

STATE-DEPENDENT NETWORKS WITH
SIDE INFORMATION AND PARTIAL STATE RECOVERY

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Shiva Prasad Kotagiri, B.E., M.S.

J. Nicholas Laneman, Director

Graduate Program in Electrical Engineering

Notre Dame, Indiana

December, 2007

© Copyright by
Shiva Prasad Kotagiri
2007
All Rights Reserved

STATE-DEPENDENT NETWORKS WITH
SIDE INFORMATION AND PARTIAL STATE RECOVERY

Abstract

by

Shiva Prasad Kotagiri

In many communication scenarios, the communicating parties typically have some knowledge or attempt to learn about the time-varying environment or the channel over which communication takes place. To understand these models, it is important to study the fundamental performance limits of state-dependent channels whose probabilistic input-output relationship depends on a time-varying random parameter called channel state. This channel state could be either fading in a wireless environment, interference, or host signal in information embedding, etc. In this thesis, we focus on studying state-dependent network models from an information theoretic perspective when the side information is available at some encoders and state recovery is considered at some decoders. In general, the side information could be either exact channel state, noisy channel state, or channel output feedback. In this thesis, we assume that the side information is exact channel state. We study single-user state-dependent models and multi-user state-dependent models such as multiple access channels (MAC) and broadcast channels (BC).

For the state-dependent MAC with non-causal side information at some encoders and without state recovery at the decoder, we study bounds on the capacity region in the case of independent messages and derive the capacity region in the case of

dependent or degraded messages. For the Gaussian case, we develop a coding scheme based on both dirty paper coding and state cancellation to obtain the inner bound on the capacity region.

For the state-dependent MAC with different non-causal side information at different encoders, we derive inner bounds on the capacity region for the case of no state recovery at the decoder and the case of some, but not all, state recovery at the decoder. For this model, we also study bounds on the capacity region for the case of all state recovery at the decoder. It turns out that the inner and outer bounds meet if all state signals are independent. Consequently, we obtain the capacity region for this case.

For the state-dependent BC with non-causal encoder side information, we consider lossless state recovery at some decoders. In the two decoder model, we study bounds on the capacity region in the case of state recovery at only the better decoder. We also study the capacity region for the case of state recovery at all decoders and the case of state recovery at only the worse decoder. Finally, we consider lossy or partial state recovery in single-user state-dependent models without side information and study bounds on the capacity region for a given state-recovery distortion constraint.

To Parents, Sisters, and Brothers-in-Law

CONTENTS

TABLES	vi
FIGURES	vii
ACKNOWLEDGMENTS	ix
CHAPTER 1: INTRODUCTION	1
1.1 General State-Dependent Network Model	1
1.2 Key Issues	4
1.3 Outline	6
CHAPTER 2: SINGLE-USER STATE-DEPENDENT MODELS	9
2.1 State Neither at the Encoder Nor at the Decoder	10
2.2 State only at the Decoder	12
2.3 State only at the Encoder	13
2.3.1 Causal Case	13
2.3.2 Non-Causal Case	14
2.4 State at the Encoder and the Decoder	19
2.5 Applications	20
2.6 Summary	22
CHAPTER 3: MULTI-USER STATE-DEPENDENT MODELS	24
3.1 Multiple Access Channel	24
3.1.1 State Neither at the Encoders Nor at the Decoder	25
3.1.2 State only at the Decoder	27
3.1.3 State only at the Encoders	28
3.1.4 State at the Encoders and the Decoder	33
3.2 Broadcast Channel	34
3.2.1 State Neither at the Encoder Nor at the Decoders	35
3.2.2 State only at the Decoders	37
3.2.3 State only at the Encoder	38
3.2.4 State at the Encoder and the Decoders	41
3.3 Other Models	42

3.4	Summary	42
CHAPTER 4: STATE-DEPENDENT MULTIPLE ACCESS CHANNELS WITH SIDE INFORMATION AT SOME ENCODERS		
4.1	Independent Messages	45
4.1.1	Discrete Memoryless Case	45
4.1.2	Gaussian Memoryless Case	52
4.2	Dependent or Degraded Messages	59
4.3	Summary	60
CHAPTER 5: STATE-DEPENDENT MULTIPLE ACCESS CHANNELS WITH ENCODER SIDE INFORMATION AND RECOVERY OF SOME STATES		
5.1	Recovery of Neither Host or State	65
5.2	Recovery of One Host or State	65
5.3	Recovery of Both Hosts or States	66
5.4	Summary	68
CHAPTER 6: STATE-DEPENDENT BROADCAST CHANNELS WITH ENCODER SIDE INFORMATION AND STATE RECOVERY AT SOME DECODERS		
6.1	No State or Host Recovery	73
6.2	State or Host Recovery at the Better Decoder	74
6.3	State or Host Recovery at Both Decoders	75
6.4	State or Host Recovery at the Worse Decoder	76
6.5	Summary	76
CHAPTER 7: MODEL EXTENSIONS		
7.1	Partial State Recovery in State-Dependent Models	77
7.1.1	Single-User Models	78
7.1.2	Multi-User Models	84
7.2	State-Dependent Models with Noisy State Information	87
7.3	Summary	91
CHAPTER 8: CONCLUSIONS		
8.1	Contributions	92
8.2	Future Directions	93
APPENDIX A:		
A.1	Notation	95
A.2	Entropy and Mutual Information	95
A.3	Definitions for Chapter 2 and Chapter 3.	96
A.3.1	Single-User Models	96
A.3.2	Multiple Access Channels	97

A.3.3 Broadcast Channels	97
A.4 Strong Typicality	97
APPENDIX B:	99
B.1 Proof of Theorem 1	99
B.2 Proof of Theorem 2	103
B.3 Proof of Theorem 5	107
B.3.1 Achievability	107
B.3.2 Converse	110
APPENDIX C:	114
C.1 Proof of Theorem 6	114
C.2 Proof of Theorem 7	118
APPENDIX D:	124
D.1 Proof of Theorem 8	124
D.2 Proof of Theorem 9	127
D.3 Proof of Theorem 10	130
BIBLIOGRAPHY	137

TABLES

1.1	VARIOUS IMPORTANT ISSUES CONSIDERED IN SINGLE-USER AND MULTI-USER STATE-DEPENDENT MODELS SUCH AS MULTIPLE ACCESS CHANNELS AND BROADCAST CHANNELS. . .	6
-----	---	---

FIGURES

1.1	Block diagram for a state-dependent network	3
2.1	State-dependent single-user model.	10
2.2	Illustration of Gel'fand-Pinsker codebooks.	15
2.3	Illustration of dirty paper coding.	17
3.1	Block diagram of a state-dependent multiple access channel	25
3.2	Illustration of the capacity region for the additive state-dependent Gaussian multiple access channel with non-causal side information at all encoders.	32
3.3	Block diagram of a state-dependent broadcast channel.	35
4.1	Block diagram of a state-dependent multiple access channel with asymmetric encoder side information.	44
4.2	A numerical example for the binary noiseless multiple access channel with $p_1 = 0.1$, $p_2 = 0.4$, $q = 0.2$	51
4.3	Gaussian multiple access channel with channel state information known at one encoder.	52
4.4	An achievable region for Gaussian multiple access channel with $P_1 = 15$, $P_2 = 50$, $Q = 20$, and $N = 60$	56
4.5	An achievable region for Gaussian multiple access channel with $P_1 = 20$, $P_2 = 50$, $Q = 20$, and $N = 60$	57
5.1	Block diagram of state-dependent multiple access channel with two components of state.	62
6.1	Block diagram for state-dependent broadcast channel.	70
7.1	Block diagram of single-user state-dependent model.	78

7.2	Information rate-distortion trade-off for $P = 1$, $Q = 1$ and $N = 1$. . .	80
7.3	Block diagram of state-dependent multi-access model.	85
7.4	Block diagram of a state-dependent broadcast channel.	86
7.5	Block diagram of a single-user, state-dependent model with two-sided side information.	88

ACKNOWLEDGMENTS

Since this thesis is a long time effort, numerous people have provided help, encouragement, and support during this time. I am grateful to you for your support and encouragement if I neglect mentioning some names in this acknowledgments.

First, I would like to thank my advisor J.Nicholas Laneman for providing me good insights in research, encouragement and support, and good perspective on graduate program. He had provided lots of freedom and support for me to choose research topic. He was always friendly with me and has helped me to develop teaching and presentation skills.

I would also like to thank Dr. Daniel Costello and Dr. Martin Haenggi for serving as candidacy committee members, and Dr. Tom Fuja and Dr. Abdellatif Zaidi for serving as doctoral defense committee members. I am very grateful to Dr. Bill Weeks because he inspired me during my masters program to pursue Ph.D. My summer internship at LG/Zenith Electronics corp. had helped me to understand concepts in Communications and Signal processing. During internship, I thoroughly enjoyed discussions with my supervisors Wayne Bretl and Ajay Gupta which helped me to get good understanding of practical systems.

I would like to thank my colleagues Wenyi Zhang, Dequiang Chen, Brian Dunn, Michael Dickens, and other group members for having interesting discussions which helped me to do good research. I would like to thank Amaresh, Jagadish, Krishnan, Sundeep, Sunil, and Shashank for having lots of enjoyable discussions with me during research breaks. I am grateful to my roommates Kameshwar Yadavalli, Ajit Nimbalker, and Rajkumar Sankaralingam who made my stay at 29 O'hara Grace

very comfortable and enjoyable with interesting non-technical discussions. I would like to specially thank 22 O'Hara Grace roommates Radha-Krishna Ganti, Srinath Puducheri, Mahesh Mahadevan, and Sundaram Vanka from whom I have learned lots of technical and non-technical concepts by having discussions with them during lunch and dinner everyday.

My childhood friends have always been very special to me. They have never disappointed me till now whenever I needed their support and help. I am viewing this as opportunity to thank them and recognize their help. I would like to thank my childhood friends Kalyan, Rakesh, Madhav, Pavan, Balaraju, and Sanjeev who always encouraged and helped me to come to this level. I am also grateful to my friends Nanda kumar, Vidyadhar, Murali Mohan, Naveen Bejugam, Suneetha Mare, Suman Penugonda, and Ramesh Alleti who encouraged me a lot during this Ph.D. program.

Last, but not least, I owe a lot of respect and thanks to my parents, sisters, brothers-in-law, brother, and sister-in-law who have given me a lot of support and freedom in my career. I would like to thank my nephews and niece who have always entertained me a lot.

CHAPTER 1

INTRODUCTION

In many communication models, the communicating parties typically have some knowledge or attempt to learn about the time-varying environment or the channel over which communication takes place. To understand such scenarios, it is important to study the fundamental performance limits of channels whose probabilistic input-output relationship depends on a time-varying random parameter called the channel state. The study of state-dependent single-user channel models was initiated by Shannon [38]. Several groups of researchers study single-user state-dependent models [38, 2, 37, 30, 19, 48, 9, 17, 10]. Although much is known about single-user state-dependent models, the theory is less well developed for multi-user state-dependent models. In the following sections, we discuss several interesting state-dependent network models under various scenarios, and highlight important issues arising in these models.

1.1 General State-Dependent Network Model

A state-dependent multi-terminal network is shown in Figure 1.1. There are m nodes in the network. In this network, node i has an associated transmitted vector \mathbf{X}_i^n and a received vector \mathbf{Y}_i^n , where $i \in \{1, 2, \dots, m\}$ ¹. Node i wants

¹ \mathbf{X} is random variable whose alphabet is \mathcal{X} and sample value is \mathbf{x} . \mathbf{X}_i^n denotes as $\{\mathbf{X}_{i,1}, \mathbf{X}_{i,2}, \dots, \mathbf{X}_{i,n}\}$

to send messages $W_i = \{W_{i,1}, \dots, W_{i,m}\}$ to the remaining nodes, where $W_{i,j} \in \mathcal{W}_{i,j}$ is the message intended for node j with $\mathcal{W}_{i,j} = \{1, 2, \dots, \lceil 2^{nR_{i,j}} \rceil\}$ and rate $R_{i,j}$. The state-dependent channel is represented by a conditional probability law $p(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m, \mathbf{s})$. This probability transition function captures the effects of propagation, interference, and noise in the network. In this model, the channel outputs $(Y_1^n, Y_2^n, \dots, Y_m^n)$ are controlled by a random parameter or channel state sequence S^n along with the channel inputs $(X_1^n, X_2^n, \dots, X_m^n)$. This random parameter or channel state sequence captures fading in a wireless environment, interference from other users [3], or the host sequence in information embedding and data hiding applications [5, 6, 34, 8, 21]. In these models, we assume that the side information T_i^n about the channel is either *causally* or *non-causally* known at node i . This side information could be the exact channel state, noisy channel state, or some form of feedback. If there is no side information at node i , i.e., $T_i^n = \emptyset$, then node i is called *uninformed* node. If the side information $T_i^n \neq \emptyset$ is available at node i , then node i is called *informed* node.

If it is informed, node i generates $X_{i,k}$ at time instant k from the available messages W_i , the side information, and the past received symbols $(Y_{i,1}, Y_{i,2}, \dots, Y_{i,k-1})$. Otherwise, node i ignores the side information while encoding its messages. Node i decodes the messages \hat{W}_i intended for it from the channel observation Y_i^n and the side information T_i^n if it is an informed node. Otherwise, node i may attempt to estimate the state \hat{S}_i^n satisfying a given distortion constraint according to a prescribed distortion measure. From the point of view of encoding, the informed nodes can help the uninformed nodes in terms of the communication rates and channel estimation using the available side information. But, from the point of view of decoding, the informed nodes view the side information as another channel observation. It is interesting to study these models from the point of view of encoding with the side

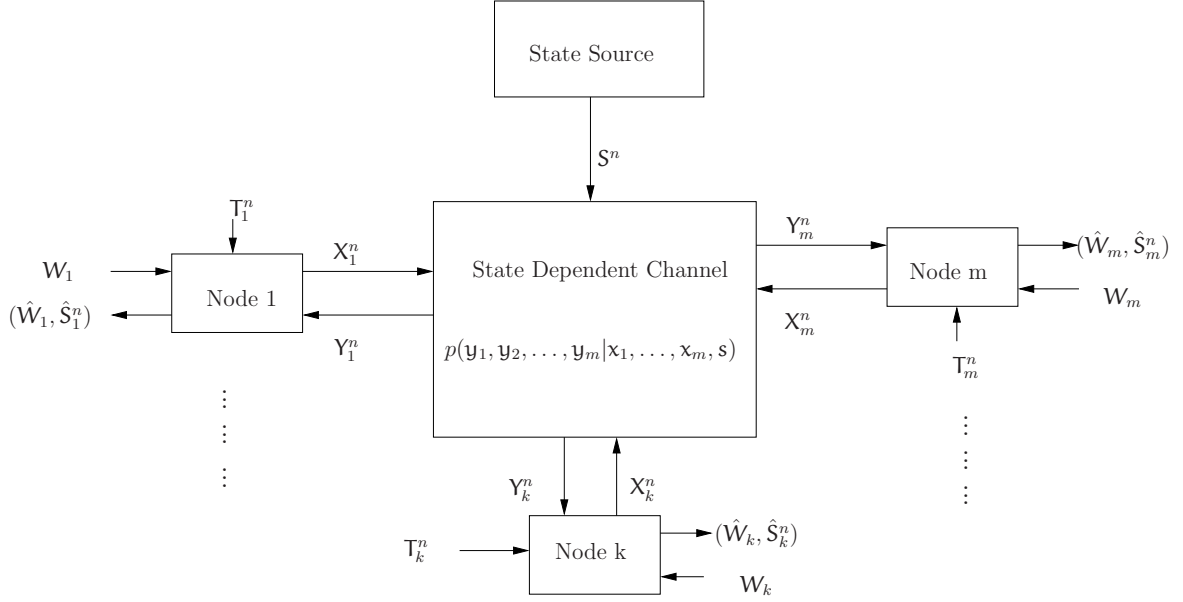


Figure 1.1. Block diagram for a state-dependent network

information and from the point of view of the channel estimation at the decoders. To understand these models, it is important to study the fundamental performance limits of these models such as achievable communications rates from an information theoretic perspective.

In this thesis, we focus mainly on studying communication models containing nodes that can either transmit or receive the information. We call the nodes that can only transmit *encoders* and the nodes that can only receive *decoders*. We also assume that the side information T_i^n is either the channel state S^n or \emptyset . From the above general state-dependent network model, we can derive state-dependent communication models containing encoders and decoders in the following situations.

- The side information is known only at the encoders [38, 30, 19, 48, 9, 46, 45, 23, 18, 24, 25, 27, 28, 17].
- The side information is known only at decoders.

- The side information is known at the encoders and the decoders [2, 37, 10].
- The side information is known neither at the encoders nor at the decoders.

1.2 Key Issues

In the state-dependent models, the following issues are very important to be considered.

- Encoding of the message using side information at the informed encoders in the presence of some uninformed encoders and uninformed decoders in the model.
- Recovery or estimation of the channel state at the uninformed decoders in the presence of informed encoders, uninformed encoders, and informed decoders.

State recovery could be either *lossless* or *lossy*. In lossless recovery case, the state estimate should be exactly same as the actual channel state. In lossy recovery case, the state estimate at the decoder should be “close” to the actual channel state according to some distortion measure $d(\cdot, \cdot)$. To make the estimate “close” to the actual channel state, per-letter average distortion between the estimate and the actual channel state should satisfy a given distortion constraint Δ . This distortion constraint determines how close the estimate should be with the actual channel state.

In this thesis, we consider state-dependent models with non-causal side information at some encoders. The state-dependent models with encoder side information can be used to model information embedding, data hiding, memories with defects, and related models [5, 6, 34, 8]. In the state-dependent models with some informed encoders, it is very challenging for the informed encoders to exploit the available

side information for removing the effect of channel state on their message transmission and uninformed encoder's message transmission, and helping the uninformed decoder in terms of channel state estimation or recovery. Under these constraints, single user models are studied in [38, 2, 37, 30, 19, 48, 9, 17, 10, 40] and multi-user models are studied in [46, 45, 23, 18, 24, 25, 27, 28, 48].

For some of the models considered in this thesis, lossless state recovery is considered along with the decoding of messages at the uninformed decoders. Lossless recovery of the channel state can not be accomplished in some channel models even with side information available to the encoders, because the uncertainty of the channel state exceeds the channel capacity. Lossless recovery is also not possible for channel states with continuous alphabet. To overcome both difficulties, it is important to consider *lossy* recovery of the channel state at the uninformed decoders because the partial knowledge about the channel state at the uninformed decoders could be helpful or required in some scenarios. Lossy recovery at the decoders is considered in [48, 11, 7].

In models with side information known neither at the encoders nor at the decoders, lossless state recovery can not be accomplished in most of the non-trivial models. In these scenarios, the decoders consider lossy state recovery so that the coherent coding techniques can be used for communication using the channel state estimate [50, 35, 50]. In general, information transmission and channel estimation are done in time-division mode. But, it is important to study the trade-off between the information transmission rate and the state-recovery distortion to understand resource overhead required for channel estimation. The estimated channel can also be fed back to the uninformed encoders so that they can adapt their encoding schemes suitable for the state of the channel.

Table 1.1

VARIOUS IMPORTANT ISSUES CONSIDERED IN SINGLE-USER AND
MULTI-USER STATE-DEPENDENT MODELS SUCH AS MULTIPLE ACCESS
CHANNELS AND BROADCAST CHANNELS.

	Single-User	MAC	BC
Causal SI at all encoders	[38, 2]	[40]	[40]
Non-causal SI (NC-SI) at all encoders	[17, 9]	[18, 24]	[18, 24]
NC-SI at some (not all) encoders	[39]	In this thesis	[12]
NC-SI at all encoders & lossless state recovery	[53]	In this thesis	In this thesis
NC-SI at some (not all) encoders & lossless state recovery	N/A	In this thesis	N/A
NC-SI at some (not all) encoders & lossy state recovery	In this thesis	Not considered	Not considered

In Table 1.2, various important issues for single-user state-dependent models, state-dependent multiple access channels, and state-dependent broadcast channels are presented; the problems that have already been considered, the problems that are considered in this thesis, the problems that have not been considered so far are mentioned.

1.3 Outline

An outline for the remainder of the thesis is as follows. Chapter 2 discusses single-user state-dependent models in various cases that have already been studied, presenting important results, random coding techniques, and their applications. Chapter 3 summarizes history of multi-user state-dependent models such as multiple access channels, broadcast channels, and so forth, highlighting important cases,

important results, and random coding techniques.

Chapter 4 focuses on studying a state-dependent multiple access channel (MAC) with side information available at some encoders and without state recovery at the decoder in the discrete memoryless case and the Gaussian memoryless case. Section 4.1 and Section 4.2 focus on studying the model in the case of independent messages and in the case of degraded messages, respectively. In Section 4.1, we derive inner and outer bounds for the capacity region of the discrete memoryless model, specialize the inner bound for binary noiseless case, and also obtain inner and outer bounds for the capacity region of the Gaussian memoryless model. Finally, in Section 4.2, we obtain the capacity region for the discrete memoryless model with degraded messages.

Chapter 5 focuses on studying a state-dependent MAC with different side information at different encoders and lossless recovery of some states at the decoder. In Section 5.1 and Section 5.2, we derive inner bounds for the capacity region in the case of no state recovery and the case of one state recovery, respectively. In Section 5.3, we derive inner and outer bounds for the capacity region and also obtain the capacity region in a special case in which side information sequences at different encoders are independent.

Chapter 6 focuses on studying a degraded broadcast channel with side information available at the encoder and lossless recovery of state at *some* decoders. In Section 6.1, we derive inner and outer bounds for the capacity region when state recovery is not considered at all decoders. In Section 6.2, we also obtain inner and outer bounds for the capacity region when state recovery is considered only at the better decoder. We derive the capacity region for this model in the case of state recovery at all decoders and in the case of state recovery only at the worse decoder in Section 6.3 and Section 6.4, respectively.

Chapter 7 discusses few extensions such as lossy state recovery in state-dependent models without side information, and state-dependent models with side information partially correlated to the channel state. Finally, Chapter 8 concludes the thesis, highlighting contributions and future directions.

CHAPTER 2

SINGLE-USER STATE-DEPENDENT MODELS

In this chapter, we present background material on single-user state-dependent models, and discuss some important results and random coding schemes. We do not give converse proofs for capacity theorems here because converse proofs do not give any insight about encoding and decoding schemes. But, the converse shows the optimality of rates achieved using the encoding and decoding schemes proposed in achievability proof.

Figure 2.1 illustrates a state-dependent model in which $X \in \mathcal{X}$ is the channel input, $S \in \mathcal{S}$ is the channel state, and $Y \in \mathcal{Y}$ is the channel output which is controlled by X and S according to a memoryless probability law ¹ $p(\mathbf{y}|\mathbf{s}, \mathbf{x})$. It is also assumed that the channel state is memoryless and drawn with probability law $p(\mathbf{s})$. The encoder wants to reliably transmit a message $W \in \mathcal{W} = \{1, 2, \dots, \lceil 2^{nR} \rceil\}$ to the decoder in n channel uses, where R is the rate of transmission. It is also assumed that W is independent of the channel state S . The decoder forms an estimate \hat{W} of the message W from the channel output Y^n .

An important case to which we will often specialize is the additive state-dependent Gaussian model whose output is given as ²

$$Y^n = X^n + S^n + Z^n, \tag{2.1}$$

¹Calligraphic letters are used to denote the random variables alphabet set, e.g., $X \in \mathcal{X}$.

²Addition in $Y^n = X^n + S^n + Z^n$ is element wise addition.

where the channel input \mathbf{X}^n should satisfy the average power constraint $\frac{1}{n} \sum_{j=1}^n X_j^2 \leq P$; the channel state \mathbf{S}^n is zero mean Gaussian vector with covariance matrix³ $\mathbf{Q}\mathbf{I}_n$; and the noise \mathbf{Z}^n is zero mean Gaussian with covariance matrix $N\mathbf{I}_n$. The noise \mathbf{Z}^n and the channel state \mathbf{S}^n are mutually independent. All alphabets are discrete and continuous in discrete and Gaussian cases, respectively. Unless indicated otherwise, results in the discrete memoryless (DM) case can readily be extended to memoryless models with discrete time and continuous alphabets using standard techniques [16].

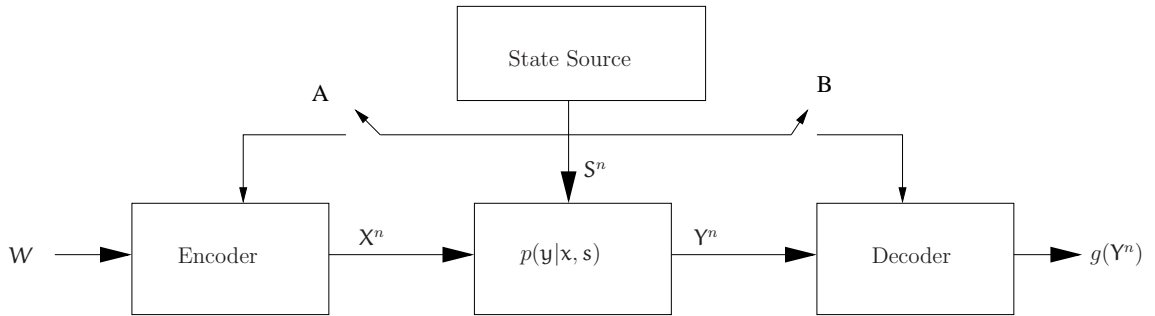


Figure 2.1. State-dependent single-user model.

2.1 State Neither at the Encoder Nor at the Decoder

In this case, switches A and B are open in Figure 2.1. For this model, the encoder is a function $f^n : \mathcal{W} \rightarrow \mathcal{X}^n$ and the decoder is a function $g^n : \mathcal{Y}^n \rightarrow \mathcal{W}$. For the DM case, the capacity⁴ is given by⁵ [39]

$$C = \max_{p(x)} \mathbb{I}(\mathbf{X}; \mathbf{Y}).$$

The following simple random coding technique is used to prove existence of codes that achieve the capacity:

³ \mathbf{I}_n denotes an identity matrix of size n .

⁴The capacity is defined in Appendix A for single-user models in various cases.

⁵ $\mathbb{I}(\cdot; \cdot)$ denotes the mutual information and $\mathbb{H}(\cdot)$ denotes the entropy; $\mathbb{I}(\cdot; \cdot)$ and $\mathbb{H}(\cdot)$ are defined in Appendix A.

- For each message $m \in \mathcal{W}$, generate random codeword $\mathbf{X}^n(m)$ whose elements are independently drawn with probability law $p(\mathbf{x})$. The codebook is revealed to both the encoder and the decoder.
- The encoder chooses the codeword $\mathbf{X}^n(W)$ to transmit the message W .
- The decoder looks for a codeword $\mathbf{X}^n(m)$, $m \in \mathcal{W}$ that is jointly typical⁶ with the channel output \mathbf{Y}^n . If such a codeword exists and is unique, then the decoder declares its index as an estimate of the transmitted message. Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R < \mathbb{I}(\mathbf{X}; \mathbf{Y})$.

The capacity of the additive state-dependent Gaussian model is given as⁷

$$C = \frac{1}{2} \log \left(1 + \frac{P}{Q + N} \right).$$

In this case, since the knowledge of channel state is known neither at the encoder nor at the decoder, then the additive channel state \mathbf{S} is treated simply as noise in encoding and decoding the message. For each message $m \in \mathcal{W}$, the encoder generates the codeword \mathbf{X}^n whose elements are independently drawn according to probability distribution⁸ $\mathcal{N}(0, P)$. The decoder applies minimum Euclidean-distance decoding to obtain the message from the channel output \mathbf{Y}^n , i.e., it chooses the codeword $\mathbf{X}^n(m)$, $m \in \mathcal{W}$ that is closest to \mathbf{Y}^n in the sense of Euclidean distance.

In the above model, the decoder only considers recovery of the message, but, in many communication models, the decoder may also want to estimate the channel state for a variety of reasons. If the decoder must also partially recover the channel state according to a prescribed distortion measure, there is trade-off between the

⁶Typicality is formally defined in Appendix A.4.

⁷ \log is logarithm with base 2.

⁸ $\mathcal{N}(0, P)$ is Gaussian distribution with zero mean and variance P .

message transmission rate from the encoder and the state-recovery distortion constraint with which the channel state is estimated or recovered at the decoder. We study the optimal trade-off between the message transmission rate and the state-recovery distortion in Chapter 7.

2.2 State only at the Decoder

In this case, switch A is open, and switch B is closed in Figure 2.1. For this model, the encoder is a function $f^n : \mathcal{W} \rightarrow \mathcal{X}^n$ and the decoder is a function $g^n : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{W}$. In the DM case, the capacity is given by

$$C = \max_{p(x)} \mathbb{I}(\mathbf{X}; \mathbf{Y}, \mathbf{S}).$$

Encoding is identical to that in Section 2.1 for the previous case. In terms of decoding, the only difference between the previous case and this case is that the decoder views the available channel state \mathbf{S}^n as another channel observation along with the channel output \mathbf{Y}^n , i.e., the effective channel output is the pair $(\mathbf{Y}^n, \mathbf{S}^n)$.

In the simple Gaussian case given in (2.1), the decoder removes the effect of \mathbf{S}^n from \mathbf{Y}^n and then decodes the message from $\mathbf{Y}^n - \mathbf{S}^n$. For this model, the channel capacity is given by

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right). \quad (2.2)$$

Thus, it can be observed that the effect of channel state can be completely removed if the channel state is additive and known at the decoder. In general, the effect of channel state can not be completely removed for all models even though the channel state is available at the decoder. We also note that causality is irrelevant for block decoding techniques in the case of decoder side information.

2.3 State only at the Encoder

In this case, switch A is closed and switch B is open in Figure 2.1. The channel state can be either *causally* or *non-causally* known at the encoder.

2.3.1 Causal Case

For this case, the encoder is a sequence of functions $f_i : \mathcal{W} \times \mathcal{S}^i \rightarrow \mathcal{X}$ for $i = 1, 2, \dots, n$ and the decoder is a function $g^n : \mathcal{Y}^n \rightarrow \mathcal{W}$. For the DM case, the capacity is given by [38, 2]

$$C = \max_{p(\mathbf{u})} \mathbb{I}(\mathbf{U}; \mathbf{Y}), \quad (2.3)$$

where $\mathbf{U} \in \mathcal{U}$, and \mathcal{U} is the set of random vectors of length $|\mathcal{S}|$ with elements in \mathcal{X} with $|\mathcal{U}| = |\mathcal{X}|^{|\mathcal{S}|}$. The random coding scheme that achieves the capacity for this model is a set of n length vectors from alphabet \mathcal{U} .

- For each message $m \in \mathcal{W}$, the encoder generates the codeword $\mathbf{U}^n(m)$ whose elements are independently drawn according to $p(\mathbf{u})$. This codebook is revealed to both the encoder and the decoder.
- The encoder wants to send the message W . The encoder, provided with S_i at time i , chooses $U_i(W, S_i)$ as the channel input at time instant i , where $U_i(W, S_i)$ is the S_i th element of the vector $U_i(W)$.
- The decoder looks for a codeword $\mathbf{U}^n(m)$, $m \in \mathcal{W}$, that is jointly typical with the channel output \mathbf{Y}^n . If such a codeword exists and is unique, then the decoder declares the index of codeword as estimate of the transmitted message.

Shannon's proof technique for the discrete case leading to (2.3) does not offer much insight for more general channels, including the simple Gaussian case. We also

observe that the state-dependent models with causal encoder side information and recovery of state at the decoder has not been considered.

2.3.2 Non-Causal Case

For these models, the decoder may or may not consider the channel state recovery. In the next sections, we discuss these models without state recovery and with state recovery.

No State Recovery

For this case, the encoder is a function $f^n : \mathcal{W} \times \mathcal{S}^n \rightarrow \mathcal{X}^n$ and the decoder is a function $g^n : \mathcal{Y}^n \rightarrow \mathcal{W}$. For the DM case, the capacity is given by [17]

$$C = \max_{p(\mathbf{u}|\mathbf{s}), h: \mathcal{U} \times \mathcal{S} \rightarrow \mathcal{X}} [\mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})], \quad (2.4)$$

where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}|$, and if $\mathbf{x} = h(\mathbf{u}, \mathbf{s})$, $p(\mathbf{s}, \mathbf{u}, \mathbf{x}, \mathbf{y}) = p(\mathbf{s})p(\mathbf{u}|\mathbf{s})p(\mathbf{y}|\mathbf{s}, \mathbf{x})$, otherwise, it is zero.

The above result and coding scheme to achieve it are very important to understand most of the problems we study. Let us discuss the random coding scheme used to achieve (2.4) [17]. We call this coding scheme *Gel'fand-Pinsker* (GP) coding throughout the thesis.

- For each message $m \in \mathcal{W}$, generate a *bin* of $\lceil 2^{nR_0} \rceil$ codewords $\mathbf{U}^n(m, j)$ whose elements are generated according to distribution $p(\mathbf{u})$, where $j \in \{1, 2, \dots, \lceil 2^{nR_0} \rceil\}$ as shown in Figure 2.2. This codebook is revealed to both the encoder and the decoder. In this case, there is more than one codeword for each message. The available channel state \mathbf{S}^n determines which codeword should be chosen for a given message.
- Given the message $W \in \mathcal{W}$ and the channel state \mathbf{S}^n , the encoder finds the codeword $\mathbf{U}^n(W, j)$ in bin W that is jointly typical with \mathbf{S}^n . If no such j

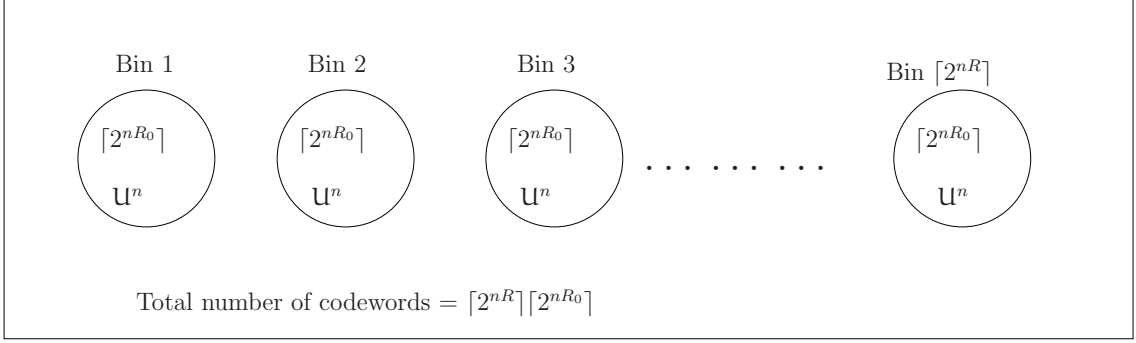


Figure 2.2. Illustration of Gel'fand-Pinsker codebooks.

exists, then the encoder declares an error. Otherwise, the encoder generates the channel input $\mathbf{X}^n = h^n(\mathbf{U}^n(\mathcal{W}, j), \mathbf{S}^n)$ ⁹. The probability of encoding error goes to zero as $n \rightarrow \infty$ provided $R_0 > \mathbb{I}(\mathbf{U}; \mathbf{S})$.

- The decoder finds $\mathbf{U}^n(m, j)$ that is jointly typical with the channel output \mathbf{Y}^n , where $m \in \mathcal{W}$ and $j \in \{1, 2, \dots, \lceil 2^{nR_0} \rceil\}$. If such a sequence \mathbf{U}^n exists and is unique, the decoder declares its bin index as the estimate for the transmitted message; otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R + R_0 < \mathbb{I}(\mathbf{U}; \mathbf{Y})$. Thus, the overall probability of error goes to zero as $n \rightarrow \infty$ provided $R < \mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})$.

For the state-dependent Gaussian model (2.1) with non-causal encoder side information, the capacity is given by [9]

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right), \quad (2.5)$$

which is same as the capacity if the channel state is known at the decoder. Interestingly, the encoder does not use its transmit power to explicitly cancel the channel state in order to achieve (2.5). Encoding and decoding are similar to those of the

⁹ $h^n(\mathbf{U}^n(\mathcal{W}, j), \mathbf{S}^n) = \{h(\mathbf{U}_1(\mathcal{W}, j), \mathbf{S}_1), h(\mathbf{U}_2(\mathcal{W}, j), \mathbf{S}_2), \dots, h(\mathbf{U}_n(\mathcal{W}, j), \mathbf{S}_n)\}$.

DM case. The new observation by Costa in [9] is an auxiliary random variable that achieves the capacity in (2.4); specifically, the optimal auxiliary random variable is $\mathbf{U} = \mathbf{X} + \alpha\mathbf{S}$, where \mathbf{X} is zero mean Gaussian with variance P and independent of \mathbf{S} , and α is a real parameter. Using this choice of auxiliary random variable, $\mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})$ is computed and optimized over α . It turns out that the optimal choice yielding (2.5) is $\alpha^* = \frac{P}{P+N}$. The GP coding scheme with the above mentioned \mathbf{U} is often called *dirty paper coding* (DPC).

DPC can be explained with the help of Figure 2.3. In this figure, four possible messages are represented by four different symbols, and each symbol represents an auxiliary codeword \mathbf{U}^n . For each message, there are more than one auxiliary codeword. These auxiliary codewords are spread over the entire \mathcal{S}^n space. For a given message \mathbf{W} and channel state \mathbf{S}^n , the channel input \mathbf{X}^n is chosen as the minimum difference between all auxiliary codewords representing the message \mathbf{W} and $\alpha\mathbf{S}^n$. We should choose these auxiliary codewords such that \mathbf{X}^n satisfies the average power constraint P for all \mathbf{S}^n and all messages $\mathbf{W} \in \mathcal{W}$. The existence of such a codebook is shown using typicality and random coding arguments.

For this model, no additional converse proof is needed because an obvious outer bound for the capacity is (2.2), and the inner bound for the capacity meets this obvious outer bound. It is very rare for such obvious outer bounds to be tight, as we will see throughout the thesis. Single-user state-dependent models with partial state at the encoder is considered in [36].

State Recovery

In this case, the decoder must recover the state as well as decode the message. State recovery could be either *lossless* or *lossy*. In this case, the encoder is a function $f^n : \mathcal{W} \times \mathcal{S}^n \rightarrow \mathcal{X}^n$ and the decoder is a function $g^n : \mathcal{Y}^n \rightarrow \mathcal{W} \times \hat{\mathcal{S}}^n$.

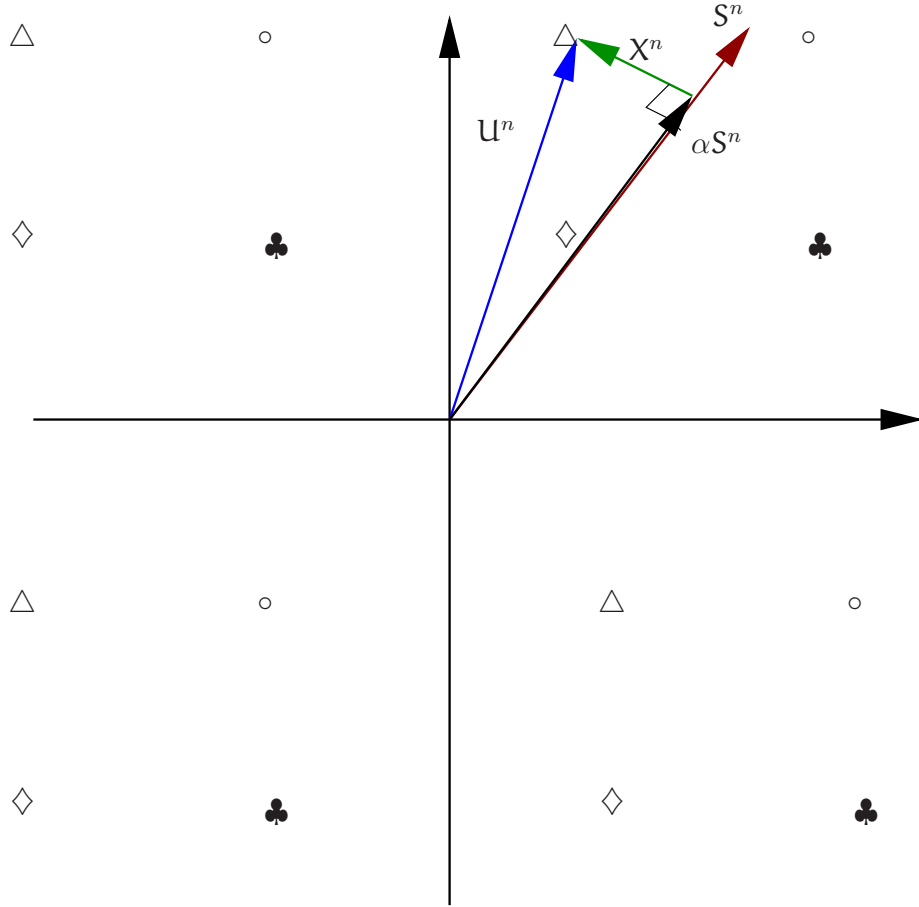


Figure 2.3. Illustration of dirty paper coding.

We first focus on lossless recovery of the channel state at the decoder. In lossless recovery case, alphabet $\hat{\mathcal{S}}$ should be same as alphabet \mathcal{S} . For the DM models, then the capacity is given by [53]

$$C = \max_{p(x|s)} [\mathbb{I}(X, S; Y) - \mathbb{H}(S)]. \quad (2.6)$$

Encoding and decoding schemes used to achieve (2.6) are quite different from GP coding. For this model, the following superposition coding scheme is used to achieve (2.6).

- Generate codewords $X^n(S^n, m)$ according to $p(x|s)$ for each typical S^n and all

$m \in \mathcal{W}$. The codebooks are revealed to both the encoder and the decoder

- The encoder, provided with \mathbf{S}^n , sends $\mathbf{X}^n(\mathbf{S}^n, \mathbf{W})$ to transmit the message \mathbf{W} .
- The decoder looks for $\mathbf{X}^n(\hat{\mathbf{S}}^n, \hat{\mathbf{W}})$ that is jointly typical with \mathbf{Y}^n for all $\hat{\mathbf{W}} \in \mathcal{W}$ and typical sequences $\hat{\mathbf{S}}^n$. If such a sequence exists and is unique, then the estimates of message and the host sequence are $\hat{\mathbf{W}}$ and $\hat{\mathbf{S}}^n$, respectively. Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R < \mathbb{I}(\mathbf{S}, \mathbf{X}; \mathbf{Y}) - \mathbb{H}(\mathbf{S})$.

If the state source $p(\mathbf{s})$ is such that (2.6) becomes negative, then lossless state recovery is impossible, because the state source entropy exceeds the channel capacity. In such scenarios, lossy state recovery would be useful to have partial knowledge about the channel state and reduce the constraints in the problem. In lossy recovery case, the decoder estimates the channel state $\hat{\mathbf{S}}^n$ according to some distortion measure $d(\cdot, \cdot)$ satisfying a given distortion constraint Δ , i.e., $\mathbb{E}(d(\mathbf{S}^n, \hat{\mathbf{S}}^n)) \leq \Delta$.

In lossy state recovery case, the capacity is not known for the DM models with encoder side information. Although the capacity $C(\Delta)$ is in general not known, it satisfies [7]

$$C(\Delta) \geq \max_{p(\mathbf{s}, \mathbf{u}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}}) \in \mathcal{P}_i(\Delta)} [\mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})], \quad (2.7)$$

where, $\mathcal{P}_i(\Delta)$ is the set of distribution functions $p(\mathbf{s}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}})$ satisfying the following conditions

- $p(\mathbf{s}, \mathbf{u}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}}) = p(\mathbf{s})p(\mathbf{u}|\mathbf{s})p(\mathbf{x}|\mathbf{s}, \mathbf{u})p(\mathbf{y}|\mathbf{s}, \mathbf{x})p(\hat{\mathbf{s}}|\mathbf{u}, \mathbf{y})$,
- $\mathbb{E}d(\mathbf{S}, \hat{\mathbf{S}}) \leq \Delta$.

For state-dependent additive Gaussian models with non-causal encoder side information, [48] obtains the capacity-distortion function $C(\Delta)$ for lossy state recovery

subject to mean-square error distortion, specifically the capacity $C(\Delta)$ is given by

$$C(\Delta) = \frac{1}{2} \log \left(1 + \frac{\gamma_{\max} P}{N} \right) \quad \text{for } \Delta \geq Q \frac{P+N}{(\sqrt{P} + \sqrt{Q})^2 + N}, \quad (2.8)$$

where

$$\gamma_{\max} = \max_{\gamma} \left\{ \gamma \in [0, 1] : Q \frac{\gamma P + N}{\left(\sqrt{(1-\gamma)P} + \sqrt{Q} \right)^2 + \gamma P + N} \leq \Delta \right\}.$$

The distortion constraint $\Delta < Q \frac{P+N}{(\sqrt{P} + \sqrt{Q})^2 + N}$ is not achievable for any rate of message transmission. From (2.8), the capacity $C(\Delta)$ increases as Δ . In random coding that achieves the above capacity, the channel input \mathbf{X}^n is decomposed into two parts, namely \mathbf{X}_w^n and \mathbf{X}_s^n with average power constraints γP and $(1-\gamma)P$, respectively, for fixed $\gamma \in [0, 1]$. \mathbf{X}_s^n is generated from the available channel state \mathbf{S}^n , i.e., $\mathbf{X}_s^n = \sqrt{\frac{(1-\gamma)P}{Q}} \mathbf{S}^n$. Then, the channel output can be written as

$$\mathbf{Y}^n = \mathbf{X}_w^n + \mathbf{X}_s^n + \mathbf{S}^n + \mathbf{Z}^n = \mathbf{X}_w^n + \left(1 + \sqrt{\frac{(1-\gamma)P}{Q}} \right) \mathbf{S}^n + \mathbf{Z}^n.$$

From the available state \mathbf{S}^n and \mathbf{W} , the encoder generates \mathbf{X}_w^n using DPC with $\alpha = \frac{\gamma P}{\gamma P + N}$. The decoder estimates the message as in DPC and generates the state estimate as the minimum mean square error (MMSE) estimate of \mathbf{S}^n from \mathbf{Y}^n .

2.4 State at the Encoder and the Decoder

In this case, switches A and B are closed in Figure 2.1. In these models, the side information, fully correlated to channel state, is *causally* or *non-causally* available at the encoder and is also known at the decoder.

For the discrete memoryless case with causal encoder side information, the capacity is given by [37, 2]

$$C = \max_{p(\mathbf{u})} \mathbb{I}(\mathbf{U}; \mathbf{Y}|\mathbf{S}), \quad (2.9)$$

where $\mathbf{U} \in \mathcal{U} = \mathcal{X}^{|\mathcal{S}|}$. The encoding scheme used to achieve (2.9) is similar to the encoding scheme with causal side information available only at the encoder. The decoding scheme is similar to that in the case of decoder side information

If the channel state is *non-causally* known at the encoder and the alphabets are finite, the capacity of this model is given by [10]

$$C = \max_{p(\mathbf{x}|\mathbf{s})} \mathbb{I}(\mathbf{X}; \mathbf{Y}|\mathbf{S}).$$

The above capacity region can be obtained by using codebooks in superposition coding scheme. The decoding scheme is similar to that in the case of state only at the decoder.

When the channel state is either *non-causally* or *causally* known at the encoder, the capacity of the additive Gaussian model is given as

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right).$$

The state-dependent models would be more interesting if the channel state is multiplicative rather than additive. For state-dependent fading models with encoder side information, it is very challenging to obtain an auxiliary random variable that maximize (2.4). DPC for fading scenarios is considered in [59].

2.5 Applications

The channel models with random parameters or channel state and non-causal channel state at the encoder can be used to study computer memories with defects. In the context of computer memories with defects, these models are also studied in [30, 19].

The channel models with random parameters or the channel state *non-causally* known at the encoder can be used to model information embedding or data hiding [5, 6, 34, 8]. Information embedding (IE) is a recent area of digital media research

with many applications, including: passive and active copyright protection (digital watermarking); embedding important control, descriptive, reference information into a given signal; steganography; backward-compatible digital upgrades of communications infrastructure; and covert communications [49, 1]. Figure 2.1 with switch A closed shows the block diagram of IE or data hiding.

In IE, the message W is embedded in to a host signal S^n (digital audio, video or images) in a manner that is essentially transparent to a conventional decoder. Thus, the embedded signal X^n should be close to the host data S^n under some prescribed distortion measure $d(\cdot, \cdot)$, i.e., $\mathbb{E}[d(S^n, X^n)] \leq \Delta$. In IE, the existence of the message in the host signal should be unknown to the conventional decoder, so the distortion constraint Δ is kept small.

Based on the recovery, in the sense of probability of error going to zero, of the message embedded in the host sequence as well as the host sequence, there are two important types of IE. They are *irreversible* IE and *reversible* IE.

- **Irreversible Information Embedding:**

In this case, the decoder is concerned with decoding only the message embedded in the host sequence from the channel output Y^n , i.e., $g(Y^n) = \hat{W}$. The irreversible information embedding capacity is given by [34]

$$C = \max_{p(\mathbf{u}|\mathbf{s}), f: \mathcal{U} \times \mathcal{S} \rightarrow \mathcal{X}: \mathbb{E}[d(\mathbf{S}, \mathbf{X})] \leq \Delta} [\mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})], \quad (2.10)$$

where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}|$, and if $\mathbf{x} = f(\mathbf{u}, \mathbf{s})$, $p(\mathbf{s}, \mathbf{u}, \mathbf{x}, \mathbf{y}) = p(\mathbf{s})p(\mathbf{u}|\mathbf{s})p(\mathbf{y}|\mathbf{s}, \mathbf{x})$, otherwise, it is zero. If we compare (2.4) and (2.10), they differ only in the set of distributions over which the maximization takes place. Indeed, encoding and decoding used to achieve (2.10) are similar to the encoding and decoding used to achieve (2.4).

- **Reversible Information Embedding :**

In this case, the decoder is concerned with the lossless recovery of the host sequence along with the decoding of the message embedded in the host sequence from the channel output Y^n i.e., $g(Y^n) = (\hat{W}, \hat{S}^n)$. The reversible information embedding capacity is given by [53]

$$C = \max_{p(x|s): E[d(S,X)] \leq \Delta} [\mathbb{I}(S, X; Y) - \mathbb{H}(S)]. \quad (2.11)$$

For this model, distortion-constrained superposition coding is used to achieve (2.11). If we compare (2.6) and (2.11), they differ only in the set of distributions over which the maximization takes place. Indeed, encoding and decoding used to achieve (2.11) are similar to the encoding and decoding used to achieve (2.6).

Information theoretic study of single-user public and private watermarking systems is considered in [41, 42, 43]. Joint IE and lossy compression is studied in [31, 32] and joint watermarking and encryption is studied in [33].

In some applications, the message embedded in the host signal should also be protected from unintended users. If unintended users know the existence of the message embedded in the host signal, they can attempt to decode the message. To protect the message from the unintended users, it is better to incorporate a secrecy constraint on the message embedded in the host sequence. Communication models with secrecy constraints are studied in [54, 13]. The capacity is not known for IE model with secrecy constraints in the DM case.

2.6 Summary

In this chapter, we briefly summarized a variety of results and coding schemes for single-user, state-dependent models. In most of the interesting cases, the capacity for these models is known. However, partial or lossy state recovery is one important

direction that is not fully understood. Though much is known about single-user models, comparatively less theory has been developed for multi-user state-dependent models, as we will see in the next chapter.

CHAPTER 3

MULTI-USER STATE-DEPENDENT MODELS

In this chapter, we present background material on some multi-user state-dependent models, and discuss some important results and random coding schemes. We focus mainly on state-dependent *multiple access channel* (MAC) and *broadcast channel* (BC) in various interesting scenarios.

3.1 Multiple Access Channel

Figure 3.1 illustrates a state-dependent MAC in which $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ are the channel inputs from Encoder 1 and Encoder 2, respectively, $S \in \mathcal{S}$ is the channel state, and $Y \in \mathcal{Y}$ is the channel output. The channel output Y is controlled by X_1 , X_2 , and S according to a memoryless probability law $p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2)$. It is also assumed that the channel state is memoryless and drawn with probability law $p(\mathbf{s})$. Encoder i wants to reliably transmit a message $W_i \in \mathcal{W} = \{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$ to the decoder, where R_i is the rate of transmission, for $i = 1, 2$. It is also assumed that the W_1 , W_2 , and S are independent. The decoder estimates the message pair (\hat{W}_1, \hat{W}_2) from the channel output Y^n .

An additive state-dependent Gaussian MAC is a channel model whose output is given as

$$Y^n = X_1^n + X_2^n + S^n + Z^n, \quad (3.1)$$

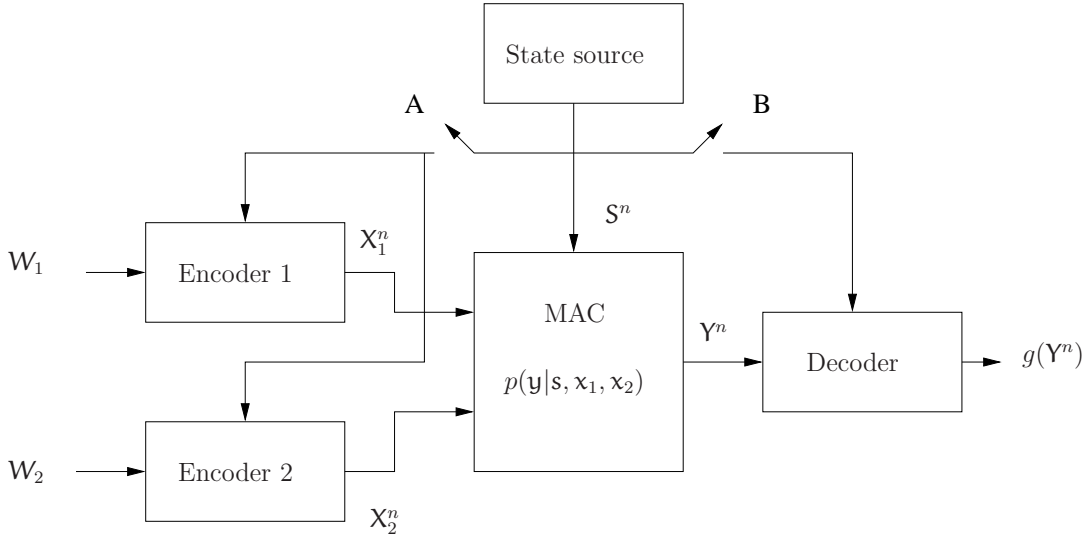


Figure 3.1. Block diagram of a state-dependent multiple access channel

where: the channel inputs X_1^n and X_2^n should satisfy the average power constraint $\frac{1}{n} \sum_{j=1}^n X_{1,j}^2 \leq P_1$ and $\frac{1}{n} \sum_{j=1}^n X_{2,j}^2 \leq P_2$, respectively; the channel state S^n is zero mean Gaussian vector with covariance matrix $Q\mathbf{I}_n$; and the noise Z^n is zero mean Gaussian with covariance matrix $N\mathbf{I}_n$. The noise Z^n and the channel state S^n are mutually independent.

3.1.1 State Neither at the Encoders Nor at the Decoder

In this case, switches A and B are open in Figure 3.1. The encoding functions at Encoder 1 and Encoder 2 are $f_1^n : \mathcal{W}_1 \rightarrow \mathcal{X}_1^n$ and $f_2^n : \mathcal{W}_2 \rightarrow \mathcal{X}_2^n$, respectively, and the decoding function is $g^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$. For the discrete memoryless (DM) case, the capacity region ¹ is closure of the set of rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \mathbb{I}(X_1; Y|X_2, Q), \quad (3.2a)$$

$$R_2 \leq \mathbb{I}(X_2; Y|X_1, Q), \quad (3.2b)$$

¹The capacity region for multiple access channel is defined in Appendix A.

$$R_1 + R_2 \leq \mathbb{I}(X_1, X_2; Y|Q), \quad (3.2c)$$

for some distribution of the form

$$p(\mathbf{q})p(\mathbf{s})p(\mathbf{x}_1|\mathbf{q})p(\mathbf{x}_2|\mathbf{q})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where $Q \in \mathcal{Q}$ is a time-sharing auxiliary random variable with $|\mathcal{Q}| \leq 4$. Since the time-sharing random variable captures time sharing of different codebooks, it can be assumed that the time-sharing sequence Q^n is available at the encoders and the decoder. For this case, the following simple random coding scheme achieves the above capacity region.

- For each message $m_i \in \mathcal{W}_i$ and each time sharing sequence $\mathbf{q}^n \in \mathcal{Q}^n$, generate codewords $\mathbf{X}_i^n(m_i, \mathbf{q}^n)$ whose elements are independently drawn with probability law $p(x_i|\mathbf{q})$, for $i = 1, 2$. These codebooks are revealed to both the encoders and the decoder.
- The time-sharing sequence $Q^n = \mathbf{q}^n$ whose elements are independently drawn with probability law $p(\mathbf{q})$ is available at the encoders and the decoders.
- Encoder i chooses the codeword $\mathbf{X}_i^n(W_i, Q^n)$ to transmit the message W_i , for $i = 1, 2$.
- The decoder looks for a codeword pair $(\mathbf{X}_1^n(m_1, Q^n), \mathbf{X}_2^n(m_2, Q^n))$ that is jointly typical with the channel output Y^n and Q^n . If such a codeword pair exists and is unique, then the decoder declares it to be the estimate of the transmitted message pair. Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided (R_1, R_2) satisfies (3.2) with strict inequalities.

For the additive state-dependent Gaussian MAC, the capacity region is the set of rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{Q + N} \right), \quad (3.3a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{Q + N} \right), \quad (3.3b)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_1 + P_2}{Q + N} \right). \quad (3.3c)$$

In this case, since the channel state is known neither at the encoders nor at the decoder, then the channel state \mathbf{S} is simply treated as noise in the encoding and decoding process. For each message $m_i \in \mathcal{W}_i$, the encoder generates the codeword \mathbf{X}_i^n whose elements are independently drawn according to probability distribution $\mathcal{N}(0, P_i)$ for $i = 1, 2$. The decoder applies minimum Euclidean distance decoding to the channel output \mathbf{Y}^n , i.e., choosing the codeword sum $\mathbf{X}^n(m_1) + \mathbf{X}^n(m_2)$, $m_1 \in \mathcal{W}_1$, $m_2 \in \mathcal{W}_2$ that is closest to \mathbf{Y}^n in the sense of Euclidean distance. In the above model, as in the single-user case, it can be practically useful to incorporate channel state estimation at the decoder. Lossy state recovery at the decoder for the state-dependent MAC will be discussed in detail in Chapter 7.

3.1.2 State only at the Decoder

In this case, switch A is open, and switch B is closed in Figure 3.1. For this model, the encoding functions at Encoder 1 and Encoder 2 are $f_1^n : \mathcal{W}_1 \rightarrow \mathcal{X}_1^n$ and $f_2^n : \mathcal{W}_2 \rightarrow \mathcal{X}_2^n$, respectively, and the decoding function is $g^n : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$. The capacity region for the DM models is closure of the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \mathbb{I}(\mathbf{X}_1; \mathbf{Y}, \mathbf{S} | \mathbf{X}_2, \mathbf{Q}), \quad (3.4a)$$

$$R_2 \leq \mathbb{I}(\mathbf{X}_2; \mathbf{Y}, \mathbf{S} | \mathbf{X}_1, \mathbf{Q}), \quad (3.4b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}, \mathbf{S} | \mathbf{Q}), \quad (3.4c)$$

for some distribution of the form

$$p(\mathbf{q})p(\mathbf{s})p(\mathbf{x}_1|\mathbf{q})p(\mathbf{x}_2|\mathbf{q})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where $\mathbf{Q} \in \mathcal{Q}$ is a time-sharing auxiliary random variable with $|\mathcal{Q}| \leq 4$. Encoding schemes are similar to that in the previous case. The decoder views the available channel state \mathbf{S}^n as another channel observation along with the channel output \mathbf{Y}^n .

In the additive state-dependent Gaussian MAC, the decoder removes the effect of \mathbf{S}^n from \mathbf{Y}^n and then decodes the messages from $\mathbf{Y}^n - \mathbf{S}^n$. For this model, the capacity region is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right), \quad (3.5a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right), \quad (3.5b)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_1 + P_2}{N} \right). \quad (3.5c)$$

As in single-user models, the effect of the channel state can be completely removed when the channel state is additive and known at the decoder. Again, causality is irrelevant for block coding techniques.

3.1.3 State only at the Encoders

In this case, switch A is closed and switch B is open in Figure 3.1. As in single-user case, the channel state is either *causally* or *non-causally* known at *all* encoders.

Causal Case

For this case, the sequences of encoding functions at Encoder 1 and Encoder 2 are $f_{1,i} : \mathcal{W}_1 \times \mathcal{S}^i \rightarrow \mathcal{X}_1$ and $f_{2,i} : \mathcal{W}_2 \times \mathcal{S}^i \rightarrow \mathcal{X}_2$, respectively, for $i = 1, 2, \dots, n$, and the decoding function is $g^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$.

For the discrete model, an inner bound for the capacity region is convex hull of

the set of rate pairs (R_1, R_2) satisfying [40]

$$R_1 < \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2), \quad (3.6a)$$

$$R_2 < \mathbb{I}(\mathbf{U}_2; \mathbf{Y} | \mathbf{U}_1), \quad (3.6b)$$

$$R_1 + R_2 < \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Y}), \quad (3.6c)$$

for some distribution of the form $p(\mathbf{u}_1)p(\mathbf{u}_2)$, $\mathbf{X}_1 = h_1(\mathbf{U}_1, \mathbf{S})$, $\mathbf{X}_2 = h_1(\mathbf{U}_2, \mathbf{S})$. Let us discuss the random coding scheme that achieves the inner bound.

- For each message $m_i \in \mathcal{W}_i$, generate codewords $\mathbf{U}_i^n(m_i)$ whose elements are independently drawn with probability law $p(\mathbf{u}_i)$, for $i = 1, 2$. These codebooks are revealed to both the encoders and the decoder.
- Encoder i , provided with \mathcal{W}_i and \mathbf{S}_j , chooses the channel input $\mathbf{X}_j = h_i(\mathbf{U}_j(\mathcal{W}_i), \mathbf{S}_j)$ at time instant j .
- The decoder looks for a codeword pair $(\mathbf{U}_1^n(m_1), \mathbf{U}_2^n(m_2))$ that is jointly typical with the channel output \mathbf{Y}^n . If such a codeword pair exists and is unique, then the decoder declares it to be the estimate of the transmitted message pair. Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided (R_1, R_2) satisfy (3.6).

For the DM model, an outer bound for the capacity region is convex hull of the set of rate pairs (R_1, R_2) satisfying [40]

$$R_1 < \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2), \quad (3.7a)$$

$$R_2 < \mathbb{I}(\mathbf{U}_2; \mathbf{Y} | \mathbf{U}_1), \quad (3.7b)$$

$$R_1 + R_2 < \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Y}), \quad (3.7c)$$

for some distribution of the form $p(\mathbf{u}_1, \mathbf{u}_2)$, $\mathbf{X}_1 = h_1(\mathbf{U}_1, \mathbf{S})$, $\mathbf{X}_2 = h_1(\mathbf{U}_2, \mathbf{S})$. Since the above inner and outer bounds differ in distributions considered, the capacity region is still not known.

Non-Causal Case

For this case, the encoding functions at Encoder 1 and Encoder 2 are $f_1^n : \mathcal{W}_1 \times \mathcal{S}^n \rightarrow \mathcal{X}_1^n$ and $f_2^n : \mathcal{W}_2 \times \mathcal{S}^n \rightarrow \mathcal{X}_2^n$, respectively, and the decoding function is $g^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$.

In the DM case, the capacity region is not known. An inner bound for the capacity region is the set of rate pairs (R_1, R_2) satisfying [24]

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{U}_2, \mathbf{Q}), \quad (3.8a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Y} | \mathbf{U}_1, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_2; \mathbf{S} | \mathbf{U}_1, \mathbf{Q}), \quad (3.8b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Y} | \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{S} | \mathbf{Q}) \quad (3.8c)$$

for some distribution of the form

$$p(\mathbf{q})p(\mathbf{s})p(\mathbf{u}_1, \mathbf{x}_1 | \mathbf{s}, \mathbf{q})p(\mathbf{u}_2, \mathbf{x}_2 | \mathbf{s}, \mathbf{q})p(\mathbf{y} | \mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where \mathbf{U}_1 and \mathbf{U}_2 are auxiliary random variables, and $\mathbf{Q} \in \mathcal{Q}$ is a time-sharing auxiliary random variable with $|\mathcal{Q}| \leq 4$. The inner bound (3.8) can be obtained by applying GP coding at both encoders and joint decoding at the decoder.

For an additive state-dependent Gaussian MAC with side information at all encoders, the capacity region is the set of rate pairs (R_1, R_2) satisfying [18, 24]

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right), \quad (3.9a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right), \quad (3.9b)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_1 + P_2}{N} \right), \quad (3.9c)$$

which is same as the capacity region of models in which the channel state is known at only the decoder. Similar to the single-user additive state-dependent Gaussian models with non-causal encoder side information, DPC at both the encoders achieves (3.9), and state cancellation is not optimal in terms of the capacity region.

Let us discuss DPC in this model to achieve the capacity region, focusing on successive decoding. The capacity region for this model is shown in Figure 3.2. Assume that one message is decoded first by treating the input from the other encoder as noise and then the other message is decoded with knowledge of the first decoded message. At Encoder 1, codewords $\mathbf{X}_1^n = \mathbf{U}_1^n - \alpha_1 \mathbf{S}^n$ are generated using DPC with $\alpha_1 = \frac{P_1}{P_1 + P_2 + N}$. At Encoder 2, codewords $\mathbf{X}_2^n = \mathbf{U}_2^n - \alpha_2 \mathbf{S}^n$ are generated using DPC with $\alpha_2 = \frac{P_2}{P_2 + N}$. If $R_1 < \frac{1}{2} \log \left(1 + \frac{P_1}{P_2 + N} \right)$, the auxiliary codeword \mathbf{U}_1^n and \mathbf{W}_1 are reliably decoded [9]. Then, we decode the message of Encoder 2 using

$$\mathbf{Y}^n - \mathbf{U}_1^n = \mathbf{X}_2^n + (1 - \alpha_1) \mathbf{S}^n + \mathbf{Z}^n.$$

Using the same argument, if $R_2 < \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$, \mathbf{W}_2 can also be decoded reliably. Thus, rate pairs (R_1, R_2) satisfying $R_1 < \frac{1}{2} \log \left(1 + \frac{P_1}{P_2 + N} \right)$ and $R_2 < \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$, i.e., shaded region \mathcal{A} in Figure 3.2, are achievable. If we change the role of the encoders and the decoding order, rate pairs (R_1, R_2) satisfying $R_1 < \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right)$ and $R_2 < \frac{1}{2} \log \left(1 + \frac{P_2}{P_1 + N} \right)$, i.e., shaded region \mathcal{B} in Figure 3.2, are also achievable. Taking the convex closure of these sets of achievable rates yields (3.9).

For this model, no additional converse proof is needed because an obvious outer bound for the capacity region is (3.5), and the inner bound for the capacity meets this obvious outer bound. Otherwise, it is very difficult to obtain simple, tight outer bounds in general for these models as we will see in the next chapters.

If the channel state is non-causally known at all encoders, DPC is optimal in terms of the capacity region. The capacity region is not known when *some* but not *all* encoders know the channel state because the encoders that are not provided with the channel state can not apply DPC. We study state-dependent MAC with asymmetric availability of side information at the encoders in Chapter 4.

In the above models, recovery of the channel state at the decoder is not considered. As in the single-user case, state-dependent MACs with non-causal encoder

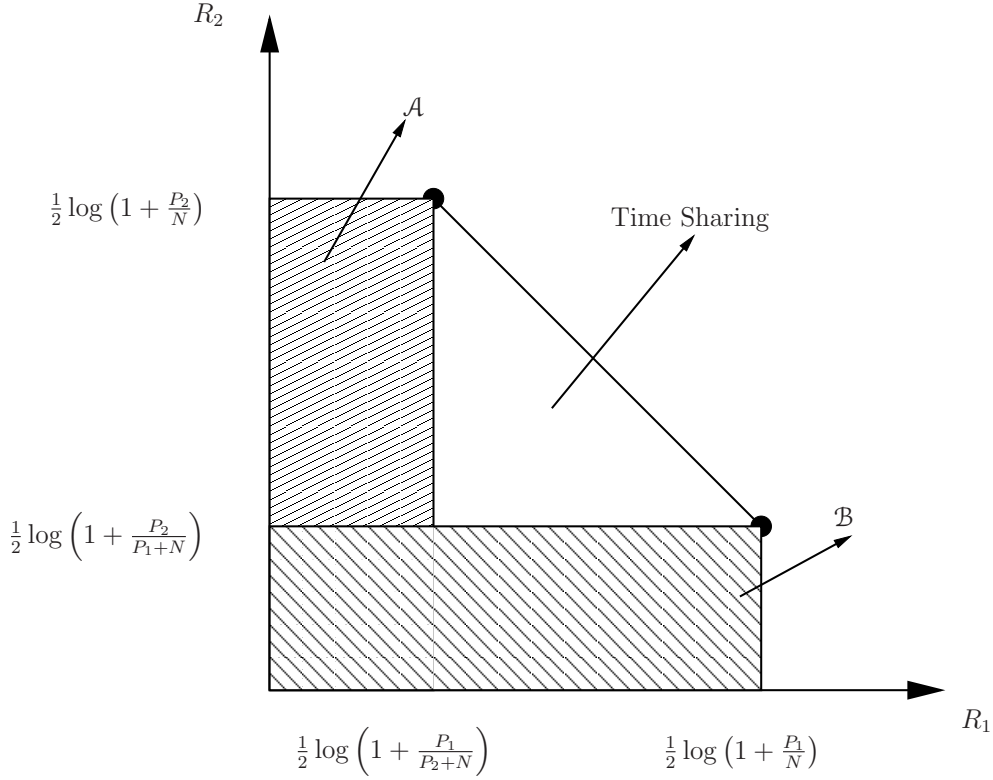


Figure 3.2. Illustration of the capacity region for the additive state-dependent Gaussian multiple access channel with non-causal side information at all encoders.

side information can be used to model distributed information embedding in which each user embeds its message into the provided host satisfying a distortion constraint and all encoders want to communicate their embedded signals to a common decoder. We study lossless recovery of the channel state at the decoder for state-dependent MAC in Chapter 5 from the view of distributed information embedding. From the view of IE, state-dependent MACs with non-causal encoder side information is studied in [55]. Lossy recovery of the channel state at the decoder in state-dependent MAC has not been considered.

3.1.4 State at the Encoders and the Decoder

In this case, switches A and B are closed in Figure 3.1. In this discussion, we focus on models in which the channel state is known *non-causally* at the encoders and is also known at the decoder.

For the DM case, the capacity region is closure of the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \mathbb{I}(X_1; Y, S | X_2, Q), \quad (3.10a)$$

$$R_2 \leq \mathbb{I}(X_2; Y, S | X_1, Q), \quad (3.10b)$$

$$R_1 + R_2 \leq \mathbb{I}(X_1, X_2; Y, S | Q), \quad (3.10c)$$

for some distribution of the form

$$p(\mathbf{q})p(\mathbf{s})p(\mathbf{x}_1|\mathbf{s}, \mathbf{q})p(\mathbf{x}_2|\mathbf{s}, \mathbf{q})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where $Q \in \mathcal{Q}$ is time-sharing auxiliary random variable with $|\mathcal{Q}| \leq 4$. The above capacity region can be obtained using superposition coding at both the encoders. The decoding scheme is similar to that in the case of decoder side information.

In this case, the capacity region for additive state-dependent Gaussian MAC is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right), \quad (3.11a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right), \quad (3.11b)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_1 + P_2}{N} \right). \quad (3.11c)$$

Cemal and Steinberg study state-dependent MAC with rate-constrained channel state known at the encoders and the channel state known at the decoder [4]. Jafar also study state-dependent MAC with encoder side information and decoder side information in [20].

3.2 Broadcast Channel

In this section, let us focus on state-dependent broadcast models in various interesting scenarios. Figure 3.3 illustrates a state-dependent broadcast channel (BC) in which $\mathbf{X} \in \mathcal{X}$ is the channel input, $\mathbf{S} \in \mathcal{S}$ is the channel state, and $\mathbf{Y} \in \mathcal{Y}$ and $\mathbf{Z} \in \mathcal{Z}$ are the channel outputs for Decoder 1 and Decoder 2, respectively. In this model, the channel outputs (\mathbf{Y}, \mathbf{Z}) are controlled by \mathbf{X} and \mathbf{S} according to a memoryless probability law $p(\mathbf{y}, \mathbf{z} | \mathbf{s}, \mathbf{x})$. It is also assumed that the channel state is memoryless and drawn with probability law $p(\mathbf{s})$. The encoder wants to reliably transmit messages $W_1 \in \mathcal{W}_1 = \{1, 2, \dots, \lceil 2^{nR_1} \rceil\}$ and $W_2 \in \mathcal{W}_2 = \{1, 2, \dots, \lceil 2^{nR_2} \rceil\}$ to Decoder 1 and Decoder 2, respectively, in n channel uses, where R_i is the rate of transmission, for $i = 1, 2$. It is also assumed that W_1 , W_2 , and \mathbf{S} are independent. The decoders estimate the messages intended for them from their respective channel outputs. In this chapter, we focus on broadcast channels that are physically degraded i.e., $p(\mathbf{y}, \mathbf{z} | \mathbf{s}, \mathbf{x}) = p(\mathbf{y} | \mathbf{s}, \mathbf{x})p(\mathbf{z} | \mathbf{y})$, in the DM case.

An additive state-dependent Gaussian BC is a channel model whose outputs are given as

$$\mathbf{Y}^n = \mathbf{X}^n + \mathbf{S}^n + \mathbf{Z}_1^n \quad \text{and} \quad \mathbf{Z}^n = \mathbf{X}^n + \mathbf{S}^n + \mathbf{Z}_2^n,$$

where \mathbf{X}^n should satisfy the average power constraint $\frac{1}{n} \sum_{j=1}^n \mathbf{X}_j^2 \leq P$; \mathbf{S}^n is zero mean Gaussian vector with covariance matrix $Q\mathbf{I}_n$; \mathbf{Z}_1^n is zero mean Gaussian with covariance matrix $N_1\mathbf{I}_n$; and \mathbf{Z}_2^n is zero mean Gaussian with covariance matrix $N_2\mathbf{I}_n$. We assume that \mathbf{Z}_1^n and \mathbf{Z}_2^n are independent and $(\mathbf{Z}_1^n, \mathbf{Z}_2^n)$ is independent of \mathbf{S}^n . We assume that $N_2 > N_1$. This Gaussian model is not necessarily physically degraded. But, by a similar argument to [12, Problem 14.10], it suffices to consider instead the following physically degraded, state-dependent broadcast channels with the same

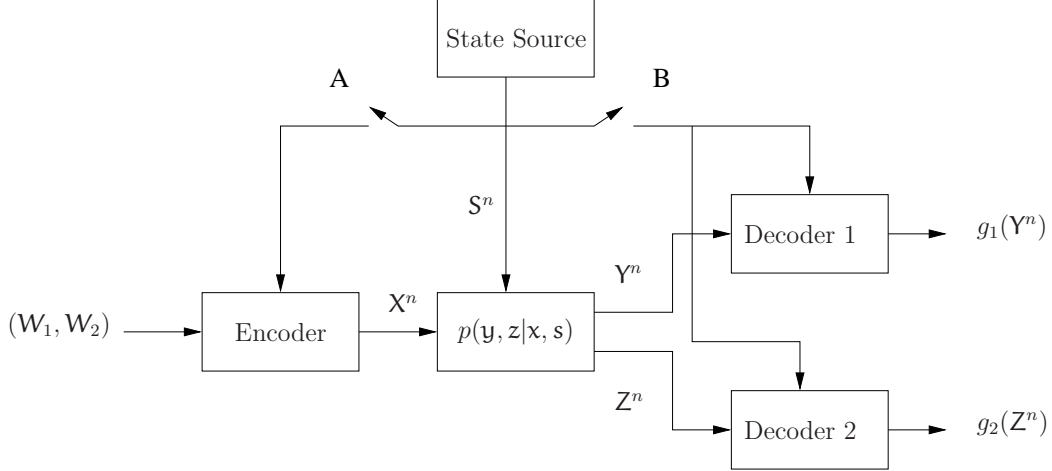


Figure 3.3. Block diagram of a state-dependent broadcast channel.

capacity region:

$$Y^n = X^n + S^n + Z_1^n \quad \text{and} \quad Z^n = Y^n + \tilde{Z}_2^n,$$

where the channel input has average power constraint $\frac{1}{n} \sum_{j=1}^n X_j^2 \leq P$; and zero mean Gaussian vectors S^n , Z_1^n , and \tilde{Z}_2^n are independent of each other and have covariance matrix $Q\mathbf{I}_n$, $N_1\mathbf{I}_n$, and $(N_2 - N_1)\mathbf{I}_n$, respectively.

3.2.1 State Neither at the Encoder Nor at the Decoders

In this case, switches A and B are open in Figure 3.3. The encoding function at the encoder is $f^n : \mathcal{W}_1 \times \mathcal{W}_2 \rightarrow \mathcal{X}^n$, and the decoding functions at the Decoder 1 and Decoder 2 are $g_1^n : \mathcal{Y}^n \rightarrow \mathcal{W}_1$ and $g_2^n : \mathcal{Z}^n \rightarrow \mathcal{W}_2$, respectively.

For the DM case, the capacity region ² is the closure of all rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y} | \mathbf{U}), \quad (3.12a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z}), \quad (3.12b)$$

²The capacity region for broadcast channel is defined in Appendix A.

for some distribution of the form $p(\mathbf{s})p(\mathbf{u})p(\mathbf{x}|\mathbf{u})p(\mathbf{y}|\mathbf{s}, \mathbf{x})p(\mathbf{z}|\mathbf{y})$, where $\mathbf{U} \in \mathcal{U}$ is an auxiliary random variable with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}|$. The following simple random coding technique is used to achieve the above capacity region.

- For each message $m_2 \in \mathcal{W}_2$, generate codewords $\mathbf{U}^n(m_2)$ whose elements are independently drawn with probability law $p(\mathbf{u})$. For each $\mathbf{U}^n(m_2)$, $m_2 \in \mathcal{W}_2$, generate codewords $\mathbf{X}^n(m_1, m_2)$, $m_1 \in \mathcal{W}_1$, whose elements are independently drawn with probability law $p(\mathbf{x}|\mathbf{u})$. The codebooks are revealed to the encoder and both decoders.
- Encoder chooses the codeword $\mathbf{X}^n(\mathcal{W}_1, \mathcal{W}_2)$ to send the message pair $(\mathcal{W}_1, \mathcal{W}_2)$.
- Decoder 2 looks for codeword $\mathbf{U}^n(m_2)$ that is jointly typical with \mathbf{Z}^n , for $m_2 \in \mathcal{W}_2$. If such a codeword exists and is unique, then the decoder declares its index to be the estimate of the message \mathcal{W}_2 . Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R_2 < \mathbb{I}(\mathbf{U}; \mathbf{Z})$.
- Decoder 1 also looks for codeword $\mathbf{U}^n(m_2)$ that is jointly typical with \mathbf{Y}^n , for $m_2 \in \mathcal{W}_2$. If such a codeword exists and is unique, then Decoder 1 declares its index to be the estimate $\hat{\mathcal{W}}_2$ of the message \mathcal{W}_2 . Otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R_2 < \mathbb{I}(\mathbf{U}; \mathbf{Y})$. Decoder 2 then looks for a codeword $\mathbf{X}^n(m_1, \hat{\mathcal{W}}_2)$ that is jointly typical with \mathbf{Y}^n , for $m_1 \in \mathcal{W}_1$. If such a codeword exists and is unique, then the decoder declares its first index to be the estimate $\hat{\mathcal{W}}_1$ of the message \mathcal{W}_1 . Otherwise, Decoder 2 declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R_1 < \mathbb{I}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$.

For the additive state-dependent Gaussian BC, the capacity region is the set of

rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{\gamma P}{Q + N_1} \right), \quad (3.13a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{(1 - \gamma)P}{\gamma P + Q + N_2} \right), \quad (3.13b)$$

for some $\gamma \in [0, 1]$. Since the channel state is known neither at the encoder nor at the decoders, then the channel state \mathbf{S} is simply treated as noise in the encoding and decoding process. Encoding and decoding at both the decoders are identical to that of DM case. But, the codewords are $\mathbf{X}^n = \mathbf{X}_1^n + \mathbf{X}_2^n$, where \mathbf{X}_1^n and \mathbf{X}_2^n carry messages \mathcal{W}_1 and \mathcal{W}_2 , respectively, \mathbf{X}_1^n is drawn with $\mathcal{N}(0, \gamma P)$, \mathbf{X}_2^n is drawn with $\mathcal{N}(0, (1 - \gamma)P)$, and \mathbf{X}_1^n and \mathbf{X}_2^n are independent of each other.

In the above model, the decoders consider only recovering the messages intended for them. Partial recovery of channel state at the decoders for state-dependent BC is considered in Chapter 7.

3.2.2 State only at the Decoders

In this case, switch A is open, and switch B is closed in Figure 3.3. The encoding function at the encoder is $f^n : \mathcal{W}_1 \times \mathcal{W}_2 \rightarrow \mathcal{X}^n$, and the decoding functions at the Decoder 1 and Decoder 2 are $g_1^n : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{W}_1$ and $g_2^n : \mathcal{Z}^n \times \mathcal{S}^n \rightarrow \mathcal{W}_2$, respectively.

For the DM case, the capacity region is the closure of all rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y}, \mathbf{S} | \mathbf{U}), \quad (3.14a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z}, \mathbf{S}), \quad (3.14b)$$

for some distribution of the form $p(\mathbf{s})p(\mathbf{u})p(\mathbf{x}|\mathbf{u})p(\mathbf{y}|\mathbf{s}, \mathbf{x})p(\mathbf{z}|\mathbf{y})$, where $\mathbf{U} \in \mathcal{U}$ is an auxiliary random variable with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}|$. The encoding scheme is similar to that in previous case. In terms of decoding, the difference between the case of side

information neither at the encoder nor at the decoders and this case is that the decoders view the available channel state \mathbf{S}^n as another channel observation along with their respective channel outputs.

In the additive state-dependent Gaussian BC, the decoders remove the effect of \mathbf{S}^n from their respective channel outputs and then decode the messages intended for them. The capacity region is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{\gamma P}{N_1} \right), \quad (3.15a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{(1 - \gamma)P}{\gamma P + N_2} \right), \quad (3.15b)$$

for some $\gamma \in [0, 1]$. As in single-user models, the effect of the channel state can be completely removed if the channel state is additive and known at the decoder.

3.2.3 State only at the Encoder

In this case, switch A is closed and switch B is open in Figure 3.3. As before, the channel state can be either *causally* or *non-causally* known at the encoder.

Causal Case

For the causal case, the sequence of encoding functions is $f_i : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^i \rightarrow \mathcal{X}$, for $i = 1, 2, \dots, n$, and the decoding functions at Decoder 1 and Decoder 2 are $g_1^n : \mathcal{Y}^n \rightarrow \mathcal{W}_1$ and $g_2^n : \mathcal{Z}^n \rightarrow \mathcal{W}_2$, respectively.

For the DM case, the capacity region is the closure of set of rate pairs (R_1, R_2) satisfying [40]

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2), \quad (3.16a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Z}), \quad (3.16b)$$

for some $p(\mathbf{u}_1, \mathbf{u}_2)$, and $\mathbf{X} = h(\mathbf{U}_1, \mathbf{U}_2, \mathbf{S})$. Let us discuss the capacity achieving random coding scheme.

- For each message $m_2 \in \mathcal{W}_2$, generate codewords $\mathbf{U}_2^n(m_2)$ whose elements are independently drawn with probability law $p(\mathbf{u}_2)$. For each $\mathbf{U}_1^n(m_2)$, $m_2 \in \mathcal{W}_2$, generate codewords $\mathbf{U}_1^n(m_1, m_2)$, $m_1 \in \mathcal{W}_1$, whose elements are drawn with probability law $p(\mathbf{u}_1|\mathbf{u}_2)$. The codebooks are revealed to the encoder and both decoders.
- At time instant i , the encoder, provided with $(\mathcal{W}_1, \mathcal{W}_2)$ and \mathbf{S}^i , chooses $\mathbf{X}_i = h(\mathbf{U}_{1,i}(\mathcal{W}_1, \mathcal{W}_2), \mathbf{U}_{2,i}(\mathcal{W}_2), \mathbf{S}_i)$.
- Decoding procedures at both decoders are similar to those in the case of side information neither at the encoder nor at the decoders. The messages \mathcal{W}_1 and \mathcal{W}_2 are decoded reliably provided (R_1, R_2) satisfies (3.16).

Non-Causal Case

For the non-causal case, the encoding function is $f^n : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \rightarrow \mathcal{X}^n$, and the decoding functions at Decoder 1 and Decoder 2 are $g_1^n : \mathcal{Y}^n \rightarrow \mathcal{W}_1$ and $g_2^n : \mathcal{Z}^n \rightarrow \mathcal{W}_2$, respectively.

In the DM case, the capacity region is still not known. An inner bound for the capacity region is the set of rate pairs (R_1, R_2) satisfying [24, 45]

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{Y}|\mathbf{U}_2) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}|\mathbf{U}_2), \quad (3.17a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Z}) - \mathbb{I}(\mathbf{U}_2; \mathbf{S}), \quad (3.17b)$$

for some distribution of the form

$$p(\mathbf{s})p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}|\mathbf{s})p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\mathbf{z}|\mathbf{y}),$$

where \mathbf{U}_1 and \mathbf{U}_2 are auxiliary random variables. The inner bound (3.17) can be obtained by combining superposition coding and GP coding.

In the DM case, an outer bound for the capacity region is the set of rate pairs (R_1, R_2) satisfying [45]

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2, \mathbf{U}_3) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{U}_2, \mathbf{U}_3), \quad (3.18a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Z}) - \mathbb{I}(\mathbf{U}_2; \mathbf{S}), \quad (3.18b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3; \mathbf{Y}) - \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3; \mathbf{S}), \quad (3.18c)$$

for some distribution of the form

$$p(\mathbf{s})p(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{x} | \mathbf{s})p(\mathbf{y} | \mathbf{x}, \mathbf{s})p(\mathbf{z} | \mathbf{y}),$$

where \mathbf{U}_1 , \mathbf{U}_2 , and \mathbf{U}_3 are auxiliary random variables.

For the additive state-dependent Gaussian BC, the capacity region is the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{\gamma P}{N_1} \right), \quad (3.19a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{(1-\gamma)P}{\gamma P + N_2} \right), \quad (3.19b)$$

for some $\gamma \in [0, 1]$. The capacity region is same as that of models when the channel state is known at only the decoder, and can be obtained using DPC.

Let us discuss the capacity achieving coding scheme. We decompose the input signal \mathbf{X}^n into two parts \mathbf{X}_1^n and \mathbf{X}_2^n with average power constraints γP and $(1-\gamma)P$, and carrying independent messages \mathbf{W}_1 and \mathbf{W}_2 , respectively. For the worse decoder, the channel output is $\mathbf{Z}^n = \mathbf{X}_2^n + \mathbf{S}^n + (\mathbf{X}_1^n + \mathbf{Z}_2^n)$. Using DPC with $\alpha_2 = \frac{(1-\gamma)P}{P+N_2}$, $\mathbf{X}_2^n = \mathbf{U}_2^n - \alpha_2 \mathbf{S}^n$ is generated. Then, \mathbf{W}_2 and auxiliary codeword \mathbf{U}_2^n are reliably decoded at Decoder 2 if $R_2 < \frac{1}{2} \log \left(1 + \frac{(1-\gamma)P}{\gamma P + N_2} \right)$ [9]. Decoder 1 can decode \mathbf{U}_2^n as well because the channel output of Decoder 2 is a degraded version of that of Decoder 1. Then Decoder 1 can remove $\mathbf{U}_2^n = \mathbf{X}_2^n + \alpha_2 \mathbf{S}^n$ from \mathbf{Y}^n to make the channel output equivalent to

$$\tilde{\mathbf{Y}}^n = \mathbf{Y}^n - \mathbf{U}_2^n = \mathbf{X}_1^n + (1 - \alpha_2) \mathbf{S}^n + \mathbf{Z}_1^n.$$

Using DPC with $\alpha_1 = \frac{\gamma P}{\gamma P + N_1}$, $\mathbf{X}_1^n = \mathbf{U}_1^n - \alpha_1 \mathbf{S}^n$ is generated, and \mathbf{W}_1 is reliably decoded at Decoder 1 if $R_1 < \frac{1}{2} \log \left(1 + \frac{\gamma P}{N_1} \right)$ [9].

In the above models, recovery of the channel state at the decoders is not considered. We study lossless recovery of the channel state at some decoders for state-dependent BC in Chapter 6. Lossy recovery of the channel state at the decoders in state-dependent BC will be discussed in 7.

In our summary here, we have focused on degraded state-dependent broadcast channels. The *general* state-dependent BC with non-causal encoder side information is studied in [46]. State-dependent Gaussian BC with non-causal encoder side information from the view of *multi-casting*, i.e., $\mathbf{W}_1 = \mathbf{W}_2$, is considered in [23].

3.2.4 State at the Encoder and the Decoders

In this case, switches A and B are closed in Figure 3.3. We again focus on the case of channel state is *non-causally* known at the encoder and is also known at the decoders.

For the DM case, the capacity region is the closure of all rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y} | \mathbf{U}, \mathbf{S}), \quad (3.20a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z} | \mathbf{S}), \quad (3.20b)$$

for some distribution of the form $p(\mathbf{s})p(\mathbf{u}|\mathbf{s})p(\mathbf{x}|\mathbf{s}, \mathbf{u})p(\mathbf{y}|\mathbf{s}, \mathbf{x})p(\mathbf{z}|\mathbf{y})$, where $\mathbf{U} \in \mathcal{U}$ is an auxiliary random variable with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}| + 2$.

For the state-dependent Gaussian BC, the capacity region is the set of rate pairs (R_1, R_2) satisfying [12]

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{\gamma P}{N_1} \right), \quad (3.21a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{(1 - \gamma)P}{\gamma P + N_2} \right), \quad (3.21b)$$

for some $\gamma \in [0, 1]$.

3.3 Other Models

State-dependent degraded relay channels with non-causal side information at the encoder and the relay are considered in [24]; the capacity is found in the Gaussian case, and DPC is optimal and removes the effect of the channel state on the capacity. State-dependent relay broadcast channels with symmetric or asymmetric non-causal side information are considered in [57]. From the view of information embedding, the state-dependent relay channels are also studied in [56].

3.4 Summary

In this chapter, we have briefly summarized a variety of existing results and coding schemes for state-dependent MAC and BC. We have also identified several interesting scenarios in which theory is less well developed. In the forthcoming chapters, we focus on studying some of the interesting scenarios discussed in this chapter for state-dependent MACs and BCs. In the next chapter, we focus on studying state-dependent MAC with asymmetric encoder side information.

CHAPTER 4

STATE-DEPENDENT MULTIPLE ACCESS CHANNELS WITH SIDE INFORMATION AT SOME ENCODERS

In this chapter, we consider a state-dependent multiple access channel (MAC) with side information fully correlated to the channel state *non-causally* known only at some encoders. The simplest example of a communication system under investigation is shown in Figure 4.1, in which two encoders communicate to a single decoder through a MAC $p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{S} \sim \mathcal{S}$ is the channel state, $\mathbf{X}_1 \in \mathcal{X}_1$ and $\mathbf{X}_2 \in \mathcal{X}_2$ are the channel inputs from the two encoders, $\mathbf{Y} \in \mathcal{Y}$ is the channel output. These alphabet sets are discrete sets and the set of real numbers for discrete models and Gaussian models, respectively. The results can in principle be extended to any number of encoders with a subset of them being informed of the side information. We assume that the channel state \mathbf{S}_i are independent and identically distributed (i.i.d.) random variables drawn according to $p(\mathbf{s})$, $i = 1, 2, \dots, n$, and one of the encoders has non-causal side information. The side information $\mathbf{T}_i \in \mathcal{T}$ is correlated to the channel state \mathbf{S}_i through a memoryless probability law $p(\mathbf{t}_i|\mathbf{s}_i)$, $i = 1, 2, \dots, n$. The informed encoder, provided with the messages $\mathbf{W}_0 \in \{1, 2, \dots, M_0\}$, $\mathbf{W}_1 \in \{1, 2, \dots, M_1\}$, and the side information \mathbf{T}^n , generates the codeword \mathbf{X}_1^n . The uninformed encoder, provided only with the messages \mathbf{W}_0 and $\mathbf{W}_2 \in \{1, 2, \dots, M_2\}$, generates the codeword \mathbf{X}_2^n . The decoder, upon receiving the channel output \mathbf{Y}^n , estimates all messages $(\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2)$ from \mathbf{Y}^n . We assume

that all messages are independent and probability of each of message $W_i = w_i$ is given by $\frac{1}{M_i}$, for $i = 0, 1, 2$.

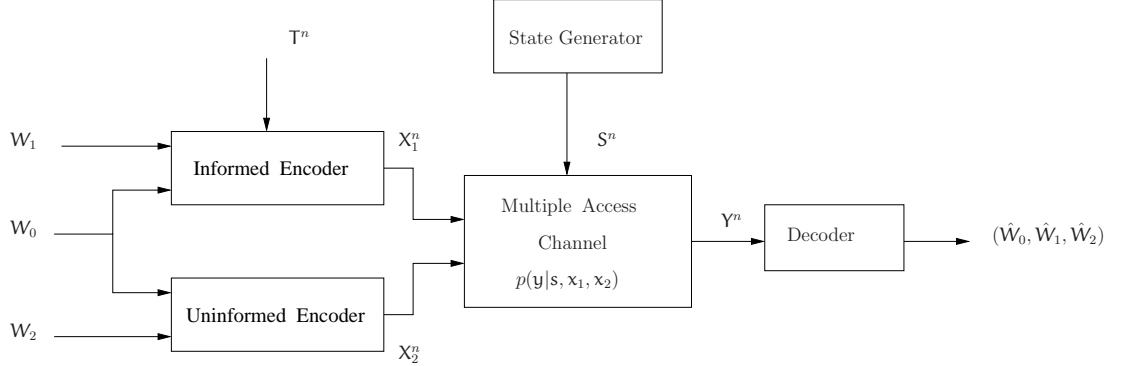


Figure 4.1. Block diagram of a state-dependent multiple access channel with asymmetric encoder side information.

Definition 1 A $(\lceil 2^{nR_0} \rceil, \lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ code consists of encoding functions

$$f_1^n : \mathcal{T}^n \times \mathcal{W}_0 \times \mathcal{W}_1 \rightarrow \mathcal{X}_1^n \text{ and } f_2^n : \mathcal{W}_0 \times \mathcal{W}_2 \rightarrow \mathcal{X}_2^n$$

at the informed encoder and the uninformed encoder, respectively, and a decoding function

$$g^n : \mathcal{Y}^n \rightarrow \mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{W}_2,$$

where $\mathcal{W}_i = \{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$ for $i = 1, 2$.

From a $(\lceil 2^{nR_0} \rceil, \lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ code, the sequences X_1^n and X_2^n from the informed encoder and the uninformed encoder, respectively, are transmitted across a state-dependent MAC without feedback modeled as a discrete memoryless conditional probability distribution $p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2)$, so that

$$\Pr(\mathbf{Y}^n = \mathbf{y}^n | \mathbf{s}^n, \mathbf{x}_1^n, \mathbf{x}_2^n) = \prod_{j=1}^n p(\mathbf{y}_j | \mathbf{s}_j, \mathbf{x}_{1,j}, \mathbf{x}_{2,j}). \quad (4.1)$$

The decoder, upon receiving the channel output \mathbf{Y}^n , reconstructs the messages. The average probability of error is defined as $P_e^n = \mathbb{P}[g(\mathbf{Y}^n) \neq (W_0, W_1, W_2)]$.

Definition 2 A rate triple (R_0, R_1, R_2) is said to be achievable if there exists a sequence of $(\lceil 2^{nR_0} \rceil, \lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes (f_1^n, f_2^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n = 0$.

Definition 3 The capacity region \mathcal{C} is the closure of the set of achievable rate triple (R_0, R_1, R_2) .

In this chapter, our goal is to study the capacity region of the model shown in Figure 4.1 when the side information is perfectly correlated to the channel state information (CSI) i.e., $\mathsf{T} = \mathsf{S}$. Results of this chapter are presented in [26, 25, 29]

4.1 Independent Messages

In this section, we consider the general model shown in Figure 4.1 with common message rate R_0 being zero (no common message case, $\mathcal{W}_0 = \emptyset$).

4.1.1 Discrete Memoryless Case

In this section, we derive inner and outer bounds for the capacity region of discrete memoryless (DM) state-dependent MAC with one encoder being informed, and specialize the inner bound to a binary noiseless MAC. In this section, we consider $\mathcal{X}_1, \mathcal{X}_2, \mathcal{S}$, and \mathcal{Y} to all be discrete and finite alphabets; and all probability laws are to be interpreted as probability mass functions.

Definition 4 For given distributions $p(\mathsf{s})$ and $p(\mathsf{y}|\mathsf{s}, \mathsf{x}_1, \mathsf{x}_2)$, let \mathcal{P}^i be the collection of random variables $(\mathsf{Q}, \mathsf{S}, \mathsf{U}_1, \mathsf{X}_1, \mathsf{X}_2, \mathsf{Y})$ with

$$p(\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = p(\mathbf{q})p(\mathbf{s})p(\mathbf{u}_1|\mathbf{q}, \mathbf{s})p(\mathbf{x}_1|\mathbf{q}, \mathbf{s}, \mathbf{u}_1)p(\mathbf{x}_2|\mathbf{q})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where Q and U_1 are auxiliary random variables.

Inner Bound for the Capacity Region

The following theorem provides an inner bound for the DM MAC with CSI non-causally known at one encoder.

Theorem 1 Let \mathcal{R}^i be the closure of the set of rate pairs (R_1, R_2) satisfying

$$R_1 < \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{X}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) \quad (4.2a)$$

$$R_2 < \mathbb{I}(\mathbf{X}_2; \mathbf{Y} | \mathbf{U}_1, \mathbf{Q}) \quad (4.2b)$$

$$R_1 + R_2 < \mathbb{I}(\mathbf{U}_1, \mathbf{X}_2; \mathbf{Y} | \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}), \quad (4.2c)$$

for some random vector $(\mathbf{Q}, \mathbf{S}, \mathbf{U}_1, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}^i$. Then, the capacity region \mathcal{C} of DM MAC with one informed encoder satisfies $\mathcal{R}^i \subseteq \mathcal{C}$.

Proof: The informed encoder applies Gel'fand-Pinsker coding [17] to encode its messages using available state and the uninformed encoder encodes its message in the same way as for a regular DM MAC [12]. The decoder uses joint decoding of messages from both the encoders. In this inner bound, the random variable \mathbf{Q} takes care of time-sharing of different codebooks. A formal proof for the above theorem considering all error events is given in Appendix B.1.

Remarks:

- The region \mathcal{R}^i in Theorem 1 is convex due to the auxiliary time-sharing random variable \mathbf{Q} .
- Since the region is convex in $p(\mathbf{x}_1 | \mathbf{u}_1, \mathbf{s}, \mathbf{q})$, it is sufficient to take \mathbf{X}_1 to be a deterministic function of $(\mathbf{S}, \mathbf{U}_1, \mathbf{Q})$.
- To compute \mathcal{R}^i in Theorem 1, it is sufficient to restrict random variables $\mathbf{Q} \in \mathcal{Q}$ and $\mathbf{U}_1 \in \mathcal{U}_1$ to alphabet sizes $|\mathcal{Q}| \leq 4$ and $|\mathcal{U}_1| \leq |\mathcal{X}_1| |\mathcal{X}_2| |\mathcal{S}| + 4$, respectively.
- The inner bound \mathcal{R}^i of Theorem 1 uses joint decoding of both encoders' messages. Let us also discuss successive decoding, i.e., decoding one encoder's message first and use the decoded codeword and the channel output to decode the other encoder's message. First, consider decoding the message of the informed encoder. Following [17], if $R_1 < \mathbb{I}(\mathbf{U}_1; \mathbf{Y}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S})$, we can decode the codeword \mathbf{U}_1^n of the informed encoder with arbitrarily low probability of error. Now, we use \mathbf{U}_1^n along with \mathbf{Y}^n to decode \mathbf{X}_2^n . Under these

conditions, if $R_2 < \mathbb{I}(X_2; Y|U_1)$, then we can decode the message of the uninformed encoder with arbitrarily low probability of error. If we change the decoding order of the two messages, the constraints are $R_2 < \mathbb{I}(X_2; Y)$ and $R_1 < \mathbb{I}(U_1; Y|X_2) - \mathbb{I}(U_1; S)$.

Outer bound for the Capacity Region

In this section, we present two outer bounds for the capacity region of the DM MAC with one informed encoder as shown in Figure 4.1. We present two outer bounds here because we do not know which one of the following outer bounds is better at this point.

Definition 5 For given distributions $p(s)$ and $p(y|s, x_1, x_2)$, let \mathcal{P}_1^o be the collection of random variables $(Q, S, U_1, U_2, V, X_1, X_2, Y)$ with

$$p(\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = p(\mathbf{q})p(\mathbf{s})p(\mathbf{u}_1|\mathbf{q}, \mathbf{s})p(\mathbf{u}_2|\mathbf{q})p(\mathbf{v}|\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{u}_2) \\ \times p(\mathbf{x}_1, \mathbf{x}_2|\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where Q, U_1, U_2 and V are auxiliary random variables.

Theorem 2 Let \mathcal{R}_1^o be the set of all (R_1, R_2) satisfying

$$R_1 \leq \mathbb{I}(U_1; Y|V, U_2, Q) - \mathbb{I}(U_1; S|V, U_2, Q) \quad (4.3a)$$

$$R_2 \leq \mathbb{I}(U_2; Y|V, U_1, Q) - \mathbb{I}(U_2; S|V, U_1, Q) \quad (4.3b)$$

$$R_1 + R_2 \leq \mathbb{I}(U_1, U_2; Y|V, Q) - \mathbb{I}(U_1, U_2; S|V, Q) \quad (4.3c)$$

for some $(Q, S, U_1, U_2, V, X_1, X_2, Y) \in \mathcal{P}_1^o$. Then, the capacity region \mathcal{C} of the DM MAC with one informed encoder satisfies

$$\mathcal{C} \subseteq \mathcal{R}_1^o.$$

Proof: A proof of this theorem is given in Appendix B.2.

Definition 6 For given distributions $p(s)$ and $p(y|s, x_1, x_2)$, let \mathcal{P}_2^o be the collection of random variables $(Q, S, U_1, U_2, X_1, X_2, Y)$ with

$$p(\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = p(\mathbf{q})p(\mathbf{s})p(\mathbf{u}_2|\mathbf{q})p(\mathbf{u}_1|\mathbf{q}, \mathbf{s}, \mathbf{u}_2)p(\mathbf{x}_1, \mathbf{x}_2|\mathbf{q}, \mathbf{s}, \mathbf{u}_1, \mathbf{u}_2)p(\mathbf{y}|\mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where Q, U_1 , and U_2 are auxiliary random variables.

Theorem 3 Let \mathcal{R}_2^o be the set of all (R_1, R_2) satisfying

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{U}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{U}_2, \mathbf{Q}) \quad (4.4a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Y} | \mathbf{U}_1, \mathbf{Q}) \quad (4.4b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Y} | \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) \quad (4.4c)$$

for some $(\mathbf{Q}, \mathbf{S}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}_2^o$. Then, the capacity region \mathcal{C} of the DM MAC with one encoder satisfies

$$\mathcal{C} \subseteq \mathcal{R}_2^o.$$

Proof: We do not give the proof of this theorem because it is similar to the proof of Theorem 2.

Remarks:

- We note that for \mathcal{R}_1^o and \mathcal{R}_2^o to be computable, we would also need to provide cardinality bounds on the auxiliary random variables. To compute \mathcal{R}_1^o in Theorem 2, it is sufficient to restrict to $\mathbf{Q} \in \mathcal{Q}$ with $|\mathcal{Q}| \leq 4$ and $\mathbf{V} \in \mathcal{V}$ with $|\mathcal{V}| \leq 4$. To compute \mathcal{R}_2^o in Theorem 3, it is sufficient to restrict to $\mathbf{Q} \in \mathcal{Q}$ with $|\mathcal{Q}| \leq 4$. In either case, we do not have the cardinality bounds on the alphabet of auxiliary random variables \mathbf{U}_1 and \mathbf{U}_2 . So, these two outer bounds can not be computable.
- In the region \mathcal{R}_1^o , we can also combine \mathbf{V} and \mathbf{Q} in (2). In that case, \mathbf{U}_1 and \mathbf{U}_2 are not conditionally independent given (\mathbf{V}, \mathbf{Q}) . The common information is captured in the random variable \mathbf{V} .
- Region \mathcal{R}_1^o in Theorem 2 and \mathcal{R}_2^o in Theorem 3 are convex due to the auxiliary time-sharing random variable \mathbf{Q} .
- Since $(\mathbf{Q}, \mathbf{S}, (\mathbf{U}_1, \mathbf{V}), \mathbf{U}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}_2^o$, it can be easily verified that $\mathcal{P}_1^o \subseteq \mathcal{P}_2^o$. If we are able to show that the rate bounds in \mathcal{R}_1^o are less than or equal to the corresponding rate bounds in \mathcal{R}_2^o , then we can conclude that $\mathcal{R}_1^o \subseteq \mathcal{R}_2^o$. The rate bounds in \mathcal{R}_1^o are less than or equal to the corresponding rate bounds in \mathcal{R}_2^o if

$\mathbb{I}(V; S | \mathbf{U}_2, Q) - \mathbb{I}(V; Y | \mathbf{U}_2, Q) \geq 0$ for every $(Q, S, (\mathbf{U}_1, V), \mathbf{U}_2, \mathbf{X}_1, \mathbf{X}_2, Y) \in \mathcal{P}_1^o$.

In this case, $\mathcal{R}_1^o \subseteq \mathcal{R}_2^o$. Otherwise, it is not trivial to compare \mathcal{R}_1^o and \mathcal{R}_2^o .

Binary Noiseless Example

In this section, we specialize Theorem 1 to a binary noiseless MAC with CSI $\mathbf{Y}^n = \mathbf{X}_1^n \oplus \mathbf{X}_2^n \oplus \mathbf{S}^n$, where \mathbf{X}_1^n and \mathbf{X}_2^n are channel inputs with constraints $\sum_{i=1}^n X_{1,i} \leq np_1$ and $\sum_{i=1}^n X_{2,i} \leq np_2$, respectively; \mathbf{S}^n is the memoryless CSI vector whose elements are *non-causally* known at *one* encoder and are i.i.d. Bernoulli(q) random variables; and \oplus represents modulo-2 addition.

To obtain an inner bound for the capacity region of the binary noiseless MAC, the informed encoder applies a slightly generalized binary DPC. In the generalized binary DPC, the informed encoder uses auxiliary random variable $\mathbf{U}_1 = \mathbf{X}_1 \oplus \mathbf{S}$ in which the channel input \mathbf{X}_1 and the channel state \mathbf{S} are correlated. The following definition is obtained by computing the inner bound in Theorem 1 using $\mathbf{U}_1 = \mathbf{X}_1 \oplus \mathbf{S}$ with \mathbf{X}_1 and \mathbf{S} being correlated and $\mathbf{X}_2 \sim \text{Bernoulli}(p_2)$.

Definition 7 Let $\mathcal{R}^i(a_{10}, a_{01})$ be the set of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq \mathbb{H}_b(a_{10}) + q[\mathbb{H}_b(a_{01}) - \mathbb{H}_b(a_{10})] \quad (4.5a)$$

$$R_2 \leq \mathbb{H}_b(p_2) \quad (4.5b)$$

$$R_1 + R_2 \leq \mathbb{H}_b(p_2 * [qa_{01} + (1-q)a_{10}]) + \mathbb{H}_b(a_{10}) \\ + q[\mathbb{H}_b(a_{01}) - \mathbb{H}_b(a_{10})] - \mathbb{H}_b(qa_{01} + (1-q)a_{10}), \quad (4.5c)$$

for $(a_{10}, a_{01}) \in \mathcal{A}$, where

$$\mathcal{A} := \{(x, y) : 0 \leq x, y \leq 1, \text{ and } (1-q)x + q(1-y) \leq p_1\},$$

and $\mathbb{H}_b(\gamma) := -\gamma \log_2(\gamma) - (1-\gamma) \log_2(1-\gamma)$, and $x * y := x(1-y) + y(1-x)$. Let

$$\mathcal{R}_{\text{BIN}}^i := \text{cl}\{\text{co}\{\cup_{(a_{10}, a_{01}) \in \mathcal{A}} \mathcal{R}^i(a_{10}, a_{01})\}\}.$$

The following corollary gives an inner bound for the capacity region of the binary noiseless MAC.

Corollary 1 *The capacity region \mathcal{C}_{BIN} for the binary noiseless MAC with CSI known at one encoder satisfies $\mathcal{R}_{\text{BIN}}^i \subseteq \mathcal{C}_{\text{BIN}}$.*

Proof: Encoding and decoding are similar to encoding and decoding explained for the general discrete memoryless case above. The informed encoder uses generalized binary dirty paper coding (DPC) which allows arbitrary correlation between the codeword and the known CSI. We consider $\mathbf{U}_1 = \mathbf{X}_1 \oplus \mathbf{S}$ and $\mathbf{X}_2 \sim \text{Bernoulli}(p_2)$, where: $\mathbf{S} \sim \text{Bernoulli}(q)$; \mathbf{X}_1 is related to \mathbf{S} by $a_{01} := P(\mathbf{X}_1 = 0 | \mathbf{S} = 1)$ and $a_{10} := P(\mathbf{X}_1 = 1 | \mathbf{S} = 0)$ with a_{01} and a_{10} chosen such that $P(\mathbf{X}_1 = 1) \leq p_1$. We compute the region $\mathcal{R}^i(a_{10}, a_{01})$ defined in (1) using the probability mass function of \mathbf{X}_2 and the auxiliary random variable \mathbf{U}_1 for all $(a_{10}, a_{01}) \in \mathcal{A}$ to obtain the region $\mathcal{R}_{\text{BIN}}^i$ in (7).

The following proposition provides an outer bound for the capacity region of the binary noiseless MAC with one informed encoder. We do not provide a proof of the following proposition because it is same as the capacity region of the binary noiseless MAC with CSI known at the decoder.

Proposition 1 *Let $\mathcal{R}_{\text{BIN}}^o$ be the set of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \mathbb{H}_b(p_1) \tag{4.6a}$$

$$R_2 \leq \mathbb{H}_b(p_2) \tag{4.6b}$$

$$R_1 + R_2 \leq \begin{cases} \mathbb{H}_b(p_1 + p_2) & \text{if } 0 \leq p_1 + p_2 < 0.5 \\ 1 & \text{if } 0.5 \leq p_1 + p_2 \leq 1 \end{cases} \tag{4.6c}$$

Then, the capacity region \mathcal{C}_{BIN} for the binary noiseless MAC with one informed encoder satisfies $\mathcal{C}_{\text{BIN}} \subseteq \mathcal{R}_{\text{BIN}}^o$.

Numerical Example:

Figure 4.2 depicts the inner bound using the generalized binary DPC specified in Corollary 1 and the outer bound specified in Proposition 1 for the case in which $p_1 = 0.1$, $p_2 = 0.4$, and $q = 0.2$. Also shown for comparison are the following: an inner bound using binary DPC alone, or the generalized DPC with $a_{10} = p_1$ and

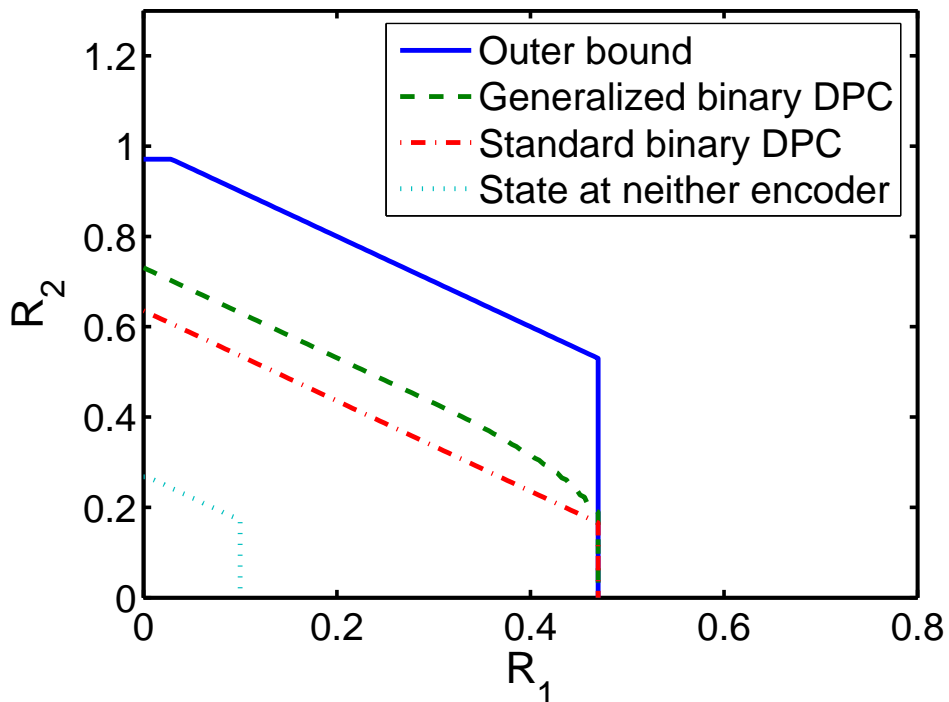


Figure 4.2. A numerical example for the binary noiseless multiple access channel with $p_1 = 0.1$, $p_2 = 0.4$, $q = 0.2$.

$a_{01} = 1 - p_1$; and the capacity region for the case in which the channel state is known at neither the encoders nor the decoder.

These results show that the inner bound obtained by using generalized binary DPC is larger than that obtained using binary DPC [58]. These results suggest that the informed encoder can help the uninformed encoder using binary DPC as well as generalized binary DPC. Even though CSI is known at only one encoder, both the encoders can benefit in terms of achievable rates compared to the case in which CSI is available nowhere. However, it does not appear that the region achieved by an informed decoder is achievable with only one informed encoder, in contrast to the single-user case.

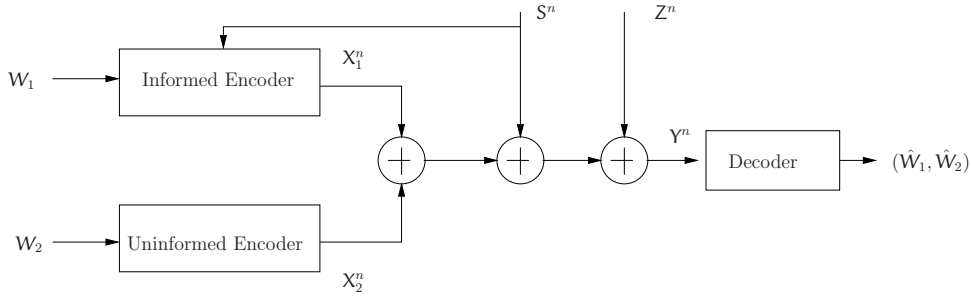


Figure 4.3. Gaussian multiple access channel with channel state information known at one encoder.

4.1.2 Gaussian Memoryless Case

In this section, we develop inner and outer bounds for the memoryless Gaussian case. The additive Gaussian MAC with one informed encoder is shown in Figure 4.3. The output of the channel is $Y^n = X_1^n + X_2^n + S^n + Z^n$, where: X_1^n and X_2^n are the channel inputs with average power constraints $\sum_{i=1}^n X_{1,i}^2 \leq nP_1$ and $\sum_{i=1}^n X_{2,i}^2 \leq nP_2$ with probability one, respectively; S^n is the memoryless channel state vector whose elements are zero-mean Gaussian random variables with variance Q ; and Z^n is the memoryless additive noise vector whose elements are zero-mean Gaussian random variables with variance N and are independent of the channel inputs and the channel state.

The following theorem gives an inner bound for the Gaussian MAC with one informed encoder. To obtain the inner bound for this case, we apply a slightly generalized DPC at the informed encoder in which the channel input X_1 and the channel state S are negatively correlated.

Definition 8 *Let*

$$r_1(\rho, \alpha) := \frac{1}{2} \log \left(\frac{P_1(1 - \rho^2)(P_1 + Q + 2\rho\sqrt{P_1Q} + N)}{P_1Q(1 - \rho^2)(1 - \alpha)^2 + N(P_1 + \alpha^2Q + 2\alpha\rho\sqrt{P_1Q})} \right) \quad (4.7a)$$

$$r_2(\rho, \alpha) := \frac{1}{2} \log \left(1 + \frac{P_2}{N + \frac{P_1Q(1 - \rho^2)(1 - \alpha)^2}{(P_1 + \alpha^2Q + 2\alpha\rho\sqrt{P_1Q})}} \right) \quad (4.7b)$$

$$r_3(\rho, \alpha) := \frac{1}{2} \log \left(\frac{P_1(1 - \rho^2)(P_1 + P_2 + Q + 2\rho\sqrt{P_1Q} + N)}{P_1Q(1 - \rho^2)(1 - \alpha)^2 + N(P_1 + \alpha^2Q + 2\alpha\rho\sqrt{P_1Q})} \right) \quad (4.7c)$$

for a given $-1 \leq \rho \leq 0$, and a given $\alpha \in \mathcal{A}(\rho)$, where

$$\mathcal{A}(\rho) = \{x \in \mathbb{R} : r_1(\rho, x) \geq 0, r_2(\rho, x) \geq 0, r_3(\rho, x) \geq 0\}.$$

Theorem 4 *Let $\mathcal{R}^i(\rho, \alpha)$ be the set of all rate pairs (R_1, R_2) satisfying $R_1 \leq r_1(\rho, \alpha)$, $R_2 \leq r_2(\rho, \alpha)$, and $R_1 + R_2 \leq r_3(\rho, \alpha)$ for given $-1 \leq \rho \leq 0$ and $\alpha \in \mathcal{A}(\rho)$. Let*

$$\mathcal{R}_{\text{GAUS}}^i = \text{cl}\{\text{co}\{\cup_{-1 \leq \rho \leq 0, \alpha \in \mathcal{A}(\rho)} \mathcal{R}^i(\rho, \alpha)\}\}. \quad (4.8)$$

Then, the capacity region $\mathcal{C}_{\text{GAUS}}$ of the Gaussian MAC with one informed encoder satisfies $\mathcal{R}_{\text{GAUS}}^i \subseteq \mathcal{C}_{\text{GAUS}}$.

Proof: Our results for the DM MAC can readily be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [16]. The informed encoder uses generalized DPC that allows arbitrary correlation between the codeword \mathbf{X}_1^n and the known CSI \mathbf{S}^n . Fix a correlation parameter $-1 \leq \rho \leq 0$. We then consider the auxiliary random variable $\mathbf{U}_1 = \mathbf{X}_1 + \alpha\mathbf{S}$, where α is a real number whose range will be discussed later, \mathbf{X}_1 and \mathbf{S} are correlated with correlation coefficient ρ , $\mathbf{X}_1 \sim \mathcal{N}(0, P_1)$, and $\mathbf{S} \sim \mathcal{N}(0, Q)$. We consider $\mathbf{X}_2 \sim \mathcal{N}(0, P_2)$. Encoding and decoding are done similar to the those for the discrete memoryless case. We evaluate (4.2) using the jointly Gaussian distribution of random variables \mathbf{S} , \mathbf{X}_1 , \mathbf{U}_1 , \mathbf{X}_2 , \mathbf{Z} , and \mathbf{Y} for a given (ρ, α) and obtain $\mathcal{R}^i(\rho, \alpha)$. Also note that we restrict α to $\mathcal{A}(\rho) = \{\alpha : \alpha \in \mathbb{R}, r_1(\rho, \alpha) \geq 0, r_2(\rho, \alpha) \geq 0, r_3(\rho, \alpha) \geq 0\}$ for a given ρ . By varying ρ and α , we obtain different achievable rate regions $\mathcal{R}^i(\rho, \alpha)$. The union of regions $\mathcal{R}^i(\rho, \alpha)$ obtained by varying ρ and α gives an inner bound for the Gaussian MAC. Finally taking the closure and the convex hull operations completes the proof.

The following proposition gives an outer bound for the capacity region of the Gaussian MAC with one informed encoder. We do not provide a proof because this bound is same as the capacity region of the additive white Gaussian MAC with *all* informed encoders [18, 24] as well as the the capacity region of the additive white Gaussian MAC with CSI known at the decoder.

Proposition 2 *Let $\mathcal{R}_{\text{GAUS}}^o$ be the set of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right) \quad (4.9a)$$

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right) \quad (4.9b)$$

$$R_1 + R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_1 + P_2}{N} \right). \quad (4.9c)$$

Then, the capacity region $\mathcal{C}_{\text{GAUS}}$ for the Gaussian MAC with one informed encoder satisfies $\mathcal{C}_{\text{GAUS}} \subseteq \mathcal{R}_{\text{GAUS}}^o$.

Numerical Example

Let us consider a numerical example with $P_2 = 50$, $Q = 20$, and $N = 60$ for two interesting cases ($P_1 < Q$ and $P_1 = Q$). Figure 4.4 and Figure 4.5 depict both the inner bound using generalized DPC given in Theorem 4 and the outer bound specified in Proposition 2 for the cases $P_1 = 15$ ($P_1 < Q$) and $P_2 = 20$ ($P_1 = Q$), respectively. In both cases, also shown for comparison are the following: an inner bound using DPC alone, or the generalized DPC with $\rho = 0$ and α as parameter at the informed encoder; and the capacity region for the case in which the CSI is known at neither the encoders nor the decoder.

These results suggest that the informed encoder can help the uninformed encoder using DPC as well as generalized DPC. Even though the CSI is known only at one encoder, both the encoders benefit from this situation by allowing negative correlation between the channel input X_1 and the CSI S at the informed encoder since the negative correlation allows the informed encoder to partially cancel the

CSI. The achievable rate region $\mathcal{R}^i(0, \alpha)$ obtained by applying DPC [9] with α as parameter is always contained in $\mathcal{R}_{\text{GAUS}}^i$ in (4.8). In contrast to the case of CSI available to both the encoders [18, 24], DPC alone is insufficient.

Asymptotic Analysis

In this section, we discuss the inner bound in Theorem 4 as $Q \rightarrow \infty$.

Definition 9 Let $\mathcal{R}^i(\rho, \alpha)$ be the set of all rate pairs (R_1, R_2) satisfying $R_1 \leq \lim_{Q \rightarrow \infty} r_1(\rho, \alpha)$, $R_2 \leq \lim_{Q \rightarrow \infty} r_2(\rho, \alpha)$, and $R_1 + R_2 \leq \lim_{Q \rightarrow \infty} r_3(\rho, \alpha)$ for a given $-1 \leq \rho \leq 0$ and $\alpha \in \mathcal{A}(\rho) = \{x \in \mathbb{R} : 0 \leq x \leq \frac{2P_1(1-\rho^2)}{P_1(1-\rho^2)+N}\}$, where

$$\lim_{Q \rightarrow \infty} r_1(\rho, \alpha) := \frac{1}{2} \log \left(\frac{P_1(1-\rho^2)}{P_1(1-\rho^2)(1-\alpha)^2 + \alpha^2 N} \right) \quad (4.10a)$$

$$\lim_{Q \rightarrow \infty} r_2(\rho, \alpha) := \frac{1}{2} \log \left(1 + \frac{P_2}{N + \frac{P_1(1-\rho^2)(1-\alpha)^2}{\alpha^2}} \right) \quad (4.10b)$$

$$\lim_{Q \rightarrow \infty} r_3(\rho, \alpha) := \frac{1}{2} \log \left(\frac{P_1(1-\rho^2)}{P_1(1-\rho^2)(1-\alpha)^2 + \alpha^2 N} \right). \quad (4.10c)$$

As the variance of the CSI becomes very large, i.e., $Q \rightarrow \infty$, the inner bound in Theorem 4 becomes

$$\mathcal{R}_{\text{GAUS}}^i = \text{cl}\{\text{co}\{\cup_{-1 \leq \rho \leq 0, \alpha \in \mathcal{A}(\rho)} \mathcal{R}^i(\rho, \alpha)\}\}.$$

Let us investigate how the uninformed encoder can benefit from the informed encoder's actions even as $Q \rightarrow \infty$. For this discussion, consider successive decoding in which the auxiliary codeword \mathbf{U}_1^n of the informed encoder is decoded first using the channel output \mathbf{Y}^n and then the codeword \mathbf{X}_2^n of the uninformed encoder is decoded using \mathbf{Y}^n and \mathbf{U}_1^n . In the limit as $Q \rightarrow \infty$, \mathbf{U}_1^n can be decoded first with arbitrary low probability of error if R_1 satisfies

$$R_1 \leq \frac{1}{2} \log \left(\frac{P_1(1-\rho^2)}{P_1(1-\rho^2)(1-\alpha)^2 + \alpha^2(P_2 + N)} \right), \quad (4.11)$$

where $\rho \in [-1, 0]$ and $0 \leq \alpha \leq \frac{2P_1(1-\rho^2)}{P_1(1-\rho^2)+P_2+N}$. The right hand side of (4.11) is obtained by calculating the expression $\mathbb{I}(\mathbf{U}_1; \mathbf{Y}) - (\mathbf{U}_1, \mathbf{S})$ for the assumed jointly

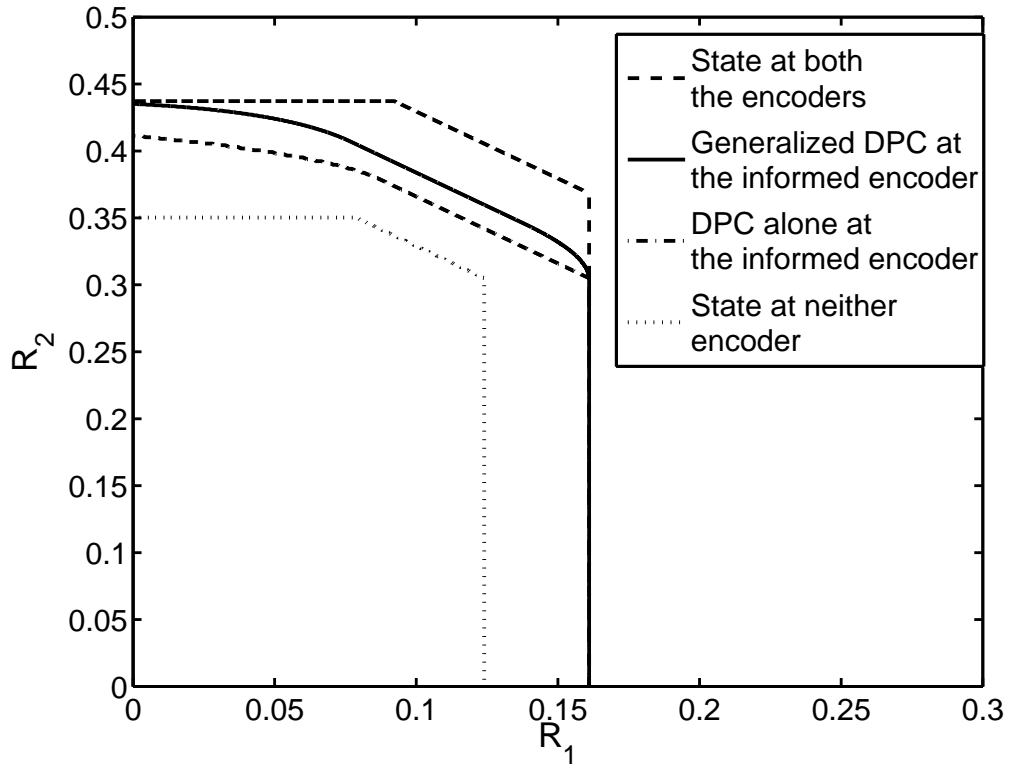


Figure 4.4. An achievable region for Gaussian multiple access channel with $P_1 = 15$, $P_2 = 50$, $Q = 20$, and $N = 60$.

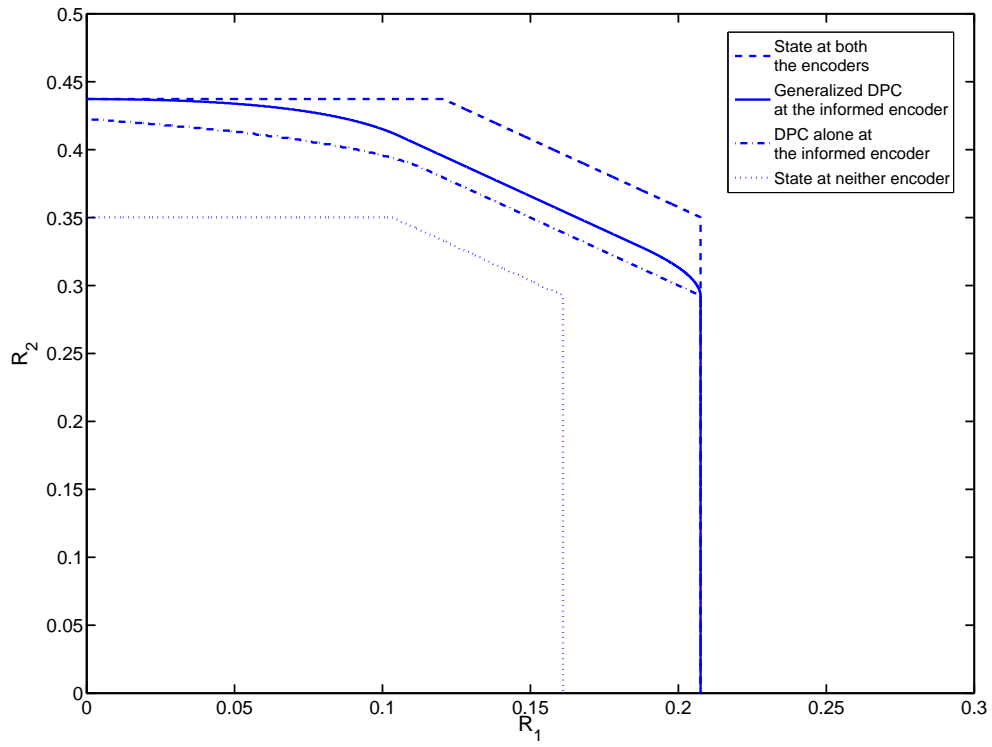


Figure 4.5. An achievable region for Gaussian multiple access channel with $P_1 = 20$, $P_2 = 50$, $Q = 20$, and $N = 60$.

Gaussian distribution and letting $Q \rightarrow \infty$. The channel output can be written as $Y_i = \mathbf{U}_{1,i} + X_{2,i} + (1 - \alpha)S_i + Z_i$ because $\mathbf{U}_{1,i} = X_{1,i} + \alpha S_i$ for $i \in \{1, 2, \dots, n\}$. The estimate of $(1 - \alpha)S_i$ using $\mathbf{U}_{1,i}$ is denoted as \hat{S}_i for $i \in \{1, 2, \dots, n\}$. Using \hat{S}^n and \mathbf{U}_1^n , we can generate a new channel output for decoding X_2^n as

$$\tilde{Y}_i = Y_i - \mathbf{U}_{1,i} - \hat{S}_i = X_{2,i} + Z_i + ((1 - \alpha)S_i - \hat{S}_i)$$

for $i \in \{1, 2, \dots, n\}$. Since all random variables are identical, we omit the subscript i for further discussion. Then, the variance of total noise present in elements of \tilde{Y}^n for decoding X_2^n is $N + \frac{P_1(1-\rho^2)(1-\alpha)^2}{\alpha^2}$, where N is the variance of Z , and $\frac{P_1(1-\rho^2)(1-\alpha)^2}{\alpha^2}$ is the error of estimating $(1 - \alpha)S$ from \mathbf{U}_1 . Then the message of the uninformed encoder can be decoded with arbitrarily low probability of error if $R_2 \leq \lim_{Q \rightarrow \infty} r_2(\rho, \alpha)$ for given $\rho \in [-1, 0]$ and $0 \leq \alpha \leq \frac{2P_1(1-\rho^2)}{P_1(1-\rho^2)+P_2+N}$. Even if the variance of the additive CSI becomes infinite, nonzero rate for the uninformed encoder can be achieved because the estimation error is finite for $\rho \in [-1, 0]$ due to the increase of the variance of \mathbf{U}_1 with the increase of the CSI variance.

Our aim is to minimize the variance of the estimation error $((1 - \alpha)S - \hat{S})$ to maximize $r_2(\rho, \alpha)$ over ρ and α . Since the right hand side of (4.11) becomes non-negative for $0 \leq \alpha \leq \frac{2P_1(1-\rho^2)}{P_1(1-\rho^2)+P_2+N}$ and $\rho \in [-1, 0]$, we consider only these values. The variance of the estimation error is decreasing in both $\rho \in [-1, 0]$ and $\alpha \in [0, 1]$ and is increasing in the remaining range of α . Then, $r_2(\rho, \alpha)$ achieves its maximum at $\rho = 0$ and $\alpha = \min\{1, \frac{2P_1}{P_1+P_2+N}\}$. If $P_1 \geq P_2 + N$, so that R_1 is nonnegative, then

$$R_2 \leq \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$$

is achievable. In this case, the uninformed encoder fully benefits from CSI at the informed encoder even though the interfering CSI is very large. If $P_1 < P_2 + N$, then $R_2 \leq \lim_{Q \rightarrow \infty} r_2(0, \alpha^*)$ is achievable where $\alpha^* = \frac{2P_1}{P_1+P_2+N}$. In both the cases, the generalized DPC with $\rho = 0$ is optimal in terms of assisting the uninformed

encoder, contrary to the finite CSI variance case. This makes sense because, if the CSI has infinite variance, then it is impossible for the informed encoder to explicitly cancel it with finite power.

Now consider successive decoding in the reverse order in which \mathbf{X}_2^n is decoded first using \mathbf{Y}^n and then \mathbf{U}_1^n is decoded using \mathbf{Y}^n and \mathbf{X}_2^n . As $Q \rightarrow \infty$, \mathbf{X}_2^n can be decoded with arbitrary low probability of error if $R_2 \leq \lim_{Q \rightarrow \infty} \mathbb{I}(\mathbf{X}_2, \mathbf{Y}) = 0$. This means that only $R_2 = 0$ is achievable. Then, $R_1 \leq \frac{1}{2} \log \left(1 + \frac{P_1}{N} \right)$ is achievable with $\rho = 0$ and $\alpha = \frac{P_1}{P_1 + N}$.

4.2 Dependent or Degraded Messages

In this section, we study the capacity region for the general model shown in Figure 4.1 with rate R_2 being zero (no private message of uninformed encoder, $\mathcal{W}_2 = \emptyset$). In this case, we consider \mathcal{X}_1 , \mathcal{X}_2 , \mathcal{S} , and \mathcal{Y} to all be discrete and finite alphabets; and all probability laws are to be interpreted as probability mass functions. This model is considered for the Gaussian case in [44].

The Capacity Region

The following theorem presents the capacity region for the model considered in this section.

Theorem 5 *The capacity region \mathcal{C} for the model considered in this section is the closure of all rate pairs (R_1, R_2) satisfying*

$$R_1 < \mathbb{I}(\mathbf{U}; \mathbf{Y} | \mathbf{X}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}; \mathbf{S} | \mathbf{X}_2, \mathbf{Q}) \quad (4.12a)$$

$$R_1 + R_2 < \mathbb{I}(\mathbf{U}, \mathbf{X}_2; \mathbf{Y} | \mathbf{Q}) - \mathbb{I}(\mathbf{U}; \mathbf{S} | \mathbf{X}_2, \mathbf{Q}) \quad (4.12b)$$

for some random variables $(\mathbf{Q}, \mathbf{S}, \mathbf{U}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ whose distribution is of the form

$$p(\mathbf{q})p(\mathbf{s})p(\mathbf{x}_2 | \mathbf{q})p(\mathbf{u}, \mathbf{x}_1 | \mathbf{s}, \mathbf{x}_2, \mathbf{q})p(\mathbf{y} | \mathbf{s}, \mathbf{x}_1, \mathbf{x}_2),$$

where, \mathbf{U} is an auxiliary random variable, and \mathbf{Q} is a time-sharing auxiliary random variable.

Proof: A proof of the above theorem is given in Appendix B.3.

Remarks:

- To compute the region in the above theorem, it is sufficient to consider the auxiliary random variables \mathbf{U} and \mathbf{Q} with $|\mathcal{U}| \leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}|$ and $|\mathcal{Q}| \leq 3$, respectively.

4.3 Summary

In the independent messages case, we derived inner and outer bounds for the capacity region in DM case and Gaussian memoryless case. Since there is gap between the inner bound and outer bounds for the capacity region, the capacity region is still not known in either the discrete memoryless case or the Gaussian case. For Gaussian models, we proposed the generalized DPC in which the channel input and the channel state are correlated. Our results suggest that DPC is not optimal in terms of the capacity region in contrast to the single-user model with CSI known at the encoder and MAC with CSI known at *all* encoders. However, in one case of degraded messages, we derived the capacity region for the discrete memoryless MAC with asymmetric encoder state.

The coding schemes, generalized DPC, and other observations made in the chapter can also be applied to understand other multi-user models with asymmetric encoder side information such as interference channels, relay channels, and so forth. In this chapter, the side information available at the encoders is fully correlated to the channel state. These models can be extended by considering the side information at the encoders partially correlated to the channel state. In this chapter, the decoder is concerned with the decoding of only the messages. In the next chapter, we consider MAC with encoder side information in which the decoder is concerned with not only decoding of messages but also recovering the channel state.

CHAPTER 5

STATE-DEPENDENT MULTIPLE ACCESS CHANNELS WITH ENCODER SIDE INFORMATION AND RECOVERY OF SOME STATES

In this chapter, we consider a state-dependent multiple access channel (MAC) with different side information available at different encoders and lossless recovery of some states at the decoder. As shown in Figure 5.1, we consider a state-dependent MAC whose output Y controlled by a state pair $(S_1, S_2) \in \mathcal{S}_1 \times \mathcal{S}_2$, the channel inputs $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ from two encoders through a memoryless probability law $p(y|x_1, s_1, x_2, s_2)$. In this thesis, we consider two encoder model but the results can be extended to any number of users. Encoder i , provided with W_i and the channel state S_i^n , generates X_i^n such that the average per-letter distortion between S_i^n and X_i^n is less than Δ_i according to a given bounded distortion measure $d_i(\cdot, \cdot)$, $i = 1, 2$.¹ This encoder constraint captures power constraints in some applications, and distortion constraints in information embedding (IE) applications.

For this model, we consider the following three cases in recovering, in the sense of probability of error going to zero, the messages and the state sequences at the decoder.

- **CaseA, Recovery of Neither State:** The decoder recovers (W_1, W_2) from Y^n .

¹ $S_i^n = \{S_{i,1}, S_{i,2}, \dots, S_{i,n}\}$ and $X_i^n = \{X_{i,1}, X_{i,2}, \dots, X_{i,n}\}$

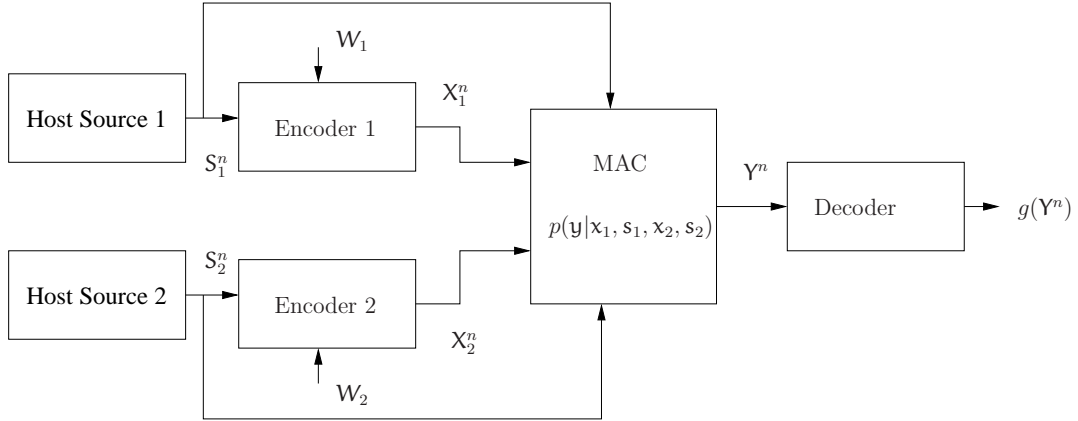


Figure 5.1. Block diagram of state-dependent multiple access channel with two components of state.

- **CaseB, Recovery of One State :** The decoder recovers (W_1, W_2) along with the one state sequence from Y^n . Without loss of generality, we can assume that the state available at Encoder 2 is recovered at the decoder.
- **CaseC, Recovery of Both States:** The decoder recovers (W_1, W_2) and (S_1^n, S_2^n) from Y^n .

As in the single-user case, the model considered in this chapter is also closely related to IE in MAC in which each user provided with the host sequence wants to embed its message into the host sequence such that the per-letter distortion between the host sequence and the embedded sequences satisfies a distortion constraint. From the view of distributed IE, the state signal in the above model is the host signal. In this chapter, we study the capacity region of the model shown in Figure 5.1 in various decoding scenarios from the view of MAC IE. So, we use the term “host signal” for the term “state”, the term “embedded signal” for the term “channel inputs” .

Since our model shown in Figure 5.1 considers scenarios in which the MAC

output potentially depends on both the embedded signals and the host signals. The model considered in the paper specializes to various distributed RIE channel models with lossless state or host signal recovery at the decoder, including

1. A discrete memoryless multi-user version of the single-user data hiding model considered in [21, 22].
2. A discrete memoryless multi-user version of the single-user information embedding model considered in [6, 8].

Let us now formally discuss the model shown in Figure 5.1. The host source i generates a sequence $\mathbf{S}_i^n = \{S_{i,1}, S_{i,2}, \dots, S_{i,n}\}$ of symbols from the discrete alphabet \mathcal{S}_i , where $i = 1, 2$. We assume that the host sequence pair $(\mathbf{S}_1^n, \mathbf{S}_2^n)$ is generated by repeated independent drawings of a pair of discrete random variables (S_1, S_2) from a given joint distribution $p(s_1, s_2)$. The host sequence \mathbf{S}_i^n is non-causally known at Encoder i , for $i = 1, 2$. The message source at Encoder i produces the message index $W_i \in \mathcal{W}_i = \{1, 2, \dots, M_i\}$ with equal probability $1/M_i$, for $i = 1, 2$. The message index at any encoder is independent of all host sequences and also independent of the messages at all other encoders. The rate at Encoder i , in bits per channel use, is defined as $R_i = (1/n) \log_2(M_i)$. In this chapter, the channel input alphabets \mathcal{X}_1 and \mathcal{X}_2 , and the channel output alphabet \mathcal{Y} are discrete.

Definition 10 A $(M_1, M_2, D_1^{(n)}, D_2^{(n)}, n)$ MAC IE code consists of sequences of encoding functions at Encoder 1 and Encoder 2,

$$f_1^n : \mathcal{W}_1 \times \mathcal{S}_1^n \rightarrow \mathcal{X}_1^n, \quad \text{and} \quad f_2^n : \mathcal{W}_2 \times \mathcal{S}_2^n \rightarrow \mathcal{X}_2^n,$$

respectively, and a sequence of decoding functions,

- **Case A, Recovery of Neither Host or State** $g_A^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$
- **Case B, Recovery of One Host or State** $g_B^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2, \mathcal{S}_2^n)$
- **Case C, Recovery of Both Hosts or States** $g_C^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{S}_1^n, \mathcal{W}_2, \mathcal{S}_2^n)$

The distortions associated with MAC IE code are defined as $D_i^{(n)} = \mathbb{E}d_i(\mathbf{S}_i^n, \mathbf{X}_i^n)$ for the additive distortion function

$$d_i(\mathbf{S}_i^n, \mathbf{X}_i^n) = \frac{1}{n} \sum_{j=1}^n d_i(\mathbf{S}_{ij}, \mathbf{X}_{ij})$$

for some non-negative, bounded distortion functions $d_i(\mathbf{S}_{ij}, \mathbf{X}_{ij})$, where $i = 1, 2$.

The embedded signals \mathbf{X}_1^n and \mathbf{X}_2^n from Encoder 1 and Encoder 2, respectively, are transmitted across a MAC $p(\mathbf{y}|\mathbf{x}_1, \mathbf{s}_1, \mathbf{x}_2, \mathbf{s}_2)$ without feedback modeled as a memoryless conditional probability distribution

$$\Pr(\mathbf{y}^n = \mathbf{y}^n | \mathbf{x}_1^n, \mathbf{s}_1^n, \mathbf{x}_2^n, \mathbf{s}_2^n) = \prod_{j=1}^n p(\mathbf{y}_j | \mathbf{x}_{1j}, \mathbf{s}_{1j}, \mathbf{x}_{2j}, \mathbf{s}_{2j}). \quad (5.1)$$

Definition 11 A rate pair (R_1, R_2) for a given distortion pair (Δ_1, Δ_2) is said to be MAC IE achievable if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D_1^{(n)}, D_2^{(n)}, n)$ MAC IE codes (f_1^n, f_2^n, g^n) with $\lim_{n \rightarrow \infty} D_i^{(n)} \leq \Delta_i$, for $i = 1, 2$, and $\lim_{n \rightarrow \infty} P_e^n = 0$, where P_e^n is the probability of error defined appropriately for each case in the sequel of this section.

Definition 12 The MAC IE capacity region is the closure of the convex hull of the set of MAC IE achievable (R_1, R_2) rate pairs for a given distortion pair (Δ_1, Δ_2) .

Definition 13 For given $p(\mathbf{s}_1, \mathbf{s}_2)$ and $p(\mathbf{y}|\mathbf{x}_1, \mathbf{s}_1, \mathbf{x}_2, \mathbf{s}_2)$, let $\mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$ be the set of all random variable tuples $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{U}_1, \mathbf{X}_1), (\mathbf{U}_2, \mathbf{X}_2), \mathbf{Y})$ taking values in finite alphabets $\mathcal{Q}, \mathcal{S}, \mathcal{U}_1 \times \mathcal{X}_1, \mathcal{U}_2 \times \mathcal{X}_2$, and \mathcal{Y} , respectively, with joint distribution satisfying the conditions

a) $\sum_{\mathbf{q}, (\mathbf{u}_1, \mathbf{x}_1), (\mathbf{u}_2, \mathbf{x}_2), \mathbf{y}} p(\mathbf{q}, \mathbf{s}_1, \mathbf{s}_2, (\mathbf{u}_1, \mathbf{x}_1), (\mathbf{u}_2, \mathbf{x}_2), \mathbf{y}) = p(\mathbf{s}_1, \mathbf{s}_2),$

b)

$$p(\mathbf{q}, \mathbf{s}_1, \mathbf{s}_2, (\mathbf{u}_1, \mathbf{x}_1), (\mathbf{u}_2, \mathbf{x}_2), \mathbf{y}) = p(\mathbf{q})p(\mathbf{s}_1, \mathbf{s}_2)p(\mathbf{u}_1, \mathbf{x}_1 | \mathbf{s}_1, \mathbf{q})p(\mathbf{u}_2, \mathbf{x}_2 | \mathbf{s}_2, \mathbf{q}) \\ \times p(\mathbf{y} | \mathbf{x}_1, \mathbf{s}_1, \mathbf{x}_2, \mathbf{s}_2)$$

c) $\mathbb{E}d_i(\mathbf{S}_i, \mathbf{X}_i) \leq \Delta_i$, for $i = 1, 2$.

Definition 14 For given $p(\mathbf{s}_1, \mathbf{s}_2)$ and $p(\mathbf{y}|\mathbf{x}_1, \mathbf{s}_1, \mathbf{x}_2, \mathbf{s}_2)$, let $\mathcal{P}_{\text{MAC}}^o(\Delta_1, \Delta_2)$ be the set of all random variable tuples $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ taking values in finite alphabets $\mathcal{Q}, \mathcal{S}, \mathcal{X}_1, \mathcal{X}_2$, and \mathcal{Y} , respectively, with distribution satisfying the conditions

a) $\sum_{\mathbf{q}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}} p(\mathbf{q}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = p(\mathbf{s}_1, \mathbf{s}_2),$

b) $p(\mathbf{q}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{y}) = p(\mathbf{q})p(\mathbf{s}_1, \mathbf{s}_2)p(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{s}_1, \mathbf{s}_2, \mathbf{q})p(\mathbf{y} | \mathbf{x}_1, \mathbf{s}_1, \mathbf{x}_2, \mathbf{s}_2),$

c) $\mathbb{E}d_i(\mathbf{S}_i, \mathbf{X}_i) \leq \Delta_i$, for $i = 1, 2$.

5.1 Recovery of Neither Host or State

In this section, we derive an inner bound on the MAC IE capacity region for Case A, in which the decoder recovers only (W_1, W_2) from Y^n . We define the MAC IE capacity region $\mathcal{C}_{\text{MAC,A}}(\Delta_1, \Delta_2)$ as the closure of the set of all achievable rates (R_1, R_2) with $P_e^{(n)} := \mathbb{P}[(g_A^n(Y^n) \neq (W_1, W_2))] \rightarrow 0$ as $n \rightarrow \infty$. The following theorem provides an inner bound on the capacity region.

Proposition 3 *Let $\mathcal{R}_{\text{MAC,A}}^i(\Delta_1, \Delta_2)$ be the closure of the set of all rate pairs (R_1, R_2) such that*

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; \mathbf{U}_2, Y|Q) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}_1|Q), \quad (5.2a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{U}_1, Y|Q) - \mathbb{I}(\mathbf{U}_2; \mathbf{S}_2|Q), \quad (5.2b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; Y|Q) - \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{S}_1, \mathbf{S}_2|Q) \quad (5.2c)$$

for some $(Q, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{U}_1, \mathbf{X}_1), (\mathbf{U}_2, \mathbf{X}_2), Y) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$, where \mathbf{U}_1 and \mathbf{U}_2 are auxiliary random variables. Then, $\mathcal{R}_{\text{MAC,A}}^i(\Delta) \subseteq \mathcal{C}_{\text{MAC,A}}$.

Remarks

- The inner bound in Proposition 3 is similar to that in [47], which considers a Gaussian MAC with no host recovery, but the result here is for the discrete memoryless case. Because the coding procedures, and error events in [47] apply, we do not provide a proof here.
- To achieve the inner bound, distortion-constrained Gel'fand-Pinsker codes can be used to embed W_1 and W_2 into the host sequences S_1^n and S_2^n such that the distortion constraints Δ_1 and Δ_2 are met, respectively.

5.2 Recovery of One Host or State

In this section, we derive inner and outer bounds on the MAC IE capacity region for Case B, in which the decoder recovers (W_1, W_2, S_2^n) from Y^n . We define the MAC IE capacity region $\mathcal{C}_{\text{MAC,B}}(\Delta_1, \Delta_2)$ as the closure of the set of all MAC IE

achievable rates (R_1, R_2) with $P_e^{(n)} := \mathbb{P}[(g_B^n(Y^n) \neq (W_1, W_2, S_2^n))] \rightarrow 0$ as $n \rightarrow \infty$.

The following theorem provides an inner bound for the capacity region.

Proposition 4 *Let $\mathcal{R}_{\text{MAC},B}^i(\Delta_1, \Delta_2)$ be the closure of the set of all rate pairs (R_1, R_2) such that*

$$R_1 \leq \mathbb{I}(\mathbf{U}_1; Y | \mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}_1 | \mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}), \quad (5.3a)$$

$$R_2 \leq \mathbb{I}(\mathbf{X}_2, \mathbf{S}_2; Y | \mathbf{U}_1, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2 | \mathbf{U}_1, \mathbf{Q}), \quad (5.3b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{X}_2, \mathbf{S}_2; Y | \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}_1 | \mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) \quad (5.3c)$$

for some $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{U}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), Y) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$, where \mathbf{U}_1 and \mathbf{Q} are auxiliary random variables. Then, $\mathcal{R}_{\text{MAC},B}^i(\Delta_1, \Delta_2) \subseteq \mathcal{C}_{\text{MAC},B}(\Delta_1, \Delta_2)$

Remarks

- The inner bound in Proposition 4 is a special case of an inner bound in [25], which considers the state-dependent MAC with state known at one encoder and recovery of only messages at the decoder. To obtain the inner bound in Proposition 4, substitute $(\mathbf{X}_2, \mathbf{S}_2)$ in place of \mathbf{X}_2 into the inner bound in [25].
- To achieve the inner bound, distortion constrained Gel'fand-Pinsker coding is used to embed \mathbf{W}_1 into the host sequence \mathbf{S}_1^n , and distortion-constrained superposition coding is used to embed \mathbf{W}_2 into the host sequence \mathbf{S}_2^n .
- If we choose $\mathbf{U}_2 = (\mathbf{X}_2, \mathbf{S}_2)$ in Proposition 3, we obtain the inner bound in Proposition 4. Thus, $\mathcal{R}_{\text{MAC},B}^i(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{\text{MAC},A}^i(\Delta_1, \Delta_2)$.

5.3 Recovery of Both Hosts or States

In this section, we derive inner and outer bounds on the MAC IE capacity region for Case C, in which the decoder recovers (W_1, S_1^n, W_2, S_2^n) from Y^n . We define the MAC IE capacity region $\mathcal{C}_{\text{MAC},C}(\Delta_1, \Delta_2)$ as the closure of all achievable rates (R_1, R_2) with $P_e^{(n)} := \mathbb{P}[(g(Y^n) \neq (W_1, S_1^n, W_2, S_2^n))] \rightarrow 0$ as $n \rightarrow \infty$. The following theorem obtains an inner bound for the capacity region.

Theorem 6 Let $\mathcal{R}_{\text{MAC,C}}^i(\Delta_1, \Delta_2)$ be the set of all rate pairs (R_1, R_2) such that

$$R_1 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_1|\mathbf{S}_2)], \quad (5.4a)$$

$$R_2 < [\mathbb{I}(\mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{S}_1, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2|\mathbf{S}_1)], \quad (5.4b)$$

$$R_1 + R_2 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1, \mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{H}(\mathbf{S}_1, \mathbf{S}_2)], \quad (5.4c)$$

for some $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$. Then,

$$\mathcal{R}_{\text{MAC,C}}^i(\Delta_1, \Delta_2) \subseteq \mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2).$$

Proof: See Appendix C.1

The following theorem gives an outer bound for the capacity region if \mathbf{S}_1 and \mathbf{S}_2 are correlated.

Theorem 7 Let $\mathcal{R}_{\text{MAC,C}}^o(\Delta_1, \Delta_2)$ be the set of all rate pairs (R_1, R_2) such that

$$R_1 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_1|\mathbf{S}_2)], \quad (5.5a)$$

$$R_2 < [\mathbb{I}(\mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{S}_1, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2|\mathbf{S}_1)], \quad (5.5b)$$

$$R_1 + R_2 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1, \mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{H}(\mathbf{S}_1, \mathbf{S}_2)], \quad (5.5c)$$

for some $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^o(\Delta_1, \Delta_2)$. If the host random variables \mathbf{S}_1 and \mathbf{S}_2 are correlated, then

$$\mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{\text{MAC,C}}^o(\Delta_1, \Delta_2).$$

If the host random variables \mathbf{S}_1 and \mathbf{S}_2 are independent, then

$$\mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{\text{MAC,C}}^i(\Delta_1, \Delta_2).$$

Proof: See Appendix C.2

The following corollary of Theorem 6 and Theorem 7 states the MAC IE capacity region for a given pair of distortion constraints (Δ_1, Δ_2) if the host random variables \mathbf{S}_1 and \mathbf{S}_2 are independent.

Corollary 2 If the host random variables \mathbf{S}_1 and \mathbf{S}_2 are independent, then the capacity region $\mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2)$ is the closure of the set of all rate pairs (R_1, R_2) such that

$$R_1 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_1|\mathbf{S}_2)], \quad (5.6a)$$

$$R_2 < [\mathbb{I}(\mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{S}_1, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2|\mathbf{S}_1)], \quad (5.6b)$$

$$R_1 + R_2 < [\mathbb{I}(\mathbf{X}_1, \mathbf{S}_1, \mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{H}(\mathbf{S}_1, \mathbf{S}_2)], \quad (5.6c)$$

for some $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$.

Remarks

- To compute either (5.4) or (5.5), it is sufficient to consider time-sharing random variable Q with $|\mathcal{Q}| \leq 4$ by Caratheodory's theorem [12].
- In most communication scenarios, message transmission rates of zero are achievable. However, in this model, message transmission rates of zero can be unachievable if the host source pair $p(\mathbf{s}_1, \mathbf{s}_2)$ is such that the upper bounds on R_1 , R_2 and $R_1 + R_2$ in (5.6) are negative. This is because we require host recovery at the decoder as well.

5.4 Summary

In this chapter, we considered the state-dependent MAC with recovery of some states at the decoder or MAC IE with recovery of some hosts. We derived inner and outer bounds for the capacity region when all states or hosts are recovered at the decoder. These inner and outer bounds coincide when all states of the channel or all hosts are independent. We derived inner bounds for the model considered in this chapter when some, but not all, hosts are recovered at the decoder.

In this chapter, we considered the model in which the decoder considers the lossless recovery of the host sequences. This model is suitable for applications with discrete alphabets. Since lossless recovery is not applicable for models with continuous alphabets, it is important to consider partial state recovery in such models.

CHAPTER 6

STATE-DEPENDENT BROADCAST CHANNELS WITH ENCODER SIDE INFORMATION AND STATE RECOVERY AT SOME DECODERS

In this chapter, we consider a state-dependent broadcast channel with non-causal encoder side information at the encoder and lossless recovery of state at *some* decoders. As shown in Figure 6.1, we consider a state-dependent broadcast channel whose outputs Y for Decoder 1 and Z for Decoder 2 are controlled by the channel state $S \in \mathcal{S}$ and the channel input $X \in \mathcal{X}$ through a memoryless probability law $p(\mathbf{y}, \mathbf{z} | \mathbf{s}, \mathbf{x})$. In this chapter, some decoders are concerned with recovering the channel state alongwith the decoding messages intended for them. We develop results for two decoder model but, in principle, results can be extended to any number of decoders. In this model, the encoder, provided with message pair (W_1, W_2) and the state S^n , generates X^n such that per-letter distortion between the state S^n and the channel input X^n satisfies a given constraint Δ according to a given bounded measure $d(\cdot, \cdot)$ ¹. This encoder constraint on the channel input captures power constraint in some applications, and distortion constraint between the host and the embedded signal in information embedding (IE) applications.

For this model, we consider the following four cases in recovering, in the sense of probability of error going to zero, the messages and the host sequences at the decoders

¹ $S^n = \{S_1, S_2, \dots, S_n\}$ and $X^n = \{X_1, X_2, \dots, X_n\}$

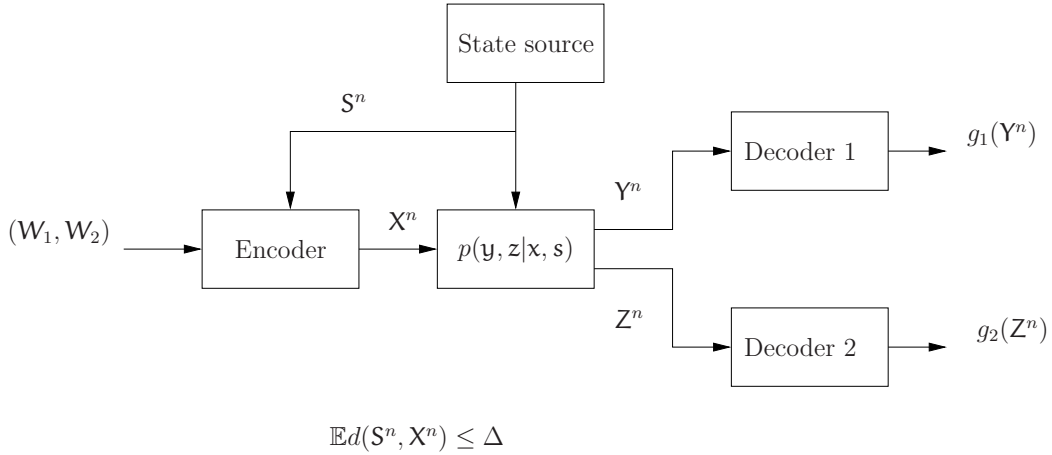


Figure 6.1. Block diagram for state-dependent broadcast channel.

- **Case A, No State Recovery:** Decoder 1 recovers (W_1, W_2) from Y^n ; Decoder 2 recovers W_2 from Z^n .
- **Case B, State Recovery at Decoder 1:** Decoder 1 recovers (W_1, W_2) and S^n from Y^n ; Decoder 2 recovers W_2 from Z^n .
- **Case C, State Recovery at Both Decoders:** Decoder 1 recovers (W_1, W_2) and S^n from Y^n ; Decoder 2 recovers W_2 and S^n from Z^n .
- **Case D, State Recovery at Decoder 2:** Decoder 1 recovers (W_1, W_2) from Y^n ; Decoder 2 recovers W_2 and S^n from Z^n .

As in the single-user case, the model considered in this chapter is also closely related to broadcast IE in which the encoder embeds information into the provided host S^n such that the distortion between the host S^n and the embedded signal X^n satisfies a given distortion constraint, each decoder is provided with the noisy version of the embedded signal. Depending on application, some decoders consider reversibility of the host along with the messages intended for them. From the view of

broadcast IE, the state signal \mathbf{S} in the above model is the host signal. In this chapter, we study the capacity region of the model shown in Figure 5.1 in various scenarios from the view of broadcast IE. So, we use the term “host” for the term “state”, the term “embedded signal” for the term “channel input”. In this chapter, we restrict broadcast probability laws to only physically degraded broadcast probability laws, i.e., one channel output is degraded or noisy version of the other channel output. Formally speaking, this means that $p(\mathbf{y}, \mathbf{z}|\mathbf{x}, \mathbf{s}) = p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\mathbf{z}|\mathbf{y})$, i.e., \mathbf{Z} is degraded versions of \mathbf{Y} . We assume that all alphabets are discrete. Since Decoder 1 receives the better channel output \mathbf{Y} and Decoder 2 receives the worse channel output \mathbf{Z} , we call Decoder 1 and Decoder 2 as *better* decoder and *worse* decoder, respectively.

Let us now formally define broadcast IE model shown in Figure 6.1. A host sequence $\mathbf{S}^n = (\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_n)$ is an independent and identically distributed (i.i.d.) discrete random sequence whose elements are drawn with probability mass function $p(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}$. We assume that the host sequence \mathbf{S}^n is non-causally known at the encoder. The encoder embeds a message pair $(\mathbf{W}_1, \mathbf{W}_2)$ into the host sequence \mathbf{S}^n such that the average distortion between \mathbf{S}^n and the embedded sequence \mathbf{X}^n satisfies a given distortion constraint Δ . The messages $\mathbf{W}_1 \in \{1, 2, \dots, M_1\}$ and $\mathbf{W}_2 \in \{1, 2, \dots, M_2\}$ are drawn equally likely with probabilities $1/M_1$ and $1/M_2$, respectively. Then the rate of message \mathbf{W}_i is given by $R_i = (1/n) \log_2 M_i$ bits per channel use, for $i = 1, 2$. It is also assumed that the message \mathbf{W}_i is independent of the other message and the host sequence for $i = 1, 2$. In this chapter, the host alphabet \mathcal{S} , the channel input alphabet \mathcal{X} , and the channel output alphabets \mathcal{Y} and \mathcal{Z} are discrete.

Definition 15 A $(M_1, M_2, D^{(n)}, n)$ broadcast IE code consists of a sequence of encoding functions at the encoder

$$f^n : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \rightarrow \mathcal{X}^n,$$

and a sequence of decoding functions at Decoder 1 and Decoder 2

- **Case A, No State or Host Recovery** $g_{1,A}^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$ and $g_{2,A}^n : \mathcal{Z}^n \rightarrow \mathcal{W}_2$
- **Case B, State or Host Recovery at the Better Decoder** $g_{1,B}^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2, \mathcal{S}^n)$ and $g_{2,B}^n : \mathcal{Z}^n \rightarrow \mathcal{W}_2$
- **Case C, State or Host Recovery at Both Decoders** $g_{1,C}^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2, \mathcal{S}^n)$ and $g_{2,C}^n : \mathcal{Z}^n \rightarrow (\mathcal{W}_2, \mathcal{S}^n)$
- **Case D, State or Host Recovery at the Worse Decoder** $g_{1,D}^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \mathcal{W}_2)$ and $g_{2,D}^n : \mathcal{Z}^n \rightarrow (\mathcal{W}_2, \mathcal{S}^n)$,

respectively. The associated distortion is defined as $D^{(n)} = \mathbb{E}d(\mathbf{S}^n, \mathbf{X}^n)$, where $d(\mathbf{S}^n, \mathbf{X}^n) = (1/n) \sum_{j=1}^n d(\mathbf{S}_j, \mathbf{X}_j)$ for given non-negative bounded distortion measure $d(\cdot, \cdot)$.

The embedded signal \mathbf{X}^n is transmitted across a discrete memoryless degraded broadcast channel (DM-DBC) with state, $p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\mathbf{z}|\mathbf{y})$, modeled as a memoryless conditional probability distribution

$$\Pr(\mathbf{Y}^n = \mathbf{y}^n, \mathbf{Z}^n = \mathbf{z}^n | \mathbf{x}^n, \mathbf{s}^n) = \prod_{j=1}^n p(\mathbf{y}_j | \mathbf{x}_j, \mathbf{s}_j) p(\mathbf{z}_j | \mathbf{y}_j). \quad (6.1)$$

Definition 16 A broadcast IE rate pair (R_1, R_2) for a given distortion Δ is said to be achievable if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D^{(n)}, n)$ broadcast IE codes (f^n, g_1^n, g_2^n) with $\lim_{n \rightarrow \infty} D^{(n)} \leq \Delta$ and $\lim_{n \rightarrow \infty} P_e^n = 0$, where P_e^n is the probability of error defined appropriately for each case in the sequel.

Definition 17 For a given $p(\mathbf{s})$ and $p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\mathbf{z}|\mathbf{y})$, let $\mathcal{P}(\Delta)$ be the collection of random variables $(\mathbf{T}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ with joint probability mass function satisfying the following conditions

- $p(\mathbf{t}, \mathbf{s}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{t}, \mathbf{s}, \mathbf{x})p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\mathbf{z}|\mathbf{y})$
- $\sum_{\mathbf{t} \in \mathcal{T}, \mathbf{x} \in \mathcal{X}} p(\mathbf{t}, \mathbf{x}, \mathbf{s}) = p(\mathbf{s})$
- $\mathbb{E}d(\mathbf{S}, \mathbf{X}) \leq \Delta$,

where \mathbf{T} is an auxiliary random variable.

6.1 No State or Host Recovery

In this section, we state inner and outer bounds for the broadcast IE capacity region in Case A, in which Decoder 1 recovers (W_1, W_2) from Y^n and Decoder 2 recovers W_2 from Z^n . The broadcast IE capacity region $\mathcal{C}_A(\Delta)$ is the closure of all achievable rates (R_1, R_2) with $P_e^{(n)} := \Pr[(g_{1,A}^n(Y^n) \neq (W_1, W_2) \text{ or } g_{2,A}^n(Z^n) \neq W_2)] \rightarrow 0$ as $n \rightarrow \infty$.

Proposition 5 *Let $\mathcal{R}_A^i(\Delta)$ be the closure of the set of all rate pairs (R_1, R_2) such that*

$$R_1 \leq \mathbb{I}(\mathbf{V}; \mathbf{Y}|\mathbf{U}) - \mathbb{I}(\mathbf{V}; \mathbf{S}|\mathbf{U}), \quad (6.2a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z}) - \mathbb{I}(\mathbf{U}; \mathbf{S}), \quad (6.2b)$$

for some $((\mathbf{U}, \mathbf{V}), \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$, where \mathbf{U} and \mathbf{V} are auxiliary random variables with alphabet sizes satisfying $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}|+1$ and $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}|+1)$, respectively. Let $\mathcal{R}_A^o(\Delta)$ be the closure of the set of all rate pairs (R_1, R_2) such that

$$R_1 \leq \mathbb{I}(\mathbf{V}; \mathbf{Y}|\mathbf{U}, \mathbf{W}) - \mathbb{I}(\mathbf{V}; \mathbf{S}|\mathbf{U}, \mathbf{W}), \quad (6.3a)$$

$$R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z}) - \mathbb{I}(\mathbf{U}; \mathbf{S}), \quad (6.3b)$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}, \mathbf{V}, \mathbf{W}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}, \mathbf{V}, \mathbf{W}; \mathbf{S}), \quad (6.3c)$$

for some $((\mathbf{U}, \mathbf{V}, \mathbf{W}), \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$, where \mathbf{U} , \mathbf{W} , and \mathbf{V} are auxiliary random variables with alphabet sizes satisfying $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}| + 2$, $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}| + 2) + 1$, and $|\mathcal{W}| \leq (|\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}| + 2) + 1)(|\mathcal{X}||\mathcal{S}| + 2)|\mathcal{X}||\mathcal{S}| + 1$, respectively. Then, $\mathcal{R}_A^i(\Delta) \subseteq \mathcal{C}_A(\Delta) \subseteq \mathcal{R}_A^o(\Delta)$.

Remarks

The inner and outer bounds in Proposition 5 are slightly different from those in [45], which does not consider an encoder distortion constraint. Although essentially the same proofs in [45] apply, here there is an additional constraint on the joint probability mass functions $\mathcal{P}(\Delta)$ to limit the average distortion between the host \mathbf{S} and the channel input \mathbf{X} to be at most Δ . To achieve the inner bound, Gel'fand-Pinsker codes can be used to embed the messages (W_1, W_2) into the host sequence \mathbf{S}^n .

6.2 State or Host Recovery at the Better Decoder

In this section, we derive inner and outer bounds on the broadcast IE capacity region in Case B, in which Decoder 1 recovers (W_1, W_2) and S^n from Y^n and Decoder 2 recovers only W_2 from Z^n . We define the broadcast IE capacity region $\mathcal{C}_B(\Delta)$ as the closure of all achievable rates (R_1, R_2) with $P_e^{(n)} := \Pr[(g_{1,B}(Y^n) \neq (W_1, W_2, \hat{S}^n) \text{ or } g_{2,B}(Z^n) \neq W_2] \rightarrow 0$ as $n \rightarrow \infty$. The following two theorems give inner and outer bounds for the capacity region in this case.

Theorem 8 *Let $\mathcal{R}_B^i(\Delta)$ be the closure of the set of all rate pairs (R_1, R_2) such that*

$$R_1 \leq \mathbb{I}(X, S; Y|U) - \mathbb{H}(S|U), \quad (6.4a)$$

$$R_2 \leq \mathbb{I}(U; Z) - \mathbb{I}(U; S), \quad (6.4b)$$

for some $(U, S, X, Y, Z) \in \mathcal{P}(\Delta)$, where U is an auxiliary random variable with alphabet size satisfying $|U| \leq |\mathcal{X}||\mathcal{S}| + 1$. Then $\mathcal{R}_B^i(\Delta) \subseteq \mathcal{C}_B(\Delta)$.

Proof: See Appendix D.1 .

Theorem 9 *Let $\mathcal{R}_B^o(\Delta)$ be the closure of the set of all rate pairs (R_1, R_2) such that*

$$R_1 \leq \mathbb{I}(X, S; Y|U) - \mathbb{H}(S|U), \quad (6.5a)$$

$$R_2 \leq \mathbb{I}(U, V; Z) - \mathbb{I}(U, V; S), \quad (6.5b)$$

for some $((U, V), S, X, Y, Z) \in \mathcal{P}(\Delta)$, where U and V are auxiliary random variables with alphabet sizes satisfying $|U| \leq |\mathcal{X}||\mathcal{S}| + 1$ and $|V| \leq |\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}| + 1)$, respectively. Then $\mathcal{C}_B(\Delta) \subseteq \mathcal{R}_B^o(\Delta)$.

Proof: See Appendix D.2.

Remarks

To obtain the above inner bound, the message W_2 is embedded into the host sequence S^n using Gel'fand-Pinsker coding, and the message W_1 is embedded into the host sequence using superposition coding such that the distortion constraint is satisfied. The above inner and outer bounds are already convex regions. So, there

is no need to introduce time-sharing auxiliary random variables. Let us write the constraint on R_2 in the outer bound given in (6.5) as follows

$$\mathbb{I}(\mathbf{U}, \mathbf{V}; Z) - \mathbb{I}(\mathbf{U}, \mathbf{V}; \mathbf{S}) = \mathbb{I}(\mathbf{U}; Z) - \mathbb{I}(\mathbf{U}; \mathbf{S}) + \{\mathbb{I}(\mathbf{V}; Z|\mathbf{U}) - \mathbb{I}(\mathbf{V}; \mathbf{S}|\mathbf{U})\}.$$

This term $\mathbb{I}(\mathbf{V}; Z|\mathbf{U}) - \mathbb{I}(\mathbf{V}; \mathbf{S}|\mathbf{U})$ is the difference between the inner and outer bounds. If \mathbf{V} is a deterministic function of \mathbf{U} , both inner and outer bounds coincide. This clearly shows that $\mathcal{R}_B^i(\Delta) \subseteq \mathcal{R}_B^o(\Delta)$.

6.3 State or Host Recovery at Both Decoders

This section derives the broadcast IE capacity region in Case C, in which Decoder 1 recovers $(\mathbf{W}_1, \mathbf{W}_2)$ and \mathbf{S}^n from \mathbf{Y}^n and Decoder 2 recovers \mathbf{W}_2 and \mathbf{S}^n from \mathbf{Z}^n . We define the broadcast IE capacity region $\mathcal{C}_C(\Delta)$ as the closure of all achievable rates (R_1, R_2) with $P_e^{(n)} := \Pr[(g_{1,C}^n(\mathbf{Y}^n) \neq (\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n) \text{ or } g_{2,C}^n(\mathbf{Z}^n) \neq (\mathbf{W}_2, \mathbf{S}^n)] \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 10 $\mathcal{C}_C(\Delta)$ is the closure of the set of all rate pairs (R_1, R_2) such that

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y}|\mathbf{U}, \mathbf{S}), \tag{6.6a}$$

$$R_2 \leq \mathbb{I}(\mathbf{X}, \mathbf{S}; Z) - \mathbb{H}(\mathbf{S}), \tag{6.6b}$$

for some $(\mathbf{U}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, Z) \in \mathcal{P}(\Delta)$, where \mathbf{U} is an auxiliary random variable with $|\mathbf{U}| \leq |\mathcal{X}||\mathcal{S}|$.

Proof: See Appendix D.3

Remarks

To achieve the broadcast IE capacity region, the messages $(\mathbf{W}_1, \mathbf{W}_2)$ are embedded into the host sequence using distortion-constrained superposition coding as in the previous cases because lossless recovery i.e., reversible embedding, of the host sequence \mathbf{S}^n is required in Case C.

6.4 State or Host Recovery at the Worse Decoder

This section derives the broadcast IE capacity region in Case D, in which Decoder 1 recovers (W_1, W_2) from Y^n and Decoder 2 recovers W_2 and S^n from Z^n . We define the broadcast IE capacity region $\mathcal{C}_D(\Delta)$ as the closure of all achievable rates (R_1, R_2) with $P_e^{(n)} := \Pr[(g_{1,D}^n(Y^n) \neq (W_1, W_2) \text{ or } g_{2,D}^n(Z^n) \neq (W_2, S^n)] \rightarrow 0$ as $n \rightarrow \infty$.

Corollary 3 $\mathcal{C}_D(\Delta) = \mathcal{C}_C(\Delta)$.

Proof: Since Z^n is a degraded version of Y^n , and (W_2, S^n) must be reliably decoded from Z^n , (W_2, S^n) can also be decoded from Y^n . This implies that the broadcast IE capacity region in Case D is the same as in Case C.

6.5 Summary

We considered IE in degraded broadcast channels. We derive inner and outer bounds for the case of no host recovery at both the decoders and no host recovery at the worse decoder. We derived the capacity region when the host recovery is considered at both the decoders and the host recovery is considered at the worse decoder. We also considered lossless state or host recovery at some decoders. If we consider partial state recovery of host or state, then this model can also be extended to continuous alphabet models such as Gaussian models.

CHAPTER 7

MODEL EXTENSIONS

In this chapter, we discuss some interesting extensions such as partial state recovery in single-user state-dependent models without side information and models with two-sided side information. Since these directions are very challenging, we focus more on single-user models in this chapter to keep development as simple as possible.

7.1 Partial State Recovery in State-Dependent Models

In some scenarios of communication over state-dependent channels, the communicating parties may want to estimate the channel state with some allowable distortion. In many cases, availability of the channel state at the decoders allows the use of less complex coding schemes for reliable communication. To obtain the channel state at the decoder, we usually allocate resources for training, but often, the resource overhead for estimating the channel state is not taken into account. To understand the resource overhead for estimating the channel, we study the trade-off between information transmission rate and distortion with which the channel state is estimated at the decoder. We present some preliminary results in the case of single-user models and we also define some multi-user problems that might be interesting from this point of view.

7.1.1 Single-User Models

In this section, we consider the communication system shown in Figure 7.1. In this model, the channel output $Y \in \mathcal{Y}$ is controlled by the channel state $S \in \mathcal{S}$ and the input $X \in \mathcal{X}$ through a probability law $p(\mathbf{y}|\mathbf{s}, \mathbf{x})$. The encoder sends message $W \in \mathcal{W}$ to the decoder by transmitting the codeword X^n , where $\mathcal{W} = \{1, 2, \dots, M\}$. We define the transmission rate as $R = (1/n) \log_2 M$ bits per channel use. The decoder, upon receiving the channel output Y^n , recovers W in the sense of probability of error going to zero as n goes to infinity and estimates the channel state such that the distortion between the channel state estimate and the actual channel state satisfies the distortion constraint Δ as n goes infinity.

For scenarios in which the channel state is available at the encoder, i.e., switch A in Figure 7.1 is closed, [7] and [48] study the problem in the discrete memoryless case and the Gaussian memoryless cases, respectively. We study the rate-distortion (R, Δ) region for scenarios in which the channel state is available at neither the encoder nor the decoder, i.e., switch A in Figure 7.1 is open. This model is also considered in [52].

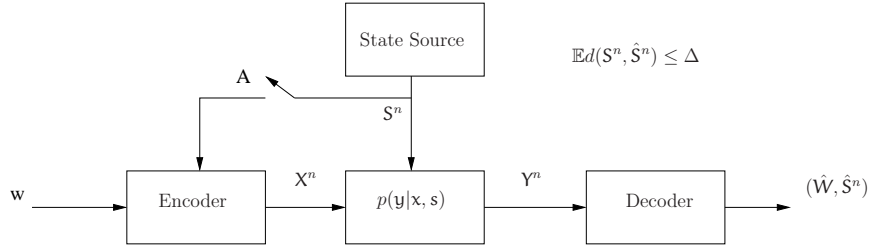


Figure 7.1. Block diagram of single-user state-dependent model.

Gaussian Case

To motivate the issues in information transmission and estimation of the channel state, let us first consider a Gaussian channel model, i.e., the channel output is given by $Y = X + S + Z$, where X is the channel input with average power constraint P , S is a zero mean Gaussian with variance Q , and Z is a zero mean Gaussian with variance N . In this model, the decoder estimates the channel state according to squared-error distortion criterion, i.e., $d(S, \hat{S}) = (S - \hat{S})^2$. This model can be viewed as Gaussian MAC in which one input is coded and one input is uncoded.

Let us discuss successive decoding in which the decoder first estimates the channel and then decodes the message with the help of the channel state estimate. In this case, the encoder transmits its message using average power γP , where $\gamma \in [0, 1]$. The decoder forms an estimate according to minimum mean squared error criterion and then decodes the message. In this case, (R, Δ) pairs are achievable if they satisfy the conditions

$$\Delta \geq Q \left[\frac{\gamma P + N}{\gamma P + Q + N} \right] \quad (7.1)$$

$$R \leq \frac{1}{2} \log \left(1 + \frac{\gamma P}{Q + N} \right), \quad (7.2)$$

where $\gamma \in [0, 1]$.

On the other hand, suppose the decoder first decodes the message and then estimates the channel state by removing the effect of the decoded codeword from the channel output. In this case, (R, Δ) pairs are achievable if they satisfy the conditions

$$\Delta \geq Q \frac{N}{Q + N} \quad (7.3)$$

$$R \leq \frac{1}{2} \log \left(1 + \frac{P}{Q + N} \right). \quad (7.4)$$

The rate-distortion pairs (R, Δ) achieved by the above successive schemes are

plotted in Figure 7.2 for $P = 1$, $Q = 1$, and $N = 1$. These results suggest that, for a given distortion constraint Δ , we can achieve better rates by first decoding the message and using it to estimate the channel than by first estimating the channel and using it to decode the message. This example suggests that, for a given distortion constraint, there is some penalty in terms of transmission rates in estimating the channel state first and then decoding the message. This observation is important, because, in many practical communication systems, the channel state is estimated first and then the estimated channel state is used to decode the message, e.g., enabling “coherent” coding techniques.

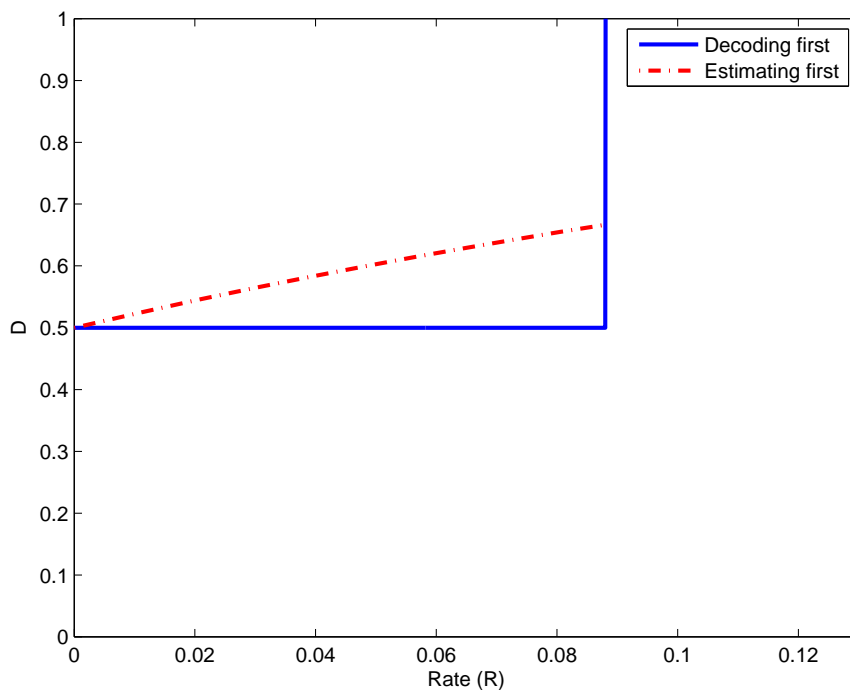


Figure 7.2. Information rate-distortion trade-off for $P = 1$, $Q = 1$ and $N = 1$

Formal Definitions

Let us now define the problem more formally and study bounds on the trade-off between information transmission rate R and state-recovery distortion constraint Δ .

Definition 18 An $(M, n, D^{(n)})$ transmission-estimation code consists of a message W uniform on $\{1, 2, \dots, M\}$, a sequence of encoding functions

$$f^n : \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n,$$

and a sequence of decoding functions

$$g^n : \mathcal{Y}^n \rightarrow (\mathcal{W}, \hat{\mathcal{S}}^n)$$

where $\hat{\mathcal{S}}$ is the reconstruction alphabet of the state. The associated message error probability is $P_e^n := \Pr[\hat{W} \neq W]$ and associated distortion is defined as $D^{(n)} = \mathbb{E}[d(\mathbf{S}^n, \hat{\mathbf{S}}^n)]$, where $d(\mathbf{S}^n, \hat{\mathbf{S}}^n) = \frac{1}{n} \sum_{i=1}^n d(\mathbf{S}_i, \hat{\mathbf{S}}_i)$ for some non-negative bounded distortion function $d(\cdot, \cdot)$.

Definition 19 A rate-distortion pair (R, Δ) is said to be achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n, D^{(n)})$ transmission-estimation codes (f^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n = 0$, and $\lim_{n \rightarrow \infty} D^{(n)} \leq \Delta$.

Definition 20 The feasible set \mathcal{D} of distortion constraints is the set of all achievable Δ for all $R \geq 0$.

Definition 21 The capacity $C(\Delta)$ for a given achievable distortion $\Delta \in \mathcal{D}$ is defined as the supremum of all achievable rates R such that (R, Δ) is achievable.

The point $(0, \Delta_{\min})$ corresponds to full training, i.e., no information transmission, where $\Delta_{\min} := \inf \mathcal{D}$. If we want to estimate the channel state to within distortion $\Delta \in \mathcal{D}$, we can find the maximum achievable rate for Δ .

Inner and Outer Bounds for $C(\Delta)$

In this section, we assume that all alphabets are discrete and develop inner and outer bounds.

Definition 22 For a given $p(\mathbf{s})$ and $p(\mathbf{y}|\mathbf{s}, \mathbf{x})$, let us define the set $\mathcal{P}_{\text{RD}}(\Delta)$ as the set of distribution functions $p(\mathbf{s}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}})$ satisfying the following conditions

1. $p(\mathbf{s}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}}) = p(\mathbf{s})p(\mathbf{x})p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\hat{\mathbf{s}}|\mathbf{y}, \mathbf{x})$,
2. $\mathbb{E}d(\mathbf{S}, \hat{\mathbf{S}}) \leq \Delta$.

Let us define \mathcal{D} as

$$\mathcal{D} := \sup\{\Delta : \mathcal{P}_{\text{RD}}(\Delta) \neq \emptyset\}$$

The following theorem gives an inner bound for $C(\Delta)$ for $\Delta \in \mathcal{D}$.

Theorem 11 *Let*

$$R_i(\Delta) := \sup_{p(\mathbf{s}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}}) \in \mathcal{P}_{\text{RD}}(\Delta)} \mathbb{I}(\mathbf{X}; \mathbf{Y})$$

for $\Delta \in \mathcal{D}$. Then,

$$R_i(\Delta) \leq C(\Delta). \tag{7.5}$$

Remarks:

- $R_i(\Delta)$ is a non-decreasing function of Δ .
- $R < R_i(\Delta)$ is achievable by jointly estimating the channel state and decoding the message.
- $R < R_i(\Delta)$ is also achievable by first decoding the message and then estimating the channel state.
- First estimating the channel state and then decoding the message,

$$R < \max_{p(\mathbf{s})p(\mathbf{x})p(\mathbf{y}|\mathbf{x}, \mathbf{s})p(\hat{\mathbf{s}}|\mathbf{y}): \mathbb{E}d(\mathbf{S}, \hat{\mathbf{S}}) \leq \Delta} \mathbb{I}(\mathbf{X}; \mathbf{Y})$$

is achievable for a given distortion constraint Δ . In this case, the maximum achievable rate is less than that achieved by joint estimation and decoding because the set of distributions is more restrictive in this case.

Proof: Fix $p(\mathbf{s}, \mathbf{x}, \mathbf{y}, \hat{\mathbf{s}}) \in \mathcal{P}_{\text{RD}}(\Delta)$ and n .

- **Codebook Generation:** Generate codewords $\mathbf{X}^n(m)$ whose elements are independently drawn according to $p(\mathbf{x})$ for each $m \in \{1, 2, \dots, \lceil 2^{nR} \rceil\}$. Generate

quantized codewords $\hat{S}^n(j)$ whose elements are independently drawn according to $p(\hat{s})$ for each $j \in \{1, 2, \dots, \lceil 2^{nR_0} \rceil\}$. These codebooks are revealed to both the encoder and the decoder.

- **Encoding & Decoding:** To send the message W , $X^n(W)$ is transmitted from the encoder. The decoder, upon receiving Y^n , looks for $(X^n(m), \hat{S}^n(j))$ that is jointly typical with Y^n . If such a pair exists and is unique, then the decoder declares that the estimate of message is m and the estimate of the channel state sequence is $\hat{S}^n(j)$. Otherwise, the decoder declares an error.
- **Probability of Error:** Without loss of generality, it can be assumed that codeword $X^n(1)$ is transmitted.
 - According to the strong asymptotic equipartition property (AEP) [12], $X^n(1)$ is jointly typical with Y^n with high probability for sufficiently large n .
 - An error occurs if there is no $\hat{S}^n(j)$ that is jointly typical with $(X^n(1), Y^n)$. The probability of this event can be made arbitrarily small if $R_0 > \mathbb{I}(\hat{S}; X, Y)$ for sufficiently large n .
 - An error also occurs if $(X^n(m), \hat{S}^n(j), Y^n)$ is jointly typical for $m \neq 1$ and for some j . The probability of this event can be made arbitrarily small if $R + R_0 < \mathbb{I}(\hat{S}; Y) + \mathbb{I}(X; \hat{S}, Y)$ for sufficiently large n .
 - An error also occurs if there no $\hat{S}^n(j)$ that is jointly typical with typical pair $(X^n(m), Y^n)$ for $m \neq 1$. The probability of this event can be made arbitrarily small if $R_0 > \mathbb{I}(\hat{S}; X, Y)$ for sufficiently large n .

Finally, the probability of error can be written as the probability of the union of the above error events. Using the union bound, it can be concluded that

the probability of error can be made arbitrarily small for sufficiently large n if $R < \mathbb{I}(X; Y)$ and $R_0 > \mathbb{I}(\hat{S}; X, Y)$.

- **Average distortion:** With high probability, $(X^n(1), \hat{S}^n(j), Y^n)$ is jointly typical for some j . According to strong AEP, $(X^n(1), S^n, Y^n)$ is jointly typical with high probability. According to Markov lemma [12], S^n and $\hat{S}^n(j)$ are jointly typical with high probability. It can be concluded that the distortion between S^n and $\hat{S}^n(j)$ satisfies a distortion constraint Δ with high probability.

The following theorem gives an outer bound for $C(\Delta)$.

Theorem 12 *Let*

$$R_o(\Delta) := \max_{p(s)p(x)p(y|x,s)p(\hat{s}|y,x,s): \mathbb{E}d(S,\hat{S}) \leq \Delta} \mathbb{I}(X; Y).$$

Then, $R_o(\Delta) \geq C(\Delta)$.

The above outer bound $R_o(\Delta)$ is different from the inner bound $R_i(\Delta)$ because the distributions are more restrictive in the inner bound.

7.1.2 Multi-User Models

In this section, we extend the framework of transmission rate and distortion trade-off to multi-user models. We define some multi-user problems that might be interesting from this point of view.

Multiple Access Channel

Consider a multiple access channel (MAC) as shown in Figure 7.3 that generates the output $Y \in \mathcal{Y}$ which is controlled by the state pair $(S_1, S_2) \in \mathcal{S}_1 \times \mathcal{S}_2$ and the input pair $(X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ from two encoders. Here, we consider a model with two encoders, but, in principle, the definitions and results can be extended to any number of encoders. We assume that the channel state is known at neither the encoders

nor the decoder. As shown in Figure 4.1, Encoder i sends message $W_i \in \mathcal{W}_i$ to the decoder by transmitting the codeword X_i^n , where $\mathcal{W}_i = \{1, 2, \dots, M_i\}$ for $i = 1, 2$. As usual, we define the transmission rate as $R_i = (1/n) \log_2 M_i$ bits per channel use for $i = 1, 2$. The decoder, upon receiving the channel output Y^n , recovers the message pair (W_1, W_2) in the sense of probability of error going to zero as n goes to infinity and estimates the channel state pair such that the distortion between the channel state estimate \hat{S}_i^n and the actual channel state S_i^n satisfies the distortion constraint Δ_i for $i = 1, 2$ as n goes infinity.

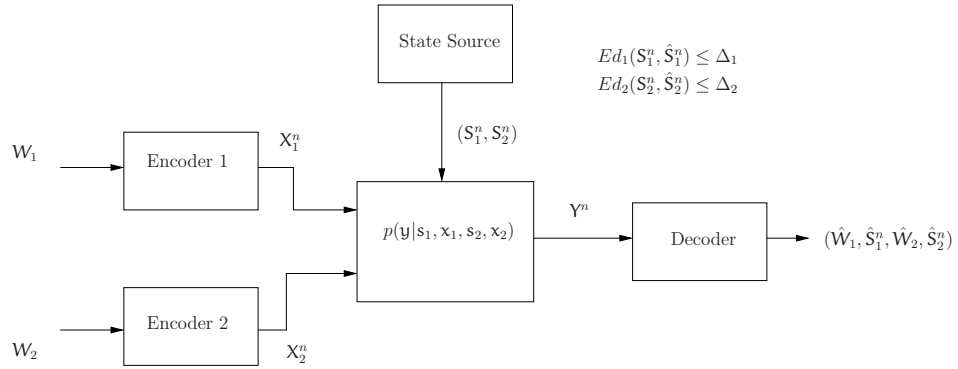


Figure 7.3. Block diagram of state-dependent multi-access model.

Definition 23 An $(M_1, M_2, n, D_1^{(n)}, D_2^{(n)})$ transmission-estimation code for the multiple access channel consists of messages W_i uniform on $\{1, 2, \dots, M_i\}$, and sequence of encoding functions

$$f_i^n : \{1, 2, \dots, M_i\} \rightarrow \mathcal{X}_i^n$$

for $i = 1, 2$; and a sequence of decoding functions at the decoder

$$g^n : \mathcal{Y}^n \rightarrow (\mathcal{W}_1, \hat{\mathcal{S}}_1^n, \mathcal{W}_2, \hat{\mathcal{S}}_2^n)$$

where $(\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2)$ is the reconstruction alphabet pair of the states. The associated probability of error is $P_e^n := \Pr[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)]$ and associated distortion is defined as $D_i^{(n)} = E[d_i(S_i^n, \hat{S}_i^n)]$, where $d_i(S_i^n, \hat{S}_i^n) = \frac{1}{n} \sum_{j=1}^n d(S_{i,j}, \hat{S}_{i,j})$ for some non-negative bounded distortion function $d_i(\cdot, \cdot)$ for $i = 1, 2$.

Definition 24 A rate-distortion tuple $(R_1, R_2, \Delta_1, \Delta_2)$ is said to be achievable for the state-dependent MAC if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n, D_1^{(n)}, D_2^{(n)})$

transmission-estimation codes (f_1^n, f_2^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n = 0$ and $\lim_{n \rightarrow \infty} D_i^{(n)} \leq \Delta_i$ for $i = 1, 2$.

Broadcast Channel

Consider a broadcast channel as shown in Figure 7.4, which is controlled by the state $S \in \mathcal{S}$ and the input $X \in \mathcal{X}$ from the encoder and generates the output pair $(Y_1, Y_2) \in (\mathcal{Y}_1, \mathcal{Y}_2)$. Here, we consider a model with two decoders, but, in principle, the definitions and results can be extended to any number of decoders. We assume that the channel state is known at neither the encoder nor the decoders. The encoder sends messages $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$ to the decoders by transmitting the codeword X^n , where $\mathcal{W}_i = \{1, 2, \dots, M_i\}$, for $i = 1, 2$. We define the transmission rate as $R_i = (1/n) \log_2 M_i$ bits per channel use. Decoder i , upon receiving the channel output Y_i^n , recovers W_i in the sense of probability of error going to zero as n goes to infinity and estimates the channel state such that the distortion between the channel state estimate and the actual channel state satisfies the distortion constraint Δ_i as n goes infinity for $i = 1, 2$.

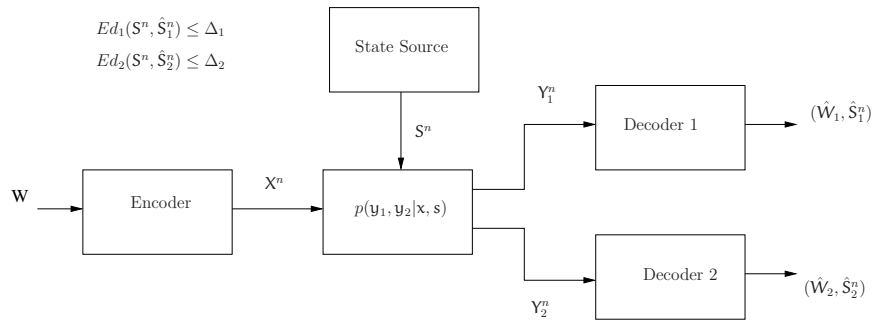


Figure 7.4. Block diagram of a state-dependent broadcast channel.

Definition 25 An $(M_1, M_2, n, D_1^{(n)}, D_2^{(n)})$ transmission-estimation code for the broadcast model consists of messages W_i uniform on $\{1, 2, \dots, M_i\}$ for $i = 1, 2$ and a

sequence of encoding functions

$$f^n : \{1, 2, \dots, M_1\} \times \{1, 2, \dots, M_2\} \rightarrow \mathcal{X}^n,$$

and sequences of decoding functions

$$g_1^n : \mathcal{Y}_1^n \rightarrow (\mathcal{W}_1, \hat{\mathcal{S}}_1^n) \quad \text{and} \quad g_2^n : \mathcal{Y}_2^n \rightarrow (\mathcal{W}_2, \hat{\mathcal{S}}_2^n),$$

at the Decoder 1 and Decoder 2, respectively, where $\hat{\mathcal{S}}_1$ and $\hat{\mathcal{S}}_2$ are the reconstruction alphabets of the state at the Decoder 1 and Decoder 2, respectively. The associated probability of error is $P_e^n := \Pr[\hat{\mathbf{W}}_1 \neq \mathbf{W}_1 \text{ or } \hat{\mathbf{W}}_2 \neq \mathbf{W}_2]$ and associated distortion is defined as $D_i^{(n)} = E[d_i(\mathbf{S}^n, \hat{\mathbf{S}}_i^n)]$, where $d_i(\mathbf{S}^n, \hat{\mathbf{S}}_i^n) = \frac{1}{n} \sum_{j=1}^n d_i(S_j, \hat{S}_{i,j})$ for some non-negative bounded distortion function $d_i(\cdot, \cdot)$ for $i = 1, 2$.

Definition 26 A rate-distortion tuple $(R_1, R_2, \Delta_1, \Delta_2)$ is said to be achievable if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n, D_1^{(n)}, D_2^{(n)})$ transmission-estimation codes for the state-dependent broadcast channel (f^n, g_1^n, g_2^n) with $\lim_{n \rightarrow \infty} P_e^n = 0$ and $\lim_{n \rightarrow \infty} D_i^{(n)} \leq \Delta_i$ for $i = 1, 2$.

7.2 State-Dependent Models with Noisy State Information

In the previous chapters, we considered several models with side information that is identical to the channel state. Let us now consider a state-dependent model with general encoder side information and decoder side information as shown in Figure 7.5. In this model, the channel output $\mathbf{Y} \in \mathcal{Y}$ is controlled by the channel state $\mathbf{S} \in \mathcal{S}$ and the input $\mathbf{X} \in \mathcal{X}$ through a memoryless probability law $p(\mathbf{y}|\mathbf{s}, \mathbf{x})$. In Figure 7.5, the encoder side information \mathbf{T}_1^n and decoder side information \mathbf{T}_2^n are related to the channel state \mathbf{S}^n through a memoryless probability law $p(\mathbf{s}, \mathbf{t}_1, \mathbf{t}_2)$. The encoder sends a message $\mathbf{W} \in \mathcal{W}$ to the decoder by transmitting the codeword \mathbf{X}^n , where $\mathcal{W} = \{1, 2, \dots, M\}$. If \mathbf{T}_2 is zero with probability one, this model is considered in [15].

Definition 27 An (M, n) code consists of message \mathbf{W} uniform on $\{1, 2, \dots, M\}$, a sequence of encoding functions

$$f^n : \{1, 2, \dots, M\} \times \mathcal{T}_1^n \rightarrow \mathcal{X}^n,$$

and a sequence of decoding functions

$$g^n : \mathcal{Y}^n \times \mathcal{T}_2^n \rightarrow \mathcal{W}.$$

The associated probability of error is defined as $P_e^n := \Pr[\hat{\mathbf{W}} \neq \mathbf{W}]$.

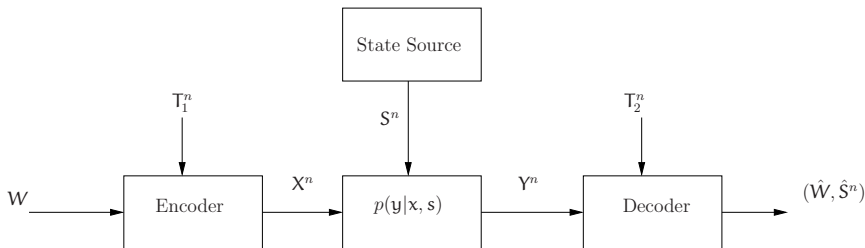


Figure 7.5. Block diagram of a single-user, state-dependent model with two-sided side information.

Definition 28 A rate R is said to be achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes (f^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n = 0$.

Definition 29 The capacity C is the supremum of all achievable rates.

The following theorem presents the capacity in the case of discrete alphabets.

Theorem 13

$$C = \max_{p(\mathbf{u}, \mathbf{x} | \mathbf{t}_1)} [\mathbb{I}(\mathbf{U}; \mathbf{Y}, \mathbf{T}_2) - \mathbb{I}(\mathbf{U}; \mathbf{T}_1)]$$

In the above expression, the decoder side information \mathbf{T}_2 is viewed as another channel output along with the usual channel output \mathbf{Y} .

Achievability:

The following coding scheme to achieve the above capacity closely follows [17] and [15].

- For each message $m \in \mathcal{W}$, generate a bin of $\lceil 2^{nR_0} \rceil$ codewords $\mathbf{U}^n(m, j)$ whose elements are generated according to distribution $p(\mathbf{u})$, where $j \in \{1, 2, \dots, \lceil 2^{nR_0} \rceil\}$. This codebook is revealed to both the encoder and the decoder. In this case, there is more than one codeword for each message. The available noisy channel state \mathbf{T}_1^n determines which codeword should be chosen for a given message.

- Given the message $W \in \mathcal{W}$ and the channel state T_1^n , the encoder finds the codeword $\mathbf{U}^n(W, j)$ in bin W that is jointly typical with T_1^n . If no such j exists, then the encoder declares an error. Otherwise, the encoder generates the channel input \mathbf{X}^n whose elements are independently and identically distributed according to $p(\mathbf{x}|\mathbf{u}, \mathbf{t}_1)$. The probability of encoding error goes to zero as $n \rightarrow \infty$ provided $R_0 > \mathbb{I}(\mathbf{U}; \mathsf{T}_1)$.
- The decoder finds $\mathbf{U}^n(m, j)$ that is jointly typical with the channel output $(\mathbf{Y}^n, \mathsf{T}_2^n)$, where $m \in \mathcal{W}$ and $j \in \{1, 2, \dots, \lceil 2^{nR_0} \rceil\}$. If such a sequence \mathbf{U}^n exists and is unique, the decoder declares its bin index as the estimate for the transmitted message; otherwise, the decoder declares an error. The probability of decoding error goes to zero as $n \rightarrow \infty$ provided $R + R_0 < \mathbb{I}(\mathbf{U}; \mathbf{Y}, \mathsf{T}_2)$. Thus, the overall probability of error goes to zero as $n \rightarrow \infty$ as long as $R < \mathbb{I}(\mathbf{U}; \mathbf{Y}, \mathsf{T}_2) - \mathbb{I}(\mathbf{U}; \mathsf{T}_1)$.

Converse : The converse proof closely follows Gel'fand-Pinsker converse [17]. In this case, for a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with probability of error P_e^n going to zero as n goes to infinity, we have

$$R \leq \max_{p(\mathbf{u}, \mathbf{x}|\mathbf{t}_1)p(\mathbf{y}|\mathbf{s}, \mathbf{x})} [\mathbb{I}(\mathbf{U}; \mathbf{Y}, \mathsf{T}_2) - \mathbb{I}(\mathbf{U}; \mathsf{T}_1)].$$

From the given code, the distribution on $(W, \mathbf{S}^n, \mathsf{T}_1^n, \mathsf{T}_2^n, \mathbf{X}^n, \mathbf{Y}^n)$ is as follows

$$p(w, \mathbf{s}^n, \mathbf{t}_1^n, \mathbf{t}_2^n, \mathbf{x}^n, \mathbf{y}^n) = p(w)p(\mathbf{s}^n, \mathbf{t}_1^n, \mathbf{t}_2^n)p(\mathbf{x}^n|w, \mathbf{t}_1^n)p(\mathbf{y}^n|\mathbf{s}^n, \mathbf{x}^n),$$

where $p(\mathbf{x}^n|w, \mathbf{t}_1^n)$ is 1 if $\mathbf{x}^n = f^n(w, \mathbf{t}_1^n)$ and 0 otherwise.

We can bound the rate R as follows.

$$\begin{aligned} nR &\leq \mathbb{H}(W) \\ &= \mathbb{I}(W; \mathbf{Y}^n, \mathsf{T}_{2,1}^n) + \mathbb{H}(W|\mathbf{Y}^n, \mathsf{T}_{2,1}^n) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}; \mathbf{Y}^n, \mathbb{T}_{2,1}^n) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n [\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i+1}^n; \mathbf{Y}^i, \mathbb{T}_{2,1}^i) - \mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i}^n; \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1})] + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n [\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i+1}^n; \mathbf{Y}_i, \mathbb{T}_{2,i} | \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}) \\
&\quad - \mathbb{I}(\mathbb{T}_{1,i}; \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1} | \mathbf{W}, \mathbb{T}_{1,i+1}^n)] + n\epsilon_n \\
&\stackrel{(d)}{\leq} \sum_{i=1}^n [\mathbb{H}(\mathbf{Y}_i) - \mathbb{H}(\mathbf{Y}_i | \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}, \mathbf{W}, \mathbb{T}_{1,i+1}^n) \\
&\quad + \mathbb{H}(\mathbb{T}_{1,i} | \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}, \mathbf{W}, \mathbb{T}_{1,i+1}^n) - \mathbb{H}(\mathbb{T}_{1,i})] + n\epsilon_n \\
&= \sum_{i=1}^n [\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i+1}^n, \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}; \mathbf{Y}_i, \mathbb{T}_{2,i}) \\
&\quad - \mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i+1}^n, \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}; \mathbb{T}_{1,i})] + n\epsilon_n \\
&\stackrel{(e)}{\leq} n[\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,k+1}^n, \mathbf{Y}^{k-1}, \mathbb{T}_{2,1}^{k-1}; \mathbf{Y}_k, \mathbb{T}_{2,k}) \\
&\quad - \mathbb{I}(\mathbf{W}, \mathbb{T}_{1,k+1}^n, \mathbf{Y}^{k-1}, \mathbb{T}_{2,1}^{k-1}; \mathbb{T}_{1,k})] + n\epsilon_n \tag{7.6}
\end{aligned}$$

where:

(a) follows from Fano's inequality with $\epsilon_n \rightarrow 0$ as $P_e^n \rightarrow 0$,

(b) follows from $\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i+1}^n; \mathbf{Y}^i, \mathbb{T}_{2,i}^n) = \mathbb{I}(\mathbf{W}; \mathbf{Y}^n, \mathbb{T}_{2,1}^n)$ for $i = n$;

$\mathbb{I}(\mathbf{W}, \mathbb{T}_{1,i}^n; \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}) = 0$ for $i = 1$; and the sum of the remaining terms equals to zero,

(c) follows from applying the chain rule for mutual information to

$\{(\mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1}), (\mathbf{Y}_i, \mathbb{T}_{2,i})\}$ in the first term and to $(\{\mathbf{W}, \mathbb{T}_{1,i+1}^n\}, \mathbb{T}_{1,i})$ in the second term,

(d) follows from $\mathbb{H}(\mathbf{Y}_i, \mathbb{T}_{2,i}) \geq \mathbb{H}(\mathbf{Y}_i, \mathbb{T}_{2,i} | \mathbf{Y}^{i-1}, \mathbb{T}_{2,1}^{i-1})$ and $\mathbb{T}_{1,i}$ being independent of $(\mathbb{T}_{1,i+1}^n, \mathbf{W})$,

(e) follows from the fact that

$$[\mathbb{I}(\mathbf{W}, \mathbf{T}_{1,i+1}^n, \mathbf{Y}^{i-1}, \mathbf{T}_{2,1}^{i-1}; \mathbf{Y}_i, \mathbf{T}_{2,i}) - \mathbb{I}(\mathbf{W}, \mathbf{T}_{1,i+1}^n, \mathbf{Y}^{i-1}, \mathbf{T}_{2,1}^{i-1}; \mathbf{T}_{1,i})]$$

is maximized over all i for the value of k .

Let us define $\mathbf{U} := (\mathbf{W}, \mathbf{T}_{1,k+1}^n, \mathbf{Y}^{k-1}, \mathbf{T}_{2,1}^{k-1})$ in (7.6). As $n \rightarrow \infty$, we obtain

$$R \leq \max_{p(\mathbf{u}, \mathbf{x} | \mathbf{t}_1)} [\mathbb{I}(\mathbf{U}; \mathbf{Y}, \mathbf{T}_2) - \mathbb{I}(\mathbf{U}; \mathbf{T}_1)].$$

7.3 Summary

In this chapter, we have studied two extensions of our modeling framework from earlier chapters. We focused on the single-user state-dependent models to keep the development as simple as possible. First, we studied bounds on the capacity region of single-user model without encoder side information and with lossy state recovery at the decoder. Since the inner and outer bounds on the capacity-distortion region do not meet, we still do not know the capacity-distortion region. In this chapter, we consider memoryless state but in practice we often see state-dependent models in which the channel state has memory. It would be therefore useful to study state with memory for the models considered. Second, we studied bounds on the capacity region of single-user, state-dependent model with noisy encoder state information and noisy decoder state information. For this model, we obtain the capacity.

In many scenarios in information theory, there is some duality between source coding problems and channel coding problems. In this thesis, we did not focus on source coding duals of the models we considered. But, study of such duals are necessary for understanding source coding with side information, and could lead to additional insights for the corresponding channel coding problems.

CHAPTER 8

CONCLUSIONS

In the final chapter, we conclude the thesis, highlighting contributions and future directions.

8.1 Contributions

In this thesis, we consider various state-dependent models in both single-user and multi-user cases. We consider multi-user models with encoder side information such as multiple access channels (MAC), and broadcast channels (BC).

For the state-dependent MAC with side information at some encoders, we discussed bounds on the capacity region. Since the inner and outer bounds do not meet, the capacity region is still not known for this problem. We discussed how the informed encoders exploit the known channel state to help uninformed encoders in terms of their information rates. However, we obtained some insights about the random coding techniques based on dirty paper coding and state cancellation at the informed encoder.

For the state-dependent MAC with different state information available at different encoders, we consider lossless recovery of some state signals at the decoder. We derived inner bounds for the case of no state recovery and the case of some, but not all, state recovery. If all state signals are recovered at the decoder, we derived inner and outer bounds for the capacity region. It turns out that the inner

and outer bounds meet if all state signals are independent, so that we obtain the capacity region for this case.

For the state-dependent BC with side information known at the encoder, we considered lossless recovery of the state at some decoders. If state recovery is not considered at any decoder, we obtained inner and outer bounds for the capacity region, but the capacity region is still not known. If the state recovery is considered at all decoders, we obtain the capacity region. If asymmetry in terms of lossless state recovery at the decoders is present in the model, i.e., lossless state recovery is considered at only the better decoder, the capacity region is still not known.

We also discussed some interesting model extensions, i.e, lossy state recovery at the decoder in single-user state-dependent models without side information and models with two-sided side information. In these extensions, we focused more on single-user models to keep the development as simple as possible.

Even though we study various state-dependent models in some interesting scenarios very well, results and observations suggest that further study of multi-user state-dependent models is required to understand them completely. Based on our study of state-dependent multi-user models, we suggest some future directions in the next section.

8.2 Future Directions

We observed that asymmetry in terms of the channel state availability at the encoders is very interesting in MAC in terms of the capacity region. We also observed that asymmetry in terms of lossless channel state recovery at the decoders is very interesting in BC in terms of the capacity region. These observations suggest the asymmetry plays a key role in making these problems interesting and challenging. In the retrospect, the original Gel'fand-Pinsker model exhibits such asymmetry, i.e.,

state available at the encoder but not the decoder. Further study of multi-user state-dependent models with asymmetry such as interference channels, etc. would be interesting and challenging.

As already mentioned in this thesis, the important issues in these state-dependent networks are encoding with the help of available side information and estimating the channel state at the decoders. In this thesis, we assume that the exact channel state is available at the encoders for the most of the models. This assumption is valid in some applications such as information embedding and data hiding. But, in some scenarios, partial channel state or rate-constrained channel state is available at the encoders. It is interesting to consider state-dependent models under these assumptions to have further understanding of these models.

In the case of channel estimation at the decoders, we considered lossless recovery of state at the decoders for models with encoder side information. But, lossless recovery can not be required in models with continuous state alphabets and is not possible in some channels. As we have seen that lossy recovery in single-user models itself is very challenging, it would be even more challenging to study lossy recovery in multi-user state-dependent models. At the end, we conclude that further study of state-dependent models is required to understand them.

APPENDIX A

A.1 Notation

In this section, we present notation that is used throughout the thesis. Random variables and sample values are denoted in a special font, e.g., the random variable \mathbf{X} and sample value \mathbf{x} . Alphabets are denoted in calligraphic font, e.g., \mathcal{X} . The shorthand \mathbf{X}_1^n represents the sequence $\mathbf{X}_{1,1}, \mathbf{X}_{1,2}, \dots, \mathbf{X}_{1,n}$, and $\mathbf{X}_{1,i}^n$ represents the sequence $\mathbf{X}_{1,i}, \mathbf{X}_{1,i+1}, \dots, \mathbf{X}_{1,n}$. Addition in $\mathbf{X}^n + \mathbf{Y}^n$ is element wise addition. We denote an identity matrix of size $n \times n$ as \mathbf{I}_n , and zero mean Gaussian distribution with variance N as $\mathcal{N}(0, N)$. All logarithms in this thesis are with base 2. Finally, $\mathbb{H}(\cdot)$ and $\mathbb{I}(\cdot; \cdot)$ denote the standard information theoretic quantities of (ensemble average) entropy and mutual information which will be defined in the next section.

A.2 Entropy and Mutual Information

In this section, we define entropy for a random variable and mutual information between two random variables.

Definition 30 *The entropy $\mathbb{H}(\mathbf{X})$ of discrete random variable \mathbf{X} is defined as*

$$\mathbb{H}(\mathbf{X}) = - \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \log p(\mathbf{x}).$$

If $(\mathbf{X}, \mathbf{Y}) \sim p(\mathbf{x}, \mathbf{y})$, the conditional entropy $\mathbb{H}(\mathbf{Y}|\mathbf{X})$ is defined as

$$\mathbb{H}(\mathbf{Y}|\mathbf{X}) = \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \mathbb{H}(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = - \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \sum_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}) \log p(\mathbf{y}|\mathbf{x}).$$

Definition 31 The entropy $\mathbb{H}(\mathbf{X})$ of continuous random variable \mathbf{X} with density function $p(\mathbf{x})$ is defined as

$$\mathbb{H}(\mathbf{X}) = - \int_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}.$$

If \mathbf{X} and \mathbf{Y} are continuous random variables and $(\mathbf{X}, \mathbf{Y}) \sim p(\mathbf{x}, \mathbf{y})$, the conditional entropy $\mathbb{H}(\mathbf{Y}|\mathbf{X})$ is defined as

$$\mathbb{H}(\mathbf{Y}|\mathbf{X}) = \int_{\mathbf{x}} p(\mathbf{x}) \mathbb{H}(\mathbf{Y}|\mathbf{X} = \mathbf{x}) d\mathbf{x} = - \int_{\mathbf{x}} p(\mathbf{x}) d\mathbf{x} \int_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \log p(\mathbf{y}|\mathbf{x}) d\mathbf{y}.$$

Definition 32 If \mathbf{X} and \mathbf{Y} are random variables and $(\mathbf{X}, \mathbf{Y}) \sim p(\mathbf{x}, \mathbf{y})$, then the mutual information $\mathbb{I}(\mathbf{X}; \mathbf{Y})$ is defined as

$$\mathbb{I}(\mathbf{X}; \mathbf{Y}) = \mathbb{H}(\mathbf{X}) - \mathbb{H}(\mathbf{X}|\mathbf{Y}) = \mathbb{H}(\mathbf{Y}) - \mathbb{H}(\mathbf{Y}|\mathbf{X}).$$

A.3 Definitions for Chapter 2 and Chapter 3.

In this section, we define the capacity for single-user models and the capacity region for multi-user models in various scenarios. These definitions are used in the background material presented in Chapter 2 and Chapter 3.

A.3.1 Single-User Models

Definition 33 A rate R is said to be achievable for single-user model without state recovery if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes (f^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n \rightarrow 0$, where $P_e^n = \Pr[\mathbf{W} \neq \hat{\mathbf{W}}]$. The capacity C is given as the supremum of the set of achievable rates.

Definition 34 A rate R is said to be achievable for single-user state-dependent model with lossless state recovery if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes (f^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n \rightarrow 0$, where $P_e^n = \Pr[g(\mathbf{Y}^n) \neq (\mathbf{W}, \mathbf{S}^n)]$. The capacity C is given as the supremum of the set of achievable rates.

Definition 35 A rate R is said to be achievable for single-user models with lossy state recovery with a given distortion constraint Δ if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes (f^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n \rightarrow 0$ and $\lim_{n \rightarrow \infty} \mathbb{E}d(\mathbf{S}^n, \hat{\mathbf{S}}^n) \leq \Delta$, where $P_e^n = \Pr[\mathbf{W} \neq \hat{\mathbf{W}}]$, and $g(\mathbf{Y}^n) = (\hat{\mathbf{W}}, \hat{\mathbf{S}}^n)$. The capacity $C(\Delta)$ is given as the supremum of the set of achievable rates.

A.3.2 Multiple Access Channels

Definition 36 A rate pair (R_1, R_2) is said to be achievable for state-dependent multiple access channel without state recovery if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes (f_1^n, f_2^n, g^n) with $\lim_{n \rightarrow \infty} P_e^n \rightarrow 0$, where $P_e^n = \Pr[g(Y^n) \neq (W_1, W_2)]$. The capacity region \mathcal{C} is given as the closure of the set of achievable rate pair (R_1, R_2) .

A.3.3 Broadcast Channels

Definition 37 A rate pair (R_1, R_2) is said to be achievable for state-dependent broadcast channel without state recovery if there exists a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes (f^n, g_1^n, g_2^n) with $\lim_{n \rightarrow \infty} P_e^n \rightarrow 0$, where $P_e^n = \Pr[(W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)]$, $\hat{W}_1 = g_1^n(Y^n)$, and $\hat{W}_2 = g_2^n(Z^n)$. The capacity region \mathcal{C} is given as the closure of the set of achievable rate pair (R_1, R_2) .

A.4 Strong Typicality

In this section, we present definitions on strong typicality [12, 14, 51] and state theorems based on strong typicality which will be used to prove theorems in this thesis.

Definition 38 A sequence $\mathbf{x}^n \in \mathcal{X}^n$ is said to be ϵ -strongly typical with respect to a distribution $p(\mathbf{x})$ on \mathcal{X} or $\mathbf{x}^n \in T_\epsilon^n(\mathbf{X})$ if

$$\left| \frac{1}{n} N(\mathbf{a} | \mathbf{x}^n) - p(\mathbf{a}) \right| < \frac{\epsilon}{|\mathcal{X}|},$$

for all $\mathbf{a} \in \mathcal{X}$ with $p(\mathbf{a}) > 0$, and $N(\mathbf{a} | \mathbf{x}^n) = 0$ for all $\mathbf{a} \in \mathcal{X}$ with $p(\mathbf{a}) = 0$, where $N(\mathbf{a} | \mathbf{x}^n)$ is the number of occurrences of the symbol \mathbf{a} in the sequence \mathbf{X}^n .

Definition 39 A pair of sequences $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be ϵ -strongly typical with respect to a distribution $p(\mathbf{x}, \mathbf{y})$ on $\mathcal{X} \times \mathcal{Y}$ or $(\mathbf{x}^n, \mathbf{y}^n) \in T_\epsilon^n(\mathbf{x}, \mathbf{y})$ if

$$\left| \frac{1}{n} N(\mathbf{a}, \mathbf{b} | \mathbf{x}^n, \mathbf{y}^n) - p(\mathbf{a}, \mathbf{b}) \right| < \frac{\epsilon}{|\mathcal{X}| |\mathcal{Y}|},$$

for all $(\mathbf{a}, \mathbf{b}) \in \mathcal{X} \times \mathcal{Y}$ with $p(\mathbf{a}, \mathbf{b}) > 0$, and $N(\mathbf{a}, \mathbf{b} | \mathbf{x}^n, \mathbf{y}^n) = 0$ for all $(\mathbf{a}, \mathbf{b}) \in \mathcal{X} \times \mathcal{Y}$ with $p(\mathbf{a}, \mathbf{b}) = 0$, where $N(\mathbf{a}, \mathbf{b} | \mathbf{x}^n, \mathbf{y}^n)$ is the number of occurrences of the symbol (\mathbf{a}, \mathbf{b}) in the pair of sequences $(\mathbf{x}^n, \mathbf{y}^n)$.

We state the following theorems on strong typicality without proofs. For proofs of the following theorems using the strong typicality definitions, look at [12, 14, 51].

Lemma 1 Suppose X^n is generated from a discrete memoryless source (DMS) $p(x)$ and $X^n \in T_\epsilon^n(\mathbf{X})$. Then, we have the following

$$2^{-n[\mathbb{H}(X)+\epsilon_1]} < P^n(\mathbf{x}^n) < 2^{-n[\mathbb{H}(X)-\epsilon_1]} \quad (\text{A.1})$$

$$(1 - \epsilon_2) 2^{n[\mathbb{H}(X)-\epsilon_1]} < |T_\epsilon^n(\mathbf{X})| < 2^{n[\mathbb{H}(X)+\epsilon_1]} \quad (\text{A.2})$$

$$(1 - \epsilon_2) \leq \Pr[\mathbf{X}^n \in T_\epsilon^n(\mathbf{X})] \leq 1 \quad (\text{A.3})$$

where $\epsilon_1 \rightarrow 0$ as $\epsilon \rightarrow 0$, and $\epsilon_2 \rightarrow 0$ as $n \rightarrow \infty$ for fixed ϵ .

Lemma 2 Suppose (X^n, Y^n) is generated from a discrete memoryless source (DMS) $p(x, y)$ and $(\mathbf{x}^n, \mathbf{y}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y})$ and Then, we have the following

$$2^{-n[\mathbb{H}(X, Y)+\epsilon'_1]} < P^n(\mathbf{x}^n, \mathbf{y}^n) < 2^{-n[\mathbb{H}(X, Y)-\epsilon'_1]} \quad (\text{A.4})$$

$$(1 - \epsilon'_2) 2^{n[\mathbb{H}(X, Y)-\epsilon'_1]} < |T_\epsilon^n(\mathbf{X}, \mathbf{Y})| < 2^{n[\mathbb{H}(X, Y)+\epsilon'_1]} \quad (\text{A.5})$$

$$(1 - \epsilon'_2) \leq \Pr[(\mathbf{X}^n, \mathbf{Y}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y})] \leq 1 \quad (\text{A.6})$$

where $\epsilon'_1 \rightarrow 0$ as $\epsilon \rightarrow 0$, and $\epsilon'_2 \rightarrow 0$ as $n \rightarrow \infty$ for fixed ϵ .

Lemma 3 Suppose (X^n, Y^n) is generated from a discrete memoryless source (DMS) $p(x, y)$ and $(\mathbf{X}^n, \mathbf{Y}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y})$. Then, we have the following

$$2^{-n[\mathbb{H}(Y|X)+\epsilon''_1]} < P^n(\mathbf{y}^n | \mathbf{x}^n) < 2^{-n[\mathbb{H}(Y|X)-\epsilon''_1]} \quad (\text{A.7})$$

$$(1 - \epsilon''_2) 2^{n[\mathbb{H}(Y|X)-\epsilon''_1]} < |T_\epsilon^n(\mathbf{X}, \mathbf{Y} | \mathbf{x}^n)| < 2^{n[\mathbb{H}(Y|X)+\epsilon''_1]} \quad (\text{A.8})$$

$$(1 - \epsilon''_2) \leq \Pr[(\mathbf{x}^n, \mathbf{Y}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y})] \leq 1 \quad (\text{A.9})$$

where $\epsilon''_1 \rightarrow 0$ as $\epsilon \rightarrow 0$, and $\epsilon''_2 \rightarrow 0$ as $n \rightarrow \infty$ for fixed ϵ , and $T_\epsilon^n(\mathbf{X}, \mathbf{Y} | \mathbf{x}^n) = \{\mathbf{y}^n : (\mathbf{x}^n, \mathbf{y}^n) \in T_\epsilon^n(\mathbf{X}, \mathbf{Y})\}$.

APPENDIX B

B.1 Proof of Theorem 1

In this section, we construct a sequence of codes $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ with $P_e^n \rightarrow 0$ as $n \rightarrow \infty$ if (R_1, R_2) satisfies Equation (4.2). The proof closely follows the proofs in [17, 10]. Fix $\epsilon > 0$ and take $(\mathbf{Q}, \mathbf{S}, \mathbf{U}_1, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}^i$, where \mathcal{P}^i is defined in Definition 4.

- **Encoding:** The encoding strategy at the two encoders is as follows. Let $M_1 = 2^{n(R_1 - 4\epsilon)}$, $M_2 = 2^{n(R_2 - 2\epsilon)}$, and $J = 2^{n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + 2\epsilon)}$. At the informed encoder, where the CSI is available, generate JM_1 sequences $\mathbf{U}_1^n(\mathbf{q}^n, m_1, j)$, whose elements are drawn i.i.d. with $p(\mathbf{u}_1 | \mathbf{q})$, for each time sharing random sequence \mathbf{Q}^n , where $1 \leq m_1 \leq M_1$, and $1 \leq j \leq J$. Here, m_1 indexes bins and j indexes sequences within a particular bin m_1 . For encoding, given CSI $\mathbf{S}^n = \mathbf{s}^n$, time sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$ and message $W_1 \in \{1, 2, \dots, M_1\}$, look in bin W_1 for a sequence $\mathbf{U}_1^n(\mathbf{q}^n, W_1, j), 1 \leq j \leq J$, such that $\mathbf{U}_1^n(\mathbf{q}^n, W_1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{S} | \mathbf{q}^n, \mathbf{s}^n]$. Then, the informed encoder chooses \mathbf{X}_1^n , whose elements are given by deterministic function $X_{1i} = f(s_i, \mathbf{q}_i, \mathbf{U}_{1i})$ for $i = 1, 2, \dots, n$.

At the uninformed encoder, sequences $\mathbf{X}_2^n(\mathbf{q}^n, m_2)$, whose elements are drawn i.i.d. with $p(x_2 | \mathbf{q})$, are generated for each time sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$, where $1 \leq m_2 \leq M_2$. The uninformed encoder chooses $\mathbf{X}_2^n(\mathbf{q}^n, W_2)$ to send the message $W_2 \in \{1, 2, \dots, M_2\}$ for a given time-sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$

and sends the codeword \mathbf{X}_2^n .

Given the inputs and the CSI, the decoder receives \mathbf{Y}^n according to conditional probability distribution $\prod_i p(\mathbf{y}_i | \mathbf{s}_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$. It is assumed that the time-sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$ is non-causally known at both the encoders and the decoder.

- **Decoding:** The decoder, upon receiving the sequence \mathbf{Y}^n , chooses a pair $(\mathbf{U}_1^n(\mathbf{q}^n, m_1, j), \mathbf{X}_2^n(m_2))$, $1 \leq m_1 \leq M_1$, $1 \leq j \leq J$, and $1 \leq m_2 \leq M_2$ such that $(\mathbf{U}_1^n(\mathbf{q}^n, m_1, j), \mathbf{X}_2^n(\mathbf{q}^n, m_2)) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y} | \mathbf{q}^n, \mathbf{Y}^n]$. If such a pair exists and is unique, the decoder declares that $(\hat{\mathbf{W}}_1, \hat{\mathbf{W}}_2) = (m_1, m_2)$. Otherwise, the decoder declares an error.
- **Analysis of Probability of Error:** The average probability of error is given by

$$\begin{aligned}
P_e^n &= \sum_{\mathbf{s}^n \in \mathcal{S}^n, \mathbf{q}^n \in \mathcal{Q}^n} p(\mathbf{s}^n) p(\mathbf{q}^n) \Pr[\text{error} | \mathbf{s}^n, \mathbf{q}^n] \\
&\leq \sum_{\mathbf{s}^n \notin T_\epsilon^n[\mathcal{S}]} p(\mathbf{s}^n) + \sum_{\mathbf{q}^n \notin T_\epsilon^n[\mathcal{Q}]} p(\mathbf{q}^n) \\
&\quad + \sum_{\mathbf{s}^n \in T_\epsilon^n[\mathcal{S}], \mathbf{q}^n \in T_\epsilon^n[\mathcal{Q}]} p(\mathbf{q}^n) \Pr[\text{error} | \mathbf{s}^n, \mathbf{q}^n]. \tag{B.1}
\end{aligned}$$

The first term, $\Pr[\mathbf{s}^n \notin T_\epsilon^n[\mathcal{S}]]$, and the second term, $\Pr[\mathbf{q}^n \notin T_\epsilon^n[\mathcal{Q}]]$, in the right hand side expression of (B.1) go to zero as $n \rightarrow \infty$ by the strong asymptotic equipartition property (AEP) [12].

Without loss of generality, we can assume that $(\mathbf{W}_1, \mathbf{W}_2) = (1, 1)$ is sent, time sharing sequence is $\mathbf{Q}^n = \mathbf{q}^n$, and state realization is $\mathbf{S}^n = \mathbf{s}^n$. The probability of error is given by the conditional probability of error given $(\mathbf{W}_1, \mathbf{W}_2) = (1, 1)$, $\mathbf{Q}^n = \mathbf{q}^n \in T_\epsilon^n[\mathcal{Q}]$, and $\mathbf{S}^n = \mathbf{s}^n \in T_\epsilon^n[\mathcal{S}]$.

Let E_1 be the event that there is no sequence $\mathbf{U}_1^n(\mathbf{q}^n, \mathbf{W}_1, j)$ such that

$$\mathbf{U}_1^n(\mathbf{q}^n, 1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{S} | \mathbf{q}^n, \mathbf{s}^n]$$

. For any $\mathbf{U}_1^n(\mathbf{q}^n, 1, j)$ and $\mathbf{S}^n = \mathbf{s}^n$ generated independently according to $\prod p(\mathbf{u}_{1i} | \mathbf{q}_i)$ and $\prod p(\mathbf{s}_i)$, respectively, the probability that there exists at least one j such that $\mathbf{U}_1^n(\mathbf{q}^n, 1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{S} | \mathbf{q}^n, \mathbf{s}^n]$ is greater than $(1 - \epsilon)2^{-n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + \epsilon)}$ for n sufficiently large. There are J number of such \mathbf{U}_1^n 's in each bin. The probability of event E_1 , the probability that there is no \mathbf{U}_1^n for a given \mathbf{s}^n in a particular bin, is therefore bounded by

$$\Pr[E_1] \leq [1 - (1 - \epsilon)2^{-n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + \epsilon)}]^{2^{n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + 2\epsilon)}}. \quad (\text{B.2})$$

Taking the natural logarithm on both sides of (B.2), we obtain

$$\begin{aligned} \ln(\Pr[E_1]) &\leq 2^{n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + 2\epsilon)} \ln[1 - (1 - \epsilon)2^{-n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + \epsilon)}] \\ &\stackrel{(a)}{\leq} -2^{n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + 2\epsilon)} (1 - \epsilon) 2^{-n(\mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{Q}) + \epsilon)} \\ &= -(1 - \epsilon) 2^{n\epsilon}, \end{aligned} \quad (\text{B.3})$$

where (a) follows from the inequality $\ln(q) \leq (q - 1)$. From (B.3), $\Pr[E_1] \rightarrow 0$ as $n \rightarrow \infty$.

Under the event E_1^c , we can also assume that a particular sequence $\mathbf{U}_1^n(\mathbf{q}^n, 1, 1)$ in bin 1 is jointly strongly typical with $\mathbf{S}^n = \mathbf{s}^n$. Thus, codewords \mathbf{X}_1^n corresponding to the pair $(\mathbf{U}_1^n(\mathbf{q}^n, 1, 1), \mathbf{s}^n)$ and \mathbf{X}_2^n corresponding to $\mathbf{X}_2^n(\mathbf{q}^n, 1)$ are sent from the informed and the uninformed encoders, respectively.

Let E_2 be the event that $(\mathbf{U}_1^n(\mathbf{q}^n, 1, 1), \mathbf{X}_2^n(\mathbf{q}^n, 1), \mathbf{Y}^n) \notin T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y} | \mathbf{q}^n]$.

The Markov lemma [12] ensures jointly strong typicality of

$$(\mathbf{q}^n, \mathbf{s}^n, \mathbf{U}_1^n(\mathbf{q}^n, 1, 1), \mathbf{X}_2^n(\mathbf{q}^n, 1), \mathbf{Y}^n)$$

with high probability if $(\mathbf{q}^n, \mathbf{s}^n, \mathbf{U}_1^n(\mathbf{q}^n, 1, 1), \mathbf{X}_1^n)$ is jointly strongly typical and $(\mathbf{q}^n, \mathbf{X}_2^n(1))$ is jointly strongly typical. We can conclude that $\Pr[E_2|E_1^c] \rightarrow 0$ as $n \rightarrow \infty$.

Let E_3 be the event that $\mathbf{U}_1^n(\mathbf{q}^n, m_1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n, \mathbf{X}_2^n(\mathbf{q}^n, 1)]$.

The probability that

$$\mathbf{U}_1^n(\mathbf{q}^n, m_1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n, \mathbf{X}_2^n(\mathbf{q}^n, 1)]$$

for $m_1 \neq 1$, $1 \leq j \leq J$, is less than $2^{-n(\mathbb{I}(\mathbf{U}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{Q}) - \epsilon)}$ for sufficiently large n .

There are approximately JM_1 (exactly $J(M_1 - 1)$) such \mathbf{U}_1^n sequences in the codebook. Thus, the conditional probability of event E_3 given E_1^c and E_2^c is upper bounded by

$$\Pr[E_3|E_1^c, E_2^c] \leq 2^{-n((\mathbb{I}(\mathbf{U}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}|\mathbf{Q})) - R_1) + \epsilon}. \quad (\text{B.4})$$

From (B.4), $\Pr[E_3|E_1^c, E_2^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < \mathbb{I}(\mathbf{U}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}|\mathbf{Q})$ and $\epsilon > 0$.

Let E_4 be the event that $\mathbf{X}_2^n(\mathbf{q}^n, m_2) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n, \mathbf{U}_1^n(\mathbf{q}^n, 1, 1)]$ for $m_2 \neq 1$. The probability that $\mathbf{X}_2^n(\mathbf{q}^n, m_2) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n, \mathbf{U}_1^n(\mathbf{q}^n, 1, 1)]$ for $m_2 \neq 1$ is less than $2^{-n(\mathbb{I}(\mathbf{X}_2; \mathbf{Y}|\mathbf{U}_1, \mathbf{Q}) - \epsilon)}$ for sufficiently large n . There are approximately $M_2 = 2^{n(R_2 - 2\epsilon)}$ such \mathbf{X}_2^n sequences in the codebook. Thus, the conditional probability of event E_4 given E_1^c and E_2^c is upper bounded by

$$\Pr[E_4|E_1^c, E_2^c] \leq 2^{-n(\mathbb{I}(\mathbf{X}_2; \mathbf{Y}|\mathbf{U}_1, \mathbf{Q}) - R_2 + \epsilon)}. \quad (\text{B.5})$$

From (B.5), the $\Pr[E_4|E_1^c, E_2^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 < \mathbb{I}(\mathbf{X}_2; \mathbf{Y}|\mathbf{U}_1, \mathbf{Q})$.

Finally, let E_5 be the event that

$$(\mathbf{U}_1^n(\mathbf{q}^n, m_1, j), \mathbf{X}_2^n(\mathbf{q}^n, m_2)) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n]$$

for $m_1 \neq 1$, $1 \leq j \leq J$, and $m_2 \neq 1$. The probability that

$$(\mathbf{U}_1^n(\mathbf{q}^n, m_1, j), \mathbf{X}_2^n(\mathbf{q}^n, m_2)) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n]$$

for $m_1 \neq 1$, $1 \leq j \leq J$, and $m_2 \neq 1$ is less than $2^{-n(\mathbb{I}(\mathbf{U}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{Q}) - \epsilon)}$, for sufficiently large n . There are approximately JM_1 sequences \mathbf{U}_1^n and M_2 sequences \mathbf{U}_2^n in the codebook. Thus, the conditional probability of event E_5 given E_1^c and E_2^c is upper bounded by

$$\Pr[E_5|E_1^c, E_2^c] \leq 2^{-n(\mathbb{I}(\mathbf{U}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}|\mathbf{Q}) - (R_1 + R_2) + 3\epsilon)}. \quad (\text{B.6})$$

From (B.6), the $\Pr[E_5|E_1^c, E_2^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 + R_2 < \mathbb{I}(\mathbf{U}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S}|\mathbf{Q})$.

In terms of these events, $\Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n]$ in (B.1) can be upper-bounded via the union bound, and the fact that probabilities are less than one, as

$$\begin{aligned} \Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n] &\leq \Pr[E_1] + \Pr[E_2|E_1^c] + \Pr[E_3|E_1^c, E_2^c] \\ &\quad + \Pr[E_4|E_1^c, E_2^c] + \Pr[E_5|E_1^c, E_2^c]. \end{aligned} \quad (\text{B.7})$$

From Equation (B.7), it can be easily seen that $\Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n] \rightarrow 0$ as $n \rightarrow \infty$. Therefore, the probability of error goes to zero as $n \rightarrow \infty$ and completes the proof.

B.2 Proof of Theorem 2

We prove that, for any sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes with $P_e^n \rightarrow 0$ as $n \rightarrow \infty$, the rate pair (R_1, R_2) must satisfy (4.3). Fix n and consider a given code of block length n . The joint distribution of random variables $\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n, \mathbf{X}_1^n$ and \mathbf{X}_2^n is given by

$$p(\mathbf{s}^n, \mathbf{w}_1, \mathbf{w}_2, \mathbf{x}_1^n, \mathbf{x}_2^n, \mathbf{Y}^n) =$$

$$\frac{1}{2^{n(R_1+R_2)}} p(\mathbf{s}^n) p(\mathbf{x}_1^n | \mathbf{w}_1, \mathbf{s}^n) p(\mathbf{x}_2^n | \mathbf{w}_2) \prod_{j=1}^n p(\mathbf{y}_j | \mathbf{s}_j, \mathbf{x}_{1j}, \mathbf{x}_{2j})$$

where, $p(\mathbf{x}_1^n | \mathbf{w}_1, \mathbf{s}^n)$ is 1 if $\mathbf{X}_1^n = f_1^n(\mathbf{W}_1, \mathbf{S}^n)$ and 0 otherwise, and $p(\mathbf{x}_2^n | \mathbf{w}_2)$ is 1 if $\mathbf{X}_2^n = f_2^n(\mathbf{W}_2)$ and 0 otherwise. Using given code sequence, it is possible to estimate $(\mathbf{W}_1, \mathbf{W}_2)$ from the received sequence \mathbf{Y}^n with arbitrarily low probability of error as $n \rightarrow \infty$. Hence, the conditional entropy of $(\mathbf{W}_1, \mathbf{W}_2)$ given \mathbf{Y}^n must be small. By Fano's inequality,

$$\mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{Y}^n) \leq \mathbb{H}(\mathbf{W}_1, \mathbf{W}_2 | \mathbf{Y}^n) \leq n\epsilon_n, \quad (\text{B.8})$$

$$\mathbb{H}(\mathbf{W}_2 | \mathbf{W}_1, \mathbf{Y}^n) \leq \mathbb{H}(\mathbf{W}_1, \mathbf{W}_2 | \mathbf{Y}^n) \leq n\epsilon_n, \quad (\text{B.9})$$

where, $\epsilon_n \rightarrow 0$ as $P_e^n \rightarrow 0$.

We can bound the rate R_1 as

$$\begin{aligned} nR_1 &= \mathbb{H}(\mathbf{W}_1) \\ &= \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2) \\ &= \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2) + \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{Y}^n) \\ &\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2) + n\epsilon_n \\ &\stackrel{(b)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}^i | \mathbf{W}_2) - \mathbb{I}(\mathbf{W}_1, \mathbf{S}_i^n; \mathbf{Y}^{i-1} | \mathbf{W}_2) + n\epsilon_n \\ &\stackrel{(c)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}) - \mathbb{I}(\mathbf{S}_i; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{H}(\mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}) - \mathbb{H}(\mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{W}_1, \mathbf{S}_{i+1}^n) \\ &\quad - \mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) + \mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}, \mathbf{W}_2) + n\epsilon_n \\ &\stackrel{(d)}{\leq} \sum_{i=1}^n \mathbb{H}(\mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}) - \mathbb{H}(\mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{W}_1, \mathbf{S}_{i+1}^n) \\ &\quad - \mathbb{H}(\mathbf{S}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}) + \mathbb{H}(\mathbf{S}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{W}_1, \mathbf{S}_{i+1}^n) + n\epsilon_n, \end{aligned} \quad (\text{B.10})$$

where,

(a) follows from Fano's inequality

(b) follows from : $\mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}^i | \mathbf{W}_2) = \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2)$ for $i = n$; $\mathbb{I}(\mathbf{W}_1, \mathbf{S}_i^n; \mathbf{Y}^{i-1} | \mathbf{W}_2) = 0$ for $i = 1$; and the sum of remaining terms equals to zero

(c) follows from applying chain rule for mutual information to $\{\mathbf{Y}^{i-1}, \mathbf{Y}_i\}$ in the first term and to $\{\{\mathbf{W}_1, \mathbf{S}_{i+1}^n\}, \mathbf{S}_i\}$ in the second term

(d) follows from $\mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) = \mathbb{H}(\mathbf{S}_i | \mathbf{W}_2) \geq \mathbb{H}(\mathbf{S}_i | \mathbf{W}_2, \mathbf{Y}^{i-1})$.

Then the rate R_2 can also be bounded as .

$$\begin{aligned}
nR_2 &= \mathbb{H}(\mathbf{W}_2) \\
&= \mathbb{H}(\mathbf{W}_2 | \mathbf{W}_1) \\
&= \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^n | \mathbf{W}_1) + \mathbb{H}(\mathbf{W}_2 | \mathbf{W}_1, \mathbf{Y}^n) \\
&\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^n | \mathbf{W}_1) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^i | \mathbf{W}_1, \mathbf{S}_{i+1}^n) - \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_i^n) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n) + \mathbb{I}(\mathbf{W}_2; \mathbf{Y}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}) \\
&\quad - \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_i^n) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_2; \mathbf{Y}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}) - \mathbb{I}(\mathbf{W}_2; \mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}) + n\epsilon_n, \quad (\text{B.11})
\end{aligned}$$

where:

(a) follows from Fano's inequality

(b) follows from : $\mathbb{I}(\mathbf{W}_2; \mathbf{Y}^i | \mathbf{W}_1, \mathbf{S}_{i+1}^n) = \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^n | \mathbf{W}_2)$ for $i = n$; $\mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_i^n) = 0$ for $i = 1$; and the sum of remaining terms equals zero

(c) follows from applying chain rule for mutual information to $\{\mathbf{Y}^{i-1}, \mathbf{Y}_i\}$ in the first term.

(d) follows from $\mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n) - \mathbb{I}(\mathbf{W}_2; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_i^n) = -\mathbb{I}(\mathbf{W}_2; \mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1})$.

Finally, the sum rate $R_1 + R_2$ can be upper bounded as

$$\begin{aligned}
n(R_1 + R_2) &= \mathbb{H}(\mathbf{W}_1, \mathbf{W}_2) \\
&= \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2; \mathbf{Y}^n) + \mathbb{H}(\mathbf{W}_1, \mathbf{W}_2 | \mathbf{Y}^n) \\
&\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2; \mathbf{Y}^n) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}^i) - \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_i^n; \mathbf{Y}^{i-1}) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{Y}^{i-1}) - \mathbb{I}(\mathbf{S}_i; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) + n\epsilon_n \\
&= \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{Y}^{i-1}) \\
&\quad - \mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) + \mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}, \mathbf{W}_2) + n\epsilon_n \\
&\stackrel{(d)}{\leq} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{Y}^{i-1}) \\
&\quad - \mathbb{H}(\mathbf{S}_i | \mathbf{Y}^{i-1}) + \mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}, \mathbf{W}_2) + n\epsilon_n, \tag{B.12}
\end{aligned}$$

where:

(a) follows from Fano's inequality

(b) follows from : $\mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}^i) = \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2; \mathbf{Y}^n)$ for $i = n$; $\mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_i^n; \mathbf{Y}^{i-1}) = 0$ for $i = 1$; and the sum of remaining terms equals to zero

(c) follows from applying chain rule for mutual information to $\{\mathbf{Y}^{i-1}, \mathbf{Y}_i\}$ in the first term and to $\{\{\mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2\}, \mathbf{S}_i\}$ in the second term.

(d) follows from $\mathbb{H}(\mathbf{S}_i | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2) = \mathbb{H}(\mathbf{S}_i) \geq \mathbb{H}(\mathbf{S}_i | \mathbf{Y}^{i-1})$.

Let $\mathbf{V}(i) := \{\mathbf{Y}^{i-1}\}$, $\mathbf{U}_1(i) := \{\mathbf{W}_1, \mathbf{S}_{i+1}^n\}$, and $\mathbf{U}_2(i) := \{\mathbf{W}_2\}$ for $i \in \{1, 2, \dots, n\}$.

We can then write (B.10)-(B.12) more compactly as

$$R_1 \leq \frac{1}{n} \sum_{i=1}^n [\mathbb{I}(\mathbf{U}_1(i); \mathbf{Y}_i | \mathbf{V}(i), \mathbf{U}_2(i)) - \mathbb{I}(\mathbf{U}_1(i); \mathbf{S}_i | \mathbf{V}(i), \mathbf{U}_2(i))] + \epsilon_n \tag{B.13a}$$

$$R_2 \leq \frac{1}{n} \sum_{i=1}^n [\mathbb{I}(\mathbf{U}_2(i); \mathbf{Y}_i | \mathbf{V}(i), \mathbf{U}_1(i)) - \mathbb{I}(\mathbf{U}_2(i); \mathbf{S}_i | \mathbf{V}(i), \mathbf{U}_1(i))] + \epsilon_n \quad (\text{B.13b})$$

$$R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^n [\mathbb{I}(\mathbf{U}_1(i), \mathbf{U}_2(i); \mathbf{S}_i | \mathbf{V}(i)) - \mathbb{I}(\mathbf{U}_1(i), \mathbf{U}_2(i); \mathbf{S}_i | \mathbf{V}(i))] + \epsilon_n. \quad (\text{B.13c})$$

Let us also define random variable \mathbf{Q} to uniformly take value in the set $\mathcal{Q} = \{1, 2, \dots, n\}$. Then (B.13a) can be written as

$$\begin{aligned} R_1 &\leq \frac{1}{n} \sum_{i=1}^n [\mathbb{I}(\mathbf{U}_1(i); \mathbf{Y}_i | \mathbf{V}(i), \mathbf{U}_2(i), \mathbf{Q} = i) - \mathbb{I}(\mathbf{U}_1(i); \mathbf{S}_i | \mathbf{V}, \mathbf{U}_2, \mathbf{Q} = i)] + \epsilon_n \\ &= \mathbb{I}(\mathbf{U}_1(\mathbf{Q}); \mathbf{Y}_{\mathbf{Q}} | \mathbf{V}(\mathbf{Q}), \mathbf{U}_2(\mathbf{Q}), \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1(\mathbf{Q}); \mathbf{S}_{\mathbf{Q}} | \mathbf{Q}) + \epsilon_n \\ &= \mathbb{I}(\mathbf{U}_1; \mathbf{Y} | \mathbf{V}, \mathbf{U}_2, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1; \mathbf{S} | \mathbf{V}, \mathbf{U}_2, \mathbf{Q}) + \epsilon_n, \end{aligned} \quad (\text{B.14})$$

Similarly, (B.13a) and (B.13b) can be written as

$$R_2 \leq \mathbb{I}(\mathbf{U}_2; \mathbf{Y} | \mathbf{V}, \mathbf{U}_1, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_2; \mathbf{S} | \mathbf{V}, \mathbf{U}_1, \mathbf{Q}) + \epsilon_n \quad (\text{B.15a})$$

$$R_1 + R_2 \leq \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{S} | \mathbf{V}, \mathbf{Q}) - \mathbb{I}(\mathbf{U}_1, \mathbf{U}_2; \mathbf{S} | \mathbf{V}, \mathbf{Q}) + \epsilon_n. \quad (\text{B.15b})$$

where, $(\mathbf{Q}, \mathbf{S}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in \mathcal{P}^o$, where \mathcal{P}_1^o is defined in Definition 5.

Hence, taking the limit as $n \rightarrow \infty$, $P_e^n \rightarrow 0$, (B.14), (B.15a), and (B.15b) become (4.3a), (4.3b), and (4.3c), respectively.

B.3 Proof of Theorem 5

B.3.1 Achievability

We denote by $T_\epsilon^n[\mathbf{X}, \mathbf{Y}]$ the set of jointly strongly typical sequences [12, 14] with distribution $p(\mathbf{x}, \mathbf{y})$. Let $T_\epsilon^n[\mathbf{X}, \mathbf{Y} | \mathbf{x}^n] := \{\mathbf{y}^n : (\mathbf{x}^n, \mathbf{y}^n) \in T_\epsilon^n[\mathbf{X}, \mathbf{Y}]\}$. In this section, we construct a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes with $P_e^n \rightarrow 0$ as $n \rightarrow \infty$ if (R_1, R_2) satisfies (4.12). Fix $\epsilon > 0$ and a distribution $p(\mathbf{q})p(\mathbf{s})p(\mathbf{x}_2 | \mathbf{q})p(\mathbf{u}, \mathbf{x}_1 | \mathbf{q}, \mathbf{s}, \mathbf{x}_2)$ on $(\mathbf{Q}, \mathbf{S}, \mathbf{U}, \mathbf{X}_1, \mathbf{X}_2)$. In this case, \mathbf{Q} is a time-sharing random variable. Generate the time-sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$ according to $\prod p(\mathbf{q}_i)$. Without loss generality,

it is assumed that the time-sharing sequence is non-causally known at the encoders and the decoder.

Encoding

The encoding strategy at the two encoders is as follows. Let $M_1 = 2^{n(R_1 - 4\epsilon)}$, $M_2 = 2^{n(R_2 - 2\epsilon)}$, and $J = 2^{n(\mathbb{I}(\mathbf{U}; \mathbf{S} | \mathbf{Q}, \mathbf{X}_2) + 2\epsilon)}$. At the uninformed encoder, generate the sequences \mathbf{X}_2^n , according to $\prod p(x_{2,i} | \mathbf{q}_i)$, where $1 \leq m_2 \leq M_2$. The uninformed encoder sends the codeword $\mathbf{X}_2^n(\mathbf{q}^n, \mathbf{w}_2)$ to send the message $\mathbf{W}_2 \in \{1, 2, \dots, M_2\}$ for a given time-sharing sequence $\mathbf{Q}^n = \mathbf{q}^n$.

At the informed encoder, generate JM_1 sequences $\mathbf{U}^n(\mathbf{q}^n, m_1, m_2, j)$, with $\prod p(\mathbf{u}_i | \mathbf{q}_i, \mathbf{x}_{2,i}(\mathbf{q}, m_2))$, where $1 \leq m_1 \leq M_1$, $1 \leq m_2 \leq M_2$ and $1 \leq j \leq J$. Here, (m_1, m_2) indexes bins and j indexes sequences within a particular bin (m_1, m_2) . To encode the message $(\mathbf{W}_1, \mathbf{W}_2) \in \{1, 2, \dots, M_1\} \times \{1, 2, \dots, M_2\}$ given $\mathbf{S}^n = \mathbf{s}^n$ and $\mathbf{Q}^n = \mathbf{q}^n$, the informed encoder looks in bin $(\mathbf{W}_1, \mathbf{W}_2)$ for a sequence $\mathbf{U}^n(\mathbf{q}^n, \mathbf{w}_1, \mathbf{w}_2, j)$, $1 \leq j \leq J$, such that

$$\mathbf{U}^n(\mathbf{q}^n, \mathbf{w}_1, \mathbf{w}_2, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{S}, \mathbf{X}_2 | \mathbf{q}^n, \mathbf{s}^n, \mathbf{X}_2^n(\mathbf{w}_2)].$$

Then, the informed encoder generates \mathbf{X}_1^n with

$$\prod p(x_{1,i} | \mathbf{q}_i, \mathbf{s}_i, \mathbf{u}_i(\mathbf{q}^n, \mathbf{w}_1, \mathbf{w}_2, j), \mathbf{x}_{2,i}(\mathbf{q}^n, \mathbf{w}_2)).$$

Given $(\mathbf{S}^n, \mathbf{X}_1^n, \mathbf{X}_2^n)$, the channel generates the output \mathbf{Y}^n according to conditional probability distribution $\prod_i p(\mathbf{y}_i | \mathbf{s}_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$.

Decoding

The decoder chooses a pair $(\mathbf{U}^n(\mathbf{q}^n, m_1, m_2, j), \mathbf{X}_2^n(m_2))$, $1 \leq m_1 \leq M_1$, $1 \leq j \leq J$, and $1 \leq m_2 \leq M_2$ such that $(\mathbf{U}^n(\mathbf{q}^n, m_1, m_2, j), \mathbf{X}_2^n(\mathbf{q}^n, m_2)) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{X}_2, \mathbf{Y} | \mathbf{q}^n, \mathbf{Y}^n]$. If such a pair exists and is unique, the decoder declares that $(\hat{\mathbf{W}}_1, \hat{\mathbf{W}}_2) = (m_1, m_2)$. Otherwise, the decoder declares an error.

Analysis of Probability of Error

The average probability of error is given by

$$\begin{aligned}
P_e^n &= \sum_{s^n \in \mathcal{S}^n, \mathbf{q}^n \in \mathcal{Q}^n} p(s^n)p(\mathbf{q}^n)\Pr[\text{error}|s^n, \mathbf{q}^n] \\
&\leq \sum_{s^n \notin T_\epsilon^n[\mathcal{S}]} p(s^n) + \sum_{\mathbf{q}^n \notin T_\epsilon^n[\mathcal{Q}]} p(\mathbf{q}^n) \\
&\quad + \sum_{s^n \in T_\epsilon^n[\mathcal{S}], \mathbf{q}^n \in T_\epsilon^n[\mathcal{Q}]} p(s^n)p(\mathbf{q}^n)\Pr[\text{error}|s^n, \mathbf{q}^n]. \tag{B.16}
\end{aligned}$$

The first term, $\Pr[s^n \notin T_\epsilon^n[\mathcal{S}]]$, and the second term, $\Pr[\mathbf{q}^n \notin T_\epsilon^n[\mathcal{Q}]]$, in the right hand side of (B.16) go to zero as $n \rightarrow \infty$ by the strong asymptotic equipartition property (AEP) [12].

Without loss of generality, we can assume that $(\mathbf{W}_1, \mathbf{W}_2) = (1, 1)$ is sent, the time sharing sequence is $\mathbf{Q}^n = \mathbf{q}^n$, and the channel state realization is $\mathbf{S}^n = s^n$. Then $\Pr[\text{error}|s^n, \mathbf{q}^n]$ is the conditional probability of error given $(\mathbf{W}_1, \mathbf{W}_2) = (1, 1)$, $\mathbf{Q}^n = \mathbf{q}^n \in T_\epsilon^n[\mathcal{Q}]$, and $\mathbf{S}^n = s^n \in T_\epsilon^n[\mathcal{S}]$.

- Let E_1 be the event that there is no sequence $\mathbf{U}^n(\mathbf{q}^n, 1, 1, j)$ such that

$$\mathbf{U}^n(\mathbf{q}^n, 1, 1, j) \in T_\epsilon^n[\mathcal{Q}, \mathcal{U}, \mathcal{S}, \mathcal{X}_2 | \mathbf{q}^n, s^n, \mathcal{X}_2^n(\mathbf{q}^n, 1)]$$

. Since $\mathbf{U}^n(\mathbf{q}^n, 1, 1, j)$ and $\mathbf{S}^n = s^n$ are generated independently according to $\prod p(\mathbf{u}_i | \mathbf{q}_i, \mathbf{x}_{2,i}(\mathbf{q}^n, 1))$ and $\prod p(s_i)$, respectively; and there are J sequences in each bin, the probability of event E_1 goes to zero as $n \rightarrow \infty$.

Under the event E_1^c , we can also assume that a particular sequence $\mathbf{U}^n(\mathbf{q}^n, 1, 1, 1)$ in bin (1, 1) is jointly strongly typical with $(s^n, \mathbf{q}^n, \mathcal{X}_2^n(\mathbf{q}^n, 1))$. Thus, \mathbf{X}_1^n corresponding to $(\mathbf{U}^n(\mathbf{q}^n, 1, 1, 1), s^n, \mathbf{q}^n, \mathcal{X}_2^n(\mathbf{q}^n, 1))$ and $\mathcal{X}_2^n(\mathbf{q}^n, 1)$ are sent from the informed and the uninformed encoders, respectively.

- Let E_2 be the event that $(\mathbf{U}^n(\mathbf{q}^n, 1, 1, 1), \mathcal{X}_2^n(\mathbf{q}^n, 1), \mathbf{Y}^n) \notin T_\epsilon^n[\mathcal{Q}, \mathcal{U}, \mathcal{X}_2, \mathcal{Y} | \mathbf{q}^n]$.

According to the Markov lemma [12], $\Pr[E_2 | E_1^c] \rightarrow 0$ as $n \rightarrow \infty$. Let E_3 be

the event that $\mathbf{U}^n(\mathbf{q}^n, m_1, 1, j) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n, \mathbf{X}_2^n(q^n, 1)]$ for $m_1 \neq 1$ and $j \neq 1$. Using properties of strongly typical sequences [12], it can be easily shown that $\Pr[E_3|E_1^c, E_2^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < \mathbb{I}(\mathbf{U}; \mathbf{Y}|\mathbf{Q}, \mathbf{X}_2) - \mathbb{I}(\mathbf{U}; \mathbf{S}|\mathbf{Q}, \mathbf{X}_2)$.

- Finally, let E_4 be the event that

$(\mathbf{U}_1^n(\mathbf{q}^n, m_1, m_2, j), \mathbf{X}_2^n(\mathbf{q}^n, m_2)) \in T_\epsilon^n[\mathbf{Q}, \mathbf{U}, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{Y}^n]$ for $m_1 \neq 1$, $1 \leq j \leq J$, and $m_2 \neq 1$. Using the properties of strongly typical sequences, it can be shown that the $\Pr[E_5|E_1^c, E_2^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 + R_2 < \mathbb{I}(\mathbf{U}, \mathbf{X}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{I}(\mathbf{U}; \mathbf{S}|\mathbf{Q}, \mathbf{X}_2)$.

In terms of these events, $\Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n]$ in (B.16) can be upper-bounded via the union bound, and the fact that probabilities are less than one, as

$$\begin{aligned} \Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n] &\leq \Pr[E_1] + \Pr[E_2|E_1^c] + \Pr[E_3|E_1^c, E_2^c] \\ &\quad + \Pr[E_4|E_1^c, E_2^c]. \end{aligned} \tag{B.17}$$

From (B.17) and the above limiting arguments, we have $\Pr[\text{error}|\mathbf{s}^n, \mathbf{q}^n] \rightarrow 0$ as $n \rightarrow \infty$. Therefore, the average probability of error $P_e^n \rightarrow 0$ as $n \rightarrow \infty$, completing the proof.

B.3.2 Converse

First, we prove that, for any sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, n)$ codes with $P_e^n \rightarrow 0$ as $n \rightarrow \infty$, the rate pair $(R_1, R_2) \in \mathcal{C}$. Fix n and consider a given code of block length n . The joint distribution of random variables \mathbf{W}_1 , \mathbf{W}_2 , \mathbf{S}^n , \mathbf{X}_1^n , and \mathbf{X}_2^n is given by

$$\begin{aligned} p(\mathbf{s}^n, \mathbf{w}_1, \mathbf{w}_2, \mathbf{x}_1^n, \mathbf{x}_2^n, \mathbf{y}^n) &= \\ &\frac{1}{\lceil 2^{nR_1} \rceil \lceil 2^{nR_2} \rceil} p(\mathbf{s}^n) p(\mathbf{x}_1^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n) p(\mathbf{x}_2^n | \mathbf{w}_2) \\ &\times p(\mathbf{y}^n | \mathbf{s}^n, \mathbf{x}_1^n, \mathbf{x}_2^n) \end{aligned}$$

where, $p(\mathbf{x}_1^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ is 1 if $\mathbf{x}_1^n = f_1^n(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ and 0 otherwise, and $p(\mathbf{x}_2^n | \mathbf{w}_2)$ is 1 if $\mathbf{x}_2^n = f_2^n(\mathbf{w}_2)$ and 0 otherwise.

We can bound the rate R_1 as

$$\begin{aligned}
nR_1 &\leq \mathbb{H}(\mathbf{W}_1) \\
&= \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2) \\
&= \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2) + \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{Y}^n) \\
&\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n [\mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}^i | \mathbf{W}_2) - \mathbb{I}(\mathbf{W}_1, \mathbf{S}_i^n; \mathbf{Y}^{i-1} | \mathbf{W}_2)] + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n [\mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}) \\
&\quad - \mathbb{I}(\mathbf{S}_i; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2)] + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n [\mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{X}_{2,i}) \\
&\quad - \mathbb{I}(\mathbf{S}_i; \mathbf{Y}^{i-1} | \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2, \mathbf{X}_{2,i})] + n\epsilon_n \\
&\stackrel{(e)}{\leq} \sum_{i=1}^n [\mathbb{H}(\mathbf{Y}_i | \mathbf{X}_{2,i}) - \mathbb{H}(\mathbf{Y}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{X}_{2,i}) \\
&\quad + \mathbb{H}(\mathbf{S}_i | \mathbf{W}_2, \mathbf{Y}^{i-1}, \mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{X}_{2,i}) \\
&\quad - \mathbb{H}(\mathbf{S}_i | \mathbf{X}_{2,i})] + n\epsilon_n \\
&= \sum_{i=1}^n [\mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}; \mathbf{Y}_i | \mathbf{X}_{2,i}) \\
&\quad - \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}; \mathbf{S}_i | \mathbf{X}_{2,i})] + n\epsilon_n \tag{B.18}
\end{aligned}$$

where:

(a) follows from Fano's inequality with $\epsilon_n \rightarrow 0$ as $P_e^n \rightarrow 0$,

(b) follows from $\mathbb{I}(\mathbf{W}_1, \mathbf{S}_{i+1}^n; \mathbf{Y}^i | \mathbf{W}_2) = \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2)$ for $i = n$; $\mathbb{I}(\mathbf{W}_1, \mathbf{S}_i^n; \mathbf{Y}^{i-1} | \mathbf{W}_2) = 0$ for $i = 1$; and the sum of the remaining terms equals to zero,

(c) follows from applying the chain rule for mutual information to (Y^{i-1}, Y_i) in the first term and to $(\{W_1, S_{i+1}^n\}, S_i)$ in the second term,

(d) follows from $X_{2,i}$ is a deterministic function of W_2 for $i = \{1, 2, \dots, n\}$,

(e) follows from $\mathbb{H}(Y_i|X_{2,i}) \geq \mathbb{H}(Y_i|W_2, Y^{i-1}, X_{2,i})$ and S_i being independent of (S_{i+1}^n, W_1, W_2) .

Finally, the sum rate $R_1 + R_2$ can be upper bounded as

$$\begin{aligned}
n(R_1 + R_2) &= \mathbb{H}(W_1, W_2) \\
&= \mathbb{I}(W_1, W_2; Y^n) + \mathbb{H}(W_1, W_2|Y^n) \\
&\stackrel{(a)}{\leq} \mathbb{I}(W_1, W_2; Y^n) + n\epsilon_n \\
&\stackrel{(b)}{=} \sum_{i=1}^n \mathbb{I}(W_1, W_2, S_{i+1}^n; Y^i) \\
&\quad - \mathbb{I}(W_1, W_2, S_i^n; Y^{i-1}) + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{i=1}^n \mathbb{I}(W_1, W_2, S_{i+1}^n; Y_i|Y^{i-1}) \\
&\quad - \mathbb{I}(S_i; Y^{i-1}|W_1, S_{i+1}^n, W_2) + n\epsilon_n \\
&= \sum_{i=1}^n \mathbb{I}(W_1, W_2, S_{i+1}^n; Y_i|Y^{i-1}) \\
&\quad - \mathbb{H}(S_i|W_1, S_{i+1}^n, W_2) \\
&\quad + \mathbb{H}(S_i|W_1, S_{i+1}^n, Y^{i-1}, W_2) + n\epsilon_n \\
&\leq \sum_{i=1}^n \mathbb{I}(W_1, W_2, S_{i+1}^n, Y^{i-1}; Y_i) \\
&\quad - \mathbb{H}(S_i) + \mathbb{H}(S_i|W_1, S_{i+1}^n, Y^{i-1}, W_2) + n\epsilon_n, \\
&\stackrel{(d)}{=} \sum_{i=1}^n \mathbb{I}(W_1, W_2, S_{i+1}^n, Y^{i-1}, X_{2,i}; Y_i) \\
&\quad - \mathbb{I}(W_1, W_2, S_{i+1}^n, Y^{i-1}, X_{2,i}; S_i),
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(e)}{=} \sum_{i=1}^n \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}, \mathbf{X}_{2,i}; \mathbf{Y}_i) \\
& \quad - \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1}; \mathbf{S}_i | \mathbf{X}_{2,i}), \tag{B.19}
\end{aligned}$$

where:

- (a) follows from Fano's inequality with $\epsilon_n \rightarrow 0$ as $P_e^n \rightarrow 0$,
- (b) follows from $\mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n; \mathbf{Y}^i) = \mathbb{I}(\mathbf{W}_1, \mathbf{W}_2; \mathbf{Y}^n)$ for $i = n$; $\mathbb{I}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_i^n; \mathbf{Y}^{i-1}) = 0$ for $i = 1$; and the sum of remaining terms equals zero,
- (c) follows from applying the chain rule for mutual information to $(\mathbf{Y}^{i-1}, \mathbf{Y}_i)$ in the first term and to $(\{\mathbf{W}_1, \mathbf{S}_{i+1}^n, \mathbf{W}_2\}, \mathbf{S}_i)$ in the second term,
- (d) follows from $\mathbf{X}_{2,i}$ being a deterministic function of \mathbf{W}_2 for $i = \{1, 2, \dots, n\}$.

Let us define $\mathbf{U}(i) := (\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_{i+1}^n, \mathbf{Y}^{i-1})$ and \mathbf{Q} to take values uniformly in the set $\mathcal{Q} = \{1, 2, \dots, n\}$. As $n \rightarrow \infty$, we obtain $(R_1, R_2) \in \mathcal{C}$ from (B.18) and (B.19) where the distribution on $(\mathbf{Q}, \mathbf{S}, \mathbf{X}_2, \mathbf{X}_1, \mathbf{U}, \mathbf{Y})$ is

$$p(\mathbf{s})p(\mathbf{q})p(\mathbf{x}_{2,\mathbf{q}}|\mathbf{q})p(\mathbf{u}_{\mathbf{q}}, \mathbf{x}_{1,\mathbf{q}}|\mathbf{q})p(\mathbf{y}|\mathbf{s}, \mathbf{x}_{1,\mathbf{q}}, \mathbf{x}_{2,\mathbf{q}}).$$

APPENDIX C

C.1 Proof of Theorem 6

In this section, we demonstrate existence of a sequence of MAC IE codes $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D_1^{(n)}, D_2^{(n)}, n)$ with $\lim_{n \rightarrow \infty} P_e^n = 0$, and $\lim_{n \rightarrow \infty} D_i^{(n)} \leq \Delta_i$ for $i = 1, 2$ if the rate pair (R_1, R_2) satisfying (5.4). Fix $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$ and n . We construct a MAC IE code $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D_1^{(n)}, D_2^{(n)}, n)$ as follows.

- **Code construction:** Throughout the achievability proof, let $i \in \mathcal{J} = \{1, 2\}$.

Generate time sharing sequence $\mathbf{Q}^n = (Q_1, Q_2, \dots, Q_n)$ whose elements are i.i.d. with distribution $p(\mathbf{q})$. At Encoder i , for each $\mathbf{s}_i^n \in \mathcal{S}_i^n$, generate 2^{nR_i} \mathbf{X}_i^n sequence drawn according to $\prod_{j=1}^n p(x_{ij} | s_{ij}, \mathbf{q}_j)$. Call these sequences $\mathbf{X}_i^n(\mathbf{Q}^n, \mathbf{S}_i^n, m_i)$ where $m_i \in \{1, 2, \dots, 2^{nR_i}\}$, $i = 1, 2$. In this way, the codebooks are generated at each encoder and revealed to the decoder.

Since the sequence \mathbf{Q}^n serves as time sharing sequence, it can be assumed that the sequence \mathbf{Q}^n is known at both the encoders and at the decoder without loss of generality.

- **Encoding:** Encoder i , upon observing \mathbf{S}_i^n at the output of host source i and time sharing random sequence \mathbf{Q}^n , sends message $W_i \in \{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$ by transmitting the codeword $\mathbf{X}_i^n(\mathbf{Q}^n, \mathbf{S}_i^n, W_i)$. In this way, the codeword \mathbf{X}_i^n is chosen and transmitted from Encoder i for a given time sharing sequence \mathbf{Q}^n ,

a given host sequence \mathbf{S}_i^n , and a message W_i .

- **Decoding:** Fix $0 < \epsilon_1 < \epsilon$. Since the decoder knows the time sharing sequence $Q^n = \mathbf{q}^n$, the decoder, upon receiving the channel output Y^n , looks for a tuple $(X_1^n(\mathbf{q}^n, \mathbf{s}_1^n, m_1), X_2^n(\mathbf{q}^n, \mathbf{s}_2^n, m_2))$ such that $(X_1^n(\mathbf{q}^n, \mathbf{s}_1^n, m_1), X_2^n(\mathbf{q}^n, \mathbf{s}_2^n, m_2), Y^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, X_1, X_2, Y | \mathbf{q}^n, \mathbf{s}_1^n, \mathbf{s}_2^n]$ for all $(\mathbf{s}_1^n, \mathbf{s}_2^n) \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2]$. If a unique vector of sequences exists, the decoder declares that $(\hat{W}_1, \hat{W}_2, \hat{\mathbf{S}}_1^n, \hat{\mathbf{S}}_2^n) = (m_1, m_2, \mathbf{s}_1^n, \mathbf{s}_2^n)$. Otherwise, the decoder declares an error. In this way, the messages and the host sequences are decoded at the decoder.

- **Probability of error:** The average probability of error is given by the following

$$\begin{aligned}
P_e^n &= \sum_{(\mathbf{s}_1^n, \mathbf{s}_2^n, \mathbf{q}^n) \in \mathcal{S}_1^n \times \mathcal{S}_2^n \times \mathcal{Q}^n} p(\mathbf{q}^n) p(\mathbf{s}_1^n, \mathbf{s}_2^n) \Pr[\text{error} | (\mathbf{s}_1^n, \mathbf{s}_2^n, \mathbf{q}^n)] \\
&\leq \sum_{(\mathbf{q}^n, \mathbf{s}_1^n, \mathbf{s}_2^n) \notin T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]} p(\mathbf{q}^n) p(\mathbf{s}_1^n, \mathbf{s}_2^n) \\
&+ \sum_{(\mathbf{q}^n, \mathbf{s}_1^n, \mathbf{s}_2^n) \in T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]} p(\mathbf{s}_1^n, \mathbf{s}_2^n) p(\mathbf{q}^n) \Pr[\text{error} | (\mathbf{s}_1^n, \mathbf{s}_2^n, \mathbf{q}^n)] \quad (\text{C.1})
\end{aligned}$$

The first term, $\Pr[(\mathbf{q}^n, \mathbf{s}_1^n, \mathbf{s}_2^n) \notin T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]]$, in the right hand side expression of (C.1) goes to zero as $n \rightarrow \infty$ by Lemma 2.

Without loss of generality, it can be assumed that the time-sharing sequence is \mathbf{q}^n , the output of the host source i is $\tilde{\mathbf{s}}_i^n$, and $W_i = 1$ is being transmitted from Encoder i . Hence, the codeword $X_i^n(\mathbf{q}^n, \tilde{\mathbf{s}}_i^n, 1)$ is transmitted from Encoder i . It is also assumed that the time-sharing random sequence $Q^n = \mathbf{q}^n$ is known at both the encoders and the decoder. Let F be the event that $(\tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n)$ and \mathbf{q}^n are the output of the host source pair and time sharing sequence, respectively and $(\mathbf{q}^n, \mathbf{s}_1^n, \mathbf{s}_2^n) \in T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]$.

The following error events are considered to compute $\Pr[\text{error}|F]$ and can be made to approach zero as $n \rightarrow \infty$.

1. $E_1: (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, 1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, 1), \mathbf{Y}^n) \notin T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F . By using Lemma 2, we can show that $\Pr[E_1|F] \rightarrow 0$ as $n \rightarrow \infty$.
2. $E_2: (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, 1), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F for all $m_1 \neq 1$. It can be shown that $\Pr(E_2|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_1 < \mathbb{I}(\mathbf{X}_1; \mathbf{Y}|\mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_2, \mathbf{Q})$.
3. $E_3: (\mathbf{X}_1^n(\mathbf{q}^n, \mathbf{s}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, 1), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{s}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F for all $m_1 \in M_1$ and for all $\mathbf{s}_1^n \neq \tilde{\mathbf{s}}_1^n$ and $\mathbf{s}_1^n \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2|\tilde{\mathbf{s}}_2^n]$. It can be shown that $\Pr(E_3|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_1 < \mathbb{I}(\mathbf{S}_1, \mathbf{X}_1; \mathbf{Y}|\mathbf{S}_2, \mathbf{X}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_1|\mathbf{S}_2)$.
4. $E_4: (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, 1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F for all $m_2 \neq 1$. It can be shown that $\Pr(E_4|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_2 < \mathbb{I}(\mathbf{X}_2; \mathbf{Y}|\mathbf{S}_1, \mathbf{X}_1, \mathbf{S}_2, \mathbf{Q})$.
5. $E_5: (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, 1), \mathbf{X}_2^n(\mathbf{q}^n, \mathbf{s}_2^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \mathbf{s}_2^n]$ under the event F for all $m_2 \in M_2$, $\mathbf{s}_2^n \neq \tilde{\mathbf{s}}_2^n$, and $\mathbf{s}_2^n \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2|\tilde{\mathbf{s}}_1^n]$. It can be shown that $\Pr(E_5|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_2 < \mathbb{I}(\mathbf{X}_2, \mathbf{S}_2; \mathbf{Y}|\mathbf{S}_1, \mathbf{X}_1, \mathbf{S}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2|\mathbf{S}_1)$.
6. $E_6: (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \mathbf{s}_2^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \mathbf{s}_2^n]$ under the event F for all $m_1 \in M_1$, $m_2 \in M_2$, $\mathbf{s}_2^n \neq \tilde{\mathbf{s}}_2^n$ and $\mathbf{s}_2^n \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2|\tilde{\mathbf{s}}_1^n]$. It can be shown that $\Pr(E_6|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $R_1 + R_2 < \mathbb{I}(\mathbf{X}_1, \mathbf{S}_2, \mathbf{X}_2; \mathbf{Y}|\mathbf{S}_1, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_2|\mathbf{S}_1)$.
7. $E_7: (\mathbf{X}_1^n(\mathbf{q}^n, \mathbf{s}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \mathbf{s}_2^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \mathbf{s}_2^n]$ under the event F for all $m_1 \in M_1$, $m_2 \in M_2$, $(\mathbf{s}_1^n, \mathbf{s}_2^n) \neq (\tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n)$, and

$(\mathbf{s}_1^n, \mathbf{s}_2^n) \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2]$. It can be shown that $\Pr(E_7|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_1 + R_2 < \mathbb{I}(\mathbf{S}_1, \mathbf{X}_1, \mathbf{S}_2, \mathbf{X}_2; \mathbf{Y}|\mathbf{Q}) - \mathbb{H}(\mathbf{S}_1, \mathbf{S}_2)$.

8. $E_8 : (\mathbf{X}_1^n(\mathbf{q}^n, \mathbf{s}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, m_2), \mathbf{Y}^n) \in T_{\epsilon}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{X}_1, \mathbf{S}_2, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \mathbf{s}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F for all $m_1 \neq 1$, $m_2 \in M_2$, $\mathbf{s}_1^n \neq \tilde{\mathbf{s}}_1^n$, and $\mathbf{s}_1^n \in T_{\epsilon_1}^n[\mathbf{S}_1, \mathbf{S}_2|\tilde{\mathbf{s}}_2^n]$. It can be shown that $\Pr(E_8|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_1 + R_2 < \mathbb{I}(\mathbf{S}_1, \mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{S}_2, \mathbf{Q}) - \mathbb{H}(\mathbf{S}_1|\mathbf{S}_2)$.
9. $E_9 : (\mathbf{X}_1^n(\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, m_1), \mathbf{X}_2^n(\mathbf{q}^n, \tilde{\mathbf{s}}_2^n, m_2), \mathbf{Y}^n) \in T_{\epsilon}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}|\mathbf{q}^n, \tilde{\mathbf{s}}_1^n, \tilde{\mathbf{s}}_2^n]$ under the event F for all $m_1 \neq 1$, and $m_2 \neq M_2$. It can be shown that $\Pr(E_9|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_1 + R_2 < \mathbb{I}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{S}_1, \mathbf{S}_2, \mathbf{Q})$.

Then by using the union bound, $\Pr[\text{error}|F] \leq \sum_{j=1}^9 \Pr[E_j|F]$. $\Pr[\text{error}|F]$ goes to zero as $n \rightarrow \infty$ since $\Pr(E_j) \rightarrow 0$, where $j = 1$ to 9 , as $n \rightarrow \infty$ if rate pair (R_1, R_2) satisfies (5.4). It can be concluded that $P_e^n \rightarrow 0$ as $n \rightarrow \infty$ if rate pair (R_1, R_2) satisfies (5.4).

- **Average distortions:** We consider two cases in calculating the average distortion between the host sequence \mathbf{S}_i^n and the codeword \mathbf{X}_i^n for any given message m_i and $\mathbf{q}^n \in T_{\epsilon}^n[\mathbf{Q}]$. If $\mathbf{X}_i^n(\mathbf{q}^n, \mathbf{S}_i^n, m_i) \in T_{\epsilon}^n(\mathbf{X}_i|\mathbf{q}^n, \mathbf{S}_i^n)$ for any $(\mathbf{q}^n, \mathbf{S}_1^n, \mathbf{S}_2^n) \in T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]$, then the distortion between \mathbf{S}_i^n and \mathbf{X}_i^n is given by

$$\begin{aligned}
d_i(\mathbf{S}_i^n, \mathbf{X}_i^n) &= \frac{1}{n} \sum_{\mathbf{x}_i, \mathbf{s}_i} N(\mathbf{x}_i, \mathbf{s}_i|\mathbf{S}_i^n, \mathbf{X}_i^n) d_i(\mathbf{s}_i, \mathbf{x}_i), \\
&\leq \sum_{\mathbf{x}_i, \mathbf{s}_i} p(\mathbf{s}_i, \mathbf{x}_i) d_i(\mathbf{s}_i, \mathbf{x}_i) + \epsilon d_{i, \max} \\
&\leq \Delta + \epsilon d_{i, \max}
\end{aligned} \tag{C.2}$$

where $d_{i, \max}$ is the maximum distortion over the set $\mathcal{S}_i \times \mathcal{X}_i$. If $\mathbf{X}_i^n(\mathbf{q}^n, \mathbf{S}_i^n, m_i) \in T_{\epsilon}^n(\mathbf{X}_i|\mathbf{q}^n, \mathbf{s}_i^n)$ for any $(\mathbf{q}^n, \mathbf{S}_1^n, \mathbf{S}_2^n) \in T_{\epsilon_1}^n[\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2]$, the distortion $d_i(\mathbf{S}_i^n, \mathbf{X}_i^n)$ can

be upper bounded by $d_{i,max}$. From error event E_1 given F , we can show that $\Pr[\mathbf{X}_i^n(\mathbf{q}^n, \mathbf{S}_i^n, m_i) \in T_\epsilon^n(\mathbf{X}_i|\mathbf{q}^n, \mathbf{S}_i^n)]$ goes to zero as $n \rightarrow \infty$. We can then conclude that $\lim_{n \rightarrow \infty} \mathbb{E}d_i(\mathbf{S}_i^n, f^n(\mathbf{S}_i^n, \mathbf{W}_i)) \leq \Delta_i$ by letting $\epsilon \rightarrow 0$ and $n \rightarrow \infty$.

This concludes that $\mathcal{R}_{\text{MAC,C}}^i(\Delta_1, \Delta_2) \subseteq C_{\text{MAC,C}}(\Delta_1, \Delta_2)$.

C.2 Proof of Theorem 7

We prove the following lemmas which will be used in the proof of Theorem 7.

Lemma 4 *Let $(\mathbf{Q}_j, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_{1j}, \mathbf{X}_{1j}), (\mathbf{X}_{2j}, \mathbf{X}_{2j}), \mathbf{Y}_j) \in \mathcal{P}_{\text{MAC}}^i(\Delta_{1j}, \Delta_{2j})$, let $\sum_{j=1}^n \lambda_j = 1$, $\lambda_j > 0$ for $j \in \{1, 2, \dots, n\}$, and let $\Delta_i = \sum_{j=1}^n \lambda_j \Delta_{ij}$ for $i \in \{1, 2\}$. Then, there exists*

$$(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$$

such that

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(\mathbf{S}_1, \mathbf{X}_{1j}; \mathbf{Y}_j | \mathbf{X}_{2j}, \mathbf{S}_2, \mathbf{Q}_j)] = \mathbb{I}(\mathbf{S}_1, \mathbf{X}_1; \mathbf{Y} | \mathbf{X}_2, \mathbf{S}_2, \mathbf{Q}) \quad (\text{C.3a})$$

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(\mathbf{S}_2, \mathbf{X}_{2j}; \mathbf{Y}_j | \mathbf{S}_1, \mathbf{X}_{1j}, \mathbf{Q}_j)] = \mathbb{I}(\mathbf{S}_2, \mathbf{X}_2; \mathbf{Y} | \mathbf{X}_1, \mathbf{S}_1, \mathbf{Q}) \quad (\text{C.3b})$$

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(\mathbf{S}_1, \mathbf{X}_{1j}, \mathbf{S}_2, \mathbf{X}_{2j}; \mathbf{Y}_j | \mathbf{Q}_j)] = \mathbb{I}(\mathbf{S}_1, \mathbf{X}_1, \mathbf{S}_2, \mathbf{X}_2; \mathbf{Y} | \mathbf{Q}) \quad (\text{C.3c})$$

Proof: If we prove the lemma for $n = 2$, then we can easily extend it to any value of n . Let $n = 2$ and let $\lambda_1 + \lambda_2 = 1$, $\lambda_j > 0$ for $j = 1, 2$. Let β be a binary random variable such that $\Pr(Z = j) = \lambda_j$ for $j = 1, 2$. Let

$$(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) = ((\mathbf{Q}_z, Z), \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_{1z}, \mathbf{X}_{1z}), (\mathbf{X}_{2z}, \mathbf{X}_{2z}), \mathbf{Y}_z).$$

$$(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) = \begin{cases} ((\mathbf{Q}_1, 1), \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_{11}, \mathbf{X}_{11}), (\mathbf{X}_{21}, \mathbf{X}_{21}), \mathbf{Y}_1) & \text{if } Z = 1 \\ ((\mathbf{Q}_2, 2), \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_{12}, \mathbf{X}_{12}), (\mathbf{X}_{22}, \mathbf{X}_{22}), \mathbf{Y}_2) & \text{if } Z = 2 \end{cases}$$

To show that $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y}) \in \mathcal{P}_{\text{MAC}}^i(\Delta_1, \Delta_2)$, we have to check the conditions in Definition (13). We can easily show that $(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_1, \mathbf{X}_1), (\mathbf{X}_2, \mathbf{X}_2), \mathbf{Y})$ satisfies the first condition. To check the second condition, we observe that the

$X_1 \leftrightarrow (S_1, S_2, Q) \leftrightarrow X_2$ follows as consequence of

$$\mathbb{I}(X_1, X_2 | S_1, S_2, Q) = \lambda_1 \mathbb{I}(X_{11}, X_{21} | S_1, S_2, Q_1) + \lambda_2 \mathbb{I}(X_{12}, X_{22} | S_1, S_2, Q_2) = 0$$

Similarly, $X_1 \leftrightarrow (S_1, Q) \leftrightarrow S_2$ and $S_1 \leftrightarrow (S_2, Q) \leftrightarrow X_2$. We can easily verify that

$$\mathbb{E}d_i(S_i, X_i) < \lambda_1 \Delta_{i1} + \lambda_2 \Delta_{i2}, \text{ for } i = 1, 2 \text{ using the distribution on}$$

$(Q, S_1, S_2, (X_1, X_1), (X_2, X_2), Y)$. Since the distribution on $(Q, S_1, S_2, (X_1, X_1), (X_2, X_2), Y)$

satisfies the conditions in Definition (13), we can conclude that

$$(Q, S_1, S_2, (X_1, X_1), (X_2, X_2), Y) \in \mathcal{P}_{MAC}^i(\Delta_1, \Delta_2).$$

We can easily derive the equations (C.3) using the distribution on

$(Q, S_1, S_2, (X_1, X_1), (X_2, X_2), Y)$. This completes the proof of Lemma.

Lemma 5 *Let $(Q_j, S_1, S_2, X_{1j}, X_{2j}, Y_j) \in \mathcal{P}_{MAC}^o(\Delta_{1j}, \Delta_{2j})$, let $\sum_{j=1}^n \lambda_j = 1$, $\lambda_j > 0$ for $j \in \{1, 2, \dots, n\}$, and let $\Delta_i = \sum_{j=1}^n \lambda_j \Delta_{ij}$ for $i \in \{1, 2\}$. Then, there exists*

$$(Q, S_1, S_2, X_1, X_2, Y) \in \mathcal{P}_{MAC}^o(\Delta_1, \Delta_2)$$

such that

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(X_{1j}, S_1; Y_j | X_{2j}, S_2, Q_j)] = \mathbb{I}(X_1, S_1; Y | X_2, S_2, Q) \quad (\text{C.4a})$$

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(X_{2j}, S_2; Y_j | X_{1j}, S_1, Q_j)] = \mathbb{I}(X_2, S_2; Y | X_1, S_1, Q) \quad (\text{C.4b})$$

$$\sum_{j=1}^n \lambda_j [\mathbb{I}(X_{1j}, S_1, X_{2j}, S_2; Y_j | Q_j)] = [\mathbb{I}(X_1, S_1, X_2, S_2; Y | Q)] \quad (\text{C.4c})$$

Proof: We do not prove the lemma because proof is similar to the proof of Lemma 4.

Lemma 6 $\mathcal{R}_{MAC,C}^i(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{MAC,C}^i(\Delta'_1, \Delta'_2)$ and $\mathcal{R}_{MAC,C}^o(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{MAC,C}^o(\Delta'_1, \Delta'_2)$ for any $\Delta_1 \leq \Delta'_1$ and $\Delta_2 \leq \Delta'_2$.

Proof: This lemma can be directly proved from the fact that $\mathcal{P}_{MAC}^i(\Delta_1, \Delta_2) \subseteq \mathcal{P}_{MAC}^i(\Delta'_1, \Delta'_2)$ and $\mathcal{P}_{MAC}^o(\Delta_1, \Delta_2) \subseteq \mathcal{P}_{MAC}^o(\Delta'_1, \Delta'_2)$.

We are now ready to prove the Theorem 7, i.e., prove that for any sequence of MAC IE codes $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D_1^{(n)}, D_2^{(n)}, n)$ with $\lim_{n \rightarrow \infty} P_e^n = 0$ and $\lim_{n \rightarrow \infty} D_i^{(n)} \leq \Delta_i$, for $i = 1, 2$, the rates must satisfy (5.6).

Consider a given code of block length n . The joint distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}_1^n \times \mathcal{S}_2^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$ is given by

$$p(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}_1^n, \mathbf{s}_2^n, \mathbf{x}_1^n, \mathbf{x}_2^n, \mathbf{y}^n) = \frac{1}{2^{nR_1}} \frac{1}{2^{nR_2}} \left(\prod_{j=1}^n p(s_{1j}, s_{2j}) \right) p(\mathbf{x}_1^n | \mathbf{w}_1, \mathbf{s}_1^n) p(\mathbf{x}_2^n | \mathbf{w}_2, \mathbf{s}_2^n) \prod_{i=1}^n p(\mathbf{y}_j | \mathbf{x}_{1j}, \mathbf{x}_{2j}, s_{1j}, s_{2j}),$$

where, $p(\mathbf{x}_i^n | \mathbf{w}_i, \mathbf{s}_i^n)$ is 1 if $\mathbf{x}_i^n = f_i^n(\mathbf{w}_i, \mathbf{s}_i^n)$ and 0 otherwise, for $i = 1, 2$. By Fano's inequality [12], the conditional entropy of $(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_1^n, \mathbf{S}_2^n)$ given \mathbf{Y}^n is bounded as

$$\mathbb{H}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}_1^n, \mathbf{S}_2^n | \mathbf{Y}^n) \leq n(R_1 + R_2 + \log_2(|\mathcal{S}_1| |\mathcal{S}_2|)) P_e^n + 1 \triangleq n\epsilon_n, \quad (\text{C.5})$$

for $i = 1, 2$, where $\epsilon_n \rightarrow 0$ as $P_e^n \rightarrow 0$. We can now bound the rate R_1 as

$$\begin{aligned} nR_1 &\leq \mathbb{H}(\mathbf{W}_1) = \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2) \\ &\stackrel{(a)}{=} \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) \\ &= \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n, \mathbf{Y}^n) \\ &\quad + \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n, \mathbf{Y}^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) \\ &\stackrel{(b)}{\leq} \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{S}_2^n, \mathbf{Y}^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) + n\epsilon_n \\ &\stackrel{(c)}{=} \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n | \mathbf{Y}^n, \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) + n\epsilon_n \\ &= \mathbb{H}(\mathbf{W}_1, \mathbf{S}_1^n; \mathbf{Y}^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) + n\epsilon_n \\ &= \mathbb{H}(\mathbf{Y}^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{Y}^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n, \mathbf{W}_1, \mathbf{S}_1^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) + n\epsilon_n \\ &\stackrel{(d)}{=} \mathbb{H}(\mathbf{Y}^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n) - \mathbb{H}(\mathbf{Y}^n | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n, \mathbf{W}_1, \mathbf{S}_1^n, \mathbf{X}_1^n) - \mathbb{H}(\mathbf{S}_1^n | \mathbf{S}_2^n) + n\epsilon_n \\ &\stackrel{(e)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n, \mathbf{Y}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n, \mathbf{W}_1, \mathbf{S}_1^n, \mathbf{X}_1^n, \mathbf{Y}^{j-1}) \\ &\quad - \mathbb{H}(\mathbf{S}_{1j} | \mathbf{S}_2^n, \mathbf{S}_1^{j-1})] + n\epsilon_n \\ &\stackrel{(f)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{X}_2^n, \mathbf{S}_2^n, \mathbf{Y}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{X}_{1j}, \mathbf{S}_{1j}, \mathbf{X}_{2j}, \mathbf{S}_{2j}) \\ &\quad - \mathbb{H}(\mathbf{S}_{1j} | \mathbf{S}_{2j})] + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&\stackrel{(g)}{\leq} \sum_{j=1}^n [\mathbb{H}(Y_j|X_{2j}, S_{2j}) - \mathbb{H}(Y_j|X_{1j}, S_{1j}, X_{2j}, S_{2j}) - \mathbb{H}(S_{1j}|S_{2j})] + n\epsilon_n \\
&= \sum_{j=1}^n [\mathbb{I}(X_{1j}, S_{1j}; Y_j|X_{2j}, S_{2j}) - \mathbb{H}(S_{1j}|S_{2j})] + n\epsilon_n,
\end{aligned}$$

where:

(a) follows from the fact that W_1 is independent of each other; and (W_1, W_2) is independent of (S_1^n, S_2^n) .

(b) follows from Fano's inequality,

(c) follows from the fact that X_2^n is a function of (W_1, S_1^n) ,

(d) follows from the fact that X_1^n is a function of (W_1, S_1^n) ,

(e) follows from the chain rule of mutual information and entropy,

(f) follows from the fact that Y_j depends only on X_{1j}, X_{2j}, S_{1j} , and S_{2j} by the memoryless property of the channel and $S_{1j} \leftrightarrow S_{2j} \leftrightarrow (S_1^{j-1}, S_2^{j-1}, S_{2,j+1}^n)$,

(g) follows from removing conditioning.

Hence, we have

$$R_1 \leq \frac{1}{n} \sum_{j=1}^n [\mathbb{I}(X_{1j}, S_{1j}; Y_j|X_{2j}, S_{2j})] - \mathbb{H}(S_1|S_2) + \epsilon_n$$

Similarly, we can bound R_2 and $R_1 + R_2$ as

$$\begin{aligned}
R_2 &\leq \frac{1}{n} \sum_{j=1}^n [\mathbb{I}(X_{2j}, S_{2j}; Y_j|X_{1j}, S_{1j})] - \mathbb{H}(S_1|S_2) + \epsilon_n, \\
R_1 + R_2 &\leq \frac{1}{n} \sum_{j=1}^n [\mathbb{I}(X_{1j}, S_{1j}, X_{2j}, S_{2j}; Y_j)] - \mathbb{H}(S_1|S_2) + \epsilon_n.
\end{aligned}$$

If the host random variables S_1 and S_2 are correlated, we can clearly see that the random vector $(Q_j, S_1, S_2, X_{1j}, X_{2j}, Y_j)$ with $p(q_j = j) = 1$ belongs to set $\mathcal{P}_{\text{MAC}}^o(\mathbb{E}[d_1(S_{1j}, X_{1j})], \mathbb{E}[d_2(S_{2j}, X_{1j})])$ for $j \in \{1, 2, \dots, n\}$. According to Lemma 5, there exists a random vector

$$(Q, S_1, S_2, \tilde{X}_1, \tilde{X}_2, \tilde{Y}) \in \mathcal{P}_{\text{MAC}}^o\left(\frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_1(S_{1j}, X_{1j})], \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_2(S_{1j}, X_{1j})]\right)$$

such that the following is true

$$\begin{aligned}\frac{1}{n} \sum_{j=1}^n [\mathbb{I}(\mathbf{X}_{1j}, \mathbf{S}_1; \mathbf{Y}_j | \mathbf{X}_{2j}, \mathbf{S}_2)] &= \mathbb{I}(\tilde{\mathbf{X}}_1, \mathbf{S}_1; \tilde{\mathbf{Y}} | \tilde{\mathbf{X}}_2, \mathbf{S}_2, \mathbf{Q}) \\ \frac{1}{n} \sum_{j=1}^n [\mathbb{I}(\mathbf{X}_{2j}, \mathbf{S}_2; \mathbf{Y}_j | \mathbf{X}_{1j}, \mathbf{S}_1)] &= \mathbb{I}(\tilde{\mathbf{X}}_2, \mathbf{S}_2; \tilde{\mathbf{Y}} | \tilde{\mathbf{X}}_1, \mathbf{S}_1, \mathbf{Q}) \\ \frac{1}{n} \sum_{j=1}^n [\mathbb{I}(\mathbf{X}_{1j}, \mathbf{S}_{1j}, \mathbf{X}_{2j}, \mathbf{S}_{2j}; \mathbf{Y}_j)] &= \mathbb{I}(\tilde{\mathbf{X}}_1, \mathbf{S}_1, \tilde{\mathbf{X}}_2, \mathbf{S}_2; \tilde{\mathbf{Y}} | \mathbf{Q})\end{aligned}$$

As $n \rightarrow \infty$, we can conclude the following

$$\begin{aligned}\mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2) &\subseteq \mathcal{R}_{\text{MAC,C}}^{\circ} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_1(\mathbf{S}_{1j}, \mathbf{X}_{1j})], \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_2(\mathbf{S}_{1j}, \mathbf{X}_{1j})] \right) \\ &\stackrel{(a)}{\subseteq} \mathcal{R}_{\text{MAC,C}}^{\circ}(\Delta_1, \Delta_2)\end{aligned}\tag{C.8}$$

where (a) follows from the Lemma 6.

If the host random variables \mathbf{S}_1 and \mathbf{S}_1 are independent, we can obtain the following from the condition that the messages \mathbf{W}_1 and \mathbf{W}_2 are independent.

$$p(\mathbf{x}_{1j}, \mathbf{x}_{2j} | \mathbf{s}_{1j}, \mathbf{s}_{2j}) = p(\mathbf{x}_{1j} | \mathbf{s}_{1j}) p(\mathbf{x}_{2j} | \mathbf{s}_{2j}).$$

Then we can clearly see that the random variable tuple

$$(\mathbf{Q}_j, \mathbf{S}_1, \mathbf{S}_2, (\mathbf{X}_{1j}, \mathbf{X}_{1j}), (\mathbf{X}_{2j}, \mathbf{X}_{2j}), \mathbf{Y}_j)$$

with $p(\mathbf{q}_j = j) = 1$ belongs to set

$$\mathcal{P}_{\text{MAC}}^i(\mathbb{E}[d_1(\mathbf{S}_{1j}, \mathbf{X}_{1j})], \mathbb{E}[d_2(\mathbf{S}_{2j}, \mathbf{X}_{1j})])$$

for $j \in \{1, 2, \dots, n\}$. According to Lemma 4, there exists a random vector

$$(\mathbf{Q}, \mathbf{S}_1, \mathbf{S}_2, (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_1), (\tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_2), \tilde{\mathbf{Y}}) \in \mathcal{P}_{\text{MAC}}^i\left(\frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_1(\mathbf{S}_{1j}, \mathbf{X}_{1j})], \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_2(\mathbf{S}_{1j}, \mathbf{X}_{1j})]\right)$$

such that (C.7) is true. As $n \rightarrow \infty$, we can conclude the following

$$\mathcal{C}_{\text{MAC,C}}(\Delta_1, \Delta_2) \subseteq \mathcal{R}_{\text{MAC,C}}^i \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_1(\mathbf{S}_{1j}, \mathbf{X}_{1j})], \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[d_2(\mathbf{S}_{1j}, \mathbf{X}_{1j})] \right)$$

$$\stackrel{(a)}{\subseteq} \mathcal{R}_{\text{MAC,C}}^i(\Delta_1, \Delta_2) \tag{C.9}$$

where (a) follows from the Lemma 6. This completes the proof of Theorem 7.

APPENDIX D

D.1 Proof of Theorem 8

In this section, we show that $\mathcal{R}_B^i(\Delta) \subseteq \mathcal{C}_B(\Delta)$. Fix the random vector

$$(\mathbf{U}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$$

. For each n , we construct a $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D^{(n)}, n)$ broadcast IE code as follows.

- **Code construction** : Generate $\lceil 2^{nR_2} \rceil 2^{n(\mathbb{I}(\mathbf{U};\mathbf{S})+\epsilon)}$ \mathbf{U}^n sequences drawn according to $\prod_{j=1}^n p(\mathbf{u}_j)$. Distribute these sequences randomly into $\lceil 2^{nR_2} \rceil$ bins such that each bin has $2^{n(\mathbb{I}(\mathbf{U};\mathbf{S})+\epsilon)}$ sequences. Label all sequences \mathbf{U}^n in bin $m_2 \in \{1, 2, \dots, \lceil 2^{nR_2} \rceil\}$ as $\mathbf{U}_1^n(m_2)$. For each $(\mathbf{S}^n, \mathbf{U}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}]$, generate $\lceil 2^{nR_1} \rceil$ \mathbf{X}^n sequences according to $\prod_{j=1}^n p(x_j|u_j, s_j)$. Label these sequences as $\mathbf{X}^n(\mathbf{S}^n, \mathbf{U}^n, m_1)$, where $(\mathbf{S}^n, \mathbf{U}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}]$ and $m_1 \in \{1, 2, \dots, \lceil 2^{nR_1} \rceil\}$. These codebooks are revealed to the encoder and both the decoders.
- **Encoder** : The encoder, upon observing $\mathbf{S}^n \in T_\epsilon^n[\mathbf{S}]$ at the output of the host source, embeds message $\mathbf{W}_2 \in \{1, 2, \dots, \lceil 2^{nR_2} \rceil\}$ into the host sequence by looking for a \mathbf{U}^n in bin \mathbf{W}_2 such that $\mathbf{U}^n(\mathbf{W}_2) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}|\mathbf{S}^n]$. If such a sequence $\mathbf{U}^n(\mathbf{W}_2)$ does not exist, the encoder declares an error; otherwise, the encoder embeds message $\mathbf{W}_1 \in \{1, 2, \dots, \lceil 2^{nR_1} \rceil\}$ into the host sequence \mathbf{S}^n by choosing the codeword $\mathbf{X}^n(\mathbf{S}^n, \mathbf{U}^n(\mathbf{W}_2), \mathbf{W}_1)$.
- **Decoder 1**: Decoder 1, upon receiving \mathbf{Y}^n , which is a distorted or attacked

version of the embedded sequence X^n , looks for $\mathbf{U}^n(m_2)$, $m_2 \in \{1, 2, \dots, \lceil 2^{nR_2} \rceil\}$ such that $(\mathbf{U}^n(m_2), Y^n) \in T_\epsilon^n[\mathbf{U}, Y]$. If a unique codeword $\mathbf{U}^n(m_2)$ does not exist, Decoder 1 declares an error; otherwise, Decoder 1 declares that $\hat{W}_2 = m_2$. Upon decoding the sequence $\mathbf{U}^n(\hat{W}_2)$, Decoder 1 looks for $X^n(s^n, \mathbf{U}^n(\hat{W}_2), m_1)$ such that $(X^n(s^n, \mathbf{U}^n(\hat{W}_2), m_1), Y^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, X, Y | s^n, \mathbf{U}^n(\hat{W}_2)]$ for each $s^n \in T_\epsilon^n[\mathbf{U}, \mathbf{S} | \mathbf{U}^n(\hat{W}_2)]$ and $m_1 \in \{1, 2, \dots, \lceil 2^{nR_1} \rceil\}$. If a unique codeword $X^n(s^n, \mathbf{U}^n(\hat{W}_2), m_1)$ exists, Decoder 1 declares that $(\hat{W}_1, \hat{S}_2^n) = (m_1, s^n)$; otherwise, it declares an error.

- **Decoder 2:** Decoder 2, up on receiving Z^n , which is a degraded version of Y^n , looks for $\mathbf{U}^n(m_2)$, $m_2 \in \{1, 2, \dots, \lceil 2^{nR_2} \rceil\}$ such that $(\mathbf{U}^n(m_2), Z^n) \in T_\epsilon^n[\mathbf{U}, Z]$. If a unique codeword $\mathbf{U}^n(m_2)$ exists, Decoder 2 declares that $\hat{W}_2 = m_2$; otherwise, Decoder 2 declares an error.

- **Probability of error:** The average probability of error is given by

$$\begin{aligned} P_e^n &= \sum_{s^n \in \mathcal{S}^n} p(s^n) \Pr[\text{error} | s^n] \\ &\leq \sum_{s^n \notin T_\epsilon^n[\mathbf{S}]} p(s^n) + \sum_{s^n \in T_\epsilon^n[\mathbf{S}]} p(s^n) \Pr[\text{error} | s^n], \end{aligned} \quad (\text{D.1})$$

where the first term, $\Pr[s^n \notin T_\epsilon^n[\mathbf{S}]]$, goes to zero as $n \rightarrow \infty$ by the strong asymptotic equipartition property (AEP). Without loss of generality, it can be assumed that the output of the host source is \tilde{s}^n , and the message pair $(W_1, W_2) = (1, 1)$ is to be embedded in to the host sequence \tilde{s}^n . Let F be the event that the host source output is \tilde{s}^n . To compute $\Pr[\text{error} | F]$, let us write the error event as $E_0 \cup E_1 \cup E_2 \cup E_3$, where:

1. E_0 is the event that there is no $\mathbf{U}^n(1)$ such that $\mathbf{U}^n(1) \in T_\epsilon^n[\mathbf{U}, \mathbf{S} | \tilde{s}^n]$.

Using well-known rate-distortion arguments, the probability of this event

approaches zero as n goes to infinity since each bin has $2^{n(\mathbb{I}(\mathbf{U};\mathbf{S})+\epsilon)}$ \mathbf{U}^n sequences.

Conditioned on the event $F \cap E_0^c$, it can also be assumed that $\tilde{\mathbf{U}}^n(1)$ is jointly strongly typical with the host sequence \tilde{s}^n . Hence, the embedded sequence $\mathbf{X}^n(\tilde{s}^n, \tilde{\mathbf{U}}^n(1), 1)$ is generated and transmitted from the encoder.

2. E_1 is the event that

$$(\tilde{\mathbf{U}}^n(1), \mathbf{X}^n(\tilde{s}^n, \tilde{\mathbf{U}}^n(1), 1), \mathbf{Y}^n, \mathbf{Z}^n) \notin T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}|\tilde{s}^n].$$

By the strong AEP, we can show that $\Pr[E_1|F \cap E_0^c] \rightarrow 0$ as $n \rightarrow \infty$.

3. $E_2 := E_{2,1} \cup (E_{2,1}^c \cap E_{2,2})$, where $E_{2,1}$ is the event that $(\mathbf{U}^n, \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{U}, \mathbf{Y}]$ for $\mathbf{U}^n \neq \tilde{\mathbf{U}}^n(1)$, and $E_{2,2}$ is the event that $(\mathbf{X}^n(\mathbf{s}^n, \tilde{\mathbf{U}}^n(1), m_1), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y}|\mathbf{S}^n, \tilde{\mathbf{U}}^n(1)]$ for $m_1 \neq 1$ or

$$\mathbf{s}^n \in \{\mathbf{s}^n : \mathbf{s}^n \neq \tilde{s}^n, \mathbf{s}^n \in T_\epsilon^n[\mathbf{U}, \mathbf{S}|\tilde{\mathbf{U}}^n(1)]\}.$$

It can be shown that $\Pr[E_{2,1}|F \cap E_0^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Y}) - \mathbb{I}(\mathbf{U}; \mathbf{S})$ and that $\Pr(E_{2,2}|F \cap E_0^c \cap E_{2,1}^c) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 \leq \mathbb{I}(\mathbf{S}, \mathbf{X}; \mathbf{Y}|\mathbf{U}) - \mathbb{H}(\mathbf{S}|\mathbf{U})$.

4. E_3 is the event that $(\mathbf{U}^n, \mathbf{Z}^n) \in T_\epsilon^n[\mathbf{U}, \mathbf{Z}]$ for $\mathbf{U}^n \neq \tilde{\mathbf{U}}^n(1)$. Using Gel'fand-Pinsker arguments, it can be shown that $\Pr[E_3|F \cap E_0^c] \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 \leq \mathbb{I}(\mathbf{U}; \mathbf{Z}) - \mathbb{I}(\mathbf{U}; \mathbf{S})$. Because the broadcast channel is degraded, this constraint on R_2 is more restrictive than the previous constraint.

Thus, by the union bound, it can be shown that P_e^n goes to zero as $n \rightarrow \infty$ if $(R_1, R_2) \in \mathcal{R}_B^i$.

- **Average distortion:** Since $(\mathbf{X}^n, \tilde{s}^n)$ is jointly strongly typical with high probability and the distribution belongs to $\mathcal{P}(\Delta)$, it can be shown that the average

distortion $D^{(n)}$ associated with the generated code satisfies the distortion constraint Δ as $n \rightarrow \infty$ as in the Proof of Theorem 6.

D.2 Proof of Theorem 9

In this section, we show that $\mathcal{C}_B(\Delta) \subseteq \mathcal{R}_B^o(\Delta)$. If we are given a sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D^{(n)}, n)$ broadcast IE codes, i.e., $\mathbf{X}^n = f(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n)$, $g_1^n(\mathbf{Y}^n) = (\hat{\mathbf{W}}_1, \hat{\mathbf{W}}_2, \hat{\mathbf{S}}^n)$, and $g_2^n(\mathbf{Z}^n) = \hat{\mathbf{W}}_2$, with $\lim_{n \rightarrow \infty} P_e^n = 0$ and $\lim_{n \rightarrow \infty} D^{(n)} \leq \Delta$, then we show that the rate pair (R_1, R_2) must satisfy (6.5) for some $((\mathbf{U}, \mathbf{V}), \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$. Consider a given code of block length n . The joint distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ induced by the code is given by

$$\begin{aligned} p(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n, \mathbf{x}^n, \mathbf{y}^n, \mathbf{z}^n) &= \\ &= \frac{1}{\lceil 2^{nR_1} \rceil \lceil 2^{nR_2} \rceil} p(\mathbf{s}^n) p(\mathbf{x}^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n) \\ &\quad \times \prod_{i=1}^n p(\mathbf{y}_i | \mathbf{x}_i, \mathbf{s}_i) p(\mathbf{z}_i | \mathbf{y}_i), \end{aligned}$$

where, $p(\mathbf{x}^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ is 1 if $\mathbf{x}^n = f^n(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ and 0 otherwise. We can bound the rate R_1 as follows:

$$\begin{aligned} nR_1 &\leq \mathbb{H}(\mathbf{W}_1) \\ &\stackrel{(a)}{=} \mathbb{H}(\mathbf{W}_1, \mathbf{S}^n | \mathbf{W}_2) - \mathbb{H}(\mathbf{S}^n | \mathbf{W}_2) \\ &= \mathbb{H}(\mathbf{W}_1, \mathbf{S}^n | \mathbf{W}_2) - \mathbb{H}(\mathbf{W}_1, \mathbf{S}^n | \mathbf{W}_2, \mathbf{Y}^n) \\ &\quad + \mathbb{H}(\mathbf{W}_1, \mathbf{S}^n | \mathbf{W}_2, \mathbf{Y}^n) - \mathbb{H}(\mathbf{S}^n | \mathbf{W}_2) \\ &\stackrel{(b)}{\leq} \mathbb{I}(\mathbf{W}_1, \mathbf{S}^n; \mathbf{Y}^n | \mathbf{W}_2) - \mathbb{H}(\mathbf{S}^n | \mathbf{W}_2) + n\epsilon_n \\ &\stackrel{(c)}{=} \sum_{j=1}^n [\mathbb{I}(\mathbf{W}_1, \mathbf{S}^n; \mathbf{Y}_j | \mathbf{W}_2, \mathbf{Y}^{j-1}) - \mathbb{H}(\mathbf{S}_j | \mathbf{W}_2)] + n\epsilon_n \\ &\stackrel{(d)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{Y}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{Y}^{j-1}, \mathbf{W}_1, \mathbf{S}^n, \mathbf{X}^n) \\ &\quad - \mathbb{H}(\mathbf{S}_j | \mathbf{W}_2)] + n\epsilon_n \end{aligned}$$

$$\begin{aligned}
&\stackrel{(e)}{=} \sum_{j=1}^n [\mathbb{H}(Y_j | \mathbf{W}_2, Y^{j-1}, Z^{j-1}) - \mathbb{H}(Y_j | \mathbf{S}_j, \mathbf{X}_j) \\
&\quad - \mathbb{H}(\mathbf{S}_j | \mathbf{W}_2)] + n\epsilon_n \\
&\stackrel{(f)}{\leq} \sum_{j=1}^n [\mathbb{H}(Y_j | \mathbf{W}_2, Z^{j-1}) - \mathbb{H}(Y_j | \mathbf{S}_j, \mathbf{X}_j, \mathbf{W}_2, Z^{j-1}) \\
&\quad - \mathbb{H}(\mathbf{S}_j | \mathbf{W}_2, Z^{j-1})] + n\epsilon_n \\
&= \sum_{j=1}^n \mathbb{I}(\mathbf{S}_j, \mathbf{X}_j; Y_j | \mathbf{W}_2, Z^{j-1}) - \mathbb{H}(\mathbf{S}_j | \mathbf{W}_2, Z^{j-1}) + n\epsilon_n \tag{D.2}
\end{aligned}$$

where, $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, and

- (a) follows from the fact that \mathbf{W}_1 , \mathbf{W}_2 and \mathbf{S}^n are mutually independent,
- (b) follows from Fano's inequality,
- (c) follows from the chain rule and the fact that \mathbf{S}^n is i.i.d. and independent of \mathbf{W}_2 ,
- (d) follows from the fact that \mathbf{X}^n is a deterministic function of $(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n)$,
- (e) follows from degraded and memoryless properties of the broadcast channel, and
- (f) follows from removing conditioning in the positive term and introducing conditioning in the negative term.

We can also bound the rate R_2 as follows:

$$\begin{aligned}
nR_2 &\leq \mathbb{H}(\mathbf{W}_2) \\
&\stackrel{(a)}{\leq} \mathbb{I}(\mathbf{W}_2; \mathbf{Z}^n) + n\epsilon_n \\
&= \sum_{j=1}^n [\mathbb{I}(\mathbf{W}_2, \mathbf{S}_{j+1}^n; \mathbf{Z}^j) - \mathbb{I}(\mathbf{W}_2, \mathbf{S}_j^n; \mathbf{Z}^{j-1})] + n\epsilon_n \\
&\stackrel{(b)}{\leq} \sum_{j=1}^n [\mathbb{I}(\mathbf{W}_2, \mathbf{S}_{j+1}^n; \mathbf{Z}^{j-1}) + \mathbb{I}(\mathbf{W}_2, \mathbf{S}_{j+1}^n; \mathbf{Z}_j | \mathbf{Z}^{j-1}) \\
&\quad - \mathbb{I}(\mathbf{W}_2, \mathbf{S}_{j+1}^n; \mathbf{Z}^{j-1}) - \mathbb{I}(\mathbf{S}_j; \mathbf{Z}^{j-1} | \mathbf{W}_2, \mathbf{S}_{j+1}^n)] + n\epsilon_n \\
&= \sum_{j=1}^n [\mathbb{I}(\mathbf{W}_2, \mathbf{S}_{j+1}^n; \mathbf{Z}_j | \mathbf{Z}^{j-1}) - \mathbb{I}(\mathbf{S}_j; \mathbf{Z}^{j-1} | \mathbf{W}_2, \mathbf{S}_{j+1}^n)] + n\epsilon_n
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n [\mathbb{H}(Z_j|Z^{j-1}) - \mathbb{H}(Z_j|W_2, Z^{j-1}, S_{j+1}^n) \\
&\quad - \mathbb{H}(S_j|W_2, S_{j+1}^n) + \mathbb{H}(S_j|W_2, Z^{j-1}, S_{j+1}^n)] + n\epsilon_n \\
&\stackrel{(c)}{\leq} \sum_{j=1}^n [\mathbb{H}(Z_j) - \mathbb{H}(Z_j|W_2, Z^{j-1}, S_{j+1}^n) \\
&\quad - \mathbb{H}(S_j) + \mathbb{H}(S_j|W_2, Z^{j-1}, S_{j+1}^n)] + n\epsilon_n \\
&= \sum_{j=1}^n [\mathbb{I}(W_2, Z^{j-1}, S_{j+1}^n; Z_j) - \mathbb{I}(W_2, Z^{j-1}, S_{j+1}^n; S_j)] + n\epsilon_n \quad (\text{D.3})
\end{aligned}$$

where, $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, and

(a) follows from Fano's inequality,

(b) follows from applying the chain rule on (Z^{j-1}, Z_j) and (S_{j+1}^n, S_j) in the first and second mutual information expressions, respectively, and

(c) follows from removing conditioning and the fact that S^n is i.i.d. and independent of W_2 .

Let $\tilde{U}_j := \{W_2, Z^{j-1}\}$ and $V_j := \{S_{j+1}^n\}$ for $j = 1, 2, \dots, n$. We can then write (D.2) and (D.3) as

$$R_1 \leq \mathbb{I}(S, X; Y|Q, \tilde{U}) - \mathbb{H}(S|Q, \tilde{U}) + \epsilon_n, \quad (\text{D.4a})$$

$$R_2 \leq \mathbb{I}(\tilde{U}, V; Z|Q) - \mathbb{I}(\tilde{U}, V; S|Q) + \epsilon_n, \quad (\text{D.4b})$$

where Q takes values in the set $Q \in \{1, 2, \dots, n\}$ with equal probability and the joint probability distribution on $(S, Q, \tilde{U}, V, X, Y, Z)$ is $p(S = s, Q = q, \tilde{U} = \tilde{u}, V = v, X = x)p(y|x, s)p(z|y)$, with

$$\begin{aligned}
p(S = s, Q = q, \tilde{U} = \tilde{u}, V = v, X = x) &= \\
&= p(s)p(q)p(\mathbf{U}_q = \tilde{u}, V_q = v|s, q)p(X_q = x|s, q, \tilde{u}, v).
\end{aligned}$$

Finally, we can write (D.4) as

$$R_1 \leq \mathbb{I}(S, X; Y|U) - \mathbb{H}(S|U) + n\epsilon_n,$$

$$R_2 \leq \mathbb{I}(\mathbf{U}, \mathbf{V}; \mathbf{Z}) - \mathbb{I}(\mathbf{U}, \mathbf{V}; \mathbf{S}) + n\epsilon_n,$$

where $\mathbf{U} := (\mathbf{Q}, \tilde{\mathbf{U}})$, since $\mathbb{I}(\tilde{\mathbf{U}}, \mathbf{V}; \mathbf{Z}|\mathbf{Q}) \leq \mathbb{I}(\mathbf{Q}, \tilde{\mathbf{U}}, \mathbf{V}; \mathbf{Z})$ and $\mathbb{I}(\mathbf{Q}; \mathbf{S}) = 0$.

Given any $\delta > 0$, the associated distortion $D^{(n)}$, for sufficiently large n , satisfies

$$\begin{aligned} \Delta + \delta &\geq D^{(n)} \\ &= \mathbb{E}d(\mathbf{X}^n, \mathbf{S}^n) \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{\mathbf{x}, \mathbf{s}} p(\mathbf{X}_j = \mathbf{x}, \mathbf{S}_j = \mathbf{s}) d(\mathbf{x}, \mathbf{s}) \\ &= \sum_{\mathbf{x}, \mathbf{s}} p(\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}) d(\mathbf{x}, \mathbf{s}) \\ &= \mathbb{E}d(\mathbf{X}, \mathbf{S}). \end{aligned}$$

As $n \rightarrow \infty$ and $\delta \rightarrow 0$, $((\mathbf{U}, \mathbf{V}), \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$ and $(R_1, R_2) \in \mathcal{R}_B^o$. Thus, $\mathcal{C}_B(\Delta) \subseteq \mathcal{R}_B^o$.

D.3 Proof of Theorem 10

Achievability

In this section, we show that $\mathcal{R}_C^i(\Delta) \subseteq \mathcal{C}_C(\Delta)$. Fix the random vector

$$(\mathbf{U}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$$

. For each n , we construct a $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D^{(n)}, n)$ broadcast IE code as follows.

- **Code construction:** At the encoder, for each $\mathbf{s}^n \in \mathcal{S}^n$, generate 2^{nR_2} \mathbf{U}^n sequences drawn according to $\prod_{j=1}^n p(\mathbf{u}_j | \mathbf{s}_j)$. Denote these sequences as $\mathbf{U}^n(\mathbf{s}^n, m_2)$, where $m_2 \in \{1, 2, \dots, 2^{nR_2}\}$. For each pair $(\mathbf{s}^n, \mathbf{U}^n)$, generate 2^{nR_1} \mathbf{X}_1^n sequences drawn according to $\prod_{j=1}^n p(\mathbf{x}_j | \mathbf{u}_j, \mathbf{s}_j)$. Call these sequences $\mathbf{X}^n(\mathbf{S}^n, m_1, m_2)$ where $m_1 \in \{1, 2, \dots, 2^{nR_1}\}$. In this way, the codebook is generated at the encoder and revealed to both the decoders.

- **Encoding:** The encoder, upon observing \mathbf{s}^n at the output of host source, sends messages $W_1 \in \{1, 2, \dots, 2^{nR_1}\}$ and $W_2 \in \{1, 2, \dots, 2^{nR_2}\}$ by transmitting the codeword $X^n(\mathbf{s}^n, W_1, W_2)$. In this way, the codeword X^n is chosen and transmitted from the encoder for a given host sequence \mathbf{S}^n , and a given message pair (W_1, W_2) .
- **Decoding at Decoder 1:** Decoder 1, up on receiving the channel output Y^n , looks for $\mathbf{U}^n(\mathbf{s}^n, m_2)$ such that $(\mathbf{U}^n(\mathbf{s}_1^n, m_2), Y^n) \in T_\epsilon^n[\mathbf{U}, Y|\mathbf{s}^n]$ for all $\mathbf{s}^n \in T_{\epsilon_1}^n[\mathbf{S}]$. If a unique codeword $\mathbf{U}^n(\mathbf{s}^n, m_2)$ exists, Decoder 1 again looks for $X^n(\mathbf{s}^n, m_1, m_2)$ such that $(X^n(\mathbf{s}^n, m_1, m_2), Y^n) \in T_\epsilon^n[X, Y|\mathbf{s}^n, \mathbf{U}^n(\mathbf{s}^n, m_2)]$. If a unique codeword $X^n(\mathbf{s}^n, m_1, m_2)$ exists, Decoder 1 declares that $(\hat{W}_1, \hat{S}_2^n) = (m_1, \mathbf{s}^n)$. In this way, the message intended for Decoder 1 and the host sequences are decoded at Decoder 1.
- **Decoding at Decoder 2:** Decoder 2, up on receiving the channel output Z^n , looks for $\mathbf{U}^n(\mathbf{s}^n, m_2)$ such that $(\mathbf{U}^n(\mathbf{s}_1^n, m_2), Z^n) \in T_\epsilon^n[\mathbf{U}, Z|\mathbf{s}^n]$ for all $\mathbf{s}^n \in T_{\epsilon_1}^n[\mathbf{S}]$. If a unique codeword $\mathbf{U}^n(\mathbf{s}^n, m_2)$ codeword exists, Decoder 2 declares that $(\hat{W}_2, \hat{S}_1^n) = (m_2, \mathbf{s}^n)$. Otherwise, Decoder 2 declares an error. In this way, the message intended for Decoder 2 and the host sequences are decoded at Decoder 2.
- **Probability of error:** The average probability of error is given by the following

$$\begin{aligned}
P_e^n &= \sum_{(\mathbf{s}^n) \in \mathcal{S}^n} p(\mathbf{s}^n) \Pr[\text{error}|\mathbf{s}^n] \\
&\leq \sum_{\mathbf{s}^n \notin T_{\epsilon_1}^n[\mathbf{S}]} p(\mathbf{s}^n) + \sum_{\mathbf{s}^n \in T_{\epsilon_1}^n[\mathbf{S}]} p(\mathbf{s}^n) \Pr[\text{error}|\mathbf{s}^n], \\
&= \sum_{\mathbf{s}^n \notin T_{\epsilon_1}^n[\mathbf{S}]} p(\mathbf{s}^n) + \sum_{\mathbf{s}^n \in T_{\epsilon_1}^n[\mathbf{S}]} p(\mathbf{s}^n) \Pr[\mathbf{E}(1) \cup \mathbf{E}((2))|\mathbf{s}^n], \quad (\text{D.5})
\end{aligned}$$

where $E(i)$ is the event that the error is made at Decoder i , for $i = 1, 2$. The first term, $\Pr[\mathbf{s}^n \notin T_{\epsilon_1}^n[\mathbf{S}]]$, in the right hand side expression of (D.5) goes to zero as $n \rightarrow \infty$ by Lemma 2.

Without loss of generality, it can be assumed that the output of the host source is $\tilde{\mathbf{s}}^n$, and $(W_1, W_2) = (1, 1)$ is being transmitted from the encoder. Hence, the codeword $\mathbf{X}^n(\tilde{\mathbf{s}}^n, 1, 1)$ is transmitted from the encoder. Let F_1 be the event that $\tilde{\mathbf{s}}^n \in T_{\epsilon_1}^n[\mathbf{S}]$ is output of the host source.

The following error events are considered to compute $\Pr[E(2)|F]$ and can be made to approach zero as $n \rightarrow \infty$.

1. E_1 : $(\mathbf{U}^n(\tilde{\mathbf{s}}^n, 1), \mathbf{X}^n(\tilde{\mathbf{s}}^n, 1, 1), \mathbf{Y}^n, \mathbf{Z}^n) \notin T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}|\tilde{\mathbf{s}}^n]$ under the event F . By using Lemma 2, we can show that $\Pr[E_1|F] \rightarrow 0$ as $n \rightarrow \infty$.
2. E_2 : $(\mathbf{U}^n(\tilde{\mathbf{s}}^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{Z}|\tilde{\mathbf{s}}^n]$ under the event $F \cap E_1^c$ for all $m_2 \neq 1$. It can be shown that $\Pr(E_2|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_2 < \mathbb{I}(\mathbf{U}; \mathbf{Z}|\mathbf{S})$.
3. E_3 : $(\mathbf{U}^n(\mathbf{s}^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{Z}|\mathbf{s}^n]$ under the event $F \cap E_1^c$ for all m_1 and $\mathbf{s}^n \neq \tilde{\mathbf{s}}^n$. It can be shown that $\Pr(E_3|F) \rightarrow 0$ as $n \rightarrow \infty$ by using Lemma 2 and Lemma 3 if $0 \leq R_2 < \mathbb{I}(\mathbf{U}, \mathbf{S}; \mathbf{Z}) - \mathbb{H}(\mathbf{S})$.

From the all above error events, it can be concluded that $\Pr[E(1)|F] \rightarrow 0$ as $n \rightarrow \infty$ if $0 \leq R_2 < \mathbb{I}(\mathbf{U}, \mathbf{S}; \mathbf{Z}) - \mathbb{H}(\mathbf{S})$. The following error events are considered to compute $\Pr[E(1)|F]$ and can be made to approach zero as $n \rightarrow \infty$.

1. E_4 : $(\mathbf{U}^n(\mathbf{s}^n, m_2), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{Y}|\mathbf{s}^n]$ for $m_1 \neq 1$ or $\mathbf{s}^n \neq \tilde{\mathbf{s}}^n$. By considering the error events similar to E_2 and E_3 , it can be shown that $\Pr(E_4|F, E_1^c) \rightarrow 0$ as $n \rightarrow \infty$ if $0 \leq R_2 < \mathbb{I}(\mathbf{U}, \mathbf{S}; \mathbf{Y}) - \mathbb{H}(\mathbf{S})$.

2. $E_5: (\mathbf{X}^n(\tilde{s}^n, m_1, 1), \mathbf{Y}^n) \in T_\epsilon^n[\mathbf{S}, \mathbf{U}, \mathbf{X}, \mathbf{Y} | \tilde{s}^n, \mathbf{U}^n(\tilde{s}^n, 1)]$ for $m_1 \neq 1$. It can be shown that $\Pr(E_5 | F, E_1^c, E_4^c) \rightarrow 0$ as $n \rightarrow \infty$ if $0 \leq R_1 < \mathbb{I}(\mathbf{X}; \mathbf{Y} | \mathbf{S}, \mathbf{U})$.

Then by using the union bound, $\Pr[E(1) \cup E(2) | F]$ goes to zero as $n \rightarrow \infty$ if rate pair (R_1, R_2) satisfies (6.6). It can be concluded that $P_e^n \rightarrow 0$ as $n \rightarrow \infty$ if rate pair (R_1, R_2) satisfies (6.6).

- **Average distortions:** Since $(\mathbf{X}^n, \tilde{s}^n)$ is jointly strongly typical with high probability and the distribution belongs to $\mathcal{P}(\Delta)$, it can be shown that the average distortion $D^{(n)}$ associated with the generated code satisfies the distortion constraint Δ as $n \rightarrow \infty$ as in the Proof of Theorem 6.

Converse

We show that any sequence of $(\lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil, D^{(n)}, n)$ codes, i.e., $\mathbf{X}^n = f(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n)$, $g_1^n(\mathbf{Y}^n) = (\hat{\mathbf{W}}_1, \hat{\mathbf{W}}_2, \hat{\mathbf{S}}^n)$, and $g_2^n(\mathbf{Z}^n) = (\hat{\mathbf{W}}_2, \hat{\mathbf{S}}^n)$, with $\lim_{n \rightarrow \infty} P_e^n = 0$ and $\lim_{n \rightarrow \infty} D^{(n)} \leq \Delta$, the rate pair (R_1, R_2) must satisfy (6.6) for some $(\mathbf{U}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$. Consider a given code of block length n . The joint distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ induced by the code is given by

$$p(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n, \mathbf{x}^n, \mathbf{y}^n, \mathbf{z}^n) = \frac{1}{\lceil 2^{nR_1} \rceil \lceil 2^{nR_2} \rceil} p(\mathbf{s}^n) p(\mathbf{x}^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n) \times \prod_{j=1}^n p(\mathbf{y}_j | \mathbf{x}_j, \mathbf{s}_j) p(\mathbf{z}_j | \mathbf{y}_j),$$

where, $p(\mathbf{x}^n | \mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ is 1 if $\mathbf{x}^n = f^n(\mathbf{w}_1, \mathbf{w}_2, \mathbf{s}^n)$ and 0 otherwise.

We can bound the rate R_1 as follows:

$$\begin{aligned} nR_1 &\leq \mathbb{H}(\mathbf{W}_1) \\ &\stackrel{(a)}{=} \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{S}^n) \\ &= \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{S}^n) - \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^n) + \mathbb{H}(\mathbf{W}_1 | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^n) \end{aligned}$$

$$\begin{aligned}
& \stackrel{(b)}{\leq} \mathbb{I}(\mathbf{W}_1; \mathbf{Y}^n | \mathbf{W}_2, \mathbf{S}^n) + n\epsilon_n \\
& = \sum_{j=1}^n \mathbb{I}(\mathbf{W}_1; \mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}) + n\epsilon_n \\
& = \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1})] + n\epsilon_n \\
& \stackrel{(c)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}, \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}, \mathbf{Z}^{j-1})] + n\epsilon_n \\
& \stackrel{(d)}{\leq} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}, \mathbf{Z}^{j-1}, \mathbf{X}^n)] + n\epsilon_n \\
& \stackrel{(e)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{Y}_j | \mathbf{X}_j, \mathbf{S}_j)] + n\epsilon_n \\
& \stackrel{(f)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Y}_j | \mathbf{S}_j, \tilde{\mathbf{U}}_j) - \mathbb{H}(\mathbf{Y}_j | \mathbf{X}_j, \mathbf{S}_j)] + n\epsilon_n \\
& = \sum_{j=1}^n \mathbb{I}(\mathbf{X}_j; \mathbf{Y}_j | \mathbf{S}_j, \tilde{\mathbf{U}}_j) + n\epsilon_n, \tag{D.6}
\end{aligned}$$

where,

- (a) follows from the fact that \mathbf{W}_1 , \mathbf{W}_2 and \mathbf{S}^n are mutually independent,
- (b) follows from Fano's inequality and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$,
- (c) follows from $\mathbf{Y}_j \leftrightarrow (\mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}) \leftrightarrow \mathbf{Z}^{j-1}$ and $\mathbf{Y}_j \leftrightarrow (\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}) \leftrightarrow \mathbf{Z}^{j-1}$,
- (d) follows from $\mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Y}^{j-1}, \mathbf{Z}^{j-1}) \leq \mathbb{H}(\mathbf{Y}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Z}^{j-1})$, and \mathbf{X}^n is a deterministic function of $(\mathbf{W}_1, \mathbf{W}_2, \mathbf{S}^n)$,
- (e) follows from memoryless properties of the broadcast channel, and
- (f) follows from $\tilde{\mathbf{U}}_j := \{\mathbf{W}_2, \mathbf{S}_1^{j-1}, \mathbf{S}_{j+1}^n\}$.

We can also bound the rate R_2 as follows:

$$\begin{aligned}
nR_2 & \leq \mathbb{H}(\mathbf{W}_2) \\
& \stackrel{(a)}{\leq} \mathbb{H}(\mathbf{W}_2, \mathbf{S}^n) - \mathbb{H}(\mathbf{S}^n) \\
& \stackrel{(b)}{\leq} \mathbb{I}(\mathbf{W}_2, \mathbf{S}^n; \mathbf{Z}^n) - \mathbb{H}(\mathbf{S}^n) + n\epsilon_n
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n [\mathbb{I}(\mathbf{W}_2, \mathbf{S}^n; \mathbf{Z}_j | \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{S}_j | \mathbf{S}^{j-1})] + n\epsilon_n \\
&\stackrel{(c)}{=} \sum_{j=1}^n [\mathbb{H}(\mathbf{Z}^j | \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{Z}_j | \mathbf{W}_2, \mathbf{S}^n, \mathbf{Z}^{j-1}) - \mathbb{H}(\mathbf{S}_j)] + n\epsilon_n \\
&\stackrel{(d)}{\leq} \sum_{j=1}^n [\mathbb{H}(\mathbf{Z}_j) - \mathbb{H}(\mathbf{Z}_j | \tilde{\mathbf{U}}_j, \mathbf{S}_j) - \mathbb{H}(\mathbf{S}_j)] + n\epsilon_n \\
&= \sum_{j=1}^n [\mathbb{I}(\tilde{\mathbf{U}}_j, \mathbf{S}_j; \mathbf{Z}_j) - \mathbb{H}(\mathbf{S}_j)] + n\epsilon_n
\end{aligned}$$

where,

- (a) follows from the fact that \mathbf{W}_1 , \mathbf{W}_2 and \mathbf{S}^n are mutually independent,
- (b) follows from Fano's inequality and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$,
- (c) follows from the fact that \mathbf{S}^n is an i.i.d. random vector,
- (d) follows from $\mathbb{H}(\mathbf{Z}_j | \mathbf{Z}^{j-1}) \leq \mathbb{H}(\mathbf{Z}_j)$, and $\tilde{\mathbf{U}}_j := \{\mathbf{W}_2, \mathbf{S}_1^{j-1}, \mathbf{S}_{j+1}^n\}$.

We can then write (D.6) and (D.7a) as

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y} | \mathbf{Q}, \mathbf{S}, \tilde{\mathbf{U}}) + \epsilon_n, \quad (\text{D.7a})$$

$$R_2 \leq \mathbb{I}(\tilde{\mathbf{U}}, \mathbf{S}; \mathbf{Z} | \mathbf{Q}) - \mathbb{H}(\mathbf{S}) + \epsilon_n, \quad (\text{D.7b})$$

where \mathbf{Q} takes values in the set $\mathcal{Q} \in \{1, 2, \dots, n\}$ with equal probability and the joint probability distribution on $(\mathbf{S}, \mathbf{Q}, \tilde{\mathbf{U}}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is $p(\mathbf{S} = \mathbf{s}, \mathbf{Q} = \mathbf{q}, \tilde{\mathbf{U}} = \tilde{\mathbf{u}}, \mathbf{X} = \mathbf{x})p(\mathbf{y} | \mathbf{x}, \mathbf{s})p(\mathbf{z} | \mathbf{y})$, with

$$\begin{aligned}
p(\mathbf{S} = \mathbf{s}, \mathbf{Q} = \mathbf{q}, \tilde{\mathbf{U}} = \tilde{\mathbf{u}}, \mathbf{X} = \mathbf{x}) &= \\
& p(\mathbf{s})p(\mathbf{q})p(\mathbf{U}_q = \tilde{\mathbf{u}} | \mathbf{s}, \mathbf{q})p(\mathbf{X}_q = \mathbf{x} | \mathbf{s}, \mathbf{q}, \tilde{\mathbf{u}}).
\end{aligned}$$

Finally, we can write (D.7) as

$$R_1 \leq \mathbb{I}(\mathbf{X}; \mathbf{Y} | \mathbf{U}, \mathbf{S}) + n\epsilon_n,$$

$$R_2 \leq \mathbb{I}(\mathbf{U}, \mathbf{S}; \mathbf{Z}) - \mathbb{H}(\mathbf{S}) + n\epsilon_n,$$

where $\mathbf{U} := (\mathbf{Q}, \tilde{\mathbf{U}})$, since $\mathbb{I}(\tilde{\mathbf{U}}, \mathbf{S}; \mathbf{Z} | \mathbf{Q}) \leq \mathbb{I}(\mathbf{Q}, \tilde{\mathbf{U}}, \mathbf{S}; \mathbf{Z})$.

Given any $\delta > 0$, the associated distortion $D^{(n)}$, for sufficiently large n , satisfies

$$\begin{aligned}
 \Delta + \delta &\geq D^{(n)} \\
 &= \mathbb{E}d(\mathbf{X}^n, \mathbf{S}^n) \\
 &= \frac{1}{n} \sum_{j=1}^n \sum_{\mathbf{x}, \mathbf{s}} p(X_j = \mathbf{x}, S_j = \mathbf{s}) d(\mathbf{x}, \mathbf{s}) \\
 &= \sum_{\mathbf{x}, \mathbf{s}} p(\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}) d(\mathbf{x}, \mathbf{s}) \\
 &= \mathbb{E}d(\mathbf{X}, \mathbf{S}).
 \end{aligned}$$

As $n \rightarrow \infty$ and $\delta \rightarrow 0$, $(\mathbf{U}, \mathbf{S}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathcal{P}(\Delta)$ and $(R_1, R_2) \in \mathcal{C}_C$.

BIBLIOGRAPHY

- [1] R. J. Anderson and F. A. P. Petitcolas. On the Limits of Steganography. *IEEE Journal of Selected Areas in Communications*, 16(4):474–484, May 1998.
- [2] G. Caire and S. Shamai. On the Capacity of Some Channels with Channel State Information. *IEEE Trans. Inform. Theory*, 45, September 1999.
- [3] G. Caire and S. Shamai (Shitz). On achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, 49(7):1691–1706, July 2003.
- [4] Y. Cemal and Y. Steinberg. Multiple Access Channel with Partial State Information at the Encoders. *IEEE Trans. Inform. Theory*, vol.IT-51:3992–4003, November 2005.
- [5] B. Chen. *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.
- [6] B. Chen and G. W. Wornell. Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. *IEEE Trans. Inform. Theory*, 47(4):1423–1443, May 2001.
- [7] M. Chiang, A. Sutivong, and T. M. Cover. Channel Capacity and State Estimation. volume 4, pages 838– 840, Honolulu, Hawaii, USA, November 2000.
- [8] A. S. Cohen. The Gaussian Watermarking Game. *IEEE Trans. Inform. Theory*, vol.48:1639–1669, June 2002.
- [9] M. H. M. Costa. Writing on Dirty Paper. *IEEE Trans. Inform. Theory*, vol.IT-29:439–441, May 1983.
- [10] T. Cover and M. Chiang. Duality Between Channel Capacity and Rate Distortion with Two-sided State Information. *IEEE Trans. Inform. Theory*, 48:1629–1638, June 2002.
- [11] T. Cover, Y. Kim, and A. Sutivong. Simultaneous Communication of Data and State. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 24 - June 29 2007.
- [12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

- [13] I. Csiszár and J. Körner. Broadcast Channels with Confidential Messages. *IEEE Trans. Inform. Theory*, 24:339–348, May 1978.
- [14] I. Csiszár and J. Körner, editors. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press Inc., New York, 1981.
- [15] S. C. Draper. *Successive Structuring of Source Coding Algorithms for Data Fusion, Buffering, and Distribution in Networks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2002.
- [16] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.
- [17] S. I. Gel'fand and M. S. Pinsker. Coding for Channel with Random Parameters. *Probl. Contr. and Information Theory*, 9(1):pp.19–31, 1980.
- [18] S. I. Gel'fand and M. S. Pinsker. On Gaussian Channels with Random Parameters. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 1983.
- [19] C. Heegard and A. El Gamal. On the Capacities of Computer Memories with Defects. *IEEE Trans. Inform. Theory*, vol.IT-29:731–739, September 1983.
- [20] S. A. Jafar. Capacity with Causal and Non-Causal Side Information - A Unified View. *IEEE Trans. Inform. Theory*, 52(12):5468–5475, December 2006.
- [21] T. Kalker and F. Willems. Capacity Bounds and Constructions for Reversible Data-hiding. In *Proc. Int. Conf. Digital Signal Processing*, pages 71–76, 2002.
- [22] T. Kalker and F. Willems. Capacity Bounds and Constructions for Reversible Data-hiding. In *Proc. SPIE Int. Conf. Security and Watermarking of Multimedia Contents*, volume 5020, pages 604–611, 2003.
- [23] A. Khisti, U. Erez, and G. W. Wornell. Writing on Many Pieces of Dirty Paper at Once: The Binary Case. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 27 – July 2 2004.
- [24] Y. H. Kim, A. Sutivong, and S. Sigurjónsson. Multiple User Writing on Dirty Paper. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 27 - July 2 2004.
- [25] S. Kotagiri and J. N. Laneman. Multiple Access Channels with State Information Known at Some Encoders. Submitted to "*EURASIP J. Wireless Comm. Net.*," , September 2007.
- [26] S. Kotagiri and J. N. Laneman. Achievable Rates for Multiple Access Channels with State Information Known at One Encoder. In *Proc. Allerton Conf. Communications, Control, and Computing*, 2004.
- [27] S. Kotagiri and J. N. Laneman. Multi-user Reversible Information Embedding. In *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, 2005.

- [28] S. Kotagiri and J. N. Laneman. Information Embedding in Degraded Broadcast Scenarios. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seattle, WA, USA, 2006.
- [29] S. Kotagiri and J. N. Laneman. Multiaccess Channels with State Known to One Encoder: A Case of Degraded Message Sets. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 24 - June 29 2007.
- [30] A. V. Kusnetsov and B. S. Tsybakov. Coding in a Memory with Defective Cells. *Probl. Peredach. Inform.*, vol.10(2):52–60, April/June 1974.
- [31] A. Maor and N. Merhav. On Joint Information Embedding and Lossy Compression. *IEEE Trans. Inform. Theory*, 51(8):2998–3008, August 2005.
- [32] A. Maor and N. Merhav. On Joint Information Embedding and Lossy Compression in the Presence of a Memoryless Attack. *IEEE Trans. Inform. Theory*, 51(9):3166–3175, 2005.
- [33] N. Merhav. On Joint Coding for Watermarking and Encryption. *IEEE Trans. Inform. Theory*, 52(1):190–205, January 2006.
- [34] P. Moulin and J. O’Sullivan. Information-theoretic Analysis of Information Hiding. *IEEE Trans. Inform. Theory*, 49:563–593, 2003.
- [35] J. G. Proakis. *Digital Communications*. McGraw-Hill, Inc., New York, Third edition, 1995.
- [36] A. Rosenzweig, Y. Steinberg, and S. Shamai(Shitz). On Channels with Partial Channel State Information at the Transmitter. *IEEE Trans. Inform. Theory*, vol.IT-51:1817–1830, May 2005.
- [37] M. Salehi. Capacity and coding for memories with real-time noisy defect information at encoder and decoder. *Proc. Inst. Elec. Eng.-Pt.I.*, vol.139:113–117, April 1992.
- [38] C. E. Shannon. Channels with Side Information at the transmitter. *IBM J. Res. Devel.*, vol.2:289–293, 1958.
- [39] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, IL, 1949.
- [40] S. Sigurjónsson and Y. H. Kim. On Multiple User Channels with State Information at the Transmitters. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2005.
- [41] A. Somekh-Baruch and N. Merhav. On the error exponent and capacity games of private watermarking systems. *IEEE Trans. Inform. Theory*, 49(3):537–562, March 2003.
- [42] A. Somekh-Baruch and N. Merhav. On the capacity game of public watermarking system. *IEEE Trans. Inform. Theory*, 50(3):511–524, March 2004.

- [43] A. Somekh-Baruch and N. Merhav. On the capacity game of private fingerprinting systems under collusion attacks. *IEEE Trans. Inform. Theory*, 51(3):884–899, March 2005.
- [44] A. Somekh-Baruch, S. Shamai, and S. Verdú. Cooperative Multiple Access Encoding with States Available at One Transmitter. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 24 - June 29 2007.
- [45] Y. Steinberg. Coding for the Degraded Broadcast Channel with Random parameters, with Causal and Noncausal Side Information. *IEEE Trans. Inform. Theory*, vol.51:2867–2877, August 2005.
- [46] Y. Steinberg and S. Shamai. Achievable Rates for the Broadcast Channel with States Known at the Transmitter. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, September 4 – September 9 2005.
- [47] W. Sun and E. Yang. *Information Hiding*, chapter On Achievable Regions of Public Multiple-Access Gaussian Watermarking Systems, pages 38–51. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Berlin / Heidelberg, December 2004.
- [48] A. Sutivong, M. Chiang, T. M. Cover, and Y. H. Kim. Channel Capacity and State Estimation for state-dependent Gaussian Channels. *IEEE Trans. Inform. Theory*, 51:1486–1495, April 2005.
- [49] M. D. Swanson, M. Kobayashi, and A. H. Tewfik. Multimedia Data-Embedding and Watermarking Technologies. In *Proc. IEEE Int. Conf. Communications (ICC)*, volume 2, pages 823–827, 1998.
- [50] L. Tong, B.M. Sadler, and M. Dong. Pilot Assisted Wireless Transmissions: General Model, Design Criteria, and Signal Processing. *IEEE Signal Processing Magazine*, 21(6):12–25, November 2004.
- [51] S. Tung. *Multiterminal Source Coding*. PhD thesis, Cornell University, Ithaca, New York, May 1978.
- [52] S. Vedantam, W. Zhang, U. Mitra, and A. Sabharwal. Joint Channel Estimation and Data Transmission: Achievable Rates. In *IEEE Information Theory Workshop*, Lake Tahoe, CA, USA.
- [53] F. M. J. Willems and T. Kalker. Coding Theorems for Reversible Embedding. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 66, pages 61–78. American Mathematical Society, 2004.
- [54] A. D. Wyner. The Wire-Tap Channel. *Bell Syst. Tech. Journal*, vol.54:1355–1387, October 1975.
- [55] A. Zaidi, P. Piantanida, and P. Duhamel. Broadcast- and MAC- Aware Coding Strategies for Multiple User Information Embedding. *IEEE Trans. Signal Processing*, 55(8):2974–2992, 2007.
- [56] A. Zaidi and L. Vandendorpe. Coding Schemes for Relay-Assisted Information Embedding. Submitted to *IEEE Trans. on Security and Forensics*.

- [57] A. Zaidi and L. Vandendorpe. Rate Regions for the Partially-Cooperative Relay Broadcast Channel with Non-Causal Side Information. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 24 - June 29 2007.
- [58] R. Zamir, S. Shamai, and U. Erez. Nested Linear/Lattice Codes for Structured Multiterminal Binning. *IEEE Trans. Inform. Theory*, vol.IT-48:1250–1276, June 2002.
- [59] W. Zhang, S. Kotagiri, and J. N. Laneman. Writing on Dirty Paper with Resizing and its Application to Quasi-Static Fading Broadcast Channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Nice, France, June 24 - June 29 2007.