INFORMATION THEORETIC LIMITS OF MULTI-USER CHANNELS WITH STATE

A Dissertation in

Electrical Engineering

by

Min Li

© 2012 Min Li

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

August 2012
The dissertation of Min Li was reviewed and approved* by the following:

Aylin Yener
Professor of Electrical Engineering
Dissertation Adviser
Chair of Committee

W. Kenneth Jenkins
Professor of Electrical Engineering

David J. Miller
Professor of Electrical Engineering

Guohong Cao
Professor of Computer Science and Engineering

Osvaldo Simeone
Assistant Professor of Electrical and Computer Engineering (NJIT)

Kultegin Aydin
Professor of Electrical Engineering
Head of the Department of Electrical Engineering

*Signatures on file in the Graduate School.
Abstract

The communication channel with a random parameter called state was first introduced by Shannon and has received considerable attention in recent years because of its usefulness to model relevant phenomena occurring in wireless communication links, such as fading and interference. In many scenarios, nodes for communication, either transmitters or receivers, are informed of knowledge about the state or are capable of measuring the state. Hence, a fundamental problem is to understand methods of leveraging state information available at nodes to enhance communication reliability over channels. This dissertation makes progress along this line of study and concentrates on exploring new problems in multi-user communication setups, in order to provide fresh insights on different channel models with state from an information theoretic point of view.

Two themes are of particular interest in the dissertation. The first theme focuses on the merits of strictly causal state information at encoders. Within this theme, two models have been considered. First, multiple access channels (MACs) with independent states, each known strictly causally to one encoder, are studied. An achievable strategy that generalizes and improves upon previous studies is identified. Capacity results are found for a class of channels. Moreover, the proposed scheme is extended to state-dependent MACs with an arbitrary number of users.

Secondly, a state-dependent relay channel is studied in which strictly causal channel state information is available at the relay and no state information is available at the
source and destination. The source and the relay are connected via two unidirectional out-of-band orthogonal links of finite capacity, and a state-dependent memoryless channel connects the source and the relay, on one side, and the destination, on the other. Two achievable schemes are proposed that exploit both message and state cooperation, possible due to the orthogonal links and the availability of state information at the relay. Capacity results are identified for some special cases.

The second theme involves hop-by-hop communication channels with state. In particular, a state-dependent parallel-relay diamond channel is studied, where the source-to-relays cut is modeled with two noiseless, finite-capacity digital links with a degraded broadcasting structure, while the relays-to-destination cut is a general multiple access channel controlled by a random state. It is assumed that the source has non-causal channel state information and the relays have no state information. In this model, first, the capacity is characterized for the case where the destination has access to the state sequence. It is demonstrated that in this case, a joint message and state transmission scheme via binning is optimal. Next, the case with state information at the source only is considered. For this scenario, lower and upper bounds on the capacity are derived for the general discrete memoryless model.

In general, this dissertation points to the advantages of strictly causal state information in multi-user communication channels. The results are in contrast to the single-user channel, in which strictly causal state information is not beneficial. Moreover, this dissertation provides methods of leveraging state information available at the source node in hop-by-hop communication channels.
# Table of Contents

List of Tables ................................................... ix

List of Figures ................................................... x

Acknowledgments ................................................. xiii

Chapter 1. Introduction ........................................ 1
  1.1 Communication Channels with State .................... 1
  1.2 Existing Work ........................................... 3
  1.3 Focus of the Dissertation ................................ 6
  1.4 Contributions of the Dissertation ....................... 6
    1.4.1 Multiple-Access Channels with Strictly Causal State Information .......................... 7
    1.4.2 Message and State Cooperation in a Relay Channel .......................... 8
    1.4.3 The State-Dependent Degraded Broadcast Diamond Channel .......................... 9
  1.5 Dissertation Road Map .................................. 10
  1.6 Notation .................................................. 10

Chapter 2. Basic Definitions and Fundamental Tools .......... 11
  2.1 Entropy and Mutual Information ......................... 11
  2.2 Joint Typicality and Useful Lemmas ...................... 13
  2.3 Gel’fand-Pinsker Binning and Dirty-Paper Coding ........ 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Basic Digital Relaying Strategies</td>
<td>19</td>
</tr>
<tr>
<td>2.5 Chapter Summary</td>
<td>21</td>
</tr>
<tr>
<td>Chapter 3. Multiple Access Channels with Strictly Causal State Information</td>
<td>22</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>22</td>
</tr>
<tr>
<td>3.1.1 Contributions</td>
<td>24</td>
</tr>
<tr>
<td>3.2 System Model and Preliminaries</td>
<td>25</td>
</tr>
<tr>
<td>3.2.1 System Model</td>
<td>25</td>
</tr>
<tr>
<td>3.2.2 Preliminaries</td>
<td>27</td>
</tr>
<tr>
<td>3.3 A New Achievable Rate Region</td>
<td>29</td>
</tr>
<tr>
<td>3.4 Capacity Result</td>
<td>33</td>
</tr>
<tr>
<td>3.5 Generalization to $M$ Users with Independent States</td>
<td>36</td>
</tr>
<tr>
<td>3.6 Introducing Output Feedback</td>
<td>38</td>
</tr>
<tr>
<td>3.7 Chapter Summary</td>
<td>41</td>
</tr>
<tr>
<td>3.8 Appendices</td>
<td>42</td>
</tr>
<tr>
<td>3.8.1 Proof of Theorem 3.4</td>
<td>42</td>
</tr>
<tr>
<td>3.8.2 Proof of Theorem 3.5</td>
<td>44</td>
</tr>
<tr>
<td>3.8.3 Proof of Theorem 3.6</td>
<td>53</td>
</tr>
<tr>
<td>Chapter 4. Message and State Cooperation in a Relay Channel</td>
<td>57</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>57</td>
</tr>
<tr>
<td>4.2 System Model</td>
<td>61</td>
</tr>
<tr>
<td>4.3 Achievable Schemes and Upper Bound</td>
<td>64</td>
</tr>
<tr>
<td>4.3.1 Scheme 1: Block-based Message and State Cooperation</td>
<td>64</td>
</tr>
</tbody>
</table>
4.3.2 Scheme 2: Burst Message Cooperation and Block-based State Cooperation .................................. 67

4.3.3 Comparison of Achievable Rates ................................................. 70

4.3.4 An Upper Bound ................................................................. 71

4.4 Special Cases and Capacity Results .............................................. 75

4.4.1 No Message and State Cooperation .......................................... 75

4.4.2 Message Cooperation Only .................................................... 80

4.4.3 State Cooperation Only ....................................................... 82

4.5 Cooperation Strategies With Total Conferencing Capacity Fixed ... 84

4.6 Gaussian Model ................................................................. 85

4.6.1 Achievable Rate ............................................................... 86

4.6.2 Special Cases and Capacity Results ........................................ 89

4.6.3 Numerical Results and Discussions ........................................ 94

4.7 Chapter Summary ............................................................... 98

4.8 Appendices ................................................................. 99

4.8.1 Proof of Proposition 4.1 .................................................... 99

4.8.2 Proof of Proposition 4.2 .................................................... 108

Chapter 5. The State-Dependent Degraded Broadcast Diamond Channel ............................ 119

5.1 Introduction ................................................................. 119

5.1.1 Background and Related Work ............................................ 121

5.1.2 Contributions ............................................................. 123

5.2 System Model and Main Definitions .......................................... 124
5.3 Non-causal CSIT and CSIR ........................................... 127
  5.3.1 Capacity Result ............................................... 127
  5.3.2 The Suboptimality of Separate Message-State Transmission . 129
5.4 Non-causal CSIT and No CSIR ................................. 134
  5.4.1 An Upper Bound ............................................. 134
  5.4.2 Achievable Scheme 1: GP Coding With Quantized States At
      The Relays .................................................... 135
  5.4.3 Achievable Scheme 2: Quantized GP Coding ............... 138
  5.4.4 Gaussian SD-DBDC ......................................... 140
      5.4.4.1 Reference Results .................................. 140
      5.4.4.2 Achievable Rates ................................... 142
      5.4.4.3 Numerical Results .................................. 144
5.5 Chapter Summary ................................................. 146
5.6 Appendices ....................................................... 147
  5.6.1 Proof of Theorem 5.1 ....................................... 147
  5.6.2 Proof of Proposition 5.4 ................................... 156
  5.6.3 Proof of Proposition 5.5 ................................... 158

Chapter 6. Conclusion ................................................. 160
  6.1 Dissertation Summary ......................................... 160
  6.2 Future Research ................................................ 162

Bibliography ............................................................... 164
List of Tables

1.1 Summary of research on the multi-user state-dependent channels with various forms of encoder state information and no decoder state information. ...................................................... 5
List of Figures

2.1 The discrete memoryless single-user channel with random state non-causally known to the encoder. .................................................. 18

3.1 The $M$-user state-dependent MAC with $M$ mutually independent states, each of which is available to its corresponding encoder in a strictly causal manner. ................................................................. 27

3.2 Capacity region for the binary modulo-additive state-dependent MAC with input constraints considered in Remark 3.5 ($p_1 = p_2 = 1/3, p_s = 1/4$). ................................................................. 37

3.3 The state-dependent MAC with strictly causal state information at TX1 and output feedback at TX2. ........................................ 39

3.4 Comparison of different capacity regions. ............................... 41

4.1 The downlink transmission to a home user in a femtocell provides an example application of the considered model illustrated in Fig. 4.2. The home base station is assumed to be able to measure the interference from outdoor users. ................................................................. 60

4.2 A state-dependent relay channel with two unidirectional out-of-band orthogonal links. ................................................................. 61

4.3 Scheme 1: Block-based message and state cooperation. .......... 66

4.4 Scheme 2: Burst message cooperation and block-based state cooperation. 69
4.5 Comparison of achievable rates between scheme 1 and scheme 2 for message cooperation only \((C_{RS} = 0, P = P_R = P_S = 1, N_0 = 0)\). ................................ 95

4.6 Comparison of achievable rates between scheme 1 and scheme 2 for state cooperation only \((C_{SR} = 0, P = P_R = P_S = 1, N_0 = 0)\). .............................. 96

4.7 Comparison of achievable rates for message cooperation only by scheme 2
\((C_{SR} = \{0.2, 0.5, 0.8, 1.2\}, C_{RS} = 0, P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0) (dB))\). 97

4.8 Comparison of achievable rates for state cooperation only by scheme
\(2 \ (C_{SR} = 0, C_{RS} = \{0, 0.2, 0.5, 0.8, 100\}, P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0) (dB))\). .............................. 98

4.9 Comparison of achievable rates for different cooperation strategies when the total conferencing capacity is fixed \((C_{sum} = 1, P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0) (dB))\). .............................. 99

5.1 A state-dependent degraded broadcast diamond channel (SD-DBDC) with non-causal channel state information (CSI) at the transmitter (CSIT) and with or without CSI at the receiver (CSIR). The CSIR switch is closed or open, respectively. .............................. 125

5.2 Performance comparison between \(C, R_{\text{separate}}, \) and \(R_{\text{pure-message}}\) for \(C_2 = 0.5, \) and \(p_{x_2} = 0.1 \) or 0.3 in the binary example of Section 5.3.2. 132

5.3 A Gaussian SD-DBDC with an additive Gaussian state. ...................... 141

5.4 Achievable rates \(R\) vs. SNR for \(C_1 = 1.5, C_2 = 1, P_1 = P_2 = 1, P_S = 0.2 \) or 0.4. ...................... 145
5.5 Achievable rates $R$ vs. SNR for $C_1 = 1.5, C_2 = 1, P_1 = P_2 = 1, P_S = 0.8$ or $1.2$. .......................................................... 146

5.6 Achievable rates $R$ vs. $C_1$ for $C_2 = 1, P_1 = P_2 = 1, N_0 = 0.1, P_S = 1.2$. 147

5.7 Achievable rates $R$ vs. $P_S$ for $C_1 = 1.5, C_2 = 1, P_1 = P_2 = 1, N_0 = 0.1$. 148
Acknowledgments

I would like to first thank my advisor, Prof. Aylin Yener, for her support, guidance and encouragements to my research. Without those, I would have never completed this dissertation. Thanks to her for introducing me to the area of network information theory and allowing me ample freedom to choose any research topic I am interested in. Thanks to her for teaching me how to be a professional researcher. Her principles and philosophy on research will definitely make a great impact on my future career.

I would like to thank my collaborator, Prof. Osvaldo Simeone, from New Jersey Institute of Technology (NJIT). I was so lucky to know such an excellent leading expert in my filed and collaborate with him for the past two years. I have enjoyed and benefited a lot from numerous technical discussions with him. Without his patience and endless help, I would have never made those accomplishments in this dissertation.

I would also like to thank Prof. W. Kenneth Jenkins, Prof. David J. Miller and Prof. Guohong Cao for serving in my Ph.D. dissertation committee and for their valuable suggestions from the initial proposal to this dissertation. My special thanks go to Prof. David J. Miller, for his efforts to carefully review the dissertation and bring up lots of constructive comments, which greatly improve the technical quality and the overall presentation of this dissertation.

I am also greatly indebted to my colleagues at WCAN including Min Chen, Xiang He, Satashu Goel, Igor Stanojev, Ertugrul Necdet Ciftcioglu, Ye Tian, Kaya Tutuncuoglu, Basak Guler and Burak Varan, who selflessly shared their knowledge with me.
I especially thank Xiang for his encouragements to my research. Special thanks also go to Ye for lots of helpful conversations and technical discussions on the blackboard.

Many thanks go to my friends at Penn State. I was so lucky to meet three of my best friends, Wenli Hu, Wei Zhong and Jingyuan Liu here. All those happy moments with you are playing back in my mind and continuing forever. You really turn my long Ph.D. journey to a short and sweet one. Thanks to all the EEers including Guixi Zou, Junyu Yao, Jue Li, Zheqian Zhang, Xiaonan Zang, Tingting Lu, Suran Cui, Xue Dong, Weizhi Zhang, Zhou Zhou, Lei Mei, Zhihao Jiang, Yuda Zhou, Yaohui Gao, Jing Jiang and Shuo Zhao, for helpful discussions and nice moments we shared. I would also like to thank my friends, Haoting Shen, Yao Luo, Xizhen Cai, Wei Luo, Ying Ye, Wei Wang, Xu Xu and many others, with whom I have had a very wonderful time at Penn State.

Lastly, I would like to thank my family, including my mother, father, sister, brother-in-law, uncle and aunt, for their sacrifice and endless support for my Ph.D. study. Thanks to my beloved mother for her dedicated and patient care of my father when he was in severe condition, which made me greatly relieved and be able to concentrate on the completion of my degree. Thanks to my father for showing me his support even though now in a silent manner. Special thanks also go to my girlfriend, Wenli Hu, for her understanding and encouragements all the time. We have shared everything from joy to sorrow for the past few years. Her timely encouragements along the road boost me to move forward and get to the final stage of my study.
1.1 Communication Channels with State

In a wireless network, the main challenges of providing reliable communications include fading [1, 2] and interference [3]. Fading is mainly due to multiple signal paths and user mobility during the communication process and normally manifests itself as a multiplicative signal attenuation. In addition, interference is a consequence of the broadcast nature unique to the wireless medium, and it represents an additive corruption of the signal besides that created by noises. Both fading and interference induce time variations for communication channels. To capture the time-varying phenomena and establish the fundamental performance limits of such channels, a useful modeling approach is to assume that at each time instant, the underlying channel is affected by a random parameter called state, which is generated according to a fixed distribution. This approach was initiated by Shannon in [4] and has received considerable attention since then. The corresponding channel, whose input-output relationship depends also on the state sequence, is conventionally named the state-dependent channel.

State-dependent channels are usually classified on the basis of the availability of state information to encoders and decoders. For decoders, it suffices to distinguish between the availability of state information or not, since decoders usually wait until the end of the complete reception to decode message. However, encoders may have no state
information, or else be informed of the state sequence in a non-causal, causal, or strictly-causal manner. Specifically, non-causal state information means that the encoders know information about the entire state sequence $s^n$ before encoding for the current block of transmission. Causal state information refers to the case in which at channel use $i$, the encoders know information about all states up to and including time instant $i$. Furthermore, strictly causal state information refers to the case in which at channel use $i$, the encoders know information about all states up to, but excluding time instant $i$.

In general, the state information can be in forms of a perfect state sequence, a noisy state sequence or a coded state sequence. This study primarily considers the state information to be the perfect state sequence similar to Shannon’s work [4]. Of note are several scenarios in which acquisition of the state sequence at encoders or decoders is possible. For instance, when the state represents fading, measurement of the induced channel variation occurs at the receiver via training signals [2], leading to the availability of state information at the decoder. Moreover, the encoder can be informed of such state information if some dedicated feedback links exist [2]. Instead, when the state models interference created by other users’ transmissions, the state can be more effectively measured at nodes that are in the vicinity of the interferers, while nodes further away cannot directly measure the state. In this case, it may happen that nodes that are currently serving as transmitters (encoders) may acquire state information, while the respective receivers (decoders) may not. As another example, under the cognitive radio paradigm, the “smart” cognitive-type device is capable of sensing the primary (licensed) transmission and exploits it as encoder side information to facilitate its own message transmission, see, e.g., [5, 6].
In the channels with state, given the state information available at the encoders and/or the decoders, the arising fundamental problem is to understand methods of leveraging such state information to enhance communication reliability. In general, when the decoder has access to the state information, it can view such information as a component of the channel output and use it to perform joint decoding. However, lack of decoder-side information changes the problem drastically, leading to the significant challenge of leveraging the information only available at the encoder to facilitate message transmission.

1.2 Existing Work

For the single-user state-dependent channels, much is understood on how to leverage the state information in an optimal way from an information theoretic point of view. Assuming that the state sequence is independent and identically distributed (i.i.d.) and that no state information exists at the decoder. Shannon [4] studied the general discrete memoryless (DM) single-user channel with causal state information at the encoder and proposed the capacity-achieving scheme, the Shannon strategy. Gel’fand and Pinsker [7] instead studied the DM channel with non-causal state information and invented the optimal Gel’fand-Pinsker (GP) binning scheme. Later, Costa [8] studied an instance of the GP model with an additive Gaussian state and proved that the capacity of the channel is the same as if no interfered state is present. Costa’s remarkable result arises from a particular form of GP binning in the Gaussian setup, the Dirty-Paper Coding (DPC). While the previous research demonstrates the benefits of causal and non-causal state information, provably, strictly causal state information is instead useless in the DM
channel. The proof follows from the arguments in [9], which show that output feedback does not increase the capacity of a single-user DM channel. In addition, other research [10, 11, 12, 13] studied the channels with state information at the decoder besides the encoder-side information. In particular, Caire and Shamai [13] showed that a simple coding scheme constructed directly over the input alphabet is optimal, when the causal state information at the encoder is a deterministic function of the information at the decoder. Focusing on the fading channel, Goldsmith and Varaiya [12] found the capacity for the case with causal state information at the encoder and decoder. The optimal transmission scheme proposed uses a strategy of water-pouring in time for power adaptation and a variable-rate multiplexed coding scheme.

In contrast, less conclusive capacity results have been developed in the multi-user communication environment, in particular when the state information is absent from the decoder. One may envision that the previously mentioned coding schemes for the single-user channel, such as the Shannon strategy and GP coding, can be extended to the multi-user setup. However, such direct extensions often only attain suboptimal solutions. The challenges stemming from the multi-user environment include not only the distributed location of transmitting nodes but also the potentially distinguished state information available at the nodes. Several groups of researchers have made attempts to understand instances of the state-dependent multi-user channels. A thorough review of the existing results appears in [14]. Here, we point to Table 1.1 for the research with focus on multiple access channels (MACs), broadcast channels (BCs) and relay channels with various forms of encoder state information and no decoder state information. To the best available knowledge, in the two-user MAC setup, the complete capacity region
is only found for the following instances: i) the state-dependent cooperative MAC, in which one encoder has both non-causal state information and the other encoder’s message [15]; ii) the binary MAC with independent binary states, each known non-causally to its encoder [16]; and iii) the Gaussian MAC with an additive state known non-causally to both encoders [17]. In addition, for the two-user BC, with the assumption of a degraded channel, the capacity region is found for: i) the case with causal state information at the encoder only; and ii) the case with causal or non-causal state information at both the encoder and the nondegraded receiver [18]. Discarding the assumption of degraded channel, recent research [19] identified the capacity region for a class of semi-deterministic BCs with non-causal state information at the encoder.

Table 1.1. Summary of research on the multi-user state-dependent channels with various forms of encoder state information and no decoder state information.

<table>
<thead>
<tr>
<th>Models/State Information</th>
<th>Non-causal</th>
<th>Causal</th>
<th>Strictly Causal</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>[17, 20, 21, 22]</td>
<td>[15, 16, 25, 26, 27]</td>
<td>[21, 23, 24, 28]</td>
</tr>
<tr>
<td>BC</td>
<td>[18, 19, 29]</td>
<td>[18]</td>
<td>N/A</td>
</tr>
<tr>
<td>Relay</td>
<td>[28, 30, 31]</td>
<td>[32, 33]</td>
<td>this work</td>
</tr>
<tr>
<td>Hop-by-hop Network</td>
<td>[34], this work</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
1.3 Focus of the Dissertation

In light of the general background, in this dissertation, the research seeks to pursue some new problems in the multi-user setup and provide fresh insights for the fundamental limits of different state-dependent channel models.

On one hand, as indicated by Table 1.1, most of the existing work focused on discovering the merits of non-causal or causal state information. Prior to the very recent research [23, 24], none explored the merit of strictly causal state information. The reason why this line of study is ignored may be largely due to the common belief that strictly causal state information is useless as seen in the single-user channel. As demonstrated by [23, 24], however, this thought is incorrect in a state-dependent MAC. Motivated by [23, 24], we make further attempts along this direction. To that end, we first revisit the state-dependent MAC model of [23, 24] and propose a new transmission scheme to leverage the strictly causal state information. We also investigate the state-dependent relay channel, in which the relay node may have state information in a strictly causal manner. On the other hand, we advance from one-hop setups in the existing studies to hop-by-hop communication channels, aiming to provide guidance for leveraging the state information available at the source node to facilitate message delivery.

1.4 Contributions of the Dissertation

In this dissertation, we have investigated a number of state-dependent channels, which include the state-dependent multiple-access channels, single-relay channels and parallel-relay diamond channels. Our main contributions are summarized as follows.
1.4.1 Multiple-Access Channels with Strictly Causal State Information

First, we study the multiple access channel with independent states, each known strictly causally to one transmitter. This model finds application in several practical systems, such as the uplink of a cellular network. In the model, the states may represent multiple interference signals that affect the destination. Each transmitter is able to measure one interference signal created by a co-located interferer in a strictly causal manner. The interesting issues involve identifying the usefulness of strictly causal state information at the transmitters and the possibility of leveraging such information.

For the two-user case, it was recently shown by Lapidoth and Steinberg [23, 24] that strictly causal state information can be beneficial in enlarging the capacity region. The basic idea is to allow the encoders to send compressed past state information to the decoder. In our first contribution [35, 36], we propose a generalization of the said strategy whereby the encoders compress also the past transmitted codewords along with the past state sequences. The proposed scheme uses a combination of long-message encoding, compression of the past state sequences and codewords without binning, and joint decoding over all transmission blocks. The scheme proposed improves upon the original one appearing in [24]. Capacity results are derived for a class of channels. Moreover, the proposed scheme is extended to state-dependent MACs with arbitrary number of users. In general, it is seen that if the state sequences observed by users are mutually independent, then conveying state information from one user only improves the other users’ message rates but brings no gain for its own rate. In contrast, if a common
state sequence is available to all the users to allow cooperative state conveying, then all users’ message rates gain a boost simultaneously.

1.4.2 Message and State Cooperation in a Relay Channel

As the second contribution [37, 38], we study a relay channel, in which the source and the relay are connected via two unidirectional out-of-band orthogonal links of finite capacity, and a state-dependent memoryless channel connects the source and the relay, on one side, and the destination, on the other. It is assumed that the relay may obtain the state sequence in a strictly causal manner, but neither the source nor the destination is aware of the state. This model fits in the downlink transmission of a femtocell network, where a micro-base station coordinates with a home base station via dedicated fiber cable for communication to a home user over the wireless channel [39, 3]. In the system, the home base station serves as a relay and aims to provide better message delivery for the home user. The state represents an interference signal created by outdoor users, and the home base station may be able to measure it in a strictly causal manner.

In such a model, via the orthogonal links, the source can convey information about the message to be delivered to the destination to the relay, while the relay can forward state information to the source. This exchange enables cooperation between the source and the relay on transmission of message and state information to the destination. We propose two transmission schemes that exploit both message and state cooperation as described. In several special cases, we find the capacity. We also study a Gaussian model with an additive Gaussian state, by which we demonstrate the relative merits of message and state cooperation. Specifically, when the orthogonal links are unidirectional
channels with fixed capacity, e.g., cables, we demonstrate that state cooperation provides further rate improvement besides that contributed by message cooperation. Conversely, assuming that the overall orthogonal-link resources are fixed and can be arbitrarily allocated between the source-to-relay and relay-to-source links, we prove that the optimum is to devote all the resources to the former, thereby leading to message cooperation only.

1.4.3 The State-Dependent Degraded Broadcast Diamond Channel

In our third contribution [40], we investigate a state-dependent parallel-relay diamond channel, where the source-to-relays cut is modeled with two noiseless, finite-capacity digital links with a degraded broadcasting structure, while the relays-to-destination cut is a general multiple access channel controlled by a random state. It is assumed that the source has non-causal channel state information, the relays have no state information and the destination may or may not have state information. This model may represent a distributed antenna system, where a central unit controls two antennas via backhaul links for communication to an active user over a wireless channel [41, 3]. In the system, the state models the fading coefficients for the MAC between the distributed antennas and the user, or an interference signal affecting this MAC.

For the model considered, we establish the capacity for the case in which both the source and destination have non-causal access to the state sequence. We demonstrate that a joint message and state transmission scheme via binning is optimal. We also propose lower and upper bounds on the capacity for the case with non-causal state information at the source only. Performance advantages accrued from the source-side information are further illustrated via a Gaussian example with an additive state.
1.5 Dissertation Road Map

The remainder of the dissertation is organized as follows. Some basic definitions and fundamental tools used throughout the research appear in Chapter 2. Chapter 3 focuses on the multiple-access channels with strictly causal state information. Chapter 4 focuses on a state-dependent relay channel. Chapter 5 studies the state-dependent degraded broadcast diamond channel. Chapter 6 concludes the research and provides suggestions directing future investigations.

1.6 Notation

Throughout the dissertation, we use the following notation.

- $p_X(x)$ or $p(x)$: the probability of random variable $X = x$
- $p_{Y|X}(y|x)$ or $p(y|x)$: the conditional probability of random variable $Y = y$ given $X = x$
- $x^i_k$: a vector $[x_{k,1}, ..., x_{k,i}]$
- $\mathbb{R}_i^+$: the set of non-negative real vectors in $i$ dimensions
- $[1 : L]$: the set of integers $\{1, ..., L\}$
- $\lceil . \rceil$: the ceiling function
- $[1 : 2^l]$: the set of integers $\{1, ..., \lceil 2^l \rceil \}$ for a positive real number $l$
- $\mathcal{N}(0, \sigma^2)$: a zero-mean Gaussian distribution with variance $\sigma^2$
- $C(x)$: Gaussian capacity function $\frac{1}{2} \log_2(1 + x)$
- $\mathbb{E}[X]$: the expectation of random variable $X$
- $p_1 \ast p_2$: the discrete convolution $p_1 \ast p_2 = p_1 \ast (1 - p_2) + p_2 \ast (1 - p_1)$
- $H_b(p)$: the binary entropy function $H_b(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$
Chapter 2

Basic Definitions and Fundamental Tools

This chapter aims to provide some basic notions in information theory and introduce some fundamental tools we use throughout the dissertation. First, the notions of entropy and mutual information are defined for both the discrete and continuous random variables. Next, the notions of typicality and joint typicality are introduced. Important lemmas on jointly typical sequences are listed including the conditional joint typicality lemma, the joint typicality lemma, the packing lemma and the covering lemma. These lemmas as well as the aforementioned entropy and mutual information play a vital role in the proofs of the results in this research. Moreover, the Gel’fand-Pinsker (GP) binning and Dirty-Paper Coding (DPC) schemes are briefly reviewed. Lastly, several relaying schemes are summarized. These coding schemes serve as fundamental tools we leverage in this research.

2.1 Entropy and Mutual Information

Let \( X \) be a discrete random variable with finite alphabet \( \mathcal{X} \) and be associated with probability mass function (PMF) \( p_X(x) \). The entropy of random variable \( X \) is defined as

\[
H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) = -\mathbb{E}_X \left[ \log p(X) \right].
\]  

(2.1)
This quantity measures the uncertainty of the outcome of random variable $X$. Similarly, let $(X, Y) \sim p(x, y)$ be a pair of random variables. The joint entropy of $X$ and $Y$ is defined as

$$H(X, Y) = -\mathbb{E}_{X,Y} \left[ \log p(X, Y) \right]. \quad (2.2)$$

Moreover, let $X \sim F(X)$ be an arbitrary random variable and $Y \mid \{X = x\} \sim p(y \mid x)$ be discrete for every $x$. Then the conditional entropy $H(Y \mid X)$ of $Y$ given $X$ is defined as

$$H(Y \mid X) = -\mathbb{E}_{X,Y} \left[ \log p(Y \mid X) \right]. \quad (2.3)$$

The above entropy-related definitions can be generalized to the continuous random variables. For example, let $X$ be a continuous random variable with probability density function (PDF) $f(x)$. Then the differential entropy of $X$ is defined as

$$h(X) = -\int f(x) \log f(x) \, dx = -\mathbb{E}_{X} \left[ \log f(X) \right]. \quad (2.4)$$

Similarly, the joint differential entropy $h(X, Y)$ of $X$ and $Y$ and the conditional differential entropy $h(Y \mid X)$ of $Y$ given $X$ are defined as

$$h(X, Y) = -\mathbb{E}_{X,Y} \left[ \log f(X, Y) \right], \quad (2.5a)$$

and

$$h(Y \mid X) = -\mathbb{E}_{X,Y} \left[ \log f(Y \mid X) \right], \quad (2.5b)$$

respectively, where $f(x, y)$ represents the joint PDF of variables $X$ and $Y$, and $f(y \mid x)$ represents the conditional PDF of variable $Y$ given $X = x$. 
Another important notion in information theory is mutual information. Let 
\((X, Y) \sim p(x, y)\) be a pair of discrete random variables. The mutual information mea-
sures the information about \(X\) obtained from observation \(Y\), and it is defined as

\[
I(X; Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{2.6}
\]

From the definition, we observe that the mutual information can be interpreted as the
relative entropy between the joint and product measures of \(X\) and \(Y\). The definition
of mutual information can be generalized to the case of continuous random variables by
calculating the measure with respect to their joint PDF. Details are omitted for the sake
of conciseness.

It is finally remarked that the mutual information \(I(X; Y)\) is a nonnegative func-
tion of \(p(x, y)\) and it equals zero if and only if \(X\) and \(Y\) are independent. Moreover,
it is concave in \(p(x)\) for a fixed \(p(y|x)\), and convex in \(p(y|x)\) for a fixed \(p(x)\). These
properties are useful in the proofs of many channel and source coding problems.

\[2.2\]  Joint Typicality and Useful Lemmas

We now introduce the notions of typicality, joint typicality and some lemmas
relevant to them. In particular, we use the definitions as in reference [42].

Let \(X\) be a discrete random variable with finite alphabet \(\mathcal{X}\) and distributed as
\(p(x)\). Let \(x^n\) be a sequence of length \(n\) whose elements are drawn from \(\mathcal{X}\). Define the
empirical PMF of $x^n$ as

$$\pi (x \mid x^n) = \frac{|\{i : x_i = x\}|}{n} \text{ for } x \in \mathcal{X}. \quad (2.7)$$

Then, by the law of large numbers, for each $x \in \mathcal{X}$, $\pi (x \mid x^n) \rightarrow p(x)$ in probability. Let $\epsilon \in (0, 1)$. The set of $\epsilon$-typical $n$-sequence $x^n$ is defined as

$$\mathcal{T}_\epsilon^n(X) = \{x^n : |\pi (x \mid x^n) - p(x)| \leq \epsilon p(x) \text{ for all } x \in \mathcal{X}\}. \quad (2.8)$$

The notion of typicality can be extended to multiple random variables. For example, let a pair of random variables $(X, Y) \sim p(x, y)$. The set of jointly $\epsilon$-typical $n$-sequence is defined as

$$\mathcal{T}_\epsilon^n(X, Y) = \{(x^n, y^n) : |\pi (x, y \mid x^n, y^n) - p(x, y)| \leq \epsilon p(x, y) \text{ for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\},$$

where the empirical joint PMF $\pi (x, y \mid x^n, y^n)$ is defined as

$$\pi (x, y \mid x^n, y^n) = \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n} \text{ for } (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (2.10)$$

While referring to [42] for a complete discussion on the properties of joint typicality and jointly typical sequences, in the following, we introduce four important lemmas in [42] that we use frequently in the proofs of the results in this research.

The first one is the conditional typicality lemma and it is stated as follows.
Lemma 2.1 (Conditional Typicality Lemma [42]). Let \((X, Y) \sim p(x, y)\). Suppose that 
\(x^n \in T^n_\epsilon (X)\) and \(Y^n \sim p(y^n | x^n) = \prod_{i=1}^{n} p_{Y|X}(y_i | x_i)\). Then, for every \(\epsilon > \epsilon'\),
\[
\Pr \left\{ (x^n, y^n) \in T^n_\epsilon (X, Y) \right\} \to 1 \tag{2.11}
\]
as \(n \to \infty\).

The second one is the joint typicality lemma, which bounds the probability of a set of joint typicality events as defined both from above and below.

Lemma 2.2 (Joint Typicality Lemma [42]). Let \((X, Y, Z) \sim p(x, y, z)\) and \(\epsilon' < \epsilon\). Let function \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\). We have that:

1. If \((\tilde{x}^n, \tilde{y}^n)\) is a pair of arbitrary sequences and \(\tilde{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(\tilde{z}_i | \tilde{x}_i)\), then
\[
\Pr \left\{ (\tilde{x}^n, \tilde{y}^n, \tilde{Z}^n) \in T^n_\epsilon (X, Y, Z) \right\} \leq 2^{-n(I(Y;Z|X) - \delta(\epsilon))}. \tag{2.12}
\]

2. If \((\tilde{x}^n, \tilde{y}^n) \in T^n_{\epsilon'} (X, Y)\) and \(\tilde{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(\tilde{z}_i | \tilde{x}_i)\), then
\[
\Pr \left\{ (\tilde{x}^n, \tilde{y}^n, \tilde{Z}^n) \in T^n_\epsilon (X, Y, Z) \right\} \geq 2^{-n(I(Y;Z|X) + \delta(\epsilon))}. \tag{2.13}
\]

Next, the packing lemma is introduced. This lemma generalizes the bound on the probability of the decoding error events for the Shannon channel coding theorem [4], and it plays an important role in the achievability proofs of many channel coding theorems. The lemma reads as follows.
Lemma 2.3 (Packing Lemma [42]). Let \((U, X, Y) \sim p(u, x, y)\). Let \((\tilde{U}^n, \tilde{Y}^n) \sim p(\tilde{u}^n, \tilde{y}^n)\) be a pair of arbitrarily distributed random sequences. Let \(X^n(w), w \in \mathcal{W}\) with \(|\mathcal{W}| \leq 2^{nR}\), be random sequences each distributed according to \(\prod_{i=1}^n p_{X|U}(x_i|u_i)\). In addition, assume that \(X^n(w)\) is pairwise conditionally independent of \(\tilde{Y}^n\) given \(\tilde{U}^n\), but is arbitrarily dependent on other \(X^n(w)\) sequences. Then there exists function \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\) such that the probability

\[
\Pr \left\{ \left( \tilde{U}^n, X^n(w), \tilde{Y}^n \right) \in T^n_\epsilon(U, X, Y) \text{ for some } w \in \mathcal{W} \right\} \to 0 \quad (2.14)
\]

as \(n \to \infty\) if \(R < I(X;Y|U) - \delta(\epsilon)\).

Finally, the covering lemma is introduced. This lemma generalizes the bound on the probability of the encoding error events for the lossy source coding theorem and is crucial for the achievability proofs of many multi-user source and channel coding theorems.

Lemma 2.4 (Covering Lemma [42]). Let \((U, X, \hat{X}) \sim p(u, x, \hat{x})\) and \(\epsilon' < \epsilon\). Let \((U^n, X^n) \sim p(u^n, x^n)\) be a pair of jointly \(\epsilon'\)-typical \(n\)-sequences. Let \(\hat{X}^n(l), l \in \mathcal{L}\) with \(|\mathcal{L}| \geq 2^{n\tilde{R}}\), be random sequences conditionally independent of each other and of \(X^n\) given \(U^n\), each distributed according to \(\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)\). Then there exists function \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\) such that the probability

\[
\Pr \left\{ \left( U^n, X^n, \hat{X}^n(l) \right) \notin T^n_\epsilon(U, X, \hat{X}) \text{ for all } l \in \mathcal{L} \right\} \to 0 \quad (2.15)
\]

as \(n \to \infty\) if \(\tilde{R} > I(X;\hat{X}|U) + \delta(\epsilon)\).
2.3 Gel’fand-Pinsker Binning and Dirty-Paper Coding

Besides the basic notions and lemmas above, we also summarize some fundamental coding schemes. We first present coding schemes for the single-user channel with non-causal state information.

The capacity of the DM single-user channel with random state was found by Gel’fand and Pinsker [7], assuming that the entire state sequence is known non-causally to the encoder but unknown to the decoder. This channel is denoted by the tuple \((X \times S, p(y|x, s), p(s), Y)\) (see Fig. 2.1), where \(X\) is the input alphabet, \(Y\) is the output alphabet, \(S\) is the state alphabet, \(p(s)\) denotes the state PMF and \(p(y|x, s)\) represents the channel PMF. The state sequence \(s^n\) is i.i.d. according to \(p(s^n) = \prod_{i=1}^{n} p_S(s_i)\). The capacity result is given as follows.

**Theorem 2.1 ([7]).** The capacity of the DM single-user channel with non-causal state information at the encoder is characterized by

\[
C = \max_{p(u|s), x(u,s)} \left( I(U; Y) - I(U; S) \right), \tag{2.16}
\]

where \(U\) is an auxiliary random variable, whose cardinality is bounded by

\[
|U| \leq \min \{ |X| |S|, |Y| + |S| - 1 \}, \tag{2.17}
\]

and \(X\) is a deterministic function of \(U\) and \(S\) and hence is denoted as \(x(u, s)\).
Fig. 2.1. The discrete memoryless single-user channel with random state non-causally known to the encoder.

The capacity is attained by GP binning [7]. The main idea is that the encoder generates a large codebook in which each message is associated with multiple $U^n$ codewords. One may visually imagine such a codebook as a set of bins, where each message represents the index of a bin and in each bin there are an equal number of $U^n$ codewords. To transmit message $w$, the encoder first looks for a $u^n$ codeword in bin $w$ that matches with the state sequence realization $s^n$. Then, based on the codeword selected and the state sequence, the encoder forms an appropriate channel input $x^n$ and transmits it over the channel. Upon receiving sequence $y^n$, the decoder looks for a $\hat{u}^n$ codeword that is jointly typical with $y^n$ and claims the message estimate as the index of the bin to which codeword $\hat{u}^n$ belongs.

Now, consider a particular instance of the GP model:

$$Y = X + S + Z,$$

(2.18)

where $Z \sim \mathcal{N}(0, N_0)$ is additive Gaussian noise, $S \sim \mathcal{N}(0, P_S)$ is additive interfered state independent of the noise, and the channel input satisfies power constraint $\sum_{i=1}^{n} \mathbb{E} \left[ X_i^2 \right] \leq nP$. This channel was studied by Costa in [8] and the setting is coined “Writing on Dirty
Paper”. Surprisingly, it was shown that with non-causal knowledge about the interfered state at the transmitter, the capacity of the channel is the same as if no interference is present in the channel, i.e., $C = C(P/N_0)$. The achievability is attained by the Dirty-Paper Coding (DPC), a particular realization of the GP binning in which input $X \sim \mathcal{N}(0, P)$, independent of $S$ and $U = X + \alpha S$, for some $\alpha > 0$. Furthermore, it is shown that the optimum $\alpha^*$ corresponds to the weight of minimum mean-square-error (MMSE) estimate of $X$ given $Y - S$, i.e., $\alpha^* = P/P + N_0$. Later we shall encounter applications of GP binning and DPC, see, e.g., in Chapter 5.

2.4 Basic Digital Relaying Strategies

In this section, basic digital relaying strategies are reviewed. We will exploit some key techniques in these transmission strategies later, e.g., in Chapter 3 and Chapter 4.

Consider a three-terminal relay channel, where the source wishes to communicate to the destination via the help of a relay node. Van der Meulen first introduced this channel model in [43]. In [44], Cover and El Gamal proposed serval coding schemes for the same model. Variations of Cover and El Gamal’s coding schemes have been developed and applied to the relay network in recent decades. A non-exhaustive list of work includes [45, 46, 47, 48, 49, 50, 51]. Following the taxonomy advocated in [51], we have two main categories of digital relaying strategies: decode-and-forward (DF) and compress-and-forward (CF).

The DF first appeared in [44] and refers to the strategy whereby the relay decodes part or full of the message and then cooperates with the source on transmitting
the previous message to the destination. In reference [44], the implementation of DF relaying involves techniques of block-Markov coding and binning along with regular block-decoding. However, one can also implement the DF without binning but coupled with backward decoding [52] or slide-window decoding [45]. The DF relaying relies on the successful decoding at the relay and hence its performance is limited by the quality of the source-to-relay link.

Instead, the CF refers to the strategy whereby the relay does not attempt to decode the message but transmits a compressed version of its noisy observation of the source signal to the destination so that the latter can first recover the compressed information and utilize it along with its own received sequence to perform message decoding. When the source-to-relay link is weak and no full message decoding is possible at the relay, the CF performs better than the DF. On the basis of compressing operations at the relay, the CF is classified into several groups: the regular CF, hashing-forward and quantized-forward relaying.

- In the regular CF relaying [44], by treating the received sequence at the destination as side information, the relay compresses its noisy observation of the source signal via Wyner-Ziv binning [53] and forwards the bin index to the destination. The destination decodes the bin index, reconstructs a quantized version of the relay’s received sequence, and then uses it along with its own received sequence to decode the message. It is noted that in the scheme, the Wyner-Ziv binning includes both compression (quantization) and hashing operations.
• In the hashing-forward (HF) relaying [47], the relay hashes its observed signal directly without quantization and forwards the bin index to the destination. For decoding, the destination first retrieves the bin index and then exploits it along with its own received sequence to perform message decoding.

• In the quantize-forward (QF) relaying [46, 48, 50], the relay instead compresses its observed signal in a standard manner and conveys the compression index to the destination. In this way, the destination exploits the quantized signal and its own received sequence to decode the message.

It is remarked that in both the regular CF and HF relaying, the message at the source is normally split into independent and short messages with each carried over one transmission block, while in the QF relaying, a long message is instead encoded over all transmission blocks. Correspondingly, the decoding at the destination associated with the regular CF and HF relaying is conventionally done in a block-by-block fashion, while the decoding associated with the QF relaying is instead done jointly over all transmission blocks. Regarding their performance, references [44, 47, 50, 51] indicate that no single scheme dominates the others at all. Interested readers may refer to [51] for further comparisons of these CF schemes in terms of reliability, complexity, delay and flexibility.

2.5 Chapter Summary

In this chapter, we have introduced some basic notions and lemmas that are useful in this research. In addition, we have reviewed several fundamental coding schemes related to the current research.
Chapter 3

Multiple Access Channels with Strictly Causal State Information

3.1 Introduction

As described in Chapter 1, in the dissertation, we are interested in identifying the benefits of state information available at nodes in state-dependent multi-user channels. In this chapter, we start with the most basic model in the category, i.e., the multiple access channels (MACs) with independent states.

While referring to [14] for a thorough review on state-dependent channels, here we summarize existing results on state-dependent MACs, which are the focus of this chapter. Reference [28] derived single-letter inner and outer bounds on the capacity region for two-user MACs with causal common state information at the encoders. Reference [21] derived a genie-aided bound to assess the capacity advantage of non-causal state over causal state information for MACs with independent state sequences available at the two encoders. Reference [15] characterized the capacity region of a cooperative MAC with state non-causally available at one encoder, while reference [22] proposed several inner and outer bounds for a MAC with states non-causally known to some encoders. A lattice coding strategy was proposed for a MAC with non-causal state information in [16] and [25].
The works summarized above demonstrate the advantages of causal and non-causal state information at the encoders for MACs. Instead, in references [23] and [24] Lapidoth and Steinberg discovered that, even with strictly causal state information at the encoders, an improvement in the capacity region is possible. By strictly causal state information, it is meant that at channel use \( i \), the encoders know a state sequence up to, but excluding channel use \( i \). This result stands in contrast to the well-known fact that strictly causal state information cannot improve the capacity of point-to-point channels with an i.i.d. state sequence. More specifically, in [23], a common state sequence is assumed to be known either strictly causally or causally at both encoders of a two-user MAC, while in [24] two independent state sequences are assumed to be available strictly causally or causally, each to a single encoder of a two-user MAC. An achievable rate region is derived in both papers and the capacity region is identified for some special cases including Gaussian models.

The main idea in the achievability proofs in [23] and [24] is to use a block Markov coding scheme in which the two users cooperatively [23] or non-cooperatively [24] transmit compressed past state information to the decoder, which in turn uses such information to perform joint decoding along with the channel output, i.e., perform coherent decoding. The results show that an increase in the capacity region can be obtained, even though transmission of the state information requires diverting part of the transmission resources from the transmission of message information.
3.1.1 Contributions

In this chapter, we propose a generalization of the strategy in [24] whereby the encoders compress also the past transmitted codewords along with the past state sequences. We first focus on the two-user MAC with independent states each strictly causally known to its corresponding encoder. The proposed scheme is based on long-message encoding [51], compression of the past state sequences and past codewords without binning, and joint decoding over all transmission blocks [50]. We also report on an example, put forth by Lapidoth and Steinberg in [54], in which the proposed scheme is shown to strictly improve upon the original strategy of [24].

We then generalize the capacity result for Gaussian channels of [24] for the case of a single state sequence to a larger class of channels that includes two-user modulo-additive state-dependent MACs. The proposed scheme is then extended to the state-dependent MAC with an arbitrary number of users. Finally, we introduce output feedback and show via a specific example that feedback allows user cooperation for the transmission of state information to the receiver, beside standard cooperation on the transmission of messages [55], and that this increases the capacity region.

The remainder of the chapter is organized as follows: In Section 3.2, we describe the general model considered in this work and summarize some of the existing results in [24]. Sections 3.3 and 3.4 focus on the two-user state-dependent MAC. Section 3.5 provides a generalization to arbitrary number of users with independent states. Section 3.6 discusses the model with output feedback. Section 3.7 concludes the chapter.
3.2 System Model and Preliminaries

In this section, we describe our channel model, formulate the problem and revisit some related results derived in previous work [24].

3.2.1 System Model

We investigate an $M$-user discrete memoryless MAC channel with $M$ mutually independent states, which is depicted in Fig. 3.1 and denoted by the tuple

$$(X_1 \times \ldots \times X_M, S_1 \times \ldots \times S_M, Y, p(s_1), \ldots, p(s_M), \ p(y|x_1, \ldots, x_M, s_1, \ldots, s_M)) (3.1)$$

with input alphabets $(X_1, \ldots, X_M)$, output alphabet $Y$ and state alphabets $(S_1, \ldots, S_M)$. The state sequences are assumed to be i.i.d. and are mutually independent, i.e.,

$$\prod_{k=1}^{M} p(s_k^n) = \prod_{k=1}^{M} \prod_{i=1}^{n} p(s_{k,i}). (3.2)$$

The state-dependent channel is memoryless in the sense that at any discrete time $i = 1, \ldots, n$, we can write:

$$p \left( y_i \bigg| x_{1,i}^{i-1}, x_{M,i}^{i-1}, s_{1,i}^{i-1}, s_{M,i}^{i-1} \right) = p \left( y_{i} \bigg| x_{1,i}^{i-1}, \ldots, x_{M,i}, x_{M,i}, s_{1,i}, \ldots, s_{M,i} \right). (3.3)$$

Each state realization is available to its corresponding encoder in a strictly causal manner as defined in Section 3.1. Transmitter $k$’s signal $x_k^n$ is subject to an average input cost
constraint:

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[c_k(X_{k,i})] \leq \Gamma_k, \quad k = 1, \ldots, M,
\]  

(3.4)

where \( c_k : \mathcal{X}_k \rightarrow \mathbb{R}^+ \) is a single-letter input cost function for transmitter \( k \) and the expectation is taken with respect to all the messages and states. We now define the following code.

**Definition 3.1.** Let \( w_k \), uniformly distributed over the set \( \mathcal{W}_k = [1 : 2^{nR_k}] \), be the message sent by transmitter \( k \). A \((2^{nR_1}, \ldots, 2^{nR_M}, n, \Gamma_1, \ldots, \Gamma_M)\) code for the MAC with strictly causal and independent state information at the encoders consists of sequences of encoder mappings:

\[
f_{k,i} : \mathcal{W}_k \times \mathcal{S}^{i-1}_k \rightarrow \mathcal{X}_k, \quad i = 1, \ldots, n, \quad k = 1, \ldots, M,
\]

(3.5)

each of which maps message \( w_k \) to a channel input such that the cost constraint (3.4) is satisfied, and a decoder mapping

\[
g : \mathcal{Y}^n \rightarrow \mathcal{W}_1 \times \cdots \times \mathcal{W}_M,
\]

(3.6)

which produces the estimate of messages \((w_1, \ldots, w_M)\).
Fig. 3.1. The $M$-user state-dependent MAC with $M$ mutually independent states, each of which is available to its corresponding encoder in a strictly causal manner.

The average probability of error, $\Pr(E)$, is defined by:

$$\Pr(E) = \frac{1}{M} \prod_{k=1}^{M} \sum_{w_1=1}^{2^{nR_k}} ... \sum_{w_M=1}^{2^{nR_M}} \Pr \left( g \left( y^n \right) \neq (w_1, ..., w_M) \mid (w_1, ..., w_M) \text{ sent} \right).$$

(3.7)

Given a cost tuple $\Gamma = (\Gamma_1, ..., \Gamma_M)$, a rate tuple $(R_1, ..., R_M)$ is said to be $\Gamma$-achievable if there exists a sequence of codes $(2^{nR_1}, ..., 2^{nR_M}, n, \Gamma_1, ..., \Gamma_M)$ as defined above such that the probability of error satisfies $\Pr(E) \to 0$ as $n \to \infty$. The capacity region $\mathcal{C}(\Gamma)$ is the closure of all the $\Gamma$-achievable rate tuples.

We first restrict our attention to a two-user MAC with two independent states, and then generalize to an arbitrary $M$-user MAC with $M$ independent states in Section 3.5.

3.2.2 Preliminaries

For comparison, we summarize a key result of [24].
Theorem 3.1 ([24]). Let $\Gamma = (\Gamma_1, \Gamma_2)$ be given. Let $\mathcal{P}_{sc}$ be the set of all random variables $(V_1, V_2, S_1, S_2, X_1, X_2, Y)$ whose joint distribution is factorized as

$$p(v_1 | s_1)p(v_2 | s_2)p(s_1)p(s_2)p(x_1)p(x_2)p(y | s_1, s_2, x_1, x_2).$$

(3.8)

For the two-user MAC with strictly causal state information, a $\Gamma$-achievable rate region, denoted as $\mathcal{R}_{\text{in1}}(\Gamma_1, \Gamma_2)$, is given by the projection in the $(R_1, R_2)$ plane of the set of rate-cost tuples $(R_1, R_1, \Gamma_1, \Gamma_2)$ belonging to the convex hull of the collection of all the tuples $(R_1, R_2, \Gamma_1', \Gamma_2')$ satisfying

$$0 \leq R_1 < I(X_1; Y | X_2, V_1, V_2) - I(V_1; S_1 | Y, V_2),$$

(3.9a)

$$0 \leq R_2 < I(X_2; Y | X_1, V_1, V_2) - I(V_2; S_2 | Y, V_1),$$

(3.9b)

$$R_1 + R_2 < I(X_1, X_2; Y | V_1, V_2) - I(V_1, V_2; S_1, S_2 | Y),$$

(3.9c)

and $\mathbb{E}[c_k(X_k)] \leq \Gamma_k'$, $k = 1, 2,$

(3.9d)

for some random variables $(V_1, V_2, S_1, S_2, X_1, X_2, Y) \in \mathcal{P}_{sc}$.

Remark 3.1. The basic idea of the achievable scheme of Theorem 3.1 is to let the transmitters convey a compressed version of the state, namely $V_1$ for $S_1$ and $V_2$ for $S_2$, to the receiver. The receiver can then use this partial information about the state to improve decoding. As an example, if the state models fading channels, state information enables partially coherent decoding. The proof of the theorem, though not available in detail in [24], is there indicated to be based on a scheme that leverages distributed Wyner-Ziv compression [56] and block Markov encoding. □
3.3 A New Achievable Rate Region

In this section, for the two-user MAC ($M = 2$), we propose a new achievable scheme. The scheme is based on the idea of letting the encoders compress also the past input codewords along with the past states. We first show that the new achievable region includes the original one. Then, we report on the example put forth in [54] that demonstrates that the inclusion can be strict.

**Theorem 3.2.** Let $\Gamma = (\Gamma_1, \Gamma_2)$ be given. Let $P_{sc}^*$ be the set of all random variables $(V_1, V_2, S_1, S_2, X_1, X_2, Y)$ whose joint distribution is factorized as

$$p(v_1 | s_1, x_1)p(v_2 | s_2, x_2)p(s_1)p(s_2)p(x_1)p(x_2)p(y | s_1, s_2, x_1, x_2).$$

(3.10)

For the two-user MAC with strictly causal state information, a $\Gamma$-achievable rate region, denoted as $R_{in2}(\Gamma_1, \Gamma_2)$, is given by the projection in ($R_1, R_2$) plane of the set of rate-cost tuples $(R_1, R_1, \Gamma_1, \Gamma_2)$ belonging to the convex hull of the collection of all the tuples $(R_1, R_2, \Gamma'_1, \Gamma'_2)$ satisfying

$$0 \leq R_1 < I(X_1, V_1; Y | X_2, V_2) - I(V_1; S_1 | X_1),$$

(3.11a)

$$0 \leq R_2 < I(X_2, V_2; Y | X_1, V_1) - I(V_2; S_2 | X_2),$$

(3.11b)

$$R_1 + R_2 < I(X_1, X_2, V_1, V_2; Y) - I(V_1; S_1 | X_1) - I(V_2; S_2 | X_2),$$

(3.11c)

and

$$E[c_k(X_k)] \leq \Gamma'_k, \ k = 1, 2,$$

(3.11d)

for some random variables $(V_1, V_2, S_1, S_2, X_1, X_2, Y) \in P_{sc}^*$.  

Proof: The theorem follows as a special case of the $M$-user result of Theorem 3.5 for $M = 2$. We refer the reader to Appendix 3.8.2 for a proof of Theorem 3.5. ■

Remark 3.2. In the proposed strategy, the transmitters convey codewords $V_1$ and $V_2$, which compress both the past state sequences and the past transmitted codewords of the previous transmission block. This difference with respect to Theorem 3.1 is reflected in the different factorizations (3.8) and (3.10). Specifically, in the latter, the test channels $p(v_k | s_k, x_k)$, $k = 1, 2$, is made to depend also on the previously transmitted symbols $X_k$. We also note that, unlike [24], our scheme uses long-message encoding, quantization without binning and joint decoding over all blocks of transmission, similar to [50] (see also [51]). □

While the joint distribution factorization (3.10) is more general than the original (3.8) used in [24], the two regions (3.9) and (3.11) are not immediately comparable given the different mutual information expressions. The next theorem shows that in fact the proposed achievable region always includes the original.

Theorem 3.3. The achievable rate region of Theorem 3.2 includes the achievable region of Theorem 3.1, i.e., $\mathcal{R}_{in2}(\Gamma_1, \Gamma_2) \supseteq \mathcal{R}_{in1}(\Gamma_1, \Gamma_2)$.

Proof: Given any constraint pair $(\Gamma_1, \Gamma_2)$, setting $p(v_1 | s_1, x_1) = p(v_1 | s_1)$ and $p(v_2 | s_2, x_2) = p(v_2 | s_2)$ in $\mathcal{R}_{in2}(\Gamma_1, \Gamma_2)$, we obtain the following.

1. For the sum-rate bound,

$$R_1 + R_2 < I(X_1, X_2; Y | V_1, V_2) + I(V_1, V_2; Y) - I(V_1; S_1 | X_1) - I(V_2; S_2 | X_2)$$

(3.12a)
\begin{align*}
&= I(X_1, X_2; Y | V_1, V_2) + H(V_1 | S_1) + H(V_2 | S_2) - H(V_1, V_2 | Y) \\
&= I(X_1, X_2; Y | V_1, V_2) + H(V_1 | S_1, S_2, Y) + H(V_2 | S_2, S_1, V_1, Y) - H(V_1, V_2 | Y) \\
&= I(X_1, X_2; Y | V_1, V_2) - I(V_1, V_2; S_1, S_2 | Y), \tag{3.12d}
\end{align*}

where (3.12c) follows from the Markov chain \((V_1, V_2) \leftrightarrow (S_1, S_2) \leftrightarrow Y\) and from the fact that \((V_1, S_1)\) are independent of \((V_2, S_2)\). Note the last equation is exactly the same sum-rate bound in \(R_{in1}(\Gamma_1, \Gamma_2)\) given by (3.9c).

2. For the individual rate bound on \(R_1\), we can write

\begin{align*}
R_1 &< I(X_1; Y | X_2, V_1, V_2) + I(V_1; Y | X_2, V_2) - I(V_1; S_1 | X_1) \tag{3.13a} \\
&= I(X_1; Y | X_2, V_1, V_2) + H(V_1 | S_1) - H(V_1 | Y, X_2, V_2) \tag{3.13b} \\
&\geq I(X_1; Y | X_2, V_1, V_2) + H(V_1 | S_1) - H(V_1 | Y, V_2) \tag{3.13c} \\
&= I(X_1; Y | X_2, V_1, V_2) + H(V_1 | S_1, Y, V_2) - H(V_1 | Y, V_2) \tag{3.13d} \\
&= I(X_1; Y | X_2, V_1, V_2) - I(V_1; S_1 | Y, V_2), \tag{3.13e}
\end{align*}

where (3.13c) follows from conditioning reduces entropy while (3.13d) follows from the Markov chain \(V_1 \leftrightarrow S_1 \leftrightarrow Y\). The last equation is exactly the same as the bound on \(R_1\) in \(R_{in1}(\Gamma_1, \Gamma_2)\) given by (3.9a).

3. A similar observation holds for \(R_2\) by symmetry.
These three facts imply the relationship $\mathcal{R}_{in2}(\Gamma_1, \Gamma_2) \supseteq \mathcal{R}_{in1}(\Gamma_1, \Gamma_2)$.

It was recently shown in [54] that the proposed region $\mathcal{R}_{in2}(\Gamma_1, \Gamma_2)$ strictly includes the original region $\mathcal{R}_{in1}(\Gamma_1, \Gamma_2)$ for some channels. The following is the example given in [54] that illustrates such inclusion.

**Example 3.1 ([54]).** Consider a MAC with two binary inputs $X_1 = X_2 = \{0, 1\}$; state $S_1 = \emptyset$ and state $S_2 = (T_0, T_1) \in \{0, 1\}^2$, where $T_0$ and $T_1$ are independent with entropies

$$H(T_0) = H(T_1) = \frac{1}{2};$$  \hspace{1cm} (3.14)

and the output $Y = (Y_1, Y_2) \in \{0, 1\}^2$ is given as

$$Y_1 = X_1 \oplus T X_2, \hspace{1cm} (3.15a)$$

$$Y_2 = X_2, \hspace{1cm} (3.15b)$$

where notation “$\oplus$” denotes the conventional modulo-sum operation. The key point of this example is that the state sequence affects the received signal in a way that depends on the transmitted symbol $X_2$. Therefore, joint compression of both the past state and the past codeword, or compression of the past state in a way that depends on the past codeword, is expected to be beneficial. To show this, following [54], it can be seen that rate pair $(1, \frac{1}{2})$ lies in the inner bounds of $\mathcal{R}_{in2}$ by setting $V_1 = \emptyset$, $V_2 = TX_2$ in (3.11). However, with $R_1 = 1$, it was demonstrated in [54] that $R_2$ is necessarily zero in $\mathcal{R}_{in1}$. 
This allows us to conclude, along with Theorem 3.3, that the region $R_{in2}$ is strictly larger than the region $R_{in1}$ for this example.

### 3.4 Capacity Result

In this section, we generalize the capacity result derived in [24] for Gaussian channels with a single state sequence to a larger class of channels.

Consider a class of discrete memoryless two-user deterministic MACs denoted by $\mathcal{D}_{MAC}$, in which the output $Y$ is a deterministic function of inputs $X_1$, $X_2$ and the channel state $S$ as

$$Y = f(X_1, X_2, S),$$

(3.16)

and where the channel state $S$, strictly causally known to encoder 1, can be calculated as a deterministic function of the inputs $X_1$, $X_2$ and the output $Y$ as

$$S = g(X_1, X_2, Y).$$

(3.17)

Then the capacity region for the class of channels $\mathcal{D}_{MAC}$ is identified as follows.
Theorem 3.4. Let \( \Gamma = (\Gamma_1, \Gamma_2) \) be given. For any MAC in the class \( \mathcal{D}_{MAC} \) defined above, the capacity region \( \mathcal{C}(\Gamma) \) is given by:

\[
\mathcal{C}(\Gamma) \triangleq \bigcup \left\{ (R_1, R_2) \in \mathbb{R}^+_2 : \begin{align*}
R_1 &\leq H(Y | X_2, Q) - H(S) \\
R_2 &\leq H(Y | X_1, S, Q) \\
R_1 + R_2 &\leq H(Y | Q) - H(S)
\end{align*}\right\} \tag{3.18}
\]

where the union is taken over all product input distributions \( p(x_1 | q)p(x_2 | q)p(q) \) satisfying \( \mathbb{E}[c_k(X_k)] \leq \Gamma_k, \ k = 1, 2, \) and \( Q \) is an auxiliary random variable with cardinality bound \( |Q| \leq 5 \).

Proof: See Appendix 3.8.1. \( \blacksquare \)

Remark 3.3. The achievability proof in Appendix 3.8.1 is based on setting \( V_1 = S_1 = S \) in the achievable region \( R_{in2} \) in Theorem 3.2, which implies that \( V_1 \) is independent of \( X_1 \). Hence, the achievable scheme proposed in [24] is also optimal for the class of channels considered here. \( \square \)

Remark 3.4. The class \( \mathcal{D}_{MAC} \) includes the Gaussian model considered in [24], which is defined as \( Y = X_1 + X_2 + S \) with input power constraints \( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_{k,i}^2] \leq P_k \) and state \( S \sim \mathcal{N}(0, \sigma^2_S) \) known strictly causally to encoder 1. The capacity region \( \mathcal{C}^{snf} \) for
this model is given by:

\[ 
\mathcal{C}_{snf} = \left\{ (R_1, R_2) \in \mathbb{R}_2^+ : \\
R_1 \leq C\left( \frac{P_1}{\sigma_s^2} \right) \\
R_1 + R_2 \leq C\left( \frac{P_1 + P_2}{\sigma_s^2} \right) \right\}. 
\]  

(3.19)

This region can be identified from Theorem 3.4 by the standard extension to continuous alphabets (see, e.g., [42, Chapter 3]) and by maximizing each bound via the maximum entropy theorem [57]. Note that when providing both \( S \) and \( X_1 \) to the receiver, the channel from user 2 to the receiver is noiseless and hence the individual bound on \( R_2 \) is redundant. □

Remark 3.5. The class \( D_{MAC} \) contains more channels along with the Gaussian model discussed in Remark 3.4. In particular, consider a class of binary modulo-additive state-dependent MAC channels, e.g., \( Y = X_1 \oplus X_2 \oplus S \), where \( S \sim Bernoulli(p_s) \), with input cost constraints \( \frac{1}{n} \sum_{i=1}^{n} E[X_{1,i}] \leq p_1 \) and \( \frac{1}{n} \sum_{i=1}^{n} E[X_{2,i}] \leq p_2 \), \( 0 < p_1, p_2, p_s \leq \frac{1}{2} \). Note that assumption (3.17) automatically holds for this class of binary deterministic channels. From Theorem 3.4, by direct evaluation, we obtain that the capacity region is:

\[ 
\mathcal{C}_{bin}^s = \left\{ (R_1, R_2) \in \mathbb{R}_2^+ : \\
R_1 \leq H_b(p_1 * p_s) - H_b(p_s) \\
R_2 \leq H_b(p_2) \\
R_1 + R_2 \leq H_b(p_1 * p_2 * p_s) - H_b(p_s) \right\}. 
\]  

(3.20)
where \( p_1 \ast p_2 \) denotes the convolution operation of two Bernoulli distributions with parameters \( p_1 \) and \( p_2 \), i.e., \( p_1 \ast p_2 = p_1(1-p_2) + p_2(1-p_1) \), and \( H_b(p) = -p \log_2 p - (1-p) \log_2 (1-p) \). It is known from [57] that, without state information, the capacity region for this MAC channel is given by:

\[
\mathcal{C}^{ns}_{bin} = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 : \begin{array}{l}
R_1 \leq H_b(p_1 \ast p_s) - H_b(p_s) \\
R_2 \leq H_b(p_2 \ast p_s) - H_b(p_s) \\
R_1 + R_2 \leq H_b(p_1 \ast p_2 \ast p_s) - H_b(p_s)
\end{array} \right\}.
\] (3.21)

Hence, we have the relationship \( \mathcal{C}^{ns}_{bin} \subseteq \mathcal{C}^s_{bin} \), which confirms the benefit of strictly causal state information in enlarging the capacity region for this channel. For a numerical example, we set \( p_1 = p_2 = \frac{1}{3} \) and \( p_s = \frac{1}{4} \). The corresponding regions (3.20) and (3.21) are depicted and compared in Fig. 3.2. It is seen that the presence of strictly causal state information at encoder 1 improves the maximum rate of user 2. \( \square \)

### 3.5 Generalization to \( M \) Users with Independent States

In this section, we generalize the proposed achievable scheme to an arbitrary number \( M \) of users with independent states, as depicted in Fig. 3.1 and described in Section 3.2.

Let \( \mathcal{A} \) denote any subset of the set of encoders \([1 : M]\), i.e., \( \mathcal{A} \subseteq [1 : M] \) and \( \mathcal{A}^c \) be the complement of \( \mathcal{A} \) with respect to the set \([1 : M]\). Define \( X(\mathcal{A}) \) to be the set of random variables \( X_k \) indexed by \( k \in \mathcal{A} \) and similarly for \( V(\mathcal{A}) \).
Fig. 3.2. Capacity region for the binary modulo-additive state-dependent MAC with input constraints considered in Remark 3.5 ($p_1 = p_2 = 1/3, p_s = 1/4$).

**Theorem 3.5.** Let $\Gamma = (\Gamma_1, ..., \Gamma_M)$ be given. Let $\mathcal{P}^s_{sc}$ be the set of all random variables $(V_1, ..., V_M, S_1, ..., S_M, X_1, ..., X_M, Y)$ whose joint distribution is factorized as

$$
\prod_{k=1}^{M} (p(v_k | s_k, x_k)p(s_k)p(x_k))p(y | x_1, ..., x_M, s_1, ..., s_M).
$$

For the $M$-user MAC with strictly causal and independent state information, a $\Gamma$-achievable rate region, denoted as $\mathcal{R}^{\Gamma}_{in}(\Gamma_1, ..., \Gamma_M)$, is given by the projection in the space $(R_1, ..., R_M)$ of the set of rate-cost tuples $(R_1, ..., R_M, \Gamma_1, ..., \Gamma_M)$ belonging to the convex hull of the tuples $(R_1, ..., R_M, \Gamma'_1, ..., \Gamma'_M)$ satisfying

$$
0 \leq \sum_{k \in T} R_k < \min_{\{S \subseteq [1:M]: T \subseteq S\}} \left( I(X(S), V(S); Y | X(S^c), V(S^c)) - \sum_{l \in S} I(V_l; S_l | X_l) \right),
$$

(3.23a)

$$
\forall T \subseteq [1 : M],
$$

(3.23b)
and $\mathbb{E}[c_k(X_k)] \leq \Gamma'_k$, $k = 1, \ldots, M$, \hspace{1cm} (3.23c)

for some random variables $(V_1, \ldots, V_M, S_1, \ldots, S_M, X_1, \ldots, X_M, Y) \in \mathcal{P}^*_sc$.

Proof: See Appendix 3.8.2. \hfill \blacksquare

### 3.6 Introducing Output Feedback

In this section, we briefly consider an extension of the model with independent states studied in Section 3.3, where output feedback is available to some encoder in addition to strictly causal state information. It is well known that the use of output feedback can enlarge the capacity region in MACs by allowing cooperation in the transmission of the encoders’ message [58, 59, 55]. Here, instead, we demonstrate that, with strictly causal state information, a different type of cooperation is enabled by feedback that concerns the transmission of the state sequence.

To this end, we focus on the two-user state-dependent Gaussian MAC shown in Fig. 3.3, for which the received signal is given by:

$$Y = X_1 + X_2 + S$$ \hspace{1cm} (3.24)

with power constraints $\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_{k,i}^2 \right] \leq P_k$, for $k = 1, 2$, and state $S \sim \mathcal{N} \left( 0, \sigma^2_s \right)$. We assume that information about $S$ is known perfectly and strictly causally to the first transmitter and a perfect output feedback link is available from the receiver to the second transmitter. More specifically, we have the following encoder and decoder mappings.
Definition 3.2. Let $w_k$, uniformly distributed over the set $\mathcal{W}_k = [1 : 2^n R_k]$, be the message sent by transmitter $k$, $k = 1, 2$. A $(2^n R_1, 2^n R_2, n, P_1, P_2)$ code for the MAC with strictly causal state information at encoder 1 and output feedback to encoder 2 consists of the sequences of encoder mappings:

\begin{align*}
  f_{1,i} : & \mathcal{W}_1 \times S^{i-1} \rightarrow \mathcal{X}_1, \\
  f_{2,i} : & \mathcal{W}_2 \times Y^{i-1} \rightarrow \mathcal{X}_2, \quad i = 1, \ldots, n,
\end{align*}

such that power constraints, i.e., $\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_{k,i}^2 \right] \leq P_k$, for $k = 1, 2$, are satisfied and a decoder mapping

\begin{equation}
  g : \mathcal{Y}^n \rightarrow \mathcal{W}_1 \times \mathcal{W}_2.
\end{equation}

Achievability and capacity region are defined in the usual way, see Section 3.2.
Theorem 3.6. The capacity region of the model in Fig. 3.3 is given by:

$$
\mathcal{C}_{sf} = \bigcup_{0 \leq \rho \leq 1} \left\{ (R_1, R_2) : \begin{array}{l}
R_1 \leq C \left( \frac{(1-\rho^2)P_1}{\sigma_s^2} \right) \\
R_1 + R_2 \leq C \left( \frac{P_1 + P_2 + 2\rho \sqrt{P_1 P_2}}{\sigma_s^2} \right) \end{array} \right\}.
$$

(3.27)

Proof: See Appendix 3.8.3. ■

Remark 3.6. Without feedback, it is known from [23] that, if the state is known strictly causally to both encoders, the capacity is given by:

$$
\mathcal{C}_{ss} = \left\{ (R_1, R_2) : \begin{array}{l}
(R_1, R_2) \in \mathbb{R}_2^+ : \\
R_1 \leq C \left( \frac{(1-\rho^2)P_1}{\sigma_s^2} \right) \\
R_1 + R_2 \leq C \left( \frac{P_1 + P_2 + 2\rho \sqrt{P_1 P_2}}{\sigma_s^2} \right) \end{array} \right\},
$$

(3.28)

whereas if the state is known strictly causally only to encoder 1, the capacity region $\mathcal{C}_{snf}$ is given by (3.19). We plot a instance of these three capacity regions by setting $P_1 = P_2 = 2$ and $\sigma_s^2 = 1$ in Fig. 3.4. As we observe, we have the inclusion relationships $\mathcal{C}_{snf} \subset \mathcal{C}_{sf} \subset \mathcal{C}_{ss}$. As it will be seen in the achievability proof in Appendix 3.8.3, the gains obtained by leveraging feedback can be ascribed to the fact that feedback enables cooperation between the encoders in transmitting the state information to the decoder. As a further remark, consider a fourth setting in which no state information is present at encoder 1 but output feedback is available to encoder 2. While the capacity region of the case is unknown in general, it can be easily seen that $R_2 \leq C \left( \frac{P_2}{\sigma_s^2} \right)$ holds for any coding scheme. This is because the capacity of user 2 cannot be improved via feedback.
Therefore, the capacity region in this case is strictly smaller than the capacity region $C^{sf}$ for the case in which the state is known at encoder 1. This demonstrates the interplay between the availability of strictly causal side information at encoder 1 and of output feedback at encoder 2 in increasing the capacity region.

3.7 Chapter Summary

In this chapter, we have studied the state-dependent MAC with strictly causal state information at the encoders, following the original work by Lapidoth and Steinberg in [23] and [24]. We have generalized the coding scheme proposed in [24] by allowing the encoders to compress jointly past states and codewords. The proposed scheme is shown to perform at least as well as the original one, and it was demonstrated in [54] that there are channels for which it outperforms the original strategy of [24]. Moreover, the capacity result for the Gaussian model of [24] for the special case of a single state sequence has
been generalized to a larger class of channels that includes two-user modulo-additive state-dependent MACs. Next, the proposed scheme has been extended to an arbitrary number of users. We have also demonstrated with an example that output feedback allows cooperation on the transmission of the state sequence in the presence of strictly causal state information.

3.8 Appendices

3.8.1 Proof of Theorem 3.4

**Achievability:**

We provide the proof of achievability for $Q = q$ for a constant value $q$ and drop the conditioning on $Q$ for simplicity. The region (3.18) then follows by using coded time-sharing [42]. We set $V_2 = S_2 = \emptyset$ and $V_1 = S_1 = S$ in the achievable region $R_{in2}$ and use the properties (3.16) and (3.17) that characterize the class of $D_{MAC}$ to obtain that a rate pair $(R_1, R_2)$ is achievable if

$$R_1 < I(X_1, S; Y | X_2) - I(S; S | X_1) = H(Y | X_2) - H(S), \quad (3.29a)$$

$$R_2 < I(X_2; Y | X_1, S) = H(Y | X_1, S), \quad (3.29b)$$

$$R_1 + R_2 < I(X_1, X_2, S; Y) - I(S; S | X_1) = H(Y) - H(S) \quad (3.29c)$$

and $E[c_k(X_k)] \leq \Gamma_k$, $k = 1, 2$, are satisfied.

**Converse:**
From Proposition 1 and 2 in [24], we have the bounds

\[ R_1 \leq I(X_1; Y | X_2, Q) + \epsilon_n \]  
(3.30a)

\[ = H(Y | X_2, Q) - H(S) + \epsilon_n \]  
(3.30b)

and

\[ R_1 + R_2 \leq I(X_1, X_2; Y | Q) + \epsilon_n \]  
(3.31a)

\[ = H(Y | Q) - H(S) + \epsilon_n \]  
(3.31b)

where \( \epsilon_n \to 0 \) as \( n \to \infty \), and we have defined \( Q \) as a uniformly distributed random variable in the set \([1 : n]\) and independent of all other variables, and also the variables \( X_1 = X_{1Q}, X_2 = X_{2Q}, Y = Y_Q \) and \( S = S_Q \). Moreover, by providing perfect state information to the receiver, one can prove the following bound by using standard arguments:

\[ R_2 \leq I(X_2; Y | X_1, S, Q) + \epsilon_n \]  
(3.32a)

\[ = H(Y | X_1, S, Q) + \epsilon_n. \]  
(3.32b)

From the definition of the code, it can be seen that the distribution on \((Q, S, X_1, X_2, Y)\) is of the form \( p(q, s, x_1, x_2, y) = p(q)p(x_1 | q)p(x_2 | q)p(s)p(y | x_1, x_2, s) \). Notice that both (3.30b) and (3.31b) leverage property (3.17) and the fact that \( S \) is independent.
of \((Q, X_1, X_2)\). For the cost constraints, starting from the definition (3.4), we easily obtain that \(\Gamma_k \geq \mathbb{E}[c_k(X_k)]\).

Finally, by the Fenchel-Eggleston-Caratheodory theorem \cite[Page 631]{42}, we establish the cardinality bound \(|Q| \leq 5\) by observing that the rate region in Theorem 3.4 is characterized by the following five continuous functions over the connected compact subset given by the product probability mass functions on \(X_1 \times X_2\): 

\[
H(Y | X_2, Q = q), \quad H(Y | X_1, S, Q = q), \quad H(Y | Q = q), \quad \mathbb{E}[c_1(X_1) | Q = q], \quad \text{and} \quad \mathbb{E}[c_2(X_2) | Q = q].
\]

### 3.8.2 Proof of Theorem 3.5

Throughout the achievability proof, we use the definition of typical sequences and typical sets as in reference \cite{42}, also see Chapter 2. The set of jointly \(\epsilon\)-typical sequences according to a joint probability distribution \(P_{X,Y}\) is denoted by \(T^n_{\epsilon}(XY)\). When the distribution with respect to which typical sequences are defined is clear from the context, we will use \(T^n_{\epsilon}\) for short. Throughout, we use capital letters to denote random variables and the corresponding lowercase letters to denote realized values.

In the proposed scheme, transmission takes place in \(b\) blocks of \(n\) channel uses each and the same message is transmitted in all blocks (long-message transmission \cite{51}). Let \(x^n_{k,j}\) be the codeword sent by user \(k\) in each block \(j \in [1:b]\). This codeword encodes both user \(k\)'s message \(w_k \in [1 : 2^{nbR_k}]\) and the index corresponding to a compressed version \(v^n_{k,j-1}\) of the state sequence \(s^n_{k,j-1}\) realized in the previous \((j-1)\)th block and of the codeword \(x^n_{k,j-1}\) transmitted in the previous block. After the \(b\) transmission blocks, based on the received signals \((y^n_1, ..., y^n_b)\), the decoder decodes the correct message tuple \(w = (w_1, ..., w_M)\) by joint typicality decoding over all blocks. We now provide details
on codebook generation, encoding and decoding operations, and probability of error analysis.

**Codebook Generation:**

Let $\epsilon > \epsilon' > 0$. Fix some probability mass function (PMF) $p_{X_k}$ such that the input cost constraint $E[c_k(X_k)] \leq \Gamma_k - \epsilon$ is satisfied, and the conditional PMFs $p(v_k | x_k, s_k)$, for all $k = 1, \ldots, M$. Define the marginal PMF

$$p_{V_k | X_k}(v_k | x_k) = \sum_{S_k \in S_k} (p_{V_k | X_k, S_k}(v_k | x_k, s_k)p_{S_k}(s_k)).$$

1. For each block $j \in [1 : b]$, randomly and independently generate $2^{nb R_k} \times 2^{\tilde{R}_k}$ i.i.d. sequences $x_{k,j}^n$ according to the PMF $p_{X_{k,j}}(x_{k,j}^n) = \prod_{i=1}^n p_{X_k}(x_{k,j,i})$, for $k = 1, \ldots, M$. Index the sequences as $x_{k,j}^n(w_k, t_{k,j-1})$, with $w_k \in [1 : 2^{nb R_k}]$ and $t_{k,j-1} \in [1 : 2^{\tilde{R}_k}]$. As it will be discussed below, index $t_{k,j-1}$ is used to encode a compressed version of past state and transmitted codeword from a codebook of rate $\tilde{R}_k$.

2. For each block $j \in [1 : b]$ and for each codeword $x_{k,j}^n(w_k, t_{k,j-1})$, randomly and independently generate $2^{\tilde{R}_k}$ i.i.d. sequences $v_{k,j}^n$ according to the marginal PMF

$$p_{V_{k,j} | X_{k,j}}(v_{k,j}^n | x_{k,j}^n) = \prod_{i=1}^n p_{V_k}(v_{k,j,i} | x_{k,j,i}),$$

for $k = 1, \ldots, M$. Index the sequences as $v_{k,j}^n(t_{k,j} | w_k, t_{k,j-1})$, with $t_{k,j} \in [1 : 2^{\tilde{R}_k}]$.

**Encoding:**

Let $w_k$ be the message sent by user $k$, where $k = 1, \ldots, M$. For block $j = 1$, codeword $x_{k,1}^n(w_k, 1)$ is transmitted by user $k$. For block $j \in [2 : b]$, instead, encoder $k$
looks for an index $t_{k,j-1}$ such that

$$
(s_{k,j-1}^n, v_{k,j-1}^n(t_{k,j-1} \mid w_k, t_{k,j-2}), x_{k,j-1}^n(w_k, t_{k,j-2})) \in T_{\epsilon'}^n(S_kV_kX_k).
$$

(3.33)

If no such index is found, then an arbitrary index $t_{k,j-1}$ is selected from the set $[1 : 2^{nR_k}]$.

If more than one such index is found, the first one in lexicographical order is selected.

Finally, the codeword $x_{k,j}^n(w_k, t_{k,j-1})$ is transmitted by user $k$ in the $j$th block.

**Decoding:**

After $b$ blocks of transmission, the decoder looks for a unique message tuple $\hat{w} = (\hat{w}_1, ..., \hat{w}_M)$, where $\hat{w}_k \in [1 : 2^{nhR_k}]$, such that there exists some tuple $(t_{1,j}, ..., t_{M,j})$, with $t_{k,j} \in [1 : 2^{nR_k}]$, $k = 1, ..., M$, and $j \in [1 : b]$, satisfying the condition

$$
\left( x_{1,j}^n(\hat{w}_1, t_{1,j-1}), ..., x_{M,j}^n(\hat{w}_M, t_{M,j-1}), v_{1,j}^n(t_{1,j} \mid \hat{w}_1, t_{1,j-1}), ..., v_{M,j}^n(t_{M,j} \mid \hat{w}_M, t_{M,j-1}), y_j^n \right) \in T_{\epsilon}^n
$$

(3.34)

for all blocks $j \in [1 : b]$.

**Probability of Error Analysis:**

We now bound the probability of error $\Pr(E)$ averaged over all distribution of the codebooks defined above. Without loss of generality, given the symmetry of the codebook generation, we assume the message tuple sent is $w = (1, ..., 1) \overset{\Delta}{=} 1_M$ and we label the compression index chosen by encoder $k$ for each block $j$ as $t_{k,j} = 1$. In the following, we first define the error events associated with the encoding and decoding operations, and then bound the corresponding probabilities of error.
Let $E_0 = \bigcup_{k=1}^{M} E_{0,k}$ denote the event corresponding to encoding errors, where
$E_{0,k}$ represents the error event at encoder $k$, for $k = 1, \ldots, M$. An encoding error at encoder $k$ occurs when in some block $j$ there is no codeword $V_{n,k,j-1}^{n}(t_{k,j-1}|1,1)$ satisfying the joint typicality rule (3.33). Therefore, the error event $E_{0,k}$ can be written as the union

$$E_{0,k} = \bigcup_{j=1}^{b} \left\{ (S_{k,j-1}^{n}, V_{n,k,j-1}^{n}(t_{k,j-1}|1,1), X_{n,k,j}^{n}(1,1)) \notin T_{c}^{n}, \text{for all } t_{k,j-1} \in [1:2^{nR_{k}}] \right\}.$$  

(3.35)

In order to define the decoding error events, we first define the event $E_{w}$ indexed by a message tuple $w = (w_1, \ldots, w_M)$ as

$$E_{w} = \left\{ \bigcap_{j=1}^{b} \left\{ \begin{array}{l}
X_{1,j}^{n}(w_1, t_{1,j-1}), \ldots, X_{M,j}^{n}(w_{M}, t_{M,j-1}), \\
V_{1,j}^{n}(t_{1,j} | w_1, t_{1,j-1}), \ldots, V_{M,j}^{n}(t_{M,j} | w_{M}, t_{M,j-1}), Y_{j}^{n}
\end{array} \right\} \in T_{c}^{n}, \text{for all } j \in [1:b] \text{ and for some } t_{b}^{\Delta} = (t_{1}, \ldots, t_{b}) \right\}.$$  

(3.36)

where we have defined that $t_{j} = (t_{1,j}, t_{2,j}, \ldots, t_{M,j})$ and $t_{k,j} \in \left[1:2^{nR_{k}}\right]$, $k = 1, \ldots, M$. Event $E_{w}$ occurs when the decoder finds a message tuple $w$ satisfying the decoding rule (3.34). Based on the decoding rule (3.34), the decoding error event can thus be expressed as the union $E_{1,M}^{c} \cup \{ \bigcup_{w \neq 1_{M}} E_{w} \}$. 

Overall, by considering both encoding and decoding errors and leveraging the union bound, the probability of error can be upper bounded as

\[
\Pr (E) \leq \sum_{k=1}^{M} \Pr \left( E_{0,k} \right) + \Pr \left( E_{1,M}^c \cap E_0^c \right) + \sum_{w \neq 1_M} \Pr (E_w). \tag{3.37}
\]

We now consider separately the terms in the sum (3.37).

1) By the covering lemma\cite{42}, we have the limit \( \Pr \left( E_{0,k} \right) \to 0 \) as long as the inequality

\[
\tilde{R}_k > I (S_k; V_k | X_k) + \delta(\epsilon') \tag{3.38}
\]

holds for sufficiently large \( n \), where \( \delta(\epsilon') \to 0 \) as \( \epsilon' \to 0 \).

2) By the conditional joint typicality lemma\cite{42}, we have that \( \Pr \left( E_{1,M}^c \cap E_0^c \right) \to 0 \) for sufficiently large \( n \).

3) To bound each term in the third summand in (3.37), for convenience, for any given \( w \neq 1_M \), \( t_j = (t_{1,j}, \ldots, t_{M,j}) \) and \( t_{j-1} = (t_{1,j-1}, \ldots, t_{M,j-1}) \), we define the event \( A_j(w, t_j, t_{j-1}) \) as

\[
A_j(w, t_j, t_{j-1}) = \left\{ \begin{array}{c}
X_{1,j}^n(w_1, t_{1,j-1}), \ldots, X_{M,j}^n(w_M, t_{M,j-1}), \\
V_{1,j}^n(t_{1,j}| w_1, t_{1,j-1}), \ldots, V_{M,j}^n(t_{M,j}| w_M, t_{M,j-1}), Y_j^n
\end{array} \right\} \in \mathcal{T}_\epsilon^n. \tag{3.39}
\]
From (3.36), we have the following

\[
\Pr (E_w) = \Pr \left( \bigcup_{t^b} b \bigcap_{j=1}^{b} A_j(w, t_j, t_{j-1}) \right) \quad (3.40a)
\]

\[
\leq \sum_{t^b} \Pr \left( \bigcap_{j=1}^{b} A_j(w, t_j, t_{j-1}) \right) \quad (3.40b)
\]

\[
\leq \sum_{t^b} \prod_{j=2}^{b} \Pr \left( A_j(w, t_j, t_{j-1}) \right), \quad (3.40c)
\]

where the union and sums over \( t^b \) are taken over all vectors \( t^b \) as defined in (3.36); and (3.40c) holds due to the independence of the codebooks generated for each block, the memoryless property of the channel and the fact that \( 0 \leq \Pr (A_1) \leq 1 \).

Next, we provide an upper bound on the probability \( \Pr \left( A_j(w, t_j, t_{j-1}) \right) \) for a given tuple \((w, t_j, t_{j-1})\). To facilitate the analysis, we introduce some useful notation. Specifically, for any given pair of vectors \((w, t_{j-1})\) with \( j \in [2 : b] \), we define the index set \( S_j(w, t_{j-1}) \), where we will drop the dependence on the arguments where necessary to simplify the notation. This set contains all the indices \( k \) for which at least one of the conditions \( w_k \neq 1 \) and \( t_{k,j-1} \neq 1 \) is satisfied for the pair of vectors \((w, t_{j-1})\), i.e.,

\[
S_j(w, t_{j-1}) = \left\{ k \in [1 : M] : w_k \neq 1 \text{ or } t_{k,j-1} \neq 1 \right\}. \quad (3.41)
\]

In addition, let \( S^c_j(w, t_{j-1}) \) denote the complement of \( S_j(w, t_{j-1}) \) with respect to the set \([1 : M]\), i.e.,

\[
S^c_j(w, t_{j-1}) = \left\{ k \in [1 : M] \setminus S_j(w, t_{j-1}) \right\}. \quad \text{Furthermore, we partition}
\]
the set $S^c_j(w, t_{j-1})$ into two subsets as follows:

$$S'_j(w, t_{j-1}, t_j) = \{ k \in S^c_j(w, t_{j-1}) : t_{k,j} \neq 1 \}, \quad (3.42a)$$

and

$$S''_j(w, t_{j-1}, t_j) = \{ k \in S^c_j(w, t_{j-1}) : t_{k,j} = 1 \}. \quad (3.42b)$$

By definition, we have that $S'_j(w, t_{j-1}, t_j) \cup S''_j(w, t_{j-1}, t_j) = S^c_j(w, t_{j-1})$. Finally, for a generic set $A_j \subseteq [1 : M]$, we define as $X(A_j)$ to be the set of variables $X_{k,j}$, for $k \in A_j$, where $X_{k,j}$ is the symbol transmitted by the $k$th user in the $j$th block. We use similar definition for $V(A_j)$.

Given the above notation and by the codebook construction, we use standard arguments on joint typicality [42] to obtain

$$\Pr\left(A_j \left(w, t_j, t_{j-1}\right)\right)$$

$$\leq 2^{-n} \left( H\left(X(S_j), V(S_j)\right) + H\left(V(S'_j) | X(S'_j)\right) + H\left(X(S^c_j), V(S''_j), Y\right) - H(X(S^c_j), X(S'_j), V(S''_j), Y) - \delta(\epsilon) \right)$$

$$= 2^{-n} \left( I\left(X(S_j), V(S_j) : X(S'_j), V(S''_j), Y | V(S'_j)\right) + I\left(V(S'_j) : X(S'_j), V(S''_j), Y\right) - \delta(\epsilon) \right)$$

$$\leq 2^{-n} \left( I\left(X(S_j), V(S_j) : X(S^c_j), V(S''_j), Y | V(S'_j)\right) - \delta(\epsilon) \right)$$

$$= 2^{-n} \left( I\left(X(S_j), V(S_j) : Y | V(S^c_j), V(S'_j)\right) - \delta(\epsilon) \right)$$

$$= 2^{-n} \left( I\left(X(S_j), V(S_j) : Y | X(S^c_j), V(S'_j)\right) - \delta(\epsilon) \right),$$
where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$; (3.43b) follows from standard steps involving mutual information; (3.43c) holds because $S'_j \subseteq S'_j$ so that $I(V(S'_j); X(S'_j), V(S''_j), Y) \geq I(V(S'_j); X(S'_j)); (3.43d)$ holds because of the fact that the tuple $(X(S_j), V(S_j))$ is independent of the tuple $(V(S'_j), X(S'_j), V(S''_j));$ and finally (3.43e) is due to the fact that $S'_j \cup S''_j = S'_j$.

It is noted that the upper bound of (3.43e) depends only on the sets $S_j(w, t_{j-1})$ and $S'_j(w, t_{j-1})$, and hence it is independent of $t_j$ for any given $w$ and $t_{j-1}$.

Given this upper bound, we then proceed with (3.40c) and obtain the following

$$\Pr(E_w)$$

$$\leq \sum_{t^b} \prod_{j=2}^b \Pr(A_j(w, t_j, t_{j-1}))$$

$$= \sum_{t_b} \sum_{t^{b-1}} \prod_{j=2}^b \Pr(A_j(w, t_j, t_{j-1}))$$

$$\leq \sum_{t_b} \sum_{t^{b-1}} \prod_{j=2}^b 2^{-\delta(\epsilon)} \left( I(X(S_j(w,t_{j-1})), V(S_j(w,t_{j-1})); Y|X(S'_j(w,t_{j-1})), V(S''_j(w,t_{j-1}))) \right)$$

$$= \sum_{t_b} \prod_{j=2}^b \sum_{t_{j-1}} 2^{-\delta(\epsilon)} \left( I(X(S_j(w,t_{j-1})), V(S_j(w,t_{j-1})); Y|X(S'_j(w,t_{j-1})), V(S''_j(w,t_{j-1}))) \right)$$

$$\leq 2^n \sum_{k \in [1:M]} \tilde{R}_k \left( \sum_{S \subseteq [1:M]} \sum_{t_{j-1}} 2^{-n(I(S_j)-\delta(\epsilon))} \right)^{b-1}$$

(3.44a) (3.44b) (3.44c) (3.44d) (3.44e)
\[ \leq 2^n \sum_{k \in [1:M]} \tilde{R}_k \left( \sum_{S \subseteq [1:M]: \mathcal{T}(w) \subseteq S} 2^n \sum_{l \in S} \tilde{R}_l 2^{-n(I(S) - \delta(\epsilon))} \right)^{b-1} \]  

(3.44f)

\[ \leq 2^n \sum_{k \in [1:M]} \tilde{R}_k \left( 2(M-1) 2^{-n(I_{\text{min}} - \delta(\epsilon))} \right)^{b-1}, \]  

(3.44g)

\[ = 2^n \sum_{k \in [1:M]} \tilde{R}_k \left( 2^{(b-1)(M-1) - n(b-1)(I_{\text{min}} - \delta(\epsilon))} \right) \]  

(3.44h)

where (3.44c) follows from (3.43e); (3.44d) holds because of the fact that the upper bound (3.43e) is independent of \( t_j \) for any given \( w \) and \( t_{j-1} \); (3.44e) also follows from (3.43e), where we have defined the index set \( \mathcal{T}(w) = \{ k \in [1:M] : w_k \neq 1 \} \) and \( I(S) = I(X(S), V(S); Y | X(S^c), V(S^c)) \); (3.44f) follows by \( t_{l,j-1} \in [1:2^n\tilde{R}_l] \) for any \( l \in S \); and (3.44g) holds because there are at most \( 2^{(M-1)} \) subsets of \([1:M]\) that contain any index set \( \mathcal{T}(w) \) given, where we have defined the term

\[ I_{\text{min}} = \min_{S \subseteq [1:M]: \mathcal{T}(w) \subseteq S} \left( I(S) - \sum_{l \in S} \tilde{R}_l \right). \]  

(3.45)

In this way, we obtain that

\[ \sum_{w \neq 1_M} \Pr(E_w) \leq \sum_{T \subseteq [1:M]} 2^n b \sum_{k \in T} R_k \left( n \sum_{k \in [1:M]} \tilde{R}_k + (b-1)(M-1) - n(b-1)(I_{\text{min}} - \delta(\epsilon)) \right). \]  

(3.46)
Therefore, we conclude that the limit
\[ \sum_{w \neq 1_M} \Pr (E_w) \to 0 \]
holds as long as the condition:
\[ nb \sum_{k \in T} R_k + n \sum_{k \in [1:M]} \tilde{R}_k + (b - 1) (M - 1) < n (b - 1) (I_{\text{min}} - \delta(\epsilon)), \forall T \subseteq [1:M], \]
\[ (3.47) \]
is satisfied, or equivalently we have
\[ \sum_{k \in T} R_k < \frac{(b - 1)}{b} (I_{\text{min}} - \delta(\epsilon)) - \frac{\sum_{k \in [1:M]} \tilde{R}_k}{b} - \frac{(b - 1) (M - 1)}{nb}, \forall T \subseteq [1:M]. \]
\[ (3.48) \]
Setting \( b \to \infty \) and \( n \to \infty \), we then have the condition
\[ \sum_{k \in T} R_k < I_{\text{min}} = \min_{S \subseteq [1:M]: T \subseteq S} \left( I(X(S), V(S); Y | X(S^C), V(S^C)) - \sum_{l \in S} \tilde{R}_l \right) \]
\[ (3.49a) \]
\[ \leq \min_{S \subseteq [1:M]: T \subseteq S} \left( I(X(S), V(S); Y | X(S^C), V(S^C)) - \sum_{l \in S} I(V_l; S_l | X_l) \right) \]
\[ (3.49b) \]
for all \( T \subseteq [1:M] \). This completes the proof of Theorem 3.5.

### 3.8.3 Proof of Theorem 3.6

**Achievability:**

The key idea of the achievable scheme is based on a variation of Schalkwijk-Kailath coding [60, 55]. User 1 divides its power into two parts. Specifically, it consumes
fraction \( \alpha (0 \leq \alpha \leq 1) \) of its power to send its message \( w_1 \) over \( n \) channel uses using a codeword drawn from a codebook whose entries are generated in an i.i.d. fashion from a zero-mean Gaussian distribution with variance \( \alpha P_1 \). Moreover, it uses the remaining portion \((1-\alpha)P_1\) to transmit state information via in cooperation with user 2, as detailed below.

**Codebook Generation:** Randomly generate \( 2^{nR_1} \) i.i.d. sequences \( x_{1p}^{n} \) with each component distributed as \( x_{1p,i} \sim \mathcal{N}(0, \alpha P_1) \), for \( i = 1, \ldots, n \). Index the sequences by \( x_{1p}^{n}(w_1) \) with \( w_1 \in [1 : 2^{nR_1}] \). Partition the interval \([-1 : 1]\) into \( 2^{nR_2} \) small intervals of equal length and map messages \( w_2 \in [1 : 2^{nR_2}] \) to the middle points of these intervals. Index these middle points by \( \theta (w_2) \).

**Encoding:**

1. **Initial channel use, \( i = 0 \):** User 1 keeps silent in this channel use. To send message \( w_2 \) to the receiver, user 2 transmits \( \theta (w_2) \);

2. **First channel use, \( i = 1 \):** By feedback, user 2 learns state \( s_0 \) after subtracting its own information. Since user 1 knows \( s_0 \) as well, it cooperates with user 2 to convey information about state \( s_0 \) to the receiver, superimposed on its private message \( w_1 \). Specifically, user 1 transmits \( x_{1,1} = x_{1p,1}(w_1) + \gamma_{1,1}s_0 \), where the scalar \( \gamma_{1,1} \) is chosen so that \( \gamma_{1,1}s_0 \sim \mathcal{N}(0, (1-\alpha)P_1) \), while user 2 transmits \( x_{2,1} = \gamma_{2,1}s_0 \), where the scalar \( \gamma_{2,1} \) is chosen so that \( \gamma_{2,1}s_0 \sim \mathcal{N}(0, P_2) \);

3. **Channel uses \( i \geq 2 \):** At each following channel use \( i \), user 2 forms the minimum mean squared error (MMSE) estimate \( \mathbb{E} \left[ s_0 \mid y_1^{i-1} \right] \) of \( s_0 \) based on the observed output symbols \( y_1^{i-1} \) at the beginning of channel use \( i \). Let \( s_{i-1}' = s_0 - \mathbb{E} \left[ s_0 \mid y_1^{i-1} \right] \).
Then user 2 transmits $x_{2,i} = \gamma_{2,i}s_{i-1}'$ over the channel use $i$, where the scalar $\gamma_{2,i}$ is selected so that $\gamma_{2,i}s_{i-1}' \sim \mathcal{N}(0, P_2)$. Given the fact that user 1 knows $s_0$, the outdated channel state $s_{i-1}'$ and its own message symbols, it equivalently knows the channel output symbols from the first channel use up to current time. Hence it can also generate the MMSE estimate of $s_0$ and thus $s_{i-1}'$ as done by user 2. User 1 then transmits $x_{1,i} = x_{1p,i}(w_1) + \gamma_{1,i}s_{i-1}'$ in channel use $i$, where the scalar $\gamma_{1,i}$ is chosen so that $\gamma_{1,i}s_{i-1}' \sim \mathcal{N}(0, (1-\alpha)P_1)$.

**Decoding:** After $n + 1$ channel uses, the receiver first estimates state $s_0$ by $\hat{s}_0 = \mathbb{E}\left[s_0 \mid y_1^n\right]$; it then estimates $\theta(w_2)$ by $\hat{\theta} = y_0 - \hat{s}_0 = \theta(w_2) + (s_0 - \mathbb{E}\left[s_0 \mid y_1^n\right])$ and declares that message $\hat{w}_2$ is sent if $\theta(\hat{w}_2)$ is the closest message point to $\hat{\theta}$. After successfully estimating state $s_0$ and decoding message $w_2$, the receiver is able to retrieve the information about $s_0$, which is conveyed from both users, so as to subtract it from the received sequence $y_1^n$. In this way, message $w_1$ is decoded based on the residual signal.

**Analysis of Probability of Error:** We note that using the union bound, we have, 

$$\Pr(E) \leq \Pr(E_2) + \Pr\left(E_1 \mid E_2^c\right),$$

where the first term corresponds to the probability of decoding error for message $w_2$, and the second term is the probability of decoding error for message $w_1$ given that message $w_2$ is correctly decoded. The probability of decoding error $\Pr(E_2)$ vanishes as the variance of estimation error of $s_0$ becomes arbitrarily small as $n \to \infty$. Similar to [55], it can be shown that we have $\Pr(E_2) \to 0$ as long as

$$R_2 \leq C\left(\frac{(1-\alpha)P_1 + P_2 + 2\sqrt{(1-\alpha)P_1P_2}}{\sigma^2_s + \alpha P_1}\right). \quad (3.50)$$
Moreover, from the standard consideration, we have $\Pr \left( E_1 \mid E_2^c \right) \to 0$ as long as the inequality

$$R_1 \leq C \left( \frac{\alpha P_1}{\sigma^2_s} \right)$$

(3.51)

holds. Setting $\alpha \Delta = (1 - \rho^2)$ concludes the proof of the achievability.

It is remarked that the achievability can also be proved by extending the scheme proposed in [23, page 15]. This scheme demonstrates that it is enough for both users to know the initial state symbol $s_0$, which can be accomplished by user 2 via feedback, in order to achieve the rate region of Theorem 3.6.

Converse:

Providing message $w_2$ to encoder 1, the channel becomes the MAC studied in [61] where encoder 1 knows both $w_1$ and $w_2$, encoder 2 knows $w_2$ and output feedback is available at the encoders. In fact, the state sequence at encoder 1 in this genie-aided model can be seen as equivalent to feedback, since via feedback, encoder 1 effectively obtains $s_{i-1}$. It is shown in [61] that feedback does not increase capacity and thus the capacity region is given by (3.27).
Chapter 4

Message and State Cooperation in a Relay Channel

4.1 Introduction

Relaying is one of the promising technologies to improve the performance of wireless communication systems in terms of coverage, throughput, and energy/cost. The first three-terminal relay channel was introduced by van der Meulen [43] in 1971. Later, in reference [44], Cover and El Gamal devised several coding schemes for the same channel. These coding schemes, along with several new variations, have been summarized in Section 2.4 of Chapter 2.

For the relay channel with state, however, research arises only recently and results are few. Reference [30] investigated the case of non-causal state information at the relay, and proposed a coding scheme that combines the strategies of decode-and-forward [44] and precoding against the state, while reference [31] studied the case of non-causal state information at the source and proposes various achievable schemes. Capacity results are identified for some special cases in [31]. Instead, when the state information is causally known to the relay, reference [32] derived achievable rates by combining the ideas of compress-and-forward [44] and adapting input codewords to the state (also known as Shannon strategies [4]).

This work also focuses on a state-dependent relay channel, but unlike previous work, assumes that state information is available only at the relay in a strictly causal
fashion. This scenario is more relevant in practical scenarios since in practice the state can be learned only in a strictly causal way. For instance, in the case of an interference network, an interfering sequence can be learned as it is observed, and, thus, in a strictly causal manner. With strictly causal state information, the strategies leveraged in [30, 32], for example, of precoding against the state or Shannon strategies cannot be applied. More fundamentally, the question arises as to whether strictly causal, and thus outdated, state information may be useful at all in a memoryless channel with i.i.d. state sequence.

As seen in references [23, 24] and our contribution in Chapter 3, for the MACs with independent (or common) state information available strictly causally at the encoders, capacity gains can be accrued by leveraging this information. Motivated by these results, in this chapter, we would like to assess the merits of strictly causal state information in a relay channel. Specifically, we consider a three-node relay channel where the source and relay are connected via two out-of-band orthogonal links of finite capacity, and a state-dependent memoryless channel connects the source and relay, on one side, and the destination, on the other. The source and destination have no state information, while the relay has access to the state information in a strictly causal manner. The channel model is shown in Fig. 4.2. This model is related to the class of relay channels, that are not state-dependent, with orthogonal links from the source to the relay and from the source and relay to the destination investigated by El Gamal and Zahedi [62]. In fact, in the scenario under study, we simplify the link from the source to the relay by modeling it as a noiseless finite-capacity link, while adding a similar backward relay-to-source link. Cooperation as enabled by orthogonal noiseless links, also referred to as conferencing, was first introduced by Willems [63] for a two-user MAC channel and was later extended
to several settings [64, 65, 66]. It is noted that, in practice, orthogonal links can be realized if nodes are connected via a number of different radio interfaces or wired links [3].

As an example, our model fits a downlink communication scenario in a cellular network where femtocells are overlaid on a microcell as shown in Fig. 4.1. Femtocells are served by home base stations, which are typically located around high user-density hot spots, that can serve as intermediate nodes or relays between users and the mobile operator network, to provide better indoor voice service or data delivery for stationary or low-mobility home users [39]. The home base station is typically connected to the outdoor base station via an out-of-band wired link, e.g., a last-mile connection followed by the Internet. The home base station may be able to measure the interference created by outdoor users, whereas this may not be possible at the base station or at indoor users. This gives rise to the system model we consider in this chapter, as can be readily observed from Figs. 4.1 and 4.2.

In the considered model, cooperation between source and relay through the conferencing links can aim at two distinct goals: 

i) Message transmission: Through the source-to-relay link, the source can provide the relay with some information about the message to be conveyed to the destination, thus enabling message cooperation; 

ii) State transmission: Through the relay-to-source link, the relay can provide the source with some information about the state, thus enabling cooperative transmission of the state information to the destination. We propose two achievable schemes, one based on conventional block Markov coding [44] and backward decoding [52] and one inspired by noisy network coding. We show that the latter outperforms the former in general. Moreover,
Fig. 4.1. The downlink transmission to a home user in a femtocell provides an example application of the considered model illustrated in Fig. 4.2. The home base station is assumed to be able to measure the interference from outdoor users.

Based on these achievable rates, we identify capacity results for some special cases of the considered model. We also investigate the optimal capacity allocation between the source-to-relay and relay-to-source links where the total conferencing capacity is fixed. Finally, we derive achievable rates and some capacity results for the Gaussian version of the system at hand and elaborate on numerical results.

The remainder of this chapter is organized as follows. Section 4.2 formally describes the relay model considered in this work. Section 4.3 illustrates two different achievable coding schemes and presents the resulting achievable rates. Section 4.4 identifies capacity results for some special cases. Section 4.5 studies the scenario in which the total conferencing capacity is fixed and elaborates on optimal capacity allocation. Section 4.6 studies the Gaussian case of our model and provides numerical results. Section 4.7 concludes the work.
4.2 System Model

In this section, we present the channel model and provide relevant definitions. As depicted in Fig. 4.2, we study a three-node relay channel where the source and relay are connected via two unidirectional out-of-band orthogonal links of finite capacity, while there is a state-dependent memoryless channel between the source and relay, on one side, and the destination, on the other. Note that the relay transmits and receives simultaneously over two orthogonal channels.

The channel is characterized by the tuple:

\[(\mathcal{X} \times \mathcal{X}_R, \mathcal{S}, \mathcal{Y}, p(s), p(y|s,x,x_R), C_{SR}, C_{RS})\]

with source input alphabet \(\mathcal{X}\), relay input alphabet \(\mathcal{X}_R\), destination output alphabet \(\mathcal{Y}\) and channel state alphabet \(\mathcal{S}\). The capacity per channel use of the source-to-relay and relay-to-source out-of-band, also known as conferencing [63], links are given by \(C_{SR}\), \(C_{RS}\) respectively. The state sequence is assumed to be i.i.d., i.e., \(p(s^n) = \prod_{i=1}^{n} p(s_i)\). The relay channel is discrete memoryless (DM) in the sense that at any discrete time
We assume that state information is available to the relay in a *strictly causal* manner while there is no state information at the source and destination.

**Definition 4.1.** Let $W$, uniformly distributed over the set $W = [1 : 2^{nR}]$, be the message sent by the source. A $(2^{nR}, n)$ code consists of:

1. Conferencing codes: Conferencing mappings are defined as

$$h_{SR,i} : W \times T_{RS}^{i-1} \rightarrow T_{SR,i}, \quad (4.3a)$$

$$h_{RS,i} : S^{i-1} \times T_{SR}^{i-1} \rightarrow T_{RS,i}, \quad (4.3b)$$

where $(4.3a)$ generates the $i$th symbol sent on the source-to-relay link based on the message and all symbols previously received on the relay-to-source link, while $(4.3b)$ generates the $i$th symbol sent on the relay-to-source link based on the strictly causal states and all symbols previously received on the source-to-relay link. Note that $T_{SR,i}$ is the alphabet of the conferencing message sent from the source to the relay, while $T_{RS,i}$ is the alphabet of the conferencing message sent from the relay to the source at time instant $i$, $i = 1, ..., n$. Such mappings are permissible if the following capacity-conserving conditions are satisfied:

$$\frac{1}{n} \sum_{i=1}^{n} \log_2 |T_{SR,i}| \leq C_{SR}. \quad (4.4a)$$
\[
\frac{1}{n} \sum_{i=1}^{n} \log_2 \left| T_{RS,i} \right| \leq C_{RS}.
\] 

(4.4b)

2. Encoder mappings at the source:

\[ f_i : \mathcal{W} \times T_{RS}^i \rightarrow \mathcal{X}, \forall i = 1, ..., n, \] 

(4.5)

which generates the channel input at the source at time \( i \) based on the message and the information received from the relay up to and including time \( i \) on the relay-to-source link.

3. Encoder mappings at the relay:

\[ f_{R,i} : \mathcal{S}^{i-1} \times T_{SR}^i \rightarrow \mathcal{X}_R, \forall i = 1, ..., n, \] 

(4.6)

which generates the channel input at the relay at time \( i \) based on the strictly causal state information and the information received from the source up to and including time \( i \) on the source-to-relay link.

4. Decoder mapping at the destination:

\[ g : \mathcal{Y}^n \rightarrow \mathcal{W}, \] 

(4.7)

which produces the estimate of message at the destination based on the received sequences.
The average probability of error, $\Pr(E)$, is defined by:

$$\Pr(E) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \Pr(g(y^n) \neq w \mid w \text{ sent}).$$

(4.8)

A rate $R$ is achievable if there exists a sequence of codes $(2^{nR}, n)$ as defined above such that the probability of error $\Pr(E) \to 0$ as $n \to \infty$. The capacity of this channel is the supremum of the set of all achievable rates.

### 4.3 Achievable Schemes and Upper Bound

In this section, we demonstrate two different coding schemes that exploit the potential benefits of message and state cooperation between source and relay. We also identify a simple upper bound on the capacity.

#### 4.3.1 Scheme 1: Block-based Message and State Cooperation

We first propose an achievable scheme based on conventional block Markov coding and backward decoding.

**Proposition 4.1.** For the DM state-dependent relay channel of Fig. 4.2, any non-negative rate smaller than $R_1$ is achievable where

$$R_1 = \max_{P_1} \min_{\mathcal{F}_1} \begin{pmatrix} I(X;Y|X_R,V,U) + C_{SR}, \\ I(X,X_R;Y|V) - I(V;S|Y), \\ I(X,X_R;Y|V,U) + C_{SR} + C_{RS} - I(V;S|Y) \end{pmatrix}$$

(4.9)
with the maximum taken over the distributions in the set of

\[
\mathcal{P}_1 = \{ p(v, u, s, x, x_R, y) : p(s)p(v|s)p(u)p(x|u)p(x_R|u)p(y|s, x, x_R) \}
\]  

subject to the constraint:

\[
I(X_R; Y | X, U) + \min (C_{RS}, I(X, U; Y)) \geq I(V; S | Y).
\]  

Sketch of Proof: The idea is to follow a natural block Markov strategy. Specifically, the message \( w \) is split by the source into \((b - 1)\) parts, \((w_1, ..., w_{b-1})\), \( w_j \in [1 : 2^{nR_1}] \), \( j = 1, ..., (b - 1) \), which are transmitted over \( b \) blocks, each block consisting of \( n \) channel uses where \( n = \frac{m}{b} \), and \( m \) is the number of total channel uses. At the end of each block, the relay compresses the state sequence that has affected the channel over the block with the aim of conveying such information to the destination in the next block. Compression is done via Wyner-Ziv coding [53] by exploiting the received sequence at the destination as side information. Conferencing takes place before the beginning of each block. Specifically, through conferencing, before the \( j \)th block, the source conveys part of the message \( w_j \) to the relay in order to enable message cooperation, while the relay sends part of the bin index produced by Wyner-Ziv coding [53, 44] to the source to enable cooperative state transmission. The exchange state and message information is sent cooperatively by the source and relay, while the remaining part of the message \( w_j \) is sent independently by the source and the remaining part of the bin index is sent by the relay alone. This strategy is referred to as block-based message and state cooperation. Decoding takes place by backward decoding [52]. Specifically, starting from the last
reception, the destination first retrieves the compressed state information for block \((b-1)\).

After that, it recovers message \(w_{b-1}\) by jointly decoding based on the \((b-1)\)th block reception and the state information retrieved, i.e., by a coherent decoding. Using the decoded message, the destination turns to retrieve the compressed state information for block \((b-2)\), and then decodes the corresponding message \(w_{b-2}\). Repeating this operation until back to the first block, the destination recovers all the messages over blocks. The transmission scheme as described is illustrated in Fig. 4.3. Details of the proof are provided in Appendix 4.8.1.
Remark 4.1. To interpret (4.9) to (4.11) in light of the transmission strategy discussed above for scheme 1, we remark that $V$ represents the compressed state information and $U$ accounts for the codeword transmitted cooperatively by the source and relay, which conveys both state and message information they share. Bound (4.11) imposes that the Wyner-Ziv rate $I(V;S|Y)$ is supported by the cooperative transmission of the source and relay, whose rate is limited by $\min(C_{RS}, I(X,U;Y))$ and the information sent independently from the relay $I(X_R;Y|X,U)$. The mutual information terms in (4.9), and in particular the conditioning on $V$, account for the fact that the destination has information about the channel via the compressed state $V$, which allows for partial or complete coherent decoding. Moreover, the second and third term in (4.9) reflect the cost in terms of rate to be paid for the transmission of compressed state information. \(\square\)

### 4.3.2 Scheme 2: Burst Message Cooperation and Block-based State Cooperation

In this subsection, we propose a second transmission scheme inspired by noisy network coding [50].

**Proposition 4.2.** For the DM state-dependent relay channel of Fig. 4.2, any non-negative rate smaller than $R_2$ is achievable where

\[
R_2 = \max_{p_2} \min \left( \begin{array}{c}
I(X;Y|X_R,V,U) + C_{SR}, \\
I(X,X_R,V;Y) - I(V;S|X_R,U), \\
I(X,X_R,V;Y|U) + C_{SR} + C_{RS} - I(V;S|X_R,U)
\end{array} \right)
\]  

(4.12)
with the maximum taken over the distributions in the set of

\[ P_2 = \{ p(v, u, s, x, x_R, y) : p(s)p(v | s, x, x_R, u)p(x | u)p(x_R | u)p(y | s, x, x_R) \} \] (4.13)

**Sketch of Proof:** Inspired by the noisy network coding scheme in [50], we use a long-message encoding strategy in our second scheme. Specifically, the same message \( w \), \( w \in [1 : 2^{nbR_2}] \), is sent at the source over all \( b \) blocks of transmission with each consisting of \( n \) channel uses. Thus, unlike scheme 1 discussed above, here information exchange about the message between source and relay takes place only once at the beginning of the first block. This way, the source shares part of the message \( w \) with the relay in order to enable message cooperation. As for the state, at the end of each block, the relay compresses the state sequence over the block \textit{without explicit} Wyner-Ziv coding, that is, without binning [50]. Moreover, \textit{the compression is done in a way that also depends on the past codeword}. Exchange of state information between relay and source takes place before the beginning of each block as for scheme 1 proposed above. Source and relay cooperatively send the message and state information they share, while the source sends the remaining part of the message independently and the relay sends the remaining part of the compression index alone for each block. This transmission scheme is referred to as burst message cooperation and block-based state cooperation strategy. At the end of \( b \) blocks of transmission, the destination performs \textit{joint decoding over all blocks of reception without explicitly decoding the compressed state information}. The transmission scheme as described is illustrated in Fig. 4.4. Details of the proof are provided in Appendix 4.8.2.
Compress the past state sequence without binning; Compression is done in a way that depends on the past codeword.

Fig. 4.4. Scheme 2: Burst message cooperation and block-based state cooperation.

Remark 4.2. To interpret (4.12) to (4.13) in light of the transmission strategy discussed above and in comparison to the one in scheme 1, we point out that, as in Remark 4.1, \( V \) represents the compressed state information while \( U \) denotes for the common message and state information. Each mutual information term in (4.12), in particular the conditioning on \( V \), has for a similar interpretation as explained in Remark 4.1.

Unlike scheme 1, however, the compressed state \( V \) is generated without explicit Wyner-Ziv coding and without requiring correct decoding of the compressed state at the receiver. This fact, as detailed in the proof, makes it possible to choose \( V \) to be dependent of \( X_R \), \( U \) and \( S \), instead of only \( S \) in scheme 1. Moreover, the rate loss due to the need to convey
state information can be smaller than \( I(V; S | Y) \) in (4.9), as discussed in Proposition 4.3. Finally, since the decoding is implemented jointly without recovering all the compressed states correctly in scheme 2, there is no explicit additional constraint (4.11). □

### 4.3.3 Comparison of Achievable Rates

Based on the discussion above, we expect scheme 2 of Proposition 4.2 to outperform scheme 1 of Proposition 4.1. This is shown by the following proposition.

**Proposition 4.3.** \( R_2 \geq R_1 \).

**Proof:** We prove the results by showing that the three terms in (4.12) are larger or equal than the ones in (4.9). This, coupled with the fact that the characterization of \( R_2 \) does not have additional constraint (4.11) and with the more general distribution \( p(v|s, x_R, u) \) allowed by scheme 2 over scheme 1 (which constrains the distribution as \( p(v|s) \)), is enough to conclude the proof. Specifically, setting \( p(v|s, x_R, u) = p(v|s) \) in \( P_2 \), we have that:

1. The first term in (4.12) is the same as the first term in (4.9).

2. The second terms are also equal since

\[
I(X, X_R; Y | V) - I(V; S | X_R, U) \quad (4.14a)
\]

\[
= I(X, X_R; Y | V) + I(V; Y) - I(V; S | X_R, U) \quad (4.14b)
\]

\[
= I(X, X_R; Y | V) + H(V | S) - H(V | Y) \quad (4.14c)
\]

\[
= I(X, X_R; Y | V) - I(V; S | Y), \quad (4.14d)
\]
where (4.14c) is because $V$ is independent of $(U, X_R)$, and (4.14d) follows from the Markov chain $V \leftrightarrow S \leftrightarrow Y$.

3. The third term of (4.12) is larger or equal than the corresponding term in (4.9) since

$$I(X, X_R; V; Y | U) + C_{SR} + C_{RS} - I(V; S | X_R, U)$$

(4.15a)

$$= I(X, X_R; Y | V, U) + C_{SR} + C_{RS} + I(V; Y | U) - I(V; S | X_R, U)$$

(4.15b)

$$= I(X, X_R; Y | V, U) + C_{SR} + C_{RS} + H(V | S) - H(V | Y, U)$$

(4.15c)

$$\geq I(X, X_R; Y | V, U) + C_{SR} + C_{RS} + H(V | S) - H(V | Y)$$

(4.15d)

$$= I(X, X_R; Y | V, U) + C_{SR} + C_{RS} - I(V; S | Y),$$

(4.15e)

where (4.15c) is because $V$ is independent of $(U, X_R)$, (4.15d) holds because conditioning reduces entropy, while (4.15e) again follows from the Markov chain $V \leftrightarrow S \leftrightarrow Y$.

4.3.4 An Upper Bound

Here we derive a simple upper bound.

**Proposition 4.4.** For the DM state-dependent relay channel of Fig. 4.2, the capacity is upper bounded by

$$R_{ub} = \max_{P_{ub}} \min \left( I(X, X_R; Y), I(X; Y | X_R, S) + C_{SR} \right)$$

(4.16)
with the maximum taken over the distributions in the set of

$$\mathcal{P}_{ub} = \{ p(s, x, x_R, y) : p(s)p(x, x_R)p(y \mid s, x, x_R) \}.$$  \hspace{1cm} (4.17)

**Proof:** The upper bound (4.16) is essentially a cut-set bound [57], where the first term corresponds to the MAC cut between source-relay and destination, and the second is the cut between source and relay-destination. Given presence of the state sequence, calculation requires some care and is detailed below.

For the first term, consider a genie-aided system in which the message is also provided to the relay and the state \( s^{i-1} \) is also provided to the source at time \( i \). The system can be now seen as being point-to-point with inputs \((X, X_R)\), output \( Y \) and with strictly causal state information. In this case, it is well known that state information does not increase capacity, which is given by the first term in (4.16). The result can also be seen from the Fano’ inequality [57] as

$$R_{ub} \leq \frac{1}{n} I(W; Y^n) + \epsilon_n$$ \hspace{1cm} (4.18a)

$$\leq \frac{1}{n} \sum_{i=1}^{n} I(W; Y_i \mid Y^{i-1}) + \epsilon_n$$ \hspace{1cm} (4.18b)

$$\leq \frac{1}{n} \sum_{i=1}^{n} I(W, Y^{i-1}; X_i, X_{R,i}; Y_i) + \epsilon_n$$ \hspace{1cm} (4.18c)

$$\leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; X_{R,i}; Y_i) + \epsilon_n$$ \hspace{1cm} (4.18d)

with \( \epsilon_n \to 0 \) as \( n \to \infty \), where (4.18c) follows from the non-negativity of mutual information, (4.18d) follows from the Markov chain \((W, Y^{i-1}) \leftrightarrow (X_i, X_{R,i}) \leftrightarrow Y_i\). This
Markov chain can be seen as a consequence of the independence of $S_i$ and $(W, Y^{i-1})$, and the Markov chain $(W, Y^{i-1}) \leftrightarrow (X_i, X_{R,i}, S_i) \leftrightarrow Y_i$.

For the second term, consider another genie-aided system in which the perfect state information is provided to the destination. Then, by the Fano’ inequality [57], we have

$$R_{ub} \leq \frac{1}{n} I(W; Y^n, S^n, T_{SR}^n) + \epsilon_n$$

$$= \frac{1}{n} I(W; Y^n, T_{SR}^n | S^n) + \epsilon_n$$

$$= \frac{1}{n} I(W; Y^n | S^n, T_{SR}^n) + \frac{1}{n} I(W; T_{SR}^n | S^n) + \epsilon_n$$

with $\epsilon_n \to 0$ as $n \to \infty$, where (4.19b) holds because $W$ and $S^n$ are independent and (4.19c) follows from the chain rule. Note that

$$\frac{1}{n} I(W; T_{SR}^n | S^n)$$

$$\leq \frac{1}{n} H(T_{SR}^n)$$

$$\leq \frac{1}{n} \sum_{i=1}^n H(T_{SR,i})$$

$$\leq C_{SR},$$

where (4.20b) is due to the fact that conditioning reduces entropy and (4.20c) follows from the definition of permissible conferencing mapping given by (4.4a). Moreover, we
have that

$$\frac{1}{n} I \left( W; Y^n \left| S^n, T^n_{SR} \right. \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} I \left( W; Y_i \left| Y^{i-1}, S^n, T^n_{SR} \right. \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} I \left( W; Y_i \left| Y^{i-1}, S^n, T^n_{SR}, X^n_R \right. \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[ H \left( Y_i \left| Y^{i-1}, S^n, T^n_{SR}, X^n_R \right. \right) - H \left( Y_i \left| W, Y^{i-1}, S^n, T^n_{SR}, X^n_R \right. \right) \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[ H \left( Y_i \left| Y^{i-1}, S^n, T^n_{SR}, X^n_R \right. \right) - H \left( Y_i \left| W, Y^{i-1}, S^n, T^n_{SR}, X^n_R, T^n_{RS}, X^n \right. \right) \right]$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left[ H \left( Y_i \left| X_{R,i}; S_i \right. \right) - H \left( Y_i \left| X_i, X_{R,i}; S_i \right. \right) \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} I \left( X_i; Y_i \left| X_{R,i}; S_i \right. \right),$$

(4.21f)

where (4.21b) holds because $X_{R,i}$ is a function of $(T^i_{SR}, S^{i-1})$, (4.21d) holds because $T_{RS,i}$ is a function of $(T^i_{SR}, S^{i-1})$ while $X_i$ is a function of $(W, T^i_{RS})$, (4.21e) follows from the memoryless property of the channel and the fact that conditioning reduces entropy. Overall, we have

$$R_{ub} \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i \left| X_{R,i}, S_i \right. ) + C_{SR} + \epsilon_n$$

(4.22)

with $\epsilon_n \to 0$ as $n \to \infty$.

Finally, from (4.18d) and (4.22), the proof is concluded using the standard approach of introducing a time-sharing variable $Q$ uniformly distributed in the set $[1 : n]$
and then arguing that one can set $Q$ to be constant without loss of optimality [57, Ch.15].

4.4 Special Cases and Capacity Results

In this section, we consider three special cases of the general model studied above, namely: i) No message and state cooperation, in which $C_{SR} = C_{RS} = 0$ and thus the source only transmits message and the relay only conveys state information to the destination; ii) Message cooperation only, in which $C_{SR} > 0, C_{RS} = 0$ so that cooperative message transmission between the source and relay is feasible; iii) State cooperation only, in which $C_{SR} = 0, C_{RS} > 0$ so that cooperative state transmission between the relay and source is feasible. We establish capacity results for a special class of channels for each case.

4.4.1 No Message and State Cooperation

Corollary 4.1. If $C_{SR} = C_{RS} = 0$, any non-negative rate smaller than $R_{21}$ is achievable where

$$R_{21} = \max_{\mathcal{P}_{21}} \min (I(X;Y|X_R,V), I(X,X_R,V;Y) - I(V;S|X_R))$$

(4.23)

with the maximum taken over the distributions in the set of

$$\mathcal{P}_{21} = \{ p(v,s,x,x_R,y) : p(s)p(v|s,x_R)p(x)p(x_R)p(y|s,x,x_R) \}.$$  \hspace{1cm} (4.24)
Proof: The achievable rate follows from $R_2$ (4.12) by setting $C_{SR} = C_{RS} = 0$ and $U = \emptyset$, since no information is shared between the source and relay.

This rate turns out to be optimal, i.e., capacity-achieving, for a special class of relay channels, which includes modulo-additive state-dependent relay channels, see Example 4.1.

**Proposition 4.5.** Let $\mathcal{P}_{21}^*$ denote the set of distributions defined by:

$$
\mathcal{P}_{21}^* = \{p(s, x, x_R, y) : p(s)p(x)p(x_R)p(y | s, x, x_R)\}.
$$

(4.25)

If $C_{SR} = C_{RS} = 0$,

$$
H(Y | X, X_R, S) = 0,
$$

(4.26a)

and

$$
H(S | X, X_R, Y) = 0
$$

(4.26b)

are satisfied for all distributions in $\mathcal{P}_{21}^*$, then the capacity is given by:

$$
C_{21} = \max_{\mathcal{P}_{21}^*} \min \left( H(Y | X_R, S), I(X, X_R; Y) \right).
$$

(4.27)

Proof: The achievability is straightforward by setting $V = S$ and applying assumptions (4.26a) and (4.26b) when evaluating (4.23). Specifically, we have

$$
I(X; Y | X_R, S) = H(Y | X_R, S),
$$

(4.28)
and

\[ I(X, X_R, S; Y) - I(S; S | X_R) \]  \hspace{1cm} (4.29a)

\[ = I(X, X_R; Y) + I(S; Y | X, X_R) - H(S) \]  \hspace{1cm} (4.29b)

\[ = I(X, X_R; Y) - H(S | X, X_R, Y) \]  \hspace{1cm} (4.29c)

\[ = I(X, X_R; Y). \]  \hspace{1cm} (4.29d)

To obtain a converse result, we follow from (4.16) and note the fact that \( X \) and \( X_R \) must be independent since source and relay cannot cooperate when \( C_{SR} = C_{RS} = 0 \). Hence, the capacity is upper bounded by (4.16) evaluated for some product input distribution \( p(x)p(x_R) \). Overall, we have:

\[ C_{21} \leq I(X; Y | X_R, S) = H(Y | X_R, S), \]  \hspace{1cm} (4.30a)

\[ C_{21} \leq I(X, X_R; Y) \]  \hspace{1cm} (4.30b)

for some input distribution \( p(x)p(x_R) \). The proof is concluded by maximizing the mutual information terms (4.30a) and (4.30b) over the same input distribution \( p(x)p(x_R) \).

\textbf{Remark 4.3.} Achievability of the capacity (4.25)–(4.27) has been proved above via scheme 2. The same capacity result \textit{cannot} be obtained by setting \( U = \emptyset, V = S \) in \( R_1 \) from scheme 1 of Proposition 4.1, since we have the additional constraint \( I(X_R; Y | X) \geq H(S | Y) \). This points to the advantage of the long-message encoding and joint decoding strategy used by scheme 2. \( \square \)
Remark 4.4. Condition (4.26a) basically states that, when fixed \(X\) and \(X_R\), there is no other source of uncertainty in the observation \(Y\) beside the state \(S\). Condition (4.26b), instead, says that the state \(S\) is perfectly determined when \(Y, X\) and \(X_R\) are known. These conditions guarantee that providing information about the state directly reduces the uncertainty about the input \(X\) and \(X_R\). The fact that the relay can increase the achievable rate up to \(I(X, X_R; Y)\) in (4.27) can be interpreted in light of this fact since the relay signal \(X_R\) directly contributes to the achievable rate even though the relay is not aware of the message. This will be further discussed in Remark 4.13 for a Gaussian model. \(\square\)

Example 4.1. Consider a binary modulo-additive state-dependent relay channel defined by

\[
Y = X \oplus X_R \oplus S,
\]

where \(S \sim Bernoulli(p_s)\). Let us further impose the cost constraints on the source and relay codewords \((x^n, x^n_R)\),

\[
\frac{1}{n} \sum_{i=1}^{n} E[X_i] \leq p, \quad \frac{1}{n} \sum_{i=1}^{n} E[X_{R,i}] \leq p_r
\]

with \(0 \leq p, p_r \leq \frac{1}{2}\). Extending the capacity result of Proposition 4.5 to channels with cost constraints is straightforward and leads simply to limiting the set of feasible distributions (4.25) by imposing the constraints that \(E[X] \leq p\) and \(E[X_R] \leq p_r\), see, e.g., [42].
Therefore the capacity is given by:

\[ C_{\text{bin}} = \min (H_b(p), H_b(p \ast p_r \ast p_s) - H_b(p_s)) \],

(4.33)

where \( p_1 \ast p_2 \) denotes the discrete convolution operation of two Bernoulli distributions with parameters \( p_1 \) and \( p_2 \), i.e., \( p_1 \ast p_2 = p_1(1 - p_2) + p_2(1 - p_1) \), and \( H_b(p) = -p \log_2 p - (1 - p) \log_2 (1 - p) \).

As a specific numerical example, setting \( p = p_r = 0.15 \) and \( p_s = 0.1 \), we have \( C_{\text{bin}} = 0.4171 \). Note that without state information at the relay, the channel can be considered as a relay channel with reversely degraded components in [44]. In this case, the best rate achieved is given by [44, Theorem 2]:

\[ C_{\text{bin}, \text{no SI}} = \max_{p(x)} \max_{x_R} I(X; Y | X_R = x_R) \]

\[ = H_b(p \ast p_s) - H_b(p_s) \]

\( (4.34a) \)

\( = 0.2912. \)

(4.34b)

Hence \( C_{\text{bin}} > C_{\text{bin}, \text{no SI}} \), which assesses the benefit of state information known at the relay even in a strictly causal manner.

Remark 4.5. The channel discussed in Example 4.1, has a close relationship with the modulo-additive state-dependent relay model considered by Aleksic, Razaghi and Yu in [67]. Therein, the relay observes a corrupted version of the noise (state) non-causally and has a separate and rate-limited digital link to communicate to the destination. For this class of channels, a compress-and-forward strategy is devised and shown to achieve
capacity. Unlike [67], the relay obtains the state information *noiselessly, strictly causally* and the relay-to-destination link is *non-orthogonal* to the source-to-destination link. We have shown in Proposition 4.5 that in this case, the proposed scheme 2 achieves capacity.

\\[\square\\]

### 4.4.2 Message Cooperation Only

With $C_{RS} = 0$, the model at hand is similar to the one studied in [62], where capacity was obtained for a *state-independent* channel in which a general noisy channel models the source-to-relay link. For this scenario, the optimal coding strategy was found to split the message into two parts, one decoded by the relay and sent cooperatively with the source to the destination and the other sent directly from the source to the destination. By setting $S = V = \emptyset$ and $C_{RS} = 0$ in (4.12), we recover a special case of the capacity obtained in [62] with noiseless source-to-relay link.

For *state-dependent* channels, a general achievable rate can be identified through $R_2$ in (4.12) by setting $C_{RS} = 0$. Moreover, when the source-to-relay capacity is large enough, we are able to characterize the capacity as follows. Notice that this capacity result holds for an arbitrary $C_{RS}$, not necessary $C_{RS} = 0$.

**Proposition 4.6.** Let $\mathcal{P}_{22}^*$ denote the set of distributions defined by:

$$
\mathcal{P}_{22}^* = \{ p(s, x, x_R, y) : p(s)p(x, x_R)p(y | s, x, x_R) \}.
$$

(4.35)
If $C_{SR} \geq \max_{\mathcal{P}_{22}} I(X, X_R; Y)$ and arbitrary $C_{RS}$, the capacity $C_{22}$ is given by:

$$C_{22} = \max_{\mathcal{P}_{22}} I(X, X_R; Y),$$

(4.36)

and is achieved by message cooperation only.

**Proof:** When $C_{SR} \geq C_{22}$, the source can share a message $w$ of rate $C_{22}$ with the relay through the conferencing link. By setting $U = X$ and $V = \emptyset$ in (4.12) and removing redundant bounds, we establish the achievability part. The converse part follows directly from (4.16).

**Remark 4.6.** The capacity identified above is the same as without any state information at the relay. This result implies that when the relay is cognizant of the entire message, message transmission always outperforms sending information about the channel states. In other words, no benefits can be reaped if the relay allocates part of its transmission resources to state forwarding. This can be seen as a consequence of the fact that in a point-to-point channel, no gain is possible by exploiting availability of strictly causal state information.

**Remark 4.7.** The capacity result of Proposition 4.6 has been proved by using scheme 2 for achievability. However, it *can* also be obtained with scheme 1 of Proposition 4.1 by setting $U = X$ and $V = \emptyset$. This may not be surprising since the two schemes differ most notably in the way state information is processed at encoder and decoder, and the capacity result of Proposition 4.6 is achieved with full message cooperation.
4.4.3 State Cooperation Only

If $C_{SR} = 0$, no cooperative message transmission is allowed. However, through the conferencing link of capacity $C_{RS}$, cooperative state transmission between the relay and source is still feasible. A general achievable rate can be identified from $R_2$ in (4.12) by setting $C_{SR} = 0$. Specifically, when $C_{RS}$ is large enough, we have the following corollary.

**Corollary 4.2.** Let $\mathcal{P}_{23}$ denote the set of distributions defined by:

$$\mathcal{P}_{23} = \{ p(s, v, x, x_R, y) : p(s)p(v | s, x_R)p(x, x_R)p(y | s, x, x_R) \}. \quad (4.37)$$

If $C_{SR} = 0$ and $C_{RS} \geq \max_{\mathcal{P}_{23}} I(X_R; Y)$, any non-negative rate smaller than $R_{23}$ is achievable where

$$R_{23} = \max_{\mathcal{P}_{23}} \min \{ I(X; Y | X_R, V), I(X, X_R; V, Y) - I(V; S | X_R) \}. \quad (4.38)$$

**Proof:** By setting $C_{SR} = 0$ and $U = X_R$ in (4.12), the set of $\mathcal{P}_2$ is specialized to $\mathcal{P}_{23}$. Fix any input distribution in $\mathcal{P}_{23}$. The first term in the min function of (4.12) is reduced to $I(X; Y | X_R, V)$ and the second term is reduced to $I(X, X_R; V, Y) - I(V; S | X_R)$. For the third term, it becomes:

$$I(X, X_R, V; Y | X_R) + C_{RS} - I(V; S | X_R) \quad (4.39a)$$

$$= I(X, V; Y | X_R) + C_{RS} - I(V; S | X_R) \quad (4.39b)$$

$$\geq I(X, V; Y | X_R) + I(X R; Y) - I(V; S | X_R) \quad (4.39c)$$
\[ I(X, X_R; V; Y) = I(V; S | X_R), \quad (4.39d) \]

where the inequality (4.39c) follows from the assumption on \( C_{RS} \). Notice that the third term cannot be smaller than the second term, hence it is redundant. Therefore, we establish the achievable rate given by (4.38).

The achievable rate (4.38) coincides with the upper bound (4.16) for the special class of relay channels characterized by (4.26a)–(4.26b).

**Proposition 4.7.** Let \( P_{23}^{*} = P_{22}^{*} \) of (4.35). If \( C_{SR} = 0, C_{RS} \geq \max_{P_{23}^{*}} I(X_R; Y) \), and (4.26a)–(4.26b) are satisfied for all distributions in \( P_{23}^{*} \), then the capacity is given by:

\[ C_{23} = \max_{P_{23}^{*}} \min \left( H(Y | X_R, S), I(X, X_R; Y) \right). \quad (4.40) \]

**Proof.** The result follows from Corollary 4.2. For the achievability, set \( V = S \) in (4.38) and apply assumptions (4.26a) and (4.26b) to obtain (4.40), by similar steps from (4.28) to (4.29d). The upper bounds follow from (4.16) and note that the second bound therein is reduced to \( H(Y | X_R, S) \) under assumption (4.26a).

**Remark 4.8.** Achievability of the capacity (4.40) has been proved above via scheme 2. It cannot be attained by scheme 1 because of the additional constraint required to support the transmission of compressed state information specified by (4.11).

**Remark 4.9.** Compared to the capacity result provided in Proposition 4.5 for the same class of channels (4.26a)–(4.26b), \( C_{23} \) is potentially larger because a general input distribution is admissible instead of the product input distribution due to state cooperation.
The resulting cooperative gain will be further discussed for the Gaussian model in Section 4.6.

Remark 4.10. The capacity result of Proposition 4.7 is derived for $C_{SR} = 0$ and is thus achieved by state cooperation only. Optimality of state cooperation only can also be concluded in some case when $C_{SR} > 0$ and thus message cooperation is possible. For instance, assume that $H(Y | X, S) \geq I(X, X_R; Y)$ for the distribution in $\mathcal{P}_{23}^*$ that maximizes (4.40). Then it can be proved, following the same bounds used in Proposition 4.7, that the capacity of channels satisfying (4.26a)–(4.26b) is given by (4.36) and is achieved by state cooperation only. An instance of this scenario will be considered in Corollary 4.4.

4.5 Cooperation Strategies With Total Conferencing Capacity Fixed

In the previous sections, we have studied system performance for given values of the link capacities $C_{SR}$ and $C_{RS}$. Here, we briefly investigate the optimal capacity allocation between the source-to-relay and relay-to-source links where the total conferencing capacity is instead fixed as $C_{SR} + C_{RS} = C_{sum}$. In particular, we compare the rates achievable when the entire capacity is allocated to message cooperation only ($C_{SR} = C_{sum}$ and $C_{RS} = 0$), to state cooperation only ($C_{SR} = 0$ and $C_{RS} = C_{sum}$), or to a combination of message and state cooperation ($C_{SR} = C_{opt} > 0$ and $C_{RS} = C_{sum} - C_{opt}$). We refer to the achievable rates corresponding to the three cases above by scheme 2 as $R_{2,M}$, $R_{2,S}$ and $R_{2,MS}$, respectively.
**Proposition 4.8.** For scheme 2, when $C_{SR} + C_{RS} = C_{sum}$ is fixed, we have

$$R_{2,S} \leq R_{2,MS} \leq R_{2,M}. \quad (4.41)$$

*Proof:* Fix any input distribution of the form (4.13) in $R_2$ of (4.12). The second term in the min function of (4.12) is independent of both $C_{SR}$ and $C_{RS}$, hence it is independent of $C_{sum}$. The third term, when $C_{SR} + C_{RS} = C_{sum}$ is fixed, is the same no matter how one allocates $C_{sum}$ between $C_{SR}$ and $C_{RS}$. Finally, the first term increases with $C_{SR}$. Since $C_{SR}$ cannot be greater than $C_{sum}$, it is optimal to set $C_{SR} = C_{sum}$. It follows that $R_{2,S} \leq R_{2,MS} \leq R_{2,M}$. □

**Remark 4.11.** For scheme 2, it is optimal to allocate all conferencing resources for message forwarding, thereby leading to message cooperation only. In other words, state cooperation is generally not advantageous when utilizing this scheme if one can arbitrarily allocate the overall conferencing capacity. Notice that this may not be always possible, as for instance, in applications where the two conferencing links are unidirectional channels with fixed capacity, e.g., cables. Assessing a similar conclusion holds for scheme 1 seems to be more difficult and is left as an open problem. □

### 4.6 Gaussian Model

In this section, we study the Gaussian model depicted in Fig. 4.2, in which the destination output $Y_i$ at time instant $i$ is related to the channel input $X_i$ from the source,
$X_{R,i}$ from the relay, and the channel state $S_i$ as

$$Y_i = X_i + X_{R,i} + S_i + Z_i,$$

(4.42)

where $S_i \sim \mathcal{N}(0, P_S)$ and $Z_i \sim \mathcal{N}(0, N_0)$, are i.i.d., mutually independent sequences. The channel inputs from the source and relay satisfy the following average power constraints

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_i^2 \right] \leq P, \quad \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_{R,i}^2 \right] \leq P_R.$$  

(4.43)

The conferencing operations, encoding and decoding functions are defined as in Definition 4.1 except that the codewords are required to guarantee the input power constraints (4.43).

### 4.6.1 Achievable Rate

First, we extend the rate (4.9) achievable by scheme 1 to the Gaussian model of (4.42)–(4.43).

**Proposition 4.9.** For the Gaussian relay channel considered, scheme 1 achieves any non-negative rate smaller than $R_1^G$ where

$$R_1^G = \max_{0 \leq \alpha, \beta, \gamma \leq 1} \min \left( A_1, A_2, A_3 \right)$$

(4.44)
with

\[ A_1 = \mathcal{C} \left( \frac{(1 - \alpha) P}{N_0 + \frac{P_S P_Q}{P_S + P_Q}} \right) + C_{SR}, \]  

(4.45a)

\[ A_2 = \mathcal{C} \left( \frac{P + P_R + 2\sqrt{\alpha \beta P P_R}}{N_0 + \frac{P_S P_Q}{P_S + P_Q}} \right) - \mathcal{C} \left( \frac{(P + P_R + 2\sqrt{\alpha \beta P P_R} + N_0) P_S}{(P + P_R + 2\sqrt{\alpha \beta P P_R} + P_S + N_0) P_Q} \right), \]  

(4.45b)

\[ A_3 = \mathcal{C} \left( \frac{(1 - \alpha) P + (1 - \beta) P_R}{N_0 + \frac{P_S P_Q}{P_S + P_Q}} \right) - \mathcal{C} \left( \frac{(P + P_R + 2\sqrt{\alpha \beta P P_R} + N_0) P_S}{(P + P_R + 2\sqrt{\alpha \beta P P_R} + P_S + N_0) P_Q} \right) + C_{SR} + C_{RS}, \]  

(4.45c)

where \( \alpha, \beta \) are the power allocation coefficients at the source and relay respectively, \( P_Q \) is the variance of compression noise selected at the relay and \( \sigma \) is a threshold defined as:

\[
\sigma = \frac{P_S (P + P_R + 2\sqrt{\alpha \beta P P_R} + N_0)}{(P + P_R + 2\sqrt{\alpha \beta P P_R} + P_S + N_0) \min \left(2^{C_{RS}} \left(1 + \frac{(1 - \beta) P_R}{P_S + N_0}\right) - 1, \frac{P + P_R + 2\sqrt{\alpha \beta P P_R}}{P_S + N_0}\right)}.
\]  

(4.46)

**Sketch of Proof:** The result follows from (4.9)–(4.11) by choosing Gaussian input signals satisfying the power constraints. Explicitly, the signals are generated as follows. First, choose \( U \sim \mathcal{N}(0, 1) \). Then, consider \( X = \sqrt{\alpha P} U + \tilde{X} \), where \( 0 \leq \alpha \leq 1 \) and \( \tilde{X} \sim \mathcal{N}(0, (1 - \alpha) P) \), independent of \( U \). Hence, \( X \sim \mathcal{N}(0, P) \). Similarly, set \( X_R = \sqrt{\beta P_R} U + \tilde{X}_R \), where \( 0 \leq \beta \leq 1 \) and \( \tilde{X}_R \sim \mathcal{N}(0, (1 - \beta) P_R) \), independent of \( U \) and \( \tilde{X} \). Hence \( X_R \sim \mathcal{N}(0, P_R) \) and \( \mathbb{E}[XX_R] = \sqrt{\alpha \beta P P_R} \). Next, set \( V = S + Q \) with
compression noise $Q \sim \mathcal{N}(0, P_Q)$ for some $P_Q \geq \sigma$. By standard techniques as in [57, Ch.8 and 9], each mutual information term in (4.9) and (4.11) can be explicitly evaluated, establishing the achievable rate given from (4.44) to (4.46).

Next, we extend the rate (4.12) achievable by scheme 2 to the Gaussian model of (4.42)–(4.43).

**Proposition 4.10.** For the Gaussian relay channel considered, scheme 2 achieves any non-negative rate smaller than $R_2^G$ where

$$R_2^G = \max_{0 \leq \alpha \leq 1} \min_{0 \leq \beta \leq 1} \min_{0 \leq P_Q} (B_1, B_2, B_3)$$

(4.47)

with

$$B_1 = C \left( \frac{(1 - \alpha) P}{N_0 + \frac{P_S P_R}{P_S + P_Q}} \right) + C_{SR},$$

(4.48a)

$$B_2 = \frac{1}{2} \log_2 \left( \frac{P + P_R + 2 \sqrt{\alpha \beta PP_R + P_S + N_0}}{N_0 + \frac{P_S P_Q}{P_S + P_Q}} \right) - C \left( \frac{P_S}{P_Q} \right),$$

(4.48b)

$$B_3 = \frac{1}{2} \log_2 \left( \frac{(1 - \alpha) P + (1 - \beta) P_R + P_S + N_0}{N_0 + \frac{P_S P_Q}{P_S + P_Q}} \right) - C \left( \frac{P_S}{P_Q} \right) + C_{SR} + C_{RS},$$

(4.48c)

where $\alpha, \beta$ are the power allocation coefficients at the source and relay respectively and $P_Q$ is the variance of compression noise selected at the relay.

**Sketch of Proof:** We use the same variable definitions as in the proof of Proposition 4.9 with exception that $P_Q$ only needs to satisfy $P_Q \geq 0$. Then we can explicitly evaluate
each mutual information term in (4.12) following standard techniques in [57, Ch.8 and 9]. Details are omitted here for the sake of conciseness.

**Remark 4.12.** If the relay ignores the available state information, it only cooperates with the source in sending the message information and does not employ the relay-to-source conferencing link. An achievable rate corresponding to this situation can be found from (4.47)−(4.48c) by setting $P_Q \to \infty$, i.e., an infinite variance for the compression of the state information, and $\beta = 1$, i.e., the relay allocates all its power to message transmission. We thus obtain

$$R_{\text{no SI}}^G = \max_{0 \leq \alpha \leq 1} \min \left( C \left( \frac{(1-\alpha)P}{N_0+P_S} + C_{SR} \right), \frac{C \left( P + P_R + 2\sqrt{\alpha P P_R} \right)}{N_0+P_S} \right).$$

(4.49)

Notice that the rate is clearly independent of $C_{RS}$. This rate will be later used for performance comparison.

### 4.6.2 Special Cases and Capacity Results

Now we focus on the special case where $N_0 = 0$ for the Gaussian model of (4.42)−(4.43). We first consider the case with no both message and state cooperation.

**Corollary 4.3.** If $N_0 = 0$ and the conferencing links satisfy $C_{SR} = C_{RS} = 0$, the capacity is given by:

$$C_{\text{no coop}}^G = C \left( \frac{P + P_R}{P_S} \right).$$

(4.50)
Proof: Notice that the channel discussed here satisfies assumptions (4.26a)−(4.26b) in Proposition 4.5. Hence, by extending the results therein to continuous alphabets and evaluating each term by the maximum entropy theorem [57], one can obtain the result claimed in this corollary. Note that when providing both \( S \) and \( X_R \) to the destination, the channel from source to destination is noiseless and hence the first bound in the \( \min \) function of (4.27) goes to infinity, and is thus redundant.

Remark 4.13. The capacity result indicates that strictly causal state information at the relay can provide power gain for the channel considered, even though the relay knows nothing about the message information intended for destination from the source. In fact, when \( N_0 = 0 \), conveying state information from the relay to destination can be considered as equivalently sending part of message for the source, as previously discussed in Remark 4.4.

To elaborate on this insight further, we sketch an alternative achievable scheme in which we explicitly split the message \( W \) from the source into two parts, \( W = (W_{s1}, W_{s2}) \), with \( W_{s1} \in \left[ 1 : 2^{nR_{s1}} \right] \) and \( W_{s2} \in \left[ 1 : 2^{nR_{s2}} \right] \). We divide interval \([-1, 1]\) into \( 2^{nR_{s1}} \) subintervals of equal length and map \( W_{s1} \) to the middle points, denoted by \( \theta(W_{s1}) \), of those subintervals. In addition, we generate \( 2^{nR_{s2}} \) i.i.d. sequences \( x^n \) with each component satisfying \( x_i \sim \mathcal{N}(0, P) \), and map \( W_{s2} \) to the sequences generated as \( x^n(W_{s2}) \). Assume that the source wishes to send \((w_{s1}, w_{s2})\) to the destination. The communication happens in \((n + 1)\) channel uses as follows. In the first channel use, the source sends out the middle point \( \theta(w_{s1}) \) corresponding to message \( w_{s1} \) while the relay sends \( x_{R,1} = 0 \). For the remaining \( n \) channel uses, the source sends out each component of \( x^n(w_{s2}) \) in order. While, for the relay, in the second channel use, it sends out a scaled
version of the state of the previous channel use such that the power constraint is satisfied at the relay, i.e., \( x_{R,2} = \mu_2 s_1 \), where the scalar \( \mu_2 \) is chosen such that \( x_{R,2} \sim \mathcal{N}(0, P_R) \);

For \( i \geq 3 \) channel uses, the relay sequentially forms the minimum mean squared error (MMSE) estimate \( \mathbb{E}[s_1 | \tilde{y}_2^{i-1}] \) with each \( \tilde{y}_k = x_{R,k} + s_k, \forall k = 2, ..., i - 1 \), based on the available states \( s_1^{-1} \) and sends out \( x_{R,i} = u_i (s_1 - \mathbb{E}[s_1 | \tilde{y}_2^{i-1}]) \), where the scalar \( \mu_i \) is chosen such that \( x_{R,i} \sim \mathcal{N}(0, P_R) \). This way, at the end of transmission, the destination first decodes message \( w_s^2 \) by treating the states and information sent by the relay as noise. Hence, as long as \( R_s^2 \leq C(\frac{P}{P_R + P_S}) \), \( w_s^2 \) can be successfully recovered as \( n \to \infty \). After decoding \( w_s^2 \), subtracting \( x_n(w_s^2) \) from the received signal, similar to the analysis of the feedback strategy for point-to-point additive Gaussian channels in [60, 42], one can show that \( w_{s1} \) can be successfully decoded at rate \( R_s^1 = \mathcal{C}\left(\frac{P_R}{P} \right) \) by the state refinement transmission from the relay as \( n \to \infty \). Overall, rate \( R_s^1 + R_s^2 = \mathcal{C}\left(\frac{P_R}{P} \right) + \mathcal{C}\left(\frac{P}{P_R + P_S} \right) = \mathcal{C}\left(\frac{P + P_R}{P_S} \right) = C_{\text{no coop}}^G \) is thus achieved for the source. It is noted that a similar feedback coding scheme can be found in [23] to achieve the maximum rate for each user in a two-user MAC with common state information. □

Next, we consider the optimality of state and message cooperation only following Proposition 4.6 and 4.7.

**Corollary 4.4.** If \( N_0 = 0 \) and the conferencing links satisfy \( C_{RS} \geq \mathcal{C}\left(\frac{P + P_R + 2\sqrt{PP_R}}{P_S} \right) \) with arbitrary \( C_{SR} \), the capacity is given by:

\[
C^G = \mathcal{C}\left(\frac{P + P_R + 2\sqrt{PP_R}}{P_S} \right),
\]  

(4.51)
and is achieved by state cooperation only. Moreover, if \( N_0 = 0 \) and the conferencing links satisfy \( C_{SR} \geq C^G \) with arbitrary \( C_{RS} \), the capacity is also given by (4.51), and is attained by message cooperation only.

Remark 4.14. Example 1 in [23] implies that, if the source knows the state sequence as well, then the maximum rate is given by (4.51). Corollary 4.4 then quantifies the minimum capacity \( C_{RS} \) necessary for this result to be attained on the relay channel of Fig. 4.2 where the source is not given the state sequence. □

Proof: To prove achievability for the case when \( C_{RS} \geq C^G \), we consider a scheme that uses only the relay-to-source conferencing link and perform no message cooperation so that we can equivalently set \( C_{SR} = 0 \). Then, we can identify the result from Proposition 4.7 by simple extension to continuous alphabets and maximizing each term by the maximum entropy theorem [57]. Alternatively, considering the achievable rate (4.47) by scheme 2 and setting \( N_0 = 0, C_{SR} = 0 \), we rewrite \( B_1 \) to \( B_3 \) in the min function as follows:

\[
B'_1 = C \left( \frac{(1 - \alpha) P}{P_S P_Q} \right),
\]

\[
B'_2 = C \left( \frac{P + P_R + 2\sqrt{\alpha \beta P P_R}}{P_S} \right),
\]

\[
B'_3 = C \left( \frac{(1 - \alpha) P + (1 - \beta) P_R}{P_S} \right) + C_{RS}.
\]

Further, setting \( \alpha \to 1, \beta \to 1 \) and \( P_Q \to 0 \) such that \( B'_1 \to \infty \), and under the assumption that \( C_{RS} \geq C \left( \frac{P + P_R + 2\sqrt{PP_R}}{P_S} \right) \), we thus get \( C^G = B'_2 = C \left( \frac{P + P_R + 2\sqrt{PP_R}}{P_S} \right) \).
For the converse part, the upper bound (4.16) reduces to $C^G$ following from the maximum entropy theorem [57].

Turning to the case when $C_{SR} \geq C^G$, for the achievable scheme, the relay simply ignores the state information, so that one can equivalently set $C_{RS} = 0$, and fully cooperates with the source to transmit the message, so that one achieves rate (4.49) with $\alpha = 1$, which reduces to (4.51) under the given condition for $C_{SR}$.

From Corollary 4.4, we immediately have the following.

**Corollary 4.5.** If $N_0 = 0$, and both $C_{RS}$ and $C_{SR}$ are large enough, both state and message cooperation only are optimal and achieve the full cooperation bound (4.51). Compared to the case without any cooperation of (4.50), they both provide cooperative gain.

**Remark 4.15.** If $C_{SR}$ is large enough, e.g., $C_{SR} \geq C^G$, scheme 1 can also achieve capacity, which is attained by setting $P_Q \to \infty$ and $\beta = 1$ in (4.44) similar to Remark 4.12. However, no matter how large $C_{RS}$ is, scheme 1 cannot achieve capacity if $C_{SR} = 0$. This can be argued by considering the extreme case with $C_{RS} \to \infty$. Examining rate (4.44)−(4.46) of scheme 1, we notice that the third term in the min function is redundant due to $C_{RS} \to \infty$. With $N_0 = 0$ and $C_{SR} = 0$, the first two terms can be instead rewritten as

\[
A'_1 = C \left( \frac{(1 - \alpha)P}{P_S P_Q} \right), \quad (4.53a)
\]

\[
A'_2 = C \left( \frac{P + P_R + 2\sqrt{\alpha \beta PP_R}}{P_S} \right), \quad (4.53b)
\]
along with an additional constraint

\[ P_Q \geq \sigma = \frac{P_S^2}{P + P_R + 2\sqrt{\alpha\beta P_P} + P_S}. \] (4.54)

To achieve capacity \( C^G \), we need to set \( \alpha \to 1, \beta \to 1 \) and \( P_Q \to 0 \) as discussed in Corollary 4.4. But, notice that \( P_Q \) is always bounded below by a nonzero threshold, which implies that \( P_Q \to 0 \) cannot be satisfied. Therefore, it can be concluded that scheme 1 cannot achieve capacity by state cooperation only no matter how large \( C_{RS} \) is.

Recall that in scheme 1, the additional constraint comes from the fact that the destination needs to decode the compressed state explicitly, as discussed in Remark 4.1. Compared to this scheme, the advantages of scheme 2 come from joint decoding of message and compression indices. □

### 4.6.3 Numerical Results and Discussions

We now present some numerical results. We start from the special case with \( N_0 = 0 \) studied in Corollary 4.4. We first compare the performance of scheme 1 and scheme 2 for message cooperation only, i.e., \( C_{RS} = 0 \). In Fig. 4.5, we plot the achievable rates versus conferencing capacity \( C_{SR} \). We also plot the rate \( R^G_{no SI} \) in (4.49) that is achieved when the relay does not use the available side information. It can be seen that scheme 2 outperforms scheme 1 in general, consistently with Proposition 4.3. Moreover, if \( C_{SR} \) is large enough, both schemes achieve the upper bound (4.51) and the optimal strategy is to let the relay ignore the state information as provided in Corollary 4.4. But
Fig. 4.5. Comparison of achievable rates between scheme 1 and scheme 2 for message cooperation only ($C_{RS} = 0, P = P_R = P_S = 1, N_0 = 0$).

this strategy is suboptimal for smaller $C_{SR}$. The benefits of state transmission from the relay to the destination are thus clear from this example.

Next, we consider state cooperation only, that is, $C_{SR} = 0$, and compare the achievable rates for two schemes in Fig. 4.6 with the upper bound (4.51). We also plot the achievable rate $C_{no coop}^G$ in (4.50) that is attained when the source transmits message only. The benefits of cooperative state transmission by the source are clear from the figure. Moreover, if $C_{RS}$ is large enough, scheme 2 is seen to achieve the upper bound, as proved in Corollary 4.4. Instead, scheme 1 cannot, as discussed in Remark 4.15.

We now get further insights into system performance by letting $N_0 \neq 0$. We set $P = P_R = P_S = 1$ and vary $N_0$ such that the resulting signal-to-noise ratio, or interfered state-to-noise ratio, $\gamma = 10 \log_{10} (1/N_0)$ lies between $[-5 : 30]$ dB.
Achievable rate $R$ (bits/channel use)

scheme 1, $R_1$

scheme 2, $R_2$

source transmits message only, $C_{no coop}$

upper bound, $R_{ub} = C_{SR}$

Fig. 4.6. Comparison of achievable rates between scheme 1 and scheme 2 for state cooperation only ($C_{SR} = 0, P = P_R = P_S = 1, N_0 = 0$).

We focus on scheme 2 and consider message cooperation only, i.e., $C_{RS} = 0$. Fig. 4.7 shows the rates achievable by scheme 2 and by the same scheme when the relay ignores the state information (4.49) versus $\gamma$. It can be seen that in general state transmission from the relay can provide rate improvement, as also shown in Fig. 4.5. With $C_{SR}$ increasing, the achievable rate increases until it saturates at the upper bound (4.16) when $C_{SR}$ is large enough. For example, as shown in Fig. 4.7, when $C_{SR} = 1.2$, the achievable rate overlaps with the upper bound.

We now consider state cooperation only, that is, $C_{SR} = 0$. Fig. 4.8 shows the rate achievable by scheme 2. The upper bound therein also refers to (4.16). It can be seen that cooperative state transmission by the source is general advantageous, as compared to the performance without cooperation, i.e., $C_{RS} = 0$. However, unlike the case of message cooperation only, even if $C_{RS}$ is large enough, e.g., $C_{RS} = 100$ in Fig. 4.8, the
Fig. 4.7. Comparison of achievable rates for message cooperation only by scheme 2 \((C_{SR} = \{0.2, 0.5, 0.8, 1.2\}, C_{RS} = 0, P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0) (dB))\).

The upper bound is not achievable in general. This is unlike the noiseless case studied in Fig. 4.6, due to the fact that noise makes the state information at the destination less valuable (see Remark 4.4 and 4.13).

Finally, we consider the case when the total conferencing capacity \(C_{SR} + C_{RS} = C_{sum}\) is fixed as discussed in Section 4.5. Under this assumption, we have shown in Section 4.5 that for scheme 2, it is enough to devote all the capacity for message conferencing, thereby leading to message cooperation only. We corroborate this analytical result via a specific example in Fig. 4.9. It can be seen that a combination of both message and state cooperation is able to provide rate improvements as compared to cooperation on state only, while message cooperation only is always optimal.
Fig. 4.8. Comparison of achievable rates for state cooperation only by scheme 2 ($C_{SR} = 0, C_{RS} = \{0, 0.2, 0.5, 0.8, 100\}$, $P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0)$ (dB)).

4.7 Chapter Summary

In this chapter, we have focused a state-dependent relay channel where state information is available at the relay in a strictly causal fashion. Assuming that source and relay can communicate via conferencing links, cooperation is enabled for both transmission of message and state information to the destination. First, we have proposed two coding schemes that exploit both message and state cooperation. The long-message coding scheme inspired by noisy network coding outperforms the more conventional strategy based on block Markov coding and backward decoding. Next, capacity results have been established for some special cases, including no cooperation, message cooperation only and state cooperation only for a class of channels. We have also elaborated on the issue
Fig. 4.9. Comparison of achievable rates for different cooperation strategies when the total conferencing capacity is fixed \( (C_{sum} = 1, P = P_R = P_S = 1, \gamma = 10 \log_{10} (1/N_0) \, (dB)) \).

of optimal capacity allocation between the source-to-relay and relay-to-source conferencing links. Finally, we have characterized achievable rates for the Gaussian model and obtained some capacity results. In general, our results point to the advantage of state information at the relay, despite it being known only strictly causally. This is unlike point-to-point channels. Moreover, for given conferencing capacities, both state and message cooperation can improve the achievable rate.

4.8 Appendices

4.8.1 Proof of Proposition 4.1

Throughout the achievability proofs we use the definition of a strong typical set [42], also see Chapter 2. In particular, the set of strongly jointly \( \epsilon \)-typical sequences
[42] according to a joint probability distribution \( p(xy) \) is denoted by \( T^n_{\epsilon}(XY) \). When the distribution, with respect to which typical sequences are defined, is clear from the context, we will use \( T^n_{\epsilon} \) for short.

Now we present the achievable scheme. Consider \( b \) blocks of transmission. We randomly and independently generate codebooks for each block.

**Codebook Generation:**

Fix a joint distribution \( p(s, v, u, x, x_R, y) = p(s)p(v \mid s)p(u)p(x \mid u)p(x_R \mid u)p(y \mid s, x, x_R) \).

Define rates \( R = R_c + R_p \) with \( 0 \leq R_c \leq \min(R, C_{SR}) \), and \( \tilde{R} = \tilde{R}_c + \tilde{R}_p \) with \( 0 \leq \tilde{R}_c \leq \min \left(\tilde{R}, C_{RS}\right) \).

1. For each block \( j, j \in [1 : b] \), generate \( 2^{nR_v} \) i.i.d. sequences \( v^n_j \) according to the marginal probability mass function (PMF) \( p(v^n_j) = \prod_{i=1}^{n} p(v_{j,i}) \) for the given \( p(v) \). Index them as \( v^n_j(l_j) \) with \( l_j \in \left[1 : 2^{nR_v}\right] \). First partition the set \( \left[1 : 2^{nR_v}\right] \) into \( 2^{n\tilde{R}_c} \) superbins of equal size with each containing \( 2^n(R_v - \tilde{R}_c) \) codewords. Then further partition the codewords in each superbin into \( 2^{n\tilde{R}_p} \) bins of equal size. Then each bin contains \( 2^n(R_v - \tilde{R}) \) codewords. Index each superbin as \( B_{s,j}(t_{c,j}) \) while index each bin as \( B_j(t_{c,j}, t_{p,j}) \) with \( t_{c,j} \in \left[1 : 2^{n\tilde{R}_c}\right], t_{p,j} \in \left[1 : 2^{n\tilde{R}_p}\right] \).

2. For each block \( j \), generate \( 2^{n\tilde{R}_c} \) i.i.d. sequences \( u^n_j \) according to \( p(u^n_j) = \prod_{i=1}^{n} p(u_{j,i}) \) for the given \( p(u) \). Index them as \( u^n_j(w_{c,j}, t_{c,j-1}) \) with \( w_{c,j} \in \left[1 : 2^{nR_c}\right] \) and \( t_{c,j-1} \in \left[1 : 2^{n\tilde{R}_c}\right] \).

3. For each block \( j \), for each \( u^n_j(w_{c,j}, t_{c,j-1}) \), generate \( 2^{nR_p} \) i.i.d. sequences \( x^n_j \) according to the conditional PMF \( p(x^n_j \mid u^n_j(w_{c,j}, t_{c,j-1})) = \prod_{i=1}^{n} p(x_{j,i} \mid u_{j,i}(w_{c,j}, t_{c,j-1})) \) for the given \( p(x \mid u) \). Index them as \( x^n_j(w_{p,j} \mid w_{c,j}, t_{c,j-1}) \) with \( w_{p,j} \in \left[1 : 2^{nR_p}\right] \).
4. For each block $j$, for each $u_j^n(w_{c,j},t_{c,j-1})$, generate $2^{nR_p}$ i.i.d. sequences $x_{R,j}^n$ according to the conditional PMF $p(x_{R,j}^n \mid u_j^n(w_{c,j},t_{c,j-1})) = \prod_{i=1}^{n} p(x_{R,j_i} \mid u_{j_i}(w_{c,j},t_{c,j-1}))$ for the given $p(x \mid u)$.

Index them as $x_{R,j}^n(t_{p,j-1} \mid w_{c,j},t_{c,j-1})$ with $t_{p,j-1} \in \left[1 : 2^{nR_p}\right]$.

**Encoding:**

At the beginning of each block, through conferencing link $C_{SR}$, the common message $w_{c,j}$ can be perfectly conveyed to the relay as long as $bnR_c \leq bnC_{SR}$, which implies that

$$R_c \leq C_{SR}. \quad (4.55)$$

Similarly, the superbin index can be delivered to the source as long as

$$\tilde{R}_c \leq C_{RS}. \quad (4.56)$$

Then we have the following encoding operations:

1. $j = 1$: To send $w_1 = (w_{c,1}, w_{p,1})$ to the destination, the source sends out codeword $x_1^n(w_{p,1} \mid w_{c,1}, 1)$ while the relay sends out codeword $x_{R,1}^n(1 \mid w_{c,1}, 1)$.

2. $j = [2 : b - 1]$: Assume $w_j = (w_{c,j}, w_{p,j})$ to be sent in the $j$th block. At the end of the $(j - 1)$th block, the relay learns the entire state sequence, i.e., $s_{j-1}^n$, and looks for an index (compression index) $l_{j-1}$ such that $(s_{j-1}^n, v_{j-1}^n(l_{j-1})) \in T_{c_l}^n$. If more than one such indices are found, choose the smallest one. If there is no such an index, choose an arbitrary index at random from $\left[1 : 2^{nR_v}\right]$. Let
\((t_{c,j-1}, t_{p,j-1})\) be the bin index pair associated with \(v_{j-1}^{n}(l_{j-1})\). Then codeword 
\[x_{j}^{n}(w_{p,j} | w_{c,j}, t_{c,j-1})\] is sent out by the source and codeword 
\[x_{R,j}^{n}(t_{p,j} | w_{c,j}, t_{c,j-1})\] is sent out by the relay in the \(j\)th block.

3. \(j = b\): No new message is sent at the source. Hence, the source sends out codeword 
\[x_{b}^{n}(1 | 1, t_{c,b-1})\] while the relay sends out codeword 
\[x_{R,b}^{n}(t_{p,b-1} | 1, t_{c,b-1})\].

**Decoding:**

Let \(\epsilon > \epsilon' > 0\). At the end of \(b\) blocks of transmission, the destination performs backward decoding. It first retrieves the bin index pair \((t_{c,b-1}, t_{p,b-1})\) through reception of \(b\)th block, then it decodes the compression index \(l_{b-1}\) by using the received signal \(y_{b-1}^{n}\) and finally it decodes the message \((w_{c,b-1}, w_{p,b-1})\) for block \((b-1)\) using the compressed state information \(v_{b-1}^{n}(l_{b-1})\). This decoding operation is repeated for all blocks back to the first.

Specifically, the decoding procedure for message \((w_{c,j}, w_{p,j})\) of block \(j\) is as follows. Assume that \((w_{c,j+1}, w_{p,j+1})\) are perfectly decoded from the previous estimate. Now the destination looks for an unique bin index pair \((\hat{t}_{c,j}, \hat{t}_{p,j})\) such that

\[
\left( x_{j+1}^{n}(w_{p,j+1} | w_{c,j+1}, \hat{t}_{c,j}), x_{R,j+1}(\hat{t}_{p,j} | w_{c,j+1}, \hat{t}_{c,j}), u_{j+1}^{n}(w_{c,j+1}, \hat{t}_{c,j}), y_{j+1}^{n} \right) \in T_{\epsilon}^{n}
\]  
(4.57)

If there is none or more than one such bin index pairs found, the destination reports an error. Once it finds such a \((t_{c,j}, t_{p,j})\), it looks for an unique compression index \(\hat{l}_{j}\) such
that

\[
\left( v^n_j(\hat{l}_j), y^n_j \right) \in \mathcal{T}_e^n, \quad (4.58)
\]

and

\[
\hat{l}_j \in B_j(t_{c,j}, t_{p,j}). \quad (4.59)
\]

If there is none or more than one such compression indices found, the destination reports an error. Once it finds such a \( l_j \), the destination looks for an unique message \( \hat{w}_j = (\hat{w}_{c,j}, \hat{w}_{p,j}) \) such that

\[
\left( x^n_j(\hat{w}_{p,j} | \hat{w}_{c,j}, t_{c,j-1}), x^n_{R,j} | t_{p,j-1} | \hat{w}_{c,j}, t_{c,j-1}, u^n_j(\hat{w}_{c,j}, t_{c,j-1}), v^n_j(l_j), y^n_j \right) \in \mathcal{T}_e^n \quad (4.60)
\]

for some \( t_{c,j-1} \in \left[ 1 : 2^{n \hat{R}_c} \right], t_{p,j-1} \in \left[ 1 : 2^{n \hat{R}_p} \right] \).

**Analysis of Probability of Error:**

Let \( \Pr(E_j) \) denote the average probability of error for each block \( j \) as defined in (4.8). To bound the overall probability of error, say \( P_e \), without loss of generality (WLOG), assume \((w_{c,j}, w_{p,j}) = (1, 1)\) are sent for each block \( j \). Also denote the compression index selected by the relay for each block by \( L_{j-1} \) and the corresponding bin index pair for each block by \((T_{c,j-1}, T_{p,j-1})\). Note that, following the chain rule,

\[
P_e = \Pr \left( \bigcup_{j=1}^{b} E_j \right) \quad (4.61a)
\]
\[ \Pr(E_b) + \sum_{j=1}^{b-1} \Pr \left( E_j \bigcap_{i=j+1}^{b} E_i^c \right), \quad (4.61b) \]

where since there is no new message sent in the last block, we have \( \Pr(E_b) = 0 \). In the following, we focus on \( \Pr \left( E_j \bigcap_{i=j+1}^{b} E_i^c \right) \), i.e., the probability of error conditioned on not having errors in block \( j+1, \ldots, b \) for each block \( j, j = 1, \ldots, b-1 \), and we show that \( \Pr \left( E_j \bigcap_{i=j+1}^{b} E_i^c \right) \to 0 \) as \( n \to \infty \) if conditions (4.70), (4.71b)–(4.72), and (4.74a)–(4.74c) are satisfied.

Define the encoding error event for each block as follows:

\[ E_{j,0} = \left\{ (v^n_j(l_j), s^n_j) \notin T^n_c, \forall l_j \in \left[ 1 : 2^n R_v \right] \right\}. \quad (4.62) \]

The error events correspond to decoding \( T_{c,j} \) and \( T_{p,j} \) based on rule (4.57) are given by:

\[ E_{j,1} = E_{j,11}^c \bigcup E_{j,12} \bigcup E_{j,13} \bigcup E_{j,14} \quad (4.63) \]

with

\[ E_{j,11} = \left\{ (x^n_{j+1}(1 | l_{c,j}), x^n_{j+1}R_{j+1} | t_{p,j}, l_{c,j}, y^n_{j+1}) \in T^n_e \right\}, \quad (4.64a) \]

\[ E_{j,12} = \left\{ (x^n_{j+1}(1 | l_{c,j}), x^n_{j+1}R_{j+1} | t_{p,j}, l_{c,j}, y^n_{j+1}) \in T^n_e \right\}, \quad (4.64b) \]

for some \( t_{c,j} \neq T_{c,j} \).
\[ E_{j,13} = \left\{ \left( x_{j+1}^n(1, T_{c,j}), x_{R,j+1}^n(t_{p,j} | 1, T_{c,j}), u_{j+1}^n(t_{p,j}, T_{c,j}), y_{j+1}^n \right) \in T_{\epsilon}^n, \right\} \]

\text{for some } t_{p,j} \neq T_{p,j} \tag{4.64c}

\[ E_{j,14} = \left\{ \left( x_{j+1}^n(1, t_{c,j}), x_{R,j+1}^n(t_{p,j} | 1, t_{c,j}), u_{j+1}^n(t_{c,j}, T_{c,j}), y_{j+1}^n \right) \in T_{\epsilon}^n, \right\} \]

\text{for some } t_{c,j} \neq T_{c,j}, t_{p,j} \neq T_{p,j} \tag{4.64d}

The error events correspond to decoding \( L_j \) according to rule (4.58)–(4.59) are given by:

\[ E_{j,2} = E_{j,21} \cup E_{j,22} \tag{4.65} \]

with

\[ E_{j,21} = \left\{ \left( v_j^n(L_j), y_j^n \right) \in T_{\epsilon}^n \right\} \tag{4.66a} \]

\[ E_{j,22} = \left\{ \left( v_j^n(l_j), y_j^n \right) \in T_{\epsilon}^n, \text{for some } l_j \neq L_j, l_j \in B_j(T_{c,j}, T_{p,j}) \right\}. \tag{4.66b} \]

The error events correspond to decoding message \((w_{c,j}, w_{p,j})\) according to rule (4.60) are given by:

\[ E_{j,3} = E_{j,31} \cup E_{j,32} \cup E_{j,33} \tag{4.67} \]
with

\[
E_{j,31} = \left\{ \begin{array}{l}
    \left( x^*_j(1 \mid \hat{w}_{c,j}, t_{c,j-1}), x^n_{R,j}(t_{p,j-1} \mid \hat{w}_{c,j}, t_{c,j-1}) , \\
    u^n_j(\hat{w}_{c,j}, t_{c,j-1}), v^n_j(L_j), y^n_j
\end{array} \right) \in T^n \epsilon , \\
    \text{for some } \hat{w}_{c,j} \neq 1, t_{c,j-1} \in \left[ 1 : 2^n \tilde{R}_c \right], t_{p,j-1} \in \left[ 1 : 2^n \tilde{R}_p \right] \right\},
\]

\[(4.68a)\]

\[
E_{j,32} = \left\{ \begin{array}{l}
    \left( x^*_j(\hat{w}_{p,j} \mid 1 , t_{c,j-1}), x^n_{R,j}(t_{p,j-1} \mid 1 , t_{c,j-1}) , \\
    u^n_j(1, t_{c,j-1}), v^n_j(L_j), y^n_j
\end{array} \right) \in T^n \epsilon , \\
    \text{for some } \hat{w}_{p,j} \neq 1, t_{c,j-1} \in \left[ 1 : 2^n \tilde{R}_c \right], t_{p,j-1} \in \left[ 1 : 2^n \tilde{R}_p \right] \right\},
\]

\[(4.68b)\]

\[
E_{j,33} = \left\{ \begin{array}{l}
    \left( x^*_j(\hat{w}_{p,j} \mid \hat{w}_{c,j}, t_{c,j-1}), x^n_{R,j}(t_{p,j-1} \mid \hat{w}_{c,j}, t_{c,j-1}) , \\
    u^n_j(\hat{w}_{c,j}, t_{c,j-1}), v^n_j(L_j), y^n_j
\end{array} \right) \in T^n \epsilon , \\
    \text{for some } \hat{w}_{p,j} \neq 1, \hat{w}_{c,j} \neq 1, t_{c,j-1} \in \left[ 1 : 2^n \tilde{R}_c \right], t_{p,j-1} \in \left[ 1 : 2^n \tilde{R}_p \right] \right\}.
\]

\[(4.68c)\]

Hence by the union bound,

\[
\Pr \left( E_j \bigcap_{i=j+1}^{b} E^n_i \right) \leq \Pr \left( E_{j,0} \right) + \Pr \left( E_{j,1} \right) + \Pr \left( E_{j,2} \right) + \Pr \left( E_{j,3} \right). \tag{4.69}
\]

1. By the covering lemma in [42], \( \Pr \left( E_{j,0} \right) \to 0 \) as long as

\[
R_v > I(V; S) + \delta \epsilon' \tag{4.70}
\]
for sufficiently large $n$, where $\delta(\epsilon') \to 0$ as $\epsilon' \to 0$.

2. By the packing lemma in [42], $\Pr(E_{j,1}) \to 0$ as long as

$$\tilde{R}_c < I(X, X_R, U; Y) - \delta(\epsilon),$$  \hspace{1cm} (4.71a)

$$\tilde{R}_p < I(X_R; Y | X, U) - \delta(\epsilon),$$  \hspace{1cm} (4.71b)

$$\tilde{R}_c + \tilde{R}_p < I(X, X_R; U; Y) - \delta(\epsilon)$$  \hspace{1cm} (4.71c)

for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Note that bound (4.71c) implies (4.71a), hence (4.71a) is redundant.

3. $\Pr(E_{j,2}) \to 0$ as long as

$$R_v - \tilde{R} < I(V; Y) - \delta(\epsilon)$$  \hspace{1cm} (4.72)

for sufficiently large $n$.

4. For $\Pr(E_{j,3})$, following from the joint typicality lemma [42], for each set of error events from (4.68a) to (4.68c), we have:

$$\Pr(E_{j,31}) \leq 2^{nR_c}2^n(\tilde{R}_c + \tilde{R}_p)2^{-n(I(X, X_R, U; Y | V) - \delta(\epsilon))},$$  \hspace{1cm} (4.73a)

$$\Pr(E_{j,32}) \leq 2^{nR_p}(2^{-n(I(X; Y | X_R, U, V) - \delta(\epsilon))} + 2^{n\tilde{R}_p}2^{-n(I(X, X_R; Y | U, V) - \delta(\epsilon))} + 2^n(\tilde{R}_c + \tilde{R}_p)2^{-n(I(X, X_R; U; Y | V) - \delta(\epsilon))},$$  \hspace{1cm} (4.73b)

$$\Pr(E_{j,33}) \leq 2^n(R_c + R_p)2^n(\tilde{R}_c + \tilde{R}_p)2^{-n(I(X, X_R, U; Y | V) - \delta(\epsilon))}$$  \hspace{1cm} (4.73c)
for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Note that $I(X, X_R, U; Y | V) = I(X, X_R; Y | V)$ because $U \leftrightarrow (X, X_R, V) \leftrightarrow Y$ forms a Markov chain. Thus, $\Pr \left( E_{j,3} \right) \leq \sum_{k=1}^{3} \Pr \left( E_{j,3k} \right) \to 0$ as long as

$$R_p < I(X; Y | X_R, U, V) - \delta(\epsilon),$$

(4.74a)

$$R_p + \tilde{R}_p < I(X, X_R; Y | U, V) - \delta(\epsilon),$$

(4.74b)

$$R_c + R_p + \tilde{R}_c + \tilde{R}_p < I(X, X_R; Y | V) - \delta(\epsilon)$$

(4.74c)

for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Therefore, if bounds (4.70), (4.71b)–(4.72), and (4.74a)–(4.74c) are satisfied and taking both $\epsilon'$ and $\epsilon$ to zero, $\Pr \left( E_j \right) \to 0$ for all $j = 1, ..., b - 1$ and for sufficiently large $n$.

Collecting bounds (4.70), (4.71b)–(4.72), and (4.74a)–(4.74c), along with (4.55), (4.56), $R = R_c + R_p$, $\tilde{R} = \tilde{R}_c + \tilde{R}_p$, applying Fourier-Motzkin elimination[42, Appendix D], and exploiting the fact that $V \leftrightarrow S \leftrightarrow Y$ forms a Markov chain, we establish the achievable rate given by (4.9)–(4.11).

**4.8.2 Proof of Proposition 4.2**

In the proposed scheme, transmission takes place in $b$ blocks of $n$ channel uses each. We randomly and independently generate codebooks for each block. Details of the codebook generation, encoding and decoding operations, and probability of error analysis are provided as follows.
Codebook Generation:

Fix a joint distribution \( p(s, v, u, x, x_R, y) = p(s)p(v | s, x_R, u)p(x | u)p(x_R | u)p(y | s, x, x_R). \)

Define rates \( R = R_c + R_p \) with \( 0 \leq R_c \leq \min(R, C_{SR}) \), and \( R_v = \tilde{R}_c + \tilde{R}_p \) with \( 0 \leq \tilde{R}_c \leq \min(R_v, C_{RS}) \).

1. For each block \( j, j \in [1 : b] \), generate \( 2^{n\left(\frac{bR_c}{n} + \tilde{R}_c\right)} \) i.i.d. sequences \( u^n_j \) according to \( p(u^n_j) = \prod_{i=1}^n p(u_{j,i}) \) for the given \( p(u) \). Index them as \( u^n_j(w_c, t_{c,j-1}) \) with \( w_c \in \left[ 1 : 2^{nR_c} \right] \) and \( t_{c,j-1} \in \left[ 1 : 2^n \tilde{R}_c \right] \).

2. For each block \( j \), for each \( u^n_j(w_c, t_{c,j-1}) \), generate \( 2^{n\tilde{R}_p} \) i.i.d. sequences \( x^n_j \) according to the conditional PMF \( p \left( x^n_j | u^n_j \right) = \prod_{i=1}^n p(x_{j,i} | u_{j,i}) \) for the given \( p(x | u) \). Index them as \( x^n_j(w_p | w_c, t_{c,j-1}) \) with \( w_p \in \left[ 1 : 2^{n\tilde{R}_p} \right] \).

3. For each block \( j \), for each \( u^n_j(w_c, t_{c,j-1}) \), generate \( 2^{n\tilde{R}_p} \) i.i.d. sequences \( x^n_{R,j} \) according to the conditional PMF \( p \left( x^n_{R,j} | u^n_j \right) = \prod_{i=1}^n p(x_{R,i} | u_{j,i}) \) for the given \( p(x_R | u) \). Index them as \( x^n_{R,j}(t_{p,j-1} | w_c, t_{c,j-1}) \) with \( t_{p,j-1} \in \left[ 1 : 2^{n \tilde{R}_p} \right] \).

4. For each block \( j \), for each \( \left( x^n_{R,j}(t_{p,j-1} | w_c, t_{c,j-1}), u^n_j(w_c, t_{c,j-1}) \right) \), generate \( 2^{nR_v} \) i.i.d. sequences \( v^n_j \) according to the conditional marginal PMF

\[
p \left( v^n_j | x^n_{R,j}, u^n_j \right) = \prod_{i=1}^n p(v_{j,i} | x_{R,i}, u_{j,i})
\]

for the given \( p(v | x_R, u) \). Index them as \( v^n_j(t_{c,j}, t_{p,j} | t_{c,j-1}, t_{p,j-1}, w_c) \) with \( t_{c,j} \in \left[ 1 : 2^{n\tilde{R}_c} \right] \) and \( t_{p,j} \in \left[ 1 : 2^{n \tilde{R}_p} \right] \).

Encoding:
The source wishes to send the same message \( w = (w_c, w_p) \) to the destination over all the blocks. At the beginning of the first block, through conferencing link \( C_{SR} \), the common message \( w_c \) can be perfectly conveyed to the relay as long as \( bnR_c \leq bnC_{SR} \), which implies that

\[
R_c \leq C_{SR}. \tag{4.75}
\]

Similarly, the partial compression index \( t_{c,j-1} \in \left[ 1 : 2^n \hat{R}_c \right] \) selected at the relay can always be delivered to the source through the conferencing link \( C_{RS} \) for each \( j \)th block as long as

\[
\hat{R}_c \leq C_{RS}. \tag{4.76}
\]

Then we have the following encoding operations:

1. \( j = 1 \): The source transmits \( x^R_1 (w_p | w_c, 1) \) while the relay transmits \( x^R_1 (1 | w_c, 1) \).

2. \( j = [2 : b] \): At the end of block \( (j - 1) \), the relay learns the entire state sequence, i.e., \( s^n_{j-1} \), and looks for a compression codeword \( v^n_{j-1} \) associated with index \( (t_{c,j-1}, t_{p,j-1}) \) such that

\[
\begin{pmatrix}
s^n_{j-1}, v^n_{j-1}(t_{c,j-1}, t_{p,j-1} | t_{c,j-2}, t_{p,j-2}, w_c), \\
x^n_{R,j-1}(t_{p,j-2} | t_{c,j-2}, w_c), u^n_{j-1}(w_c, t_{c,j-2})
\end{pmatrix} \in T^n_\epsilon.
\]

If more than one codewords are found, choose the first one in the list. If there is no such a codeword, choose an arbitrary one at random from the compression
codebook. Then codeword \( x^n_j(w_p \mid w_c, t_{c,j-1}) \) is transmitted by the source and codeword \( x^n_{R,j}(t_{p,j-1} \mid w_c, t_{c,j-1}) \) is transmitted by the relay in the \( j \)th block.

**Decoding:**

At the end of \( b \) blocks of transmission, the destination performs joint decoding over all blocks by looking for a unique message \( \hat{w} = (\hat{w}_c, \hat{w}_p) \) with \( \hat{w}_c \in [1 : 2^{nbR_c}] \) and \( \hat{w}_p \in [1 : 2^{nbR_p}] \) such that:

\[
\left( x^n_j(\hat{w}_p \mid \hat{w}_c, t_{c,j-1}), x^n_{R,j}(t_{p,j-1} \mid \hat{w}_c, t_{j-1}), v^n_j(t_{c,j}, t_{p,j} \mid t_{c,j-1}, t_{p,j-1}, \hat{w}_c), u^n_j(\hat{w}_c, t_{c,j-1}), y^n_j \right) \in T^n_{\epsilon} \tag{4.77}
\]

for all \( j = 1, \ldots, b \) and some \( t^b \Delta = (t_1, t_2, \ldots, t_b) = (t_{c,1}, t_{p,1}, t_{c,2}, t_{p,2}, \ldots, t_{c,b}, t_{p,b}) \).

**Analysis of Probability of Error:**

To bound the probability of error \( \Pr(E) \), WLOG, assume \( (w_c, w_p) = (1, 1) \) are sent for all blocks. Also denote the indices selected by the relay for each block by \( (T_{c,j-1}, T_{p,j-1}) \).

Define the following encoding error events:

\[
E_0 = \bigcup_{j=1}^{b} \left\{ \left( v^n_{j}(t_{c,j}, t_{p,j} \mid T_{c,j-1}, T_{p,j-1}), u^n_{j}(1, T_{c,j-1}), x^n_{R,j}(T_{p,j-1} \mid T_{c,j-1}, 1), v^n_{j}(1, 1) \right) \notin T^n_{\epsilon}, \forall t_{c,j} \in [1 : 2^{n\hat{R}_c}], \forall t_{p,j} \in [1 : 2^{n\hat{R}_p}] \right\} \tag{4.78}
\]
Define the following decoding events:

\[
E_{(w_c, w_p)} = \left\{ \bigcap_{j=1}^{b} A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \right\},
\]

(4.79)

where each \( A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \) is defined by:

\[
A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \triangleq \left\{ \begin{array}{l}
\left( x_j^n(w_p \mid w_c, t_{c,j-1}), x_{R,j}^{n}(t_{p,j-1} \mid w_c, t_{c,j-1}) , \\
v_j^n(t_{c,j} t_{p,j} \mid t_{c,j-1}, t_{p,j-1}, w_c) , u_j^n(w_c, t_{c,j-1}), y_j^n \right) \in T^n \\
\end{array} \right\}.
\]

(4.80)

Hence by the union bound,

\[
\Pr(E) \leq \Pr(E_0) + \Pr\left( E_{(1,1)}^c \cap E_0^c \right) + \Pr\left( \bigcup_{(w_c, w_p) \neq (1,1)} E_{(w_c, w_p)} \right)
\]

(4.81a)

\[
\leq \Pr(E_0) + \Pr\left( E_{(1,1)}^c \cap E_0^c \right) + \sum_{w_c \neq 1, w_p \neq 1} \Pr\left( E_{(w_c, w_p)} \right)
\]

\[
+ \sum_{w_c \neq 1} \Pr\left( E_{(w_c, 1)} \right) + \sum_{w_p \neq 1} \Pr\left( E_{(1, w_p)} \right).
\]

(4.81b)

1. By the covering lemma in [42], \( \Pr(E_0) \to 0 \) as long as

\[
R_v > I(V; S \mid X_{R}, U) + \delta(\epsilon')
\]

(4.82)

for sufficiently large \( n \), where \( \delta(\epsilon') \to 0 \) as \( \epsilon' \to 0 \).
2. By the conditional joint typicality lemma in [42], \( \text{Pr}\left(E^c_{(1,1)} \mid E^c_0\right) \to 0 \) for sufficiently large \( n \).

3. For \( \sum_{w_c \neq 1, w_p \neq 1} \text{Pr}\left(E(w_c, w_p)\right) \), we have

\[
\sum_{w_c \neq 1, w_p \neq 1} \text{Pr}\left(E(w_c, w_p)\right) = \sum_{w_c \neq 1, w_p \neq 1} \text{Pr}\left(\bigcup_{b} A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1})\right) \leq \sum_{w_c \neq 1, w_p \neq 1} \sum_{b} \prod_{j=1}^{b} \text{Pr}\left(A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1})\right)
\]

\[
\leq 2^{nb(R_c + R_p)} 2^{nb(\tilde{R}_c + \tilde{R}_p)} 2^{-n(b-1)(I(X,X_R,V,U;Y) - \delta(\epsilon))}
\]

where (4.83b) holds by the union bound; (4.83c) holds due to the independence of codebook for each block and the memoryless property of the channel; (4.83d) follows from \( 0 \leq \text{Pr}(A_j) \leq 1 \); and (4.83e) follows from the fact that

\[
\text{Pr}\left(A_j(w_c, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1})\right) \leq 2^{-n(I(X,X_R,V,U;Y) - \delta(\epsilon))}
\]

for sufficiently large \( n \), when \( w_c \neq 1 \) and \( w_p \neq 1 \), by the standard joint typicality lemma.
Note that $I(X, X_R, V; U; Y) = I(X, X_R, V; Y)$ because $U \leftrightarrow (X, X_R, V) \leftrightarrow Y$ forms a Markov chain. Thus $\sum_{w_c \neq 1, w_p \neq 1} \Pr\left( E_{(w_c, w_p)} \right) \to 0$ as long as

$$R_c + R_p < \frac{b - 1}{b} (I(X, X_R, V; Y) - \delta(\epsilon)) - (\tilde{R}_c + \tilde{R}_p)$$

(4.84)

for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Setting $b \to \infty$, we have

$$R_c + R_p < I(X, X_R, V; Y) - (\tilde{R}_c + \tilde{R}_p).$$

(4.85)

4. For $\sum_{w_c \neq 1} \Pr\left( E_{(w_c,1)} \right)$, following similar arguments from (4.83a) to (4.83d), we have

$$\sum_{w_c \neq 1} \Pr\left( E_{(w_c,1)} \right)$$

$$\leq \sum_{w_c \neq 1} \sum_{j=2}^{b} \prod_{t=1}^{b} \Pr\left( A_j(w_c, 1, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \right)$$

(4.86a)

$$\leq 2^{nR_c} 2^{nb(\tilde{R}_c + \tilde{R}_p)} 2^{-n(b-1)(I(X, X_R, V; Y) - \delta(\epsilon))}$$

(4.86b)

where (4.86b) follows from the fact that

$$\Pr\left( A_j(w_c, 1, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \right) \leq 2^{-n(I(X, X_R, V; Y) - \delta(\epsilon))}$$

for sufficiently large $n$ when $w_c \neq 1$, by the standard joint typicality lemma.
Thus \( \sum_{w_c \neq 1} \Pr \left( E_{(w_c, 1)} \right) \to 0 \) as long as

\[
R_c < \frac{b - 1}{b} \left( I(X, X_R; V; Y) - \delta(\epsilon) \right) - (\tilde{R}_c + \tilde{R}_p) \tag{4.87}
\]

for sufficiently large \( n \), where \( \delta(\epsilon) \to 0 \) as \( \epsilon \to 0 \).

Setting \( b \to \infty \), we have

\[
R_c < I(X, X_R; V; Y) - (\tilde{R}_c + \tilde{R}_p). \tag{4.88}
\]

But notice that (4.85) implies this bound, hence it is redundant.

5. For \( \sum_{w_p \neq 1} \Pr \left( E_{(1, w_p)} \right) \), again following similar arguments from (4.83a) to (4.83d), we have

\[
\sum_{w_p \neq 1} \Pr \left( E_{(1, w_p)} \right) \leq \sum_{w_p \neq 1} \sum \prod_{j=2}^{b} \Pr \left( A_j(1, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \right) \tag{4.89}
\]

By the standard argument on joint typicality [42] for enumerations over all \( (t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \) given any fixed \( (w_c = 1, w_p \neq 1) \), we have

\[
\Pr \left( A_j(1, w_p, t_{c,j}, t_{p,j}, t_{c,j-1}, t_{p,j-1}) \right) \tag{4.90a}
\]
where the upper bound $Q_j$ is dependent on $(t_{c,j-1}, t_{p,j-1})$ only given $w_p$, for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Then we have

\[
\sum_{t_{j-1}} Q_j(1, w_p, t_{c,j-1}, t_{p,j-1}) \leq 2^{-n(I(X;Y|X_R,V,U) - \delta(\epsilon))} + 2^n \bar{R}_p 2^{-n(I(X,X_R,V|U) - \delta(\epsilon))}
\]

\[
+ 2^n(\bar{R}_c + \bar{R}_p) 2^{-n(I(X,X_R,V;Y) - \delta(\epsilon))}
\]

\[
\leq 3 \times 2^{-n \min(I_1, I_2, I_3) - \delta(\epsilon)}
\]

with

\[
I_1 = I(X;Y|X_R,V,U),
\]

\[
I_2 = I(X, X_R; V; Y|U) - \bar{R}_p,
\]

\[
I_3 = I(X, X_R; V; Y) - (\bar{R}_c + \bar{R}_p),
\]

for sufficiently large $n$. 

\[
(4.90b)\quad \Delta \triangleq Q_j(1, w_p, t_{c,j-1}, t_{p,j-1})
\]

\[
(4.90c)
\]
Hence,

$$\sum_{w_p \neq 1} \Pr \left( E_{(1,w_p)} \right)$$

$$\leq \sum_{w_p \neq 1} \sum_{t_b} \sum_{j=2}^{b} Q_j(1, w_p, t_{c,j-1}, t_{p,j-1})$$  \hspace{1cm} (4.93a)

$$= \sum_{w_p \neq 1} \sum_{t_b} \sum_{j=2}^{b} Q_j(1, w_p, t_{c,j-1}, t_{p,j-1})$$  \hspace{1cm} (4.93b)

$$\leq 2^{n b R_p} 2^{n \left( \tilde{R}_c + \tilde{R}_p \right) 3 (b-1) 2 - n (b-1) (\min(I_1, I_2, I_3) - \delta(\epsilon))}$$  \hspace{1cm} (4.93c)

where (4.93b) holds because $Q_j(1, w_p, t_{c,j-1}, t_{p,j-1})$ is dependent on $(t_{c,j-1}, t_{p,j-1})$ only when $w_p \neq 1$.

Thus $\sum_{w_p \neq 1} \Pr \left( E_{(1,w_p)} \right) \to 0$ as long as

$$R_p < \frac{b-1}{b} \left( \min(I_1, I_2, I_3) - \delta(\epsilon) \right) - \frac{1}{b} (\tilde{R}_c + \tilde{R}_p) - \frac{(b-1) \log_2 3}{nb}$$  \hspace{1cm} (4.94)

for sufficiently large $n$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Setting $b \to \infty$ and $n \to \infty$, we have

$$R_p < I_1 = I(X; Y | X_R, V, U),$$  \hspace{1cm} (4.95a)

$$R_p < I_2 = I(X, X_R, V; Y | U) - \tilde{R}_p,$$  \hspace{1cm} (4.95b)

$$R_p < I_3 = I(X, X_R, V; Y) - (\tilde{R}_c + \tilde{R}_p).$$  \hspace{1cm} (4.95c)

Again notice that (4.85) implies (4.95c), hence (4.95c) is redundant.
Collecting all the necessary constraints (4.75), (4.76), (4.82), (4.85), (4.95a) and (4.95b), combining with $R = R_c + R_p$, $R_v = \tilde{R}_c + \tilde{R}_p$, and applying Fourier-Motzkin elimination [42, Appendix D], we finally establish the achievable rate given by (4.12)–(4.13).
Chapter 5

The State-Dependent Degraded Broadcast Diamond Channel

In this chapter, we turn to hop-by-hop communication scenarios. In particular, we study a state-dependent parallel-relay diamond channel, in which the state information is non-causally known to the source and it may or may not be available to the destination. We aim to provide new methods of leveraging the non-causal state information in this two-hop channel, besides the idea of state conveying in the previous two chapters.

5.1 Introduction

We consider a communication channel in which the source wishes to communicate to the destination via the help of two parallel relays and there is no direct link between the source and the destination, as shown in Fig. 5.1. The first hop, from the source to the relays, consists of two noiseless digital links of finite capacity: a common link of capacity $C_1$ (bits per channel use) from the source to both relays and a private link of capacity $C_2$ (bits per channel use) from the source to relay 2. The first hop has thus a degraded broadcast channel (BC) structure. The second hop, from the relays to the destination, is a general multiple access channel (MAC) controlled by a random state \cite{7}. It is assumed that (i) the entire state sequence that affects the MAC is known to the source before transmission, (ii) the state is not available at the relays, and (iii) it
may or may not be known at the destination. We term this channel model as the state-dependent degraded broadcast diamond channel (SD-DBDC) with non-causal channel state information (CSI) at the transmitter (i.e., CSIT) and with or without CSI at the receiver (CSIR).

The motivation to study this channel stems from the downlink of a distributed antenna system, in which a central unit controls two antennas, e.g., two pico-base stations, via backhaul links, for communication to an active user over a wireless channel, see for example [3]. The backhaul communication may be received by both antennas over a wireless broadcast channel modeled by $C_1$, or received by one of antennas via a dedicated optical fiber cable modeled by $C_2$. In such a system, the state may model the fading coefficients for the MAC between the distributed antennas and the user, or an interference signal affecting this MAC. In the first case, the user can typically measure the fading channels of the MAC, thus obtaining CSIR, while the central unit may be informed about such fading channels, e.g., via dedicated feedback links, thus obtaining CSIT. The pico-base stations, serving as the relays, are not expected to decode the feedback signal from the user, due to a design choice or insufficient signal-to-noise ratio, and thus CSI is assumed to be unavailable at the relays. In the latter case of an interfering signal affecting the MAC, the interference signal may be communicated to the central unit via backhaul links from the interfering transmitters, e.g., another central unit, thus obtaining CSIT, while relays and the user are not informed, thus having no CSIR.
5.1.1 Background and Related Work

The diamond channel, in which a source communicates to two relays via a general broadcast channel and the relays are connected to the destination via a general state-independent MAC, was introduced by Schein and Gallager in [68] and has been widely studied ever since. For the discrete memoryless diamond channel, several achievability results were established in [68], while for the Gaussian case, it was shown by [48] that partial-decode-and-forward relaying achieves a rate within one bit of the cut-set bound. Despite all the activity, the capacity of this channel in general is open except for some particular instances [69, 70, 71].

A relevant special case of the diamond channel is obtained when the BC in the first hop is modeled as two orthogonal, noiseless digital links of finite capacity. We refer to this model as orthogonal broadcast diamond channel (OBDC). The OBDC was first studied by Traskov and Kramer in [72], where upper and lower bounds on the capacity of the DM OBDC were derived. Recently, Kang and Liu [73] proposed a single-letter upper bound for the OBDC with a Gaussian MAC and established the capacity for a special subclass of Gaussian OBDCs. The SD-DBDC studied here is related to the OBDC, with the differences that the first hop is modeled as a degraded noiseless broadcast channel and that the MAC in the second hop is state-dependent.

A comprehensive review of previous work on channels with states can be found in [14], while the discussion here focuses only on work directly related to the present contribution. Consider first a system as in Fig. 5.1, but with a single relay and with the relay having full knowledge of the message intended for the destination. Note that in
this case, the source-to-relay link, unlike the SD-DBDC, only carries state information and not the message. This channel, which can be seen as a point-to-point system with *encoded* CSIT, was studied by Heegard and El Gamal in [10] under the assumption of CSIR. Therein, a general lower bound was derived and shown to be tight for some special cases. In [20], Cemal and Steinberg studied the extension of this single-relay setting to the case with two relays, under the assumption that the relays are informed about the two independent messages to be delivered to the destination and that there is full CSIR. This model can be seen as a MAC with coded CSIT. Assuming that the source-to-relays links are modeled as in Fig. 5.1 with degraded noiseless channels, the capacity region for this model was characterized. Additionally, inner and outer bounds on the capacity region were derived for the case where the source-to-relays cut consists of separate noiseless links. A related work is also that of Permuter et al. [27], which derived the capacity region for a MAC where the encoders, i.e., the relays of Fig. 5.1, are connected by finite-capacity links to one another, and the MAC channel depends on two correlated state sequences, each known to only one encoder, and there is full CSIR.

We now focus on related studies that assume no CSIR. For the set-up with a single relay and where the relay is informed about the message, i.e., the coded CSIT problem, an upper bound on the capacity was found in reference [74] and proved to be achievable in some special cases. It is noted that, if the relay was informed about both state and message, the optimal strategy would be GP encoding [7], which reduces to DPC [8] in the corresponding Gaussian model with an additive state. The state-dependent MAC with various form of CSIT and no CSIR has been studied in [17, 21, 15, 22, 16, 35]. Assuming non-causal CSIT, the capacity regions for such MAC models are still unknown except
some special instances [15, 16, 17]. Relay channels with state have also been investigated with various type of state information at the nodes, see for example, [30, 31, 37]. In particular, in reference [31], Zaidi et al. studied a single relay channel with non-causal CSI at the source and proposed various achievable schemes. Capacity results were also identified for some special cases [31].

5.1.2 Contributions

In this chapter, we study the SD-DBDC model illustrated in Fig. 5.1 with non-causal CSIT and with or without CSIR. Our contributions are summarized as follows:

- For the DM SD-DBDC with non-causal CSIT and CSIR, we find the capacity. The key ingredient of the achievability is a form of binning inspired by [27], whereby the source selects directly the codewords to be transmitted by the relays in such a way as to adapt them to the given realization of the state sequence. It is demonstrated, similar to [27], that such a joint message and state transmission scheme from the source to the relays is optimal and that it generally outperforms a simple scheme whereby the source sends separate message and state descriptions to the relays, see Section 5.3;

- For the DM SD-DBDC with non-causal CSIT and no CSIR, we first derive an upper bound on the capacity and then propose two transmission strategies. The first proposed strategy operates by sending separate message and state descriptions over the digital links to the relays so as to allow each relay to perform GP coding using the quantized state sequence it reconstructs. We refer the scheme to as GP
coding with quantized states (GP-QS) at the relays. The second scheme, inspired by [75, 31], instead works by having the source first encode the message via GP coding as if the relays had perfect message and state information. Then it sends one common description of the resulting GP sequence to both relays and one refinement description to relay 2. We refer this scheme to as quantized GP coding (QGP). The corresponding lower bounds are derived and presented in Section 5.4.2 to 5.4.3;

- For the case with non-causal CSIT and no CSIR, we also study the Gaussian SD-DBDC with an additive Gaussian state. Achievable rates based on the proposed GP-QS and QGP schemes are evaluated. Numerical results illuminate the merits of non-causal CSIT at the source node and demonstrate the relative performance between the GP-QS and QGP schemes for the Gaussian SD-DBDC, see Section 5.4.4.

## 5.2 System Model and Main Definitions

In this section, we introduce the model studied in this work. Specifically, the SD-DBDC model, depicted in Fig. 5.1, is denoted by the tuple

\[
(C_1, C_2, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S}, p(y|x_1, x_2, s), \mathcal{Y}),
\]  

(5.1)

where \(C_1\) and \(C_2\) are the capacities in bits per channel use of the common link from the source to both the relays, and the private link from the source to relay 2, respectively, \(\mathcal{X}_1\) and \(\mathcal{X}_2\) are the two input alphabets, \(\mathcal{S}\) is the state alphabet, \(\mathcal{Y}\) is the output alphabet and \(p(y|x_1, x_2, s)\) represents the channel probability mass functions (PMFs) describing
Fig. 5.1. A state-dependent degraded broadcast diamond channel (SD-DBDC) with non-causal channel state information (CSI) at the transmitter (CSIT) and with or without CSI at the receiver (CSIR). The CSIR switch is closed or open, respectively.

the MAC between the relays and the destination. The state sequence $s^n$ is generated in an independent and identically distributed (i.i.d.) fashion according to a fixed PMF $p(s)$, i.e.,

$$p(s^n) = \prod_{i=1}^{n} p(s_i).$$

(5.2)

The channel is memoryless in the usual sense and the entire state sequence $s^n$ is assumed to be non-causally known to the source node, i.e., we assume non-causal CSIT. However, sequence $s^n$ may or may not be available at the decoder, i.e., we may or may not have CSIR.

Let $W$ be the message that the source wishes to send to the destination, which is uniformly distributed over the set $\mathcal{W} = [1 : 2^{nR}]$. We define the code as follows.

**Definition 5.1.** A $(2^{nR}, n)$ code for the SD-DBDC consists of:
1. An encoding function at the source node

\[ f : \mathcal{W} \times \mathcal{S}^n \rightarrow [1 : 2^{nC_1}] \times [1 : 2^{nC_2}], \]

(5.3)

which maps the message and the state sequence into two indices \( M_1 \) and \( M_2 \) transmitted over the source-to-relays links;

2. Two encoding functions at the relays

\[ h_1 : [1 : 2^{nC_1}] \rightarrow \mathcal{X}_1^n, \]

(5.4a)

and

\[ h_2 : [1 : 2^{nC_1}] \times [1 : 2^{nC_2}] \rightarrow \mathcal{X}_2^n, \]

(5.4b)

that map the information received by each relay, namely \( M_1 \) by relay 1 and \( (M_1, M_2) \) by relay 2, into the corresponding sequences transmitted by the two relays;

3. A decoding function at the destination. For the case of no CSIR, we have

\[ g : \mathcal{Y}^n \rightarrow \mathcal{W}, \]

(5.5)

which maps the received sequence into a message estimate \( \hat{W} \in \mathcal{W} \), while with CSIR, we have

\[ g : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{W}, \]

(5.6)
which maps the received sequence and the state sequence into a message estimate $\hat{W} \in \mathcal{W}$.

The average probability of error, $P_e^{(n)}$, is defined as $P_e^{(n)} = \Pr[\hat{W} \neq W]$. A rate $R$ is achievable if there exists a sequence of codes $(2^{nR}, n)$ as defined above such that the probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity $C$ of this channel is the supremum of the set of all achievable rates [57].

5.3 Non-causal CSIT and CSIR

In this section, the capacity is established for the DM SD-DBDC with non-causal CSIT and CSIR. The capacity-achieving transmission scheme is presented in Section 5.3.1. For comparison, a straightforward transmission strategy is also considered and its suboptimality is then shown in Section 5.3.2.

5.3.1 Capacity Result

The achievability is based on a scheme in which the source encoder directly selects the codewords to be transmitted by the relays so as to adapt them to the given realization of the state sequence. This is accomplished via a strategy, inspired by [27], in which the codebooks for the transmitted signals $X_1^n$ and $X_2^n$, are binned so that the bin index is identified by the message to be delivered to the destination, and the codewords within the bin are chosen to match the state sequence. Moreover, given the degraded broadcast channel between source and relays, the codebooks for $X_1^n$ and $X_2^n$ are superimposed, so that the codeword for $X_1^n$ is known at both relays, while the codeword for $X_2^n$ is only transmitted, superimposed on $X_1^n$, by relay 2. The following theorem presents the result.
Theorem 5.1. For the DM SD-DBDC model with non-causal CSIT and CSIR, the capacity is given by

\[
C = \max_{\mathcal{P}} \min \begin{pmatrix}
C_1 + C_2 - I(X_1, X_2; S), \\
C_1 - I(X_1; S) + I(X_2; Y | X_1, S), \\
I(X_1, X_2; Y | S)
\end{pmatrix}
\]

with the maximum taken over the distributions in the set

\[
\mathcal{P} = \{ p(s, x_1, x_2, y) : p(s)p(x_1, x_2 | s)p(y | x_1, x_2, s) \}
\]

subject to

\[
\begin{align*}
C_1 & \geq I(X_1; S), \\
C_1 + C_2 & \geq I(X_1, X_2; S).
\end{align*}
\]

Proof: We provide here a sketch of the proof of achievability. Details are provided in Appendix 5.6.1, along with the proof of converse. Let \( \epsilon_2 > \epsilon_1 \), and define functions \( \delta(\epsilon_1) \) and \( \delta(\epsilon_2) \) such that \( \delta(\epsilon_1) \to 0 \) as \( \epsilon_1 \to 0 \) and \( \delta(\epsilon_2) \to 0 \) as \( \epsilon_2 \to 0 \). The source splits message \( w \in [1 : 2^nR] \) into two independent parts \( w_1 \in [1 : 2^{nR_1}] \) and \( w_2 \in [1 : 2^{nR_2}] \). Message \( w_1 \) is associated with a bin \( B_1(w_1) \), that contains \( 2^n(I(X_1; S) + \delta(\epsilon_1)) \) i.i.d. generated codewords indexed by \( x_1^n(w_1, l_1) \), with \( l_1 \in [1 : 2^n(I(X_1; S) + \delta(\epsilon_1))] \), while message \( w_2 \) is associated with a bin \( B_2(w_2 | w_1, l_1) \) for all pairs \( (w_1, l_1) \), that contains \( 2^n(I(X_2; S | X_1) + \delta(\epsilon_2)) \) i.i.d. generated codewords indexed by \( x_2^n(w_2, l_2 | w_1, l_1) \), where...
Given a message pair \( w = (w_1, w_2) \) and a state realization \( s^n \), the source encoder first looks for an index \( l_1 \in [1 : 2^n(I(X_1;S) + \delta(\epsilon_1))] \) such that codeword \( x_1^n(w_1, l_1) \in B_1(w_1) \) is jointly typical with \( s^n \); it then looks for an index \( l_2 \in [1 : 2^n(I(X_2;S|X_1) + \delta(\epsilon_2))] \) such that codeword \( x_2^n(w_2, l_2 | w_1, l_1) \in B_2(w_2 | w_1, l_1) \) is jointly typical with \( (x_1^n(w_1, l_1), s^n) \). Thus, index \( m_1 = (w_1, l_1) \), is conveyed to both relays and index \( m_2 = (w_2, l_2) \), is only conveyed to relay 2 over the digital links. Upon receiving the index and retrieving its corresponding components, relay 1 forwards \( x_1^n(w_1, l_1) \) and relay 2 forwards \( x_2^n(w_2, l_2 | w_1, l_1) \) to the destination. Observing the output sequence \( y^n \) and the state sequence \( s^n \), the decoder chooses a unique tuple of \( (\hat{w}_1, \hat{w}_2, \hat{l}_1, \hat{l}_2) \) such that \( (x_1^n(\hat{w}_1, \hat{l}_1), x_2^n(\hat{w}_2, \hat{l}_2 | \hat{w}_1, \hat{l}_1), s^n, y^n) \) are jointly typical. In this way, the final message estimate \( \hat{w} \) is uniquely determined by \( \hat{w}_1 \) and \( \hat{w}_2 \).

### 5.3.2 The Suboptimality of Separate Message-State Transmission

In the capacity-achieving scheme discussed above, the source encoder selects the codewords for the relay directly based on both message and state sequence in a joint fashion. One can consider, for comparison purposes, a scheme in which the source encoder sends message and state information to the relays separately. The suboptimality of such an approach for a related model was discussed in [27]. We emphasize, however, that, while related, the model considered here is not subsumed by, nor does it subsume, the model in [27].

To elaborate, assume that the source splits the message as \( w = (w_1, w_2) \), as done above, and describes the state sequence using a successive refinement code \( (S_1, S_2) \) [76], where \( S_1 \) represents the base state description and \( S_2 \) represents the refined description.
Message $w_1$ and state description $S_1$ are sent to both relays, while message $w_2$ and state description $S_2$ are sent only to relay 2. A coding scheme, similar to that of Theorem 1 of [20], can now be devised in which message $w_1$ is transmitted by using a codebook, conditioned on the description $S_1$, while message $w_2$ is encoded by relay 2, superimposed on the codeword encoding $w_1$ and is conditioned on state descriptions $(S_1, S_2)$. The corresponding achievable rate is characterized as

$$R_{\text{separate}} = \max_{\mathcal{P}'} \min \begin{pmatrix}
C_1 + C_2 - I(S_1, S_2; S), \\
C_1 - I(S_1; S) + I(X_2; Y | X_1, S, S_1, S_2), \\
I(X_1, X_2; Y | S, S_1, S_2)
\end{pmatrix}$$

with the maximum taken over the distributions in the set

$$\mathcal{P}' = \{p(s, s_1, s_2, x_1, x_2, y) : p(s)p(s_1, s_2 | s)p(x_1 | s_1)p(x_2 | x_1, s_1, s_2)p(y | x_1, x_2, s)\}$$

subject to

$$C_1 \geq I(S_1; S), \quad (5.12a)$$

and

$$C_1 + C_2 \geq I(S_1, S_2; S), \quad (5.12b)$$

where the alphabet size of $S_1$ is bounded as $|S_1| \leq |S| + 3$ and the alphabet size of $S_2$ is bounded as $|S_2| \leq |S|(|S| + 3) + 2$, by standard cardinality bounding techniques [42, Appendix C]. Note that the constraints (5.12a) and (5.12b) represent the well-known
conditions that allow the construction of a successive refinement code with test channel $p(s_1, s_2 | s)$ [76].

We now show that we have in general $R_{\text{separate}} \leq C$ and that this inequality can be strict. In particular, for a fixed $p(s)$ and channel PMF $p(y | x_1, x_2, s)$, considering any PMF in the set $\mathcal{P}'$ of (5.11), we have the following Markov chains: $S - S_1 - X_1$, $S - (S_1, S_2) - (X_1, X_2)$ and $(S_1, S_2) - (S, X_1, X_2) - Y$. Based on these chains, we have the following inequalities

\begin{align*}
C_1 &\geq I(S_1; S) \geq I(X_1; S), \quad (5.13a) \\
C_1 + C_2 &\geq I(S_1, S_2; S) \geq I(X_1, X_2; S), \quad (5.13b) \\
I(X_2; Y | X_1, S, S_1, S_2) &\leq I(X_2; Y | X_1, S), \quad (5.13c)
\end{align*}

and

\begin{align*}
I(X_1, X_2; Y | S, S_1, S_2) &\leq I(X_1, X_2; Y | S), \quad (5.13d)
\end{align*}

which imply that $R_{\text{separate}} \leq C$. We now show with an example that this inequality can be strict.
Fig. 5.2. Performance comparison between $C$, $R_{\text{separate}}$, and $R_{\text{pure-message}}$ for $C_2 = 0.5$, and $p_{x_1^2} = 0.1$ or 0.3 in the binary example of Section 5.3.2.

For the example, we consider the special case of our model obtained with $C_1 = 0$ and $X_1$ taken as a constant, so that the model reduces to the two-hop line network, consisting of the source, relay 2 and the destination (studied also in [27], see Fig. 2 of [27] if $R_2 = 0$ and $p(y|x_1, x_2, s) = p(y|x_2, s)$). Inspired by the binary example considered in [27] in a slightly different context, we then concentrate on the binary model described by

$$Y = S X_2 \oplus Z,$$

where the state $S \sim \text{Bernoulli}(\frac{1}{2})$, the noise $Z \sim \text{Bernoulli}(p_z)$ with $p_z \overset{\Delta}{=} \text{Pr}[Z = 1] \in [0, \frac{1}{2}]$, independent of $S$, and $\oplus$ denotes the modulo-sum operation. We further impose
a cost constraint on the binary input $X_2$ at relay 2 as $\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_{2,i}] \leq p_{x_2}$ with $p_{x_2} \in [0, \frac{1}{2}]$, where $\mathbb{E}[]$ denotes the expectation operation. The capacity of this binary example can be derived from Theorem 5.1 along with the additional input constraint and is given by

\[
C = \max \min \left( C_2 - H_b(\frac{1}{2}(p_0 + p_1)) + \frac{1}{2} H_b(p_0) + \frac{1}{2} H_b(p_1), \right. \\
\left. \frac{1}{2} H_b(p_1 * p_z) - \frac{1}{2} H_b(p_z) \right), 
\]  
(5.15)

subject to constraints $H_b(\frac{1}{2}(p_0 + p_1)) - \frac{1}{2} H_b(p_0) - \frac{1}{2} H_b(p_1) \leq C_2$ and $\frac{1}{2}(p_0 + p_1) \leq p_{x_2}$, where $p_0 \overset{\Delta}{=} \Pr[X_2 = 1 | S = 0] \in [0, 1]$, $p_1 \overset{\Delta}{=} \Pr[X_2 = 1 | S = 1] \in [0, 1]$, $H_b(p) \overset{\Delta}{=} -p \log_2(p) - (1-p) \log_2(1-p)$, and “*” denotes the convolution operation, e.g., $p_1 * p_z = p_1 (1 - p_z) + (1 - p_1) p_z$. Similarly, rate $R_{separate}$ can be obtained from (5.10). We also consider a special case of the “separate” scheme, in which only message information is sent to the relays, so that we set $S_1, S_2$ to a constant in (5.10) (rate $R_{pure\text{-}message}$ in the figure).

Numerical results are provided in Fig. 5.2, where $C$, $R_{separate}$ and $R_{pure\text{-}message}$ are plotted versus $p_z$ for $C_2 = 0.5$, $p_{x_2} = 0.1$ or 0.3, and the cardinality of $S_2$ is assumed to be $m = 2$ in $R_{separate}$ (increasing $m$ to 3, 4 or 5 did not boost the numerical rates of $R_{separate}$). It is clearly seen that $C$ strictly improves upon $R_{separate}$ and the latter strictly outperforms $R_{pure\text{-}message}$ for a wide range of $p_z$ values in this example.
5.4 Non-causal CSIT and No CSIR

In this section, we turn to the SD-DBDC with non-causal CSIT and without CSIR. In the absence of CSIR, the capacity is difficult to establish. In the following, we thus first present an upper bound on the capacity and then illustrate two achievable schemes for the DM model in Section 5.4.1 to 5.4.3. Results are then extended to a Gaussian SD-DBDC with an additive Gaussian state in Section 5.4.4.

5.4.1 An Upper Bound

Proposition 5.1. For the DM SD-DBDC model with non-causal CSIT and no CSIR, the capacity is upper bounded by

\[
R_{ub} = \max_{P_{ub}} \min_{\mathcal{P}_{ub}} \left[ \begin{array}{c}
C_1 + C_2 - I(X_1, X_2; S), \\
C_1 - I(X_1; S) + I(X_2; Y | X_1, S), \\
I(U; Y) - I(U; S)
\end{array} \right]
\]  \hspace{1cm} (5.16)

with the maximization taken over the distributions in the set

\[
\mathcal{P}_{ub} = \{p(s, u, x_1, x_2, y) : p(s)p(u|s)p(x_1, x_2|u, s)p(y|x_1, x_2, s)\}.
\]  \hspace{1cm} (5.17)

Proof: Since the capacity with CSIR cannot be smaller than without CSIR, the first two bounds follow from the converse proof of Theorem 5.1. The third bound in (5.16) is instead obtained by providing message and state information to the relays and thus the proof can be derived as in [7] with the identification of auxiliary random variable as

\[
U_i = (W, S_{i+1}^n, Y^{i-1}).
\]
5.4.2 Achievable Scheme 1: GP Coding With Quantized States At The Relays

In the absence of CSIR, the source can provide information about the state to the relays so as to allow the latter to perform GP coding. Following this idea and an appropriate combination of message splitting, superposition coding and successive refinement coding [76], similar to the discussion in the previous section, we can devise a scheme detailed below, which is referred to as GP coding with quantized states (GP-QS) at the relays. The GP-QS leads to an achievable rate given as follows.

**Proposition 5.2.** For the DM SD-DBDC model with non-causal CSIT and no CSIR, a lower bound on the capacity is given by

\[
R_{\text{GP-QS}} = \max_{\mathcal{P}_1} \min \left\{ \begin{array}{c}
C_1 + C_2 - I(S_1, S_2; S), \\
C_1 - I(S_1; S) + I(U_2; Y | U_1) - I(U_2; S_1, S_2 | U_1), \\
I(U_1, U_2; Y) - I(U_1; S_1) - I(U_2; S_1, S_2 | U_1)
\end{array} \right\}
\]  

(5.18)

with the maximum taken over the distributions in the set

\[
\mathcal{P}_1 = \left\{ \begin{array}{c}
p(s, s_1, s_2, u_1, u_2, x_1, x_2, y):
p(s)p(s_1 | s)p(u_1 | s_1)p(u_2 | u_1, s_1, s_2)
p(x_1 | u_1, s_1)p(x_2 | x_1, u_1, u_2, s_1, s_2)p(y | x_1, x_2, s)
\end{array} \right\}
\]  

(5.19)

subject to

\[
I(S_1; S) \leq C_1,
\]  

(5.20a)
and \( I(S_1, S_2; S) \leq C_1 + C_2 \). \hfill (5.20b)

**Sketch of Proof:** The proof follows from rather standard arguments, see, e.g., [7, 76] and the GP binning coding reviewed in Chapter 2, and thus it is only sketched here. Let \( \epsilon_2 > \epsilon_1 \), and define functions \( \delta(\epsilon_1) \) and \( \delta(\epsilon_2) \) such that \( \delta(\epsilon_1) \to 0 \) as \( \epsilon_1 \to 0 \) and \( \delta(\epsilon_2) \to 0 \) as \( \epsilon_2 \to 0 \). As done in the “separate” strategy discussed in the previous section, the source encoder splits message \( w \in [1 : 2^nR] \) into a common message \( w_1 \in [1 : 2^nR_1] \), to be delivered to both relays, and a private message \( w_2 \in [1 : 2^nR_2] \), to be delivered to relay 2 (so that \( w = (w_1, w_2) \)). Moreover, a successive refinement code \((S_1, S_2)\) is used to describe the state sequence, where the description \( S_1 \), of rate \( R_{s1} \), is delivered to both relays, and the description \( S_2 \), of rate \( R_{s2} \), which refines the first, is communicated only to relay 2. As discussed around conditions (5.12a) and (5.12b), the following conditions guarantee the existence of a successive refinement code with test channel \( p(s_1, s_2 | s) \)

\[
R_{s1} > I(S_1; S), \tag{5.21a}
\]

and \( R_{s2} > I(S_2; S | S_1) \). \hfill (5.21b)

Moreover, in order to guarantee the successful delivery of the messages and state descriptions, the following conditions are sufficient

\[
R_1 + R_{s1} \leq C_1, \tag{5.22a}
\]

and \( R_1 + R_{s1} + R_2 + R_{s2} \leq C_1 + C_2 \). \hfill (5.22b)
Given the messages and quantized state sequences, GP coding is performed by the relays. Specifically, an auxiliary codebook of \( 2^n(R_1 + I(U_1; S_1) + \delta_1) \) i.i.d. codewords \( u^n_1 \) is generated, and then partitioned into \( 2^{nR_1} \) bins indexed by \( B_1(w_1) \), where \( w_1 \in [1 : 2^{nR_1}] \). Using superposition coding, for each codeword \( u^n_1(w_1, l_1) \), where \( l_1 \in [1 : 2^n(I(U_1; S_1) + \delta_1)] \) is the index of the codeword \( u^n_1 \) in the bin \( B_1(w_1) \), a second auxiliary codebook of \( 2^n(R_2 + I(U_2; S_1, S_2 | U_1) + \delta_2) \) i.i.d. codewords \( u^n_2(w_2, l_2 | w_1, l_1) \) is generated, and then partitioned into \( 2^{nR_2} \) bins indexed by \( B_2(w_2 | w_1, l_1) \), where \( w_2 \in [1 : 2^{nR_2}] \) and \( l_2 \in [1 : 2^n(I(U_2; S_1, S_2 | U_1) + \delta_2)] \) is the index of the codeword \( u^n_2 \) in the bin \( B_2(w_2 | w_1, l_1) \). With these codebooks, GP coding of a message \( w = (w_1, w_2) \) takes place as follows. Relay 1 and relay 2 encode \( w_1 \) via the selection of a codeword \( u^n_1(w_1, l_1) \) that is jointly typical with the common quantized state sequence \( s^n_1 \). Then, relay 2 encodes message \( w_2 \) by choosing a codeword \( u^n_2(w_2, l_2 | w_1, l_1) \) jointly typical with \( (u^n_1(w_1, l_1), s^n_1, s^n_2) \). Appropriate channel inputs \( x^n_1 \) and \( x^n_2 \) are then formed by relay 1 and relay 2, respectively, based on the binning codeword(s) selected and the available quantized state(s).

At the destination, upon observing the channel output \( y^n \), the decoder looks for a unique pair of \( (u^n_1(\hat{w}_1, \hat{l}_1), u^n_2(\hat{w}_2, \hat{l}_2 | \hat{w}_1, \hat{l}_1)) \), that is jointly typical with \( y^n \), and assigns the message estimate as \( \hat{w} = (\hat{w}_1, \hat{w}_2) \). If none or more than one such pair is found, an error is declared. By the packing lemma [42, Chapter 3], it is shown that the probability of decoding error vanishes if

\[
R_2 + I(U_2; S_1, S_2 | U_1) < I(U_2; Y | U_1), \quad (5.23a)
\]

and

\[
R_1 + I(U_1; S_1) + R_2 + I(U_2; S_1, S_2 | U_1) < I(U_1, U_2; Y). \quad (5.23b)
\]
Finally, combining the constraints above and using the Fourier-Motzkin procedure [42, Appendix D] to eliminate \((R_{s_1}, R_{s_2})\) and then \((R_1, R_2)\) completes the proof of achievability.

5.4.3 Achievable Scheme 2: Quantized GP Coding

In the GP-QS scheme, a separate description of state and message is conveyed to the relays. Based on the results with CSIR, one might envision that a scheme in which selection of the relays’ codewords is done directly at the source based on both message and state information could be instead advantageous. One such scheme is described here. As further discussed below, however, without CSIR, this scheme is generally not optimal and might even be outperformed by the “separate” GP-QS strategy.

In the second scheme proposed here, inspired by [75, 31], GP coding is done by the source encoder, as if the source encoder had direct access to the relays. Given the finite-capacity link between source and relays, the source encoder then quantizes the resulting GP sequence using a successive refinement code [76], and conveys a common description to both relays and a private description to relay 2. Upon receiving the descriptions and hence having the reconstructed sequences, the relays simply forward them to the destination. Observing the channel output, the decoder looks for a GP codeword that is jointly typical with the received sequence, and obtains the message estimate as the index of the bin to which such codeword belongs. This scheme is referred to as the quantized GP coding (QGP). It leads to the following achievable rate.
Proposition 5.3. For the DM SD-DBDC model with non-causal CSIT and no CSIR, a lower bound on the capacity is given by

$$R_{QGP} = \max_{\mathcal{P}_2} (I(U;Y) - I(U;S))$$

with the maximum taken over the distributions in the set

$$\mathcal{P}_2 = \left\{ p(s, u, v, x_1, x_2, y) : \\
p(s)p(u|s)p(v|u, s)p(x_1, x_2|v)p(y|x_1, x_2, s) \right\}$$

subject to

$$I(X_1; V) \leq C_1,$$  \hspace{1cm} (5.26a)

and

$$I(X_1, X_2; V) \leq C_1 + C_2.$$  \hspace{1cm} (5.26b)

Remark 5.1. The proof of the proposition follows from the discussion above and standard arguments [7, 76] and hence details are omitted for brevity. In the achievable rate derived, we remark that as in [7], $U^n$ denotes the auxiliary binning codewords, while $V^n$ denotes the (auxiliary) analog input sequence, produced by GP encoding at the source encoder. A common description of $V^n$ is carried via both $X_1^n$ and $X_2^n$, a private one is carried via $X_2^n$ only. Inequalities (5.26a)–(5.26b) impose the rates at which the descriptions can be generated. The rate (5.24) is the rate achievable by GP coding on the virtual channel that connects the source to the destination. \qed
Remark 5.2. While a general performance comparison between the GP-QS and QGP schemes does not seem to be easy to establish, it can be seen that when the link capacities are arbitrarily large, either the state sequence or the GP analog sequence can be perfectly conveyed to the relays, and thus both the GP-QS and QGP schemes achieve the upper bound (5.16), and specifically the third bound in (5.16), thus giving the capacity.

5.4.4 Gaussian SD-DBDC

We now study a Gaussian SD-DBDC as depicted in Fig. 5.3. In particular, we assume that the destination output $Y_i$ at time instant $i$ is related to the channel inputs $X_{1,i}, X_{2,i}$ at the relays and the channel state $S_i$ as

$$Y_i = X_{1,i} + X_{2,i} + S_i + Z_i,$$

(5.27)

where $S_i \sim \mathcal{N}(0, P_S)$ and $Z_i \sim \mathcal{N}(0, N_0)$, are i.i.d., mutually independent sequences. The channel inputs at the relays satisfy the following average power constraints

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_{k,i}^2] \leq P_k, k = 1, 2.$$

(5.28)

The encoding and decoding functions are defined as in Definition 5.1 except that the codewords are required to guarantee the input power constraints (5.28).

5.4.4.1 Reference Results

For reference, we first consider the performance of a simple scheme that does not leverage the non-causal CSIT. In particular, the source splits again the message $u$ into
two independent parts $w = (w_1, w_2)$ and sends $w_1$ at rate $R_1$ to both the relays and $w_2$ at rate $R_2$ to the relay 2 via the digital links. In this way, the model at hand is converted into a Gaussian MAC channel with degraded message sets [77, 78]. The decoder simply treats the state as noise. The maximum message rates supported by the first hop are given by: $R_1 \leq C_1$ and $R_1 + R_2 \leq C_1 + C_2$, while the capacity region for the MAC cut is obtained from [78] as

$$R_2 \leq C \left( \frac{(1 - \rho^2)P_2}{N_0 + P_S} \right), \quad R_1 + R_2 \leq C \left( \frac{P_1 + P_2 + 2\rho\sqrt{P_1P_2}}{N_0 + P_S} \right)$$

for $0 \leq \rho \leq 1$, where we recall that $C(x)$ is defined as $C(x) = \frac{1}{2} \log_2(1 + x)$. Therefore, the overall achievable rate without using CSIT is given by

$$R^G_{\text{no SI}} = \max_{0 \leq \rho \leq 1} \min \begin{pmatrix}
C_1 + C_2, \\
C \left( \frac{P_1 + P_2 + 2\rho\sqrt{P_1P_2}}{N_0 + P_S} \right), \\
C \left( \frac{(1 - \rho^2)P_2}{N_0 + P_S} \right) + C_1
\end{pmatrix},$$

which serves as a natural lower bound for the capacity of our example considered.
A simple upper bound $R_{ub}^G$ can be instead obtained by providing the decoder with the interference sequence so that it can be cancelled. The capacity region of the corresponding state-independent system can be found from [78] and is given by (5.30) with $N_0$ in lieu of $N_0 + P_S$.

5.4.4.2 Achievable Rates

We now apply the GP-QS and QGP schemes discussed above to the given Gaussian model.

**Proposition 5.4.** For the Gaussian SD-DBDC model, the following rate is achievable by the GP-QS scheme:

$$
R_{GP-QS}^G = \max_{0 \leq \rho \leq 1, (D_1, D_2) \in \mathcal{A}} \min \left( C_1 + C_2 - \frac{1}{2} \log_2 \left( \frac{P_S}{D_2} \right), C_1 - \frac{1}{2} \log_2 \left( \frac{P_S}{D_1} \right) + C \left( \frac{\rho P_2}{D_2 + N_0} \right), C \left( \sqrt{P_1 + \rho P_2} + \frac{N_0}{2} \right) \right),
$$

(5.31)

where $\bar{\rho} = 1 - \rho$ and the set of $\mathcal{A}$ is defined as

$$
\mathcal{A} \triangleq \left\{ (D_1, D_2) : P_S \geq D_1 \geq D_2 \geq 0, D_1 \geq P_S 2^{-2C_1}, D_2 \geq P_S 2^{-2(C_1 + C_2)} \right\}.
$$

\( \text{Proof:} \) Note that the result of Proposition 5.2 can be extended to the continuous channel by standard techniques [42, Chapter 3]. Thus, one can obtain the achievable rate in this proposition through evaluation of the general result therein by identifying appropriate inputs. Details of the proof are provided in Appendix 5.6.2. We remark that
$(D_1, D_2)$ in (5.32) represent the distortions at which the state $S$ is described to the two relays via the successive refinement code $(S_1, S_2)$ used in GP-QS.

Next, we derive the achievable rate based on the QGP scheme.

**Proposition 5.5.** For the Gaussian SD-DBDC model, the following rate is achievable by the QGP scheme:

$$R_{QGP}^G = C \left( \frac{\left( \sqrt{P_1 (1 - 2^{-2C_1})} + \sqrt{P_2 \left(1 - 2^{-2(C_1 + C_2)}\right)} \right)^2}{P_1 2^{-2C_1} + P_2 2^{-2(C_1 + C_2)} \left(1 + 2 \frac{P_1 (1 - 2^{-2C_1})}{P_2 \left(1 - 2^{-2(C_1 + C_2)}\right)} + N_0\right)} \right) + N_0.$$  

(5.33)

**Proof:** The proof is obtained from Proposition 5.3, similar to the proof of Proposition 5.4 (see Appendix 5.6.3).

**Remark 5.3.** As the digital link capacity $C_1$ becomes arbitrarily large, it is easy to see that both schemes GP-QS and QGP attain the upper bound $R_{ub}^G$, leading to the capacity $C = C \left( \frac{P_1 + P_2 + 2\sqrt{P_1 P_2}}{N_0} \right)$. Note that the capacity is the same as if the interference at the destination was not present and if full cooperation was possible at the relays. The benefit of utilizing the non-causal CSIT is therefore evident from this example. We also emphasize that letting capacity $C_2$ alone grow to infinity is not enough to obtain the upper bound above, as in this case only relay 2 can be fully informed by the central unit.

**Remark 5.4.** The achievable rate $R_{GP-QS}^G$ of scheme GP-QS is generally dependent on the interference power $P_S$, while the achievable rate $R_{QGP}^G$ of scheme QGP is not.
This is because in the GP-QS scheme, the state sequence needs to be described to the relays on the finite-capacity links, and thus the stronger is the power $P_S$ of the interfered state, the larger are the feasible distortions $(D_1, D_2)$ in (5.32) for reproducing the state sequence at the relays. As a result, in the extreme case in which the state power $P_S$ becomes arbitrarily large, the rate $R^G_{GP-QS}$ reduces to rate $R^G_{no SI}$ (5.30) obtained when the decoder simply treats the state as noise. On the other hand, in the QGP scheme, the source compresses directly the appropriate GP sequence, whose power does not depend on $P_S$. Given the fact that the performance of QGP is not dependent on $P_S$, it is expected that the QGP scheme outperforms the GP-QS scheme in case $P_S$ is sufficiently large.

5.4.4.3 Numerical Results

We now further investigate the performance of the proposed schemes via numerical results. We first fix the digital link capacities as $C_1 = 1.5$ and $C_2 = 1$. We also set $P_1 = P_2 = P = 1$, and vary $N_0$ so that the signal-to-noise ratio (SNR), defined as $\text{SNR} = 10 \log_{10}(P/N_0)$, lies between $[-10 : 30]$ dB. Fig. 5.4 and Fig. 5.5 illustrate the corresponding achievable rates versus SNR, given $P_S = 0.2$ or $0.4$, and $P_S = 0.8$ or 1.2, respectively. It can be seen that with a small state power $P_S$, e.g., $P_S = 0.2$ as in Fig. 5.4, rate $R^G_{GP-QS}$ of scheme GP-QS improves upon rate $R^G_{no SI}$ of the simple scheme without using CSIT, while rate $R^G_{QGP}$ of scheme QGP is smaller than both. This is due to the fact that, when $P_S$ is relatively small, it is more effective to describe the state sequence to the relays, as done with GP-QS. In the case of moderate $P_S$, e.g., $P_S = 0.4$ as in Fig. 5.4, we observe that both the GP-QS and QGP schemes outperform
the simple scheme. In the case of moderate-to-strong \( P_S \), e.g., \( P_S = 0.8 \) or \( 1.2 \) as in Fig. 5.5, as explained in Remark 5.4, scheme QGP is generally advantageous over scheme GP-QS.

We now plot in Fig. 5.6 the achievable rates versus \( C_1 \), for \( C_2 = 1, P_1 = P_2 = 1, N_0 = 0.1 \) and \( P_S = 1.2 \). It can be seen that, when \( C_1 \) is large enough, both the GP-QS and QGP schemes attain the upper bound \( R_{\text{ub}}^G \), hence giving the capacity, as discussed in Remark 5.3. Next, the achievable rates are plotted versus \( P_S \) in Fig. 5.7, for fixed link capacities \( C_1 = 1.5, C_2 = 1 \) and \( P_1 = P_2 = 1, N_0 = 0.1 \). This figure further confirms the discussion in Remark 5.4, by showing that both rates \( R_{\text{GP-QS}}^G \) and \( R_{\text{no SI}}^G \) decrease as \( P_S \) increases.
Fig. 5.5. Achievable rates $R$ vs. SNR for $C_1 = 1.5, C_2 = 1, P_1 = P_2 = 1, P_S = 0.8$ or $1.2$.

5.5 Chapter Summary

In this chapter, we have studied a state-dependent diamond channel, in which the broadcast channel between source and relays is defined by a noiseless degraded broadcast channel, and the multiple access channel between relays and destination is state-dependent. For the case with non-causal channel state information at the transmitter (CSIT) and at the receiver (CSIR), we have established the capacity and shown that a joint message and state transmission scheme via binning is optimal and superior to the scheme that performs separate message and state description transmission. For the case without CSIR, we have proposed an upper bound and two transmission schemes, and applied the results to a Gaussian model with an additive Gaussian state. For the
Fig. 5.6. Achievable rates $R$ vs. $C_1$ for $C_2 = 1, P_1 = P_2 = 1, N_0 = 0.1, P_S = 1.2$.

Gaussian model, numerical results demonstrate the merit of the non-causal CSIT, and indicate that the best available transmission scheme generally depends on the power of the state.

5.6 Appendices

5.6.1 Proof of Theorem 5.1

Throughout the achievability proofs we use the definition of a strong typical set[42], also see Chapter 2. In particular, the set of strongly jointly $\epsilon$-typical sequences [42] according to a joint probability distribution $p(xy)$ is denoted by $T^n_\epsilon(XY)$. When the distribution, with respect to which typical sequences are defined, is clear from the context, we will use $T^n_\epsilon$ for short.
Achievable rates $R$ vs. $P_S$ for $C_1 = 1.5$, $C_2 = 1$, $P_1 = P_2 = 1$, $N_0 = 0.1$.

Achievability:

**Codebook generation**: Fix a joint distribution $p(s)p(x_1, x_2 | s)p(y | x_1, x_2, s)$ where $p(s)$ and $p(y | x_1, x_2, s)$ are defined by the channel. Let $R = R_1 + R_2$, $\hat{R}_1 > R_1 \geq 0$ and $\hat{R}_2 > R_2 \geq 0$. Randomly and independently generate $2^{n\hat{R}_1}$ i.i.d. $x_1^n$ sequences, each according to $\prod_{i=1}^{n} p(x_{1,i})$ and then partition them into $2^{nR_1}$ bins $B_1(w_1)$, with $w_1 \in [1 : 2^{nR_1}]$. Hence, there are $2^{n(\hat{R}_1 - R_1)}$ $x_1^n$ codewords in each bin, which are indexed by $x_1^n(w_1, l_1)$ with $l_1 \in [1 : 2^n(\hat{R}_1 - R_1)]$. Moreover, for any given $x_1^n(w_1, l_1)$, generate $2^{n\hat{R}_2}$ i.i.d. $x_2^n$ sequences, each according to $\prod_{i=1}^{n} p(x_{2,i} | x_{1,i}(w_1, l_1))$ and then partition them into $2^{nR_2}$ bins $B_2(w_2 | w_1, l_1)$, with $w_2 \in [1 : 2^{nR_2}]$. Hence, there are $2^{n(\hat{R}_2 - R_2)}$ $x_2^n$ codewords in each bin, which are further indexed by $x_2^n(w_2, l_2 | w_1, l_1)$ with $l_2 \in [1 : 2^n(\hat{R}_2 - R_2)]$. Reveal the whole codebook generated to all parties involved.
Encoding: Let $\epsilon_3 > \epsilon_2 > \epsilon_1$, and define functions $\delta(\epsilon_k)$ such that $\delta(\epsilon_k) \rightarrow 0$ as $\epsilon_k \rightarrow 0$ for $k = 1, 2, 3$. The source encoder splits message $w \in [1 : 2^{nR}]$ into two independent parts $w_1 \in [1 : 2^{nR_1}]$ and $w_2 \in [1 : 2^{nR_2}]$. Message $w_1$ is associated with each bin $B_1(w_1)$, while message $w_2$ is associated with each bin $B_2(w_2 | w_1, l_1)$ for any fixed $(w_1, l_1)$. Given the message pair $(w_1, w_2)$ and non-causal state information $s^n$, the source encoder first looks for a codeword $x^n_1(w_1, l_1) \in B_1(w_1)$ such that $(x^n_1(w_1, l_1), s^n) \in T^n_1(X_1 S)$; if there are more than one, choose the first one according to the lexicographic order; if there is none, set $l_1 = 1$. Given the $x^n_1(w_1, l_1)$ found, the source encoder further looks for $x^n_2(w_2, l_2 | w_1, l_1) \in B_2(w_2 | w_1, l_1)$ such that $(x^n_2(w_2, l_2 | w_1, l_1), x^n_1(w_1, l_1), s^n) \in T^n_2(X_2 X_1 S)$; if there are more than one, choose the first one according to the lexicographic order; if there is none, set $l_2 = 1$. Then the source conveys index $m_1 = (w_1, l_1)$ and index $m_2 = (w_2, l_2)$ to the relays via the digital links. In particular, index $m_1$ is intended for both relays and $m_2$ only for relay 2. Upon receiving the index and retrieving its corresponding components from the source, relay 1 transmits $x^n_1(w_1, l_1)$, while relay 2 transmits $x^n_2(w_2, l_2 | w_1, l_1)$ to the destination.

Decoding: Given $(s^n, y^n)$, the decoder looks for a unique tuple of $(\hat{w}_1, \hat{l}_1, \hat{w}_2, \hat{l}_2)$ such that $(x^n_1(\hat{w}_1, \hat{l}_1), x^n_2(\hat{w}_2, \hat{l}_2 | \hat{w}_1, \hat{l}_1), s^n, y^n) \in T^n_3(X_1 X_2 S Y)$; if there is none or more than one such tuples, an error is reported. Then the final message estimate is assigned as $\hat{w} = (\hat{w}_1, \hat{w}_2)$.

Analysis of probability of error: Without loss of generality, assume that $w = (w_1, w_2) = (1, 1)$ is sent by the source and the indices conveyed to the relays are $M_1 = (1, L_1)$ and $M_2 = (1, L_2)$. The analysis of probability of error mainly follows from the covering lemma and the packing lemma [42, Chapter 3]. Specifically, by the covering
lemma, given any typical sequence $s^n$, the source encoding error vanishes as $n \to \infty$ if

$$\tilde{R}_1 - R_1 > I(X_1; S) + \delta(\epsilon_1), \quad (5.34a)$$

and

$$\tilde{R}_2 - R_2 > I(X_2; S | X_1) + \delta(\epsilon_2). \quad (5.34b)$$

Moreover, the indices $M_1$ and $M_2$ can be perfectly conveyed to both relays and relay 2, respectively, as long as the digital link capacities satisfy

$$\tilde{R}_1 \leq C_1, \quad (5.35a)$$

and

$$\tilde{R}_1 + \tilde{R}_2 \leq C_1 + C_2. \quad (5.35b)$$

By the packing lemma, the probability of decoding error event \{$(w_1, l_1) \neq (1, L_1)$, for all $(w_2, l_2)$\} vanishes as $n \to \infty$ if

$$\tilde{R}_1 + \tilde{R}_2 < I(X_1, X_2; Y, S) - \delta(\epsilon_3). \quad (5.36)$$

Similarly, the probability of decoding error event \{$(w_1, l_1) = (1, L_1), (w_2, l_2) \neq (1, L_2)$\} vanishes as $n \to \infty$ if

$$\tilde{R}_2 < I(X_2; Y, S | X_1) - \delta(\epsilon_3). \quad (5.37)$$

Finally, combining the above conditions (5.34a)–(5.37) and using the Fourier-Motzkin procedure to eliminate $(\tilde{R}_1, \tilde{R}_2)$ and then $(R_1, R_2)$ completes the proof of achievability.
The Converse: Let $M_1$ be the common index conveyed to both relays and $M_2$ be the private index conveyed to relay 2 only. First, considering the digital link capacity constraint, we have that

\begin{align*}
nC_1 & \geq H(M_1) \quad (5.38a) \\
& \geq I(M_1; S^n) \quad (5.38b) \\
& = \sum_{i=1}^{n} I(M_1; S_i | S^{i-1}) \quad (5.38c) \\
& = \sum_{i=1}^{n} I(M_1, S^{i-1}, X_{1,i}; S_i) \quad (5.38d) \\
& \geq \sum_{i=1}^{n} I(X_{1,i}; S_i), \quad (5.38e)
\end{align*}

where (5.38d) holds because of the facts that $S_i$ is independent of $S^{i-1}$ and $X_{1,i}$ is a deterministic function of $M_1$. By the same reasoning, we can show that

\begin{align*}
n(C_1 + C_2) & \geq H(M_1, M_2) \quad (5.39a) \\
& \geq I(M_1, M_2; S^n) \quad (5.39b) \\
& \geq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; S_i). \quad (5.39c)
\end{align*}

We can also write

\begin{align*}
nR & = H(W) \quad (5.40a) \\
& \leq I(W; Y^n | S^n) + n\epsilon_n \quad (5.40b)
\end{align*}
\[ \sum_{i=1}^{n} I(W; Y_i | Y^{i-1}, S^n) + n \varepsilon_n \quad (5.40c) \]

\[ \sum_{i=1}^{n} \left[ H(Y_i | Y^{i-1}, S^n) - H(Y_i | Y^{i-1}, S^n, W, M_1, M_2, X_{1,i}, X_{2,i}) \right] + n \varepsilon_n \quad (5.40d) \]

\[ \sum_{i=1}^{n} \left[ H(Y_i | Y^{i-1}, S^n) - H(Y_i | S_i, X_{1,i}, X_{2,i}) \right] + n \varepsilon_n \quad (5.40e) \]

\[ \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i | S_i) + n \varepsilon_n \quad (5.40f) \]

with \( \varepsilon_n \to 0 \) as \( n \to \infty \), where (5.40b) is due to Fano’s inequality, i.e., \( H(W | Y^n, S^n) \leq n \varepsilon_n \); (5.40d) holds because \((M_1, M_2)\) is a deterministic function of \((W, S^n)\), \(X_{1,i}\) is a deterministic function of \(M_1\) and \(X_{2,i}\) is a deterministic function of \((M_1, M_2)\); (5.40e) follows from the memoryless property of the channel; and (5.40f) follows from the fact that conditioning reduces entropy.

Next, we can prove a second bound on the rate as

\[ nR = H(W) \quad (5.41a) \]

\[ = H(W | S^n) \quad (5.41b) \]

\[ = H(W, M_1, M_2 | S^n) \quad (5.41c) \]

\[ = H(M_1, M_2) - I(M_1, M_2; S^n) + H(W | M_1, M_2, S^n) \quad (5.41d) \]

\[ = H(M_1, M_2) - \sum_{i=1}^{n} I(M_1, M_2, X_{1,i}, X_{2,i}, S^{i-1}; S_i) + H(W | M_1, M_2, S^n) \quad (5.41e) \]

\[ \leq n(C_1 + C_2) - \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; S_i) + H(W | M_1, M_2, S^n) \quad (5.41f) \]

\[ = n(C_1 + C_2) - \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; S_i) + H(W | M_1, M_2, S^n, Y^n) \quad (5.41g) \]
\[ \leq n(C_1 + C_2) - \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; S_i) + n \epsilon_n \tag{5.41h} \]

with \( \epsilon_n \to 0 \) as \( n \to \infty \), where (5.41b) is due to the independence between \( W \) and \( S^n \); (5.41c) holds because \( (M_1, M_2) \) is a deterministic function of \( (W, S^n) \); (5.41e) follows from the facts that \( S_i \) is independent of \( S^{i-1} \), \( X_{1,i} \) is a deterministic function of \( M_1 \) and \( X_{2,i} \) is a deterministic function of \( (M_1, M_2) \); (5.41f) follows because of the capacity constraints on the links between source and relays, and because of the chain rule and the non-negativity of mutual information; (5.41g) holds due to the Markov chain \( W - (M_1, M_2, S^n) - Y^n \) so that \( I(W; Y^n | M_1, M_2, S^n) = 0 \); and (5.41h) follows from Fano’s inequality.

Moreover, we have the third bound

\[ nR = H(W) \tag{5.42a} \]

\[ = H(W, M_1 | S^n) \tag{5.42b} \]

\[ = H(M_1) - I(M_1; S^n) + H(W | M_1, S^n) \tag{5.42c} \]

\[ \leq nC_1 - \sum_{i=1}^{n} I(X_{1,i}; S_i) + H(W | M_1, S^n) \tag{5.42d} \]

\[ \leq nC_1 - \sum_{i=1}^{n} I(X_{1,i}; S_i) + I(W; Y^n | M_1, S^n) + n \epsilon_n \tag{5.42e} \]

\[ = nC_1 - \sum_{i=1}^{n} I(X_{1,i}; S_i) \]

\[ + \sum_{i=1}^{n} \left[ H(Y_i | Y^{i-1}, M_1, S^n, X_{1,i}) - H(Y_i | Y^{i-1}, M_1, S^n, W, M_2, X_{1,i}, X_{2,i}) \right] + n \epsilon_n \tag{5.42f} \]
\[ nC_1 - \sum_{i=1}^{n} I(X_{1,i}; S_i) + \sum_{i=1}^{n} \left[ H(Y_i | Y_{i-1}, M_1, S^n, X_{1,i}) - H(Y_i | S_i, X_{1,i}, X_{2,i}) \right] + n\epsilon_n \]  

(5.42g) 

\[ \leq nC_1 - \sum_{i=1}^{n} I(X_{1,i}; S_i) + \sum_{i=1}^{n} I(X_{2,i}; Y_i | X_{1,i}, S_i) + n\epsilon_n \]  

(5.42h) 

with \( \epsilon_n \to 0 \) as \( n \to \infty \), where lines (5.42b) to (5.42d) are obtained by similar reasonings for lines (5.41b) to (5.41f) in the previous bound; (5.42e) is due to Fano’s inequality, i.e., \( H(W | Y^n, S^n, M_1) \leq n\epsilon_n \); (5.42f) holds by the chain rule and also because \( M_2 \) is a deterministic function of \( (W, S^n) \), \( X_{1,i} \) is a deterministic function of \( M_1 \) and \( X_{2,i} \) is a deterministic function of \( (M_1, M_2) \); (5.42g) follows from the memoryless property of the channel; and (5.42h) holds due to the fact that conditioning reduces entropy. 

Finally, let \( Q \) be a random variable uniformly distributed over the set \([1 : n] \). Define random variables \( S = S_Q, X_1 = X_{1,Q}, X_2 = X_{2,Q} \) and \( Y = Y_Q \). Then, bounds (5.38e), (5.39c), (5.40f), (5.41h) and (5.42h) can be written as 

\[ C_1 \geq I(X_{1,Q}; S_Q | Q) = I(X_1; S | Q), \]  

(5.43a) 

\[ C_1 + C_2 \geq I(X_{1,Q}, X_{2,Q}; S_Q | Q) = I(X_1, X_2; S | Q), \]  

(5.43b) 

and 

\[ R - \epsilon_n \leq I(X_{1,Q}, X_{2,Q}; Y_Q | S_Q, Q) = I(X_1, X_2; Y | S, Q), \]  

(5.44a) 

\[ R - \epsilon_n \leq (C_1 + C_2) - I(X_{1,Q}, X_{2,Q}; S_Q | Q) = (C_1 + C_2) - I(X_1, X_2; S | Q), \]  

(5.44b) 

\[ R - \epsilon_n \leq C_1 - I(X_{1,Q}; S_Q | Q) + I(X_{2,Q}; Y_Q | S_Q, X_{1,Q}, Q) \]
\[ C_1 - I(X_1; S|Q) + I(X_2; Y|S, X_1, Q), \]  
where the distribution on \((Q, S, X_1, X_2, Y)\) from a given code is of the form

\[ p(q, s, x_1, x_2, y) = p(q)p(s)p(x_1, x_2 | s, q)p(y | x_1, x_2, s). \]  

To eliminate the variable \(Q\) from bounds (5.43a) to (5.44c), we note that

\[ I(X_1; S|Q) = H(S|Q) - H(S|X_1, Q) \]  
\[ = H(S) - H(S|X_1, Q) \]  
\[ \geq I(X_1; S), \]

where (5.46b) follows from the fact that the symbols \(S_i\) with \(i \in [1 : n]\) are i.i.d. and hence \(S = S_Q\) is independent of \(Q\). Similarly, we can prove that

\[ I(X_1, X_2; S|Q) \geq I(X_1, X_2; S). \]  

Moreover, the inequalities

\[ I(X_1, X_2; Y|S, Q) \leq I(X_1, X_2; Y|S), \]  
and \[ I(X_2; Y|S, X_1, Q) \leq I(X_2; Y|S, X_2), \]
hold because of the Markov chain $Q - (X_1, X_2, S) - Y$. Given the facts above, the bounds corresponding to (5.7)–(5.9b) are recovered by noticing that the distribution of the random variables $(S, X_1, X_2, Y)$ obtained by marginalizing (5.45) over $Q$ is of the exact form given in $P$ of (5.8). This concludes the converse proof and also the proof of Theorem 5.1.

5.6.2 Proof of Proposition 5.4

Based on the GP-QS scheme described in Section 5.4.2 and whose achievable rate is given by (5.18)–(5.20b), for state encoding, we consider the following cascade of backward channels: $S = S_2 + Z_2 = (S_1 + Z_1) + Z_2$, where $S_1 \sim \mathcal{N}(0, P_S - D_1)$, $Z_1 \sim \mathcal{N}(0, D_1 - D_2)$ and $Z_2 \sim \mathcal{N}(0, D_2)$, are independent, and $P_S \geq D_1 \geq D_2 \geq 0$. This construction implies the Markov chain: $S_1 - S_2 - S$. Hence, we have that

$$I(S_1; S) = \frac{1}{2} \log_2 \left( \frac{P_S}{D_1} \right),$$  
(5.49a)

and $I(S_1, S_2; S) = I(S_2; S) = \frac{1}{2} \log_2 \left( \frac{P_S}{D_2} \right).$  
(5.49b)

And the constraints of (5.20a) and (5.20b) become

$$D_1 \geq P_S 2^{-2C_1}, \quad D_2 \geq P_S 2^{-2(C_1 + C_2)}.$$  
(5.50)

For message encoding, we let $X_1 \sim \mathcal{N}(0, P_1)$, independent of $(S, S_1, S_2)$; $X_2 = \sqrt{\frac{\rho P_2}{P_1}} X_1 + V_2$, where $0 \leq \rho \leq 1$, and $V_2 \sim \mathcal{N}(0, \rho P_2)$ is also independent of $(S, S_1, S_2)$. 
The auxiliary random variables $U_1$ and $U_2$ are defined as

$$U_1 = X_1 + \alpha_1 S_1, \quad (5.51a)$$

$$U_2 = V_2 + \alpha_2 \left( S_2 - \alpha_1 \left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right) S_1 \right), \quad (5.51b)$$

for some $\alpha_1, \alpha_2 \geq 0$ to be specified later. Note that, with these choices, the channel output $Y$ becomes

$$Y = X_1 + X_2 + S + Z \quad (5.52a)$$

$$= \left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right) X_1 + V_2 + S + Z \quad (5.52b)$$

$$= \left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right) X_1 + V_2 + S_1 + Z_1 + Z_2 + Z. \quad (5.52c)$$

Therefore, with the choice of $U_1$ given above, we have that

$$I(U_1; Y) - I(U_1; S_1) \leq C \left( \frac{\left( \sqrt{P_1} + \sqrt{\rho P_2} \right)^2}{\rho P_2 + D_1 + N_0} \right), \quad (5.53)$$

where the equality is achieved by setting

$$\alpha_1^* = \frac{\left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right) P_1}{\left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right)^2 P_1 + \bar{\rho} P_2 + D_1 + N_0} \quad (5.54)$$

in (5.51a), which is such that $\alpha_1^*(Y - S_1)$ is the minimum Mean-Square-Error (MSE) estimate of $X_1$ given $Y - S_1$, similar to Costa’s DPC [8]. Next, to decode the private message carried over $U_2$, the decoder subtracts $\left( 1 + \sqrt{\frac{\rho P_2}{P_1}} \right) U_1$ from $Y$ obtaining the
received signal

\[ Y' = V_2 + S_2 - \alpha_1^* \left( 1 + \sqrt{\frac{P_2}{P_1}} \right) S_1 + Z_2 + Z. \quad (5.55) \]

Now, with the choice of \( U_2 \) in (5.51b), we have that

\[
I(U_2; Y | U_1) - I(S_1, S_2; U_2 | U_1) \\
= I(U_2; Y') - I(U_2; S_2 - \alpha_1^* \left( 1 + \sqrt{\frac{P_2}{P_1}} \right) S_1) \\
\leq C \left( \frac{\bar{\rho}P_2}{D_2 + N_0} \right),
\]

where the equality is achieved by setting

\[ \alpha_2^* = \frac{\bar{\rho}P_2}{\bar{\rho}P_2 + D_2 + N_0}. \quad (5.57) \]

This concludes the proof.

### 5.6.3 Proof of Proposition 5.5

Based on the QGP scheme described in Section 5.4.3 and whose achievable rate is given by (5.24)–(5.26b), we let the auxiliary random variable \( V \sim \mathcal{N}(0, P_v) \) for some \( P_v > 0 \), independent of \( S \). Consider the following cascade of forwarding channels: \( X_2 = V + Z_2 \), and \( X_1 = \alpha_1 X_2 + Z_1 \), where \( X_1 \sim \mathcal{N}(0, P_1) \) and \( X_2 \sim \mathcal{N}(0, P_2) \); \( Z_1 \sim \mathcal{N}(0, \sigma_1^2) \), \( Z_2 \sim \mathcal{N}(0, \sigma_2^2) \), which are independent of each other and also of \( V \); Parameters \( \alpha_1, \sigma_1^2 \) and \( \sigma_2^2 \) are to be specified. Following this construction, note that \( X_1 - X_2 - V \) forms a Markov chain. Therefore, the constraint of (5.26b) becomes \( I(X_1, X_2; V) = I(X_2; V) = \)
\[ \frac{1}{2} \log_2 \left( \frac{P_2}{\sigma_2^2} \right) \leq C_1 + C_2. \] Thus, one can choose \( \sigma_2^2 = P_2 2^{-2(C_1+C_2)} \). Then, \( P_v = P_2 \left( 1 - 2^{-2(C_1+C_2)} \right) \) due to the power constraint on \( X_2 \). Moreover, noting that \( \alpha_1^2 P_2 + \sigma_2^2 = P_1 \) and \( \frac{1}{2} \log_2 \left( \frac{P_1}{\alpha_1^2 \sigma_2^2 + \sigma_1^2} \right) \leq C_1 \) due to constraint (5.26a), we thus choose \( \sigma_1^2 = \frac{P_1 2^{-2C_1} \left( 1 - 2^{-2C_2} \right)}{1 - 2^{-2(C_1+C_2)}} \) and \( \alpha_1 = \frac{P_1 \left( 1 - 2^{-2C_1} \right)}{P_2 \left( 1 - 2^{-2(C_1+C_2)} \right)} \). The auxiliary random variable \( U \) is defined as \( U = V + \beta^* S \), where \( \beta^* \) is chosen to be the weight of the minimum MSE estimate of \( V \) given \( Y - S = X_1 + X_2 + Z \), similar to Costa’s DPC [8]. In this way, the message rate \( R_{QGP}^G = I(U; Y) - I(U; S) = I(V; X_1 + X_2 + Z) \) which equals (5.33). This completes the proof.
Chapter 6

Conclusion

6.1 Dissertation Summary

As introduced in Chapter 1, state-dependent channels are useful to model relevant phenomena in wireless communication links, such as fading and interference. The standard model prescribes existence of a state sequence with each symbol denoting the channel’s variation at an instant in time. In many communication scenarios, nodes gain knowledge about the state sequence or are capable of sensing the state sequence. Hence, a key problem of both theoretical and practical interest is identifying methods for leveraging the available information about the state sequence at the nodes to ensure reliable communication over the state-dependent channels.

In light of this background, this dissertation is dedicated to exploring some new problems in multi-user channels and providing fresh insights for the fundamental limits of different channel models. The particular focus is on channels with strictly causal state information and hop-by-hop state-dependent channels, for which existing results are quite limited.

First, in Chapter 3, we have investigated the state-dependent MAC channels with strictly causal state information at the transmitters only. To leverage the outdated state information, we have used the idea of compressing the state and devoting part of the resources to transmitting the compressed state information from each transmitter.
to the receiver. We have proposed a new approach to implement the said idea. The scheme proposed uses a combination of long-message encoding, compression of the past state sequences and codewords without binning, and joint decoding over all transmission blocks. Our scheme generalizes and improves upon the existing work. Capacity results are identified for a class of channels. We have also extended the proposed scheme to state-dependent MAC with an arbitrary number of users.

Next, in Chapter 4, we have studied a state-dependent relay channel where only the relay has strictly causal state information. In the model considered, the source can convey information about the message to the relay while the relay can forward state information to the source. This exchange enables cooperation between the source and the relay on transmission of message and state information to the destination. We have devised two transmission schemes for the general model and demonstrated the relative merits of message cooperation and state cooperation via a Gaussian model.

In general, the above works point to the advantages of strictly causal state information in multi-user communication channels. The results are in contrast to the single-user channel, in which the strictly causal state information is useless.

In Chapter 5, we have considered a two-hop communication scenario, that is the state-dependent parallel-relay diamond channel. In the model, the first hop from the source to the relays is modeled as a degraded broadcasting channel and the second hop from the relays to the destination is a general state-dependent MAC. Under the model, we have established the capacity for the case with non-causal state information known to both the source and destination but unknown to the relays. We have demonstrated that a joint message and state transmission strategy is optimal and attains the capacity.
We have also proposed two transmission strategies for the case with non-causal state information at the source only and computed the achievable rates for a Gaussian model with an additive state.

The results in the dissertation have been presented in part at conferences [36, 38] and submitted as journal papers [35, 37, 40].

6.2 Future Research

In terms of future research, several directions are worthy of pursuit.

First, the merits of strictly causal state information have further exploration in other multi-user communication settings. Moreover, besides the discrete memoryless and single-antenna Gaussian channels considered in the current research, studying the impact of strictly causal state information in the Gaussian multi-antenna multi-user communication settings is worthwhile. This direction is closely related to the recent work [79], which showed that the outdated transmitter state information can still greatly improve the degrees of freedom for the two-user Gaussian Multi-Input Single-Output (MISO) broadcast channel.

Next, it is noticed that the current research primarily focuses on exploiting the state information to improve the reliability of message delivery. But in some scenarios, see, e.g., [80, 81], the destination not only intends to recover the message but also attempts to reconstruct the state sequence. Consequently, this desire leads to the problem of finding the fundamental trade-off between message rate and state reconstruction distortion. Research along this direction appears in [80, 81] for the single user channel with
non-causal or causal state information. An interesting investigation would be extension to the multi-user setup.

Lastly, in the current research, the state sequence, being either the interference or the fading sequence, is governed by the nature of wireless channels. It is possible that such a state sequence is generated and controlled by a communication node, which intentionally introduces some time-varying property to the channels as seen by other nodes. This scenario, referred to as the channel with action-dependent state [82], opens an interesting avenue for assessing the roles of such action as well as the induced state in this class of communication settings.
Bibliography


Symposium on Information Theory, pages 2806–2810, Saint Petersburg, Russia, August 2011.


Vita

Min Li received his B.E. and M.E. degrees from the Department of Information Science and Electronic Engineering at Zhejiang University, Hangzhou, China, in 2006 and 2008, respectively. Since August 2008, he has been pursuing a Ph.D. degree at the Department of Electrical Engineering, Penn State University. He has been a member of the Wireless Communications and Networking Laboratory (WCAN) at Penn State, working with Professor Aylin Yener as a research assistant. His research interests include network information theory, coding theory, wireless communication theory and system designs, VLSI designs and optimization techniques. He has served as journal reviewers for IEEE Transactions on Information Theory, IEEE Transactions on Wireless Communication, IEEE Communications Letters, and IEEE Journal on Selected Areas in Communications. He is one of the recipients of 2007 ASUS Distinguished Graduate Student Scholarship at Zhejiang University in China.

Publications:


J.1 Min Li, Osvaldo Simeone and Aylin Yener, “Multiple Access Channels with States Causally Known at Transmitters”, submitted to IEEE Transactions on Information Theory, November 2010, revised in April 2012.
