", *SIAM J. Appl. Math.*, vol. 28, pp. 100-113, Jan. 1975.

[25] G. Kramer and C. Nair, "Comments on broadcast channels with arbitrarily correlated sources", In *Proc. IEEE Int. Symp. on Information Theory (ISIT09)*, Seoul, Korea, June-July 2009.

[26] K. Marton, "A coding theorem for the discrete memoryless broadcast channel", *IEEE. Trans. Inform. Theory*, vol. 25, no. 1, pp. 306-311, May 1979.

[27] P. Minero and Y. H. Kim, "Correlated sources over broadcast channels", In *Proc. IEEE Int. Symp. on Information Theory (ISIT09)*, Seoul, Korea, June-July 2009.

[28] M. Salehi and E. Kurtas, "Interference channels with correlated sources", In *Proc. IEEE Int. Symp. on Information Theory (ISIT93)*, San Antorio, Texas, Jan. 1993.

[29] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory", *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 396-412, May 1983.

[30] J. Jiang, Y. Xin, and H. K. Garg, "An achievable rate region for interference channels with common information", In *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2006.

[31] Y. Cao, B. Chen, and J. Zhang, "A new achievable rate region for interference channels with common information", In *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC 07)*, Hong Kong, Mar. 2007.

[32] J. Jiang, Y. Xin, and H. K. Garg, "Interference channels with common information", *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 171-187, Jan. 2008.

[33] A. Lapidoth and S. Tinguely, "Sending a bi-variate Gaussian source over a Gaussian MAC", In *Proc. IEEE Int. Symp. Infor.Theory (ISIT06)*, Seattle, USA, Jul. 2006.

[34] D. Gunduz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels", *IEEE. Trans. Inform. Theory*, vol. 55, no. 9, pp. 3927-3944, Sep. 2009.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, New York, 1991.

[36] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources", *IEEE. Trans. Inform. Theory*, vol. 11, no. 4, pp. 558-567, Oct. 1965.

[37] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network", *IEEE. Trans. Inform. Theory*, vol. 54, no. 11, pp. 2254-2261, Nov. 2008.

[38] S. Bross, A. Lapidoth, and S. Tinguely, "Broadcasting correlated Gaussians", In *Proc. IEEE Int. Symp. Infor.Theory (ISIT08)*, Toronto, Canada, Jul. 2008.

[39] T. Berger, "Multiterminal source coding," in *lecture notes presented at CISM*, Udine, Italy, 1977.

[40] S. Y. Tung, *Multiterminal rate-distortion theory*, Ph.D. thesis, School of Elec. Eng., Cornell Univ., Ithaca, NY, 1977.

[41] Y. Oohama, "Gaussian multiterminal source coding", *IEEE. Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912-1923, Nov. 1997.

[42] U. Mittal and N. Phamdo, "Hybrid digital analog (HDA) joint source channel codes for broadcasting and robust communications", *IEEE. Trans. Inform. Theory*, vol. 48, no. 5, pp. 1082-1102, May 2002.

[43] V. Prabhakaran, R. Puri, and K. Ramchandran, "Hybrid digital-analog strategies for source-channel broadcast", In *Proc. Allerton Conf. Communication, Control, and Computing*, 2005.

[44] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion", *IEEE. Trans. Inform. Theory*, vol. 52, no. 8, pp. 3778-3788, Aug. 2006.

[45] K. Narayanan, G. Caire, and M. Wilson, "Duality between broadcasting with bandwidth expansion and bandwidth compression", In *Proc. IEEE Int. Symp. Infor.Theory*, Nice, France, June 2007.

[46] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder", *IEEE. Trans. Inform. Theory*, vol. 51, no. 7, pp. 2577-2593, Jul. 2005.

[47] V. Prabhakaran, D. N. C. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem", In *Proc. IEEE Int. Symp. Infor.Theory*, Chicago, IL, USA, June-July 2004.

[48] O. Simeone, "Source and channel coding for homogeneous sensor networks with partial cooperation", *IEEE Trans. on Wireless Communications*, vol. 8, no. 3, pp. 1113-1117, March 2009.

[49] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit", *IEEE. Trans. Inform. Theory*, vol. 54, no. 12, pp. 5534-5562, Dec. 2008.

[50] M. Salehi, "Multiple-access channels with correlated sources - coding subject to a fidelity criterion", In *Proc. IEEE Int. Symp. on Information Theory (ISIT95)*, Sep. 1995.

[51] R. Rajesh and V. Sharma, "Source-channel coding for Gaussian sources over a Gaussian multiple access channel", In *Proc. Allerton Conf. Communication, Control, and Computing*, 2007.

[52] R. Soundararajan and S. Vishwanath, "Hybrid Coding for Gaussian Broadcast Channels with Gaussian Sources", In *Proc. IEEE Int. Symp. on Information Theory (ISIT09)*, Seoul, Korea, June-July 2009.

[53] H. Behroozi, F. Alajaji, and T. Linder, "Hybrid digital-analog joint source-channel coding for broadcasting correlated Gaussian sources", In *Proc. IEEE Int. Symp. on Information Theory (ISIT09)*, Seoul, Korea, June-July 2009.

[54] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem", *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1938-1961, May 2008.

[55] S. Bross, A. Lapidoth, and S. Tinguely, "Superimposed coded and uncoded transmissions of a Gaussian source over the Gaussian channel", In *Proc. IEEE Int. Symp. Infor.Theory (ISIT06)*, Seattle, USA, July 2006.

[56] P. Gács and J. Körner, "Common information is much less than mutual information", *Problems Contr. Inform. Theory*, vol. 2, pp. 149-162, 1973.

[57] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources", *Proc. of the 7th Prague Conference of Information Theory*, 1974.

[58] H. S. Witsenhausen, "On sequences of pairs of dependent random variables", *SIAM J. Appl. Math.*, vol. 28, pp. 100-113, Jan. 1975.

[59] A. D. Wyner, "The common information of two dependent random variables", *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163-179, March. 1975.

[60] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables", *SIAM J. Appl. Math.*, vol. 31, no. 2, pp. 313-333, 1976.

[61] R. Ahlswede and I. Csiszár, "Common randomness in information thoery and cryptography, Part I: Secret sharing", *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121-1132, July 1993.

[62] R. Ahlswede and I. Csiszár, "Common randomness in information thoery and cryptography, Part I: CR capacity", *IEEE Trans. Inf. Theory*, vol. 44, pp. 225-240, Jan. 1998.

[63] U. M. Maurer, "Secret key agreement by public discussion from common information", *IEEE Trans. Inf. Theory*, vol. 39, pp. 733-742, May 1993.

[64] G. Hu, "On the amount of information", *Teor. Veroyatnost. i Primenen.*, vol. 4, pp. 447-455, 1962 (in Russian).

[65] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic, New York, 1981.

[66] R. W. Yeung, *Information Theory and Network Coding*, Springer, 2008.

[67] H. Tyagi and P. Narayan and P. Gupta, "When is a function securely computable? ", *http://arxiv.org/abs/1007.2945*, 2010.

[68] R. M. Gray and A. D. Wyner, "Source coding for a simple network", *Bell Syst. Tech. J.*, vol. 58, pp. 1681-1721, Nov. 1974.

[69] A. El Gamal and Y. H. Kim, *Lecture notes on network information theory*, http://arxiv.org/abs/1001.3404, 2010.

[70] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, pp. 480-489, Jul. 1973.

[71] T. Berger, *Rate Distortion Theory*, Prentice-Hall, 1971.

[72] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and control*, vol. 38, pp. 60-80, 1978.

[73] J. A. Thomas, "Feedback can at most double Gaussian multiple access channel capacity." *IEEE Trans. Inf. Theory*,vol. 33, no. 5, pp. 711-716, Sep. 1987.

[74] W. Hoeffding, "A non-parametric test of independence," *The Annals of Mathematical Statistics*, vol. 19, no. 4, pp. 546-557, 1948.

[75] J. G. Daugman, "High confidence visual recognition of persons by a test ofstatistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148-1161, Nov 1993.

[76] W. A. Broock, J. A. Scheinkman, W. D. Dechert, and B. LeBaron, "A test for independence based on the correlation dimension," *Econometric Reviews*, vol. 15, no. 3, pp. 197-235, 1996.

[77] D.R. Cox and D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall Ltd, New York, 1974.

[78] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533-542, July 1986.

[79] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759-772, Nov. 1987.

[80] T. S. Han and S. I. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300-2324, Oct. 1998.

[81] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag New York, Inc, New York, NY, 1997.

[82] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

VITA

NAME OF AUTHOR: Wei Liu

MAJOR: Electrical and Computer Engineering

EDUCATION:

Ph.D.	May 2011	Syracuse University, NY, USA (expected)
M.S.	Aug. 2009	Syracuse University, NY, USA
B.S.	July 2006	University of Sci. & Tech. of China, Hefei, China

PUBLICATIONS:

Journal

1. G. Xu, W. Liu and B. Chen, "The common information of N dependent random variables", in preparation.

2. W. Liu, M. Chen, B. Chen, and J. Matyjas, "Quantization for distributed testing of independence", submitted to International Journal of Distributed Sensor Networks, 2011.

3. W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources", submitted to IEEE Trans. Information Theory, revised 2010.

Conference

1. G. Xu, W. Liu and B. Chen, "Wyner's common information for continuous random variables - a lossy source coding interpretation", Proc. Annual Conference on Information Sciences and Systems (CISS 2011), Baltimore, MD, March, 2011.

2. W. Liu, C. Duan, T. Koike-Akino, Y. Wang, R. Annavajjala and J. Zhang, "A Practical Secure Wireless Transmission Scheme", submitted to The Fifth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2011).

3. W. Liu, G. Xu and B. Chen, "The Common Information of N Dependent Random Variables", Proc. Annual Allerton Conference on Communication, Control, and Computing, Montecillo, IL, Oct. 2010 (invited).

4. M. Chen, W. Liu, B. Chen and J. Matyjas, "Quantization for distributed testing of independence", Proc. International Conference on Information Fusion (Fusion2010), Edinburgh, UK, July 2010.

5. W. Liu and B. Chen, "Communicating correlated sources over interference channels: the lossy case", Proc. IEEE International Symposium on Information Theory (ISIT2010), Austin, TX, June 2010.

6. W. Liu and B. Chen, "Communicating correlated Gaussian sources over Gaussian Z interference channels", Proc. IEEE Globecom, Honolulu, Hawaii, December 2009.

7. W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources", Proc. Annual Allerton Conference on Communication, Control, and Computing, Montecillo, IL, Oct. 2009 (invited).

8. W. Liu and B. Chen, "Message transmission and state estimation over Gaussian broadcast channels", Proc. Annual Conference on Information Sciences and Systems (CISS2009), Baltimore, MD, March, 2009.

9. W. Liu and B. Chen, "Wiretap channel with two-sided state information", Proc. 41st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2007.

AWARDS AND HONORS

- Graduate Research Assistantship, Syracuse University, 2006-2011
- Graduate School Master Prize, Syracuse University, 2010
- Graduate Teaching Assistantship, Syracuse University, 2007
- Outstanding Student Scholarship, University of Sci. & Tech. of China, 2002-2006