TOPICS IN NETWORK INFORMATION THEORY

JIANG JINHUA

NATIONAL UNIVERSITY OF SINGAPORE

2008

TOPICS IN NETWORK INFORMATION THEORY

JIANG JINHUA
(*B. Eng., National University of Singapore*)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2008

To my parents

Acknowledgements

First and Foremost, I would like to thank my advisor, Prof. Yan Xin, for his invaluable guidance and constant support throughout my Ph.D. study. I am grateful to him for introducing me to the fields of communications and information theory and teaching me how to do research. His valuable advises and incisive criticisms have greatly influenced my thinking and writing. I would like to thank my co-advisor, Prof. Hari Krishna Garg for his advice and encouragement.

I would like to thank Profs. Meixia Tao and Arumugam Nallanathan for serving my oral qualification exam panel members.

I also would like to thank Hon-Fah Chong, Lawrence Ong, Feifei Gao, Jianwen Zhang, Yonglan Zhu, Lan Zhang, Seyed Hossein Seyedmehdi, Zhongjun Wang, Le Cao, Wei Cao, Rong Li, Yan Li, Qi Zhang, Jun He, Yang Lu, Lokesh Bheema Thiagarajan, Mingsheng Gao, Xin Kang, Elisa Mo, Mingwei Wu, Qian Chen, Anwar Halim, Eric Siow, Rick Zheng, Dexter Wang, Yantao Yu, Litt Teen Hiew, Fei Wang, and many others, for their help in either my research or other matters.

Last but not least, I would like to thank my parents, Haitian Jiang and Qiaoyun Zhang, for their love, encouragement, and support.

Contents

Summary **vi**

List of Tables **viii**

List of Figures **ix**

Abbreviations **xi**

1 Introduction **1**

1.1 Preliminary Background 2

1.2 Motivations and Challenges 4

1.3 Contributions and Organization of the Thesis 6

2 Interference Channels With Common Information **11**

2.1 Introduction 11

2.2 Channel Models and Preliminaries 14

2.2.1 Discrete Memoryless Interference Channel With Common Information 14

2.2.2 Modified Discrete Memoryless Interference Channel With Common Information 16

2.3 Discrete Memoryless ICC 18

2.3.1 An Achievable Rate Region for the Discrete Memoryless ICC 18

2.3.2 Explicit Description of the Achievable Rate Region 21

2.4 Relations between $\mathcal{R}_{\text{impl}}$ and Some Existing Results 22

2.4.1 Achievable Rate Region for the ICC by Tan 23

2.4.2 Strong Interference Channel With Common Information . . 25

2.4.3 Interference Channel Without Common Information 27

2.5 Two Special Cases of the ICC 28

2.5.1 Asymmetric Interference Channel With Common Information 28

2.5.2 Deterministic Interference Channel With Common Information 29

2.6 Gaussian Interference Channel With Common Information 33

2.6.1 Channel Model for the Gaussian ICC 33

2.6.2 An Achievable Rate Region for the GICC 34

2.7 Conclusions 36

3 Interference Channels With Degraded Message Sets 37

3.1 Introduction 37

3.2 Channel Model 40

3.3 An Achievable Rate Region for the Discrete Memoryless IC-DMS . 42

3.4 Relating With Some Existing Rate Regions 50

3.4.1 A Subregion of $\mathcal{R}$ 50

3.4.2 A Subregion of $\mathcal{R}_{\text{sim}}$ 52

3.5 Gaussian IC-DMS 55

3.5.1 Channel Model of the GIC-DMS 55

3.5.2 Achievable Rate Regions for the GIC-DMS 56

3.5.2.1 Gaussian Extension of $\mathcal{R}$ 56

3.5.2.2 Gaussian Extension of $\mathcal{R}_{\text{suc}}$ 58

3.5.3 Numerical Examples 61

3.6 Conclusions 63

4 Discrete Memoryless Interference Channels With Perfect Feedback 65

4.1 Introduction 65

4.2 Channel Model and Preliminaries 67

4.3 An Achievable Rate Region for the ICF 68

4.4 Conclusions 77

5 Relay Channels With Generalized Feedback 78

5.1 Introduction 78

5.2 Channel Models and Preliminaries 81

5.2.1 Relay Channel With Generalized Feedback at the Source . . 81

5.2.2 Relay Channel With Generalized Feedback from the Destination 83

5.2.3 Strong Typicality 84

5.3 Achievable Rates for $\mathcal{C}_{\text{SFB}}$ 86

5.3.1 Rates Achieved by Decode-and-Forward / Partially-Decode-and-Forward 87

5.3.2 A Rate Achieved by Compress-and-Forward 99

5.3.3 Special Cases 100

5.4 Achievable Rates for $\mathcal{C}_{\text{DFB}}$ 103

5.5 Concluding Remarks 106

6 Summary of Contributions and Future Work **107**

6.1 Summary of Contributions 107

6.2 Future Work 109

A Appendices to Chapter 2 **111**

A.1 Proof of Lemma 2.2 111

A.2 Proof of the Convexity of $\mathcal{R}_m$ 114

A.3 Proof of Corollary 2.1 115

A.4 Proof of the Converse Part of Theorem 2.4 124

B Appendices to Chapter 5 **134**

B.1 Proof of Theorem 5.2 134

B.2 Proof of Theorem 5.3 139

B.3 Proof of Theorem 5.5 142

Bibliography **147**

List of Publications **154**

Summary

This thesis studies a number of topics in network information theory. Four channel models in a wireless network, including the interference channel with common information (ICC), the interference channel with degraded message sets (IC-DMS), the interference channel with perfect feedback (ICF), and the relay channel with generalized feedback, are investigated. Three major challenging issues in a wireless network, *correlated sources*, *interference*, and *feedback*, are involved in these models. New coding schemes are developed for each channel model, based on the existing coding techniques: superposition coding, collaborative coding (also referred to as rate splitting), Gel'fand-Pinsker coding, decode-and-forward (DF), and compress-and-forward (CF). Corresponding new achievable rates/rate regions are obtained for these channels.

Specifically, a cascaded superposition coding scheme for the ICC is proposed, and a new achievable rate region is obtained for the channel. The new achievable rate region offers strict improvements over one existing rate region for the channel, which is demonstrated using a Gaussian example. The new rate region is also shown to be tight for a class of deterministic ICCs (DICCs) by establishing an outer-bound of the capacity region that meets the inner bound defined by our new rate region. For the IC-DMS, collaborative coding, Gel'fand-Pinsker coding, and superposition coding are applied collectively to develop a new coding scheme for the channel, which allows the senders and the receivers to collaborate in combating against the interference, and also allows one sender to help the other through cooperation. The obtained achievable rate region also offers strict improvements over the existing results, which is shown by using Gaussian examples.

Causal perfect feedback and generalized feedback are then considered for the interference channel and relay channel, respectively. For the ICF, partially-decode-and-forward together with the collaborative coding is applied to exploit the feedback and induce cooperation between the senders. With the proposed block Markov coding scheme, a new achievable rate region is obtained for this channel in the discrete memoryless case. The relay channels with generalized feedback investigated include two cases: 1) the source and the relay both operate in full duplex mode; 2) the relay and the destination both operate in full duplex mode. Coding schemes based on the ideas of DF and CF are developed for each case, aiming to fully exploit the feedback to improve the transmission rates between the source and the destination. It is shown that the new achievable rates obtained for the first case include the existing results on the relay channel with perfect feedback as special cases, and the new achievable rates for the second case are asymptotically tight for the extreme case.

List of Tables

5.1 Codewords transmitted in each block to achieve R_{SFB0}. 90

B.1 Codewords transmitted in each block to achieve R_{SFB2}. 140
B.2 Codewords transmitted in each block to achieve R_{SFB1}. 146
B.3 Codewords transmitted in each block to achieve R_{DFB2}. 146

List of Figures

1.1 A simple wireless network of six nodes. 2

1.2 A four-node WSN scenario: a common event is captured by two source nodes. 6

1.3 A four-node WSN scenario: the event captured by one source node is completely captured by the other source node. 8

2.1 Interference channel with common information. 15

2.2 Modified interference channel with common information. 16

2.3 Asymmetric interference channel with common information. 28

2.4 A class of deterministic interference channels with common information. 30

2.5 Gaussian interference channel with common information. 33

2.6 $P_1 = 6$, $P_2 = 0.5$, $c_{21} = 1$, $c_{12} = 0.25$. The dashed lines characterize the rate regions of $\mathcal{G}_{\text{Tan}}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively, and the solid lines characterize the rate regions of $\mathcal{G}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively. 36

3.1 An interference channel with degraded message sets in which sender 2 is close to receiver 1. 39

3.2 Interference channel with degraded message sets. 41

3.3 Gaussian interference channel with degraded message sets. 55

3.4 Achievable rate regions for GIC-DMS with setting: $P_1 = P_2 = 6$, $c_{21} = 0.3$, $c_{12} = 0$. (i) gives the rate region $\mathcal{G}_{\text{dmt1}}$; (ii) gives the rate region $\mathcal{G}_{\text{dmt2}}$; (iii) gives our rate region $\mathcal{G}_{\text{sp1}}$. 62

3.5 Achievable rate regions for the GIC-DMS two different settings: (I) $P_1 = 6$, $P_2 = 6$, $c_{21} = 2$, $c_{12} = 0.3$; (II) $P_1 = 6$, $P_2 = 1$, $c_{21} = 2$, $c_{12} = 0.3$. (i) gives the rate regions, $\mathcal{G}$, Gaussian counter part of Theorem 3.1; (ii) gives the rate regions $\mathcal{R}_{\text{sp1}}$; (iii) gives the rate regions $\mathcal{R}_{\text{sp2}}$. 63

4.1 Interference channel with perfect feedback. 67

5.1 Three-node relay channel with generalized feedback available at the source. 82

5.2 Three-node relay channel with generalized feedback from the destination. 83

5.3 Three-node relay channel in which all nodes are in full duplex mode. 106

Abbreviations

3G Third Generation.

AEP Asymptotic equipartition property.

AICC Asymmetric Interference Channels with Common Information.

BC Broadcast Channel.

CF Compress-and-Forward.

CMG Chong-Motani-Garg.

DF Decode-and-Forward.

DICC Deterministic Interference Channels with Common Information.

GIC-DMS Gaussian Interference Channel with Degraded Message Sets.

GICC Gaussian Interference Channel with Common Information.

GSM Global System for Mobile.

HK Han-Kobayashi.

IC Interference Channel.

IC-DMS Interference Channel with Degraded Message Sets.

ICC Interference Channel with Common Information.

ICF Interference Channel with Perfect Feedback.

MAC Multiple Access Channel.

MACC Multiple Access Channel with Common Information.

MIMO Multiple-Input Multiple-Output.

RC Relay Channel.

SIC Strong Interference Channel.

SICC Strong Interference Channel with Common Information.

TWC Two-way Channel.

WSN Wireless Sensor Network.

Chapter 1

Introduction

Wireless communication devices, ranging from mobile phones to laptops and other hand-held devices, have gradually become ubiquitous in our modern daily life. The unprecedented convenience and mobility brought by these devices are built on various wireless networks, such as the GSM or 3G network, as well as wireless local area networks. Although these wireless networks have been widely deployed and used, it is generally an open question whether the current design of the network is optimum in terms of either power efficiency or data transmission rate. Network information is being developed with the aim to answer this question. On the other hand, information theoretic study provides constructive insights on the design of various coding strategies to achieve the limit and unleash the potential of a network.

This thesis investigates several topics in network information theory, including the interference channel with common information (ICC), the interference channel with degraded message sets (IC-DMS), the interference channel with perfect feedback (ICF), and the relay channel with generalized feedback. Several coding schemes for these models are developed. With these coding schemes, achievability results serving as the lower or inner bounds of the capacity or capacity regions are derived. Moreover, the capacity region for a class of deterministic ICCs is established.

1.1 Preliminary Background

In general, a network consists of multiple source nodes that have certain information to transmit, and multiple destination nodes to which the information from the source nodes are to be conveyed. Moreover, between the source nodes and destination nodes, there may exist a number of relay nodes that can aid the intended transmissions between source nodes and destination nodes. The long standing open problem in network information theory is how to characterize and determine the fundamental performance limit of a general network. Efforts and advancements have been consistently made by information theorists towards addressing this problem.

Primary focuses are on relatively simple network models, including the two-way channel (TWC) [1], the multiple access channel (MAC) [2], the broadcast channel (BC) [3], the relay channel (RC) [4], and the interference channel (IC) [5], which are typically considered to be the fundamental building blocks of a network.

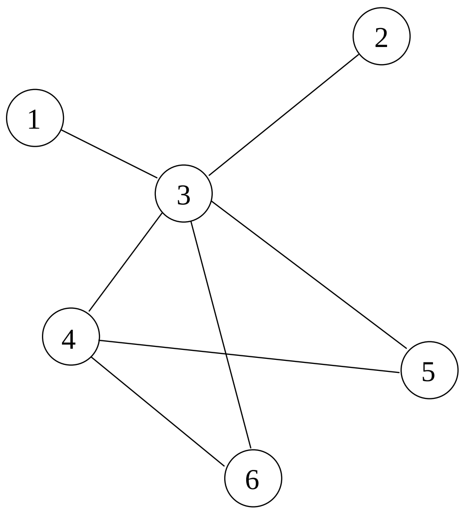

Figure 1.1: A simple wireless network of six nodes.

A generic two-user MAC consists of three nodes: two senders and one common receiver. Both senders wish to convey certain information to the common receiver. As depicted in Fig. 1.1, when both node 1 and node 4 wish to send certain information to node 3, the three nodes form such a MAC. To date, amongst the five elementary channels, the MAC is the most thoroughly studied one with the capacity regions being found for both the generic case [2, 6] and most of its variants including the MAC with common information (MACC) [7], the MAC with conferencing encoders [8], the Gaussian MAC with perfect feedback [9], and so on. One of the remaining challenging open problem regarding the MAC is to find the capacity region of the general discrete memoryless MAC with perfect feedback, for which only achievable rate regions have been obtained in [10] and [11].

In contrast to the MAC, a generic two-user BC also consists of three nodes: one sender and two receivers. In Fig. 1.1, node 3, node 2, and node 5 form a two-user BC, when node 3 wants to simultaneously transmit two different messages to node 2 and node 5. For the general BC, the capacity region has remained open for many years since the introduction of this channel [12] in 1972. The best achievable rate region for the general BC was obtained by Marton in [13]. Capacity regions have been established only for several special cases including the degraded BC [14, 15], the BC with degraded message sets [16], etc. One of the very recent breakthroughs is made on the Gaussian Multiple-Input Multiple-Output (MIMO) BC. Sum-rate capacity for the MIMO BC has been found in [17, 18, 19], while the entire capacity region has been established in [20].

Referring to Fig. 1.1, a simple RC is formed by node 4, node 3, and node 5, when node 4 wishes to send certain information to node 5 with the aid from node 3. In such a three-node RC, node 4, node 3, and node 5 are usually termed as the source, relay, and destination, respectively. Similar to the BC, the capacity of the general RC has also remained an open problem for long since its invention [4] in 1971. Nevertheless, many results have been obtained on this channel. In particular, two well-known coding strategies, the decode-and-forward (DF) strategy and the compressed-and-forward (CF) strategy, were introduced in [21] for RC. A hybrid of these two strategies leads to the best achievable rate for the generic RC [21, 22]. Both strategies have also been extended to large networks consisting of multiple relays [23, 24]. Capacity results have been established for some special cases, e.g., the degraded RC and reversely degraded RC [21], the semi-deterministic RC [25], the RC with phase fading [23], etc.

In Fig. 1.1, when node 3 and node 4 simultaneously transmit some information to node 5 and node 6 respectively, they form a simple two-user IC. The two simultaneous transmissions would interfere with each other due to broadcasting nature of wireless networks. The capacity region of the general IC is also not found, while capacity regions have been characterized for a number of special cases, e.g., the strong IC (SIC) [26, 27, 28, 29, 30], a class of discrete additive degraded ICs [31], and a class of deterministic ICs [32], etc. For the general case, various inner and outer bounds of the capacity region have been obtained [28, 33, 34]. In particular, the achievable rate region obtained in [28, 34] is by far the largest one, or the tightest as an inner bound of the capacity region for the IC.

Notably, Gupta, and Kumar investigated the throughput and delay of a wireless network consisting of a large number of randomly distributed but immobile nodes [35] (a large scale wireless network) , which paves the way to a new research area in network information theory. Following their seminal work, considerable research attention has been received on the large scale wireless network (see [36, 37, 38, 39] and references therein).

This thesis will present our work on subjects in the domain of the conventional network information theory rather than the new direction on the large scale wireless networks. Specifically, several variants of the IC and the RC are investigated from the conventional information theoretic perspective.

1.2 Motivations and Challenges

As mentioned earlier, the fundamental limit of a wireless network is the ultimate question to be answered by information theoretic studies. Towards answering such a question, three major challenging issues have to be addressed: 1) correlated sources, 2) interference, and 3) feedback [40]. As basic building blocks of a network, the simple network models introduced in the previous section usually involve only one or none of the three issues, i.e., the IC explicitly involves the issue of inference, while the MAC does not involve any of the three. Nevertheless, it is indeed a common phenomenon to have two or three issues involved altogether in wireless networks, especially the wireless sensor networks (WSN).

Emerged as one of the hottest research topics in recent years, the WSN [41] refers to a type of wireless network consisting of a large number of small sensor nodes that are equipped with three basic functions: sensing, data processing, and wireless networking. The sensor nodes are usually randomly located. Each node monitors its own nearby environment to capture the events in the monitored area, and then conveys the information about the captured events to some other nodes or a fusion center. These special characteristics make all the three challenging issues prominent in a WSN, which urges us to consider following sensor network scenarios and the related questions.

First, as the sensors are randomly located, it is likely that two neighboring sensors are near enough such that the events or source messages that they captured or obtained are correlated. Efficient schemes need to be designed to explore the correlation and convey the correlated information through the channel.

Due to the inherent broadcasting nature of wireless channels, every node that has a receiver will be affected by any signals that are being transmitted on the air. For example, when two sensor nodes has two different messages to send to two different receiver nodes, each receiver will suffer certain interference from the non-pairing transmitting senor node. This is, in fact, the generic IC when the two messages are statistically independent. This type of interference is the most common one in a wireless network, while in some other cases, the interference caused by one transmitting node can be non-causally known at another transmitting node. It is necessary to design coding schemes to allow the interfered receiving node to reduce the effect of the interference to a certain extent, or allow the pairing sender of the interfered receiving node to effectively utilize the non-causally known interference.

When some senor nodes are full duplex nodes, which can simultaneously transmit and receive signals, each of them will receive real-time feedback from the channel while they are trying to send certain information to other nodes in the network. We term this type of feedback the passive feedback, as the transmitting sensor nodes are passively receiving the feedback from the channel. The other type of feedback is termed the active feedback, as a data collecting node or destination node can actively send certain feedback to the nodes that are trying to convey information to it. How to effectively exploit the passive feedback signals, how to design active feedback schemes, and what information to be carried by the active feedback, are interesting questions to be studied.

A detailed description of the problems motivated by the WSN and our respective contributions are given the next section.

1.3 Contributions and Organization of the Thesis

The main contributions of this thesis can be summarized as developing new coding strategies for various wireless channel models using some existing coding techniques to effectively deal with the correlation, interference, and feedback, with the objective to achieve better transmission rates than existing ones.

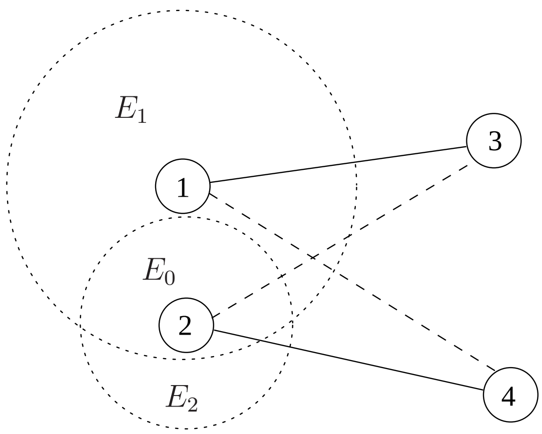

Figure 1.2: A four-node WSN scenario: a common event is captured by two source nodes.

In Chapter 2, we investigate a four-node network, where two sensor nodes monitor the environment and capture the events in the respective monitored region, and try to send the information of the captured events to the respective destination node. This communication scenario is shown in Fig. 1.2. We can see that node 1 detects two events E_0 and E_1, while node 2 detects E_0 and E_2. Both nodes have captured a common event E_0 besides the individual private event E_1 or E_2. Node 1 is required to deliver the information about the events that it has captured to its pairing destination node, node 3. Node 2 needs perform a similar task. In other words, node 1 and node 2 need to send certain correlated information to node 3 and node 4, while the correlation is in the form of common information. We term this type of channel as the interference channel with common information (ICC). In Chapter 2, we first develop a cascaded superposition coding scheme for the ICC, and obtain an achievable rate region for the channel in the general discrete memoryless case.

The coding scheme effectively deals with the common information by allowing the two source nodes to fully cooperate to send the common information. On top of that, the coding scheme also allows the destination nodes to partially decode the private information from the non-pairing source nodes, which aims to reduce the effective interference suffered by each destination node. The corresponding achievable rate region is shown to reduce to several known ones under the respective channel settings. We also investigate two special classes of this channel, including a class of channels where one sender has no private information to send, and a class of deterministic channels. For the first special case, we obtain an achievable rate region with simple description, and this rate region has been shown to be the capacity region in a recent paper [42]. For the second special case, we establish the converse for our achievable rate region, resulting a full characterization of the capacity region of this class of channels. We also extend our achievable rate region from the discrete memoryless case to the Gaussian case, and we are able to demonstrate strict improvement of our rate region over the existing result using a numerical example.

In Chapter 3, we also investigate a four-node network, but the scenario is differ-

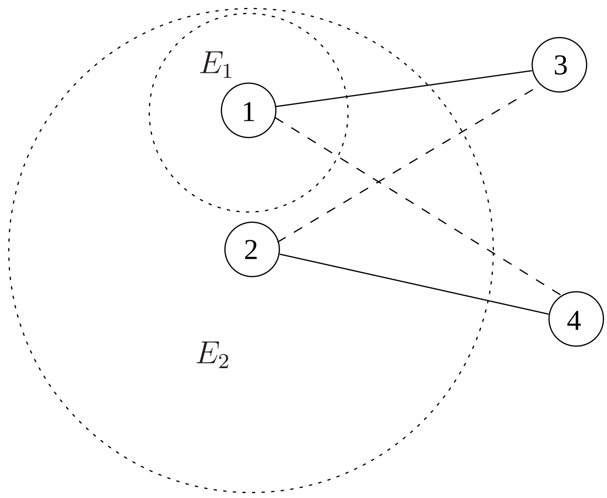

Figure 1.3: A four-node WSN scenario: the event captured by one source node is completely captured by the other source node.

ent. First, there is no common event captured by both node 1 and node 2, so the two destination nodes need to decode E_1 or E_2 only. Second, node 2 is assumed to be powerful enough such that it not only can capture the event E_2, but also can capture the event E_1. A graphical model of this communication scenario is presented in Fig. 1.3. Assume that both node 1 and node 2 would apply block coding on the information to be sent through the channel, and assume that the codebooks are revealed to all the nodes. We can observe that, although the signals transmitted from node 1 would interfere the signal reception at node 4 (the pairing destination node of node 2), node 2 has *a priori* knowledge of the interference that node 4 would suffer from. At the same time, node 3 would also suffer the interference from node 2 which is trying to convey its intended information to node 4. We refer to this type of channel as the interference channel with degraded message sets (IC-DMS), which is also known as the *cognitive radio channel or Genie-aided cognitive radio channel* [43]. In such a channel, two kinds of interference coexists. We develop a coding scheme which are based on Gel'fand-Pinsker coding, collaborative coding (or rate splitting), and superposition coding for the channel. With resort to this coding scheme, we obtain a new achievable rate region for the discrete memoryless IC-DMS, which generalizes several existing regions. We also extend the new achievable rate region to the Gaussian case. One of the existing rate regions has been proven to be the capacity region for certain class of channels, i.e., the weak IC-DMS or the IC-DMS in the *low-interference-gain* regime. Nevertheless, our achievable rate region offers strict improvement over those regions in the *high-interference-gain* regime, which is demonstrated using Gaussian numerical examples.

Having investigated the aspects of correlation (in the form of common information) and interference in a four-node WSN model, we further study the situations of perfect feedback and generalized feedback in a four-node network model and a three-node network model in Chapters 4 and 5.

Specifically, we first study a four-node case with perfect feedback in Chapter 4. Two sensor nodes monitor the nearby environment and send the information about the detected events to their respective destination node, while we assume that the two destination nodes are able to causally send the received channel outputs back perfectly to their respective source node. This is termed as the interference channel with perfect feedback (ICF). We develop a block Markov coding scheme based on rate splitting and the DF coding strategy for the channel. The coding scheme allows the senders to perform cross decoding of the information sent by each other in one block, such that the two senders can fully cooperate to transmit the crossly decoded information in the next block. We derive a corresponding new achievable rate region for the discrete memoryless ICF.

A three-node wireless network, namely the RC, with generalized feedback is considered in Chapter 5. We consider two difference feedback configurations. We first assume that the source of the RC is a full duplex node, which not only can transmit signals to other nodes, but can simultaneously receive signals induced by transmissions in the channel. We develop several coding schemes for this configuration which allow us to exploit the feedback received at the source node. The coding schemes are mainly based on the ideas of DF and CF coding strategies developed for the generic RC. Corresponding achievable rates are derived with the respective coding schemes. We show that the derive achievable rates for this generalized feedback setting reduce to the existing results for the perfect feedback setting under the specific channel assumptions.

We then consider a different scenario, where the destination is assumed to be a full duplex node. The destination can now actively send feedback to the relay. For this configuration, we construct coding schemes based on the DF and CF strategies as well. The achievable rates are shown to be asymptotically optimal, i.e., our achievable rates become the capacity for the extreme case.

In Chapter 6, we summarize our contributions, and point out some of the possible extensions of the work in this thesis.

Notation: Throughout the thesis, we apply the notations described as follows. Random variables and their realizations are denoted by upper case letters and lower case letters respectively, e.g., X and x. Bold fonts are used to indicate vectors, e.g., $\mathbf{X}$ and $\mathbf{x}$. Sets are denoted by calligraphic letters, e.g., $\mathcal{X}$.

Chapter 2

Interference Channels With Common Information

In this chapter, the ICC, in which two senders need deliver not only private messages but also certain common messages to their *corresponding* receivers, is investigated. An achievable rate region for such a channel is obtained by applying a superposition coding scheme that consists of successive encoding and simultaneous decoding. It is shown that the derived achievable rate region includes or extends several existing results for the ICs with or without common information. The rate region is then specialized to a class of ICCs in which one sender has no private information to transmit, and a class of deterministic interference channels with common information (DICCs). In particular, the derived rate region is found to be the capacity region for this class of DICCs. Lastly, the achievable rate region derived for the discrete memoryless ICC is extended to the Gaussian case, in which a numerical example is provided to illustrate the improvement of our rate region over an existing result.

2.1 Introduction

The generic IC is one of the fundamental building blocks in communication networks, in which the transmissions between each sender and its corresponding receiver (each sender-receiver pair) take place simultaneously and interfere with each other. The information-theoretic study of such a channel was initiated by Shannon [1], and has been continued by many others [5, 26, 44, 45, 27, 46, 31, 28, 29, 32, 30, 47, 33, 48, 34]. So far, the capacity region of the general IC remains unknown except for some special cases, such as the IC with strong interference (SIC) [26, 27, 28, 29, 30], a class of discrete additive degraded ICs [31], and a class of deterministic ICs [32]. However, various achievable rate regions serving as inner bounds on the capacity region have been derived for the general IC [46, 45, 28, 48].

Notably, Carleial [46] obtained an achievable rate region for the discrete memoryless IC by employing a limited form of the superposition coding scheme [3], *successive* encoding and decoding. Subsequently, Han and Kobayashi [28] established the best achievable rate region known to date by applying the superposition coding scheme comprising of *simultaneous* encoding and decoding. Indeed, the improvement of the Han-Kobayashi (HK) region [28] over the Carleial region [46] is primarily due to the use of the simultaneous decoding. This has been validated in [48, 34], in which Chong *et al.* obtained a so called Chong-Motani-Garg (CMG) rate region identical with the HK region but with a much simplified description, by using a hybrid of the successive encoding and simultaneous decoding. Moreover, Carleial [46] introduced the notion of the partial cross-observability of each sender's private information, which means that each receiver is able to decode part of the private information sent from its non-pairing sender. The derivation of the HK region and the CMG region followed this notion but Chong *et al.* have made the important observation that the decoding errors of the crossly observed information can be excluded in computing the probability of error [48]. With an introduction of the partial cross-observability, the IC can be viewed as a compound channel consisting of two associated MACs (strictly speaking, MAC-like channels), and thus its achievable rate region can be obtained by exploiting existing techniques used for MACs. However, the converse for either the HK region or the CMG region has not been established. Very recently, a notable variant of the IC, namely the

IC-DMS [43, 49, 50, 51, 52], has attracted considerable research attention due to its applicability to model certain realistic communication scenarios in cognitive radio networks or wireless sensor networks. From an information-theoretic viewpoint, the IC-DMS is fundamentally different from the IC since the capacity regions of IC and IC-DMS, if any, do not necessarily imply each other. In fact, we will also investigate this channel in Chapter 3.

Most of the prior work on the ICs assumes the statistical independence of the source messages [5, 26, 44, 45, 27, 46, 31, 29, 28, 32, 30, 47, 33, 48, 34]. However, the assumption becomes invalid in an IC where the senders need transmit not only the private information but also certain common information to their corresponding receivers. Such a scenario is generally modeled as the ICC [53, 54, 55]. The ICC was first studied by Tan in his original work [53], where inner and outer bounds on the capacity region have been derived. In particular, when no common information is present, the inner bound (the achievable rate region) in [53] reduces to the Carleial region in [26]. More recently, Maric *et al.* [54] derived the capacity region for a special case of the ICC, the strong interference channel with common information (SICC), and showed that the derived capacity result includes the capacity region of the strong interference channel (with no common information) [30] as a special case. Parallel to the case of the IC, the study of the ICC is closely related to the previous work on the MAC with common information (MACC) that has been thoroughly studied by Slepian and Wolf [7] and Willems [56]. As an example, an achievable rate region for the SICC is an intersection of the rate regions for its two corresponding MACCs, and the capacity region of the SICC is the union of all such achievable rate regions.

In this chapter, we begin with studying the *general* two-user ICC problem. We propose an encoding scheme that extends the idea of the Carleial's successive encoding for the ICC. With this encoding scheme, we allow the senders' common information to be conveyed through the channel in a cooperative manner. Exploiting the proposed encoding scheme along with the simultaneous decoding scheme

[28, 48], we derive a new achievable rate region for the discrete memoryless ICC. We show that the derived achievable rate region contains the one in [53] as a *proper* subregion under some specific setting, and reduces to the CMG region [48] as well as the capacity region of the SICC [54] in their respective channel settings. We further investigate a class of DICCs, which can be viewed as a generalization of the class of deterministic ICs in [32]. We show that under certain assumptions, our achievable rate region is the capacity region for this class of the DICCs.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the channel models. In Section 2.3, we present the achievable rate region for the general discrete memoryless ICC in both implicit and explicit forms. In Section 2.4, we discuss the relations between our achievable rate region and several existing results in [53, 54, 48, 57]. In Section 2.5, we investigate two special cases of the ICC. In Section 2.6, we extend our achievable rate region for the discrete memoryless ICC to the Gaussian case. Lastly, we conclude the chapter in Section 2.7.

2.2 Channel Models and Preliminaries

In this section, we present the channel models of the ICC, including the general ICC and a modified ICC. The modified ICC serves to reveal the information flow through its associated ICC, and facilitates the derivation of the achievable rate region for the associated ICC.

2.2.1 Discrete Memoryless Interference Channel With Common Information

A discrete memoryless IC is usually defined by a quintuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{P}, \mathcal{Y}_1, \mathcal{Y}_2)$, where $\mathcal{X}_t$ and $\mathcal{Y}_t$, $t = 1, 2$, denote the finite channel input and output alphabets respectively, and $\mathcal{P}$ denotes the collection of the conditional probabilities $p(y_1, y_2|x_1, x_2)$ on $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. The channel is memoryless in the sense that for n channel uses, we have

$$p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^{n} p(y_{1i}, y_{2i}|x_{1i}, x_{2i}),$$

where $\mathbf{x}_t := (x_{t1}, ..., x_{tn}) \in \mathcal{X}_t^n$ and $\mathbf{y}_t := (y_{t1}, ..., y_{tn}) \in \mathcal{Y}_t^n$ for $t = 1, 2$. The marginal distributions of y_1 and y_2 are given by

$$p_1(y_1|x_1, x_2) = \sum_{y_2 \in \mathcal{Y}_2} p(y_1, y_2|x_1, x_2),$$
$$p_2(y_2|x_1, x_2) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1, y_2|x_1, x_2).$$

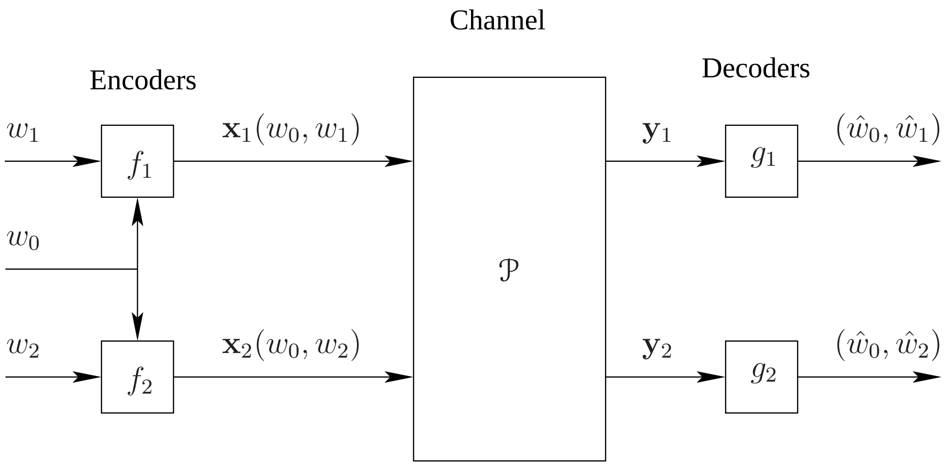

Figure 2.1: Interference channel with common information.

Building upon an IC, we depict an ICC in Fig. 2.1. Sender t, $t = 1, 2$, is to send a private message $w_t \in \mathcal{M}_t := \{1, 2, ..., M_t\}$ together with a common message $w_0 \in \mathcal{M}_0 := \{1, 2, ..., M_0\}$ to its pairing receiver. All the three messages are assumed to be independently and uniformly generated over their respective ranges.

Let C denote the discrete memoryless ICC defined above. An (M_0, M_1, M_2, n, P_e) code exists for the channel C, if and only if there exist two encoding functions

$$f_1 : \mathcal{M}_0 \times \mathcal{M}_1 \to \mathcal{X}_1^n, \quad f_2 : \mathcal{M}_0 \times \mathcal{M}_2 \to \mathcal{X}_2^n,$$

and two decoding functions

$$g_1 : \mathcal{Y}_1^n \rightarrow \mathcal{M}_0 \times \mathcal{M}_1, \quad g_2 : \mathcal{Y}_2^n \rightarrow \mathcal{M}_0 \times \mathcal{M}_2,$$

such that $\max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} \leq P_e$, where $P_{e,t}^{(n)}$, $t = 1, 2$, denotes the average decoding error probability of decoder t, and is computed by one of the following expressions:

$$P_{e,1}^{(n)} = \frac{1}{M_{\text{Prod}}} \sum_{w_0,w_1,w_2} \Pr((\hat{w}_0, \hat{w}_1) \neq (w_0, w_1)|(w_0, w_1, w_2)),$$
$$P_{e,2}^{(n)} = \frac{1}{M_{\text{Prod}}} \sum_{w_0,w_1,w_2} \Pr((\hat{w}_0, \hat{w}_2) \neq (w_0, w_2)|(w_0, w_1, w_2)),$$

where $M_{\text{Prod}} := M_0 M_1 M_2$.

A non-negative rate triple (R_0, R_1, R_2) is achievable for the channel C if for any given $0 < P_e < 1$, and for any sufficiently large n, there exists a $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e)$ code.

The capacity region for the channel C is defined as the closure of the set of all the achievable rate triples, while an achievable rate region for the channel C is a subset of the capacity region.

2.2.2 Modified Discrete Memoryless Interference Channel With Common Information

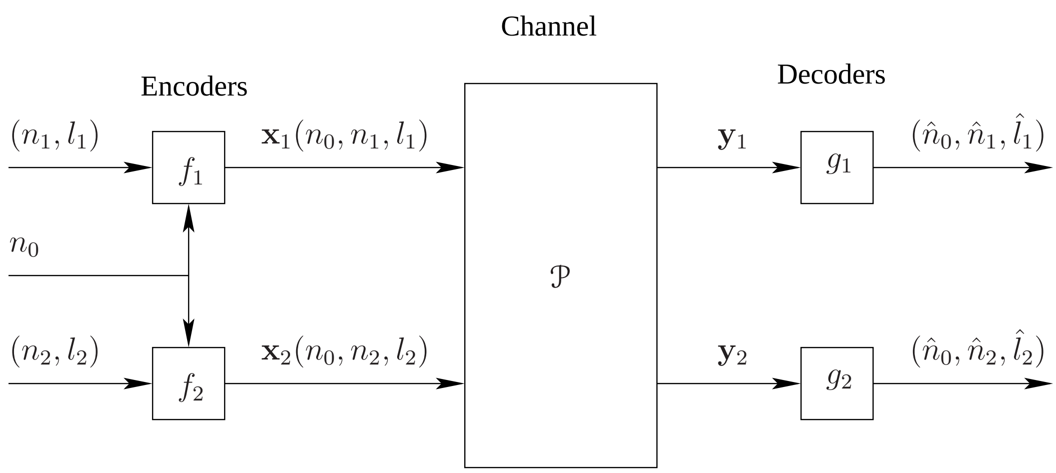

Figure 2.2: Modified interference channel with common information.

The modified ICC, as depicted in Fig. 2.2, inherits the same channel characteristics from its associated ICC, but it has five streams of messages instead of three in the associated ICC. The five streams of messages n_0, n_1, l_1, n_2, and l_2 are assumed to be independently and uniformly generated over the finite sets $\mathcal{N}_0 := \{1, ..., N_0\}$, $\mathcal{N}_1 := \{1, ..., N_1\}$, $\mathcal{L}_1 := \{1, ..., L_1\}$, $\mathcal{N}_2 := \{1, ..., N_2\}$, and $\mathcal{L}_2 := \{1, ..., L_2\}$, respectively. Denote the modified ICC by C_m.

An $(N_0, N_1, L_1, N_2, L_2, n, P_e)$ code exists for the channel C_m if and only if there exist two encoding functions

$$f_1 : \mathcal{N}_0 \times \mathcal{N}_1 \times \mathcal{L}_1 \to \mathcal{X}_1^n, \quad f_2 : \mathcal{N}_0 \times \mathcal{N}_2 \times \mathcal{L}_2 \to \mathcal{X}_2^n,$$

and two decoding functions

$$g_1 : \mathcal{Y}_1^n \to \mathcal{N}_0 \times \mathcal{N}_1 \times \mathcal{L}_1, \quad g_2 : \mathcal{Y}_2^n \to \mathcal{N}_0 \times \mathcal{N}_2 \times \mathcal{L}_2,$$

such that $\max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} \leq P_e$, where the average probabilities of decoding error denoted by $P_{e,1}^{(n)}$ and $P_{e,2}^{(n)}$ are computed as

$$P_{e,1}^{(n)} = \frac{1}{N_{\text{Prod}}} \sum_{n_0,n_1,l_1,n_2,l_2} \Pr((\hat{n}_0, \hat{n}_1, \hat{l}_1) \neq (n_0, n_1, l_1) | (n_0, n_1, l_1, n_2, l_2)),$$
$$P_{e,2}^{(n)} = \frac{1}{N_{\text{Prod}}} \sum_{n_0,n_1,l_1,n_2,l_2} \Pr((\hat{n}_0, \hat{n}_2, \hat{l}_2) \neq (n_0, n_2, l_2) | (n_0, n_1, l_1, n_2, l_2)),$$

where $N_{\text{Prod}} := N_0 N_1 L_1 N_2 L_2$.

A non-negative rate quintuple $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ is achievable for the channel C_m if for any given $0 < P_e < 1$ and any sufficiently large n, there exists a $(2^{nR_0}, 2^{nR_{12}}, 2^{nR_{11}}, 2^{nR_{21}}, 2^{nR_{22}}, n, P_e)$ code for the channel C_m.

Remark 2.1 It should be noted that compared with Fig. 2 in [28], our modified channel depicted in Fig. 2.2 does not include the index $\hat{n}_2$ (or $\hat{n}_1$) in the decoded message vector at decoder 1 (or decoder 2). This is due to the observation made in [48] that, although receiver 1 (or receiver 2) attempts to decode the crossly observ-

able private message n_2 *(or* n_1*), it is not necessary to include decoding errors of such information in calculating probability of error at the respective receiver. This is also the reason why we term the two associated channels of an ICC as MACC-like channels instead of MACCs.*

The following lemma is a straightforward consequence of the definitions of the rate triple (R_0, R_1, R_2) and the rate quintuple $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$.

Lemma 2.1 If $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ *is achievable for the channel* C_m*, then* $(R_0, R_{12}+R_{11}, R_{21}+R_{22})$ *is achievable for the associated ICC.*

Remark 2.2 With the aid of Lemma 2.1, an achievable rate region for the modified ICC can be easily extended to one for the associated ICC.

2.3 Discrete Memoryless ICC

In this section, we derive a new achievable rate region for the discrete memoryless ICC introduced in Section 2.2. The derived rate region is presented in both implicit and explicit forms.

2.3.1 An Achievable Rate Region for the Discrete Memoryless ICC

We first introduce three auxiliary random variables U_0, U_1, and U_2 that are defined over arbitrary finite sets $\mathcal{U}_0$, $\mathcal{U}_1$, and $\mathcal{U}_2$, respectively. Denote by $\mathcal{P}^*$ the set of all joint probability distributions $p(\cdot)$ that factor as

$$\begin{aligned} p(u_0, u_1, u_2, x_1, x_2, y_1, y_2) = {} & p(u_0)p(u_1|u_0)p(u_2|u_0) \\ & \cdot p(x_1|u_1, u_0)p(x_2|u_2, u_0)p(y_1, y_2|x_1, x_2). \end{aligned} \tag{2.1}$$

Let $\mathcal{R}_m(p)$ denote the set of all non-negative rate quintuples $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ such that

$$R_{11} \leq I(X_1; Y_1|U_0, U_1, U_2), \tag{2.2}$$

$$R_{12} + R_{11} \leq I(X_1; Y_1|U_0, U_2), \tag{2.3}$$

$$R_{11} + R_{21} \leq I(X_1, U_2; Y_1|U_0, U_1), \tag{2.4}$$

$$R_{12} + R_{11} + R_{21} \leq I(X_1, U_2; Y_1|U_0), \tag{2.5}$$

$$R_0 + R_{12} + R_{11} + R_{21} \leq I(U_0, X_1, U_2; Y_1); \tag{2.6}$$

$$R_{22} \leq I(X_2; Y_2|U_0, U_2, U_1), \tag{2.7}$$

$$R_{21} + R_{22} \leq I(X_2; Y_2|U_0, U_1), \tag{2.8}$$

$$R_{22} + R_{12} \leq I(X_2, U_1; Y_2|U_0, U_2), \tag{2.9}$$

$$R_{21} + R_{22} + R_{12} \leq I(X_2, U_1; Y_2|U_0), \tag{2.10}$$

$$R_0 + R_{21} + R_{22} + R_{12} \leq I(U_0, X_2, U_1; Y_2), \tag{2.11}$$

for some fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$. Note that each of the mutual information terms is computed with respect to the given fixed joint distribution.

Lemma 2.2 Any element $(R_0, R_{12}, R_{11}, R_{21}, R_{22}) \in \mathcal{R}_m(p)$ *is achievable for the modified ICC* C_m *for a fixed joint probability distribution* $p(\cdot) \in \mathcal{P}^*$.

Remark 2.3 The lengthy proof is relegated to Appendix A.1. Lemma 2.2 lays a foundation for us to establish an achievable rate region for the general ICC. One can interpret this achievable rate region as an intersection between the achievable rate regions of the two associated MACC-like channels. Specifically, inequalities (2.2)–(2.6) *depict an achievable rate region for one MACC-like channel, and inequalities* (2.7)–(2.11) *depict one for the other.*

Theorem 2.1 The rate region $\mathcal{R}_m$ *is achievable for the channel* C_m *with*

$$\mathcal{R}_m := \bigcup_{p(\cdot) \in \mathcal{P}^*} \mathcal{R}_m(p).$$

Remark 2.4 Theorem 2.1 is a direct extension of Lemma 2.2. The proof is straightforward and thus omitted. Note that the rate region $\mathcal{R}_m$ *is convex, and therefore no convex hull operation or time sharing is necessary. The proof of the convexity is given in Appendix A.2.*

Let us fix a joint distribution $p(\cdot) \in \mathcal{P}^*$, and denote by $\mathcal{R}_{\text{impl}}(p)$ the set of all the non-negative rate triples (R_0, R_1, R_2) such that $R_1 = R_{12} + R_{11}$ and $R_2 = R_{21} + R_{22}$ for some $(R_0, R_{12}, R_{11}, R_{21}, R_{22}) \in \mathcal{R}_m(p)$.

Theorem 2.2 $\mathcal{R}_{impl}$ *is an achievable rate region for the channel C with*

$$\mathcal{R}_{impl} := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_{impl}(p).$$

Proof: It suffices to prove that $\mathcal{R}_{\text{impl}}(p)$ is an achievable rate region for C for any fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$, while the achievability of any rate triple $(R_0, R_1, R_2) \in \mathcal{R}_{\text{impl}}(p)$ follows immediately from Lemma 2.1 and Lemma 2.2. ■

Remark 2.5 The main idea, as mentioned before, is that we allow the common information (of rate R_0*) to be cooperatively transmitted by the two senders, on top of which we treat the private information at each sender as two parts. One part (of rate* R_{12} *or* R_{21}*) of the private information at each sender is crossly observable to the non-pairing receiver, but not the other part (of rate* R_{11} *or* R_{22}*)*[1]*. However, for each receiver, the crossly observed information is not required to be decoded correctly [48]. Details can be found in the proof of Lemma 2.2 in Appendix A.1.*

[1] After finishing the work in this chapter, we learned of independent work by Cao *et al.* [58]. The achievable rate region in [58] is essentially the same as ours, even though, compared with the one presented in [58], the description of our achievable rate region is more compact in view of the number of constraints involved.

Remark 2.6 One can observe that the rate of the common information, R_0, is bounded by only one inequality at each decoder. This is similar to the case of the MACC [7, 56], where the rate of the common information is bounded by only one inequality. This is due to the perfect cooperation of the two senders in transmitting the common information, and the simultaneous decoding. Details are illustrated in the proof of Lemma 2.2.

Remark 2.7 The region $\mathcal{R}_{impl}$ is also convex, which can be proven by following the same procedure in the proof of the convexity of $\mathcal{R}_m$ in Appendix A.2.

2.3.2 Explicit Description of the Achievable Rate Region

In order to reveal the geometric shape of the region $\mathcal{R}_{\text{impl}}$ depicted in Theorem 2.2, we derive an explicit description of the region by applying Fourier-Motzkin elimination [59, 48, 57].

Let $\mathcal{R}(p)$ denote the set of all non-negative rate triples (R_0, R_1, R_2) such that

$$R_1 \le I(X_1; Y_1|U_0, U_2), \tag{2.12}$$

$$R_2 \le I(X_2; Y_2|U_0, U_1), \tag{2.13}$$

$$R_0 + R_1 \le I(U_0, X_1, U_2; Y_1), \tag{2.14}$$

$$R_0 + R_2 \le I(U_0, X_2, U_1; Y_2), \tag{2.15}$$

$$R_1 + R_2 \le I(X_1, U_2; Y_1|U_0, U_1) + I(X_2, U_1; Y_2|U_0, U_2), \tag{2.16}$$

$$R_1 + R_2 \le I(X_1; Y_1|U_0, U_1, U_2) + I(X_2, U_1; Y_2|U_0), \tag{2.17}$$

$$R_0 + R_1 + R_2 \le I(X_1; Y_1|U_0, U_1, U_2) + I(U_0, X_2, U_1; Y_2), \tag{2.18}$$

$$R_1 + R_2 \le I(X_2; Y_2|U_0, U_1, U_2) + I(X_1, U_2; Y_1|U_0), \tag{2.19}$$

$$R_0 + R_1 + R_2 \le I(X_2; Y_2|U_0, U_1, U_2) + I(U_0, X_1, U_2; Y_1), \tag{2.20}$$

$$2R_1 + R_2 \le I(X_1; Y_1|U_0, U_1, U_2) + I(X_1, U_2; Y_1|U_0) + I(X_2, U_1; Y_2|U_0, U_2), \tag{2.21}$$

$$R_0 + 2R_1 + R_2 \leq I(X_1; Y_1|U_0, U_1, U_2) + I(U_0, X_1, U_2; Y_1) + I(X_2, U_1; Y_2|U_0, U_2), \tag{2.22}$$

$$R_1 + 2R_2 \leq I(X_2; Y_2|U_0, U_1, U_2) + I(X_2, U_1; Y_2|U_0) + I(X_1, U_2; Y_1|U_0, U_1), \tag{2.23}$$

$$R_0 + R_1 + 2R_2 \leq I(X_2; Y_2|U_0, U_1, U_2) + I(U_0, X_2, U_1; Y_2) + I(X_1, U_2; Y_1|U_0, U_1), \tag{2.24}$$

for some fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, and define $\mathcal{R} := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p)$.

Corollary 2.1 The rate region $\mathcal{R}$ is achievable for the channel C, and $\mathcal{R} = \mathcal{R}_{impl}$.

Remark 2.8 The proof of this corollary is given in Appendix A.3. In fact, the explicit rate region obtained by applying Fourier-Motzkin elimination on (2.2)–(2.11) *contains two extra constraints:*

$$R_1 \leq I(X_1; Y_1|U_0, U_1, U_2) + I(X_2, U_1; Y_2|U_0, U_2),$$
$$R_2 \leq I(X_2; Y_2|U_0, U_1, U_2) + I(X_1, U_2; Y_1|U_0, U_1).$$

However, these two constraints are redundant and thus are excluded. This is shown in the second part of Appendix A.3 by applying the technique introduced in [34].

The close tie between the explicit CMG region and the capacity region of a class of deterministic ICs in [32] was pointed out in [59]. Similarly, we will disclose that the explicit region for the ICC is also closely related to the capacity region of a class of DICCs investigated in Section 2.5.2.

2.4 Relations between $\mathcal{R}_{\mathbf{impl}}$ and Some Existing Results

In this section, we discuss the relations between the achievable rate region derived in the preceding section and several previously known results [53][54][48].

2.4.1 Achievable Rate Region for the ICC by Tan

We show that the achievable rate region $\mathcal{R}_{\text{impl}}$ includes the one given in [53, Theorem 1] as a subregion. Note that a similar result is presented in [58, Corollary 1].

Let $\mathcal{P}^*_{\text{Tan}}$ denote the set of all the joint distributions $p(\cdot)$ that factors as

$$p(u_0,u_1,u_2,x_1,x_2,y_1,y_2) = p(u_0)p(u_1|u_0)p(u_2|u_0)p(x_1|u_1)p(x_2|u_2)p(y_1,y_2|x_1,x_2).$$

Let $\mathcal{R}^i_{\text{Tan}}(p)$, $i = 1,2,3,4$, denote the set of all non-negative rate triples (R_0, R_1, R_2) satisfying

$$\begin{aligned}
R_1 &\le I(X_1;Y_1|U_1,U_2) + s_i,\\
R_2 &\le I(X_2;Y_2|U_1,U_2) + t_i,\\
R_0 + R_1 + \frac{t_i}{I(X_2;Y_2|U_1,U_2)+t_i} R_2 &\le I(U_2,X_1;Y_1),\\
R_0 + R_2 + \frac{s_i}{I(X_1;Y_1|U_1,U_2)+s_i} R_1 &\le I(U_1,X_2;Y_2);
\end{aligned}$$

where s_i and t_i are computed as

$$\begin{aligned}
s_1 &= \min\{I(U_1;Y_1|U_0), I(U_1;Y_2|U_0)\},\\
t_1 &= \min\{I(U_2;Y_1|U_0,U_1), I(U_2;Y_2|U_0,U_1)\},\\
s_2 &= \min\{I(U_1;Y_1|U_0,U_2), I(U_1;Y_2|U_0,U_2)\},\\
t_2 &= \min\{I(U_2;Y_1|U_0), I(U_2;Y_2|U_0)\},\\
s_3 &= \min\{I(U_1;Y_1|U_0), I(U_1;Y_2|U_0,U_2)\},\\
t_3 &= \min\{I(U_2;Y_1|U_0,U_1), I(U_2;Y_2|U_0)\},\\
s_4 &= \min\{I(U_1;Y_1|U_0,U_2), I(U_1;Y_2|U_0)\},\\
t_4 &= \min\{I(U_2;Y_1|U_0), I(U_2;Y_2|U_0,U_1)\},
\end{aligned}$$

for a joint distribution $p(\cdot) \in \mathcal{P}^*_{\text{Tan}}$.

Denote the closed convex hull operation[1] by $\mathrm{co}(\cdot)$, and define

$$\mathcal{R}_{\mathrm{Tan}}(p) := \mathrm{co}\left(\bigcup_{i=1}^{4}\mathcal{R}^{i}_{\mathrm{Tan}}(p)\right).$$

In the following, we restate the achievable result obtained by Tan [53, Theorem 1], and further show that our achievable rate region includes this result as a subregion.

Corollary 2.2 ([53, Theorem 1]) Any rate triple

$$(R_0, R_1, R_2) \in \mathcal{R}_{Tan} := \bigcup_{p(\cdot)\in\mathcal{P}^*_{Tan}} \mathcal{R}_{Tan}(p)$$

is achievable for the ICC, i.e., $\mathcal{R}_{Tan} \subseteq \mathcal{R}_{impl}$.

Proof: It suffices to show that each $\mathcal{R}^{i}_{\mathrm{Tan}}(p)$, $i = 1, 2, 3, 4$, is achievable for any joint distribution $p(\cdot) \in \mathcal{P}^*_{\mathrm{Tan}}$. Let $\mathcal{R}^{i}_{\mathrm{sub}}(p)$ be the set of all rate triples (R_0, R_1, R_2) such that $R_1 = R_{12} + R_{11}$ and $R_2 = R_{21} + R_{22}$ with non-negative rate quadruples $(R_{12}, R_{11}, R_{21}, R_{22})$ satisfying

$$\begin{aligned}
R_{11} &\leq I(X_1; Y_1|U_0, U_1, U_2),\\
R_{22} &\leq I(X_2; Y_2|U_0, U_1, U_2),\\
R_{12} &\leq s_i,\\
R_{21} &\leq t_i,\\
R_0 + R_{12} + R_{11} + R_{21} &\leq I(U_0, X_1, U_2; Y_1),\\
R_0 + R_{21} + R_{22} + R_{12} &\leq I(U_0, X_2, U_1; Y_2),
\end{aligned}$$

for a joint distribution $p(\cdot) \in \mathcal{P}^*$.

It is easy to check that for each $i \in \{1, 2, 3, 4\}$, the rate region $\mathcal{R}^{i}_{\mathrm{sub}}(p)$ is a subset of our achievable rate region $\mathcal{R}_{\mathrm{impl}}(p)$. Note that $\mathcal{P}^*_{\mathrm{Tan}} \subseteq \mathcal{P}^*$. For a distribution $p(\cdot) \in \mathcal{P}^*_{\mathrm{Tan}}$, the rate region $\mathcal{R}^{i}_{\mathrm{sub}}(p)$ reduces to the region with $(R_{12}, R_{11}, R_{21}, R_{22})$

[1] The convex hull of a set $\mathcal{S}$ can be described constructively as the set of convex combinations of finite subsets of points from $\mathcal{S}$.

satisfying

$$\begin{aligned}
R_{11} &\leq I(X_1; Y_1|U_1, U_2),\\
R_{22} &\leq I(X_2; Y_2|U_1, U_2),\\
R_{12} &\leq s_i,\\
R_{21} &\leq t_i,\\
R_0 + R_{12} + R_{11} + R_{21} &\leq I(X_1, U_2; Y_1),\\
R_0 + R_{21} + R_{22} + R_{12} &\leq I(X_2, U_1; Y_2).
\end{aligned}$$

This is due to the fact that $p(\cdot) \in \mathcal{P}^*_{\text{Tan}}$ induces a Markov chain $U_0 \to (U_1, U_2) \to (X_1, X_2) \to (Y_1, Y_2)$. It is now clear that $\mathcal{R}^i_{\text{sub}}(p) = \mathcal{R}^i_{\text{Tan}}(p)$ for any joint distribution $p(\cdot) \in \mathcal{P}^*_{\text{Tan}}$. Therefore, the rate region $\mathcal{R}_{\text{Tan}}$ is achievable, and $\mathcal{R}_{\text{Tan}} \subseteq \mathcal{R}_{\text{impl}}$. ■

It should be noted that the corollary does not indicate that the inclusion, $\mathcal{R}_{\text{Tan}} \subseteq \mathcal{R}_{\text{impl}}$, is strict. Whether this inclusion is strict deserves further investigation. However under some specific setting, the region $\mathcal{R}_{\text{impl}}$ strictly contains $\mathcal{R}_{\text{Tan}}$, which can be justified as follows. In the case of no common information, $\mathcal{R}_{\text{Tan}}(p)$ reduces to $\mathcal{R}_0(Z)$ in Corollary 3.1 of [28], while $\mathcal{R}_{\text{impl}}(p)$ reduces to the CMG region (or the HK region). When the channel is Gaussian and the time-sharing variable is fixed as a constant, the HK region demonstrates strict inclusion over $\mathcal{R}_0(Z)$ in [28]. We will show that under this setting, $\mathcal{R}_{\text{impl}}$ improves $\mathcal{R}_{\text{Tan}}$ similarly in Section 2.6.2.

2.4.2 Strong Interference Channel With Common Information

Let $\mathcal{P}_s$ denote the set of all joint distributions $p(u_0, x_1, x_2, y_1, y_2)$ that factor as

$$p(u_0, x_1, x_2, y_1, y_2) = p(u_0)p(x_1|u_0)p(x_2|u_0)p(y_1, y_2|x_1, x_2).$$

As defined in [54], an ICC is considered as a SICC if

$$I(X_1; Y_1|X_2, U_0) \leq I(X_1; Y_2|X_2, U_0),$$
$$I(X_2; Y_2|X_1, U_0) \leq I(X_2; Y_1|X_1, U_0),$$

for all joint probability distributions $p(\cdot) \in \mathcal{P}_s$.

Let $\mathcal{R}_s(p)$ denote the set of all non-negative rate triples (R_0, R_1, R_2) such that

$$R_1 \leq I(X_1; Y_1|X_2, U_0), \tag{2.25}$$
$$R_2 \leq I(X_2; Y_2|X_1, U_0), \tag{2.26}$$
$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_1|U_0), I(X_2, X_1; Y_2|U_0)\}, \tag{2.27}$$
$$R_0 + R_1 + R_2 \leq \min\{I(X_1, X_2; Y_1), I(X_2, X_1; Y_2)\}, \tag{2.28}$$

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

Corollary 2.3 ([54, Achievability of Theorem 1]) Any rate triple

$$(R_0, R_1, R_2) \in \bigcup_{p(\cdot)\in\mathcal{P}_s} \mathcal{R}_s(p)$$

is achievable for the SICC.

Remark 2.9 By setting $U_t = X_t$, $t = 1, 2$, and $R_{11} = R_{22} = 0$ in (2.2)–(2.11), *and removing two redundant ones from the resulting inequalities due to the channel assumptions of the SICC, we can easily obtain* (2.25)–(2.28).

Remark 2.10 By letting $U_t = X_t$, $t = 1, 2$, we treat the private information at each sender as a whole instead of two parts. This differs from what was mentioned earlier in Remark 2.5. In this case the full private information at each sender is allowed to be crossly observed by the respective non-pairing receivers due to the strong interference.

2.4.3 Interference Channel Without Common Information

We now consider the general IC (without common information) as a special case of the ICC, and demonstrate that our achievable rate region for the ICC reduces to the CMG region [48] for the IC.

Let Q denote the time-sharing random variable, and $\mathcal{P}_o$ denote the set of all joint distributions that factor as

$$\begin{aligned} p(q, u_1, u_2, x_1, x_2, y_1, y_2) &= p(q)p(u_1|q)p(u_2|q) \\ &\quad \cdot p(x_1|u_1, q)p(x_2|u_2, q)p(y_1, y_2|x_1, x_2). \end{aligned}$$

Define $\mathcal{R}_o(p)$ as the set of all rate pairs (R_1, R_2) such that $R_1 = R_{12} + R_{11}$ and $R_2 = R_{21} + R_{22}$ with any non-negative rate quintuple $(R_{12}, R_{11}, R_{21}, R_{22})$ satisfying

$$R_{11} \leq I(X_1; Y_1|U_1, U_2, Q), \tag{2.29}$$

$$R_{12} + R_{11} \leq I(X_1; Y_1|U_2, Q), \tag{2.30}$$

$$R_{11} + R_{21} \leq I(X_1, U_2; Y_1|U_1 Q), \tag{2.31}$$

$$R_{12} + R_{11} + R_{21} \leq I(X_1, U_2; Y_1|Q); \tag{2.32}$$

$$R_{22} \leq I(X_2; Y_2|U_2, U_1, Q), \tag{2.33}$$

$$R_{21} + R_{22} \leq I(X_2; Y_2|U_1, Q), \tag{2.34}$$

$$R_{22} + R_{12} \leq I(X_2, U_1; Y_2|U_2, Q), \tag{2.35}$$

$$R_{21} + R_{22} + R_{12} \leq I(X_2, U_1; Y_2|Q), \tag{2.36}$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}_o$, and define $\mathcal{R}_o := \bigcup_{p(\cdot)\in\mathcal{P}_o} \mathcal{R}_o(p)$.

Corollary 2.4 ([48, Theorem 3]) $\mathcal{R}_o$ *is an achievable rate region for the IC.*

Remark 2.11 Since no common information is involved, we can set $U_0 = Q$ *and* $R_0 = 0$ *in* (2.2)–(2.11)*, and obtain* (2.29)–(2.36)*. On the other hand, one can readily obtain the explicit CMG region ([57, Theorem D] and [48, Theorem 4]) by setting* $U_0 = Q$ *and* $R_0 = 0$ *in* (2.12)–(2.24)*.*

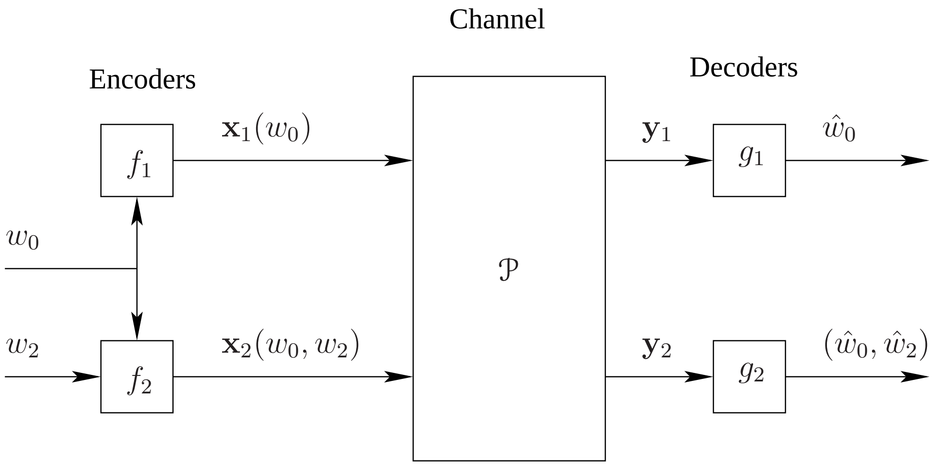

Figure 2.3: Asymmetric interference channel with common information.

2.5 Two Special Cases of the ICC

In this section, we specialize our achievable results in Section 2.3 to the following two cases.

2.5.1 Asymmetric Interference Channel With Common Information

We first introduce the channel model of this class of the ICCs, namely the asymmetric interference channel with common information (AICC), where one sender does not have private information to transmit. Without loss of generality, we assume that sender 1 only has the common message w_0 to be transmitted to receiver 1, while sender 2 needs transmit both the common message w_0 and the private message w_2 to receiver 2. Fig. 2.3 depicts the channel model for the AICC, which we denote by C_a. We follow the definitions introduced in Section 2.2, and define the capacity region of the channel C_a as the set of all achievable rate pairs (R_0, R_2) for this channel.

Let $\mathcal{P}_a$ denote the set of all joint distributions that factor as

$$p(u_0, x_1, u_2, x_2, y_1, y_2) = p(u_0, x_1)p(u_2, x_2|u_0)p(y_1, y_2|x_1, x_2).$$

Theorem 2.3 $\mathcal{R}_a := \bigcup_{p(\cdot)\in\mathcal{P}_a} \mathcal{R}_a(p)$ *is an achievable rate region for the channel* C_a*, where* $\mathcal{R}_a(p)$ *is the set of all non-negative rate pairs* (R_0, R_2) *such that*

$$\begin{aligned} R_0 &\leq I(U_0, U_2; Y_1), \\ R_2 &\leq \min\{I(X_2; Y_2|U_0), I(U_2; Y_1|U_0) + I(X_2; Y_2|U_2, U_0)\}, \\ R_0 + R_2 &\leq \min\{I(U_0, X_2; Y_2), I(U_0, U_2; Y_1) + I(X_2; Y_2|U_2, U_0)\}, \end{aligned}$$

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

Remark 2.12 1) It is straightforward to obtain Theorem 2.3 from Corollary 2.1 by letting $R_1 = 0$*,* $U_1 = U_0$*, and* $X_1 = U_0$*. 2) The coding strategy for this channel remains basically the same as the one for the general ICC: both senders first need cooperate to transmit the common information, while sender 2 treats the private information as two parts, of which only one part is crossly observable to receiver 1. 3) Although the description of this rate region appears simple, establishing the converse is still extremely difficult.*

In addition, by letting $U_0 = X_1$ and $U_2 = X_2$, the rate region $\mathcal{R}_a$ reduces to the capacity region for the strong interference channel with unidirectional cooperation [60, 52].

2.5.2 Deterministic Interference Channel With Common Information

We next investigate a class of discrete memoryless DICCs as depicted in Fig. 2.4. The major attributes of the DICCs remain the same as those of an ICC, i.e., the source messages (w_0, w_1, w_2), the channel input and output alphabets $\mathcal{X}_t$ and $\mathcal{Y}_t$, $t = 1, 2$, the encoding functions ($f_1(\cdot)$ and $f_2(\cdot)$) and decoding functions ($g_1(\cdot)$ and $g_2(\cdot)$), the existence of codes, and the achievable rates are defined in the same way as those for the general ICC. The distinction lies in the channel transition, which

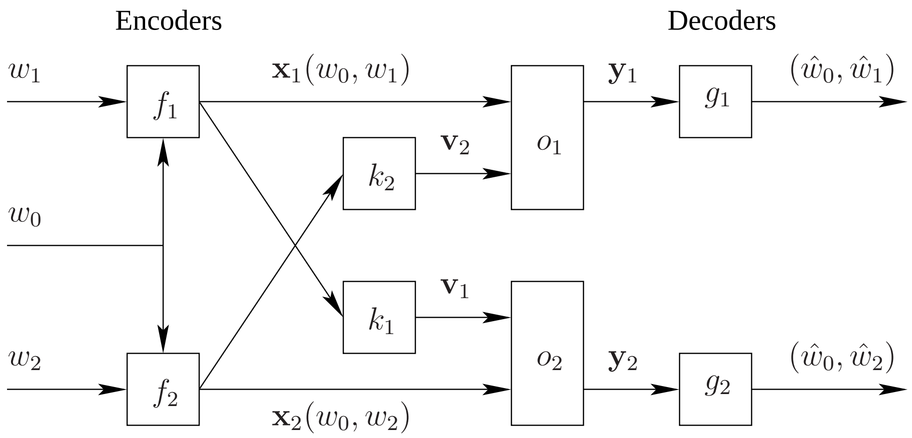

Figure 2.4: A class of deterministic interference channels with common information.

is governed by the following deterministic functions:

$$V_t = k_t(X_t), \quad t = 1, 2;$$

$$Y_1 = o_1(X_1, V_2), \text{ and } Y_2 = o_2(X_2, V_1),$$

where V_1 and V_2 represent the interference signals caused by X_1 and X_2 at the corresponding receivers. Furthermore, we assume that there exist two more deterministic functions, $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. We denote this class of DICCs by C_d.

The channel defined above is similar to the one investigated in [32], but there is a slight difference. In [32], it is assumed that $H(Y_1|X_1) = H(V_2)$ and $H(Y_2|X_2) = H(V_1)$ for all product distributions of X_1X_2. It has also been pointed out in [32] that this assumption is equivalent to assuming the existence of $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. Nevertheless, we assume the latter rather than the former since the former is not satisfied in our case. We will demonstrate that $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$ are the actual governing conditions for this class of DICCs.

Let $\mathcal{P}_d$ denote the set of all joint distributions $p(\cdot)$ that factor as

$$p(v_0, x_1, x_2) = p(v_0)p(x_1|v_0)p(x_2|v_0), \tag{2.37}$$

where v_0 is the realization of an auxiliary random variable V_0 defined over an arbitrary finite set $\mathcal{V}_0$. Let $\mathcal{R}_d(p)$ denote the set of all non-negative rate triples (R_0, R_1, R_2) such that

$$R_1 \leq H(Y_1|V_0, V_2), \tag{2.38}$$

$$R_2 \leq H(Y_2|V_0, V_1), \tag{2.39}$$

$$R_0 + R_1 \leq H(Y_1), \tag{2.40}$$

$$R_0 + R_2 \leq H(Y_2), \tag{2.41}$$

$$R_1 + R_2 \leq H(Y_1|V_0, V_1) + H(Y_2|V_0, V_2); \tag{2.42}$$

$$R_1 + R_2 \leq H(Y_1|V_0) + H(Y_2|V_0, V_1, V_2), \tag{2.43}$$

$$R_0 + R_1 + R_2 \leq H(Y_1) + H(Y_2|V_0, V_1, V_2); \tag{2.44}$$

$$R_1 + R_2 \leq H(Y_1|V_0, V_1, V_2) + H(Y_2|V_0), \tag{2.45}$$

$$R_0 + R_1 + R_2 \leq H(Y_1|V_0, V_1, V_2) + H(Y_2); \tag{2.46}$$

$$2R_1 + R_2 \leq H(Y_1|V_0) + H(Y_1|V_0, V_1, V_2) + H(Y_2|V_0, V_2), \tag{2.47}$$

$$R_0 + 2R_1 + R_2 \leq H(Y_1) + H(Y_1|V_0, V_1, V_2) + H(Y_2|V_0, V_2); \tag{2.48}$$

$$R_1 + 2R_2 \leq H(Y_2|V_0) + H(Y_2|V_0, V_1, V_2) + H(Y_1|V_0, V_1), \tag{2.49}$$

$$R_0 + R_1 + 2R_2 \leq H(Y_2) + H(Y_2|V_0, V_1, V_2) + H(Y_1|V_0, V_1), \tag{2.50}$$

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

Theorem 2.4 The capacity region of the channel C_d is the closure of $\bigcup_{p(\cdot)\in\mathcal{P}_d} \mathcal{R}_d(p)$.

Proof: 1) [**Achievability**.] It suffices to show that $\mathcal{R}_d(p)$ is achievable for the channel C_d for a fixed joint distribution $p(\cdot) \in \mathcal{P}_d$. As the joint distribution $p(\cdot) \in \mathcal{P}_d$ does not involve V_1 and V_2, it appears difficult to directly apply the superposition coding strategy developed for the general ICC to this channel. Nevertheless, because the interferences V_1 and V_2 are determined by the channel inputs X_1 and X_2, we can extend the joint distribution in the form of (2.37) to one containing V_1 and

V_2 as

$$p(v_0,x_1,x_2,v_1,v_2) = p(v_0)p(x_1|v_0)p(x_2|v_0)\delta(v_1 - k_1(x_1))\delta(v_2 - k_1(x_2)), \quad (2.51)$$

where $\delta(\cdot)$ is the Kronecker delta function. Since X_1 and X_2 are conditionally independent given V_0, the interferences V_1 and V_2 also become conditionally independent given V_0. Therefore, the extended joint distribution (2.51) can be factored as

$$p(v_0,x_1,x_2,v_1,v_2) = p(v_0)p(v_1|v_0)p(v_2|v_0)p(x_1|v_1,v_0)p(x_2|v_2,v_0),$$

and the achievability of the region $\mathcal{R}_d(p)$ follows readily from Corollary 2.1.

2) [**Converse**.] We first prove that for any non-deterministic (stochastic) $(M_0, M_1, M_2, n, P_e^*)$ code for the channel, there exists a deterministic (M_0, M_1, M_2, n, P_e) code such that $P_e \leq P_e^*$. We then upper bound the rates of any deterministic code having $P_e \to 0$ as $n \to \infty$. The detailed steps of the derivations are presented in Appendix A.4.

The upper bound meets with the inner bound (or the achievable rate region), and thus the theorem follows. ■

As mentioned earlier, we assume that there exist $h_1(\cdot,\cdot)$ and $h_2(\cdot,\cdot)$ such that $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. With this assumption, we have two equalities, $H(V_2^n|W_0) = H(Y_1^n|W_0, W_1)$ and $H(V_1^n|W_0) = H(Y_2^n|W_0, W_2)$. As can be observed from the converse part of the proof in Appendix A.4, these two equalities are crucial for us to establish the converse. Moreover, in the absence of common information our assumptions reduce to those made in [32]. In this sense, our assumption is slightly more general compared with the one made in [32]. It is also noteworthy that in the case of no common information, the capacity region of this class of DICCs reduces to the one obtained in [32].

2.6 Gaussian Interference Channel With Common Information

In this section, we show how to extend the achievable rate region $\mathcal{R}$ derived for the discrete memoryless ICC to the Gaussian case. We also present a numerical example to illustrate to what extent our region improves the Tan region in [53, Theorem 1].

2.6.1 Channel Model for the Gaussian ICC

We consider a Gaussian ICC (GICC) in standard form since any GICC can be transformed to one in standard form with the capacity region unchanged [46, 59, 49]. As depicted in Fig. 2.5, a GICC in standard form can be mathematically expressed as

$$Y_1 = X_1 + \sqrt{c_{21}} X_2 + Z_1, \tag{2.52}$$

$$Y_2 = X_2 + \sqrt{c_{12}} X_1 + Z_2, \tag{2.53}$$

where Z_i, $i = 1, 2$, is the additive white Gaussian noise with zero mean and unit variance, and $\sqrt{c_{21}}$ and $\sqrt{c_{12}}$ are the normalized channel gains of the respective interference links.

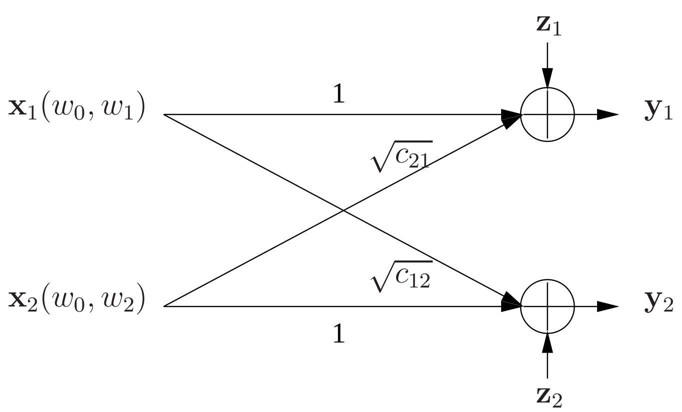

Figure 2.5: Gaussian interference channel with common information.

In addition, the codewords used for this channel are subject to the average power constraint given by $\sum_{t=1}^{n} \|x_{it}\|^2/n \leq P_i,\ i = 1, 2$. We only consider Gaussian codewords X_i^n, $i = 1, 2$, for the GICC, since it is shown in the Maximum-Entropy Theorem [40] that Gaussian inputs are optimal for Gaussian channels. Furthermore, we also fix the time-sharing random variable Q as a constant. Regarding how the choice of Q affects the size of the achievable rate region, interested readers are referred to [47] for a detailed exposition.

2.6.2 An Achievable Rate Region for the GICC

We first define the following mappings of random variables with respect to the joint probability distribution (2.1):

M1) W_0, W_{12}, W_{11}, W_{21}, and W_{22}, distributed according to $\mathcal{N}(0, 1)$,

M2) $X_1 = \sqrt{\alpha_1 P_1} W_0 + \sqrt{\bar{\alpha}_1 \beta_1 P_1} W_{12} + \sqrt{\bar{\alpha}_1 \bar{\beta}_1 P_1} W_{11}$,

M3) $X_2 = \sqrt{\alpha_2 P_2} W_0 + \sqrt{\bar{\alpha}_2 \beta_2 P_2} W_{21} + \sqrt{\bar{\alpha}_2 \bar{\beta}_2 P_2} W_{22}$,

M4) $U_0 = (\sqrt{\alpha_1 P_1} + \sqrt{\alpha_2 P_2}) W_0$,

M5) $U_1 = \sqrt{\alpha_1 P_1} W_0 + \sqrt{\bar{\alpha}_1 \beta_1 P_1} W_{12}$,

M6) $U_2 = \sqrt{\alpha_2 P_2} W_0 + \sqrt{\bar{\alpha}_2 \beta_2 P_2} W_{21}$,

where $\alpha_i, \beta_i \in [0, 1]$, $\bar{\alpha}_i = 1 - \alpha_i$, and $\bar{\beta}_i = 1 - \beta_i$, for $i = 1, 2$.

Based on these mappings, and the channel model described by (2.52) and (2.53), we obtain

$$
\begin{aligned}
Y_1 = {} & (\sqrt{\alpha_1 P_1} + \sqrt{c_{21} \alpha_2 P_2}) W_0 + \sqrt{\bar{\alpha}_1 \beta_1 P_1} W_{12} + \sqrt{\bar{\alpha}_1 \bar{\beta}_1 P_1} W_{11} + \sqrt{c_{21} \bar{\alpha}_2 \beta_2 P_2} W_{21} \\
& + \sqrt{c_{21} \bar{\alpha}_2 \bar{\beta}_2 P_2} W_{22} + Z_1, \qquad (2.54) \\
Y_2 = {} & (\sqrt{\alpha_2 P_2} + \sqrt{c_{12} \alpha_1 P_1}) W_0 + \sqrt{\bar{\alpha}_2 \beta_2 P_2} W_{21} + \sqrt{\bar{\alpha}_2 \bar{\beta}_2 P_2} W_{22} + \sqrt{c_{12} \bar{\alpha}_1 \beta_1 P_1} W_{12} \\
& + \sqrt{c_{12} \bar{\alpha}_1 \bar{\beta}_1 P_1} W_{11} + Z_2. \qquad (2.55)
\end{aligned}
$$

With the relations between the random variables defined by the mappings, M1–M6, and the channel input-output relations described by (2.54) and (2.55), we evaluate the mutual information terms $I(\cdot)$ in (2.12)–(2.24) as follows:

$$I(X_1; Y_1|U_0, U_2) = \gamma(\bar{\alpha}_1 P_1/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)), \tag{2.56}$$

$$I(X_2; Y_2|U_0, U_1) = \gamma(\bar{\alpha}_2 P_2/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)), \tag{2.57}$$

$$I(U_0, X_1, U_2; Y_1) = \gamma\left(\frac{(\sqrt{\alpha_1 P_1} + \sqrt{c_{21}\alpha_2 P_2})^2 + \bar{\alpha}_1 P_1 + c_{21}\bar{\alpha}_2 P_2}{c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1}\right), \tag{2.58}$$

$$I(U_0, X_2, U_1; Y_2) = \gamma\left(\frac{(\sqrt{\alpha_2 P_2} + \sqrt{c_{12}\alpha_1 P_1})^2 + \bar{\alpha}_2 P_2 + c_{12}\bar{\alpha}_1 P_1}{c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1}\right), \tag{2.59}$$

$$I(X_1, U_2; Y_1|U_0, U_1) = \gamma((\bar{\alpha}_1\bar{\beta}_1 P_1 + c_{21}\bar{\alpha}_2\beta_2 P_2)/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)), \tag{2.60}$$

$$I(X_2, U_1; Y_2|U_0, U_2) = \gamma((\bar{\alpha}_2\bar{\beta}_2 P_2 + c_{12}\bar{\alpha}_1\beta_1 P_1)/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)), \tag{2.61}$$

$$I(X_1; Y_1|U_0, U_1, U_2) = \gamma(\bar{\alpha}_1\bar{\beta}_1 P_1/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)), \tag{2.62}$$

$$I(X_2; Y_2|U_0, U_1, U_2) = \gamma(\bar{\alpha}_2\bar{\beta}_2 P_2/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)), \tag{2.63}$$

$$I(X_1, U_2; Y_1|U_0) = \gamma((\bar{\alpha}_1 P_1 + c_{21}\bar{\alpha}_2\beta_2 P_2)/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)), \tag{2.64}$$

$$I(X_2, U_1; Y_2|U_0) = \gamma((\bar{\alpha}_2 P_2 + c_{12}\bar{\alpha}_1\beta_1 P_1)/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)), \tag{2.65}$$

where $\gamma(x) := \frac{1}{2}\log_2(1 + x)$.

Replacing each mutual information term in (2.12)–(2.24) with its corresponding one from (2.56)–(2.65), we can obtain the Gaussian counterpart of $\mathcal{R}$, namely $\mathcal{G}$.

We next compare the obtained achievable rate region $\mathcal{G}$, with $\mathcal{G}_{\text{Tan}}$, the Gaussian counterpart of $\mathcal{R}_{\text{Tan}}$, in Fig. 2.6. It is difficult to show the comparison in a three-dimensional (3D) plot. Thus, we slice the 3D rate regions $\mathcal{G}$ and $\mathcal{G}_{\text{Tan}}$ at different values of R_0, and obtain a number of sliced views as shown in Fig. 2.6. As can be seen from Fig. 2.6, the improvement of $\mathcal{G}$ over $\mathcal{G}_{\text{Tan}}$ for $R_0 = 0.0$ is significant, which matches exactly with the result presented in Fig. 10 of [28]. It can also be observed that when R_0 is relatively high (e.g., $R_0 = 1.0$), the two regions coincide with each other. This is because most of the power of the two senders is allocated to transmit the high rate common information, while the remaining power for the private information becomes relatively small such that the improvement, primarily gained from allowing cross observation of the private information, diminishes. Note that a similar example has also been given in [58].

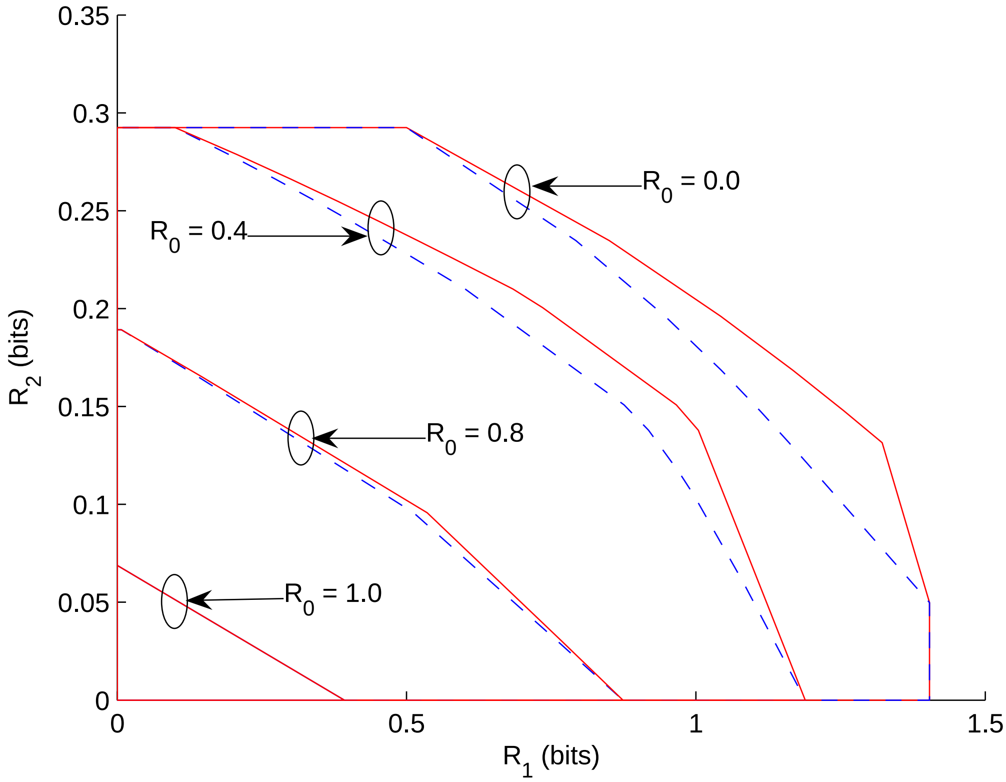

Figure 2.6: $P_1 = 6$, $P_2 = 0.5$, $c_{21} = 1$, $c_{12} = 0.25$. The dashed lines characterize the rate regions of $\mathcal{G}_{\text{Tan}}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively, and the solid lines characterize the rate regions of $\mathcal{G}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively.

2.7 Conclusions

We derived in this chapter a new achievable rate region for the two user discrete memoryless ICC. We have shown that the derived achievable rate region contains the one established in [53], and reduces to some other existing results developed for the ICC or IC. We also investigated two special cases of the ICC. For the first case in which only one sender has private information to send, we obtained an achievable rate region with a fairly simple description; while for the second case, a class of DICCs, we show that our achievable region is the capacity region. Nevertheless, in a general ICC setting, the tightness of our achievable rate region as an inner bound of the capacity region is unknown.

Chapter 3

Interference Channels With Degraded Message Sets

The IC-DMS refers to a communication model, in which two senders attempt to communicate with their respective receivers simultaneously through a common medium, and one sender has *complete* and *a priori* (non-causal) knowledge about the message being transmitted by the other. A coding scheme that collectively has advantages of cooperative coding, collaborative coding, and dirty paper coding, is developed for such a channel. By resorting to this coding scheme, achievable rate regions of the IC-DMS in both discrete memoryless and Gaussian cases are derived. The derived achievable rate regions generally include several previously known rate regions as special cases. A numerical example for the Gaussian case further demonstrates that the derived achievable rate region offers considerable improvements over these existing results in the *high-interference-gain* regime.

3.1 Introduction

The interference channel with degraded message sets (IC-DMS) refers to a communication model, in which two senders attempt to communicate with their respective receivers simultaneously through a common medium, and one sender has *complete* and *a priori* (non-causal) knowledge about the message being transmitted by the other. Such a model generically characterizes some realistic communication scenarios taking place in cognitive radio networks or in wireless sensor networks over a correlated field [43, 49, 50, 51, 52].

From an information-theoretic perspective, the IC-DMS have been investigated in [43, 51, 49, 50, 52]. Specifically, several achievable results have been obtained in [43, 51, 49, 50, 52], and the capacity regions for two special cases have been characterized in [51, 49, 50, 52]. The main achievable rate region in [43, Theorem 1] was obtained by incorporating Gel'fand-Pinsker coding [61] into the well-known coding scheme applied to the IC [46, 28]. In this coding scheme, each sender splits its message into two sub-messages, and allows its non-pairing receiver to decode one of the sub-messages. Knowing the two sub-messages and the corresponding codewords which sender 1 wishes to transmit, sender 2 applies Gel'fand-Pinsker coding to encode its own sub-messages by treating the codewords of sender 1 as known interferences. It has been also shown in [43, Corollary 2] that, an improved achievable rate region can be attained by time-sharing between the main rate region [43, Theorem 1] and a so called fully-cooperative rate point achieved by letting sender 2 use all of its power to transmit messages of sender 1. A different coding scheme was proposed in [49] and [50], in which neither of the senders splits its message into sub-messages, and receiver 2 does not decode any transmitted information from sender 1. Since sender 2 knows what sender 1 wishes to transmit, sender 2 is allowed to: 1) apply Gel'fand-Pinsker coding to encode its own message; and 2) partially cooperate with sender 1 using superposition coding. It has been proven in [49, 50] that, this is a capacity-achieving scheme for the Gaussian IC-DMS (GIC-DMS) in the *low-interference-gain* regime, in which the normalized link gain between sender 2 and receiver 1 is less than or equal to 1.

However, in practice, due to the mobility of wireless users or random geographic distributions of wireless sensors, sender 2 may be located near to receiver 1, as illustrated in Fig. 3.1. It is likely, in such a situation, that the normalized link gain between sender 2 and receiver 1 is greater than 1, which we term the

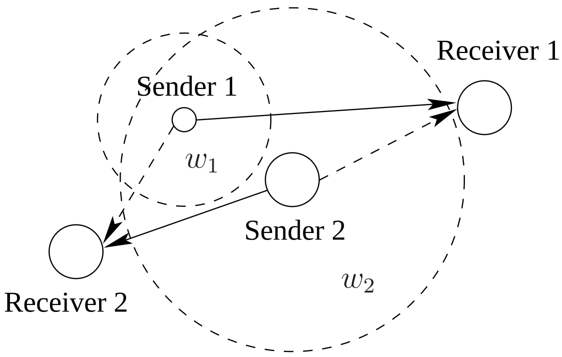

Figure 3.1: An interference channel with degraded message sets in which sender 2 is close to receiver 1.

high-interference-gain regime. In fact, the findings in this chapter reveal that the achievable rate region that has been proven to be the capacity region in the low-interference-gain regime in [49] and [50], is *strictly* non-optimal for the Gaussian IC-DMS in the high-interference-gain regime.

In this chapter, we develop a new coding scheme for the IC-DMS to improve existing achievable rate regions. Our coding scheme differs from the one proposed in [49, 50] in the way that, sender 2 splits its own message into two sub-messages, and encodes both sub-messages using Gel'fand-Pinsker coding. Moreover, receiver 1 is required to jointly decode the message from sender 1 and one sub-message from sender 2. Rate splitting is applied to enable receiver 1 to crossly observe partial information from sender 2, thus reducing the effective interference at receiver 1, whereas Gel'fand-Pinsker coding is applied to exploit the known interference at sender 2. With this coding scheme, we derive an achievable rate region for the discrete memoryless case, which is the main achievable rate result in the chapter. We further show that our region includes several existing regions as special cases. Lastly, we extend the obtained regions from the discrete memoryless case to the Gaussian case, and show by a numerical example that our achievable rate region strictly improves the existing ones [49, 50] in the *high-interference-gain* regime.

Recently, a similar coding scheme has been proposed for the IC-DMS in the independent work [62, 63, 64]. The main differences between the coding scheme [62, 63, 64] and our coding scheme can be described as follows: 1) rate splitting is employed at both senders in [62, 63, 64], whereas in our coding scheme rate splitting is only employed at sender 2; 2) Gel'fand-Pinsker coding is applied twice in a *successive* manner in [62, 63, 64], whereas Gel'fand-Pinsker coding is applied twice in a *parallel* manner in our coding scheme. However, it is not clear how the differences in coding schemes affect achievable rate results.

The rest of the chapter is organized as follows. In Section 3.2, we introduce the channel model of the IC-DMS, and the related definitions. In Section 3.3, we present the main achievable result under the discrete memoryless setting. Detailed proof is provided for this first theorem only, from which the proofs for other theorems in this chapter can be easily obtained with minor modifications. In Section 3.4, we derive two subregions of our main achievable rate region, and the second subregion is shown to be sufficient to include some existing results as special cases. Lastly, in Section 3.5, we extend our results from the discrete memoryless case to the Gaussian case, and illustrate improvements of the Gaussian rate results by numerical examples.

3.2 Channel Model

Consider the IC-DMS (also termed as the genie-aided cognitive radio channel in [43]) depicted in Fig. 3.2, where sender 1 wishes to transmit a message (or a message index), $w_1 \in \mathcal{M}_1 := \{1, ..., M_1\}$, to receiver 1, and sender 2 wishes to transmit its message, $w_2 \in \mathcal{M}_2 := \{1, ..., M_2\}$, to receiver 2. Typically, the discrete memoryless IC-DMS is described by a tuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2, p(y_1, y_2|x_1, x_2))$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are the channel input alphabets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are the channel output alphabets, and $p(y_1, y_2|x_1, x_2)$ denotes the conditional probability of $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. The channel is discrete memoryless in the sense that

$$p(y_{1,t}, y_{2,t}|x_{1,t}, x_{2,t}, x_{1,t-1}, x_{2,t-1}, ...) = p(y_{1,t}, y_{2,t}|x_{1,t}, x_{2,t}),$$

for every discrete time instant t in a synchronous transmission.

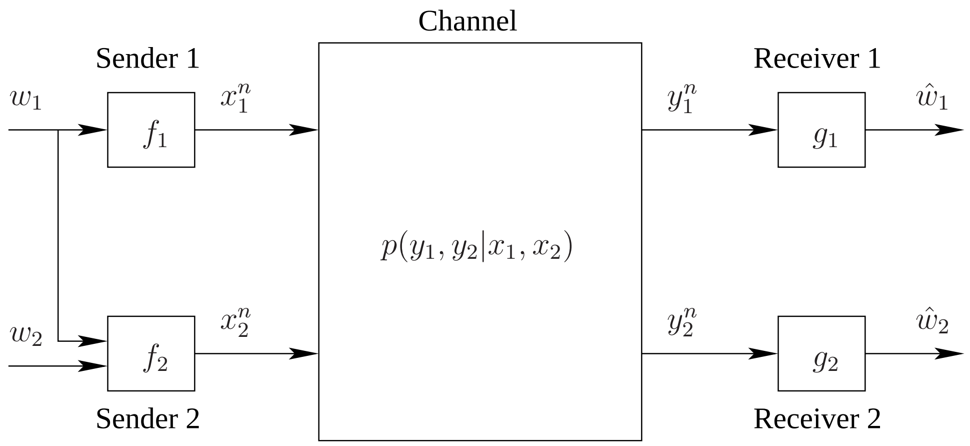

Figure 3.2: Interference channel with degraded message sets.

In view of the channel input-output relationship, the IC-DMS is the same as the IC. However, in the IC-DMS, sender 2 is able to non-causally obtain the knowledge of the message w_1, which will be transmitted from sender 1. This is the key difference between the IC-DMS and IC in terms of the information flow. We next present the following standard definitions with regard to the existence of codes and achievable rates for the discrete memoryless IC-DMS channel.

Definition 3.1 An $(M_1, M_2, n, P_e^{(n)})$ *code for the discrete memoryless IC-DMS consists of*

i) two encoding functions

$$f_1 : \mathcal{M}_1 \to \mathcal{X}_1^n, \text{ and } f_2 : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}_2^n;$$

ii) two decoding functions

$$g_1 : \mathcal{Y}_1^n \to \mathcal{M}_1, \text{ and } g_2 : \mathcal{Y}_2^n \to \mathcal{M}_2;$$

iii) the average probability of error

$$P_e^{(n)} := \max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\},$$

where $P_{e,i}^{(n)}$ *denotes the average probability of error at decoder* i*, and is com-*

puted as

$$P_{e,i}^{(n)} = \frac{1}{M_1 M_2} \sum_{w_1, w_2} \Pr(\hat{w}_i \neq w_i | (w_1, w_2) \ sent), \quad i = 1, 2.$$

Definition 3.2 A non-negative rate pair (R_1, R_2) *is said to be achievable for the IC-DMS, if there exists a sequence of* $(M_1, M_2, n, P_e^{(n)})$ *codes with*

$$R_1 \leq \frac{\log M_1}{n}, \text{ and } R_2 \leq \frac{\log M_2}{n},$$

such that $P_e^{(n)}$ *approaches zero as* $n \to \infty$. *The capacity region of the IC-DMS is the set of all the achievable rate pairs, and an achievable rate region is a subset of the capacity region.*

3.3 An Achievable Rate Region for the Discrete Memoryless IC-DMS

In this section, we present a new achievable rate region for the discrete memoryless IC-DMS, which is the primary result in this chapter.

Consider auxiliary random variables W, U, V, and a time-sharing random variable Q defined on arbitrary finite sets $\mathcal{W}$, $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{Q}$, respectively. Let $\mathcal{P}$ denote the set of all joint probability distributions $p(\cdot)$ that factor in the form of

$$\begin{aligned} p(q, w, x_1, u, v, x_2, y_1, y_2) = p(q)p(w, x_1|q)p(u|w, q)p(v|w, q) \\ \cdot p(x_2|u, v, w, q)p(y_1, y_2|x_1, x_2), \end{aligned} \tag{3.1}$$

where w, u, v, and q are realizations of random variables W, U, V, and Q.

Let $\mathcal{R}(p)$ denote the set of all non-negative rate pairs (R_1, R_2) such that the following inequalities hold simultaneously

$$R_1 \leq \min\{I(W;U,Y_1|Q), I(W,U;Y_1|Q)\}, \tag{3.2}$$

$$R_2 \leq I(U,V;Y_2|Q) + I(U;V|Q) - I(U;W|Q) - I(V;W|Q), \tag{3.3}$$

$$R_1 + R_2 \leq I(W,U;Y_1|Q) + I(V;U,Y_2|Q) - I(V;W|Q); \tag{3.4}$$

$$0 \leq I(U;Y_2,V|Q) - I(U;W|Q), \tag{3.5}$$

$$0 \leq I(V;Y_2,U|Q) - I(V;W|Q), \tag{3.6}$$

for a given joint distribution $p(\cdot) \in \mathcal{P}$.

Let $\mathcal{C}$ denote the capacity region of the discrete memoryless IC-DMS, and let

$$\mathcal{R} := \bigcup_{p(\cdot)\in\mathcal{P}} \mathcal{R}(p).$$

Theorem 3.1 The region $\mathcal{R}$ *is achievable for the discrete memoryless IC-DMS, i.e.,*

$$\mathcal{R} \subseteq \mathcal{C}.$$

Coding Scheme Outline: our coding scheme is mainly based on the ideas of superposition coding [3] and Gel'fand-Pinsker coding [61]. Specifically, sender 1 independently encodes its message w_1 as a whole; while sender 2 needs split its message into two parts, i.e., $w_2 = (w_{21}, w_{22})$, and encode them separately. Both w_{21} and w_{22} are encoded using the Gel'fand-Pinsker coding scheme, but they are processed differently at the receivers. The message w_{22} will be decoded by receiver 2 only, while w_{21} will be decoded by both receivers. Moreover, knowing the message and the corresponding codeword which sender 1 is going to transmit, sender 2 not only can apply Gel'fand-Pinsker coding to deal with the known interference, but also can cooperate with sender 1 to transmit w_1 using superposition coding. Let R_{21} and R_{22} denote the rates of w_{21} and w_{22} respectively, i.e., $w_{21} \in \{1, \ldots, 2^{nR_{21}}\}$

and $w_{22} \in \{1, \ldots, 2^{nR_{22}}\}$. If receiver 1 can decode w_1 and receiver 2 can decode both w_{21} and w_{22} with vanishing probabilities of error, then $(R_1, R_{21} + R_{22})$ is an achievable rate pair for the IC-DMS.

In the following proof, we will frequently use the notion of jointly typical sequences and joint asymptotic equipartition property [40, Section 14.2].

Proof: To prove that the entire region $\mathcal{R}$ is achievable for the channel, it is sufficient to prove that $\mathcal{R}(p)$ is achievable for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}$.

Random Codebook Generation

Consider a fixed joint distribution $p(\cdot) \in \mathcal{P}$, and a random time-sharing codeword $\mathbf{q}$ of length n. The codeword $\mathbf{q}$ that is revealed to both the senders and receivers, is assumed to be generated according to $\prod_{t=1}^{n} p(q_t)$.

Generate 2^{nR_1} independent codewords $\mathbf{W}(j)$, $j \in \{1, \ldots, 2^{nR_1}\}$, according to $\prod_{t=1}^{n} p(w_t|q_t)$, and for each $\mathbf{w}(j)$ generate one codeword $\mathbf{X}_1(j)$, according to $\prod_{t=1}^{n} p(x_{1,t}|w_i(j), q_t)$. Generate $2^{n(R_{21}+I(W;U|Q)+4\epsilon)}$ independent codewords $\mathbf{U}(l_1)$, $l_1 \in \{1, \ldots, 2^{n(R_{21}+I(W;U|Q)+4\epsilon)}\}$, according to $\prod_{t=1}^{n} p(u_t|q_t)$, and generate $2^{n(R_{22}+I(W;V|Q)+4\epsilon)}$ independent codewords $\mathbf{V}(l_2)$, $l_2 \in \{1, \ldots, 2^{n(R_{22}+I(W;V|Q)+4\epsilon)}\}$, according to $\prod_{t=1}^{n} p(v_t|q_t)$, where ϵ denotes an arbitrarily small positive number. Lastly, for each codeword triple $(\mathbf{u}(l_1), \mathbf{v}(l_2), \mathbf{w}(j))$, generate one codeword $\mathbf{X}_2(l_1, l_2, j)$ according to $\prod_{t=1}^{n} p(x_{2,t}|u_t(l_1), v_t(l_2), w_t(j), q_t)$. Now uniformly distribute the $2^{n(R_{21}+I(W;U|Q)+4\epsilon)}$ codewords $\mathbf{u}(l_1)$ into $2^{nR_{21}}$ bins indexed by $k_1 \in \{1, \ldots, 2^{nR_{21}}\}$, such that each bin contains $2^{n(I(W;U|Q)+4\epsilon)}$ codewords, and uniformly distribute the $2^{n(R_{22}+I(W;V|Q)+4\epsilon)}$ codewords $\mathbf{v}(l_2)$ into $2^{nR_{22}}$ bins indexed by $k_2 \in \{1, \ldots, 2^{nR_{22}}\}$ such that each bin contains $2^{n(I(W;V|Q)+4\epsilon)}$ codewords. The entire codebook is revealed to both the senders and receivers.

Encoding and Transmission

We assume that two senders wish to transmit a message vector $(w_1, w_{21}, w_{22}) = (j, k_1, k_2)$. Sender 1 simply encodes the message as a codeword $\mathbf{x}_1(j)$ and sends the codeword in n channel uses. Let $A_\epsilon^{(n)}$ denote a jointly typical set. Sender 2 first needs to look for a codeword $\mathbf{u}(\hat{l}_1)$ in bin k_1 such that $(\mathbf{u}(\hat{l}_1), \mathbf{w}(j), \mathbf{q}) \in A_\epsilon^{(n)}$, and a codeword $\mathbf{v}(\hat{l}_2)$ in bin k_2 such that $(\mathbf{v}(\hat{l}_2), \mathbf{w}(j), \mathbf{q}) \in A_\epsilon^{(n)}$. If sender 2 finds such $\mathbf{u}(\hat{l}_1)$ and $\mathbf{v}(\hat{l}_2)$ successfully, the codeword $\mathbf{x}_2(\hat{l}_1, \hat{l}_2, j)$ is sent through n channel uses. Otherwise, sender 2 declares an encoding error.

Decoding

Receiver 1 first looks for all the index pairs $(\hat{j}, \hat{\hat{l}}_1)$ such that $(\mathbf{w}(\hat{j}), \mathbf{u}(\hat{\hat{l}}_1), \mathbf{y}_1, \mathbf{q}) \in A_\epsilon^{(n)}$. If $\hat{j}$ in all the index pairs found are the same, receiver 1 declares $w_1 = \hat{j}$. Otherwise, receiver 1 declares a decoding error. Receiver 2 will first look for all index pairs $(\bar{\hat{l}}_1, \hat{\hat{l}}_2)$ such that $(\mathbf{u}(\bar{\hat{l}}_1), \mathbf{v}(\hat{\hat{l}}_2), \mathbf{y}_2, \mathbf{q}) \in A_\epsilon^{(n)}$. If $\bar{\hat{l}}_1$ in all the index pairs found are indices of codewords $\mathbf{u}(\bar{\hat{l}}_1)$ from the same bin with index $\hat{k}_1$, and $\hat{\hat{l}}_2$ in all the index pairs found are indices of codewords $\mathbf{v}(\hat{\hat{l}}_2)$ from the same bin with index $\hat{k}_2$, then receiver 2 declares that the messages $(w_{21}, w_{22}) = (\hat{k}_1, \hat{k}_2)$ were transmitted. Otherwise, a decoding error is declared.

Evaluation of Probabilities of Error

We now derive upper bounds for the probabilities of the respective error events which may happen during the encoding and decoding processes. Due to the symmetry of the codebook generation and encoding processing, the probability of error is not codeword dependent. Without loss of generality, we assume that the messages $(w_1, w_{21}, w_{22}) = (1, 1, 1)$ are encoded and sent. We further assume that the codewords $\mathbf{u}(\hat{l}_1)$ and $\mathbf{v}(\hat{l}_2)$ found in the respective bin 1 during the encoding process are $\mathbf{u}(1)$ and $\mathbf{v}(1)$, respectively. Hence, $\mathbf{x}_1(1)$ and $\mathbf{x}_2(1,1,1)$ are transmitted. We next define the following four types of events:

$$E^u_{a,b} := (\mathbf{U}(a), \mathbf{W}(b), \mathbf{q}) \in A^{(n)}_\epsilon,$$
$$E^v_{a,b} := (\mathbf{V}(a), \mathbf{W}(b), \mathbf{q}) \in A^{(n)}_\epsilon,$$
$$\dot{E}_{a,b} := (\mathbf{W}(a), \mathbf{U}(b), \mathbf{Y}_1, \mathbf{q}) \in A^{(n)}_\epsilon,$$
$$\ddot{E}_{a,b} := (\mathbf{U}(a), \mathbf{V}(b), \mathbf{Y}_2, \mathbf{q}) \in A^{(n)}_\epsilon.$$

Let $P_e(\text{enc2})$, $P_e(\text{dec1})$, and $P_e(\text{dec2})$ denote the probabilities of error at the encoder of sender 2, the decoder of receiver 1, and the decoder of receiver 2, respectively.

[**Evaluation of** $P_e(\text{enc2})$.] An error is made if 1) the encoder at sender 2 cannot find $\mathbf{u}(\hat{l}_1)$ in bin 1 such that $(\mathbf{u}(\hat{l}_1), \mathbf{w}(1), \mathbf{q}) \in A^{(n)}_\epsilon$, and/or 2) it cannot find $\mathbf{v}(\hat{l}_2)$ in bin 1 such that $(\mathbf{v}(\hat{l}_2), \mathbf{w}(1), \mathbf{q}) \in A^{(n)}_\epsilon$. The probability of error at the encoder of sender 2 is bounded as

$$\begin{aligned} P_e(\text{enc2}) &\leq \Pr\Big(\bigcap_{\mathbf{U}(\hat{l}_1) \in \text{bin } 1} (\mathbf{U}(\hat{l}_1), \mathbf{W}(1), \mathbf{q}) \notin A^{(n)}_\epsilon \Big) \\ &\quad + \Pr\Big(\bigcap_{\mathbf{V}(\hat{l}_2) \in \text{bin } 1} (\mathbf{V}(\hat{l}_2), \mathbf{W}(1), \mathbf{q}) \notin A^{(n)}_\epsilon \Big) \\ &= \prod_{\mathbf{u}(\hat{l}_1) \in \text{bin } 1} (1 - \Pr(E^u_{\hat{l}_1,1})) + \prod_{\mathbf{v}(\hat{l}_2) \in \text{bin } 1} (1 - \Pr(E^v_{\hat{l}_2,1})) \\ &\leq (1 - \Pr(E^u_{1,1}))^{2^{n(I(U;W|Q)+4\epsilon)}} + (1 - \Pr(E^v_{1,1}))^{2^{n(I(V;W|Q)+4\epsilon)}}. \end{aligned} \tag{3.7}$$

As the time-sharing sequence $\mathbf{q}$ is predetermined, we have

$$\begin{aligned} \Pr(E^u_{1,1}) &= \sum_{(\mathbf{u},\mathbf{w},\mathbf{q}) \in A^{(n)}_\epsilon} \Pr(\mathbf{U}(1) = \mathbf{u}|\mathbf{q}) \Pr(\mathbf{W}(1) = \mathbf{w}|\mathbf{q}) \\ &\geq 2^{n(H(U,W|Q-\epsilon)} 2^{-n(H(U|Q)+\epsilon)} 2^{-n(H(W|Q)+\epsilon)} \\ &= 2^{-n(I(U;W|Q)+3\epsilon)}. \end{aligned}$$

Similarly, we can obtain $\Pr(E^v_{1,1}) \geq 2^{-n(I(V;W|Q)+3\epsilon)}$. From (3.7), we have

$$P_e(\text{enc2}) \leq (1-2^{-n(I(U;W|Q)+3\epsilon)})^{2^{n(I(U;W|Q)+4\epsilon)}} + (1-2^{-n(I(V;W|Q)+3\epsilon)})^{2^{n(I(V;W|Q)+4\epsilon)}},$$

where the first term can be upper bounded as

$$\begin{aligned}
(1-2^{-n(I(U;W|Q)+3\epsilon)})^{2^{n(I(U;W|Q)+4\epsilon)}} &= e^{2^{n(I(U;W|Q)+4\epsilon)}\ln(1-2^{-n(I(U;W|Q)+3\epsilon)})} \\
&\overset{(a)}{\leq} e^{2^{n(I(U;W|Q)+4\epsilon)}(-2^{-n(I(U;W|Q)+3\epsilon)})} \\
&= e^{-2^{n\epsilon}}.
\end{aligned}$$

Note that (a) follows from the Mercator series of $\ln(1+x)$ with $x = -2^{-n(I(U;W|Q)+3\epsilon)}$ being a negative real number that approaches 0. The same argument was used in the proof of [65, Lemma 2.1.3]. Hence, we can readily conclude that $P_e(\text{enc2}) \to 0$ as $n \to \infty$.

[**Evaluation of** $P_e(\text{dec1})$.] An error is made if 1) $\dot{E}^c_{1,1}$ happens, and/or 2) there exists some $\hat{j} \neq 1$ such that $\dot{E}_{\hat{j},\hat{l}_1}$ happens. Note that the error events $\dot{E}_{1,\hat{l}_1}$ with $\hat{l}_1 \neq 1$ are not considered as error events, and are excluded from the computation of the probability of error, as it is unnecessary for receiver 1 to correctly decode l_1. The probability of error at the decoder of receiver 1 can be upper bounded as

$$\begin{aligned}
P_e(\text{dec1}) &\leq \Pr(\dot{E}^c_{1,1} \bigcup \cup_{\hat{j}\neq 1} \dot{E}_{\hat{j},\hat{l}_1}) \\
&\leq \Pr(\dot{E}^c_{1,1}) + \sum_{\hat{j}\neq 1} \Pr(\dot{E}_{\hat{j},\hat{l}_1}) \\
&= \Pr(\dot{E}^c_{1,1}) + \sum_{\hat{j}\neq 1} \Pr(\dot{E}_{\hat{j},1}) + \sum_{\hat{j}\neq 1, \hat{l}_1\neq 1} \Pr(\dot{E}_{\hat{j},\hat{l}_1}) \\
&\leq \Pr(\dot{E}^c_{1,1}) + 2^{nR_1}\Pr(\dot{E}_{2,1}) + 2^{n(R_1+R_{21}+I(U;W|Q)+4\epsilon)}\Pr(\dot{E}_{2,2}). \qquad (3.8)
\end{aligned}$$

The term $\Pr(\dot{E}_{2,1})$ can first be upper bounded as

$$\begin{aligned}\Pr(\dot{E}_{2,1}) &= \sum_{(\mathbf{w},\mathbf{u},\mathbf{y}_1,\mathbf{q})\in A_\epsilon^{(n)}} \Pr(\mathbf{W}(2)=\mathbf{w}|\mathbf{q})\Pr(\mathbf{U}(1)=\mathbf{u},\mathbf{Y}_1=\mathbf{y}_1|\mathbf{q}) \\ &\le 2^{n(H(W,U,Y_1|Q)+\epsilon)}2^{-n(H(W|Q)-\epsilon)}2^{-n(H(U,Y_1|Q)-\epsilon)} \\ &= 2^{-n(I(W;U,Y_1|Q)-3\epsilon)}. \end{aligned} \tag{3.9}$$

Similarly, we obtain

$$\begin{aligned}\Pr(\dot{E}_{2,2}) &= \sum_{(\mathbf{w},\mathbf{u},\mathbf{y}_1,\mathbf{q})\in A_\epsilon^{(n)}} \Pr(\mathbf{W}(2)=\mathbf{w}|\mathbf{q})\Pr(\mathbf{U}(2)=\mathbf{u}|\mathbf{q})\Pr(\mathbf{Y}_1=\mathbf{y}_1|\mathbf{q}) \\ &\le 2^{n(H(W,U,Y_1|Q)+\epsilon)}2^{-n(H(W|Q)-\epsilon)}2^{-n(H(U|Q)-\epsilon)}2^{-n(H(Y_1|Q)-\epsilon)} \\ &= 2^{-n(I(W,U;Y_1|Q)+I(W;U|Q)-4\epsilon)}. \end{aligned} \tag{3.10}$$

Substituting (3.9) and (3.10) into (3.8), we obtain

$$P_e(\text{dec1}) \le \epsilon + 2^{-n(I(W;U,Y_1|Q)-R_1-3\epsilon)} + 2^{-n(I(W,U;Y_1|Q)-R_1-R_{21}-5\epsilon)}.$$

Since $\epsilon > 0$ can be arbitrarily small, $P_e(\text{dec1})$ tends to zero as $n \to \infty$ if

$$R_1 \le I(W;U,Y_1|Q), \tag{3.11}$$

$$R_1 + R_{21} \le I(W,U;Y_1|Q), \tag{3.12}$$

are satisfied.

[**Evaluation of** $P_e(\text{dec2})$.] An error is made if 1) $\ddot{E}^c_{1,1}$ happens, and/or 2) there exists some $(\bar{\hat{l}}_1, \hat{\hat{l}}_2)$ in which either $\bar{\hat{l}}_1$ or $\hat{\hat{l}}_2$ is not an index of any codeword from the respective bin 1 such that $\ddot{E}_{\bar{\hat{l}}_1,\hat{\hat{l}}_2}$ happens. The probability of the second case is upper bounded by the probability of the event, $\ddot{E}_{\bar{\hat{l}}_1,\hat{\hat{l}}_2}$, for some $(\bar{\hat{l}}_1, \hat{\hat{l}}_2) \neq (1,1)$.

Thus, the probability of error at the decoder of receiver 2 is bounded as

$$\begin{aligned}
P_e(\text{dec2}) &\leq \Pr(\ddot{E}^c_{1,1} \bigcup \cup_{(\bar{\hat{l}}_1,\hat{\hat{l}}_2)\neq(1,1)} \ddot{E}_{\bar{\hat{l}}_1,\hat{\hat{l}}_2}) \\
&\leq \Pr(\ddot{E}^c_{1,1}) + \sum_{(\bar{\hat{l}}_1,\hat{\hat{l}}_2)\neq(1,1)} P(\ddot{E}_{\bar{\hat{l}}_1,\hat{\hat{l}}_2}) \\
&= \Pr(\ddot{E}^c_{1,1}) + \sum_{\bar{\hat{l}}_1\neq 1} \Pr(\ddot{E}_{\bar{\hat{l}}_1,1}) + \sum_{\hat{\hat{l}}_2\neq 1} \Pr(\ddot{E}_{1,\hat{\hat{l}}_2}) + \sum_{(\bar{\hat{l}}_1\neq 1,\hat{\hat{l}}_2\neq 1)} \Pr(\ddot{E}_{\bar{\hat{l}}_1,\hat{\hat{l}}_2}) \\
&\leq \Pr(\ddot{E}^c_{1,1}) + 2^{n(R_{21}+I(W;U|Q)+\epsilon)}\Pr(\ddot{E}_{2,1}) + 2^{n(R_{22}+I(W;V|Q)+\epsilon)}\Pr(\ddot{E}_{1,2}) \\
&\quad + 2^{n(R_{21}+R_{22}+I(W;U|Q)+I(W;V|Q)+2\epsilon)}\Pr(\ddot{E}_{2,2}). \qquad (3.13)
\end{aligned}$$

Following the same way as we derived upper-bounds of $\Pr(\dot{E}_{2,1})$ and $\Pr(\dot{E}_{2,2})$ in (3.9) and (3.10), we upper bound $\Pr(\ddot{E}_{2,1})$, $\Pr(\ddot{E}_{1,2})$, and $\Pr(\ddot{E}_{2,2})$ as

$$\Pr(\ddot{E}_{2,1}) \leq 2^{-n(I(U;V,Y_2|Q)-3\epsilon)}, \qquad (3.14)$$

$$\Pr(\ddot{E}_{1,2}) \leq 2^{-n(I(V;U,Y_2|Q)-3\epsilon)}, \qquad (3.15)$$

$$\Pr(\ddot{E}_{2,2}) \leq 2^{-n(I(U,V;Y_2|Q)+I(U;V|Q)-4\epsilon)}. \qquad (3.16)$$

Substituting (3.14)–(3.16) into (3.13), we conclude that $P_e(\text{dec2}) \to 0$ as $n \to \infty$ if

$$R_{21} \leq I(U;V,Y_2|Q) - I(W;U|Q), \qquad (3.17)$$

$$R_{22} \leq I(V;U,Y_2|Q) - I(W;V|Q), \qquad (3.18)$$

$$R_{21} + R_{22} \leq I(U,V;Y_2|Q) + I(U;V|Q) - I(W;U|Q) - I(W;V|Q), \qquad (3.19)$$

are satisfied.

If (3.11)–(3.12) and (3.17)–(3.19) are satisfied, the average probabilities of error at both decoders diminish as $n \to \infty$. We hence conclude that a $(2^{nR_1}, 2^{n(R_{21}+R_{22})}, n, P_e^{(n)})$ code with $P_e^{(n)} \to 0$ exists for the channel. Furthermore, we obtain (3.2)–(3.6) by applying Fourier-Motzkin elimination [57, 59, 66] on (3.11)–(3.12), (3.17)–(3.19), $R_{21} \geq 0$ and $R_{22} \geq 0$. Therefore, the rate region $\mathcal{R}(p)$ is achievable for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}$, and Theorem 3.1 follows.

■

Remark 3.1 The proposed coding scheme exploits three coding methods to achieve any rate pair in the rate region, $\mathcal{R}$. The first method is cooperation that is realized by applying the superposition relationship between $\mathbf{w}$ *and* $\mathbf{x}_2$ *via* $p(x_2|u, v, w, q)$*. The second one is collaboration, by which we mean that sender 2 separates its own message into two parts, i.e.,* $w_2 = (w_{21}, w_{22})$*, and encodes* w_{21} *at a possibly low rate such that receiver 1 can decode it. By doing so, the effective interference caused by the signals carrying the information from sender 2 may be reduced. The third one is Gel'fand-Pinsker coding [61], which we apply to encode both messages,* w_{21} *and* w_{22}*, from sender 2 by treating the codeword* $\mathbf{w}$ *as known interference. This perhaps allows receiver 2 to be able to decode the messages from sender 2 at the same rate as if the interference caused by sender 1 was not present.*

3.4 Relating With Some Existing Rate Regions

In this section, we will show that Theorem 3.1 includes the achievable rate regions presented in [49, 50]. To demonstrate it, we first compromise the advantages of the coding scheme developed in Section 3.3 to obtain the two subregions of $\mathcal{R}$.

3.4.1 A Subregion of $\mathcal{R}$

Let $\mathcal{P}^*$ denote the set of all joint probability distributions $p(\cdot)$ that factor in the form of

$$\begin{aligned} p(q, w, x_1, u, v, x_2, y_1, y_2) = {} & p(q)p(x_1, w|q)p(u|q)p(v|w, q) \\ & \cdot p(x_2|u, w, q)p(y_1, y_2|x_1, x_2). \end{aligned} \tag{3.20}$$

Note that the joint distribution (3.20) differs from (3.1) in the way that U is now independent of any other auxiliary random variables conditioned on Q.

Let $\mathcal{R}_{\text{sim}}(p)$ denote the set of all non-negative rate pairs (R_1, R_2) such that

$$R_1 \leq I(W; Y_1|U, Q), \tag{3.21}$$

$$R_2 \leq I(U, V; Y_2|Q) - I(V; W|Q), \tag{3.22}$$

$$R_1 + R_2 \leq I(W, U; Y_1|Q) + I(V; Y_2|U, Q) - I(V; W|Q); \tag{3.23}$$

$$0 \leq I(V; Y_1|U, Q) - I(V; W|Q), \tag{3.24}$$

for a joint probability distribution $p(\cdot) \in \mathcal{P}^*$. Furthermore, let

$$\mathcal{R}_{\text{sim}} := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_{\text{sim}}(p).$$

Theorem 3.2 The rate region $\mathcal{R}_{sim}$ is achievable for the discrete memoryless IC-DMS, i.e., $\mathcal{R}_{sim} \subseteq \mathcal{R} \subseteq \mathcal{C}$.

Proof: The proof can be devised from the proof of Theorem 3.1 by customizing the original coding scheme for the new joint distribution (3.20). We change the encoding and decoding method for the message w_{21} (corresponding to U), i.e., the Gel'fand-Pinsker coding used in the proof of Theorem 3.1 is replaced by conventional random coding. Specifically, we generate $2^{nR_{21}}$ independent codewords $\mathbf{U}(k_1)$, $k_1 \in \{1, \ldots, 2^{nR_{21}}\}$, according to $\prod_{i=1}^{n} p(u_i|q_i)$. The encoding and decoding are then adapted to the new codebook accordingly. Evaluating the probability of error in the same way as was done in the proof of Theorem 3.1, we can obtain

$$R_1 \leq I(W; Y_1|U, Q), \tag{3.25}$$

$$R_1 + R_{21} \leq I(W, U; Y_1|Q); \tag{3.26}$$

$$R_{21} \leq I(U; Y_2|V, Q), \tag{3.27}$$

$$R_{22} \leq I(V; Y_2|U, Q) - I(W; V|Q), \tag{3.28}$$

$$R_{21} + R_{22} \leq I(U, V; Y_2|Q) - I(W; V|Q). \tag{3.29}$$

Again, for the purpose of simplification, substitute R_{21} with $R_2 - R_{22}$ in the group of (3.25)–(3.29). By applying Fourier-Motzkin elimination on the resulting inequalities to remove R_{22}, and adding the constraints that ensure the respective rates R_1, R_{21} and R_{22} are non-negative, we can obtain (3.21)–(3.24). Therefore, the region $R_{\text{sim}}(p)$ is achievable for a given $p(\cdot) \in \mathcal{P}^*$, and the theorem follows. ■

Remark 3.2 Note that simultaneous decoding (simultaneous joint typicality) is applied at both decoders. The advantage of simultaneous decoding over successive decoding is well demonstrated on the interference channel by Han and Kobayashi in [28]. We next modify the coding scheme by applying successive decoding instead of simultaneous decoding at both decoders to derive a subregion of $\mathcal{R}_{sim}$, which, of course, is also a subregion of the achievable rate region $\mathcal{R}$.

3.4.2 A Subregion of $\mathcal{R}_{\text{sim}}$

Let $\mathcal{R}_{\text{suc}}(p)$ denote the set of all achievable rate pairs (R_1, R_2) such that

$$R_1 \leq I(W; Y_1|U, Q), \tag{3.30}$$

$$R_2 \leq \min\{I(U; Y_1|Q), I(U; Y_2|Q)\} + I(V; Y_2|U, Q) - I(V; W|Q), \tag{3.31}$$

$$0 \leq I(V; Y_1|U, Q) - I(V; W|Q), \tag{3.32}$$

for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$. Define

$$\mathcal{R}_{\text{suc}} := \bigcup_{p(\cdot) \in \mathcal{P}^*} \mathcal{R}_{\text{suc}}(p).$$

Theorem 3.3 The rate region $\mathcal{R}_{suc}$ is achievable for the discrete memoryless IC-DMS, i.e., $\mathcal{R}_{suc} \subseteq \mathcal{R}_{sim} \subseteq \mathcal{R} \subseteq \mathcal{C}$.

Proof: The codebook generation, encoding and transmission remain the same as those used to prove Theorem 3.2. The changes are made to the decoding at both decoders. Both decoders try to decode w_{21} first, and then try to decode w_1 and

w_{22} respectively. We can easily arrive

$$R_{21} \leq I(U; Y_1|Q), \tag{3.33}$$

$$R_1 \leq I(W; Y_1|U, Q), \tag{3.34}$$

$$R_{21} \leq I(U; Y_2|Q), \tag{3.35}$$

$$R_{22} \leq I(V; Y_2|U, Q) - I(W; V|Q). \tag{3.36}$$

From (3.33)–(3.36), it is straightforward to obtain (3.30)–(3.32). Therefore, the region $\mathcal{R}_{\text{suc}}(p)$ is achievable, and the theorem follows immediately. ■

Remark 3.3 Note that (3.33) *is only necessary when successive decoding is applied. This is because every decoding step in a successive decoding scheme is expected to have a vanishing probability of error.*

In what follows, we further specialize the subregion $\mathcal{R}_{\text{suc}}$ to obtain two more achievable rate regions $\mathcal{R}_{\text{sp1}}$ and $\mathcal{R}_{\text{sp2}}$.

Let $\mathcal{P}_1^*$ denote the set of all joint probability density distributions $p(\cdot)$ that factor in the form of

$$p(q, w, x_1, v, x_2, y_1, y_2) = p(q)p(x_1, w|q)p(v|w, q)p(x_2|w, q)p(y_1, y_2|x_1, x_2). \tag{3.37}$$

Let $\mathcal{R}_{\text{sp1}}(p)$ denote the set of all non-negative rate pairs (R_1, R_2) such that

$$R_1 \leq I(W; Y_1|Q), \tag{3.38}$$

$$R_2 \leq I(V; Y_2|Q) - I(V; W|Q), \tag{3.39}$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}_1^*$. Define

$$\mathcal{R}_{\text{sp1}} := \bigcup_{p(\cdot) \in \mathcal{P}_1^*} \mathcal{R}_{\text{sp1}}(p).$$

Corollary 3.1 The region $\mathcal{R}_{sp1}$ *is an achievable rate region for the discrete memoryless IC-DMS, i.e.,* $\mathcal{R}_{sp1} \subseteq \mathcal{R}_{suc} \subseteq \mathcal{R}_{sim} \subseteq \mathcal{R} \subseteq \mathcal{C}$.

Proof: Fixing the auxiliary random variable U as a constant, we reduce (3.30) and (3.31) to (3.38) and (3.39), and the corollary follows immediately. ■

Remark 3.4 The achievable rate region $\mathcal{R}_{sp1}$ *is identical to the region* $\mathcal{R}_{in}$ *reported in [50, Theorem 3.1], which is the discrete memoryless counterpart of the region given in [49, Theorem 4.1] and [50, Theorem 3.5]. It is shown in both [49] and [50] that* $\mathcal{R}_{sp1}$ *is the capacity region for the IC-DMS in the low-interference-gain regime.*

Let $\mathcal{P}_2^*$ denote the set of all joint probability distributions $p(\cdot)$ that factor in the form of

$$p(q, w, x_1, u, x_2, y_1, y_2) = p(q)p(x_1, w|q)p(u|q)p(x_2|u, w, q)p(y_1, y_2|x_1, x_2). \quad (3.40)$$

Let $\mathcal{R}_{\mathrm{sp2}}(p)$ denote the set of all non-negative rate pairs (R_1, R_2) such that

$$R_1 \leq I(W; Y_1|U, Q), \quad (3.41)$$

$$R_2 \leq \min\{I(U; Y_1|Q), I(U; Y_2|Q)\}, \quad (3.42)$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}_2^*$. Define

$$\mathcal{R}_{\mathrm{sp2}} := \bigcup_{p(\cdot)\in\mathcal{P}_2^*} \mathcal{R}_{\mathrm{sp2}}(p).$$

Corollary 3.2 The region $\mathcal{R}_{sp2}$ *is an achievable rate region for the discrete memoryless IC-DMS, i.e.,* $\mathcal{R}_{sp2} \subseteq \mathcal{R}_{suc} \subseteq \mathcal{R}_{sim} \subseteq \mathcal{R} \subseteq \mathcal{C}$.

Proof: The proof can be devised from the proof of Theorem 3.3 easily by fixing V as a constant. ■

Remark 3.5 The Gaussian counterpart of $\mathcal{R}_{sp2}$ includes the set of achievable rate pairs in [49, Lemma 4.2] as a subset.

3.5 Gaussian IC-DMS

In the preceding sections, we obtain achievable rate regions for the discrete memoryless IC-DMS. We now extend these results to obtain achievable rate regions for the *Gaussian* IC-DMS (GIC-DMS).

3.5.1 Channel Model of the GIC-DMS

With no loss of the information-theoretic optimality, any GIC-DMS can be converted to the GIC-DMS in standard form through invertible transformations [49, 59, 46]. Hence, we only need to consider the GIC-DMS in standard form described in the following

$$Y_1 = X_1 + \sqrt{c_{21}} X_2 + Z_1,$$
$$Y_2 = X_2 + \sqrt{c_{12}} X_1 + Z_2,$$

where Z_i, $i = 1, 2$, is the additive white Gaussian noise with zero mean and unit variance, and $\sqrt{c_{21}}$ and $\sqrt{c_{12}}$ are the *normalized* link gains in the GIC-DMS depicted in Fig. 3.3.

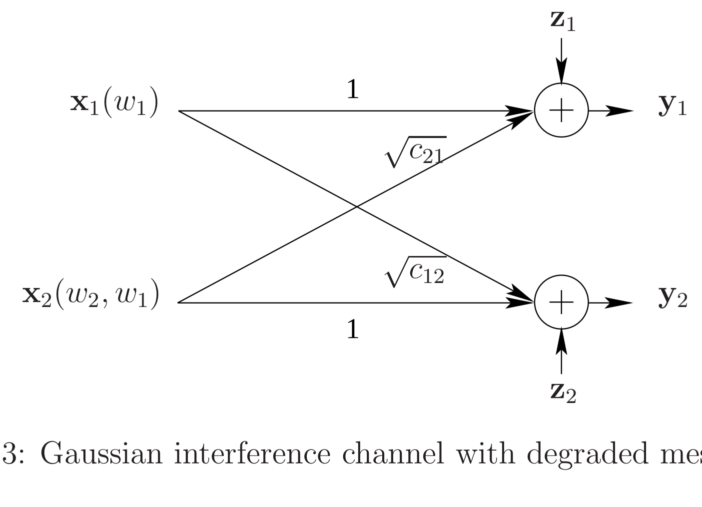

Figure 3.3: Gaussian interference channel with degraded message sets.

The transmitted codeword $\mathbf{x}_i := (x_{i1}, \ldots, x_{in})$, $i = 1, 2$, is subject to an average power constraint given by $n^{-1} \sum_{t=1}^{n} \|x_{it}\|^2 \leq P_i$, $i = 1, 2$. Furthermore, we restrict our attention to the Gaussian codewords X_i^n, $i = 1, 2$, for the convenience of evaluation and comparison with existing results.

3.5.2 Achievable Rate Regions for the GIC-DMS

3.5.2.1 Gaussian Extension of $\mathcal{R}$

Generally speaking, the achievable regions in Theorem 3.1, Corollary 3.1, and Corollary 3.2 can be extended to the discrete time Gaussian memoryless case by quantizing the channel inputs and outputs [67, Chapter 7]. In particular, the Gaussian extension of the rate region $\mathcal{R}_{\mathrm{sp1}}$ in Corollary 3.1 has been given in [49, Theorem 4.1] and [50, Theorem 3.5]. We next outline how to extend $\mathcal{R}$ to its Gaussian counterpart, while the Gaussian extension of $\mathcal{R}_{\mathrm{sp2}}$ can be obtained in a similar manner. We first map the random variables involved in the joint distribution (3.1) to a set of Gaussian random variables with the following customary constraints:

P1) W, distributed according to $\mathcal{N}(0, 1)$,

P2) $X_1 = \sqrt{P_1} W$,

P3) $\tilde{U}$, distributed according to $\mathcal{N}(0, \alpha\beta P_2)$,

P4) $\tilde{V}$, distributed according to $\mathcal{N}(0, \alpha\bar{\beta} P_2)$,

P5) $U = \tilde{U} + \lambda_1 W$,

P6) $V = \tilde{V} + \lambda_2 W$,

P7) $X_2 = \tilde{U} + \tilde{V} + \sqrt{\bar{\alpha} P_2} W$,

where $\alpha, \beta \in [0, 1]$, $\alpha + \bar{\alpha} = 1$, $\beta + \bar{\beta} = 1$, and $\lambda_1, \lambda_2 \in [0, +\infty)$. The variables W, $\tilde{U}$, and $\tilde{V}$ are assumed to be mutually statistically independent. The mappings P3)–P6) are used to extend the Gel'fand-Pinsker coding to the Gaussian case. The coefficient λ_1 (or λ_2) determines the degree of correlation between the Gaussian random variables W and U (or V), which plays the same role as the variable α in [68]. The input-output relationship of the GIC-DMS can now be described by

$$Y_1 = \left(\sqrt{P_1} + \sqrt{c_{21}\bar{\alpha}P_2}\right) W + \sqrt{c_{21}}\tilde{U} + \sqrt{c_{21}}\tilde{V} + Z_1, \tag{3.43}$$

$$Y_2 = \tilde{U} + \tilde{V} + \left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right) W + Z_2. \tag{3.44}$$

We fix the time sharing random variable Q as a constant. The issue of how this time-sharing random variable affects the achievable rate region is well addressed in [47]. The rate region $\mathcal{R}$ can be extended to its Gaussian counterpart, $\mathcal{G}$, by evaluating the respective mutual information terms in (3.2)–(3.6) with respect to the mappings defined by P1)–P7), (3.43), and (3.44). The evaluation can be readily done with the following two covariance matrices:

$$\begin{aligned}\mathbf{\Sigma}_{WUY_1} &:= \begin{pmatrix} E\{W^2\} & E\{WU\} & E\{WY_1\} \\ E\{WU\} & E\{U^2\} & E\{UY_1\} \\ E\{WY_1\} & E\{UY_1\} & E\{Y_1^2\} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \lambda_1 & \eta_1 \\ \lambda_1 & \alpha\beta P_2 + \lambda_1^2 & \sqrt{c_{21}}\alpha\beta P_2 + \eta_1 \\ \eta_1 & \sqrt{c_{21}}\alpha\beta P_2 + \eta_1 & \eta_1^2 + c_{21}\alpha P_2 + 1 \end{pmatrix},\end{aligned}$$

$$\begin{aligned}\mathbf{\Sigma}_{UVY_2} &:= \begin{pmatrix} E\{U^2\} & E\{UV\} & E\{UY_2\} \\ E\{UV\} & E\{V^2\} & E\{VY_2\} \\ E\{UY_2\} & E\{VY_2\} & E\{Y_2^2\} \end{pmatrix} \\ &= \begin{pmatrix} \alpha\beta P_2 + \lambda_1^2 & \lambda_1\lambda_2 & \alpha\beta P_2 + \lambda_1\eta_2 \\ \lambda_1\lambda_2 & \alpha\bar{\beta}P_2 + \lambda_2^2 & \alpha\bar{\beta}P_2 + \lambda_2\eta_2 \\ \alpha\beta P_2 + \lambda_1\eta_2 & \alpha\bar{\beta}P_2 + \lambda_2\eta_2 & \alpha P_2 + \eta_2^2 + 1 \end{pmatrix},\end{aligned}$$

where

$$\eta_1 := \sqrt{P_1} + \sqrt{c_{21}\bar{\alpha}P_2},$$
$$\eta_2 := \sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1},$$

and $E\{\cdot\}$ denotes the expectation of a random variable.

3.5.2.2 Gaussian Extension of $\mathcal{R}_{\text{suc}}$

For illustration and comparison purpose, we next show how to obtain the Gaussian counterpart of $\mathcal{R}_{\text{suc}}$ in detail. Following the first step in the previous derivation, we also map the random variables involved in (3.20) to the Gaussian ones with the following constraints:

M1) W, distributed according to $\mathcal{N}(0, 1)$,

M2) $X_1 = \sqrt{P_1}W$,

M3) U, distributed according to $\mathcal{N}(0, \alpha\beta P_2)$,

M4) $\tilde{V}$, distributed according to $\mathcal{N}(0, \alpha\bar{\beta}P_2)$,

M5) $V = \tilde{V} + \lambda W$,

M6) $X_2 = U + \tilde{V} + \sqrt{\bar{\alpha}P_2}W$,

where $\alpha, \beta \in [0, 1]$, $\alpha + \bar{\alpha} = 1$, $\beta + \bar{\beta} = 1$, $\lambda \in [0, +\infty)$, and W, U and $\tilde{V}$ are mutually independent. Using the mappings defined by M1)–M6), we express the input-output relationship for the GIC-DMS as:

$$Y_1 = \left(\sqrt{P_1} + \sqrt{c_{21}\bar{\alpha}P_2}\right)W + \sqrt{c_{21}}U + \sqrt{c_{21}}\tilde{V} + Z_1, \tag{3.45}$$

$$Y_2 = U + \tilde{V} + \left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right)W + Z_2. \tag{3.46}$$

Let $\mathcal{G}_{\text{suc}}(\alpha,\beta)$ denote the set of all non-negative rate pairs (R_1, R_2) such that

$$R_1 \leq \frac{1}{2}\log_2\left(1+\frac{\left(\sqrt{P_1}+\sqrt{c_{21}\bar{\alpha}P_2}\right)^2}{c_{21}\alpha\bar{\beta}P_2+1}\right), \tag{3.47}$$

$$R_2 \leq \frac{1}{2}\log_2(1+\alpha\bar{\beta}P_2)+\min\left\{\frac{1}{2}\log_2\left(1+\frac{c_{21}\alpha\beta P_2}{\left(\sqrt{P_1}+\sqrt{c_{21}\bar{\alpha}P_2}\right)^2+c_{21}\alpha\bar{\beta}P_2+1}\right),\right.$$
$$\left.\frac{1}{2}\log_2\left(1+\frac{\alpha\beta P_2}{\alpha\bar{\beta}P_2+\left(\sqrt{\bar{\alpha}P_2}+\sqrt{c_{12}P_1}\right)^2+1}\right)\right\}. \tag{3.48}$$

Define

$$\mathcal{G}_{\text{suc}} := \bigcup_{\alpha,\beta\in[0,1]} \mathcal{G}_{\text{suc}}(\alpha,\beta).$$

Theorem 3.4 The region $\mathcal{G}_{suc}$ is an achievable rate region for the GIC-DMS in standard form.

Proof: It suffices to prove that $\mathcal{G}_{\text{suc}}(\alpha,\beta)$ is achievable for any given $\alpha,\beta\in[0,1]$. Since this region $\mathcal{G}_{\text{suc}}$ is extended from $\mathcal{R}_{\text{suc}}$, we need evaluate the mutual information terms in (3.30) and (3.31). The righthand side of (3.47) can be readily obtained through a straightforward evaluation of $I(W;Y_1|U,Q)$ in (3.30). Recall that Q is a constant. It is also fairly straightforward to compute $I(U;Y_1|Q)$ and $I(U;Y_2|Q)$ in (3.31) to obtain the second term (the term involved in the minimum operation) in the righthand side of (3.48). We next evaluate the only remaining term $I(V;Y_2|U,Q)-I(V;W|Q)$ for a constant Q. Defining

$$\tilde{Y}_2 := \tilde{V}+\left(\sqrt{\bar{\alpha}P_2}+\sqrt{c_{12}P_1}\right)W+Z_2,$$

we have

$$\begin{aligned}I(V;Y_2|U)-I(V;W) &= h(Y_2|U)-h(Y_2|U,V)-I(V;W)\\ &= h(\tilde{Y}_2)-h(\tilde{Y}_2|V)-I(V;W)\end{aligned}$$

$$= h(\tilde{Y}_2) + h(V) - h(\tilde{Y}_2, V) - I(V;W). \tag{3.49}$$

With $V = \tilde{V} + \lambda W$, we evaluate (3.49) as

$$\begin{aligned} &I(V;Y_2|U) - I(V;W) \\ =&\frac{1}{2}\log_2\left(2\pi e\left(\alpha\bar{\beta}P_2 + \left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right)^2 + 1\right)\right) + \frac{1}{2}\log_2(2\pi e(\alpha\bar{\beta}P_2 + \lambda^2)) \\ &- \frac{1}{2}\log_2\left((2\pi e)^2\left[\left(\alpha\bar{\beta}P_2 + \left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right)^2 + 1\right)(\alpha\bar{\beta}P_2 + \lambda^2 P_1)\right.\right. \\ &\qquad \left.\left. - \left(\alpha\bar{\beta}P_2 + \lambda\left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right)\right)^2\right]\right) - \frac{1}{2}\log_2\left(1 + \frac{\lambda^2}{\alpha\bar{\beta}P_2}\right). \end{aligned} \tag{3.50}$$

Now, by simple calculus, we can find that when

$$\lambda = \frac{\alpha\bar{\beta}P_2\left(\sqrt{\bar{\alpha}P_2} + \sqrt{c_{12}P_1}\right)}{\alpha\bar{\beta}P_2 + 1}, \tag{3.51}$$

the term $I(V;Y|U) - I(V;W)$ can be maximized, and the maximum value is

$$\max[I(V;Y_2|U) - I(V;W)] = \frac{1}{2}\log_2(1 + \alpha\bar{\beta}P_2). \tag{3.52}$$

This is in parallel with the result in [68].

Therefore, the rate region $\mathcal{G}_{\text{suc}}(\alpha, \beta)$ is achievable for any pair $\alpha, \beta \in [0, 1]$, and the theorem follows. ■

In the following, we obtain two corollaries by setting $\beta = 0$ and $\beta = 1$ in Theorem 3.4, respectively.

Corollary 3.3 The rate region $\mathcal{G}_{sp1}$ is an achievable rate region for the GIC-DMS in standard form with

$$\mathcal{G}_{sp1} := \bigcup_{\alpha\in[0,1]} \mathcal{G}_{suc}(\alpha, 0),$$

i.e., $\mathcal{G}_{sp1}$ *is the union of the sets of non-negative rate pairs* (R_1, R_2) *satisfying*

$$\begin{aligned} R_1 &\le \frac{1}{2}\log_2\left(1+\frac{\left(\sqrt{P_1}+\sqrt{c_{21}\bar{\alpha}P_2}\right)^2}{c_{21}\alpha P_2+1}\right), \\ R_2 &\le \frac{1}{2}\log_2(1+\alpha P_2), \end{aligned}$$

over all $\alpha \in [0,1]$.

Corollary 3.4 The rate region $\mathcal{G}_{sp2}$ *is an achievable rate region for the GIC-DMS in standard form with*

$$\mathcal{G}_{sp2} := \bigcup_{\alpha\in[0,1]} \mathcal{G}_{suc}(\alpha, 1),$$

i.e., $\mathcal{G}_{sp2}$ *is the union of the sets of non-negative rate pairs* (R_1, R_2) *satisfying*

$$\begin{aligned} R_1 &\le \frac{1}{2}\log_2\left(1+\left(\sqrt{P_1}+\sqrt{c_{21}\bar{\alpha}P_2}\right)^2\right), \\ R_2 &\le \min\left\{\frac{1}{2}\log_2\left(1+\frac{c_{21}\alpha P_2}{\left(\sqrt{P_1}+\sqrt{c_{21}\bar{\alpha}P_2}\right)^2+1}\right),\right. \\ &\qquad\left.\frac{1}{2}\log_2\left(1+\frac{\alpha P_2}{\left(\sqrt{\bar{\alpha}P_2}+\sqrt{c_{12}P_1}\right)^2+1}\right)\right\}, \end{aligned}$$

over all $\alpha \in [0,1]$.

Remark 3.6 Corollaries 3.3 and 3.4 correspond to the Gaussian extensions of Corollaries 3.1 and 3.2 respectively. Particularly, the rate region depicted by Corollary 3.3 is the same as the rate regions given in [49, Theorem 4.1] and [50, Theorem 3.5]. It has been proven in both [49] and [50] that the rate region $\mathcal{G}_{sp1}$ *is indeed the capacity region for the GIC-DMS in the low interference gain regime, i.e.,* $c_{21} \le 1$. *In addition, Corollary 3.4 also implies Lemma 4.2 of [49].*

3.5.3 Numerical Examples

We next provide several numerical examples to illustrate the improvements of our achievable rate regions over the previously known results in [43, 49, 50]. We denote

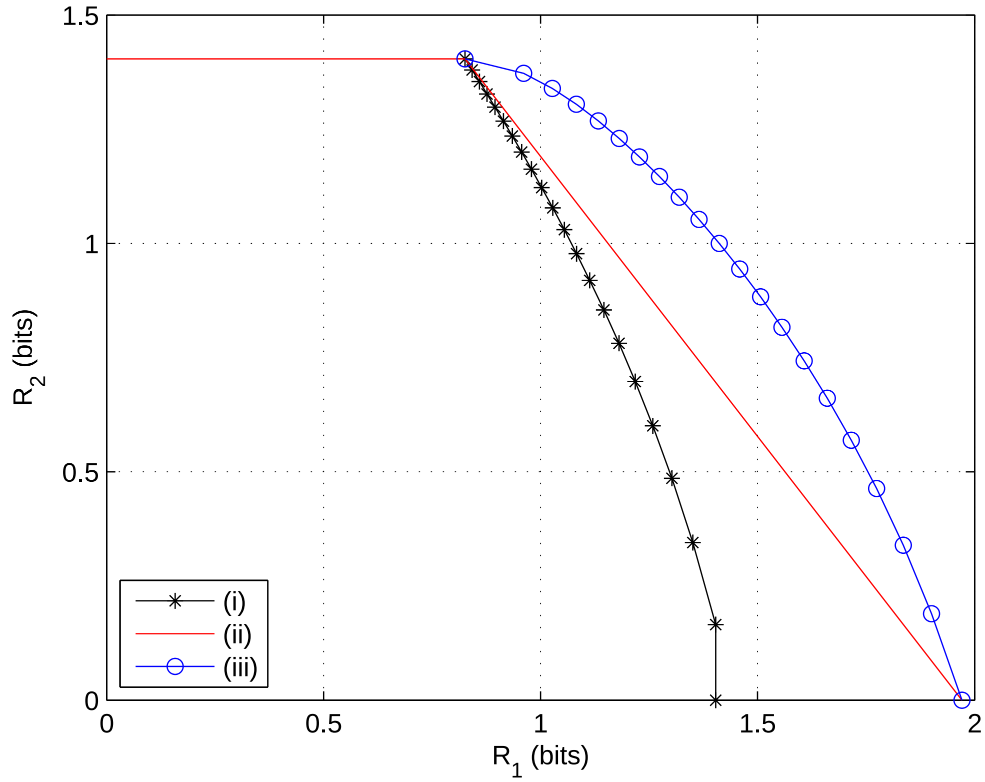

Figure 3.4: Achievable rate regions for GIC-DMS with setting: $P_1 = P_2 = 6$, $c_{21} = 0.3$, $c_{12} = 0$. (i) gives the rate region $\mathcal{G}_{\rm dmt1}$; (ii) gives the rate region $\mathcal{G}_{\rm dmt2}$; (iii) gives our rate region $\mathcal{G}_{\rm sp1}$.

the achievable rate regions obtained in [43, Theorem 1] and [43, Corollary2] by $\mathcal{G}_{\rm dmt1}$ and $\mathcal{G}_{\rm dmt2}$ respectively.

Comparing with Rate Regions in [43]: Fig. 3.4 compares the rate regions $\mathcal{G}_{\rm dmt1}$, $\mathcal{G}_{\rm dmt2}$, and $\mathcal{G}_{\rm sp1}$ for an extreme case in which receiver 2 does not experience any interference from sender 1, i.e., $c_{12} = 0$. As can be seen from Fig. 3.4, the rate region $\mathcal{G}_{\rm sp1}$ strictly includes $\mathcal{G}_{\rm dmt1}$, as well as $\mathcal{G}_{\rm dmt2}$ which is obtained through time-sharing between $\mathcal{G}_{\rm dmt1}$ and the fully-cooperative rate point. The coding scheme used to establish $\mathcal{G}_{\rm dmt1}$ incurs certain rate loss due to the fact that sender 2 does not use its power to help sender 1's transmissions even though it has complete and non-causal knowledge about the message being transmitted by sender 1. In contrast, our proposed coding scheme allows sender 2 to use superposition coding to help sender 1, and thus yields an improved rate region.

Comparing with Rate Regions in [49, 50]: Fig. 3.5 compares the Gaussian extensions of rate regions $\mathcal{R}$, $\mathcal{R}_{\text{sp1}}$, and $\mathcal{R}_{\text{sp2}}$ in the high-interference-gain regime, i.e., $c_{21} > 1$. Note that the Gaussian counterpart of $\mathcal{R}_{\text{sp2}}$ includes the set of achievable rate pairs in [49, Lemma 4.2] as a subset. As can be observed in Fig. 3.5, our achievable rate region in Theorem 3.1 offers considerable improvements over the rate regions in [49] and [50] under two different parameter settings.

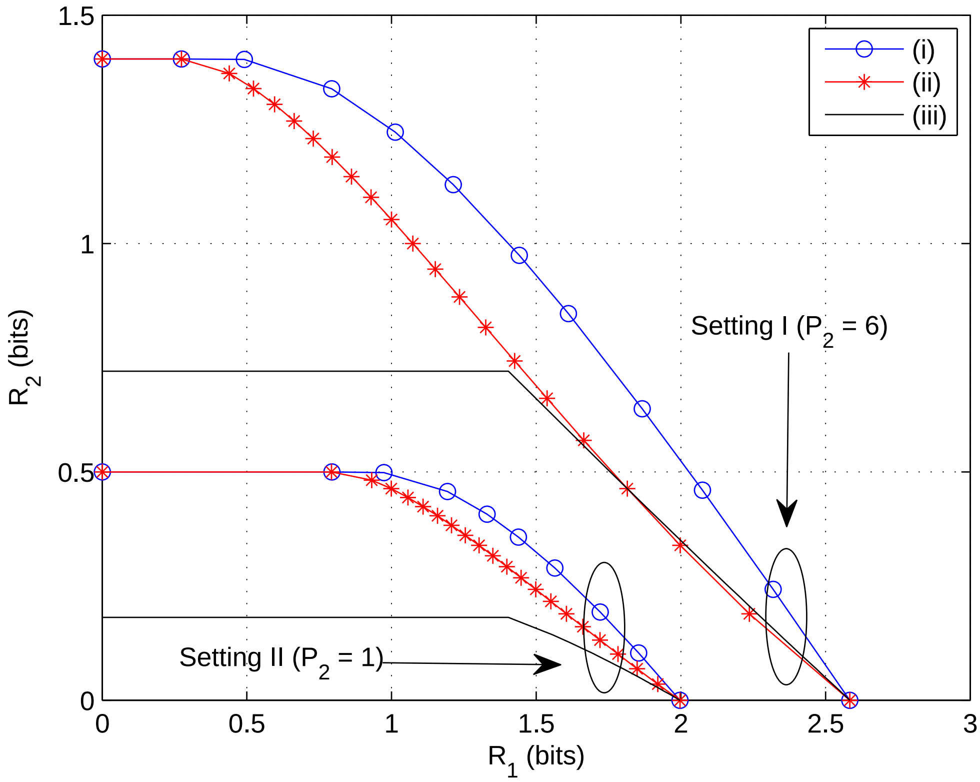

Figure 3.5: Achievable rate regions for the GIC-DMS two different settings: (I) $P_1 = 6$, $P_2 = 6$, $c_{21} = 2$, $c_{12} = 0.3$; (II) $P_1 = 6$, $P_2 = 1$, $c_{21} = 2$, $c_{12} = 0.3$. (i) gives the rate regions, $\mathcal{G}$, Gaussian counter part of Theorem 3.1; (ii) gives the rate regions $\mathcal{R}_{\text{sp1}}$; (iii) gives the rate regions $\mathcal{R}_{\text{sp2}}$.

3.6 Conclusions

In this chapter, we have investigated the IC-DMS from an information theoretic perspective. We have developed a coding scheme that combines the advantages of cooperative coding, collaborative coding, and Gel'fand-Pinsker coding. With this coding scheme, we have derived a new achievable rate region for such a channel, which not only includes two existing results as special cases, but also exceeds them in the high-interference-gain regime.

Chapter 4

Discrete Memoryless Interference Channels With Perfect Feedback

In this chapter, we investigate the ICF, in which the receivers are able to send the received channel outputs to their respective pairing sender perfectly. We develop a block Markov coding scheme which exploits the perfect feedback at each sender in order to achieve cooperation between the two senders. We also derive a corresponding achievable rate region for the ICF by analyzing the probabilities of error of the proposed coding scheme, and both the implicit and explicit descriptions of the achievable rate region are presented.

4.1 Introduction

Since Gaarder and Wolf revealed in [69] that feedback can increase the capacity region of a class of MACs, the information-theoretic study of multi-terminal networks with feedback has attracted considerable attention, and many results have been obtained for feedback settings of MACs [10, 9, 70], BCs [71], RCs[21, 72] and ICs [73, 74]. However, the capacity regions for those channels with feedback in *general* settings remains open except for some special cases, such as the Gaussian MAC with perfect feedback [9] and the degraded BC with perfect feedback [75]. For the IC, feedback has been considered mostly for the Gaussian case [73, 74], while there were few achievable rate results obtained for the discrete memoryless case prior to the work in this chapter.

We first review some well-known information-theoretic results on the genetic IC without feedback. The to date best achievable result for the IC was obtained by Han and Kobayashi in [28], which is recently simplified by Chong, Motani, and Garg in [48]. Their result is attributable to Carleial's notion [46] that each receiver is allowed to crossly observe partial information from non-pairing senders, given that the entire codebook is exposed to every receiver. One can interpret their coding schemes in [28, 48] as a type of collaborative coding (in contrast with the correlation induced cooperative coding) in the sense that, the senders sperate the information to transmit into two parts and encode each differently such that each receiver can decode the two part of information from its pairing sender and one of the two parts from the other sender. Being able to decode part of the information from the interference signal, each receiver can obtain a channel with weakened effective interference for the intended information from its pairing sender. The coding scheme is also sometimes termed as the rate splitting.

Without feedback, the current best achievable rate region for the IC is obtained with the collaborative coding scheme described above. Now, we consider the discrete memoryless ICF. When it has access to the channel output of its pairing receiver, each sender is naturally more capable of decoding the crossly observable information from the other sender than the its own pairing receiver, because the sender knows what it transmitted as additional side information. Due to this, each sender can now send the part of information, which is crossly observable, at a high rate such that, the corresponding intended receiver is not able to decode at the end of transmission of the current block but the other sender can. In the next block, after decoding from the perfect feedback the crossly observable information transmitted by the other sender, the two senders can cooperate to help the two receivers to resolve the remaining uncertainty in the previous block. A block Markov coding scheme can be developed for the discrete memoryless ICF based on this idea.

4.2 Channel Model and Preliminaries

A generic two-user discrete memoryless ICF is characterized by two finite input alphabets, $\mathcal{X}_1$ and $\mathcal{X}_2$; two finite output alphabets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$; and the conditional probabilities $p(y_1, y_2|x_1, x_2)$ on $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. The channel is memoryless in the sense that

$$p(y_{1,t}, y_{2,t}|x_{1,t}, x_{2,t}, x_{1,t-1}, x_{2,t-1}, ...) = p(y_{1,t}, y_{2,t}|x_{1,t}, x_{2,t}) \tag{4.1}$$

for every discrete time instant t in a synchronous transmission. Each receiver is assumed to feed back its received signal to the pairing sender in a causal and noiseless manner.

As shown in Figure 4.1, sender i, $i = 1, 2$, wishes to transmit a message (message index), $w_i \in \mathcal{W}_i = \{1, ..., M_i\}$, to receiver i. The message W_i is independently and uniformly generated over its index set $\mathcal{W}_i$.

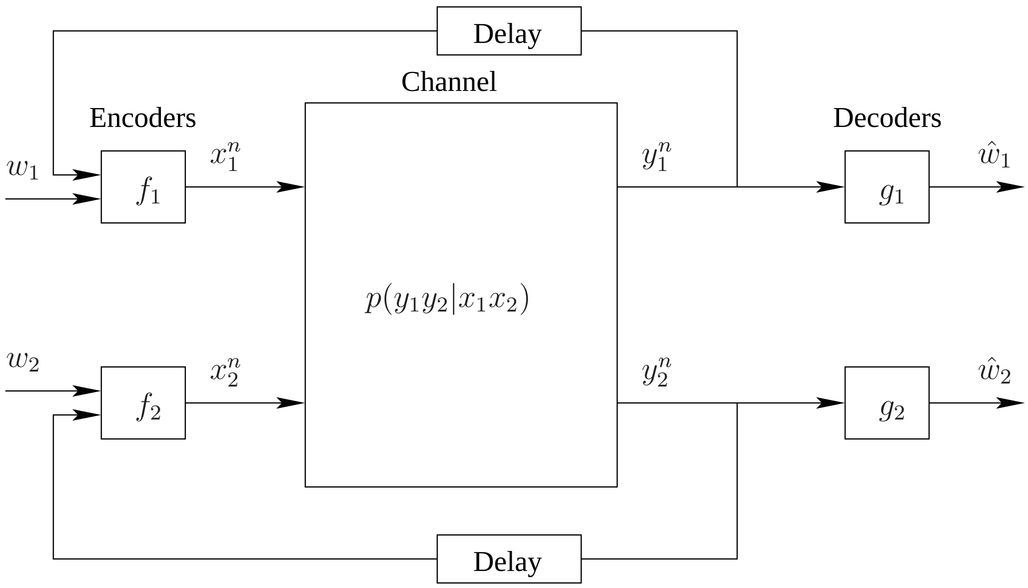

Figure 4.1: Interference channel with perfect feedback.

Definition 4.1 We claim that an $(M_1, M_2, n, P_e^{(n)})$ *feedback code exists for the dis-*

crete memoryless ICF, if and only if there exists a collection of encoding functions

$$f_{1,t} : \mathcal{W}_1 \times \{\mathcal{Y}_{1,1}, ..., \mathcal{Y}_{1,t-1}\} \rightarrow \mathcal{X}_{1,t},$$
$$f_{2,t} : \mathcal{W}_2 \times \{\mathcal{Y}_{2,1}, ..., \mathcal{Y}_{2,t-1}\} \rightarrow \mathcal{X}_{2,t},$$

where $t = 1, ..., n$*; and two decoding functions*

$$g_1 : \mathcal{Y}_1^n \rightarrow \hat{\mathcal{W}}_1, \quad g_2 : \mathcal{Y}_2^n \rightarrow \hat{\mathcal{W}}_2,$$

such that $\max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} \leq P_e^{(n)}$*, where* $P_{e,i}^{(n)}$*,* $i = 1, 2$*, denotes the average decoding error probability of decoder* i*, and is computed by one of the following:*

$$P_{e,1}^{(n)} = \frac{1}{M_1 M_2} \sum_{w_1 w_2} \Pr(\hat{w}_1 \neq w_1 | (w_1, w_2) \text{ were sent}),$$
$$P_{e,2}^{(n)} = \frac{1}{M_1 M_2} \sum_{w_1 w_2} \Pr(\hat{w}_2 \neq w_2 | (w_1, w_2) \text{ were sent}).$$

Definition 4.2 A non-negative rate pair (R_1, R_2) *is achievable for the discrete memoryless ICF, if there exists a sequence of* $(2^{nR_1}, 2^{nR_2}, n, P_e^{(n)})$ *codes such that* $P_e^{(n)} \rightarrow 0$ *as* $n \rightarrow \infty$*.*

Definition 4.3 The capacity region for the discrete memoryless ICF is defined as the closure of the set of all the achievable rate pairs, while an achievable rate region for the channel is a subset of the capacity region.

4.3 An Achievable Rate Region for the ICF

As mentioned in Section 4.1, we apply rate splitting at both senders, i.e., we let $w_1 = (w_{12}, w_{11})$, $w_{12} \in \{1, ..., 2^{nR_{12}}\}$, $w_{11} \in \{1, ..., 2^{nR_{11}}\}$; and $w_2 = (w_{21}, w_{22})$, $w_{21} \in \{1, ..., 2^{nR_{21}}\}$, $w_{22} \in \{1, ..., 2^{nR_{22}}\}$. We also follow Carleial's notion that the receivers are allowed to crossly observe partial information from the non-pairing senders; and in our setting receiver 1 will be able to decode (w_{12}, w_{11}) as well as w_{21}, while receiver 2 will be able to decode (w_{21}, w_{22}) as well as w_{12} similarly. If the receivers can successfully decode (w_{12}, w_{11}) and (w_{21}, w_{22}) respectively with high probability, we can claim that $(R_1, R_2) = (R_{12} + R_{11}, R_{21} + R_{22})$ is achievable for the discrete memoryless ICF. We present our main result as follows.

Denote by $\mathcal{P}^*$ the set of joint probability distributions $p(\cdot)$ that can factor in the form

$$\begin{aligned} p(u_0, u_1, u_2, x_1, x_2, y_1, y_2) = &\, p(u_0)p(u_1|u_0)p(u_2|u_0) \\ &\cdot p(x_1|u_1, u_0)p(x_2|u_2, u_0)p(y_1, y_2|x_1, x_2), \end{aligned} \tag{4.2}$$

where u_0, u_1 and u_2 are realizations of three auxiliary random variables U_0, U_1 and U_2 defined on arbitrary finite sets $\mathcal{U}_0$, $\mathcal{U}_1$ and $\mathcal{U}_2$.

Next, denote by $\mathcal{R}^{(1)}_{fm}(p)$ the set of all non-negative quadruples $(R_{12}, R_{11}, R_{21}, R_{22})$ such that

$$R_{21} \leq I(U_2; Y_1|X_1, U_1, U_0), \tag{4.3}$$

$$R_{11} \leq I(X_1; Y_1|U_1, U_2, U_0), \tag{4.4}$$

$$R_{12} + R_{11} \leq \min\{I(U_0; Y_1), I(U_0; Y_2)\} + I(U_1X_1; Y_1|U_2, U_0), \tag{4.5}$$

$$R_{11} + R_{21} \leq \min\{I(U_0; Y_1), I(U_0; Y_2)\} + I(U_2, X_1; Y_1|U_1, U_0), \tag{4.6}$$

$$R_{12} + R_{11} + R_{21} \leq \min\{I(U_0; Y_1), I(U_0; Y_2)\} + I(U_2, U_1, X_1; Y_1|U_0), \tag{4.7}$$

for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$. Similarly, we denote by $\mathcal{R}^{(2)}_{fm}(p)$ the set of all non-negative quadruples $(R_{12}, R_{11}, R_{21}, R_{22})$ such that inequalities (4.3)–(4.7) with subscripts 1 and 2 swapped everywhere are satisfied for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$. We define $\mathcal{R}_f(p)$ as the set of all rate pairs (R_1, R_2) such that $(R_1, R_2) = (R_{12} + R_{11}, R_{21} + R_{22})$ and

$$(R_{12}, R_{11}, R_{21}, R_{22}) \in \mathcal{R}^{(1)}_{fm}(p) \cap \mathcal{R}^{(2)}_{fm}(p), \tag{4.8}$$

for a given $p(\cdot) \in \mathcal{P}^*$.

Theorem 4.1 The rate region

$$\mathcal{R}_f := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_f(p)$$

is an achievable rate region for the discrete memoryless ICF.

Proof: To prove the entire region $\mathcal{R}_f$ is achievable, it is sufficient to prove the region $\mathcal{R}_f(p)$ is achievable for some fixed joint distribution $p(\cdot) \in \mathcal{P}^*$. Let us fix a joint distribution $p(\cdot)$ that factors in the form of (4.2). We construct a block Markov coding scheme as follows. The coding scheme consists of $B+1$ blocks of transmissions, with each block consisting of n channel uses. We first generate two statistically independent codebooks, by repeating the codebook generation process described below twice. The two codebooks are used in an alternative manner such that the error events that may happen during the decoding in two consecutive blocks are independent of each other.

[Codebook Generation.] First, generate 2^{nR_0} independent codewords $\mathbf{U}_0(i)$, $i \in \{1,...,2^{nR_0}\}$, according to $\prod_{t=1}^{n} p(u_{0,t})$. At encoder 1, for each codeword $\mathbf{u}_0(i)$, $i \in \{1,...,2^{nR_0}\}$, generate $2^{nR_{12}}$ independent codewords $\mathbf{U}_1(i,j)$, $j \in \{1,...,2^{nR_{12}}\}$ according to $\prod_{t=1}^{n} p(u_{1,t}|u_{0,t})$. Subsequently, for each pair of codewords $(\mathbf{u}_0(i), \mathbf{u}_1(i,j))$, $i \in \{1,...,2^{nR_0}\}$, $j \in \{1,...,2^{nR_{12}}\}$, generate $2^{nR_{11}}$ independent codewords $\mathbf{X}_1(i,j,k)$, $k \in \{1,...,2^{nR_{11}}\}$, according to $\prod_{t=1}^{n} p(x_{1,t}|u_{1,t},u_{0,t})$. Similarly at encoder 2, for each codeword $\mathbf{u}_0(i)$, $i \in \{1,...,2^{nR_0}\}$, generate $2^{nR_{21}}$ independent codewords $\mathbf{U}_2(i,l)$, $l \in \{1,...,2^{nR_{21}}\}$ according to $\prod_{t=1}^{n} p(u_{2,t}|u_{0,t})$. Subsequently, for each codeword pair $(\mathbf{u}_0(i), \mathbf{u}_2(i,l))$, $i \in \{1,...,2^{nR_0}\}$, $l \in \{1,...,2^{nR_{21}}\}$, generate $2^{nR_{22}}$ independent codewords $\mathbf{X}_2(i,l,m)$, $m \in \{1,...,2^{nR_{22}}\}$, according to $\prod_{t=1}^{n} p(x_{2,t}|u_{2,t},u_{0,t})$.

Note that the above codebook generation process is repeated twice.

Next, uniformly distribute the $2^{n(R_{12}+R_{21})}$ index pairs (j,l) into 2^{nR_0} bins. The entire codebook is then revealed to both receivers.

[Encoding and transmission.] Assume that transmission of block $b-1$ is just finished. Before the transmission of the next block, bth block, sender 1 will try to decode the message $w_{21}^{(b-1)}$ sent from sender 2 in the $(b-1)$th block from the feedback $\mathbf{y}_1^{(b-1)}$, by looking for a *unique* $\hat{w}_{21}^{(b-1)}$ such that

$$\begin{aligned}(\mathbf{u}_0(\hat{i}^{(b-2)}), \mathbf{u}_1(\hat{i}^{(b-2)}, w_{12}^{(b-1)}), \mathbf{x}_1(\hat{i}^{(b-2)}, w_{12}^{(b-1)}, w_{11}^{(b-1)}),\\ \mathbf{u}_2(\hat{i}^{(b-2)}, \hat{w}_{21}^{(b-1)}), \mathbf{y}_1^{(b-1)}) \in A_\epsilon^{(n)}.\end{aligned} \tag{4.9}$$

If successful, sender 1 decodes $w_{21}^{(b-1)} = \hat{w}_{21}^{(b-1)}$; otherwise, an error is declared.

Sender 1 then looks for the index $i^{(b-1)}$ of the bin which contains the index pair $(w_{12}^{(b-1)}, w_{21}^{(b-1)})$.

Similarly, sender 2 needs to decode the message from sender 1, $w_{12}^{(b-1)}$, and obtain the same bin index $i^{(b-1)}$ in a symmetrical manner.

Assume that $(w_{12}^{(b)}, w_{11}^{(b)})$ and $(w_{21}^{(b)}, w_{22}^{(b)})$ are the messages to be transmitted in the bth block. Sender 1 will transmit $\mathbf{x}_1(i^{(b-1)}, w_{12}^{(b)}, w_{11}^{(b)})$, and sender 2 will transmit $\mathbf{x}_2(i^{(b-1)}, w_{21}^{(b)}, w_{22}^{(b)})$. The transmissions are assumed to be perfectly synchronized.

[Decoding.] At the end of the bth block transmission, receiver 1 tries to decode the bin index $i^{(b-1)}$ from the channel output of current block $\mathbf{y}_1^{(b)}$ first. Receiver 1 declares $i^{(b-1)} = \hat{i}^{(b-1)}$ if there exists a unique bin index $\hat{i}^{(b-1)}$ such that we have

$$(\mathbf{u}_0(\hat{i}^{(b-1)}), \mathbf{y}_1^{(b)}) \in A_\epsilon^{(n)},$$

where $A_\epsilon^{(n)}$ is the typical set [40]. Otherwise, an error is declared.

Next, receiver 1 will decode $(w_{12}^{(b-1)}, w_{11}^{(b-1)}) = (\hat{w}_{12}, \hat{w}_{11})$ if there exists a unique triple $(\hat{w}_{12}, \hat{w}_{11}, \hat{w}_{21})$ such that $(\hat{w}_{12}, \hat{w}_{21})$ are from bin with index $i^{(b-1)}$ and

$$(\mathbf{u}_0(i^{(b-2)}), \mathbf{u}_1(i^{(b-2)}, \hat{w}_{12}), \mathbf{x}_1(i^{(b-2)}, \hat{w}_{12}, \hat{w}_{11}), \mathbf{u}_2(i^{(b-2)}, \hat{w}_{21}), \mathbf{y}_1^{(b-1)}) \in A_\epsilon^{(n)}; \tag{4.10}$$

otherwise, a decoding error is declared. Note that the bin index $i^{(b-2)}$ is assumed to have been decoded successfully at the end of the $(b-1)$th block transmission.

In a similar manner, receiver 2 will first decode the bin index $i^{(b-1)}$ and then decode the messages $(w_{21}^{(b-1)}, w_{22}^{(b-1)})$.

[Evaluation of probability of error.] Assume that there is no decoding error made by both senders and receivers at the end of transmission of the bth block, i.e., sender 1 decodes $w_{21}^{(1)}, ..., w_{21}^{(b-1)}$ and obtains $i^{(1)}, ..., i^{(b-1)}$, and sender 2 decodes $w_{12}^{(1)}, ..., w_{12}^{(b-1)}$ and obtains $i^{(1)}, ..., i^{(b-1)}$, correctly; receiver 1 decodes $i^{(1)}, ..., i^{(b-2)}$ and $(w_{12}^{(1)}, w_{11}^{(1)}, w_{21}^{(1)}), ..., (w_{12}^{(b-2)}, w_{11}^{(b-2)}, w_{21}^{(b-2)})$, and receiver 2 decodes $i^{(1)}, ..., i^{(m-2)}$ and $(w_{21}^{(1)}, w_{22}^{(1)}, w_{12}^{(1)}), ..., (w_{21}^{(b-2)}, w_{22}^{(b-2)}, w_{12}^{(b-2)})$, correctly. Since the channel is symmetric, we can analyze the probability of error made by sender 1 and receiver 1 only. Similarly due to the symmetry of the codebook generation, the probability of error is codeword independent. We further assume that the message vector $(w_{12}^{(b-1)}, w_{11}^{(b-1)}, w_{21}^{(b-1)}, w_{22}^{(b-1)}) = (1, 1, 1, 1)$ was encoded and sent through $(b-1)$th block, and we assume that $(w_{12}^{(b-1)}, w_{21}^{(b-1)}) = (1, 1)$ are in bin 1, i.e., $i^{(b-1)} = 1$.

Before we proceed, we define the following events:

$$
\begin{aligned}
&F_{j,l}^{b} := (j, l) \in \text{ bin } b, \\
&E_{j,k,l} := \\
&\quad (\mathbf{U}_0(i^{(b-2)}), \mathbf{U}_1(i^{(b-2)}, j), \mathbf{X}_1(i^{(b-2)}, j, k), \mathbf{U}_2(i^{(b-2)}, l), \mathbf{Y}_1^{(b-1)}) \in A_\epsilon^{(n)}.
\end{aligned}
$$

At the end of the transmission of the bth block, sender 1 needs to decode $w_{21}^{(b)}$ correctly to sustain the block transmission. Decoding error occurs when one of the following two events happens: 1) the codewords transmitted are not jointly typical with the channel output sequence, i.e.,

$$(\mathbf{u}_0(1), \mathbf{u}_1(1, w_{12}^{(b)}), \mathbf{x}_1(1, w_{12}^{(b)}, w_{11}^{(b)}), \mathbf{u}_2(1, w_{21}^{(b)}), \mathbf{y}_1^{(b)}) \notin A_\epsilon^{(n)};$$

2) sender 1 finds another $\hat{w}_{21}^{(b)} \neq w_{21}^{(b)}$ such that

$$(\mathbf{u}_0(1), \mathbf{u}_1(1, w_{12}^{(b)}), \mathbf{x}_1(1, w_{12}^{(b)}, w_{11}^{(b)}), \mathbf{u}_2(1, \hat{w}_{21}^{(b)}), \mathbf{y}_1^{(b)}) \in A_\epsilon^{(n)}.$$

According to the asymptotic equipartition property (AEP), when the code length n is sufficiently large, the probability of event 1 becomes arbitrarily small. The probability of event 2 also becomes arbitrarily small, when the code length n is sufficiently large and the following is satisfied:

$$R_{21} \leq I(U_2; Y_1|X_1, U_1, U_0).$$

Similarly, a constraint on the rate R_{12}:

$$R_{12} \leq I(U_1; Y_2|X_2, U_2, U_0),$$

also needs to be satisfied such that sender 2 is able to decode $w_{12}^{(b)}$ correctly.

Receiver 1 first tries to decode the bin index $i^{(b-1)}$ from $\mathbf{y}_1^{(b)}$ and then decodes $(w_{12}^{(b-1)}, w_{11}^{(b-1)}, w_{21}^{(b-1)})$ from $\mathbf{y}_1^{(b-1)}$. Recall that we assume that $(w_{12}^{(b-1)}, w_{11}^{(b-1)}, w_{21}^{(b-1)}) = (1, 1, 1, 1)$ and $i^{(b-1)} = 1$. An error may be made by receiver 1 when one of the following two situation occurs: 1) the transmitted codewords are not jointly typical with the channel output sequence, i.e.,

$$(\mathbf{u}_0(1), \mathbf{y}_1^{(b)}) \notin A_\epsilon^{(n)};$$

2) receiver 1 finds another $\hat{i}^{(b-1)} \neq 1$ such that $(\mathbf{u}_0(\hat{i}^{(b-1)}), \mathbf{y}_1^{(b)}) \in A_\epsilon^{(n)}$. According to AEP, when the code length n is sufficiently large, the probability of situation 1 becomes arbitrarily small. To drive the probability of situation 2 also arbitrarily small, the following rate constraint needs to be satisfied:

$$R_0 \leq I(U_0; Y_1).$$

Similarly, the rate constraint:

$$R_0 \leq I(U_0; Y_2),$$

also need to be satisfied for decoder 2 to successfully decode $i^{(b-1)}$.

Next, receiver 1 tries to decode $(w_{12}^{(b-1)}, w_{11}^{(b-1)}, w_{21}^{(b-1)})$ from the receiver output

$\mathbf{y}_1^{(b-1)}$. An error is made if one of the following two situation occurs: 1) the transmitted codewords are not jointly typical with the received channel output sequence, i.e.,

$$(\mathbf{u}_0(i^{(b-2)}), \mathbf{u}_1(i^{(b-2)}, 1), \mathbf{x}_1(i^{(b-2)}, 1, 1), \mathbf{u}_2(i^{(b-2)}, 1), \mathbf{y}_1^{(b-1)}) \notin A_\epsilon^{(n)};$$

2) there exists another message vector $(\hat{w}_{12}^{(b-1)}, \hat{w}_{11}^{(b-1)}, \hat{w}_{21}^{(b-1)}) \neq (1, 1, 1)$ such that the two events

$$F^1_{\hat{w}_{12}^{(b-1)} \hat{w}_{21}^{(b-1)}} \text{ and } E_{\hat{w}_{12}^{(b-1)} \hat{w}_{11}^{(b-1)} \hat{w}_{21}^{(b-1)}},$$

happen simultaneously. When code length n is sufficiently large, the probability of situation 1 can be made arbitrarily small. The probability of situation 2 can be bounded as follows:

$$\begin{aligned}
&\Pr(\text{situation 2}) \\
&= \Pr\left(\bigcup_{(\hat{w}_{12}^{(b-1)}, \hat{w}_{11}^{(b-1)}, \hat{w}_{21}^{(b-1)}) \neq (1,1,1)} \left(F^1_{\hat{w}_{12}^{(b-1)} \hat{w}_{21}^{(b-1)}} \cap E_{\hat{w}_{12}^{(b-1)} \hat{w}_{11}^{(b-1)} \hat{w}_{21}^{(b-1)}} \right) \right) \\
&\leq 2^{nR_{12}} \Pr(F^1_{2,1}) \Pr(E_{2,1,1}) + 2^{nR_{11}} \Pr(E_{1,2,1}) + 2^{nR_{21}} \Pr(F^1_{1,2}) \Pr(E_{1,1,2}) \\
&\quad + 2^{n(R_{12}+R_{11})} \Pr(F^1_{2,1}) \Pr(E_{2,2,1}) + 2^{n(R_{11}+R_{21})} \Pr(F^1_{1,2}) \Pr(E_{1,2,2}) \\
&\quad + 2^{n(R_{12}+R_{21})} \Pr(F^1_{2,2}) \Pr(E_{2,1,2}) + 2^{n(R_{12}+R_{11}+R_{21})} \Pr(F^1_{2,2}) \Pr(E_{2,2,2}). \qquad (4.11)
\end{aligned}$$

Since the index pairs (j, l) are uniformly distributed into 2^{nR_0} bins, we have

$$\Pr(F^1_{1,2}) = \Pr(F^1_{2,1}) = \Pr(F^1_{2,2}) = 2^{-nR_0}. \qquad (4.12)$$

We next evaluate $E_{2,1,1}$, $E_{1,2,1}$, $E_{1,1,2}$, $E_{2,2,1}$, $E_{1,2,2}$, $E_{2,1,2}$, and $E_{2,2,2}$ by repeatedly applying Theorem 14.2.3 in [40], and then substitute the results into inequality (4.11). It follows that

$$\Pr(\text{situation 2}) \leq 2^{nR_{12}} 2^{-nR_0} 2^{-nI(U_1, X_1; Y_1 | U_2, U_0)} + 2^{nR_{11}} 2^{-nI(X_1; Y_1 | U_1, U_2, U_0)}$$

$$+ 2^{nR_{21}}2^{-nR_0}2^{-nI(U_2;Y_1|X_1,U_1,U_0)} + 2^{n(R_{12}+R_{11})}2^{-nR_0}2^{-nI(U_1,X_1;Y_1|U_2,U_0)}$$
$$+ 2^{n(R_{11}+R_{21})}2^{-nR_0}2^{-nI(U_2,X_1;Y_1|U_1,U_0)} + 2^{n(R_{12}+R_{21})}2^{-nR_0}2^{-nI(U_1,X_1,U_2;Y_1|U_0)}$$
$$+ 2^{n(R_{12}+R_{11}+R_{21})}2^{-nR_0}2^{-nI(U_1,X_1U_2;Y_1|U_0)}. \tag{4.13}$$

It is then straightforward to verify that when inequalities (4.7)–(4.3) are satisfied and code length n is sufficiently large, Pr(situation 2) $\to 0$. The same analysis is applied to receiver 2 similarly.

Therefore, we can conclude that the region $\mathcal{R}_f(p)$ is achievable for some fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, and the theorem follows. ■

The description of the achievable rate region given in Theorem 4.1 is in an implicit form where R_1 or R_2 are not present explicitly. Nevertheless, an explicit region can be obtained from the implicit one as follows.

Theorem 4.2 The rate region

$$\mathcal{R}_f := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_f(p)$$

is an achievable rate region for the discrete memoryless ICF, where $\mathcal{R}_f(p)$ *is the set of all rate pairs* (R_1, R_2) *such that*

$$R_1 \le C_0 + I(U_1, X_1; Y_1|U_2, U_0), \tag{4.14}$$
$$R_1 \le I(U_1; Y_2|X_2, U_2, U_0) + I(X_1; Y_1|U_2, U_1, U_0), \tag{4.15}$$
$$R_2 \le C_0 + I(U_2, X_2; Y_2|U_1, U_0), \tag{4.16}$$
$$R_2 \le I(U_2; Y_1|X_1, U_1, U_0) + I(X_2; Y_2|U_1U_2, U_0), \tag{4.17}$$
$$R_1 + R_2 \le C_0 + I(U_1, U_2, X_2; Y_2|U_0) + I(X_1; Y_1|U_1, U_2, U_0), \tag{4.18}$$
$$R_1 + R_2 \le C_0 + I(U_2, U_1, X_1; Y_1|U_0) + I(X_2; Y_2|U_2, U_1, U_0), \tag{4.19}$$
$$R_1 + R_2 \le 2C_0 + I(U_1, X_2; Y_2|U_2, U_0) + I(U_2, X_1; Y_1|U_1, U_0), \tag{4.20}$$
$$\begin{aligned} R_1 + R_2 \le\ & C_0 + I(X_2; Y_2|U_2, U_1, U_0) + I(U_1; Y_2|X_2, U_2, U_0) \\ & + I(U_2, X_1; Y_1|U_1, U_0), \end{aligned} \tag{4.21}$$

$$R_1 + R_2 \leq C_0 + I(X_1; Y_1|U_1, U_2, U_0) + I(U_2; Y_1|X_1, U_1, U_0)$$
$$+ I(U_1, X_2; Y_2|U_2, U_0), \tag{4.22}$$
$$R_1 + 2R_2 \leq 2C_0 + I(X_2; Y_2|U_2, U_1, U_0) + I(U_1, U_2, X_2; Y_2|U_0)$$
$$+ I(U_2, X_1; Y_1|U_1, U_0), \tag{4.23}$$
$$2R_1 + R_2 \leq 2C_0 + I(X_1; Y_1|U_1, U_2, U_0) + I(U_2, U_1 X_1; Y_1|U_0)$$
$$+ I(U_1, X_2; Y_2|U_2, U_0), \tag{4.24}$$
$$C_0 := \min\{I(U_0; Y_1), I(U_0; Y_2)\} \tag{4.25}$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}^*$.

Proof: The main task is to apply Fourier-Motzkin elimination to the following list of inequalities:

$$R_{21} \leq I(U_2; Y_1|X_1, U_1, U_0), \tag{4.26}$$
$$R_{11} \leq I(X_1; Y_1|U_1, U_2, U_0), \tag{4.27}$$
$$R_{12} + R_{11} \leq C_0 + I(U_1, X_1; Y_1|U_2, U_0), \tag{4.28}$$
$$R_{11} + R_{21} \leq C_0 + I(U_2, X_1; Y_1|U_1, U_0), \tag{4.29}$$
$$R_{12} + R_{11} + R_{21} \leq C_0 + I(U_2, U_1, X_1; Y_1|U_0); \tag{4.30}$$
$$R_{12} \leq I(U_1; Y_2|X_2, U_2, U_0), \tag{4.31}$$
$$R_{22} \leq I(X_2; Y_2|U_2, U_1, U_0), \tag{4.32}$$
$$R_{21} + R_{22} \leq C_0 + I(U_2, X_2; Y_2|U_1, U_0), \tag{4.33}$$
$$R_{22} + R_{12} \leq C_0 + I(U_1, X_2; Y_2|U_2, U_0), \tag{4.34}$$
$$R_{21} + R_{22} + R_{12} \leq C_0 + I(U_1, U_2, X_2; Y_2|U_0); \tag{4.35}$$
$$R_1 - R_{12} - R_{21} \leq 0, \tag{4.36}$$
$$R_2 - R_{21} - R_{22} \leq 0, \tag{4.37}$$
$$-R_{12} \leq 0, \tag{4.38}$$
$$-R_{11} \leq 0, \tag{4.39}$$

$$-R_{21} \leq 0, \tag{4.40}$$

$$-R_{22} \leq 0, \tag{4.41}$$

where $C_0 = \min\{I(U_0; Y_1), I(U_0; Y_2)\}$.

Specifically, the elimination takes four major steps to remove R_{12}, R_{21}, R_{11} and R_{22}, respectively and successively. The first step is to remove R_{12}. We exhaustively combine (sum) any inequality with term $+R_{12}$ with any one with $-R_{12}$, and keep the resulting new inequalities and the inequalities which does not contain R_{12}. The second step operates on the remaining inequalities from the first step, and removes R_{21} in the same way as step one does. The rest is done similarly. Finally, the resulting inequalities become (4.14)–(4.24), and the theorem follows. Interested readers can refer to Section 1 of Appendix A.3 for a detailed procedure of the Fourier-Motzkin elimination applied on the implicit achievable rate region for the ICC. ■

Remark 4.1 The achievable rate region $\mathcal{R}_f$ is convex due to the existence of U_0, therefore time-sharing is not necessary. Moreover, $\mathcal{R}_f$ includes the Han-Kobayashi or Chong-Motani-Garg region for the IC as a special case.

4.4 Conclusions

In this chapter, we have investigated the discrete memoryless ICF. We have developed a block Markov coding scheme for the channel, which allows each sender to partially decode certain information from the other such that cooperation can be induced. We have also obtained a corresponding achievable rate region in its the implicit form. Moreover, we have also obtained an explicit description of this rate region by applying the Fourier-Motzkin elimination on the implicit rate region.

Chapter 5

Relay Channels With Generalized Feedback

This chapter considers the three-node relay channel with *generalized* feedback. In particular, two generalized feedback configurations are investigated. In the first configuration, the source operates in full duplex mode, thereby being able to receive signals. The received signals at the source can be considered to be a form of generalized feedback. In the second configuration, the destination operates in full duplex mode, thereby being able to transmit signals. The transmitted signals from the destination can be considered to be a form of generalized feedback. For both configurations, coding schemes that are based on the ideas of *decode-and-forward* and *compress-and-forward*, are developed to exploit the feedback in their respective forms, and corresponding achievable rates are derived. It is shown that the derived achievable rates include some existing results for perfect feedback settings as special cases.

5.1 Introduction

Information theoretic study of the relay channel was initiated by van der Meulen [4], and was further expanded upon by many others [21, 24, 23, 72]. In particular, in the late 1970s Cover and El Gamal [21] developed two well-known coding strategies for the three-node relay channel: decode-and-forward (DF) and compress-and-forward (CF), which are named according to the specific operations at the relay. Both strategies involve block Markov coding. In the DF strategy, at the end of a transmission block, the relay decodes the source message, and in the subsequent transmission block, the relay forwards the decoded message to the destination; knowing what the relay intends to transmit to the destination, the source cooperates with the relay to help the destination to resolve the uncertainty about the message delivered in the present block. Such *cooperation* is not possible in the CF strategy, as neither the source nor the relay knows what the other node will transmit in the next block. In the CF strategy, at the end of a transmission block, the relay compresses its channel output sequence instead of trying to decode any information from it. The relay then forwards the compressed sequence to the destination in the subsequent transmission block. By providing the compressed channel output sequence as side information for the destination to decode the source message, the relay facilitates the transmission of the message. This approach is called *facilitation* in contrast to *cooperation*. A hybrid coding strategy that combines the DF and CF strategies was proposed in [21, Theorem 7].

After many years of relative quiescence, there has been considerable recent interest in the relay channel, primarily due to the emergence of cooperative communication networks [24, 23]. Another topic that has received considerable recent attention is communication with feedback [75, 10, 72], as it has been demonstrated that feedback can be used to improve information throughput in various communication scenarios such as in multi-user networks. As an intersection of these two topics, the relay channel with feedback was first investigated in [21], where complete and causal knowledge about the channel output at the destination is assumed to be available at both the relay and the source. Capacity is shown to be achieved in this situation with the DF strategy. Such a feedback setting describes one of the possible feedback configurations for the relay channel. Two other feedback configurations for the relay channel were investigated in [72], including cases of perfect causal feedback from the relay to the source and perfect causal feedback from the destination to the source. Several coding strategies based on DF and CF were developed in [72], which shows improvements in the achievable rates for the relay channel with partial feedback over the relay channel without feedback.

In this chapter, we consider the three-node relay channel with generalized feedback, in which the feedback is obtained in the same way as the intended information is delivered, rather than assuming 'perfect' feedback to be causally available at the source or the relay. In particular, we investigate two generalized feedback configurations. In the first configuration, the source is able to receive signals, while in the second configuration the destination is able to transmit signals. Although they have not been explicitly investigated in literature, these two settings can be considered to be component settings or partial configurations of the fully cooperative broadcast relay channel [76] or the two-way relay channel [77, 78]. By extending the respective results in [76, 77, 78] to the two feedback configurations investigated in this chapter, we can obtain corresponding achievability results. However, the results obtained in this way are the same as the achievable rates for the generic relay channel [21], because the feedback is not exploited in the respective proposed coding strategies of [76, 77, 78]. Aiming to improve the achievable rates by exploiting the feedback at/from the respective nodes, new coding schemes are developed based on the existing DF and CF strategies, which are the main contributions of this chapter.

The rest of this chapter is organized as follows. In Section 5.2, we introduce our channel models and related definitions. We also present three lemmas for strong typicality, which are frequently applied in our proofs. In Sections 5.3, we present our main results regarding the first feedback configuration, including three achievable rates. We show that our results generalizes some existing results. In Section 5.4, we present two achievable rates for the second feedback configuration. Last, we conclude this chapter in Section 5.5.

5.2 Channel Models and Preliminaries

In this section, we introduce two models of the three-node relay channel with generalized feedback, in both of which the relay operates in full duplex mode. We define the existence of codes and achievable rates for the channels. As preliminaries, we discuss the notion of strong typicality and list several important lemmas which will be used frequently in our proofs.

5.2.1 Relay Channel With Generalized Feedback at the Source

With reference to Fig. 5.1, a discrete memoryless relay channel with generalized feedback available at the source consists of a source (node 0), a relay (node 1), and a destination (node 2). The channel is defined by a tuple $(\mathcal{X}_0 \times \mathcal{X}_1, p(y_0, y_1, y_2|x_0, x_1), \mathcal{Y}_0 \times \mathcal{Y}_1 \times \mathcal{Y}_2)$ with $\mathcal{X}_t$, $t = 0, 1$, denoting the channel input alphabet of node t, $\mathcal{Y}_t$, $t = 0, 1, 2$, denoting the channel output alphabet of node t, and $p(y_0, y_1, y_2|x_0, x_1)$ denoting the collections of probabilities of the channel outputs $(y_0, y_1, y_2) \in \mathcal{Y}_0 \times \mathcal{Y}_1 \times \mathcal{Y}_2$ being received conditioned on channel inputs $(x_0, x_1) \in \mathcal{X}_0 \times \mathcal{X}_1$ being transmitted. The channel is assumed to be memoryless in the sense that the channel outputs, $y_{0,k}$, $y_{1,k}$ and $y_{2,k}$, in one channel use depend only on the channel inputs, $x_{0,k}$, and $x_{1,k}$, i.e.,

$$p(y_{0,k}, y_{1,k}, y_{2,k}|x_0^k, x_1^k, y_0^{k-1}, y_1^{k-1}, y_2^{k-1}) = p(y_{0,k}, y_{1,k}, y_{2,k}|x_{0,k}, x_{1,k}),\ k = 1, ..., n,$$

where $x_t^k := (x_{t,1}, x_{t,2}, ..., x_{t,k})$, $t = 1, 2$, and $y_t^k := (y_{t,1}, y_{t,2}, ..., y_{t,k})$, $t = 0, 1, 2$. Through such a channel, the source wishes to send a message $w \in \mathcal{W} := \{1, ..., M\}$ to the destination. The message is further assumed to be generated uniformly over its range. Let us denote this channel as $\mathcal{C}_{\mathrm{SFB}}$.

Definition 5.1 An $(M, n, P_e^{(n)})$ *code exists for* $\mathcal{C}_{\mathrm{SFB}}$ *if there exist*

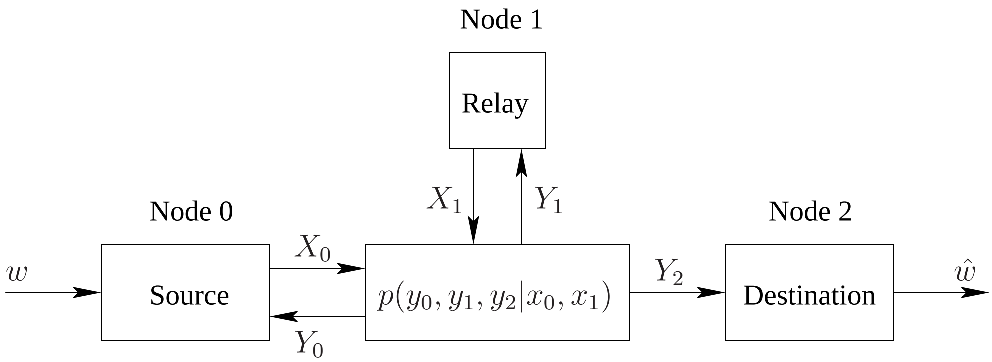

Figure 5.1: Three-node relay channel with generalized feedback available at the source.

1. *a set of encoding functions at the source*

$$f_{0,k} : \mathcal{M} \times \mathcal{Y}_0^{k-1} \mapsto \mathcal{X}_0, \quad k = 1, ..., n,$$

2. *a set of relaying functions at the relay*

$$f_{1,k} : \mathcal{Y}_1^{k-1} \mapsto \mathcal{X}_1, \quad k = 1, ..., n,$$

3. *and a decoding function at the destination*

$$g_2 : \mathcal{Y}_2^n \mapsto \mathcal{W},$$

such that the probability of decoding error is computed as

$$P_e^{(n)} = \frac{1}{M} \sum_{w \in \mathcal{W}} \Pr(g_2(Y_2^n) \neq w | W = w).$$

Definition 5.2 A non-negative rate R is achievable for $\mathcal{C}_{\mathrm{SFB}}$ if there exists a sequence of codes $(2^{nR}, n, P_e^{(n)})$ such that $P_e^{(n)}$ approaches 0, as $n \to \infty$.

5.2.2 Relay Channel With Generalized Feedback from the Destination

Fig. 5.2 depicts the second feedback configuration that will be studied in the chapter: the discrete memoryless relay channel with generalized feedback from the destination. The channel is defined by a tuple $(\mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_0, x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ with definitions analogous to those in $\mathcal{C}_{\mathrm{SFB}}$. The channel is also assumed to be memoryless. Via this channel, the source wishes to send a message $\dot{w} \in \dot{\mathcal{W}} := \{1, ..., \dot{M}\}$ to the destination. The message is assumed to be generated uniformly over its range. We denote a channel with this feedback configuration by $\mathcal{C}_{\mathrm{DFB}}$.

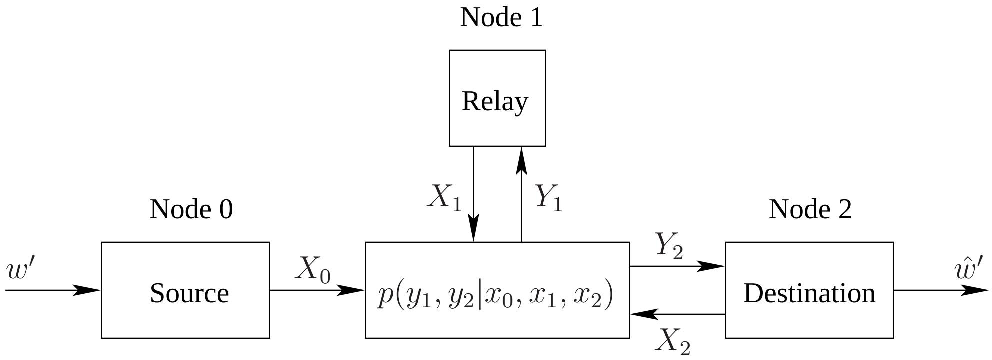

Figure 5.2: Three-node relay channel with generalized feedback from the destination.

Definition 5.3 An $(\dot{M}, n, P_e^{(n)})$ code exists for $\mathcal{C}_{\mathrm{DFB}}$ if there exist

1. *an encoding function at the source*

$$\dot{f}_0 : \dot{\mathcal{W}} \mapsto \mathcal{X}_0^n, \tag{5.1}$$

2. *a set of relaying functions at the relay*

$$\dot{f}_{1,k} : \mathcal{Y}_1^{k-1} \mapsto \mathcal{X}_1, \quad k = 1, ..., n,$$

3. *a set of feedback functions at the destination*

$$\dot{f}_{2,k} : \mathcal{Y}_2^{k-1} \mapsto \mathcal{X}_2, \quad k = 1, ..., n.$$

4. and a decoding function at the destination

$$\dot{g}_2 : \mathcal{Y}_2^n \mapsto \dot{\mathcal{W}},$$

such that the probability of decoding error is computed as

$$P_e^{(n)} = \frac{1}{\dot{M}} \sum_{\dot{w} \in \dot{\mathcal{W}}} \Pr(\dot{g}_2(Y_2^n) \neq \dot{w} | \dot{W} = \dot{w}).$$

Definition 5.4 A non-negative rate $\dot{R}$ is achievable for $\mathcal{C}_{\text{DFB}}$ if there exists a sequence of codes $(2^{n\dot{R}}, n, P_e^{(n)})$ such that $P_e^{(n)}$ approaches 0, as $n \to \infty$.

5.2.3 Strong Typicality

In this chapter, we will frequently use the notion of strong typicality [40, Section 13.6]. As preliminaries, some basic results concerning strong typicality are given as follows.

Let $\{Z_1, Z_2, ..., Z_m\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(z_1, z_2, ..., z_m)$ for $(z_1, z_2, ..., z_m) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \times ... \times \mathcal{Z}_m$. Let S denote an arbitrary ordered subset of these random variables, and consider a set $\mathbf{S}$ of n independent copies of S, i.e.,

$$\Pr(\mathbf{S} = \mathbf{s}) = \prod_{t=1}^{n} \Pr(S_t = s_t), \ \mathbf{s} \in \mathcal{S}^n,$$

where $\mathcal{S}$ is the alphabet of random variable S.

Define $N(s; \mathbf{s})$ as the number of indices $t \in \{1, 2, ..., n\}$ such that $S_t = s$. By the law of large numbers, for all $s \in \mathcal{S}$, we have

$$\frac{1}{n} N(s; \mathbf{s}) \to p(s), \tag{5.2}$$

and

$$-\frac{1}{n}\log p(s_1, s_2, ..., s_n) = -\frac{1}{n}\sum_{t=1}^{n}\log p(s_t) \rightarrow H(S). \tag{5.3}$$

Convergence in (5.2) and (5.3) occurs simultaneously with probability one for all subsets of the random variables $S \subseteq \{Z_1, Z_2, ..., Z_m\}$.

Definition 5.5 The set $\mathcal{A}_\epsilon^{(n)}$ of ϵ-typical n-sequences $(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m)$ is defined by

$$\begin{aligned}
&\mathcal{A}_\epsilon^{(n)}(Z_1, Z_2, ..., Z_m) := \mathcal{A}_\epsilon^{(n)} := \\
&\quad\Bigg\{(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m) : \\
&\quad\left|\frac{1}{n}N(z_1, z_2, ..., z_m; \mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_m) - p(z_1, z_2, ..., z_m)\right| < \frac{\epsilon}{\|\mathcal{Z}_1 \times \mathcal{Z}_2 \times ... \times \mathcal{Z}_m\|}, \\
&\quad\forall\ (z_1, z_2, ..., z_m) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \times ... \times \mathcal{Z}_m\Bigg\},
\end{aligned}$$

where $\|\mathcal{Z}\|$ denotes the cardinality of the set $\mathcal{Z}$.

For each non-empty subset $S \subseteq \{Z_1, Z_2, ..., Z_m\}$, the set $\mathcal{A}_\epsilon^{(n)}(S)$ of ϵ-typical n-sequences $\mathbf{s}$ can be similarly defined with respect to their individual joint distributions. The following lemmas present some important properties of $\mathcal{A}_\epsilon^{(n)}(S)$, which will be frequently used in our proofs.

Lemma 5.1 For any $\epsilon > 0$, there exists an integer n such that

1. $\Pr(\mathcal{A}_\epsilon^{(n)}(S)) \geq 1 - \epsilon$,

2. $2^{-n(H(S)+\epsilon)} \leq \Pr(\mathbf{S} = \mathbf{s}) \leq 2^{-n(H(S)-\epsilon)}$,

3. $(1 - \epsilon)2^{n(H(S)-\epsilon)} \leq \|\mathcal{A}_\epsilon^{(n)}(S)\| \leq 2^{n(H(S)+\epsilon)}$, *and*

4. *if $S_1, S_2 \subseteq \{Z_1, Z_2, ..., Z_m\}$, and $(\mathbf{s}_1, \mathbf{s}_2) \in \mathcal{A}_\epsilon^{(n)}(S_1 \cup S_2)$, then we have*

$$2^{-n(H(S_1|S_2)+2\epsilon)} \leq \Pr(\mathbf{S}_1 = \mathbf{s}_1|\mathbf{S}_2 = \mathbf{s}_2) \leq 2^{-n(H(S_1|S_2)-2\epsilon)}.$$

Lemma 5.2 Let S_1, S_2, and S_3 be three non-empty subsets of the set of random variables $\{Z_1, Z_2, ..., Z_m\}$. Let $S_1^{'}$ and $S_2^{'}$ be conditionally independent given S_3 with the marginal distributions

$$\Pr(S_1^{'} = s_1 | S_3 = s_3) = \sum_{\mathbf{s}_2} p(s_1, s_2, s_3)/p(s_3), \text{ and}$$
$$\Pr(S_2^{'} = s_2 | S_3 = s_3) = \sum_{\mathbf{s}_3} p(s_1, s_2, s_3)/p(s_3).$$

Let $(\mathbf{S}_1^{'}, \mathbf{S}_2^{'}, \mathbf{S}_3^{'})$ be generated according to $\prod_{t=0}^{n} p(s_3)p(s_{1,t}|s_{3,t})p(s_{2,t}|s_{3,t})$, with $\mathbf{s}_i^{'} := (s_{i,1}, s_{i,2}, ..., s_{i,n})$, $i = 1, 2, 3$. For sufficiently large n, we have

$$(1-\epsilon)2^{-n(I(S_1;S_2|S_3)+7\epsilon)} \leq \Pr\big((\mathbf{S}_1^{'}, \mathbf{S}_2^{'}, \mathbf{S}_3^{'}) \in \mathcal{A}_\epsilon^{(n)}(S_1, S_2, S_3)\big) \leq 2^{-n(I(S_1;S_2|S_3)-7\epsilon)}.$$

Lemma 5.3 (Markov Lemma, [65, Lemma 4.1]) Let S_1, S_2, and S_3 form a Markov chain, i.e., $S_1 \to S_2 \to S_3$. Let $(\mathbf{S}_1, \mathbf{S}_2)$ be generated according to $\prod_{t=1}^{n} p(s_{1,t}, s_{2,t})$ with $p(s_1, s_2) = \sum_{s_3} p(s_1, s_2, s_3)$. Given $(\mathbf{S}_2, \mathbf{s}_3) \in \mathcal{A}_\epsilon^{(n)}(S_2, S_3)$, we have

$$\Pr\big((\mathbf{S}_1, \mathbf{S}_2, \mathbf{s}_3) \in \mathcal{A}_\epsilon^{(n)}(S_1, S_2, S_3)\big) > 1 - \epsilon,$$

for sufficiently large n.

5.3 Achievable Rates for $\mathcal{C}_{\text{SFB}}$

In this section, we present our achievablility results for the first feedback configuration depicted in Fig. 5.1, in which the source operates in full duplex mode. It can be observed that the feedback received at the source consists of signals from itself and from the relay. Obviously, only the information contained in the signal from the relay is useful to the source, as the source knows its own transmitted message. Therefore, what the relay transmits determines whether the feedback can be exploited and how the feedback can be used at the source.

As mentioned earlier, the relay in a generic relay channel can perform either DF or CF to aid the transmission between the source and destination via either cooperation or facilitation. When the DF strategy is applied, the relay transmits information sent by the source, and thus its transmission is known to the source. Therefore, if we merely apply DF at the relay, the source can only obtain from the feedback information which it already knows. As a result, the feedback is wasted in the sense that the source is able to obtain the information contained in the feedback in the absence of the feedback.

On the contrary, if we apply the CF strategy, what the relay transmits contains certain information that the source is unable to obtain without exploiting the feedback. This makes the feedback potentially useful to the source. We know that in the generic relay channel, we can only achieve facilitation rather than cooperation under the CF strategy [21], due to the fact that the source does not know which codeword (carrying the compressed version of the channel output) the relay wishes to send. Nevertheless, in the current feedback figuration, the source may decode which codeword the relay wants to transmit, and thus cooperation between the two nodes in transmitting this compressed information becomes possible.

5.3.1 Rates Achieved by Decode-and-Forward / Partially-Decode-and-Forward

Based on the above discussion, we next present our first coding scheme in which the source acts as a DF relay in the way that it decodes which codeword, as the compressed version of the received sequence, the relay wishes to send, and then cooperates with the relay to forward this information to the destination. The destination then performs joint decoding over its own received sequence and the compressed version of the received sequence at the relay to obtain the message sent by the source. The following achievable rate can be established with such a coding scheme.

Let U and $\hat{Y}_1$ be auxiliary random variables defined over arbitrary finite alphabets $\mathcal{U}$ and $\hat{\mathcal{Y}}_1$, respectively. Let $\mathcal{P}_0$ denote the set of joint distributions $p(\cdot)$ that factor as follows

$$p(u, x_0, x_1, y_0, y_1, y_2, \hat{y}_1) = p(u)p(x_0|u)p(x_1|u)p(y_0, y_1, y_2|x_0, x_1)p(\hat{y}_1|y_1, x_1, u). \tag{5.4}$$

Theorem 5.1 For the discrete memoryless relay channel with generalized feedback at the source, $\mathcal{C}_{\mathrm{SFB}}$*, the following rate is achievable:*

$$R_{\mathrm{SFB0}} := \sup_{p(\cdot)\in\mathcal{P}_0} I(X_0; Y_2, \hat{Y}_1|X_1, U),$$

subject to the following two constraints

$$I(\hat{Y}_1; Y_1|X_1, U) \leq I(X_1; Y_0|X_0, U) + I(\hat{Y}_1; Y_0, X_0|X_1, U), \text{ and} \tag{5.5}$$

$$I(\hat{Y}_1; Y_1|X_1, U) \leq I(\hat{Y}_1, X_1, U; Y_2). \tag{5.6}$$

Proof: We consider a block Markov superposition coding scheme consisting of regular encoding and sliding window decoding [79]. Block Markov superposition refers to the relation between two consecutive coding blocks with the later one depending on the other, i.e., the later block is generated by superimposing the new information onto the previous block, and all the encoded blocks form a Markov chain. Sliding window decoding refers to the decoding method, where we apply a window which takes three consecutive blocks of channel outputs, and perform the decoding over these three blocks. After the decoding over the current window is finished, we slide the window forward by one block, and so on and so forth.

Under our proposed coding scheme, the successive transmissions consist of $B+2$ blocks, each of which is of length n. In each of the first B blocks, a message $w \in [1, 2^{nR_{\mathrm{SFB0}}}]$ will be sent to the destination with probability of error approaching 0. The average rate of transmission is thus $R_{\mathrm{SFB0}}B/(B+2)$, which is arbitrarily close to R_{SFB0} as $B \to \infty$.

A random coding argument is applied to show the achievability of R_{SFB0}. First, fix a joint distribution $p(\cdot) \in \mathcal{P}_0$.

Random codebook generation: Generate three statistically independent codebooks, namely codebook 1, codebook 2, and codebook 3, by repeating the following procedures three times.

1. Generate 2^{nR_0} independent and identically distributed (i.i.d.) codewords $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, according to the joint distribution $\prod_{t=1}^{n} p(u_t)$.

2. For each $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, generate $2^{nR_{\mathrm{SFB0}}}$ i.i.d. codewords $\mathbf{X}_0(i,j)$, $j \in [1, 2^{nR_{\mathrm{SFB0}}}]$, according to $\prod_{t=1}^{n} p(x_{0,t}|u_t(i))$.

3. For each $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, generate 2^{nR_0} i.i.d. codewords $\mathbf{X}_1(i,k)$, $k \in [1, 2^{nR_0}]$, according to $\prod_{t=1}^{n} p(x_{1,t}|u_t(i))$.

4. For each codeword pair $(\mathbf{U}(i), \mathbf{X_1}(i,k))$, $i, k \in [1, 2^{nR_0}]$, generate 2^{nR_0} i.i.d. codewords $\hat{\mathbf{Y}}_1(i,k,l)$, $l \in [1, 2^{nR_0}]$, according to $\prod_{t=1}^{n} p(\hat{y}_{1,t}|u_t(i), x_{1,t}(i,k))$.

Encoding and transmission: We use the three codebooks in a periodic manner such that any adjacent three blocks are encoded using the three different codebooks respectively, i.e., we use codebook 1 to encode block 1, codebook 2 to encode block 2, codebook 3 to encode block 3, and codebook 1 to encode block 4, and so on. By doing so, the mutual independence of the error events among any three consecutive blocks can be ensured [24].

Assume that the transmission of block $b-1$ has just ended, and the message $w^{(b)}$ is to be transmitted in the current block, block b. Also assume that the following information (the set of message indices) is now available at the respective nodes:

1. At the source: $l^{(1)}$, $l^{(2)}$, ..., $l^{(b-3)}$; $w^{(1)}$, $w^{(2)}$, ..., $w^{(b)}$.

2. At the relay: $l^{(1)}$, $l^{(2)}$, ..., $l^{(b-2)}$.

Table 5.1: Codewords transmitted in each block to achieve R_{SFB0}.

	Block 1	Block 2	Block 3	...	Block b	...
$\mathbf{U}(i)$	$\mathbf{u}(1)$	$\mathbf{u}(2)$	$\mathbf{u}(l^{(1)})$	...	$\mathbf{u}(l^{(b-2)})$	...
$\mathbf{X}_0(i,j)$	$\mathbf{x}_0(1,w^{(1)})$	$\mathbf{x}_0(2,w^{(2)})$	$\mathbf{x}_0(l^{(1)},w^{(3)})$	...	$\mathbf{x}_0(l^{(b-2)},w^{(b)})$	...
$\mathbf{X}_1(i,k)$	$\mathbf{x}_1(1,2)$	$\mathbf{x}_1(2,l^{(1)})$	$\mathbf{x}_1(l^{(1)},l^{(2)})$	...	$\mathbf{x}_1(l^{(b-2)},l^{(b-1)})$	...
$\hat{\mathbf{Y}}_1(i,k,l)$	$\emptyset$	$\hat{\mathbf{y}}_1(1,2,l^{(1)})$	$\hat{\mathbf{y}}_1(2,l^{(1)},l^{(2)})$	...	$\hat{\mathbf{y}}_1(l^{(b-3)},l^{(b-2)},l^{(b-1)})$	...

[Source.] If the source decodes $l^{(b-2)}$, the index of the compressed version of the channel output sequence $\mathbf{y}_1^{(b-2)}$, the source transmits the codeword $\mathbf{x}_0(l^{(b-2)},w^{(b)})$ with n channel uses in block b; otherwise it transmits $\mathbf{x}_0(1,w^{(b)})$.

[Relay.] The relay first needs to compress the newly received channel output sequence $\mathbf{y}_1^{(b-1)}$ by applying Wyner-Ziv coding [21]. If the relay successfully finds a codeword of index $l^{(b-1)}$ as the compressed version of $\mathbf{y}_1^{(b-1)}$, then it transmits $\mathbf{x}_1(l^{(b-2)},l^{(b-1)})$ with n channel uses in block b; otherwise, it transmits $\mathbf{x}_1(l^{(b-2)},1)$.

Table 5.1 lists the codewords transmitted in each block.

Decoding: [Source.] The source first needs to decode $l^{(b-2)}$ (or $\hat{\mathbf{y}}_1(l^{(b-4)},l^{(b-3)},l^{(b-2)})$), the compressed version of the channel output sequence $\mathbf{y}_1^{(b-2)}$, from its channel output sequence $\mathbf{y}_0^{(b-1)}$ received in block $b-1$, by looking for an index $\hat{l}^{(b-2)}$ such that

$$(\mathbf{u}(l^{(b-3)}),\mathbf{x}_0(l^{(b-3)},w^{(b-1)}),\mathbf{x}_1(l^{(b-3)},\hat{l}^{(b-2)}),\mathbf{y}_0^{(b-1)})\in\mathcal{A}_\epsilon^{(n)},\quad\text{and}$$
$$(\hat{\mathbf{y}}_1(l^{(b-4)},l^{(b-3)},\hat{l}^{(b-2)}),\mathbf{y}_0^{(b-2)},\mathbf{u}(l^{(b-4)}),\mathbf{x}_0(l^{(b-4)},w^{(b-2)}),\mathbf{x}_1(l^{(b-4)},l^{(b-3)}))\in\mathcal{A}_\epsilon^{(n)}.$$

If the index found is unique, the source declares $l^{(b-2)}=\hat{l}^{(b-2)}$; otherwise, an error is declared. The probability of error can be shown to approach 0 for sufficiently large n, when

$$R_0\le I(X_1;Y_0|X_0,U)+I(\hat{Y}_1;Y_0,X_0|X_1,U).\tag{5.7}$$

[Relay.] The relay first needs to compress the newly received channel output sequence $\mathbf{y}_1^{(b-1)}$ by applying Wyner-Ziv coding. To do so, the relay looks for an index (not necessarily unique) $\hat{l}^{(b-1)}$ such that

$$(\hat{\mathbf{y}}_1(l^{(b-3)}, l^{(b-2)}, \hat{l}^{(b-1)}), \mathbf{y}_1^{(b-1)}, \mathbf{u}(l^{(b-3)}), \mathbf{x}_1(l^{(b-3)}, l^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)}.$$

If any such index is found, the relay declares $l^{(b-1)} = \hat{l}^{(b-1)}$; otherwise, an error is declared. When n is sufficiently large, the probability of error approaches 0 if the following constraint is satisfied:

$$R_0 \geq I(\hat{Y}_1; Y_1 | X_1, U), \tag{5.8}$$

which follows from [65, Lemma 2.1.3] directly.

[Destination.] We now describe the decoding procedure at the end of block b. Assume that at this instant, the destination has already successfully decoded the following information: 1) $w^{(1)}$, $w^{(2)}$, ..., $w^{(b-3)}$; and 2) $l^{(1)}$, $l^{(2)}$, ..., $l^{(b-3)}$.

Before decoding the message $w^{(b-2)}$, the destination first needs to decode $l^{(b-2)}$ (equivalently, $\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)})$, the compressed version of $\mathbf{y}_1^{(b-2)}$), from its channel output sequences received in the last three blocks, $\mathbf{y}_2^{(b-2)}$, $\mathbf{y}_2^{(b-1)}$, and $\mathbf{y}_2^{(b)}$. The destination declares $l^{(b-2)} = \hat{l}^{(b-2)}$ if there exists a unique index $\hat{l}^{(b-2)}$ such that the following three events happen simultaneously

$$\begin{aligned}
&(\mathbf{u}(\hat{l}^{(b-2)}), \mathbf{y}_2^{(b)}) \in \mathcal{A}_\epsilon^{(n)}, \\
&(\mathbf{u}(l^{(b-3)}), \mathbf{x}_1(l^{(b-3)}, \hat{l}^{(b-2)}), \mathbf{y}_2^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}, \quad \text{and} \\
&(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, \hat{l}^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)};
\end{aligned}$$

otherwise, an error is declared. The probability of error in this step can be shown to approach 0 for sufficiently large n, when the following holds:

$$R_0 \leq I(U; Y_2) + I(X_1; Y_2 | U) + I(\hat{Y}_1; Y_2 | X_1, U). \tag{5.9}$$

The destination lastly decodes the message $w^{(b-2)}$ from $\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)})$

and $\mathbf{y}_2^{(b-2)}$. It is declared that $w^{(b-2)} = \hat{w}^{(b-2)}$ if $\hat{w}^{(b-2)}$ is the unique message index such that

$$(\mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, \hat{w}^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)}), \hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}), \mathbf{y}_2^{(b-2)}) \in \mathcal{A}_\epsilon^{(n)}.$$

The decoding error probability of this step can be readily shown to approach 0 for sufficiently large n, when

$$R_{\text{SFB0}} \leq I(X_0; Y_2, \hat{Y}_1 | X_1, U). \tag{5.10}$$

Analysis of probabilities of error: We first list all the events that possibly happen in the decoding process at the respective nodes as follows:

At the source:

1. E_1: $(\mathbf{u}(l^{(b-3)}), \mathbf{x}_0(l^{(b-3)}, w^{(b-1)}), \mathbf{x}_1(l^{(b-3)}, l^{(b-2)}), \mathbf{y}_0^{(b-1)}) \notin \mathcal{A}_\epsilon^{(n)}$.
2. E_2: $(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}), \mathbf{y}_0^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, w^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)}))$ $\notin \mathcal{A}_\epsilon^{(n)}$.
3. $E_3(\hat{\hat{l}}^{(b-2)})$: $(\mathbf{u}(l^{(b-3)}), \mathbf{x}_0(l^{(b-3)}, w^{(b-1)}), \mathbf{x}_1(l^{(b-3)}, \hat{\hat{l}}^{(b-2)}), \mathbf{y}_0^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}$.
4. $E_4(\hat{\hat{l}}^{(b-2)})$: $(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, \hat{\hat{l}}^{(b-2)}), \mathbf{y}_0^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, w^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}$.

At the relay:

1. $E_5(\hat{\hat{l}}^{(b-1)})$: $(\hat{\mathbf{y}}_1(l^{(b-3)}, l^{(b-2)}, \hat{\hat{l}}^{(b-1)}), \mathbf{y}_1^{(b-1)}, \mathbf{u}(l^{(b-3)}), \mathbf{x}_1(l^{(b-3)}, l^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)}$.

At the destination:

1. E_6: $(\mathbf{u}(l^{(b-2)}), \mathbf{y}_2^{(b)}) \notin \mathcal{A}_\epsilon^{(n)}$.
2. E_7: $(\mathbf{u}(l^{(b-3)}), \mathbf{x}_1(l^{(b-3)}, l^{(b-2)}), \mathbf{y}_2^{(b-1)}) \notin \mathcal{A}_\epsilon^{(n)}$.
3. E_8: $(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \notin \mathcal{A}_\epsilon^{(n)}$.
4. $E_9(\hat{\hat{l}}^{(b-2)})$: $(\mathbf{u}(\hat{\hat{l}}^{(b-2)}), \mathbf{y}_2^{(b)}) \in \mathcal{A}_\epsilon^{(n)}$.

5. $E_{10}(\hat{\hat{l}}^{(b-2)})$: $(\mathbf{u}(l^{(b-3)}), \mathbf{x}_1(l^{(b-3)}, \hat{\hat{l}}^{(b-2)}), \mathbf{y}_2^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}$.

6. $E_{11}(\hat{\hat{l}}^{(b-2)})$: $(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, \hat{\hat{l}}^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}$.

7. E_{12}: $(\mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, w^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)}), \hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}), \mathbf{y}_2^{(b-2)})$ $\notin \mathcal{A}_\epsilon^{(n)}$.

8. $E_{13}(\hat{\hat{w}}^{(b-2)})$: $(\mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, \hat{\hat{w}}^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)}), \hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}),$ $\mathbf{y}_2^{(b-2)}) \in \mathcal{A}_\epsilon^{(n)}$.

Note that $\hat{\hat{(\cdot)}}$ denotes an estimate of $(\cdot)$, e.g. $\hat{\hat{l}}^{(b-2)}$ is an estimate of $l^{(b-2)}$.

[Source.] When decoding $l^{(b-2)}$, the source may make a decoding error, which must fall into one or more of the following three error events: 1) E_1 happens; 2) E_2 happens; 3) there exists $\hat{\hat{l}}^{(b-2)} \neq l^{(b-2)}$ such that $E_3(\hat{\hat{l}}^{(b-2)})$ and $E_4(\hat{\hat{l}}^{(b-2)})$ happen simultaneously. Hence, we have the probability of error at the source expressed and bounded as

$$\begin{aligned} P_e(\text{source}) &= \Pr\left\{E_1 \bigcup E_2 \bigcup \cup_{\hat{\hat{l}}^{(b-2)} \neq (l^{(b-2)})} \left(E_3(\hat{\hat{l}}^{(b-2)}) \cap E_4(\hat{\hat{l}}^{(b-2)})\right)\right\} \\ &\overset{(a)}{\leq} \Pr(E_1) + \Pr(E_2) + 2^{nR_0} \Pr(E_3(\hat{\hat{l}}^{(b-2)})) \Pr(E_4(\hat{\hat{l}}^{(b-2)})), \end{aligned} \tag{5.11}$$

where (a) follows from the union bound and the statistical independence between the two events $E_3(\hat{\hat{l}}^{(b-2)})$ and $E_4(\hat{\hat{l}}^{(b-2)})$.

It follows from Lemma 5.1 that

$$\Pr(E_1) \leq \epsilon, \tag{5.12}$$

for sufficiently large n. For E_2, we have

$$(\mathbf{y}_0^{(b-2)}, \mathbf{u}(l^{(b-4)}), \mathbf{x}_0(l^{(b-4)}, w^{(b-2)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}$$

by Lemma 5.1, and

$$(\hat{\mathbf{y}}_1(l^{(b-4)}, l^{(b-3)}, l^{(b-2)}), \mathbf{u}(l^{(b-4)}), \mathbf{x}_1(l^{(b-4)}, l^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}$$

according to the encoding scheme at the relay. Applying Lemma 5.3 (Markov Lemma) to the case of $S_1 = (Y_0, X_0)$, $S_2 = (U, X_1)$, and $S_3 = \hat{Y}_1$, we have

$$\Pr(E_2) \leq \epsilon, \tag{5.13}$$

when n is sufficiently large.

Next, applying Lemma 5.2 to the case of $S_1^{'} = X_1$, $S_2^{'} = (Y_0, X_0)$, and $S_3^{'} = U$, we obtain

$$\Pr(E_3(\hat{\hat{l}}^{(b-2)})) \leq 2^{-n(I(X_1;Y_0,X_0|U)-7\epsilon)} \overset{(b)}{=} 2^{-n(I(X_1;Y_0|X_0,U)-7\epsilon)}, \tag{5.14}$$

where (a) follows from $I(X_1; X_0|U) = 0$ which is due to the Markov chain relationship $X_1 \rightarrow U \rightarrow X_0$. Similarly, applying Lemma 5.2 to the case of $S_1^{'} = \hat{Y}_1$, $S_2^{'} = (Y_0, X_0)$, and $S_3^{'} = (U, X_1)$, we have

$$\Pr(E_4(\hat{\hat{l}}^{(b-2)})) \leq 2^{-n(I(\hat{Y}_1;Y_0,X_0|X_1,U)-7\epsilon)}. \tag{5.15}$$

Applying the upper bounds (5.12)–(5.15) to the corresponding probability terms in (5.11), we have

$$P_e(\text{source}) \leq \epsilon + \epsilon + 2^{nR_0} 2^{-n(I(X_1;Y_0|X_0,U)-7\epsilon)} 2^{-n(I(\hat{Y}_1;Y_0,X_0|X_1,U)-7\epsilon)} \leq 3\epsilon,$$

when R_0 satisfies the constraint (5.7) and n is sufficiently large.

[Relay.] Now consider the relay. An error will be declared at the relay when for all $\hat{\hat{l}}^{(b-1)} \in [1, 2^{nR_0}]$, event $E_5^c(\hat{\hat{l}}^{(b-1)})$ happens, where $(\cdot)^c$ denotes the complement of a set. We express and upper-bound the probability of error at the relay in the following:

$$\begin{aligned}
P_e(\text{relay}) &= \Pr\{\cap_{\hat{\hat{l}}^{(b-1)}} E_5^c(\hat{\hat{l}}^{(b-1)})\} \\
&= (1 - \Pr\{E_5(\hat{\hat{l}}^{(b-1)})\})^{2^{nR_0}} \\
&\overset{(c)}{\leq} (1 - 2^{-n(I(\hat{Y}_1;Y_1|U,X_1)+7\epsilon)})^{2^{nR_0}} \\
&= e^{2^{nR_0}\ln(1-2^{-n(I(\hat{Y}_1;Y_1|U,X_1)+7\epsilon)})} \\
&\overset{(d)}{\leq} e^{2^{nR_0}2^{-n(I(\hat{Y}_1;Y_1|U,X_1)+7\epsilon)}},
\end{aligned}$$

where (c) follows from $\Pr\{E_5(\hat{\hat{l}}^{(b-1)})\} \geq 2^{-n(I(\hat{Y}_1;Y_1|U,X_1)+7\epsilon)}$; (d) follows from the Mercator series of $\ln(1+x)$ with $x = -2^{-n(I(\hat{Y}_1;Y_1|U,X_1)+7\epsilon)}$ being a negative real number that approaches 0. Note that the same argument was used in [65, Lemma 2.1.3]. Hence, we conclude that we have

$$P_e(\text{relay}) \leq \epsilon,$$

for sufficiently large n, as long as R_0 satisfies the constraint (5.8).

[Destination.] The destination may make an error when it tries to decode $l^{(b-2)}$, and also when it tries to decode $w^{(b-2)}$. By following the lines of the analysis of P_e(Source), the probability of error can be expressed and upper-bounded as follows

$$\begin{aligned}
&P_e(\text{destination}) \\
&\leq P_e(\text{decoding } l^{(b-2)}) + P_e(\text{decoding } w^{(b-2)}) \\
&= \Pr\left\{E_6 \bigcup E_7 \bigcup E_8 \bigcup \cup_{\hat{\hat{l}}^{(b-2)} \neq l^{(b-2)}} \left(E_9(\hat{\hat{l}}^{(b-2)}) \cap E_{10}(\hat{\hat{l}}^{(b-2)} \cap E_{11}(\hat{\hat{l}}^{(b-2)})\right)\right\} \\
&\quad + \Pr\left\{E_{12} \bigcup \cup_{\hat{\hat{w}}^{(b-2)} \neq w^{(b-2)}} E_{13}(\hat{\hat{w}}^{(b-2)})\right\} \\
&\leq 4\epsilon + 2^{nR_0}\Pr\{E_9(\hat{\hat{l}}^{(b-2)})\}\Pr\{E_{10}(\hat{\hat{l}}^{(b-2)})\}\Pr\{E_{11}(\hat{\hat{l}}^{(b-2)})\} \\
&\quad + 2^{nR_{\mathrm{SFB0}}}\Pr\{E_{13}(\hat{\hat{w}}^{(b-2)})\}.
\end{aligned} \tag{5.16}$$

Applying Lemma (5.2) repeatedly, we have

$$\Pr\{E_9(\hat{\hat{l}}^{(b-2)})\} \le 2^{-n(I(U;Y_2)-7\epsilon)}, \tag{5.17}$$

$$\Pr\{E_{10}(\hat{\hat{l}}^{(b-2)})\} \le 2^{-n(I(X_1;Y_2|U)-7\epsilon)}, \tag{5.18}$$

$$\Pr\{E_{11}(\hat{\hat{l}}^{(b-2)})\} \le 2^{-n(I(\hat{Y}_1;Y_2|X_1,U)-7\epsilon)}, \text{ and} \tag{5.19}$$

$$\Pr\{E_{13}(\hat{\hat{l}}^{(b-2)})\} \le 2^{-n(I(X_0;Y_2,\hat{Y}_1|X_1,U)-7\epsilon)}. \tag{5.20}$$

By substituting (5.17)–(5.20) into (5.16), we can conclude that when R_0 satisfies the constraint (5.9), and the information rate R_{SFB0} satisfies (5.10), the probability of error at the destination can be made arbitrarily small, i.e.,

$$P_e(\text{destination}) \le 5\epsilon,$$

as long as n is sufficiently large.

Therefore, any rate $R_{\text{SFB0}} \le I(X_0;Y_2,\hat{Y}_1|X_1,U)$ is achievable subject to constraints (5.7), (5.8), and (5.9) for a fixed joint distribution $p(\cdot) \in \mathcal{P}_0$. This completes the proof of the theorem. ■

Remark 5.1 In the above, constraint (5.5) *is applied such that the source is able to fully decode* $\hat{\mathbf{Y}}_1$*, the compressed version of the channel output sequence* $\mathbf{Y}_1$*. Similarly, constraint* (5.6) *needs to be satisfied such that the destination is able to decode* $\hat{\mathbf{Y}}_1$ *as well. Following [21, Theorem 6], with both its own channel output sequence* $\mathbf{Y}_2$ *and the compressed version of the channel output sequence at the relay,* $\hat{\mathbf{Y}}_1$*, the destination can achieve rate* $I(X_0;Y_2,\hat{Y}_1|X_1,U)$*.*

Remark 5.2 It is easily observed that when constraint (5.6) *dominates constraint* (5.5)*, i.e.,*

$$I(\hat{Y}_1,X_1,U;Y_2) \le I(X_1;Y_0|X_0,U) + I(\hat{Y}_1;Y_0,X_0|X_1,U),$$

we have R_{SFB0} *achievable subject to the single constraint*

$$I(\hat{Y}_1; Y_1|X_1, U) \leq I(\hat{Y}_1, X_1, U; Y_2) = I(X_1, U; Y_2) + I(\hat{Y}_1; Y_2|X_1, U). \tag{5.21}$$

Hence, the achievable rate R_{SFB0} *is potentially larger than the one in [21, Theorem 6], since our constraint* (5.21) *is more relaxed than the constraint in [21, Theorem 6]:*

$$I(\hat{Y}_1; Y_1|X_1) \leq I(X_1; Y_2) + I(\hat{Y}_1; Y_2|X_1). \tag{5.22}$$

Specifically, the relaxation is mainly due to the relationship $I(X_1, U; Y_2) \geq I(X_1; Y_2)$*. Having the uncertainty carried by* (X_1, U) *or* X_1 *resolved, the amount of remaining uncertainty about* $\hat{Y}_1$ *that we can extract from* Y_2 *should be the same for our coding scheme and the one in [21, Theorem 6], i.e.,* $I(\hat{Y}_1; Y_2|X_1, U) = I(\hat{Y}_1; Y_2|X_1)$*. Note that* U *is the auxiliary random variable introduced to induce the cooperation between the source and the relay in sending the codeword* $\hat{\mathbf{Y}}_1$*.*

However, when the feedback channel from the relay and the source is weak such that constraint (5.6) *is dominated by constraint* (5.5)*, the advantage of CF at the relay would be undermined due to the requirement of fully decoding the codeword* $\hat{\mathbf{Y}}_1$ *at the source. This is because in such a weak feedback situation, the pure CF strategy without requiring the source to decode the compressed version of the channel output sequence at the relay can enjoy a more relaxed constraint* (5.22)*. This is similar to the problem of the DF strategy when it is applied in a generic relay channel. When the link from the source to the relay is weak, the rate achievable with DF could be less than the capacity of the direct link from the source to the destination.*

To overcome this problem, we propose an extended coding scheme of the one in Theorem 5.1 to compress each received channel output sequence at the relay into two different versions of different rates. One version will be sent via relaying by the source in the manner of Theorem 5.1, while the other version will be sent solely through the direct link from the relay to the destination in the same manner as in

[21, Theorem 6]. This coding scheme establishes an improved achievable rate in Theorem 5.2.

With a slight abuse of notation, we reuse some of the auxiliary random variable names in each of the theorems in the rest of this chapter. Let U, V, $\hat{Y}_1$, and $\check{Y}_1$ be auxiliary random variables defined over arbitrary finite alphabets $\mathcal{U}$, $\mathcal{V}$, $\hat{\mathcal{Y}}_1$, and $\check{\mathcal{Y}}_1$, respectively. Let $\mathcal{P}_1$ denote the set of joint distributions $p(\cdot)$ that factor as follows

$$\begin{aligned} p(u, v, x_0, x_1, y_0, y_1, y_2, \hat{y}_1, \check{y}_1) = &\, p(u)p(x_0|u)p(v|u)p(x_1|v) \\ &\cdot (y_0, y_1, y_2|x_0, x_1)p(\hat{y}_1|y_1, u, v)p(\check{y}_1|y_1, x_1, u, v). \end{aligned}$$

Theorem 5.2 For the discrete memoryless relay channel with generalized feedback at the source, $\mathcal{C}_{\text{SFB}}$, the following rate is achievable

$$R_{\text{SFB1}} := \sup_{p(\cdot)\in\mathcal{P}_1} I(X_0; Y_2, \hat{Y}_1, \check{Y}_1|X_1, U, V), \tag{5.23}$$

subject to the following three constraints:

$$\begin{aligned} I(\hat{Y}_1; Y_1|U, V) &\leq I(V; Y_0|X_0, U) + I(\hat{Y}_1; Y_0, X_0|U, V), && (5.24)\\ I(\hat{Y}_1; Y_1|U, V) &\leq I(\hat{Y}_1, U, V; Y_2), \text{ and} && (5.25)\\ I(\check{Y}_1; Y_1|X_1, U, V) &\leq I(\check{Y}_1, X_1; Y_2|U, V). && (5.26) \end{aligned}$$

Proof: An outline of the proof is provided in Appendix B.1, based on which the detailed proof can be obtained by following the lines in the proof of Theorem 5.1. ∎

Remark 5.3 It can be observed that $\hat{\mathbf{Y}}_1$ serves as one compressed version of the channel output sequence at the relay, $\mathbf{Y}_1$, and it is sent to the destination with the aid of the source. On the other hand, being the other compressed version of the channel output sequence at the relay, $\check{\mathbf{Y}}_1$ is sent to the destination through *the link from the relay to the destination directly. With this coding flexibility, the transmission of the compressed codeword* $\check{\mathbf{Y}}_1$ *will not be bottle-necked even when the feedback link from the relay to the source is weak. At the same time, the coding scheme allows the source to cooperate with the relay in transmitting* $\hat{\mathbf{Y}}_1$ *to the destination, which exploits the feedback at the source in the same manner as in Theorem 5.1.*

Remark 5.4 It is evident that when we set $\check{Y}$ *to a constant, and choose* $V = X_1$*, Theorem 5.2 reduces to Theorem 5.1, i.e.,* $R_{\mathrm{SFB0}} \leq R_{\mathrm{SFB1}}$*.*

Remark 5.5 The coding scheme developed in Theorem 5.1 can be considered as a fully-decode-and-forward scheme applied at the source with respect to the compressed codewords $\hat{\mathbf{Y}}_1$*, while the coding scheme in Theorem 5.2 can be considered as a partially-decode-and-forward one with respect to the compressed codewords* $(\hat{\mathbf{Y}}_1, \check{\mathbf{Y}}_1)$*.*

5.3.2 A Rate Achieved by Compress-and-Forward

In contrast with the two DF-alike coding schemes developed above, a coding scheme similar to the CF coding strategy can also be developed at the source to exploit the feedback and *facilitate* the relay in forwarding the compressed version of the channel output sequence at the relay to the destination. An achievable rate established with such a coding scheme is presented as follows.

Let U, $\hat{Y}_0$, and $\hat{Y}_1$ be auxiliary random variables defined over arbitrary finite alphabets $\mathcal{U}$, $\hat{\mathcal{Y}}_0$, and $\hat{\mathcal{Y}}_1$, respectively. Let $\mathcal{P}_2$ denote the set of joint distributions $p(\cdot)$ that factor as follows

$$p(u,x_0, x_1, y_0, y_1, y_2, \hat{y}_0, \hat{y}_1) = p(u)p(x_0|u)p(x_1)p(y_0, y_1, y_2|x_0, x_1)p(\hat{y}_0|y_0, u)p(\hat{y}_1|y_1, x_1).$$

Theorem 5.3 For the discrete memoryless relay channel with generalized feedback *at the source, $\mathcal{C}_{\text{SFB}}$, the following rate R_{SFB2} is achievable:*

$$R_{\text{SFB2}} := \sup_{p(\cdot)\in\mathcal{P}_2} I(X_0; Y_2, \hat{Y}_1, \hat{Y}_0 | X_1, U),$$

subject to the following constraints

$$I(\hat{Y}_0; Y_0|U) \leq I(U; Y_2) + I(\hat{Y}_0; Y_2|U), \text{ and} \tag{5.27}$$

$$I(\hat{Y}_1; Y_1|X_1) \leq I(X_1; Y_2, \hat{Y}_0|U) + I(\hat{Y}_1; Y_2, \hat{Y}_0, U|X_1). \tag{5.28}$$

Proof: An outline of the proof is presented in Appendix B.2. ■

Remark 5.6 The source acts as a CF relay that compresses the channel output sequence $\mathbf{Y}_0$ at the source to obtain $\hat{\mathbf{Y}}_0$, and then forwards the compressed version to the destination. The destination decodes $\hat{\mathbf{Y}}_1$, the compressed version of the channel output sequence at the relay, from both $\mathbf{Y}_2$ and $\hat{\mathbf{Y}}_0$. Lastly, the destination decodes the source messages from its channel output sequence $\mathbf{Y}_2$, and the two compressed channel output sequences, $\hat{\mathbf{Y}}_1$ and $\hat{\mathbf{Y}}_0$, which achieves the rate $I(X_0; Y_2, \hat{Y}_1, \hat{Y}_0|X_1, U)$.

Remark 5.7 In this coding scheme, the link from the relay to the source will not become a bottleneck, since the source only needs to compress whatever it has received, which does not impose any constraint on the rate of the compressed channel output sequences at the relay.

5.3.3 Special Cases

By the definition of $\mathcal{C}_{\text{SFB}}$, $p(y_0, y_1, y_2|x_0, x_1)$, the feedback at the source Y_0 is arbitrarily correlated with Y_1 and Y_2. It is therefore natural for us to consider the following two special cases: 1) $Y_0 = Y_2$, i.e., the feedback is the same as the channel output at the destination; and 2) $Y_0 = Y_1$, i.e., the feedback is the same as the channel output at the relay. These two cases are in fact the two perfect feedback cases studied in [72]. In what follows, we show that Theorem 5.1 implies the results in Theorems 1 and 2 of [72].

Consider the first case of $Y_0 = Y_2$. We specialize the result in Theorem 5.1 by substituting Y_0 with Y_2 the corresponding terms, and we have an achievable rate for this special case obtained as follows.

Let $\mathcal{P}_0^{\text{D}}$ denote the set of joint distributions $p(\cdot)$ that factor as follows:

$$p(u, x_0, x_1, y_1, y_2, \hat{y}_1) = p(u)p(x_0|u)p(x_1|u)p(y_1, y_2|x_0, x_1)p(\hat{y}_1|y_1, x_1, u).$$

Corollary 5.1 ([72, Theorem 1]) For the discrete memoryless relay channel with perfect feedback from the destination to the source, i.e., $\mathcal{C}_{\text{SFB}}$ *with* $Y_0 = Y_2$*, the following rate is achievable:*

$$R_{\text{SFB0}}^{\text{D}} := \sup_{p(\cdot) \in \mathcal{P}_0^{\text{D}}} I(X_0; Y_2, \hat{Y}_1 | X_1, U),$$

subject to the following two constraints

$$I(\hat{Y}_1; Y_1 | X_1, U) \leq I(X_1; Y_2 | X_0, U) + I(\hat{Y}_1; Y_2, X_0 | X_1, U), \text{ and} \tag{5.29}$$

$$I(\hat{Y}_1; Y_1 | X_1, U) \leq I(\hat{Y}_1, X_1, U; Y_2). \tag{5.30}$$

Remark 5.8 First note that there is a slight difference between the notation used in this chapter and that in [72]: we use X_0 *and* X_1 *to denote the inputs at the source and relay respectively, while [72] uses* X_1 *and* X_2*. Also note that constraints* (5.29) *and* (5.30) *are equivalent to the corresponding two constraints in [72, Theorem 1]. We demonstrate this by transforming these two constraints as follows. For* *constraint* (5.29)*, we have*

$$\begin{aligned} I(X_1; Y_2|X_0, U) &\geq I(\hat{Y}_1; Y_1|X_1, U) - I(\hat{Y}_1; Y_2, X_0|X_1, U) \\ &\geq -H(\hat{Y}_1|Y_1, X_1, U) + H(\hat{Y}_1|Y_2, X_0, X_1, U) \\ &\overset{(d)}{=} -H(\hat{Y}_1|Y_1, Y_2, X_0, X_1, U) + H(\hat{Y}_1|Y_2, X_0, X_1, U) \\ &= I(\hat{Y}_1; Y_1|Y_2, X_0, X_1, U), \end{aligned}$$

where (d) follows from the conditional independence between $\hat{Y}_1$ *and* (Y_2, X_0) *given* (Y_1, X_1, U) *such that* $H(\hat{Y}_1|Y_1, Y_2, X_0, X_1, U) = H(\hat{Y}_1|Y_1, X_1, U)$*. Similarly, for constraint* (5.30) *we have*

$$\begin{aligned} I(\hat{Y}_1, X_1, U; Y_2) &\geq I(\hat{Y}_1; Y_1|X_1, U) \longleftrightarrow \\ I(X_1, U; Y_2) &\geq I(\hat{Y}_1; Y_1|X_1, U) - I(\hat{Y}_1; Y_2|X_1, U) \\ &= -H(\hat{Y}_1|Y_1, X_1, U) + H(\hat{Y}_1|Y_2, X_1, U) \\ &= -H(\hat{Y}_1|Y_1, Y_2, X_1, U) + H(\hat{Y}_1|Y_2, X_1, U) \\ &= I(\hat{Y}_1; Y_1|Y_2, X_1, U). \end{aligned}$$

We now consider the second case of $Y_0 = Y_1$. By substituting Y_0 with Y_1 and letting $U = X_1$ in Theorem 5.1, we have the following achievable rate for this case.

Let $\mathcal{P}_0^{\mathrm{R}}$ denote the set of joint distributions $p(\cdot)$ that factor as follows:

$$p(x_0, x_1, y_1, y_2, \hat{y}_1) = p(x_1)p(x_0|x_1)p(y_1, y_2|x_0, x_1)p(\hat{y}_1|y_1, x_1).$$

Corollary 5.2 ([72, Theorem 2]) For the discrete memoryless relay channel with perfect feedback from the relay to the source, i.e., $\mathcal{C}_{\mathrm{SFB}}$ *with* $Y_0 = Y_1$*, the following rate is achievable:*

$$R_{\mathrm{SFB0}}^{\mathrm{R}} := \sup_{p(\cdot) \in \mathcal{P}_0^{\mathrm{R}}} I(X_0; Y_2, \hat{Y}_1|X_1, U),$$

subject to the following constraint

$$I(\hat{Y}_1; Y_1|X_1) \leq I(\hat{Y}_1, X_1; Y_2). \tag{5.31}$$

Remark 5.9 Note that the constraint (5.31) *is equivalent to the corresponding one in [72, Theorem 2], which can be shown by following the lines in Remark 5.8.*

Remark 5.10 It has been shown in [72] that with perfect feedback, both achievable rates $R^{\text{D}}_{\text{SFB0}}$ *and* $R^{\text{R}}_{\text{SFB0}}$ *strictly improve the various achievable rates for the generic relay channel under some specific settings. Since our result in Theorem 5.1 includes both* $R^{\text{D}}_{\text{SFB0}}$ *and* $R^{\text{R}}_{\text{SFB0}}$ *as special cases, we can claim that our coding scheme in Theorem 5.1 indeed exploits the feedback, and can also achieve improved information rates over those for the generic relay channel under the same channel settings. Note that Theorem 5.1 is included as a special case of Theorem 5.2.*

5.4 Achievable Rates for $\mathcal{C}_{\text{DFB}}$

In this section, we consider the second feedback setting, $\mathcal{C}_{\text{DFB}}$, in which the destination works in full duplex mode whereas the source can only transmit signals but not receive them. This feedback configuration can be considered as a generalization of the perfect feedback case investigated in [21]. It has been shown in [21] that when the feedback is perfect, i.e., the relay has perfect causal knowledge of Y_2, the channel output at the destination Y_2 can be considered to be a degraded version of the channel output pair available at the relay, (Y_1, Y_2). Hence, this perfect feedback configuration can be considered to be a degraded relay channel [21]. Consequently, the capacity for the relay channel with perfect causal feedback from the destination to the relay can be achieved by solely applying the DF strategy.

In our current feedback configuration $\mathcal{C}_{\text{DFB}}$, the destination may not be able to send the perfect channel output Y_2 to the relay, which depends on the condition of the channel from the destination to the relay. However, it is natural to apply the CF strategy at the destination to send the compressed version of its channel output sequence to the relay, and let the relay perform DF in the next block. We present an achievable rate established by this coding scheme as follows.

Let $\hat{Y}_2$ be an auxiliary random variable defined over an arbitrary finite alphabet $\hat{\mathcal{Y}}_2$. Let $\mathcal{P}_1^*$ denote the set of all the joint distributions $p(\cdot)$ that factor in the form

$$p(x_0, x_1, x_2, y_1, y_2, \hat{y}_2) = p(x_1)p(x_0|x_1)p(x_2)p(y_1, y_2|x_0, x_1, x_2)p(\hat{y}_2|y_2, x_2).$$

Theorem 5.4 For the discrete memoryless relay channel with generalized feedback transmitted from the destination, $\mathcal{C}_{\mathrm{DFB}}$, the rate R_{DFB1} is achievable:

$$R_{\mathrm{DFB1}} := \sup_{p(\cdot)\in\mathcal{P}_1^*} \min\{I(X_0, X_1; Y_2|X_2), I(X_0; Y_1, \hat{Y}_2|X_1, X_2)\},$$

subject to the constraint

$$I(\hat{Y}_2; Y_2|X_2) \leq I(X_2; Y_1|X_1) + I(\hat{Y}_2; Y_1, X_1|X_2). \tag{5.32}$$

Proof: By setting $U = X_0$ in Theorem 5.5, the theorem follows immediately. A detailed proof can be obtained by following similar steps in the proof of Theorem 5.5 provided in Appendix B.3. ∎

Remark 5.11 When the channel from the destination to the relay is strong enough such that the following inequality holds:

$$H(Y_2|X_2) \leq I(X_2; Y_1|X_1) + I(Y_2; Y_1, X_1|X_2),$$

it follows that the channel output sequence of the destination $\mathbf{Y}_2$ can be sent to the relay without compression. Then, the rate R_{DFB1} reduces to the capacity result in [21, Theorem 3] for the relay channel with perfect causal feedback from the *destination to the relay.*

Remark 5.12 On the other hand, when the channel from the source to the relay and the channel from the destination to the relay are both weak, i.e., Y_1 *and* $\hat{Y}_2$ *are both 'noisy' such that the term* $I(X_0; Y_1, \hat{Y}_2 | X_1, X_2)$ *is limited, requiring the relay to fully decode the message from the source would result in a bottleneck for the achievable rate* R_{DFB1}*. One possible way to overcome this is to let the relay perform partially-decode-and-forward, and we have an extended result as follows.*

Let U and $\hat{Y}_2$ be two auxiliary random variables defined over arbitrary finite alphabets $\mathcal{U}$ and $\hat{\mathcal{Y}}_2$. Let $\mathcal{P}_2^*$ denote the set of all the joint distributions $p(\cdot)$ that factor in the form

$$p(u, x_0, x_1, x_2, y_1, y_2, \hat{y}_2) = p(x_1)p(u|x_1)p(x_0|u)p(x_2)p(y_1, y_2|x_0, x_1, x_2)p(\hat{y}_2|y_2, x_2). \tag{5.33}$$

Theorem 5.5 For the discrete memoryless relay channel with generalized feedback sent from the destination, $\mathcal{C}_{\mathrm{DFB}}$*, the following rate* R_{DFB2} *is achievable:*

$$R_{\mathrm{DFB2}} := \sup_{p(\cdot) \in \mathcal{P}_2^*} \min\{I(U, X_0, X_1; Y_2|X_2), I(U; Y_1, \hat{Y}_2|X_2, X_1) + I(X_0; Y_2|X_2, X_1, U)\},$$

subject to the constraint

$$I(\hat{Y}_2; Y_2|X_2) \leq I(X_2; Y_1|X_1) + I(\hat{Y}_2; Y_1, X_1|X_2). \tag{5.34}$$

Proof: An outline of the proof is provided in Appendix B.3. ■

Remark 5.13 In both Theorems 5.4 and 5.5, the destination compresses its channel output sequence to obtain $\hat{\mathbf{Y}}_2$ *and sends it to the relay. The relay hence can decode more information from both* $\mathbf{Y}_1$ *and* $\hat{\mathbf{Y}}_2$ *than it can decode from* $\mathbf{Y}_1$ *only. This allows more information to be sent to the destination through the cooperation between the source and the relay, which improves the achievable rate.*

5.5 Concluding Remarks

In this chapter, we have derived achievable rates for the discrete memoryless relay channel with generalized feedback, in which either the source or the destination operates in full duplex mode. We have further shown that the derived achievable rate results include several previously known results as special cases.

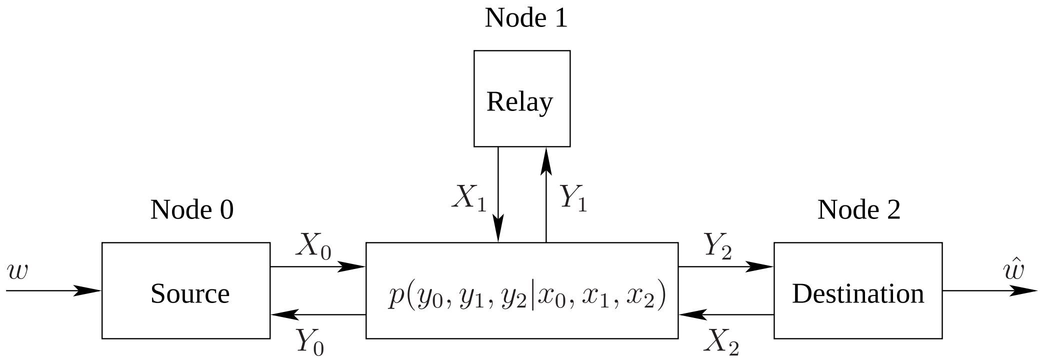

Figure 5.3: Three-node relay channel in which all nodes are in full duplex mode.

An interesting problem, as a natural extension of the two configurations, is the relay channel with all three nodes operating in full duplex mode, as illustrated in Fig. 5.3. We can directly extend the coding schemes developed in this chapter to this problem, but apparently the resultant coding scheme and achievable rate would be rather complicated. It is hence interesting future work to develop new and simpler coding schemes which can exploit the generalized feedback.

Chapter 6

Summary of Contributions and Future Work

Aiming to tackle the three challenging issues in wireless networks, correlated sources, interference, and feedback, we investigated four different multi-nodal communication scenarios in the domain of wireless networks from an information theoretic perspective. The first three scenarios or channel models can be considered as variants of the IC: 1) ICC, 2) IC-DMS, and 3) the ICF. The last one that we investigated is the relay channel with generalized feedback, for which two different feedback configurations are considered. In the following sections, we summarize our contributions on each of these four channel models, and we also point out some of the possible future work as extensions of our work in this thesis.

6.1 Summary of Contributions

We first investigated the ICC, in which two senders wish to transmit to their respective receivers some private information as well as some common information. We proposed a superposition coding scheme that consists of cascaded superposition encoding at the sender side and simultaneous decoding at the receiver side. This coding scheme allows the common information to be transmitted to the receivers in a fully cooperative manner, upon which the private information is sent using the rate splitting technique. We derived a new achievable rate region for the discrete memoryless ICC, which is shown to include a number of existing results as special cases. We also extended our rate region to two special cases, a class of ICCs in which one sender has no private information to transmit, and a class of DICCs. We found that our achievable rate region is the capacity for this class of deterministic channels, as we were able to establish the corresponding converse. We further extended our result from the discrete memoryless case to the Gaussian case. Moreover, we showed that our rate region strictly improves an existing result using a Gaussian example.

We then investigated the IC-DMS (also termed as the cognitive radio channel), in which two senders wish to send some private information (without common information) to their respective receivers, whereas one sender (sender 2) is assumed to have the *a priori* knowledge of the information that the other one wishes to send. In such a channel, two different types of interference are involved: 1) receiver 1 suffers the first type of interference from the sender 2's transmissions of its own information; 2) receiver 2 suffers the other type of interference, as the interference is non-casually known at the sender 2. We proposed a new coding scheme for this channel. In this coding scheme, sender 1 encodes its message independently, while sender 2 first performs rate splitting, and then applies Gel'fand-Pinsker coding to encode the two sub-messages by treating the codeword to be transmitted by sender 1 as known interference. Receiver 1 is required to perform a joint decoding of the intended message from sender 1 and a sub-message from sender 1, whereas receiver 2 only needs to decode the two intended sub-messages from sender 2. We derived a new achievable rate region for the discrete memoryless IC-DMS with the proposed coding scheme. We showed that our rate region includes two existing results as special cases. We also extended our rate region to the Gaussian case, and we further demonstrated that our rate region offers strict improvements over the existing ones in the *high-interference-gain* regime using Gaussian examples.

We then considered the discrete memoryless ICF. This channel model is obtained by assuming that in an IC, the channel outputs of the receivers are made causally and perfectly available at the respective senders. We proposed a block Markov coding scheme for the channel based on partially DF strategy and rate splitting. The coding scheme allows the two senders to partially decode the information sent by each other from the feedback such that, the two senders can then cooperate with each other to resolve the remaining uncertainty at the receivers. We derived a corresponding new achieved rate region for the discrete memoryless ICF, for which both the implicit and explicit descriptions were presented.

We lastly investigated the relay channel with generalized feedback. Two different feedback configurations were considered. In the first configuration, the source is assumed to be able to operate in the full duplex mode, i.e., both the source and the relay can receive and transmit signals simultaneously. We proposed to let the relay perform CF strategy, such that the source can extract new information from the feedback in order to exploit the feedback. We proposed three coding strategies for this feedback configuration, by allowing the source node to act as a relay to the original relay in the channel. The resultant achievable rates were shown to include existing results for the relay channel with perfect feedback as special cases. In the second configuration, both the relay and the destination are assumed to operate in the full duplex mode, but not the source. Our proposed scheme for this channel allows the destination node to perform CF, which facilitates the relay to decode more information from the source. The relay then performs DF or partial DF to cooperate with the source to resolve the remaining uncertainty at the destination about earlier block transmissions. We demonstrated that the corresponding achievable rates are asymptotically optimal for the extreme case.

6.2 Future Work

We note that the IC-DMS contains the BC as a special case, but our achievable rate region for the IC-DMS does not include the best achievable rate region, Marton's region [13], for the BC as a special case. This problem is mainly due to the fact that we are still viewing the IC-DMS as a variant of IC, and thus we have applied the coding technique invented for the IC, the rate splitting technique. This problem also suggests that there is certain room for us to improve our current coding scheme. A further investigation of the IC-DMS under the BC framework could potentially lead to a new or better achievable rate region. Trying to integrate both the IC and BC frameworks may result in an even better result.

To the other end, we also suspect our achievable result for the IC-DMS could be optimal for a certain class of channels in the high-interference-gain regime. This requires us to investigate the outer bounds of the IC-DMS, or derive new outer bounds for the channel. This could be another possible direction of extension for our current work on the IC-DMS.

It can be observed that our achievable rates derived for the relay channel with generalized feedback all involve multiple auxiliary random variables such that, it becomes very hard to perform evaluation of the respective rate region's Gaussian counterpart. We may wish to construct simpler coding schemes, which can also exploit the generalized feedback but involve fewer auxiliary random variables. Also, as mentioned in Section 5.5 of Chapter 5, the relay channel with three full duplex nodes would be an interesting extension of our work on the two feedback configurations. However, direct extension of our current coding schemes could lead to much more complex coding schemes for the setting with three full duplex nodes. This still urges us to seek for simpler coding schemes that can help to exploit the generalized feedback at/from each node.

Appendix A

Appendices to Chapter 2

A.1 Proof of Lemma 2.2

As the following lemma will be frequently used, we state it before the proof of Lemma 2.2.

Lemma A.1 ([40, Theorem 14.2.3]) Let $A_\epsilon^{(n)}$ *denote the typical set for the probability distribution* $p(s_1, s_2, s_3)$*, and let*

$$\Pr(\mathbf{S}_1' = \mathbf{s}_1, \mathbf{S}_2' = \mathbf{s}_2, \mathbf{S}_3' = \mathbf{s}_3) = \prod_{i=1}^{n} p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i}),$$

then $\Pr\{(\mathbf{S}_1', \mathbf{S}_2', \mathbf{S}_3') \in A_\epsilon^{(n)}\} \doteq 2^{-n(I(S_1;S_2|S_3)\pm 6\epsilon)}$.

Proof of Lemma 2.2: [**Codebook Generation**.] Let us fix a joint distribution $p(\cdot)$ that factors in the form of (2.1). We first generate 2^{nR_0} independent codewords $\mathbf{U}_0(i)$, $i \in \{1, ..., 2^{nR_0}\}$, according to $\prod_{t=1}^{n} p(u_{0,t})$. At encoder 1, for each codeword $\mathbf{u}_0(i)$, generate $2^{nR_{12}}$ independent codewords $\mathbf{U}_1(i, j)$, $j \in \{1, ..., 2^{nR_{12}}\}$, according to $\prod_{t=1}^{n} p(u_{1,t}|u_{0,t})$. Subsequently, for each pair of codewords $(\mathbf{u}_0(i), \mathbf{u}_1(i, j))$, generate $2^{nR_{11}}$ independent codewords $\mathbf{X}_1(i, j, k)$, $k \in \{1, ..., 2^{nR_{11}}\}$, according to $\prod_{t=1}^{n} p(x_{1,t}|u_{1,t}, u_{0,t})$. Similarly at encoder 2, for each codeword $\mathbf{u}_0(i)$, generate $2^{nR_{21}}$ independent codewords $\mathbf{U}_2(i, l)$, $l \in \{1, ..., 2^{nR_{21}}\}$, according to $\prod_{t=1}^{n} p(u_{2,t}|u_{0,t})$. Subsequently, for each codeword pair $(\mathbf{u}_0(i), \mathbf{u}_2(i, l))$, generate $2^{nR_{22}}$ independent codewords $\mathbf{X}_2(i,l,m)$, $m \in \{1,...,2^{nR_{22}}\}$, according to $\prod_{t=1}^{n} p(x_{2,t}|u_{2,t},u_{0,t})$. The entire codebook consisting of all the codewords $\mathbf{u}_0(i)$, $\mathbf{u}_1(i,j)$, $\mathbf{x}_1(i,j,k)$, $\mathbf{u}_2(i,l)$ and, $\mathbf{x}_2(i,l,m)$ with $i \in \{1,...,2^{nR_0}\}$, $j \in \{1,...,2^{nR_{12}}\}$, $k \in \{1,...,2^{nR_{11}}\}$, $l \in \{1,...,2^{nR_{21}}\}$, and, $m \in \{1,...,2^{nR_{22}}\}$ is revealed to both receivers.

[**Encoding** & **Transmission**.] Suppose that the source message vector generated at the two senders is $(n_0,n_1,l_1,n_2,l_2)=(i,j,k,l,m)$. Sender 1 transmits codeword $\mathbf{x}_1(i,j,k)$ with n channel uses, while sender 2 transmits codeword $\mathbf{x}_2(i,l,m)$ with n channel uses. The transmissions are assumed to be perfectly synchronized.

[**Decoding**.] Each receiver accumulates an n-length channel output sequence, $\mathbf{y}_1$ (receiver 1) or $\mathbf{y}_2$ (receiver 2). Let $A_\epsilon^{(n)}$ denote the typical sets of the respective joint distributions. Decoder 1 declares that $(\hat{i},\hat{j},\hat{k})$ is received, if $(\hat{i},\hat{j},\hat{k})$ is the unique message vector such that $(\mathbf{u}_0(\hat{i}),\mathbf{u}_1(\hat{i},\hat{j}),\mathbf{x}_1(\hat{i},\hat{j},\hat{k}),\mathbf{u}_2(\hat{i},l),\mathbf{y}_1) \in A_\epsilon^{(n)}$ for some l; otherwise declare an error. Similarly, decoder 2 looks for a unique message vector $(\hat{i},\hat{l},\hat{m})$ such that $(\mathbf{u}_0(\hat{i}),\mathbf{u}_2(\hat{i},\hat{l}),\mathbf{x}_2(\hat{i},\hat{l},\hat{m}),\ \mathbf{u}_1(\hat{i},j),\mathbf{y}_2) \in A_\epsilon^{(n)}$ for some j; otherwise declare an error.

[**Analysis of the Probability of Decoding Error**.] Because of the symmetry of the codebook generation, the probability of error does not depend on which message vector is encoded and transmitted. Since the messages are uniformly generated over their respective ranges, the average error probability over all the possible messages is equal to the probability of error incurred when any message vector is encoded and transmitted. We hence only analyze the probability of error at decoder 1 in details, since the same analysis can be performed for decoder 2. Without loss of generality, we assume that a source message vector (n_0,n_l,l_1,n_2,l_2)=$(1,1,1,1,1)$ is encoded and transmitted for the subsequent analysis. We first define the event

$$E_{ijkl} := \{(\mathbf{U}_0(i),\mathbf{U}_1(i,j),\mathbf{X}_1(i,j,k),\mathbf{U}_2(i,l),\mathbf{Y}_1) \in A_\epsilon^{(n)}\}.$$

The possible error events can be grouped into two classes: 1) the codewords transmitted are not jointly typical, i.e., E_{1111}^c happens; 2) there exist some $(i,j,k) \neq$

$(1,1,1)$ such that E_{ijkl} happens (l may not be 1). Thus the probability of error at decoder 1 can be expressed as

$$P_{e,1}^{(n)} = \Pr(E_{1111}^c \bigcup \cup_{(i,j,k)\neq(1,1,1)} E_{ijkl}). \tag{A.1}$$

By applying the union bound, we can upper-bound (A.1) as

$$\begin{aligned} P_{e,1}^{(n)} &\leq \Pr(E_{1111}^c) + \Pr(\cup_{(i,j,k)\neq(1,1,1)} E_{ijkl}) \\ &\leq \Pr(E_{1111}^c) + \sum_{i\neq 1} \Pr(E_{i111}) + \sum_{i\neq 1, l\neq 1} \Pr(E_{i11l}) + \sum_{j\neq 1} \Pr(E_{1j11}) \\ &\quad + \sum_{j\neq 1, l\neq 1} \Pr(E_{1j1l}) + \sum_{k\neq 1} \Pr(E_{11k1}) + \sum_{k\neq 1, l\neq 1} \Pr(E_{11kl}) + \sum_{i\neq 1, j\neq 1} \Pr(E_{ij11}) \\ &\quad + \sum_{i\neq 1, j\neq 1, l\neq 1} \Pr(E_{ij1l}) + \sum_{i\neq 1, k\neq 1} \Pr(E_{i1k1}) + \sum_{i\neq 1, k\neq 1, l\neq 1} \Pr(E_{i1kl}) \\ &\quad + \sum_{j\neq 1, k\neq 1} \Pr(E_{1jk1}) + \sum_{j\neq 1, k\neq 1, l\neq 1} \Pr(E_{1jkl}) + \sum_{i\neq 1, j\neq 1, k\neq 1} \Pr(E_{ijk1}) \\ &\quad + \sum_{i\neq 1, j\neq 1, k\neq 1, l\neq 1} \Pr(E_{ijkl}). \end{aligned} \tag{A.2}$$

Due to the asymptotic equipartition property (AEP) [40], $\Pr(E_{1111}^c)$ in (A.2) can be made arbitrarily small as long as n is sufficiently large. The rest of the fourteen probability terms in (A.2) can be evaluated through a standard procedure, which is demonstrated as follows. To evaluate $\Pr(E_{i111})$, we apply Lemma A.1 by letting $\mathbf{S}_1' = (\mathbf{U}_0(i), \mathbf{U}_1(i,1), \mathbf{X}_1(i,1,1), \mathbf{U}_2(i,1))$, $\mathbf{S}_2' = \mathbf{Y}_1$, and $\mathbf{S}_3' = \emptyset$ with $\emptyset$ denoting the empty set. Since the assumption of Lemma 3 on the joint distribution of $\mathbf{S}_1'$, $\mathbf{S}_2'$, and $\mathbf{S}_3'$ is satisfied, we have

$$\Pr(E_{i111}) \leq 2^{-n(I(U_0,U_1,X_1,U_2;Y_1)-6\epsilon)} \overset{(a)}{=} 2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)}.$$

Note that (a) follows from the fact that $I(U_1;Y_1|U_0,U_2,X_2) = 0$, which is because U_1, (U_0,U_2,X_1), and Y_1 forms a Markov chain $U_1 \rightarrow (U_0,U_2,X_1) \rightarrow Y_1$. Since the case with $\mathbf{S}_3' = \emptyset$ seems not archetypal, we evaluate one more probability term, $P(E_{1jk1})$. Again, we use Lemma A.1 by letting $\mathbf{S}_1' = (\mathbf{U}_1(1,j), \mathbf{X}_1(1,j,k))$,

$\mathbf{S}'_2 = \mathbf{Y}_1$, and $\mathbf{S}'_3 = (\mathbf{U}_0(1), \mathbf{U}_2(1,1))$ to obtain

$$\Pr(E_{1jk1}) \leq 2^{-n(I(U_1,X_1;Y_1|U_0,U_2)-6\epsilon)}.$$

By repeatedly applying Lemma A.1, we obtain upper-bounds of the remaining twelve probability terms. Further, we employ these bounds to derive an upper-bound of the probability of error at decoder 1 as

$$\begin{aligned}
P_{e,1}^{(n)} \leq& \epsilon + 2^{nR_0}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} + 2^{n(R_0+R_{21})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} \\
&+ 2^{nR_{12}}2^{-n(I(X_1;Y_1|U_0,U_2)-6\epsilon)} + 2^{n(R_{12}+R_{21})}2^{-n(I(X_1,U_2;Y_1|U_0)-6\epsilon)} \\
&+ 2^{nR_{11}}2^{-n(I(X_1;Y_1|U_0,U_1,U_2)-6\epsilon)} + 2^{n(R_{11}+R_{21})}2^{-n(I(X_1,U_2;Y_1|U_0,U_1)-6\epsilon)} \\
&+ 2^{n(R_0+R_{12})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} + 2^{n(R_0+R_{12}+R_{21})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} \\
&+ 2^{n(R_0+R_{11})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} + 2^{n(R_0+R_{11}+R_{21})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} \\
&+ 2^{n(R_{12}+R_{11})}2^{-n(I(X_1;Y_1|U_0,U_2)-6\epsilon)} + 2^{n(R_{12}+R_{11}+R_{21})}2^{-n(I(X_1,U_2;Y_1|U_0)-6\epsilon)} \\
&+ 2^{n(R_0+R_{12}+R_{11})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)} \\
&+ 2^{n(R_0+R_{12}+R_{11}+R_{21})}2^{-n(I(U_0,X_1,U_2;Y_1)-6\epsilon)}.
\end{aligned} \tag{A.3}$$

It is now easy to check that when (2.2)–(2.6) hold and n is sufficiently large, we have $P_{e,1}^{(n)} \leq 15\epsilon$. By symmetry, we have $P_{e,2}^{(n)} \leq 15\epsilon$ for decoder 2, when (2.7)–(2.11) hold and n is sufficiently large. Hence, $\max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} \leq 15\epsilon$, and Lemma 2 readily follows. ■

A.2 Proof of the Convexity of $\mathcal{R}_m$

Let $(R_0^1, R_{12}^1, R_{11}^1, R_{21}^1, R_{22}^1)$ and $(R_0^2, R_{12}^2, R_{11}^2, R_{21}^2, R_{22}^2)$ be two arbitrary rate quintuples belonging to $\mathcal{R}_m$. It suffices to show that for given any $\alpha \in [0,1]$, $(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2) \in \mathcal{R}_m$. Note that the rate region $\mathcal{R}_m$ is the union of regions $\mathcal{R}_m(p)$ over all $p(\cdot) \in \mathcal{P}^*$. Thus, there must exist two sets of auxiliary random variables, namely (U_0^1, U_1^1, U_2^1)

and (U_0^2, U_1^2, U_2^2) such that their joint distributions $p_1(\cdot)$ and $p_2(\cdot)$ factor as

$$
\begin{aligned}
p_1(u_0^1, u_1^1, u_2^1, x_1, x_2, y_1, y_2) &= p(u_0^1)p(u_1^1|u_0^1)p(u_2^1|u_0^1) \\
&\quad \cdot p(x_1|u_1^1, u_0^1)p(x_2|u_2^1, u_0^1)p(y_1, y_2|x_1, x_2), \\
p_2(u_0^2, u_1^2, u_2^2, x_1, x_2, y_1, y_2) &= p(u_0^2)p(u_1^2|u_0^2)p(u_2^2|u_0^2) \\
&\quad \cdot p(x_1|u_1^2, u_0^2)p(x_2|u_2^2, u_0^2)p(y_1, y_2|x_1, x_2).
\end{aligned}
$$

Let T be the independent random variable, taking the value 1 with probability α and 2 with probability $1-\alpha$. We define a new set of auxiliary random variables (U_0, U_1, U_2) such that $U_0 = (U_0^T, T)$, $U_1 = U_1^T$, and $U_2 = U_2^T$, and then their joint distribution $p_3(\cdot)$ can factor

$$
\begin{aligned}
p_3(u_0, u_1, u_2, x_1, x_2, y_1, y_2) &= p(u_0)p(u_1|u_0)p(u_2|u_0) \\
&\quad \cdot (x_1|u_1, u_0)p(x_2|u_2, u_0)p(y_1, y_2|x_1, x_2).
\end{aligned}
$$

Since $p_3(\cdot) \in \mathcal{P}^*$, we have $\mathcal{R}_m(p_3) \subseteq \mathcal{R}_m$. It is easy to show that $(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2) \in \mathcal{R}_m(p_3)$ by following the steps used to prove the convexity of the capacity region for the MACC in Appendix A of [56]. Therefore, we conclude $(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2) \in \mathcal{R}_m(p_3) \subseteq \mathcal{R}_m$, which proves the convexity.

A.3 Proof of Corollary 2.1

1. Fourier-Motzkin Elimination

We next show in details, how to apply Fourier-Motzkin elimination to obtain the explicit rate region depicted by (2.12)–(2.24).

Step 1: By defining

$$
\begin{aligned}
a_1 &:= I(X_1; Y_1|U_0, U_1, U_2),\\
b_1 &:= I(X_1; Y_1|U_0, U_2),\\
c_1 &:= I(X_1, U_2; Y_1|U_0, U_1),\\
d_1 &:= I(X_1, U_2; Y_1|U_0),\\
e_1 &:= I(U_0, X_1, U_2; Y_1),\\
a_2 &:= I(X_2; Y_2|U_0, U_2, U_1),\\
b_2 &:= I(X_2; Y_2|U_0, U_1),\\
c_2 &:= I(X_2, U_1; Y_2|U_0, U_2),\\
d_2 &:= I(X_2, U_1; Y_2|U_0),\\
e_2 &:= I(U_0, X_2, U_1; Y_2),
\end{aligned}
$$

and substituting R_{11} with $R_1 - R_{12}$ and R_{22} with $R_2 - R_{21}$, we can rewrite the implicit rate region (2.2)–(2.11) for a fixed joint distribution $p(\cdot) \in \mathcal{P}^*$ as

$$R_1 - R_{12} \leq a_1, \tag{A.4}$$

$$R_1 \leq b_1, \tag{A.5}$$

$$R_1 - R_{12} + R_{21} \leq c_1, \tag{A.6}$$

$$R_1 + R_{21} \leq d_1, \tag{A.7}$$

$$R_0 + R_1 + R_{21} \leq e_1; \tag{A.8}$$

$$R_2 - R_{21} \leq a_2, \tag{A.9}$$

$$R_2 \leq b_2, \tag{A.10}$$

$$R_2 - R_{21} + R_{12} \leq c_2, \tag{A.11}$$

$$R_2 + R_{12} \leq d_2, \tag{A.12}$$

$$R_0 + R_2 + R_{12} \leq e_2; \tag{A.13}$$

$$-R_{12} \le 0, \tag{A.14}$$

$$-R_1 + R_{12} \le 0, \tag{A.15}$$

$$-R_{21} \le 0, \tag{A.16}$$

$$-R_2 + R_{21} \le 0, \tag{A.17}$$

Categorize (A.4)–(A.17) into the following three groups such that the inequalities in group 1 do not contain the R_{12} term, those in group 2 contain the negative R_{12} term, and those in group 3 contain the positive R_{12} term.

$$R_1 \le b_1, \tag{A.18}$$

$$R_1 + R_{21} \le d_1, \tag{A.19}$$

$$R_0 + R_1 + R_{21} \le e_1; \tag{A.20}$$

$$R_2 - R_{21} \le a_2, \tag{A.21}$$

$$R_2 \le b_2, \tag{A.22}$$

$$-R_{21} \le 0, \tag{A.23}$$

$$-R_2 + R_{21} \le 0; \tag{A.24}$$

and

$$R_1 - R_{12} \le a_1, \tag{A.25}$$

$$R_1 - R_{12} + R_{21} \le c_1, \tag{A.26}$$

$$-R_{12} \le 0; \tag{A.27}$$

and

$$R_2 - R_{21} + R_{12} \le c_2, \tag{A.28}$$

$$R_2 + R_{12} \le d_2, \tag{A.29}$$

$$R_0 + R_2 + R_{12} \le e_2, \tag{A.30}$$

$$-R_1 + R_{12} \le 0. \tag{A.31}$$

Step 2: By adding each inequality from (A.25)–(A.27) and each one from

(A.28)–(A.31), we eliminate R_{12} and obtain the following new inequalities

$$R_1 + R_2 - R_{21} \leq a_1 + c_2, \tag{A.32}$$

$$R_1 + R_2 \leq a_1 + d_2, \tag{A.33}$$

$$R_0 + R_1 + R_2 \leq a_1 + e_2, \tag{A.34}$$

$$0 \leq a_1; \tag{A.35}$$

$$R_1 + R_2 \leq c_1 + c_2, \tag{A.36}$$

$$R_1 + R_2 + R_{21} \leq c_1 + d_2, \tag{A.37}$$

$$R_0 + R_1 + R_2 + R_{21} \leq c_1 + e_2, \tag{A.38}$$

$$R_{21} \leq c_1; \tag{A.39}$$

$$R_2 - R_{21} \leq c_2, \tag{A.40}$$

$$R_2 \leq d_2, \tag{A.41}$$

$$R_0 + R_2 \leq e_2, \tag{A.42}$$

$$-R_1 \leq 0. \tag{A.43}$$

Observing that (A.35) always holds, we exclude it first. It is straightforward to verify that (A.41) is implied by (A.22), and (A.40) is implied by (A.21). We therefore also exclude both (A.41) and (A.40). We then categorize the remaining inequalities in (A.18)–(A.24) and (A.32)–(A.43) into the following three groups according to the different involvement of R_{21}:

$$R_1 \leq b_1, \tag{A.44}$$

$$R_2 \leq b_2, \tag{A.45}$$

$$R_1 + R_2 \leq a_1 + d_2, \tag{A.46}$$

$$R_1 + R_2 \leq c_1 + c_2, \tag{A.47}$$

$$R_0 + R_1 + R_2 \leq a_1 + e_2, \tag{A.48}$$

$$-R_1 \leq 0, \tag{A.49}$$

$$R_0 + R_2 \leq e_2; \tag{A.50}$$

and

$$R_2 - R_{21} \leq a_2, \tag{A.51}$$

$$-R_{21} \leq 0, \tag{A.52}$$

$$R_1 + R_2 - R_{21} \leq a_1 + c_2; \tag{A.53}$$

and

$$R_1 + R_{21} \leq d_1, \tag{A.54}$$

$$R_0 + R_1 + R_{21} \leq e_1, \tag{A.55}$$

$$-R_2 + R_{21} \leq 0; \tag{A.56}$$

$$R_1 + R_2 + R_{21} \leq c_1 + d_2, \tag{A.57}$$

$$R_0 + R_1 + R_2 + R_{21} \leq c_1 + e_2, \tag{A.58}$$

$$R_{21} \leq c_1. \tag{A.59}$$

Step 3: By adding each inequality from (A.51)–(A.53) and each one from (A.54)–(A.59), we eliminate R_{21} and obtain the following new inequalities

$$R_1 + R_2 \leq a_2 + d_1, \tag{A.60}$$

$$R_0 + R_1 + R_2 \leq a_2 + e_1, \tag{A.61}$$

$$0 \leq a_2, \tag{A.62}$$

$$R_1 + 2R_2 \leq a_2 + c_1 + d_2, \tag{A.63}$$

$$R_0 + R_1 + 2R_2 \leq a_2 + c_1 + e_2, \tag{A.64}$$

$$R_2 \leq a_2 + c_1; \tag{A.65}$$

$$R_1 \leq d_1, \tag{A.66}$$

$$R_0 + R_1 \leq e_1 \tag{A.67}$$

$$-R_2 \leq 0, \tag{A.68}$$

$$R_1 + R_2 \leq c_1 + d_2, \tag{A.69}$$

$$R_0 + R_1 + R_2 \leq c_1 + e_2, \tag{A.70}$$

$$0 \le c_1; \tag{A.71}$$

$$2R_1 + R_2 \le a_1 + c_2 + d_1, \tag{A.72}$$

$$R_0 + 2R_1 + R_2 \le a_1 + c_2 + e_1, \tag{A.73}$$

$$R_1 \le a_1 + c_2, \tag{A.74}$$

$$2R_1 + 2R_2 \le a_1 + c_2 + c_1 + d_2, \tag{A.75}$$

$$R_0 + 2R_1 + 2R_2 \le a_1 + c_2 + c_1 + e_2, \tag{A.76}$$

$$R_1 + R_2 \le a_1 + c_2 + c_1. \tag{A.77}$$

We now group (A.44)–(A.50) and (A.60)–(A.77) together, and we can observe that: i) (A.62) and (A.71) always hold, ii) (A.66) is implied by (A.44), iii) (A.69) is implied by (A.46), iv) (A.77) is implied by (A.47), v) (A.75) is implied by (A.46) and (A.47), vi) (A.70) is implied by (A.48), and vii) (A.76) is implied by (A.47) and (A.48). By removing the redundant inequalities and reordering the remaining ones, we have

$$R_1 \le b_1, \tag{A.78}$$

$$R_1 \le a_1 + c_2, \tag{A.79}$$

$$R_2 \le b_2, \tag{A.80}$$

$$R_2 \le a_2 + c_1, \tag{A.81}$$

$$R_0 + R_1 \le e_1, \tag{A.82}$$

$$R_0 + R_2 \le e_2, \tag{A.83}$$

$$R_1 + R_2 \le c_1 + c_2, \tag{A.84}$$

$$R_1 + R_2 \le a_1 + d_2, \tag{A.85}$$

$$R_0 + R_1 + R_2 \le a_1 + e_2, \tag{A.86}$$

$$R_1 + R_2 \le a_2 + d_1, \tag{A.87}$$

$$R_0 + R_1 + R_2 \le a_2 + e_1, \tag{A.88}$$

$$2R_1 + R_2 \leq a_1 + c_2 + d_1, \tag{A.89}$$

$$R_0 + 2R_1 + R_2 \leq a_1 + c_2 + e_1, \tag{A.90}$$

$$R_1 + 2R_2 \leq a_2 + c_1 + d_2, \tag{A.91}$$

$$R_0 + R_1 + 2R_2 \leq a_2 + c_1 + e_2, \tag{A.92}$$

$$-R_1 \leq 0, \tag{A.93}$$

$$-R_2 \leq 0. \tag{A.94}$$

Let $\mathcal{R}^*(p)$ denote the rate region defined by (A.78)–(A.94) for a fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, and let $\mathcal{R}^*$ be defined as

$$\mathcal{R}^* := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p).$$

Note that $\mathcal{R}^*(p)$ has two additional rate constraints (A.79) and (A.81), compared with $\mathcal{R}(p)$ (explicitly given in Corollary 2.1). We next show that both (A.79) and (A.81) are redundant by establishing the following equivalence:

$$\mathcal{R}^* = \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p) \equiv \mathcal{R} = \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p). \tag{A.95}$$

2. Equivalence between $\mathcal{R}$ and $\mathcal{R}^*$

For any fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, $\mathcal{R}^*(p)$ involves two additional rate constraints compared to $\mathcal{R}(p)$. It implies that $\mathcal{R}^*(p) \subseteq \mathcal{R}(p)$ and $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p)$. To show the equivalence, we need prove $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p)$. It is sufficient to show that for any given joint distribution $p(\cdot) \in \mathcal{P}^*$, we have $\mathcal{R}(p) \subseteq \mathcal{R}^*(p) \cup \mathcal{R}^*(p_1) \cup \mathcal{R}^*(p_2)$, where $p_1(\cdot)$ and $p_2(\cdot)$ are defined as

$$p_1(u_0, u_2, x_1, x_2, y_1, y_2) = \sum_{u_1\in\mathcal{U}_1} p(u_0, u_1, u_2, x_1, x_2, y_1, y_2),$$
$$p_2(u_0, u_1, x_1, x_2, y_1, y_2) = \sum_{u_2\in\mathcal{U}_2} p(u_0, u_1, u_2, x_1, x_2, y_1, y_2).$$

Without loss of generality, suppose that $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2)$ is a rate triple such that $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in \mathcal{R}(p)$ and $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \notin \mathcal{R}^*(p)$ due to

$$I(X_1; Y_1|U_0, U_1, U_2) + I(X_2, U_1; Y_2|U_0, U_2) < \tilde{R}_1, \tag{A.96}$$

for the same given joint distribution $p(\cdot) \in \mathcal{P}^*$.

Since $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in R(p)$, from (2.12), we have

$$\tilde{R}_1 \leq I(X_1; Y_1|U_0, U_2). \tag{A.97}$$

From (2.17) and (A.96), we obtain

$$\begin{aligned}\tilde{R}_2 &< I(X_2, U_1; Y_2|U_0) - I(X_2, U_1; Y_2|U_0, U_2)\\ &= I(U_2, X_2, U_1; Y_2|U_0) - I(X_2, U_1; Y_2|U_0, U_2)\\ &= I(U_2; Y_2|U_0)\\ &\leq I(U_2, X_2; Y_2|U_0)\\ &= I(X_2; Y_2|U_0).\end{aligned} \tag{A.98}$$

From (2.16) and (A.96), we have

$$\begin{aligned}\tilde{R}_2 &< I(X_1, U_2; Y_1|U_0, U_1) - I(X_1; Y_1|U_0, U_1, U_2)\\ &= I(U_2; Y_1|U_0, U_1)\\ &\leq I(X_1, U_2; Y_1|U_0, U_1) + I(X_2; Y_2|U_0, U_2).\end{aligned} \tag{A.99}$$

From (2.14), we immediately have

$$\tilde{R}_0 + \tilde{R}_1 \leq I(U_0, U_2, X_1; Y_1). \tag{A.100}$$

From (2.18) and (A.96), we obtain

$$
\begin{aligned}
\tilde{R}_0 + \tilde{R}_2 &< I(U_0, X_1, U_2; Y_1) - I(X_2, U_1; Y_2|U_0, U_2) \\
&= I(U_0, U_1, X_1, U_2; Y_1) - I(X_2, U_1; Y_2|U_0, U_2) \\
&= I(U_0, U_2; Y_1) \\
&\leq I(U_0, U_2, X_2; Y_1) \\
&= I(U_0, X_2; Y_1). \qquad \text{(A.101)}
\end{aligned}
$$

From (2.21) and (A.96), we obtain

$$
\begin{aligned}
\tilde{R}_1 + \tilde{R}_2 &< I(X_1, U_2; Y_1|U_0) \\
&\leq I(X_1, U_2; Y_1|U_0) + I(X_2; Y_2|U_0, U_2). \qquad \text{(A.102)}
\end{aligned}
$$

Similarly, from (2.22) and (A.96), we have

$$
\begin{aligned}
\tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 &< I(U_0, X_1, U_2; Y_1) \\
&\leq I(U_0, X_1, U_2; Y_1) + I(X_2; Y_2|U_0, U_2). \qquad \text{(A.103)}
\end{aligned}
$$

Setting $\mathcal{U}_1 = \emptyset$ in (A.78)–(A.94), we obtain $\mathcal{R}^*(p_1)$ with (R_0, R_1, R_2) satisfying

$$
\begin{aligned}
R_1 &\leq I(X_1; Y_1|U_0, U_2), \\
R_2 &\leq I(X_2; Y_2|U_0), \\
R_2 &\leq I(X_2; Y_2|U_0, U_2) + I(X_1, U_2; Y_1|U_0), \\
R_0 + R_1 &\leq I(U_0, X_1, U_2; Y_1), \\
R_0 + R_2 &\leq I(U_0, X_2; Y_2), \\
R_1 + R_2 &\leq I(X_1, U_2; Y_1|U_0) + I(X_2; Y_2|U_0, U_2), \\
R_0 + R_1 + R_2 &\leq I(U_0, X_1, U_2; Y_1) + I(X_2; Y_2|U_0, U_2).
\end{aligned}
$$

Since the rate triple $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2)$ satisfies (A.97)–(A.103), we have $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in$

$\mathcal{R}^*(p_1)$.

Similarly, if $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in \mathcal{R}(p)$ and $I(X_2; Y_2|U_0, U_1, U_2) + I(X_1, U_2; Y_1|U_0, U_1) < \tilde{R}_2$, i.e., $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \notin \mathcal{R}^*(p)$, then we have $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in \mathcal{R}^*(p_2)$.

Hence, we have $\mathcal{R}(p) \subseteq \mathcal{R}^*(p) \cup \mathcal{R}^*(p_1) \cup \mathcal{R}^*(p_2)$ and $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p)$. The equivalence is thus proven.

A.4 Proof of the Converse Part of Theorem 2.4

1. Nondeterministic Codes and Deterministic Codes

In this part, we show that for any nondeterministic (or stochastic) $(M_0, M_1, M_2, n, P_e^*)$ code for the ICC, there exists a deterministic (M_0, M_1, M_2, n, P_e) code for the same channel such that $P_e \leq P_e^*$, by applying the technique introduced in [80].

Assign each codeword $\mathbf{x}_i \in \mathcal{X}_i^n$ an index $\mu_i \in \mathcal{I}_{x_i} = \{1, 2, ..., |\mathcal{X}_i|^n\}$, $i = 1, 2$, assign each channel output sequence $\mathbf{y}_i \in \mathcal{Y}_i^n$ an index $\nu_i \in \mathcal{I}_{y_i} = \{1, 2, ..., |\mathcal{Y}_i|^n\}$, $i = 1, 2$, and assign each message pair (w_0, w_i) an index $\vartheta_i \in \mathcal{I}_{w_i} = \{1, 2, ..., M_0 M_i\}$, $i = 1, 2$.

Consider one nondeterministic $(M_0, M_1, M_2, n, P_e^*)$ code with the encoders and the decoders being defined by the following probability matrices

$$\text{Encoder } i: \ P_{E_i}(\mu_i|\vartheta_i), \ i = 1, 2,$$
$$\text{Decoder } i: \ P_{D_i}(\hat{\vartheta}_i|\nu_i), \ i = 1, 2.$$

We next show that there exist the following lists of random variables

$$A_{E_i}^{M_0 M_i} = (A_{E_i}(1), A_{E_i}(2), ..., A_{E_i}(M_0 M_i)), \ i = 1, 2,$$
$$A_{D_i}^{|\mathcal{Y}_i|^n} = (A_{D_i}(1), A_{D_i}(2), ..., A_{D_i}(|\mathcal{Y}_i|^n)), \ i = 1, 2,$$

such that the encoding and decoding functions of the nondeterministic $(M_0, M_1, M_2,$

$n, P_e^*)$ code can be expressed as

$$\mu_i = f_i(\vartheta_i, A_{E_i}(\vartheta_i)),\ i = 1, 2,$$
$$\hat{\vartheta}_i = g_i(\nu_i, A_{D_i}(\nu_i)),\ i = 1, 2.$$

Let all the elements of $A_{E_i}^{M_0 M_i}$ and $A_{D_i}^{|\mathcal{Y}_i|^n}$, $i = 1, 2$, be independent of each other and all other random variables, and uniformly distributed over the interval $[0, 1)$.

With respect to encoder 1, for each $\vartheta_1 \in \mathcal{I}_{w_i}$ and each $m \in \mathcal{I}_{x_1}$ we define

$$B_{E_1}(\vartheta_1, m) = \sum_{j=1}^{m} p(j|\vartheta_1),\ \text{and}\ B_{E_1}(\vartheta_1, 0) = 0.$$

Suppose that a message pair $(\tilde{w}_0, \tilde{w}_1)$ indexed by $\tilde{\vartheta}_1$ is to be encoded. We let $f_1(\cdot)$ output $\mu_1 = t$, if $A_{E_1}(\tilde{\vartheta}_1)$ falls into the interval $[B_{E_1}(\tilde{\vartheta}_1, t-1), B_{E_1}(\tilde{\vartheta}_1, t))$. Hence, we have

$$\Pr\Big(A_{E_1}(\tilde{\vartheta}_1) \in \Big[B_{E_1}(\tilde{\vartheta}_1, t-1), B_{E_1}(\tilde{\vartheta}_1, t)\Big)\Big) = p(t|\tilde{\vartheta}_1),$$

which means that the constructed encoding function $f_1(\cdot)$ is equivalent to the original encoding probability matrix $P_{E_1}(\mu_1|\vartheta_1)$. Similar constructions can be done for encoder 2 and the two decoders.

We define the random variable $A := (A_{E_1}^{M_0 M_1}, A_{E_2}^{M_0 M_2}, A_{D_1}^{|\mathcal{Y}_1|^n}, A_{D_2}^{|\mathcal{Y}_2|^n})$, which has a joint probability distribution $p(a)$ over range $\mathcal{A}$.

The probabilities of error in decoding the given nondeterministic $(M_0, M_1, M_2, n, P_e^*)$ code can now be expressed as

$$\begin{aligned} P_{e,i} &= \Pr((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)) \\ &= \int_{a \in \mathcal{A}} \Pr((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)|a) da,\ i = 1, 2. \end{aligned}$$

Therefore, there always exists some $a \in \mathcal{A}$ such that

$$\Pr((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)|a) \leq \max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\}, \; i = 1, 2. \tag{A.104}$$

Let $P_e = \max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\}$. From (A.104), we have a deterministic (M_0, M_1, M_2, n, P_e) code. By the definition of the $(M_0, M_1, M_2, n, P_e^*)$ code, we have

$$\max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} \leq P_e^*,$$

and thus the claim follows immediately.

2. A Converse Based on Deterministic Codes

Based on the discussion in the previous section, it suffices to show that for any deterministic $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e)$ code with $P_e \to 0$, the rate triple (R_0, R_1, R_2) must satisfy (2.38)–(2.50) for some joint distribution $p(v_0)p(x_1|v_0)p(x_2|v_0)$, to establish the converse.

Consider a deterministic $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e)$ code with $P_e \to 0$. Note that $P_e \to 0$ implies $P_{e,1}^{(n)} \to 0$ and $P_{e,2}^{(n)} \to 0$. Applying Fano-inequality [40] for decoder 1, we obtain

$$H(W_0, W_1|Y_1^n) \leq n(R_0 + R_1)P_{e,1}^{(n)} + h(P_{e,1}^{(n)}) := n\epsilon_{1n},$$

where $h(\cdot)$ is the binary entropy function. Note that $\epsilon_{1n} \to 0$ as $P_{e,1}^{(n)} \to 0$. It easily follows that

$$H(W_1|Y_1^n, W_0) \leq H(W_0, W_1|Y_1^n) \leq n\epsilon_{1n}. \tag{A.105}$$

By symmetry, we also have

$$H(W_2|Y_2^n, W_0) \leq H(W_0, W_2|Y_2^n) \leq n\epsilon_{2n}. \tag{A.106}$$

We now expand the entropy term $H(Y_1^n, V_2^n|W_0, W_1)$ as

$$\begin{aligned}
H(Y_1^n, V_2^n|W_0, W_1) &\overset{(a)}{=} H(Y_1^n, V_2^n|X_1^n, W_0, W_1) \\
&\overset{(b)}{=} H(V_2^n|X_1^n, W_0, W_1) + H(Y_1^n|V_2^n, X_1^n, W_0, W_1) \\
&\overset{(c)}{=} H(Y_1^n|X_1^n, W_0, W_1) + H(V_2^n|Y_1^n, X_1^n, W_0, W_1),
\end{aligned}$$

where (a) follows from the fact that $X_1^n = f_1(W_0, W_1)$ is a deterministic function of W_0 and W_1 for a given $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e)$ code, and both (b) and (c) are based on the chain rule. Since Y_1 is a deterministic function of X_1 and V_2, $H(Y_1^n|V_2^n, X_1^n, W_0, W_1) = 0$. Similarly, due to $V_2 = h_1(Y_1, X_1)$, we have $H(V_2^n|Y_1^n, X_1^n, W_0, W_1) = 0$. Hence, we obtain the following

$$\begin{aligned}
H(V_2^n|X_1^n, W_0, W_1) &= H(Y_1^n|X_1^n, W_0, W_1), \\
H(V_2^n|W_0, W_1) &\overset{(a)}{=} H(Y_1^n|W_0, W_1), \\
H(V_2^n|W_0) &\overset{(b)}{=} H(Y_1^n|W_0, W_1), \qquad \text{(A.107)}
\end{aligned}$$

where (a) follows from the deterministic relation between X_1^n and (W_0, W_1), and (b) follows from the conditional independence between V_2^n and W_1 given W_0. Analogously, we have

$$H(V_1^n|W_0) = H(Y_2^n|W_0, W_2). \qquad \text{(A.108)}$$

Before proceeding to the main part of the converse, we need show the following two inequalities

$$\begin{aligned}
I(W_1; Y_1^n|W_0) &\le I(W_1; Y_1^n, V_1^n|V_2^n, W_0), \qquad \text{(A.109)} \\
I(W_2; Y_2^n|W_0) &\le I(W_2; Y_2^n, V_2^n|V_1^n, W_0). \qquad \text{(A.110)}
\end{aligned}$$

Inequality (A.109) can be derived as follows:

$$
\begin{aligned}
I(W_1;Y_1^n|W_0) &= H(W_1|W_0) - H(W_1|Y_1^n,W_0) \\
&\overset{(a)}{\leq} H(W_1|V_2^n,W_0) - H(W_1|Y_1^n,V_2^n,W_0) \\
&\overset{(b)}{\leq} H(W_1|V_2^n,W_0) - H(W_1|Y_1^n,V_1^n,V_2^n,W_0) \\
&= I(W_1;Y_1^n,V_1^n|V_2^n,W_0),
\end{aligned}
$$

where (a) follows from the facts that $H(W_1|W_0) = H(W_1|V_2^n,W_0)$ which is due to the conditional independence between W_1 and V_2^n given W_0, and "conditioning reduces entropy", i.e., $H(W_1|Y_1^n,V_2^n,W_0) \leq H(W_1|Y_1^n,W_0)$, and (b) follows from "conditioning reduces entropy" as well. Similarly, we can obtain (A.110).

We now prove each of (2.38)–(2.50) by using (A.105)–(A.110).

For (2.38), we have

$$
\begin{aligned}
nR_1 &= H(W_1) = H(W_1|W_0) \\
&\overset{(a)}{=} H(W_1|W_0,V_2^n) \\
&= I(W_1;Y_1^n|W_0,V_2^n) + H(W_1|Y_1^n,W_0,V_2^n) \\
&\overset{(b)}{\leq} H(Y_1^n|W_0,V_2^n) - H(Y_1^n|W_0,W_1,V_2^n) + n\epsilon_{1n} \\
&\overset{(c)}{=} H(Y_1^n|W_0,V_2^n) + n\epsilon_{1n} \\
&\leq \sum_{i=1}^{n} H(Y_{1i}|V_{2i},W_0) + n\epsilon_{1n},
\end{aligned}
\tag{A.111}
$$

where (a) follows from the fact that W_1 and V_2^n are conditionally independent given W_0, (b) follows from $H(W_1|Y_1^n,W_0,V_2^n) \leq H(W_1|Y_1^n,W_0) \leq n\epsilon_{1n}$, and (c) follows from $H(Y_1^n|W_0,W_1,V_2^n) = H(Y_1^n|X_1^n,V_2^n,W_0,W_1) = 0$.

Analogously, for (2.39) we have

$$
nR_2 \leq \sum_{i=1}^{n} H(Y_{2i}|V_{1i}W_0) + n\epsilon_{2n}. \tag{A.112}
$$

Regarding (2.40), we have

$$
\begin{aligned}
n(R_0+R_1) &= H(W_0,W_1) \\
&= I(W_0,W_1;Y_1^n) + I(W_0,W_1|Y_1^n) \\
&\overset{(a)}{\le} I(W_0,W_1;Y_1^n) + n\epsilon_{1n} \\
&\le H(Y_1^n) + n\epsilon_{1n} \\
&\le \sum_{i=1}^{n} H(Y_{1i}) + n\epsilon_{1n},
\end{aligned} \tag{A.113}
$$

where (a) follows from (A.105).

Similarly, for (2.41) we have

$$
n(R_0+R_2) \le \sum_{i=1}^{n} H(Y_{2i}) + n\epsilon_{2n}. \tag{A.114}
$$

With respect to (2.42), we have

$$
\begin{aligned}
& n(R_1+R_2) = H(W_1) + H(W_2) \\
&= H(W_1|W_0) + H(W_2|W_0) \\
&= I(W_1;Y_1^n|W_0) + H(W_1|Y_1^n,W_0) + I(W_2;Y_2^n|W_0) + H(W_2|Y_2^n,W_0) \\
&\overset{(a)}{\le} H(Y_1^n|W_0) - H(Y_1^n|W_0,W_1) + H(Y_2^n|W_0) - H(Y_2^n|W_0,W_2) + n(\epsilon_{1n}+\epsilon_{2n}) \\
&\overset{(b)}{=} H(Y_1^n|W_0) - H(V_2^n|W_0) + H(Y_2^n|W_0) - H(V_1^n|W_0) + n(\epsilon_{1n}+\epsilon_{2n}) \\
&\le H(Y_1^n,V_1^n|W_0) - H(V_1^n|W_0) + H(Y_2^n,V_2^n|W_0) - H(V_2^n|W_0) + n(\epsilon_{1n}+\epsilon_{2n}) \\
&= H(Y_1^n|V_1^n,W_0) + H(Y_2^n|V_2^n,W_0) + n(\epsilon_{1n}+\epsilon_{2n}) \\
&\le \sum_{i=1}^{n} H(Y_{1i}|V_{1i},W_0) + \sum_{i=1}^{n} H(Y_{2i}|V_{2i},W_0) + n(\epsilon_{1n}+\epsilon_{2n}),
\end{aligned} \tag{A.115}
$$

where (a) follows from inequalities (A.105) and (A.106), and (b) follows from equalities (A.107) and (A.108).

Regarding (2.43), we have

$$
\begin{aligned}
n(R_1+R_2) &= H(W_1|W_0)+H(W_2|W_0)\\
&\overset{(a)}{\le} I(W_1;Y_1^n|W_0)+I(W_2;Y_2^n|W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(b)}{\le} I(W_1;Y_1^n|W_0)+I(W_2;Y_2^n,V_2^n|V_1^n,W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&= I(W_1;Y_1^n|W_0)+I(W_2;V_2^n|V_1^n,W_0)+I(W_2;Y_2^n|V_1^n,V_2^n,W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&\le H(Y_1^n|W_0)-H(Y_1^n|W_0W_1)+H(V_2^n|V_1^n,W_0)-H(V_2^n|V_1^n,W_2,W_0)\\
&\quad +H(Y_2^n|V_1^n,V_2^n,W_0)-H(Y_2^n|V_1^n,V_2^n,W_2,W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(c)}{=} H(Y_1^n|W_0)+H(Y_2^n|V_1^n,V_2^n,W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&\le \sum_{i=1}^n H(Y_{1i}|W_0)+\sum_{i=1}^n H(Y_{2i}|V_{1i},V_{2i},W_0)+n(\epsilon_{1n}+\epsilon_{2n}),
\end{aligned}
\tag{A.116}
$$

in which (a) follows from (A.105) and (A.106), (b) follows from (A.109), and (c) follows from $H(Y_1^n|W_0,W_1)=H(V_2^n|V_1^n,W_0)$, $H(V_2^n|V_1^n,W_2,W_0)=0$ which is because V_2^n is determined by X_2^n and X_2^n is determined by (W_0,W_2), and $H(Y_2^n|V_1^n,V_2^n,W_2,W_0)=H(Y_2^n|X_2^n,V_1^n,V_2^n,W_2,W_0)=0$.

Similarly, we have

$$
n(R_1+R_2)\le\sum_{i=1}^n H(Y_{2i}|W_0)+\sum_{i=1}^n H(Y_{1i}|V_{1i},V_{2i},W_0)+n(\epsilon_{1n}+\epsilon_{2n}),
\tag{A.117}
$$

which corresponds to (2.45).

For (2.44), we obtain

$$
\begin{aligned}
n(R_0+R_1+R_2) &= H(W_0,W_1)+H(W_2|W_0)\\
&\overset{(a)}{\le} I(W_0,W_1;Y_1^n)+I(W_2;Y_2^n|W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(b)}{\le} I(W_0,W_1;Y_1^n)+I(W_2;Y_2^n,V_2^n|V_1^n,W_0)+n(\epsilon_{1n}+\epsilon_{2n})\\
&= I(W_0,W_1;Y_1^n)+I(W_2;V_2^n|V_1^n,W_0)+I(W_2;Y_2^n|V_1^n,V_2^n,W_0)\\
&\quad +n(\epsilon_{1n}+\epsilon_{2n})
\end{aligned}
$$

$$
\begin{aligned}
&\leq H(Y_1^n) - H(Y_1^n|W_0W_1) + H(V_2^n|V_1^n, W_0) - H(V_2^n|V_1^n, W_2, W_0) \\
&\quad + H(Y_2^n|V_1^n, V_2^n, W_0) - H(Y_2^n|V_1^n, V_2^n, W_2, W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\
&\overset{(c)}{=} H(Y_1^n) + H(Y_2^n|V_1^n, V_2^n, W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\
&\leq \sum_{i=1}^{n} H(Y_{1i}) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}, V_{2i}, W_0) + n(\epsilon_{1n} + \epsilon_{2n}),
\end{aligned}
\tag{A.118}
$$

where (a), (b), and (c) follow from the same arguments for (A.116). Note that the proof for (A.118) and the one for (A.116) only differ in the first few steps, and the rest follows from the same set of arguments and procedures.

Instead of expressing $n(R_0+R_1+R_2)$ as $H(W_0, W_1)+H(W_2|W_0)$, we set $n(R_0+R_1 + R_2) = H(W_0|W_1) + H(W_0, W_2)$. Following the similar steps used in deriving (A.118), we obtain

$$
n(R_0 + R_1 + R_2) \leq \sum_{i=1}^{n} H(Y_{2i}) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}, V_{2i}, W_0) + n(\epsilon_{1n} + \epsilon_{2n}), \tag{A.119}
$$

which corresponds to (2.46).

Now for (2.47), we have

$$
\begin{aligned}
n(2R_1 + R_2) &= H(W_1|W_0) + H(W_1|W_0) + H(W_2|W_0) \\
&\overset{(a)}{\leq} I(W_1; Y_1^n|W_0) + I(W_1; Y_1^n|W_0) + I(W_2; Y_2^n|W_0) + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&\overset{(b)}{\leq} I(W_1; Y_1^n|W_0) + I(W_1; Y_1^n, V_1^n|V_2^n, W_0) + I(W_2; Y_2^n|W_0) \\
&\quad + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&= I(W_1; Y_1^n|W_0) + I(W_1; V_1^n|V_2^nW_0) + I(W_1; Y_1^n|V_1^n, V_2^n, W_0) \\
&\quad + I(W_2; Y_2^n|W_0) + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&= H(Y_1^n|W_0) - H(Y_1^n|W_0, W_1) + H(V_1^n|V_2^n, W_0) - H(V_1^n|V_2^n, W_0, W_1) \\
&\quad + H(Y_1^n|V_1^n, V_2^n, W_0) - H(Y_1^n|V_1^n, V_2^n, W_0, W_1) + H(Y_2^n|W_0) \\
&\quad - H(Y_2^n|W_0, W_2) + n(2\epsilon_{1n} + \epsilon_{2n})
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(c)}{=} H(Y_1^n|W_0) - H(Y_1^n|W_0, W_1) + H(Y_1^n|V_1^n, V_2^n, W_0) \\
&\quad + H(Y_2^n|W_0) + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&\overset{(d)}{=} H(Y_1^n|W_0) - H(V_2^n|W_0) + H(Y_1^n|V_1^n, V_2^n, W_0) \\
&\quad + H(Y_2^n|W_0) + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&\leq H(Y_1^n|W_0) - H(V_2^n|W_0) + H(Y_1^n|V_1^n, V_2^n, W_0) \\
&\quad + H(Y_2^n, V_2^n|W_0) + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&= H(Y_1^n|W_0) + H(Y_1^n|V_1^n, V_2^n, W_0) + H(Y_2^n|V_2^n, W_0) \\
&\quad + n(2\epsilon_{1n} + \epsilon_{2n}) \\
&\leq \sum_{i=1}^n H(Y_{1i}|W_0) + \sum_{i=1}^n H(Y_{1i}|V_{1i}, V_{2i}, W_0) + \sum_{i=1}^n H(Y_{2i}|V_{2i}, W_0) \\
&\quad + n(2\epsilon_{1n} + \epsilon_{2n}), \qquad \text{(A.120)}
\end{aligned}
$$

where (a) follows from (A.105) and (A.106), (b) follows from (A.109), (c) follows from the facts that $H(V_1^n|V_2^n, W_0) = H(V_1^n|W_0) = H(Y_2^n|W_0, W_2)$, $H(V_1^n|V_2^n, W_0, W_1) = H(V_1^n|X_1^n, V_2^n, W_0, W_1) = 0$, and $H(Y_1^n|V_1^n, V_2^n, W_0, W_1) = H(Y_1^n|V_1^n, X_1^n, V_2^n, W_0, W_1) = 0$, and (d) follows from $H(V_2^n|W_0) = H(Y_1^n|W_0, W_1)$. Following similar procedures, we obtain

$$
\begin{aligned}
n(R_1 + 2R_2) &\leq \sum_{i=1}^n H(Y_{2i}|W_0) + \sum_{i=1}^n H(Y_{2i}|V_{1i}, V_{2i}, W_0) + \sum_{i=1}^n H(Y_{1i}|V_{1i}, W_0) \\
&\quad + n(\epsilon_{1n} + 2\epsilon_{2n}), \qquad \text{(A.121)} \\
n(R_0 + 2R_1 + R_2) &\leq \sum_{i=1}^n H(Y_{1i}) + \sum_{i=1}^n H(Y_{1i}|V_{1i}, V_{2i}, W_0) + \sum_{i=1}^n H(Y_{2i}|V_{2i}, W_0) \\
&\quad + n(2\epsilon_{1n} + \epsilon_{2n}), \qquad \text{(A.122)} \\
n(R_0 + R_1 + 2R_2) &\leq \sum_{i=1}^n H(Y_{2i}) + \sum_{i=1}^n H(Y_{2i}|V_{1i}, V_{2i}, W_0) + \sum_{i=1}^n H(Y_{1i}|V_{1i}, W_0) \\
&\quad + n(\epsilon_{1n} + 2\epsilon_{2n}), \qquad \text{(A.123)}
\end{aligned}
$$

which correspond to (2.49), (2.48), and (2.50) respectively.

We have derived a number of inequalities (A.111)–(A.123) which upper bound the rate triple (R_0, R_1, R_2) of a given code for the DICC channel. We now adopt the technique which was used to prove the converse of the capacity region of the MACC in [7] and [56]. Define $V_0 = W_0$, equivalently $p(v_{0i}) = p(w_0)$, i.e., V_0 or V_{0i} is an auxiliary random variable uniformly distributed over the common message set $\mathcal{W}_0 = \{1, ..., M_0\}$. Since X_1 and X_2 are conditionally independent given W_0, i.e., $p(x_{1i}, x_{2i}|w_0) = p(x_{1i}|w_0)p(x_{2i}|w_0)$, we can write $p(x_{1i}, x_{2i}|v_{0i}) = p(x_{1i}|v_{0i})p(x_{2i}|v_{0i})$. Due to the introduction of V_0, the region inherits the convexity from the achievable rate region for the general ICC. We now can conclude that the rate of the given code (R_0, R_1, R_2) is upper bounded by (2.38)–(2.50) for some choice of joint distribution $p(v_0)p(x_1|v_0)p(x_2|v_0)$. This completes the proof of the converse.

Appendix B

Appendices to Chapter 5

B.1 Proof of Theorem 5.2

To show the achievability of R_{SFB1}, we develop a block Markov supposition coding scheme consisting of regular encoding and sliding window decoding. The successive transmissions consist of $B+2$ blocks, and each of n symbols. In each of the first B blocks, a message $w \in [1, 2^{nR_{\text{SFB1}}}]$ is encoded and sent to the destination with probability of error approaching 0. The average rate of transmission is thus $R_{\text{SFB1}}B/(B+2)$, which approaches R_{SFB1} as $B \to \infty$.

Let us fix a joint distribution $p(\cdot) \in \mathcal{P}_1$.

[Random Codebook Generation.] Generate three statistically independent codebooks by repeating the following procedures for three times.

1. Generate $2^{n\hat{R}_0}$ i.i.d. codewords $\mathbf{U}(i)$, $i \in [1, 2^{n\hat{R}_0}]$, according to the joint distribution $\prod_{t=1}^{n} p(u_t)$.

2. For each $\mathbf{U}(i)$, $i \in [1, 2^{n\hat{R}_0}]$, generate $2^{nR_{\text{SFB1}}}$ i.i.d. codewords $\mathbf{X}_0(i,j)$, $j \in [1, 2^{nR_{\text{SFB1}}}]$, according to $\prod_{t=1}^{n} p(x_{0,t}|u_t(i))$.

3. For each $\mathbf{U}(i)$, $i \in [1, 2^{n\hat{R}_0}]$, generate $2^{n\hat{R}_0}$ i.i.d. codewords $\mathbf{V}(i,k)$, $k \in [1, 2^{n\hat{R}_0}]$, according to $\prod_{t=1}^{n} p(v_t|u_t(i))$.

4. For each $\mathbf{V}(i,k)$, $i,k \in [1, 2^{n\hat{R}_0}]$, generate $2^{n\hat{\hat{R}}_0}$ i.i.d. codewords $\mathbf{X}_1(i,k,l)$,

$l \in [1, 2^{n\hat{\hat{R}}_0}]$, according to $\prod_{t=1}^{n} p(x_{1,t}|v_t(i,k))$.

5. For each codeword pair $(\mathbf{U}(i), \mathbf{V}(i,k))$, $i, k \in [1, 2^{n\hat{R}_0}]$, generate $2^{n\hat{R}_0}$ i.i.d. codewords $\hat{\mathbf{Y}}_1(i,k,m_1)$, $m_1 \in [1, 2^{n\hat{R}_0}]$, according to

$$\prod_{t=1}^{n} p(\hat{y}_{1,t}|u_t(i), v_t(i,k)).$$

6. For each codeword triple $(\mathbf{U}(i), \mathbf{V}(i,k), \mathbf{X_1}(i,k,l))$, $i, k \in [1, 2^{n\hat{R}_0}]$, $l \in [1, 2^{n\hat{\hat{R}}_0}]$, generate $2^{n\hat{\hat{R}}_0}$ i.i.d. codewords $\check{\mathbf{Y}}_1(i,k,l,m_2)$, $m_2 \in [1, 2^{n\hat{\hat{R}}_0}]$, according to

$$\prod_{t=1}^{n} p(\check{y}_{1,t}|u_t(i), v_t(i,k), x_{1,t}(i,k,l)).$$

[Encoding and Transmission.] We use the three codebooks in a periodic manner such that any adjacent three blocks are encoded using the three different codebooks respectively, to ensure the mutual independence of the error events among any consecutive three blocks.

Assume that at the end of the transmission of block $b-1$, a new message $w^{(b)}$ is to be transmitted by the source in block b. Further assume that the following messages are now available or have been decoded at the respective nodes:

1. At the source: $m_1^{(1)}$, $m_1^{(2)}$, ..., $m_1^{(b-3)}$; $w^{(1)}$, $w^{(2)}$, ..., $w^{(b)}$.

2. At the relay: $m_1^{(1)}$, $m_1^{(2)}$, ..., $m_1^{(b-2)}$; $m_2^{(1)}$, $m_2^{(2)}$, ..., $m_2^{(b-2)}$.

The source first needs to decode $m_1^{(b-2)}$ (equivalently , $\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_1^{(b-2)})$), the compressed version of the channel output sequence $\mathbf{y}_1^{(b-2)}$, from its received channel output sequence $\mathbf{y}_0^{(b-1)}$ in block $b-1$. To do so, it looks for an index $\hat{m}_1^{(b-2)}$ such that

$$\begin{aligned}
&(\mathbf{u}(m_1^{(b-3)}), \mathbf{x}_0(m_1^{(b-3)}, w^{(b-1)}), \mathbf{v}(m_1^{(b-3)}, \hat{m}_1^{(b-2)}), \mathbf{y}_0^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}, \quad \text{and} \\
&(\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, \hat{m}_1^{(b-2)}), \mathbf{y}_0^{(b-2)}, \mathbf{u}(m_1^{(b-4)}), \mathbf{x}_0(m_1^{(b-4)}, w^{(b-2)}), \\
&\quad \mathbf{v}(m_1^{(b-4)}, m_1^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.
\end{aligned}$$

If the index found is unique, the source declares $m_1^{(b-2)} = \hat{m}_1^{(b-2)}$; otherwise, it declares an error. The probability of error approaches 0 for sufficiently large n, when the following constraint holds:

$$\hat{R}_0 \leq I(V; Y_0|X_0, U) + I(\hat{Y}_1; Y_0, X_0|V, U). \tag{B.1}$$

If the index $m_1^{(b-2)}$ is successfully found, the source transmits the codeword $\mathbf{x}_0(m_1^{(b-2)}, w^{(b)})$ in block b; otherwise, it sends $\mathbf{x}_0(1, w^{(b)})$.

The relay first needs to apply Wyner-Ziv coding twice to compress the newly received channel output sequence $\mathbf{y}_1^{(b-1)}$ into two different versions:

$$\hat{\mathbf{y}}_1(m_1^{(b-3)}, m_1^{(b-2)}, m_1^{(b-1)}), \text{ and } \check{\mathbf{y}}_1(m_1^{(b-3)}, m_1^{(b-2)}, m_2^{(b-2)}, m_2^{(b-1)}).$$

The relay looks for an index $\hat{m}_1^{(b-1)}$ such that

$$(\hat{\mathbf{y}}_1(m_1^{(b-3)}, m_1^{(b-2)}, \hat{m}_1^{(b-1)}), \mathbf{y}_1^{(b-1)}, \mathbf{u}(m_1^{(b-3)}), \mathbf{v}(m_1^{(b-3)}, m_1^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)}.$$

If such an index is found, the relay declares $m_1^{(b-1)} = \hat{m}_1^{(b-1)}$; otherwise, an error is declared when there is no such index found. The probability of error tends to 0 for sufficiently large n, as long as

$$\hat{R}_0 \geq I(\hat{Y}_1; Y_1|V, U). \tag{B.2}$$

The relay next looks for an index $\hat{m}_2^{(b-1)}$ such that

$$\begin{aligned}&(\check{\mathbf{y}}_1(m_1^{(b-3)}, m_1^{(b-2)}, m_2^{(b-2)}, \hat{m}_2^{(b-1)}), \mathbf{y}_1^{(b-1)}, \mathbf{u}(m_1^{(b-3)}), \mathbf{v}(m_1^{(b-3)}, m_1^{(b-2)}),\\ &\quad \mathbf{x}_1(m_1^{(b-3)}, m_1^{(b-2)}, m_2^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)}.\end{aligned}$$

Similarly, if such an index is found, the relay declares $m_2^{(b-1)} = \hat{m}_2^{(b-1)}$; otherwise an error is declared. The probability of this error tends to 0 when n is sufficiently large, as long as

$$\hat{\hat{R}}_0 \geq I(\check{Y}_1; Y_1|X_1, V, U). \tag{B.3}$$

If both $m_1^{(b-1)}$ and $m_2^{(b-1)}$ are successfully determined, the relay transmits the codeword $\mathbf{x}_1(m_1^{(b-2)}, m_1^{(b-1)}, m_2^{(b-1)})$ in block b; otherwise, it sends $\mathbf{x}_1(m_1^{(b-2)}, 1, 1)$.

In Table B.2, we list all the corresponding codewords being sent in each block for the current coding scheme.

[Decoding.] Assume that the transmission of block b is just finished, and assume that the destination has successfully decoded the message indices: 1) $w^{(1)}$, $w^{(2)}$, ..., $w^{(b)}$; 2) $m_1^{(1)}$, $m_1^{(2)}$, ..., $m_1^{(b-3)}$; and 3) $m_2^{(1)}$, $m_2^{(2)}$, ..., $m_2^{(b-3)}$. To decode the message $w^{(b-2)}$, the destination first needs to find the indices $m_1^{(b-2)}$ and $m_2^{(b-2)}$. Equivalently, it needs to determine the codewords $\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_1^{(b-2)})$ and $\check{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)}, m_2^{(b-2)})$, which are the two compressed versions of $\mathbf{y}_1^{(b-2)}$, from the channel output sequences $\mathbf{y}_2^{(b-2)}$, $\mathbf{y}_2^{(b-1)}$, and $\mathbf{y}_2^{(b)}$.

The destination declares $m_1^{(b-2)} = \hat{m}_1^{(b-2)}$ if there exists a unique index $\hat{m}_1^{(b-2)}$ such that the following three events happen simultaneously

$$\begin{aligned}
&(\mathbf{u}(\hat{m}_1^{(b-2)}), \mathbf{y}_2^{(b)}) \in \mathcal{A}_\epsilon^{(n)},\\
&(\mathbf{u}(m_1^{(b-3)}), \mathbf{v}(m_1^{(b-3)}, \hat{m}_1^{(b-2)}), \mathbf{y}_2^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}, \quad \text{and}\\
&(\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, \hat{m}_2^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{u}(m_1^{(b-4)}), \mathbf{v}(m_1^{(b-4)}, m_1^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.
\end{aligned}$$

Otherwise, an error is declared. The probability of error can be shown to approach 0 when

$$\hat{R}_0 \leq I(U; Y_2) + I(V; Y_2|U) + I(\hat{Y}_1; Y_2|V, U), \tag{B.4}$$

as $n \to \infty$.

Upon finding $m_1^{(b-2)}$, the destination declares $m_2^{(b-2)} = \hat{m}_2^{(b-2)}$ if there exists a unique index $\hat{m}_2^{(b-2)}$ such that the following two events happen simultaneously

$$(\mathbf{u}(m_1^{(b-3)}), \mathbf{v}(m_1^{(b-3)}, m_1^{(b-2)}), \mathbf{x}_1(m_1^{(b-3)}, m_1^{(b-2)}, \hat{m}_2^{(b-2)}), \mathbf{y}_2^{(b-1)}) \in \mathcal{A}_\epsilon^{(n)}, \quad \text{and}$$
$$(\check{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)}, \hat{m}_2^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{u}(m_1^{(b-4)}), \mathbf{v}(m_1^{(b-4)}, m_1^{(b-3)}),$$
$$\mathbf{x}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.$$

Otherwise, an error is declared. The probability of error can be shown to approach 0 when

$$\hat{\hat{R}}_0 \leq I(X_1; Y_2|U, V) + I(\check{Y}_1; Y_2|X_1, U, V), \tag{B.5}$$

as $n \to \infty$.

Finally, the destination decodes the message $w^{(b-2)}$ from the compressed versions of the channel output sequences at the relay and its own channel output sequence: $\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_1^{(b-2)})$, $\check{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)}, m_2^{(b-2)})$, and $\mathbf{y}_2^{(b-2)}$. It declares that $w^{(b-2)} = \hat{w}^{(b-2)}$ if $\hat{w}^{(b-2)}$ is the unique message index such that

$$(\mathbf{u}(m_1^{(b-4)}), \mathbf{x}_0(m_1^{(b-4)}, \hat{w}^{(b-2)}), \mathbf{v}(m_1^{(b-4)}, m_1^{(b-3)}), \mathbf{x}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)}),$$
$$\hat{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_1^{(b-2)}), \check{\mathbf{y}}_1(m_1^{(b-4)}, m_1^{(b-3)}, m_2^{(b-3)}, m_2^{(b-2)}), \mathbf{y}_2^{(b-2)}) \in \mathcal{A}_\epsilon^{(n)};$$

otherwise, an error is declared. For sufficiently large n, the decoding error probability of this step can be readily shown to approach 0 when

$$R_{\text{SFB1}} \leq I(X_0; Y_2, \hat{Y}_1, \check{Y}_1|X_1, V, U).$$

Therefore, any rate $R_{\text{SFB1}} \leq I(X_0; Y_2, \hat{Y}_1|X_1, U)$ is achievable subject to constraints (B.1)–(B.5) for a fixed joint distribution $p(\cdot) \in \mathcal{P}_1$, and the theorem follows.

■

B.2 Proof of Theorem 5.3

A block Markov coding scheme consisting $B+2$ block transmissions is developed to achieve the rate R_{SFB2}. Regular encoding and slide window decoding are applied. Fix a joint probability distribution $p(\cdot) \in \mathcal{P}_2$.

[Codebook Generation.] Generate three statistically independent codebooks by repeating the following procedure for three times.

1. Generate 2^{nR_0} i.i.d. codewords $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, according to the joint distribution $\prod_{t=1}^{n} p(u_t)$.

2. For each $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, generate $2^{nR_{\mathrm{SFB2}}}$ i.i.d. codewords $\mathbf{X}_0(i, j)$, $j \in [1, 2^{nR_{\mathrm{SFB2}}}]$, according to $\prod_{t=1}^{n} p(x_{0,t}|u_t(i))$.

3. For each codeword $\mathbf{U}(i)$, $i \in [1, 2^{nR_0}]$, generate 2^{nR_0} i.i.d. codewords $\hat{\mathbf{y}}_0(i, k)$, $k \in [1, 2^{nR_0}]$, according to $\prod_{t=1}^{n} p(\hat{y}_{0,t}|u_t(i))$.

4. Generate $2^{nR_0'}$ i.i.d. codewords $\mathbf{X}_1(l)$, $l \in [1, 2^{nR_0'}]$, according to the joint distribution $\prod_{t=1}^{n} p(x_{1,t})$.

5. For each $\mathbf{X}_1(l)$, $l \in [1, 2^{nR_0'}]$, generate $2^{nR_0'}$ i.i.d. codewords $\hat{\mathbf{Y}}_1(l, m)$, $m \in [1, 2^{nR_0'}]$, according to $\prod_{t=1}^{n} p(\hat{y}_{1,t}|x_{1,t}(l))$.

[Encoding and Transmission.] Of the $B+2$ blocks, any three adjacent blocks are encoded using the three different codebooks in a periodic manner, to ensure the mutual independence of the error events among any consecutive three blocks.

Consider the encoding procedure at the end of the transmission of block $b-1$. At the respective nodes, we assume that the following messages are available or have been successfully decoded:

1. At the source: $w^{(1)}$, $w^{(2)}$, ..., $w^{(b)}$; $k^{(1)}$, $k^{(2)}$, ..., $k^{(b-2)}$.

2. At the relay: $m^{(1)}$, $m^{(2)}$, ..., $m^{(b-2)}$.

To encode $w^{(b)}$ and transmit the corresponding codeword, the source first needs to compress its channel output sequence $\mathbf{y}_0^{(b-1)}$ received during block $b-1$ to obtain

Table B.1: Codewords transmitted in each block to achieve R_{SFB2}.

	Block 1	Block 2	Block 3	...	Block b	...
$\mathbf{U}(i)$	$\mathbf{u}(1)$	$\mathbf{u}(k^{(1)})$	$\mathbf{u}(k^{(2)})$	...	$\mathbf{u}(k^{(b-1)})$	...
$\mathbf{X}_0(i,j)$	$\mathbf{x}_0(1,w^{(1)})$	$\mathbf{x}_0(k^{(1)},w^{(2)})$	$\mathbf{x}_0(k^{(2)},w^{(3)})$	...	$\mathbf{x}_0(k^{(b-1)},w^{(b)})$	...
$\mathbf{X}_1(l)$	$\mathbf{x}_1(2)$	$\mathbf{x}_1(m^{(1)})$	$\mathbf{x}_1(m^{(2)})$	...	$\mathbf{x}_1(m^{(b-1)})$	...
$\hat{\mathbf{Y}}_0(i,k)$	$\emptyset$	$\hat{\mathbf{y}}_0(1,k^{(1)})$	$\hat{\mathbf{y}}_0(k^{(1)},k^{(2)})$	...	$\hat{\mathbf{y}}_0(k^{(b-2)},k^{(b-1)})$	...
$\hat{\mathbf{Y}}_1(l,m)$	$\emptyset$	$\hat{\mathbf{y}}_1(2,m^{(1)})$	$\hat{\mathbf{y}}_1(m^{(1)},m^{(2)})$	...	$\hat{\mathbf{y}}_1(m^{(b-2)},m^{(b-1)})$	...

$\hat{\mathbf{y}}_0(k^{(b-2)},k^{(b-1)})$. The source looks for an index $\hat{k}^{(b-1)}$ such that

$$(\hat{\mathbf{y}}_0(k^{(b-2)},\hat{k}^{(b-1)}),\mathbf{y}_0^{(b-1)},\mathbf{u}(k^{(b-2)}))\in\mathcal{A}_\epsilon^{(n)}.$$

If such an index is found, the source declares $k^{(b-1)}=\hat{k}^{(b-1)}$, and transmits the codeword $\mathbf{x}_0(k^{(b-1)},w^{(b)})$ through n channel uses; otherwise an error is declared, and it sends the codeword $\mathbf{x}_0(1,w^{(b)})$. The probability of not being able to find such an index approaches 0 as the code length $n\to\infty$, if the following holds:

$$R_0\geq I(\hat{Y}_0;Y_0|U). \tag{B.6}$$

The relay needs to compress its received channel output sequence $\mathbf{y}_1^{(b-1)}$ as well. It looks for an index $\hat{m}^{(b-1)}$ such that

$$(\hat{\mathbf{y}}_1(m^{(b-2)},\hat{m}^{(b-1)}),\mathbf{y}_1^{(b-1)},\mathbf{x}_1(m^{(b-2)}))\in\mathcal{A}_\epsilon^{(n)}.$$

If successful, the relay declares $m^{(b-1)}=\hat{m}^{(b-1)}$ and sends $\mathbf{x}_1(m^{(b-1)})$ with n channel uses. The probability of not being able to find such an index $\hat{m}^{(b-1)}$ approaches 0 for sufficiently large n, when R_0' satisfies

$$R_0'\geq I(\hat{Y}_1;Y_1|X_0). \tag{B.7}$$

The codewords being sent in each block for this coding scheme is listed in Table B.1.

[Decoding.] At the end of transmission of block b, three-step successive sliding window decoding is applied at the destination to determine the message $w^{(b-2)}$ sent in block $b-2$ as follows. Assume that the destination has successfully decoded the following information:1) $w^{(1)}$, $w^{(2)}$, ..., $w^{(b-3)}$; 2) $k^{(1)}$, $k^{(2)}$, ..., $k^{(b-2)}$; and 3) $m^{(1)}$, $m^{(2)}$, ..., $m^{(b-3)}$.

The destination first looks for a unique index $\hat{k}^{(b-1)}$ such that

$$
\begin{aligned}
&(\mathbf{u}(\hat{k}^{(b-1)}), \mathbf{y}_2^{(b)}) \in \mathcal{A}_\epsilon^{(n)}, \text{ and} \\
&(\hat{\mathbf{y}}_0(k^{(b-2)}, \hat{k}^{(b-1)}), \mathbf{y}_2^{(b-1)}, \mathbf{u}(k^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)}.
\end{aligned}
$$

If successful, the destination declares $k^{(b-1)} = \hat{k}^{(b-1)}$; otherwise it declares an error. The probability of this decoding error can be shown to approach 0 for sufficiently large n, when the following is satisfied:

$$
R_0 \leq I(U; Y_2) + I(\hat{Y}_0; Y_2|U). \tag{B.8}
$$

The destination next looks for a unique index $\hat{m}^{(b-2)}$ such that

$$
\begin{aligned}
&(\mathbf{x}_1(\hat{m}^{(b-2)}), \mathbf{y}_2^{(b-1)}, \hat{\mathbf{y}}_0(k^{(b-2)}, k^{(b-1)}), \mathbf{u}(k^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)} \text{ and} \\
&(\hat{\mathbf{y}}_1(m^{(b-3)}, \hat{m}^{(b-2)}), \mathbf{y}_2^{(b-2)}, \hat{\mathbf{y}}_0(k^{(b-3)}, k^{(b-2)}), \mathbf{x}_1(m^{(b-3)}), \mathbf{u}(k^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.
\end{aligned}
$$

If successful, the destination declares $m^{(b-2)} = \hat{m}^{(b-2)}$, i.e., the codeword $\hat{\mathbf{y}}_1(m^{(b-3)}, \hat{m}^{(b-2)})$ is the compressed version of $\mathbf{y}_1^{(b-2)}$. Otherwise, an error is declared. As $n \to \infty$, the probability of error in this step approaches 0 when the following inequality is satisfied:

$$
R_0^{'} \leq I(X_1; Y_2, \hat{Y}_0|U) + I(\hat{Y}_1; Y_2, \hat{Y}_0, U|X_1). \tag{B.9}
$$

Lastly, the destination looks for a unique message index $\hat{w}^{(b-2)}$ such that

$$(\mathbf{x}_0(k^{(b-3)}, \hat{w}^{(b-2)}), \mathbf{y}_2^{(b-2)}, \hat{\mathbf{y}}_1(m^{(b-3)}, m^{(b-2)}), \hat{\mathbf{y}}_0(k^{(b-3)}, k^{(b-2)}), \mathbf{x}_1(m^{(b-3)}),$$
$$\mathbf{u}(k^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.$$

If such a message index is found and is unique, the destination declares $w^{(b-2)} = \hat{w}^{(b-2)}$; otherwise, an error is declared. It can be readily shown that when the information rate satisfies

$$R_{\text{SFB2}} \leq I(X_1; Y_2, \hat{Y}_1, \hat{Y}_0 | X_1, U),$$

the probability of decoding error approaches 0 as $n \to \infty$.

Therefore, any rate $R_{\text{SFB2}} \leq I(X_1; Y_2, \hat{Y}_1, \hat{Y}_0 | X_1, U)$ is achievable subject to the constraints (B.6)–(B.9), and the theorem follows. ■

B.3 Proof of Theorem 5.5

We also consider a block Markov supposition coding scheme consisting of regular encoding and sliding window decoding. The successive transmissions again consist of $B+2$ blocks, each of which has length n. In each of the first B blocks, a message $w = (w_\alpha, w_\beta)$, $w_\alpha \in [1, 2^{nR_\alpha}]$, $w_\beta \in [1, 2^{nR_\beta}]$, such that $R_\alpha + R_\beta = R_{\text{DFB2}}$, will be sent to the destination with probability of error approaching 0. The average rate of transmission is thus $R_{\text{DFB2}} B/(B+2)$, which approaches R_{DFB2} as $B \to \infty$.

We apply a random coding argument to show the achievability of R_{DFB2}. First fix a joint distribution $p(\cdot) \in \mathcal{P}_2^*$.

[Random Codebook Generation.] Generate three statistically independent codebooks by repeating the following procedures three times.

1. Generate 2^{nR_α} i.i.d. codewords $\mathbf{X}_1(i)$, $i \in [1, 2^{nR_\alpha}]$, according to $\prod_{t=1}^n p(x_{1,t})$.

2. For each $\mathbf{X}_1(i)$, $i \in [1, 2^{nR_\alpha}]$, generate 2^{nR_α} i.i.d. codewords $\mathbf{U}(i,j)$, $j \in [1, 2^{nR_\alpha}]$, according to $\prod_{t=1}^n p(u_t | x_{1,t}(i))$.

3. For each $\mathbf{U}(i,j)$, $i,j \in [1, 2^{nR_\alpha}]$, generate 2^{nR_β} i.i.d. codewords $\mathbf{X}_0(i,j,k)$, $k \in [1, 2^{nR_\beta}]$, according to $\prod_{t=1}^{n} p(x_{0,t}|v_t(i,j))$.

4. Generate 2^{nR_0} i.i.d. codewords $\mathbf{X}_2(l)$, $l \in [1, 2^{nR_0}]$, according to $\prod_{t=1}^{n} p(x_{2,t})$.

5. For each $\mathbf{X}_2(l)$, $l \in [1, 2^{nR_0}]$, generate 2^{nR_0} i.i.d. codewords $\hat{\mathbf{Y}}_2(l,m)$, $m \in [1, 2^{nR_0}]$, according to $\prod_{t=1}^{n} p(\hat{y}_{2,t}|x_{2,t}(l))$.

[Encoding and Transmission.] To ensure the mutual independence of the error event among any consecutive three blocks, the three previously generated codebooks are applied in a periodic manner such that three adjacent blocks are encoded with three independent codebooks.

Assume that at the end of the transmission of block $b-1$, a source message $w^{(b)} = (w_\alpha^{(b)}, w_\beta^{(b)})$ is to be sent in block b. The source transmits the codeword $\mathbf{x}_0(w_\alpha^{(b-2)}, w_\alpha^{(b)}, w_\beta^{(b)})$ using n channel uses.

Assume that the relay has successfully decoded: $w_\alpha^{(1)}$, $w_\alpha^{(2)}$, ..., $w_\alpha^{(b-3)}$; $m^{(1)}$, $m^{(2)}$, ..., $m^{(b-3)}$. The relay first needs to decode $m^{(b-2)}$, or equivalently $\hat{\mathbf{y}}_2(m^{(b-3)}, m^{(b-2)})$, the compressed version of the channel output sequence $\mathbf{y}_2^{(b-2)}$, from its own channel output sequences accumulated during the previous two blocks, $\mathbf{y}_1^{(b-2)}$ and $\mathbf{y}_1^{(b-1)}$. It declares $m^{(b-2)} = \hat{m}^{(b-2)}$ if $\hat{m}^{(b-2)}$ is the unique index such that the following two joint typicality are satisfied simultaneously:

$$(\mathbf{x}_2(\hat{m}^{(b-2)}), \mathbf{y}_1^{(b-1)}, \mathbf{x}_1(w_\alpha^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}, \text{ and}$$
$$(\hat{\mathbf{y}}_2(m^{(b-3)}, \hat{m}^{(b-2)}), \mathbf{y}_1^{(b-2)}, \mathbf{x}_1(w_\alpha^{(b-4)}), \mathbf{x}_2(m^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)};$$

otherwise, an error is declared. The probability of error in this step approaches 0 for sufficiently large n, if the following constraint is satisfied:

$$R_0 \leq I(X_2, Y_1|X_1) + I(\hat{Y}_2; Y_1, X_1|X_2). \tag{B.10}$$

Next, the relay determines $w_\alpha^{(b-2)} = \hat{w}_\alpha^{(b-2)}$ if $\hat{w}_\alpha^{(b-2)}$ is the unique index such that

$$(\mathbf{u}(w_\alpha^{(b-4)}, \hat{w}_\alpha^{(b-2)}), \mathbf{x}_1(w_\alpha^{(b-4)}), \mathbf{y}_1^{(b-2)}, \hat{\mathbf{y}}_2(m^{(b-3)}, m^{(b-2)}), \mathbf{x}_2(m^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)}.$$

An error is declared if no such index found or the index found is not unique. As $n \to \infty$, the probability of decoding error approaches 0 when the following is satisfied:

$$R_\alpha \leq I(U; Y_1, \hat{Y}_2 | X_1, X_2). \tag{B.11}$$

If the message $w_\alpha^{(b-2)}$ is successfully decoded, the relay sends $\mathbf{x}_1(w_\alpha^{(b-2)})$ with n channel uses in block b; otherwise $\mathbf{x}_1(1)$ is sent.

The destination performs CF on its newly received channel output sequence, $\mathbf{y}_2^{(b-1)}$. Assume that it has decoded the indices: $m^{(1)}$, $m^{(2)}$, ..., $m^{(b-2)}$. The destination first looks for some index $\hat{m}^{(b-1)}$ such that

$$(\hat{\mathbf{y}}_2(m^{(b-2)}, \hat{m}^{(b-1)}), \mathbf{y}_2^{(b-1)}, \mathbf{x}_2(m^{(b-2)})) \in \mathcal{A}_\epsilon^{(n)};$$

otherwise, an error is declared. As $n \to \infty$, the probability of finding such an index approaches 1 when the following inequality holds:

$$R_0 \geq I(\hat{Y}_2, Y_2 | X_2). \tag{B.12}$$

If one such index is found, the destination declares $\hat{\mathbf{y}}_2(m^{(b-2)}, m^{(b-1)})$ as the compressed version of $\mathbf{y}_2^{(b-1)}$ with $m^{(b-1)} = \hat{m}^{(b-1)}$, and sends $\mathbf{x}_2(m^{(b-1)})$; otherwise, $\mathbf{x}_2(1)$ is sent.

Table B.3 lists the codewords transmitted in each block using the current coding scheme.

[Decoding.] The decoding procedures at the end of the transmission of block are described in the following. Assume that the destination has the following information available or decoded: 1) $w_\alpha^{(1)}$, $w_\alpha^{(2)}$, ..., $w_\alpha^{(b-3)}$; 2) $w_\beta^{(1)}$, $w_\beta^{(2)}$, ..., $w_\beta^{(b-3)}$;

and 3) $m^{(1)}$, $m^{(2)}$, ..., $m^{(b-1)}$.

The destination first decodes $w_\alpha^{(b-2)} = \hat{w}_\alpha^{(b-2)}$ if there exists a unique index $\hat{w}_\alpha^{(b-2)}$ such that we have

$$\begin{aligned}
&(\mathbf{x}_1(\hat{w}_\alpha^{(b-2)}), \mathbf{y}_2^{(b)}, \mathbf{x}_2(m^{(b-1)})) \in \mathcal{A}_\epsilon^{(n)}, \text{ and} \\
&(\mathbf{u}(w_\alpha^{(b-4)}, \hat{w}_\alpha^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{x}_1(w_\alpha^{(b-4)}), \mathbf{x}_2(m^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)};
\end{aligned}$$

otherwise, an error is declared. As $n \to \infty$, the probability error of this step approaches 0 when the following inequality holds:

$$R_\alpha \leq I(X_1; Y_2|X_2) + I(U; Y_2|X_1, X_2). \tag{B.13}$$

The destination next decodes $w_\beta^{(b-2)} = \hat{w}_\beta^{(b-2)}$ if there exists a unique message index $\hat{w}_\beta^{(b-2)}$ such that

$$(\mathbf{x}_0(w_\alpha^{(b-4)}, w_\alpha^{(b-2)}, \hat{w}_\beta^{(b-2)}), \mathbf{u}(w_\alpha^{(b-4)}, w_\alpha^{(b-2)}), \mathbf{y}_2^{(b-2)}, \mathbf{x}_1(w_\alpha^{(b-4)}), \mathbf{x}_2(m^{(b-3)})) \in \mathcal{A}_\epsilon^{(n)};$$

otherwise, an error is declared. As $n \to \infty$, the probability error of this step approaches 0 when the following inequality is satisfied:

$$R_\beta \leq I(X_0; Y_2|X_1, X_2, U). \tag{B.14}$$

Therefore, subject to constraint (B.10) and (B.12), the sub-rates R_α and R_β satisfying (B.11), (B.13), and (B.14) are achievable for a given joint distribution $p(\cdot) \in \mathcal{P}_2^*$. The theorem follows. ■

Table B.2: Codewords transmitted in each block to achieve R_{SFB1}.

	Block 1	Block 2	Block 3	...	Block b	...
$\mathbf{U}(i)$	$\mathbf{u}(1)$	$\mathbf{u}(2)$	$\mathbf{u}(m_1^{(1)})$	...	$\mathbf{u}(m_1^{(b-2)})$	...
$\mathbf{X}_0(i,j)$	$\mathbf{x}_0(1,w^{(1)})$	$\mathbf{x}_0(2,w^{(2)})$	$\mathbf{x}_0(m_1^{(1)},w^{(3)})$	...	$\mathbf{x}_0(m_1^{(b-2)},w^{(b)})$	...
$\mathbf{V}(i,k)$	$\mathbf{v}(1,2)$	$\mathbf{v}(2,m_1^{(1)})$	$\mathbf{v}(m_1^{(1)},m_1^{(2)})$	...	$\mathbf{v}(m_1^{(b-2)},m_2^{(b-1)})$	...
$\mathbf{X}_1(i,k,l)$	$\mathbf{x}_1(1,2,3)$	$\mathbf{x}_1(2,m_1^{(1)},m_2^{(1)})$	$\mathbf{x}_1(m_1^{(1)},m_1^{(2)},m_2^{(2)})$	...	$\mathbf{x}_1(m_1^{(b-2)},m_1^{(b-1)},m_2^{(b-1)})$	...
$\hat{\mathbf{Y}}_1(i,k,m_1)$	$\emptyset$	$\hat{\mathbf{y}}_1(1,2,m_1^{(1)})$	$\hat{\mathbf{y}}_1(2,m_1^{(1)},m_1^{(2)})$	...	$\hat{\mathbf{y}}_1(m_1^{(b-3)},m_1^{(b-2)},m_1^{(b-1)})$	...
$\check{\mathbf{Y}}_1(i,k,l,m_2)$	$\emptyset$	$\check{\mathbf{y}}_1(1,2,3,m_2^{(1)})$	$\check{\mathbf{y}}_1(2,m_1^{(1)},m_2^{(1)},m_2^{(2)})$	...	$\hat{\mathbf{y}}_1(m_1^{(b-3)},m_1^{(b-2)},m_2^{(b-2)},m_2^{(b-1)})$	...

Table B.3: Codewords transmitted in each block to achieve R_{DFB2}.

	Block 1	Block 2	Block 3	...	Block b	...
$\mathbf{X}_1(i)$	$\mathbf{x}_1(1)$	$\mathbf{x}_1(3)$	$\mathbf{x}_1(w_\alpha^{(1)})$	...	$\mathbf{x}_1(w_\alpha^{(b-2)})$	...
$\mathbf{U}(i,j)$	$\mathbf{u}(1,w_\alpha^{(1)})$	$\mathbf{u}(3,w_\alpha^{(2)})$	$\mathbf{u}(w_\alpha^{(1)},w_\alpha^{(3)})$	...	$\mathbf{u}(w_\alpha^{(b-2)},w_\alpha^{(b)})$	...
$\mathbf{X}_0(i,j,k)$	$\mathbf{x}_0(1,w_\alpha^{(1)},w_\beta^{(1)})$	$\mathbf{x}_0(3,w_\alpha^{(2)},w_\beta^{(2)})$	$\mathbf{x}_0(w_\alpha^{(1)},w_\alpha^{(3)},w_\beta^{(3)})$	...	$\mathbf{x}_0(w_\alpha^{(b-2)},w_\alpha^{(b)},w_\beta^{(b)})$	...
$\mathbf{X}_2(l)$	$\mathbf{x}_2(2)$	$\mathbf{x}_2(m^{(1)})$	$\mathbf{x}_2(m^{(2)})$	...	$\mathbf{x}_2(m^{(b-1)})$	...
$\hat{\mathbf{Y}}_2(l,m)$	$\emptyset$	$\hat{\mathbf{y}}_2(2,m^{(1)})$	$\hat{\mathbf{y}}_2(m^{(1)},m^{(2)})$	...	$\hat{\mathbf{y}}_2(m^{(b-2)},m^{(b-1)})$	...

Bibliography

[1] C. E. Shannon, “Two-way communication channels,” in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, 1961, pp. 611–644.

[2] R. Ahlswede, “Multiway communication channel,” in *Proc. 2nd. Int. Symp. Information theory (Tsahkadsor, Arminian S.S.R.)*, Prague, 1971, pp. 23–52.

[3] T. M. Cover, “An achievable rate region for the broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-21, no. 4, pp. 399–404, July 1975.

[4] E. C. van der Meulen, “Three-terminal communication channels,” *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[5] R. Ahlswede, “The capacity region of a channel with two senders and two receivers,” *Annals of Probability*, vol. 2, no. 5, pp. 805–814, 1974.

[6] H. Liao, “Multiple access channels,” in *Ph.D. dissertation*, University of Hawaii, Honolulu, 1972.

[7] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell System Tech. J.*, vol. 52, pp. 1037–1076, Sept. 1973.

[8] F. M. J. Willems, “The discrete memoryless multiple access channel with partially cooperating encoders,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 441–445, 1983.

[9] L. Ozarow, “The capacity of the white Gaussian multiple access channel with feedback,” *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 623–629, July 1984.

[10] T. M. Cover and C. Leung, “An achievable rate region for the multiple-access channel with feedback,” *IEEE Trans. Inform. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[11] S. I. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 811–833, Mar. 2005.

[12] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[13] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[14] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[15] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informacii*, vol. 10, no. 3, pp. 3–14, 1974.

[16] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60– 64, Jan. 1977.

[17] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[18] G. Caire and S. Shamai (shitz), "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[19] Y. Wei and J. M. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[20] H. Weingarten, Y. Steinberg, and S. Shamai (shitz), "The capacity region of the Gaussian Multiple-Input Multiple-Output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[21] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[22] H. F. Chong, M. Motani, and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 394–401, Jan. 2007.

[23] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037 – 3063, Sept. 2005.

[24] L. L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[25] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 3, p. 536, May 1982.

[26] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 5, pp. 569–570, Sept. 1975.

[27] H. Sato, "On the capacity region of a discrete two-user channel for strong interference," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 377–379, May 1978.

[28] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[29] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.

[30] M. Costa and A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 710–711, Sept. 1987.

[31] R. Benzel, "The capacity region of a class of discrete additive degraded interference channels," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 2, pp. 228–231, Mar. 1979.

[32] A. El Gamal and M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 343–346, Mar. 1982.

[33] G. Kramer, "Genie-aided outer bounds on the capacity of interference channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 3, pp. 581–586, Mar. 2004.

[34] H. F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On a simplification of the Han-Kobayashi rate region for the interference channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[35] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[36] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad-hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[37] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, no. 82, pp. 1877–1894, Aug. 2003.

[38] L. L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[39] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, 1994.

[41] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 40–50, Mar. 2002.

[42] Y. Liang, A. S. Baruch, H. V. Poor, S. Shamai (shitz), and S. Verdu, "Cognitive interference channels with confidential messages," 2008. [Online]. Available: http://arxiv.org/PS_cache/arxiv/pdf/0710/0710.2018v1.pdf

[43] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[44] H. Sato, "Two-user communication channels," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 3, pp. 295–304, May 1977.

[45] ——, "On degraded Gaussian two-user channels," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 5, pp. 637–640, May 1978.

[46] A. B. Carleial, "Interference channels," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 1, pp. 60–70, Jan. 1978.

[47] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1345–1356, June 2004.

[48] H. F. Chong, M. Motani, and H. K. Garg, "A comparison of two achievable rate regions for the interference channel," in *Proc. ITA Workshop*, San Diego, Feb. 6–10, 2006.

[49] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," 2006. [Online]. Available: http://arxiv.org/abs/cs/0604107

[50] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.

[51] I. Maric, R. D. Yates, and G. Kramer, "The strong interference channel with unidirectional cooperation," in *Proc. UCSD Workshop on Information Theory and its Applications*, San Diego, CA, USA, Feb. 6–10, 2006.

[52] ——, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.

[53] H. H. Tan, "Two-user interference channels with correlated information sources," *Information and Control*, vol. 44, no. 1, pp. 77–104, 1980.

[54] I. Maric, R. D. Yates, and G. Kramer, "The capacity region of the strong interference channel with common information," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 30–Nov. 2, 2005, pp. 1737–1741.

[55] J. Jiang, Y. Xin, and H. K. Garg, "An achievable rate region for interference channels with common information," in *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 30–Nov. 2, 2006.

[56] F. M. J. Willems, "Information theoretical results for the discrete memoryless multiple access channel," in *Ph.D. dissertation*, Katholieke Universiteit Leuven, Belgium, Oct. 1982.

[57] K. Kobayashi and T. S. Han, "A further consideration on the HK and the CMG regions for the interference channel," in *Proc. ITA Workshop*, San Diego, Jan. 29–Feb. 2, 2007.

[58] Y. Cao, B. Chen, and J. Zhang, "A new achievable rate region for interference channels with common information," in *Proc. IEEE Wireless Communication & Networking Conference (WCNC'07)*, Hong Kong, Mar. 11–15, 2007.

[59] G. Kramer, "Review of rate regions for interference channels," in *Proc. IZS Workshop*, Zurich, Switzerland, Feb. 22–24, 2006.

[60] I. Maric, R. D. Yates, and G. Kramer, "The strong interference channel with common information," in *Proc. 43th Allerton Conference on Communications, Control and Computing*, Monticello, IL, Sept. 28–30, 2005.

[61] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[62] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with a cognitive transmitter," in *Proc. ITA Workshop*, San Diego, Jan. 29–Feb. 2, 2007.

[63] ——, "On the capacity of interference channels with a partially-cognitive transmitter," in *Proc. IEEE International Symposium On Information Theory (ISIT 2007)*, Nice, France, June 24–29, 2007.

[64] ——, "On the capacity of interference channels with one cooperating transmitter," 2007. [Online]. Available: http://arxiv.org/pdf/0710.3375

[65] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications, ser. CISM Courses and Lectures, G. Longo, Ed. New York: Springer-Verlag*, vol. 229, 1978, pp. 171–231.

[66] J. Jiang, Y. Xin, and H. K. Garg, "Interference channels with common information," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 171–187, Jan. 2008.

[67] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[68] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[69] T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 100–102, Jan. 1975.

[70] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Inform. Theory*, vol. 28, no. 1, pp. 93–95, Jan. 1982.

[71] L. Ozarow and S. Leung, "Achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 667–670, July 1984.

[72] Y. Gabbai and S. I. Bross, "Achievable rates for the discrete memoryless relay channel with partial feedback configurations," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 4989 – 5007, Nov. 2006.

[73] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1423–1438, June 2002.

[74] M. Gastpar and G. Kramer, "On noisy feedback for interference channels," in *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 30–Nov. 2, 2006.

[75] A. El Gamal, "The feedback capacity of degraded broadcast channels (corresp.)," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 379 – 381, May 1978.

[76] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 900 – 928, Mar. 2007.

[77] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. 2006 IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, Washington, July 2006.

[78] L. Xie, "Network coding and random binning for multi-user channels," in *Proc. the 10th Canadian Workshop on Information Theory*, Alberta, Canada, June 6–8, 2007.

[79] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[80] F. M. J. Williems, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

List of Publications

Journal Publications

1. J. Jiang, Y. Xin, and H. V. Poor “Achievable Rates for the Discrete Memoryless Relay Channel With Generalized Feedbackk,” *IEEE Trans. Inform. Theory*, submitted in July 2008.

2. J. Jiang, and Y. Xin, “On the achievable rate regions for interference channels with degraded message sets,” *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4707–4712, Oct. 2008.

3. Y. Xin, S. A. Mujtaba, T. Zhang, and J. Jiang, “Bypass decoding – a reduced complexity decoding technique for LDPC coded MIMO-OFDM systems,” *IEEE Trans. Vehi. Tech*, vol. 57, no. 4, pp. 2319–2333, July 2008.

4. J. Jiang, Y. Xin, and H. K. Garg, “Interference channels with common information,” *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 171–187, Jan. 2008.

Conference Publications

1. J. Jiang, and Y. Xin, “A New Achievable Rate Region for the Cognitive Radio Channel,” in *Proc. 2008 IEEE International Conference on Communications (ICC 2008)*, Beijing, China, May 19–23, 2008.

2. J. Jiang, and Y. Xin, “On the Achievable Rates for the Relay Channels With Generalized Feedback,” in *Proc. 42nd Annual CISS 2008 Conference on Information Sciences and Systems*, Princeton, NJ, March 19–21, 2008.

3. J. Jiang, Y. Xin, and H. K. Garg, “The capacity region of a class of deterministic interference channels with common information,” in *Proc. 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, April 15–20, 2007.

4. J. Jiang, Y. Xin, and H. K. Garg, “Discrete memoryless interference channels with feedback,” in *Proc. 41st Annual CISS 2007 Conference on Information Sciences and Systems*, Baltimore, MD, March 14–16, 2007.

5. J. Jiang, Y. Xin, and H. K. Garg, “An achievable rate region for interference channels with common information,” in *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 30–Nov. 2, 2006.