

ENHANCING RATES IN RELAY CHANNELS

by

Wen Luo

A Dissertation submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy
in

Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada
September 2015

Copyright ©
2015 - Wen Luo

Abstract

The main objective of this study is to investigate the structures and procedures which are promising to achieve higher rates and larger rate regions in the relay channels and the communication networks that contain relay nodes. In particular, the study places particular emphasis on the relaying techniques pertaining to compress-and-forward.

The thesis first examines a generalization of decode-and-forward (DF) and compress-and-forward (CF) in Gaussian channels. Although this generalization has been known for over thirty years, the result in this thesis is the first to illustrate the signal-to-noise ratio (SNR) regions in which the generalization reduces to constituent DF or CF schemes. In particular, the thesis demonstrates the existence of SNR regions in which the generalization is guaranteed to supersede both DF and CF, but with a gain of within 0.5 bits per channel use.

Having gained insight into the random binning in the CF scheme, this thesis argues that a new decoding procedure exploiting the N -to-1 mapping based on binning is able to relax the rate constraint on the relay transmission, and generalize the noisy network coding based schemes which are constrained to the 1-to-1 mapping. This thesis identifies two instances in which exploiting the N -to-1 mapping inherent in this generalization yields rate gains, even though it does not yield such a gain in other multmessage networks.

In the analysis of a secure communication problem, the thesis introduces the concept of “friendly” eavesdropper in a broadcast channel in the presence of a malicious Gaussian jammer. Taking advantage of the N -to-1 mapping, it is shown that the new decoding procedure enables CF to achieve the capacity of the channel.

Dedicated to Kate Yang and Jeffrey Luo

Acknowledgments

My foremost thank goes to my direct supervisor, Professor Halim Yanikomeroglu at the Department of Systems and Computer Engineering of Carleton University. He provided invaluable guidance throughout the development of this study. Without his precious support and patience it would not be possible to conduct this research.

My sincere thanks also go to my co-supervisor, Dr. Ramy Gohary, at the Department of Systems and Computer Engineering of Carleton University. Through many late day discussions his insightful advice encouraged me to continue on this journey, and his hard questions inspired me to widen and deepen my research in various perspectives.

I am also very grateful to my parents who instilled in me the enthusiasm about science and gave me the courage to follow my dreams.

A special gratitude and love goes to my son, Jeffrey, for his sweet smile and understanding. My deepest gratitude is reserved for my lovingly wife, Kate. Without her tolerance, patience and unfailing support, completing this thesis would not have been possible.

Table of Contents

Abstract	ii
Acknowledgments	v
Table of Contents	vi
List of Figures	x
1 Introduction	1
1.1 Motivation	1
1.2 Background	4
1.3 Overview	5
1.3.1 Generalized DF-CF Relaying	5
1.3.2 Exploiting the N -to-1 Mapping in CF Relaying	6
1.3.3 Military Application of the Advanced Relaying Scheme	8
1.4 Contributions	8
1.5 Organization	10
1.6 Notation	10
2 Analysis of the Generalization of DF and CF in Gaussian Relay Channels: An Optimization-Based Approach	11
2.1 Introduction	12

2.2	System Model	15
2.3	Generalization of DF and CF for Gaussian Relay Channels	17
2.3.1	An Information-Theoretic Background	17
2.3.2	Application to Gaussian Signals	18
2.4	The Generalized DF-CF: Achievable Rate and Analysis	19
2.4.1	The Case of $\gamma_0(1 + \gamma_2) < \gamma_1 < \infty$	23
2.4.2	The Case of $\gamma_1 \leq \gamma_0$	26
2.4.3	The Case of SNRs Yielding $R_{\text{DF}}^* = R_{\text{CF}}^*$	28
2.4.4	An Upper Bound on the Gain of the Generalized DF-CF	33
2.5	Numerical Results	33
3	A New Decoding Procedure to Support the N-to-1 Mapping in Compress-and-Forward Relaying	37
3.1	Introduction	37
3.2	Preliminaries	41
3.3	A Layered Forward Decoding Procedure for Multimessage Network	44
3.4	Conclusion	52
4	Exploiting the N-to-1 Mapping in Compress-and-Forward Relaying	53
4.1	Introduction	54
4.1.1	The Achievable Rate of A Broadcast Relay Chain Network	55
4.1.2	Achievable Rate Region of A Partially Cooperative Multimessage Network	62
4.2	Conclusion	69
5	Military Application of the Advanced Relaying Strategy	71
5.1	Introduction	72
5.2	Channel Model	75

5.3	Capacity Results	76
5.4	Comparison with Other Eavesdropping Signalling Schemes	79
5.4.1	No Eavesdropper Case	80
5.4.2	CF With Standard Decoding Case	80
5.4.3	Amplify-and-Forward Case	81
5.5	Numerical Comparison	81
5.6	Conclusion	82
6	Conclusions and Future Work	84
6.1	Conclusions and Contributions	84
6.2	Future Work	85
6.3	Thesis Summary	87
	Appendices	88
A.1	Proof of Proposition 1	89
A.2	Proof of the achievable rate expressions of DF and CF in the Gaussian channel	91
A.3	Proof of Theorem 1	94
A.4	Proof of Lemma 2	96
A.5	Proof of Theorem 2	98
A.6	Proof of Theorem 4	100
B.1	Proof of Theorem 5	102
B.2	Proof of Remark 6	114
B.3	Proof of Theorem 6	116
B.4	Proof of Remark 9	127
C.1	Proof of Corollary 2	129
D.1	Analysis of Probability of Error for Theorem 7	134
D.2	Proof of the Achievable Rate in Section 5.4.3	139

List of Figures

1	Gaussian three-node relay channel.	16
2	Maximum achievable rate of the generalized DF-CF scheme.	35
3	Asymptotic optimality of the generalized DF-CF scheme	36
4	Maximum achievable rate in quasi-static Rayleigh fading channel. . .	36
5	Standard three-node relay channel (SNNC or CF).	43
6	An N -node multmessage network.	46
7	A two-destination broadcast relay chain network.	56
8	A two-destination Gaussian broadcast relay network with a relay chain, where Z' is an independent additive noise and Z_2, Z_3, Z_{D_1} and $Z_{D_2} \sim$ $\mathcal{N}(0, 1)$	60
9	Achievable rates of a two-destination Gaussian broadcast relay chain network in Fig. 8, $\gamma_{SR} = 2$, $\gamma_{SD_1} = 1$, $\gamma_{SD_2} = 2$, $\gamma_{RD_1} = 2$, $\gamma_{RD_2} = 1$, $\gamma_{RS_1} = 2$ and $\gamma_{RS_2} = 1$	61
10	A two-message three-receiver partially cooperative network.	63
11	Gaussian particularization of the network in Fig. 10.	67
12	Rate region for the network in Figure 11 ($\gamma_{SR} = 2$, $\gamma_{SD_1} = 1$, $\gamma_{RD_1} = 4$, $\gamma_{SD_2} = 2$, $\gamma_{RD_2} = 1$, $\gamma_{RD_3} = 1$).	70
13	An illustration of the considered communication system.	72
14	A broadcast channel in the presence of a Gaussian jammer with a friendly eavesdropper, $J \sim \mathcal{N}(0, P_J)$	76

15	Comparison between capacity (Theorem 7) and the rates achievable using the schemes in Section 5.4, for $\gamma_2 = 4$, $\gamma_{e,1} = 3$, $\gamma_{e,2} = 2$	82
----	---	----

Chapter 1

Introduction

1.1 Motivation

Wireless communication has become an integral part of our daily lives, covering applications that range from entertainment-centric to military-centric ones. The main feature that distinguishes wireless communications from wired ones, is that in the former, the information signals are air-borne. Hence, these systems not only alleviate a considerable portion of the infrastructure required to establish communication links, but also offer flexible connectivity between communicating parties.

Despite its advantages, wireless communications suffer from more adverse propagation conditions than their wired counterparts. This is because signals emitted from a wireless transmitter usually take indirect routes to reach the receiver. Each of these routes is usually composed of a cascade of multiple links, and each link is usually composed of multiple parallel paths. The number of paths of each link and the number of links of each route depend on the locations, obstructions and objects surrounding the transmitter and the receiver. The variations in the number and lengths of routes and the number of paths that each route comprises result in reverberations in the signal impinging on the antennas of the receiver. These reverberations have different

phases and may combine constructively to increase the signal strength, or destructively to decrease it. The fluctuations in the signal strength caused by the various paths traveled by the signal is collectively known as fading. To facilitate its analysis, fading is usually classified into various types, fast and slow, depending on the rate at which the signal fluctuates, and frequency-flat and frequency-selective depending on the delay between reverberations arriving at the receiver.

Analyzing fading reveals that it can have a fundamental impact on the data rate that can be reliably communicated over the channel. This impact depends, naturally, on the type of fading experienced by the signal; fast frequency-selective fading is the most challenging one. This type of fading arises in high mobility communication scenarios with multiple geographically dispersed scatterers.

Combating fading in point-to-point communications revolves around techniques by which the receiver combines phase-aligned versions of the signals arriving at the receiver at different time slots. Despite their advantages, these techniques are only capable of recovering the transmitted signals when the channel between the transmitter and the receiver is relatively strong. Unfortunately, in many practical scenarios, this is not the case. For instance, in high density city centres, buildings and other objects such as trees, cars and pedestrians in the street canyon create multiple signal paths. The mobility of such objects causes the paths to vary from time to time, and subsequently causes, in many cases, the signal from the transmitter to the receiver to experience severe fast and frequency-selective fading. In those cases, using traditional point-to-point techniques to combat fading are likely to fail to provide reliable communication at satisfactory data rates. However, an effective means for combating the deleterious effect of fading in these situations is through the deployment of communication assisting nodes, known as relays. The presence of such nodes gives rise to what is arguably the simplest multinode communication model. Studying such models in their general form from an information-theoretic perspective has proven difficult over

the years. However, even with our incomplete understanding of these models, the role of using relays can be appreciated for instance by considering communication scenarios in which the geometry of the propagation environment is extremely hostile to the wireless signal emitted from the transmitter, e.g. in mountainous terrains and underground tunnels. In fact, without relays, wireless communications in many such scenarios would be practically impossible.

Because of the fundamental impact that a relay has on the communication channel, using relays in cellular systems provides a cost-effective means for extending the coverage area and the number of users in each cell; the cost of a relay is usually negligible compared to that of a base station. Various kinds of relays are currently available. Among these kinds are fixed relays which are usually deployed by the system operator at specific well-chosen locations to serve a particular region, nomadic relays which are usually deployed by the user to enhance his/her own communication channel, and terminal relays which comprise of the communication devices of idle users.

Owing to their effectiveness in improving coverage and increasing the number of the users at low cost, relay-assisted communications have been widely adopted in many wireless standards, including IEEE 802.11ah (Wi-Fi), IEEE 802.16j (WiMAX) and 3GPP (LTE-A), and is likely to be included in future standards as well.

In order to extract the potential gains offered by a relay-assisted communication channel, the role of the relay ought to be carefully contemplated. Generally speaking, the relay cooperates with the transmitter by providing the receiver with additional information about the transmitter's signal or, in some cases, the channel condition. Such information can be constructed either in analog or in digital domains. For instance, in the analog domain the relay may only amplify its received signal, and forwards it to the destination without recovering the original message. In contrast, in the digital domain, the relay may use various procedures to extract and possibly

reconstruct and communicate a description of the original signal to the receiver. The understanding of the performance of these procedures under various conditions is key to realizing the gain offered by relays. The focus of this thesis is to analyze these structures and procedures.

1.2 Background

The standard three-node relay channel includes one source node that wishes to transmit messages to a receiver at one destination node. A relay node, which does not have its own message to transmit, assists the communication from the source to destination. The relay channel was originally studied in [1] in 1971 and two fundamental relaying schemes, namely decode-and-forward (DF) and compress-and-forward (CF) were proposed in [2] in 1979. In DF, the relay recovers the source message, whereas in CF, the relay provides a description of its received signal. Although the analysis of relaying schemes started more than forty years ago, it was not until the early twenty-first century that they attracted significant attention from the cooperative network communication point of view [3, 4].

In DF and CF, the cooperation between the source and the relay differs in the way the codewords they transmit are constructed. In many cases, to cooperate with the source, the relay transmitted codeword contains the information about the received signal at the destination. This information may be regenerative information of the codeword transmitted by the source, or non-regenerative information that only contains some description of relay received signal. The codebook structure and encoding procedure at the source and the relay enable the relay to convey such information to the destination. The destination exploits the source-relay cooperation through the decoding procedure that it uses to recover the source message. It was shown in [2] that relay node enables more rate-efficient source-to-destination communication to be

achieved.

Despite the progress that has been made in the relay channel in the past forty years, especially the significant advances in the past fifteen years, there are still many unknowns about the relay channel at large. In fact, the capacity of the general three-node relay channel remains an open problem. Achieving higher rates, let alone finding the capacity, for the channels and networks with relay nodes poses great challenges to the information theory community.

The focus of this thesis is on investigating the codebook structure, encoding procedure and decoding procedure, in particular, those that incorporate CF constituents, to achieve higher rate or larger rate region in the communication network with relay nodes.

1.3 Overview

1.3.1 Generalized DF-CF Relaying

In [2, Theorem 7], a generalization of the DF and CF relaying is proposed. This generalization uses both the DF and CF codebook structures to generate a composite codeword. This codeword is in fact the superposition of the CF codeword on the DF codeword. In decoding, the receiver recovers the DF codeword and the CF codeword successively. This generalization provides a lower bound on the capacity of the general three-node relay channel. It was shown in [2] that using a specific codebook structure, the generalized DF-CF reduces to either DF or CF, and hence in those cases, the generalization does not provide a rate advantage over the constituent DF or CF schemes for the three-node relay channel. In contrast, in [5], numerical instances for which the generalization yields higher rate than both DF and CF were provided. However, no analytical guarantees to ensure that the generalized DF-CF can provide

a rate gain over DF and CF are available in the literature.

To study the gain of the generalized DF-CF scheme, we particularize the three-node relay channel to the Gaussian case. The source and the relay use Gaussian codebooks with average transmit power constraint. We then formulate the achievable rate problem as an optimization problem of power allocations. Albeit the problem is not convex, we are able to analyze the power allocation and the achievable rate by using the Karush-Kuhn-Tucker (KKT) [6] conditions. Using the KKT conditions, we identify the optimal power allocation in various signal-to-noise ratio (SNR) regions of the source-to-relay link. In these SNR regions, the optimal power allocation at the source and relay reduces the generalized DF-CF to the underlying DF or CF schemes. On the other hand, using the KKT conditions, we showed that there exist SNR regions, in which the generalized DF-CF is guaranteed to yield strictly higher rates than the underlying DF and CF schemes. Furthermore, this rate gain was shown to be upper bounded by 0.5 bits per channel-use (bpcu).

Although the rate gain of the generalized DF-CF scheme over individual schemes was shown not to be substantial, the analysis of the power allocation in various SNR regions reveals that a strategy of switching between DF and CF is able to yield significant gain over fixed DF or CF relaying. Using this insight, we evaluated the performance of DF-CF switching in the Rayleigh fading channel [7] through simulations. The numerical results confirmed our analysis and a significant rate gain was observed.

1.3.2 Exploiting the N -to-1 Mapping in CF Relaying

In CF, the relay does not decode the source transmitted codeword. Instead, the relay provides a description of its received signal. The information of this description is sent to the destination and “facilitates” the decoding at the receiver. CF relaying uses Wyner-Ziv binning [8] to encode this information in its transmitted codeword. In

particular, each description codeword is mapped to a bin index. In general, Wyner-Ziv binning results in an N -to-1 mapping from the codewords of the relay description of its received signal to its transmitted codewords that represents the bin indices. In decoding [2], the receiver successively recovers the bin index, the relay description of its received signal and then the source message.

In the recent development of CF-based relaying schemes, namely noisy network coding (NNC) [9] and short message NNC (SNNC) [10], the N -to-1 mapping from Wyner-Ziv binning has been simplified to a 1-to-1 mapping. It has been shown that there is no rate loss due to such simplification in the network considered therein, and NNC and SNNC can generally yield larger rate regions in comparison with CF in the multmessage networks considered therein.

To understand the impact of different mappings, we analyze the rate constraint on the relay transmission, i.e., the rate of the codewords that represent the bin indices, in these CF-based relaying schemes. In conventional CF, a rate constraint is imposed on the bin indices due to the fact that the receiver decodes the bin index first. In contrast, NNC and SNNC do not impose the same constraint. However, the 1-to-1 mapping implies that the rate of the bin indices equals that of the codewords representing the relay description of its received signal. This can be seen as an implicit constraint on the bin indices.

In this thesis, we propose a modified CF decoding procedure. This procedure uses the N -to-1 mapping as opposed to the 1-to-1 mapping. It uses forward decoding and is based on a layered framework. Using this procedure, we were able to obtain the rate expression for the same multmessage network in [9, 10]. Although the new procedure does not provide rate gain in that network, its rate expression subsumes that of NNC/SNNC and the conventional CF [2]. In other words, it unifies the conventional CF and NNC/SNNC.

An interesting outcome of this decoding procedure is that it is able to eliminate the

rate constraint on the relay transmission imposed by the decoding procedure of the conventional CF while it does not have the constraint inherent in the 1-to-1 mapping in NNC and SNNC. Hence, the proposed decoding procedure is able to relax the rate constraint on the relay transmission.

To exploit the gain of the N -to-1 mapping, we investigate two communication networks. The first network is a broadcast relay chain network, whereas the second considered network is a partial cooperative network, in which the relay has its own independent message to send to a destination over a point-to-point channel. We provided the achievable rate expressions for these networks for various relaying strategies. When particularized to Gaussian, we show that in both cases, adopting the proposed CF-based relaying scheme to exploit the N -to-1 mapping yields either a higher achievable rate or a larger rate region in comparison with using either NNC or SNNC which are constrained to the 1-to-1 mapping.

1.3.3 Military Application of the Advanced Relaying Scheme

A novel application of the proposed CF-based decoding procedure is studied in this thesis. In the military communication, a jammer attempts to disrupt the source-to-receivers communication by transmitting an independent high power Gaussian signal. We introduce a “friendly” eavesdropper in this channel. The friendly eavesdropper observes the jammer’s signal and assists the source-to-receivers communication. To analyze the data rate in this channel, we conceive the role of the friendly eavesdropper as a standard relay. We provide an upper bound on the achievable rate of this channel and we show that this bound can be achieved using the proposed relaying scheme.

1.4 Contributions

The key contributions of this thesis are summarized as the following:

- On Generalized DF-CF:
 - Development of a mechanism to analyze the power allocation and achievable rate of the generalized DF-CF scheme in the three-node Gaussian relay channel. The mechanism can be readily applied to the recently developed generalization in [11]. Using this mechanism, various SNR regions are identified in which the generalized DF-CF reduces to its underlying DF and CF. It is proved that there exists an SNR region in which the generalized DF-CF is guaranteed to yield strictly higher rates than the underlying DF and CF. The rate gain is shown to be upper bounded by 0.5 bpcu. This analysis implies that DF-CF switching can yield significant rate gain over fixed DF or CF, which is confirmed by the simulations.
 - These results appear in [12] and [13].
- On Exploiting the N -to-1 Mapping in CF:
 - Development of a structured forward decoding procedure with a layered framework that exploits the N -to-1 mapping in the CF codebook structure. This decoding procedure is able to relax the rate constraint on the relay transmitted codewords that represent the bin indices. Applying this procedure in two particular networks, we obtain
 - * a higher achievable rate in a broadcast Gaussian relay chain network,
 - * a larger achievable rate region in a Gaussian partial cooperative network,
 in comparison with NNC and SNNC.
 - The procedure and results are published in part in [14], and are described in [15], which is currently undergoing a second review cycle.
- On A Military Application of the New Decoding Procedure:

- Introducing of the concept of a “friendly” eavesdropper in a broadcast channel in the presence of jamming and showing that the capacity of this channel can be achieved by the proposed decoding procedure for CF relaying.
- The results are presented in [16], and are currently in preparation for a journal submission.

1.5 Organization

In Chapter 2, we will discuss the achievable rate of the generalized DF-CF relaying in [2, Theorem 7]. In Chapter 3, we will focus on the CF-based relaying scheme and propose a modified decoding procedure for CF relaying. We will show the advantage of this new procedure in two particular networks in Chapter 4. In Chapter 5, we will study a military application of the proposed relaying scheme in battle field communications in the presence of a malicious jammer.

1.6 Notation

In this thesis, regular-face upper and lower case letters will refer to random variables, and their corresponding realizations, respectively. Boldface letters will refer to length- n sequences, and the calligraphic font will be used to refer to sets of nodes. A sequence \mathbf{x} of an index s transmitted or selected by a node d_k in block b is denoted by $\mathbf{x}_k(s_{k,b})$. The jointly ϵ -typical set of length- n sequences is denoted by $\mathcal{A}_\epsilon^{(n)}$.

Chapter 2

Analysis of the Generalization of DF and CF in Gaussian Relay Channels: An Optimization-Based Approach

In this chapter, we consider a quasi-static communication system in which a relaying node is used to assist communication between a source-destination pair. Our goal is to determine the relaying mode that enables rate-efficient communication under given channel conditions. To achieve this goal, the generalized decode-and-forward and compress-and-forward relaying scheme is studied when the source and relay signals are synthesized from commonly-used Gaussian codebooks. These signals are shown to enable the capacity of the considered channel to be achieved in two asymptotic SNR regions. For two non-asymptotic SNR regions, the generalized DF-CF scheme is shown to reduce to either DF or CF, depending on which scheme yields a higher rate. For another non-asymptotic SNR region, the generalized DF-CF is shown to yield strictly higher rates than both DF and CF. Despite the coding and decoding complexity of the generalized DF-CF, its rate advantage over DF and CF is shown to be upper bounded by 0.5 bits per channel use. This indicates that the practical benefit of the analysis of this generalization is to enable appropriate selection of either

the DF or CF relaying scheme for any given channel realization. Numerical results show that, under Rayleigh fading channel conditions, this selection yields significant gains over fixed DF and CF. The work in this chapter was published, in part, in [12] and [13].

2.1 Introduction

Future advances in wireless communications are expected to bank on the substantial gains provided by effective cooperation between multiple nodes in the network [17]. In one form, cooperation is established when a single relay node assists communication between a source-destination pair [1, 18]. Relaying operation schemes can be categorized into either full-duplex operation, in which the relay uses the same physical channel for transmission and reception, and half-duplex operation, in which these channels are orthogonal. Due to its simplicity, half-duplex relaying is better understood than its full-duplex counterpart. However, half-duplex operation is generally wasteful of the resources available for communication, and higher data rates can be reliably communicated if the relay were to operate in a full-duplex mode [19].

Among the various cooperation schemes are the DF and CF techniques [2]. In DF, the relay decodes its observed signal and generates auxiliary information that assists the decoding of the transmitted signals at the destination. In contrast, in CF the relay does not decode its observed signal, but uses a description of its received signal to generate auxiliary information that facilitates decoding at the destination [2]. When the relay channel is Gaussian, the optimal distribution of the DF signals is Gaussian, whereas the optimal distribution of the CF signals is not known [20], except for a special asymptotic case wherein that distribution is also Gaussian [21]. Despite the potential rate loss incurred by using Gaussian signals with CF relaying at non-asymptotic SNRs, their asymptotic optimality and simplicity have made them

commonly used for studying CF relaying in Gaussian channels.

Performance analyses of the DF and CF schemes with Gaussian signals were conducted in [20] for special classes of relay channels. For the case of low-power relays and fading channels, an analysis of the rates achievable by DF and time-division multiplexing was conducted in [22], and comparisons of the DF scheme with direct transmission and other relaying schemes were considered in [23]. In [24], it was shown that, for Gaussian relay channels and Gaussian signals, the gap between the rate achieved by the CF scheme and capacity is upper bounded by a constant. A relaying scheme in which the relay switches to a CF operating mode if decoding fails was analyzed in [25]. Further extensions of the DF and CF schemes are considered in [26] for Gaussian channels with correlated relay and destination noises. The application of DF and CF schemes have been extended to scalar Rayleigh fading channels in [20,27] and to multiple antenna channels in [28]. The DF and CF schemes with Gaussian signals have also been used in more general multi-terminal networks with multiple sources, relays and destinations [29].

Despite the envisioned advantages of cooperative communications, the capacity of relay channels, including scalar ones with one relay, remains an open problem, and only partial results are available. An upper bound on the capacity of relay channels is given by the cut-set bound [18]. This bound is attained by most of the relaying strategies that are known to achieve capacity [2,21,30–33]. In contrast, an interesting result in [34] showed that the capacity of the relaying channel considered therein can be strictly below the cut-set bound. Another relaying technique known as quantize-map-forward was proposed in [35] and [36] for Gaussian relay networks and was shown to achieve a rate within a constant gap to the cut-set bound, and hence to capacity.

The generalized DF-CF scheme considered in this thesis was originally proposed in [2, Theorem 7]. The key idea of this generalization is to explore the potential of

achieving higher rates by superimposing the DF and CF coding strategies. In particular, in this generalization the relay combines partial decoding and the description of its received signal to generate the auxiliary information to assist decoding at the destination. This generalization resembles, to some extent, the philosophy that underlies the generalization in [25]. However, it avoids unsuccessful decoding attempts at the relay and provides deterministic thresholds beyond which the relay ought to switch its operating mode.

The generalized DF-CF scheme has received significantly less attention than either DF or CF. An instance in which a special case of this scheme was used to develop a lower bound on the capacity of the static relay channel was considered in [21]. Another instance was considered in [37], wherein a particular geometric setup of a Gaussian relay network was considered and a numerical demonstration of the rate advantage of the generalized DF-CF scheme over DF and CF was provided. Building on the generalized DF-CF scheme, other potentially more advantageous, mixed DF and CF schemes have been recently proposed in [5, 38, 39]. Similar to the generalization proposed in [2, Theorem 7], the mixed schemes proposed in [5, 38, 39] rely on the fundamental mechanism of superimposing DF and CF. Hence, the analysis methodology proposed herein for the generalized DF-CF scheme in [2, Theorem 7] can be extended to those mixed strategies. Numerical examples that demonstrate the potential advantage of the generalized DF-CF [2, Theorem 7] with Gaussian signals over the individual DF and CF schemes were provided in [5]. However, in that work, the SNR regions at which the generalized DF-CF reduces to either the DF or the CF scheme were not identified, nor those at which the generalization is guaranteed to yield strictly higher rates. In fact, the analysis of the generalized DF-CF in [5] did not prove the existence of distinct SNR operation regions and did not bound the maximum rate advantage.

In this chapter, we consider a full-duplex single relay communication system in

which the source transmissions are received by both the relay and the destination. The relay does not have its own message and its output depends either deterministically or stochastically on its observed signal [1,18]. The destination uses a noisy combination of the source and relay signals to decode its intended message. We consider the generalized DF-CF scheme in [2, Theorem 7]. Since this scheme subsumes both DF and CF, it will enable us to determine the appropriate relaying mode corresponding to a given SNR and will also enable us to obtain explicit expressions for SNR thresholds at which the relay ought to switch its operating mode, depending on the given channel realization. Numerical results show that, for quasi-static Rayleigh fading channel conditions, this selection can yield significant gains over fixed DF and CF. The SNR thresholds and the bound on the rate of the generalized scheme are obtained by particularizing the generalized DF-CF in [2, Theorem 7] to the case in which the source and relay signals are Gaussian and the relay channel is also Gaussian. The KKT optimality conditions corresponding to the problem of maximizing the achievable rate are then analyzed to draw insight into the generalized DF-CF scheme and to obtain the aforementioned results. A similar approach can be used to analyze the mixed DF-CF strategies in [5,38,39], but this is beyond the scope of this thesis. To obtain a better understanding of the generalized DF-CF, we further analyze its maximum achievable rate and show that it is at most 0.5 bpcu higher than that achieved by the underlying DF and CF schemes.

2.2 System Model

We consider the standard three-node Gaussian relay channel depicted in Fig. 1.

In this channel, the gain of the source-to-destination link is normalized, and the gains of the source-to-relay and the relay-to-destination links are denoted by a and b , respectively. The source and relay transmit signals are denoted by X_1 and X_2 ,

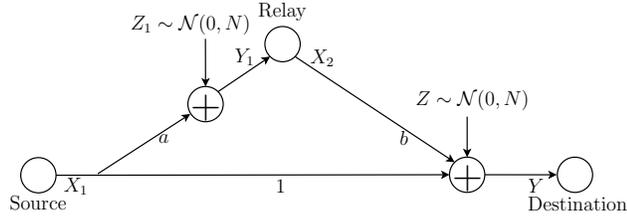


Figure 1: Gaussian three-node relay channel.

respectively, and the additive Gaussian noises at the relay and destination are independent and are denoted by $Z_1 \sim \mathcal{N}(0, N)$ and $Z \sim \mathcal{N}(0, N)$, respectively. The relay and destination received signals are respectively denoted by Y_1 and Y , and can be expressed as

$$Y_1 = aX_1 + Z_1, \quad (1a)$$

$$Y = X_1 + bX_2 + Z. \quad (1b)$$

The transmit power budgets at the source and relay are denoted by P_1 and P_2 , respectively. Using this notation, we define

$$\gamma_0 \triangleq \frac{P_1}{N}, \gamma_1 \triangleq \frac{a^2 P_1}{N} \text{ and } \gamma_2 \triangleq \frac{b^2 P_2}{N} \quad (2)$$

to be the SNR of the source-to-destination, the source-to-relay, and the relay-to-destination links, respectively.

We assume that the source and relay transmit signals are Gaussian distributed. Next, we will proceed to review the generalized DF-CF scheme described in [2, Theorem 7] and particularize it to the Gaussian channel in Fig. 1.

2.3 Generalization of DF and CF for Gaussian Relay Channels

In this section, we provide a concise information theoretic background on the generalized DF-CF scheme and its application to Gaussian signals.

2.3.1 An Information-Theoretic Background

Codebook structure: The generalized DF-CF codebook structure is described in [2,38] and is reviewed here for completeness. The codebooks \mathcal{V} and \mathcal{U} are available at both the source and the relay. The codebook \mathcal{X}_1 is available at the source, and the codebook \mathcal{X}_2 is available at the relay. At the source, each $\mathbf{v} \in \mathcal{V}$ represents a centre of a partition of \mathcal{U} , and each $\mathbf{u} \in \mathcal{U}$ represents a centre of a partition of \mathcal{X}_1 , that is, each $\mathbf{u} \in \mathcal{U}$ plays a dual role: a satellite of a codeword $\mathbf{v} \in \mathcal{V}$ and a centre of a partition of \mathcal{X}_1 . At the relay, each $\mathbf{v} \in \mathcal{V}$ represents a centre of two partitions: a partition of \mathcal{U} and a partition of \mathcal{X}_2 . Each $(\mathbf{u}, \mathbf{x}_2)$ pair corresponds to a centre of a partition of an estimation codebook $\hat{\mathcal{Y}}_1$. The partitions of \mathcal{U} , \mathcal{X}_1 , \mathcal{X}_2 and $\hat{\mathcal{Y}}_1$ are non-overlapping and the codebooks are the unions of the respective partitions.

We note that the message indices are encoded in the auxiliary information in \mathbf{u} and \mathbf{x}_1 .

An achievable rate: In the structure described above, the cooperation between the source, relay and destination is established using three auxiliary random variables, which are denoted by U , V and \hat{Y}_1 . Using these variables and the construction , the

following achievable rate expression is provided in [2]:

$$R_G^* = \sup \left\{ \min \left\{ I(X_1; Y, \hat{Y}_1 | X_2, U) + I(U; Y_1 | X_2, V), \right. \right. \\ \left. \left. I(X_1, X_2; Y) - I(\hat{Y}_1; Y_1 | X_2, X_1, U, Y) \right\} \right\}, \quad (3a)$$

where $I(X; Y)$ is the standard mutual information functional for the random variables X and Y [18], and the supremum is taken over joint distributions of $(V, U, X_1, X_2, Y, Y_1, \hat{Y}_1)$ of the form:

$$p(u, v, x_1, x_2, y, y_1, \hat{y}_1) = p(v)p(u|v)p(x_1|u)p(x_2|v)p(y, y_1|x_1, x_2)p(\hat{y}_1|x_2, y_1, u).$$

that satisfy

$$I(\hat{Y}_1; Y_1 | Y, X_2, U) \leq I(X_2; Y | V). \quad (3b)$$

The cut-set bound: An upper bound on the capacity, C , of a general multi-terminal channel is given by the cut-set bound in [18]. Applying this bound to the standard three-node relay channel yields [2, Theorem 4]

$$C \leq \sup_{p(x_1, x_2)} \min \left\{ I(X_1; Y, Y_1 | X_2), I(X_1, X_2; Y) \right\}. \quad (4)$$

2.3.2 Application to Gaussian Signals

We now derive the expressions corresponding to (3a) and (3b) when the relay channel and the signals are Gaussian. To map the codebook structure described in [2, 38] to Gaussian signals, we define the power partitions $\{\alpha_i\}_{i=0}^2$ to be used at the source for constructing U and X_1 , and $\{\beta_j\}_{j=0}^1$ to be used at the relay for constructing V and X_2 . The power partitions satisfy $\alpha_i, \beta_j \geq 0$, $i = 0, 1, 2$, $j = 0, 1$, $\sum_{i=0}^2 \alpha_i = 1$, and $\sum_{j=0}^1 \beta_j = 1$. For the Gaussian signals, we let $V \sim \mathcal{N}(0, 1)$, $V_1 = \sqrt{\alpha_0 P_1} V$, and

$$V_2 = \sqrt{\beta_0 P_2} V.$$

Consider the construction of the auxiliary codebook \mathcal{U} described in Section 2.3.1. This construction implies that each signal U can be expressed as the superposition of a (base) component V_1 and another independent component that contains incremental information that is decodable at the relay, which we denote by X_{11} . The corresponding codebook is denoted by \mathcal{X}_{11} . The signals X_1 , X_2 are constructed similarly using base components U and V_2 and incremental components X_{12} and X_{22} , respectively.

Using the approach in [20] and [21], the auxiliary signal \hat{Y}_1 can be expressed as the superposition of Y_1 , X_2 and U , as a base component, and a statistically independent estimation noise Z' , as an incremental component. In particular, we can write

$$U = V_1 + X_{11}, \quad X_1 = U + X_{12}, \quad (5a)$$

$$X_2 = V_2 + X_{22}, \quad \hat{Y}_1 = Y_1 + X_2 + U + Z', \quad (5b)$$

where $Z' \sim \mathcal{N}(0, N')$. Note that, in this construction, V , X_{11} and X_{22} are mutually statistically independent, and U and X_{12} are statistically independent; X_1 and X_2 are correlated through V . The source uses the power fraction $\alpha_0 P_1$ to transmit V_1 , the power fraction $\alpha_1 P_1$ to transmit X_{11} and the power fraction $\alpha_2 P_1$ to transmit X_{12} . The relay uses the power fraction $\beta_0 P_2$ to transmit V_2 and the power fraction $\beta_1 P_2$ to transmit X_{22} .

2.4 The Generalized DF-CF: Achievable Rate and Analysis

We begin the analysis of the achievable rate of the generalized DF-CF scheme.

Using the notation of Section 2.2 and the construction of the Gaussian signals in Section 2.3.2, we obtain the following result.

Proposition 1. *Using $\mathcal{C}(x) \triangleq \frac{1}{2} \log_2(1+x)$, $x \geq 0$ and $\gamma' \triangleq N'/N$ and applying the generalized DF-CF scheme with Gaussian signal components to the Gaussian relay channel in Figure 1 yields the following rate:*

$$R_G^* = \max_{\{\alpha_i\}_{i=0}^2, \{\beta_j\}_{j=0}^1, \gamma'} \min\{R_1, R_2\}, \quad (6a)$$

subject to

$$\gamma' \geq \frac{(1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_0)}{(1 + \alpha_2\gamma_0)\beta_1\gamma_2}, \quad (6b)$$

$$R_1 = \mathcal{C}\left(\frac{\alpha_2\gamma_1}{1 + \gamma'} + \alpha_2\gamma_0\right) + \mathcal{C}\left(\frac{\alpha_1\gamma_1}{1 + \alpha_2\gamma_1}\right), \quad (6c)$$

$$R_2 = \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\beta_0\gamma_0\gamma_2}) - \mathcal{C}(1/\gamma'), \quad (6d)$$

$$\sum_{i=0}^2 \alpha_i = 1, \quad \sum_{j=0}^1 \beta_j = 1, \quad (6e)$$

$$\alpha_i \geq 0, \quad \beta_j \geq 0, \quad \forall i, j, \quad (6f)$$

Proof. See details in Appendix A.1. □

In subsequent analysis, we will seek solutions of the optimization problem in (6) for different SNR regions. Towards that end, we will recast (6) in a more convenient form. First, we introduce a new variable $t \geq 1$ such that $\frac{1}{2} \log(t)$ is a lower bound on R_1 and R_2 in the objective in (6a). Then, invoking the fact that the $\log(\cdot)$ function is monotonically increasing, the optimization problem in (6) is recast as

$$\max \quad t, \quad (7a)$$

$$\text{subject to} \quad g_i(t, \alpha_0, \alpha_2, \beta_0, \gamma') \leq 0, \quad i = 1, \dots, 8, \quad (7b)$$

where

$$\begin{aligned}
g_1(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq t - \left(1 + \frac{\alpha_2\gamma_1}{1 + \gamma'} + \alpha_2\gamma_0\right) \frac{1 + (1 - \alpha_0)\gamma_1}{1 + \alpha_2\gamma_1}, \\
g_2(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq t - \left(1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\beta_0\gamma_0\gamma_2}\right) \frac{\gamma'}{1 + \gamma'}, \\
g_3(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq 1 + \frac{\alpha_2\gamma_1}{1 + \alpha_2\gamma_0} - \frac{(1 - \beta_0)\gamma_2}{1 + (1 - \alpha_0)\gamma_0} \gamma', \\
g_4(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq -\alpha_2, \\
g_5(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq -\alpha_0, \\
g_6(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq \alpha_0 + \alpha_2 - 1, \\
g_7(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq \beta_0 - 1, \\
g_8(t, \alpha_0, \alpha_2, \beta_0, \gamma') &\triangleq -\beta_0.
\end{aligned}$$

In the formulation in (7) we have used the fact that $\alpha_i \geq 0$, $i = 0, 1, 2$ and $\beta_j \geq 0$, $j = 0, 1$ and the fact that $\sum_{i=0}^2 \alpha_i = 1$ and $\sum_{j=0}^1 \beta_j = 1$ to eliminate α_1 and β_1 .

It can be readily verified that the optimization in (7) is not convex. However, it satisfies the qualification conditions and hence the KKT conditions are necessary for optimality [40]. Letting λ_i denote the Lagrange multipliers corresponding to the constraint g_i in (7b), respectively, $i = 1, \dots, 8$, the Lagrangian can be written as To write these conditions, we consider the Lagrangian function, which is given by

$$\begin{aligned}
L = &(\lambda_1 + \lambda_2 - 1)t + \lambda_1 \left(1 + \frac{\alpha_2\gamma_1}{1 + \gamma'} + \alpha_2\gamma_0\right) \frac{1 + (1 - \alpha_0)\gamma_1}{1 + \alpha_2\gamma_1} \\
&+ \lambda_2 \left(1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\beta_0\gamma_0\gamma_2}\right) \frac{\gamma'}{1 + \gamma'} \\
&+ \lambda_3 \left(1 + \frac{\alpha_2\gamma_1}{1 + \alpha_2\gamma_0} - \frac{(1 - \beta_0)\gamma_2}{1 + (1 - \alpha_0)\gamma_0} \gamma'\right) \\
&- \lambda_4\alpha_2 - \lambda_5\alpha_0 + \lambda_6(\alpha_0 + \alpha_2 - 1) + \lambda_7(\beta_0 - 1) - \lambda_8\beta_0.
\end{aligned}$$

Using this Lagrangian, the KKT system can be written as

$$\lambda_1 + \lambda_2 - 1 = 0, \quad (8a)$$

$$\lambda_1 \left(1 + \alpha_2 \gamma_0 + \frac{\alpha_2 \gamma_1}{1 + \gamma'} \right) \left(\frac{\gamma_1}{1 + \alpha_2 \gamma_1} \right) - \lambda_2 \sqrt{\gamma_0 \gamma_2} \sqrt{\frac{\beta_0}{\alpha_0}} \left(\frac{\gamma'}{1 + \gamma'} \right) - \lambda_3 \frac{\gamma_0 \gamma_2 \gamma' (1 - \beta_0)}{(1 + (1 - \alpha_0) \gamma_0)^2} - \lambda_5 + \lambda_6 = 0, \quad (8b)$$

$$- \lambda_1 \frac{(1 + \gamma_1 (1 - \alpha_0)) (\gamma_0 (1 + \gamma') - \gamma_1 \gamma')}{(1 + \gamma') (1 + \alpha_2 \gamma_1)^2} + \lambda_3 \frac{\gamma_1}{(1 + \alpha_2 \gamma_0)^2} - \lambda_4 + \lambda_6 = 0, \quad (8c)$$

$$- \lambda_2 \sqrt{\gamma_0 \gamma_2} \sqrt{\frac{\alpha_0}{\beta_0}} \left(\frac{\gamma'}{1 + \gamma'} \right) + \lambda_3 \frac{\gamma_2 \gamma'}{1 + (1 - \alpha_0) \gamma_0} + \lambda_7 - \lambda_8 = 0, \quad (8d)$$

$$\lambda_1 \frac{\gamma_1 \alpha_2 (1 + \gamma_1 (1 - \alpha_0))}{(1 + \gamma')^2 (1 + \gamma_1 \alpha_2)} - \lambda_2 \frac{1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}}{(1 + \gamma')^2} - \lambda_3 \frac{\gamma_2 (1 - \beta_0)}{1 + \gamma_0 (1 - \alpha_0)} = 0, \quad (8e)$$

$$g_i(t, \alpha_0, \alpha_2, \beta_0, \gamma') \leq 0, \quad \lambda_i g_i(t, \alpha_0, \alpha_2, \beta_0, \gamma') = 0, \quad \lambda_i \geq 0, \quad i = 1, 2, \dots, 8, \quad (8f)$$

where (8a), (8b), (8c), (8d) and (8e) can be obtained, respectively, by setting $\frac{\partial L}{\partial t}$, $\frac{\partial L}{\partial \alpha_0}$, $\frac{\partial L}{\partial \alpha_2}$, $\frac{\partial L}{\partial \beta_0}$ and $\frac{\partial L}{\partial \gamma'}$ to zero.

To study the KKT system for different SNR regions, we begin by determining the values of α_0 , α_2 , β_0 and γ' for which the generalized DF-CF scheme reduces to the DF and CF schemes.

We begin by providing achievable rates of the DF and CF schemes for the Gaussian channel in Figure 1. For the DF scheme, the maximum achievable rate is achieved with Gaussian signals [20, Proposition 2] and is given by

$$R_{\text{DF}}^* = \max_{\{\rho_i\}_{i=0}^1} \min \{ R_{\text{DF},1}, R_{\text{DF},2} \}, \quad (9)$$

where $R_{\text{DF},1} = \mathcal{C}(\rho_1 \gamma_1)$ and $R_{\text{DF},2} = \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\rho_0 \gamma_0 \gamma_2})$, ρ_0 and ρ_1 satisfy $\sum_{i=0}^1 \rho_i = 1$, and $\sqrt{\rho_0} = \text{E}\{X_1 X_2\}$, where $\text{E}\{\cdot\}$ denotes the expectation operator;

cf. [2, Theorem 1]. See details in Appendix A.2. In contrast with the DF scheme, in the CF scheme it is not known if using Gaussian signalling achieves its maximum rate. However, the maximum rate achieved with Gaussian signals is given by

$$R_{\text{CF}}^* = \mathcal{C}\left(\gamma_0 + \frac{\gamma_1\gamma_2}{1 + \gamma_0 + \gamma_1 + \gamma_2}\right), \quad (10)$$

where the normalized variance of the estimation noise that enables this rate to be achieved is given by

$$\gamma_{\text{CF}}^* \triangleq \frac{1 + \gamma_0 + \gamma_1}{\gamma_2}. \quad (11)$$

See details in Appendix A.2.

Remark 1. Let $\gamma_i \in (0, \infty)$, $i = 0, 1, 2$. Then, setting $\alpha_2 = 0$ and $\beta_0 = 1$, the generalized DF-CF scheme yields the maximum achievable rate of the DF scheme, R_{DF}^* . Furthermore, setting $\alpha_2 = 1$ and $\beta_1 = 1$, the generalized DF-CF scheme yields the maximum achievable rate of the CF scheme with Gaussian signals, R_{CF}^* . In particular, $R_{\text{G}}^* \Big|_{\substack{\alpha_2=0 \\ \beta_1=0}} = R_{\text{DF}}^*$, and $R_{\text{G}}^* \Big|_{\substack{\alpha_2=1 \\ \beta_1=1}} = R_{\text{CF}}^*$, with $\gamma'^* \Big|_{\substack{\alpha_2=1 \\ \beta_1=1}} = \gamma_{\text{CF}}^*$.

Proof. Direct substitution in (6) yields the statements of the remark. \square

We now analyze the rates yielded by the problem in (6) for different choices of γ_0 , γ_1 and γ_2 . In particular, we will partition γ_1 into different regions and consider the case of $\gamma_0(1 + \gamma_2) < \gamma_1 < \infty$, the case of $\gamma_1 \leq \gamma_0$, and the case of SNRs that yield $R_{\text{DF}}^* = R_{\text{CF}}^*$ to draw insight into the gains of the generalized DF-CF scheme. We will also touch upon two asymptotic cases, viz., $\gamma_2 \rightarrow \infty$ and $\gamma_0 \rightarrow 0$.

2.4.1 The Case of $\gamma_0(1 + \gamma_2) < \gamma_1 < \infty$

In this region the gain of the source-to-relay link is greater than that of the source-destination link; i.e., $a > 1$ in Fig. 1. For this case, we first record an observation in the following lemma.

Lemma 1. *When $\gamma_1 > \gamma_0(1 + \gamma_2)$, Proposition 1 yields that γ' must satisfy*

$$\gamma' > \frac{\gamma_0}{\gamma_1 - \gamma_0}. \quad (12)$$

Proof. We proceed by contradiction. Suppose that $\gamma' \leq \frac{\gamma_0}{\gamma_1 - \gamma_0}$. In this case, the constraint in (6b) yields $\frac{(1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_0)}{(1 + \alpha_2\gamma_0)\beta_1\gamma_2} \leq \frac{\gamma_0}{\gamma_1 - \gamma_0}$, which, by the fact that $\beta_1 \leq 1$, implies that

$$(1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_0) \leq \frac{\gamma_0}{\gamma_1 - \gamma_0}(1 + \alpha_2\gamma_0)\gamma_2.$$

Manipulating this inequality and using $\sum_{i=0}^2 \alpha_i = 1$ yields

$$\frac{\gamma_0(1 + \gamma_2) - \gamma_1}{\gamma_0(\gamma_1 - \gamma_0)} - \frac{\alpha_2\gamma_1\gamma_2}{(\gamma_1 - \gamma_0)(1 + \alpha_2(\gamma_0 + \gamma_1))} \geq 1 - \alpha_0,$$

which requires $\frac{\gamma_0(1 + \gamma_2) - \gamma_1}{\gamma_0(\gamma_1 - \gamma_0)} \geq 1 - \alpha_0$. If $\gamma_1 > \gamma_0(1 + \gamma_2)$, this inequality implies that $\alpha_0 > 1$, which violates the constraint that $\alpha_0 \leq 1$ in (6e). Hence, we conclude that when $\gamma_1 > \gamma_0(1 + \gamma_2)$, we must have $\gamma' > \frac{\gamma_0}{\gamma_1 - \gamma_0}$. \square

Now, we provide our main result for the case $\gamma_0(1 + \gamma_2) < \gamma_1 < \infty$ in the following theorem.

Theorem 1. *When the channel gains in Fig. 1 satisfy $\gamma_1 > \gamma_0(1 + \gamma_2)$, the generalized DF-CF scheme using Gaussian signals reduces to the DF scheme. In particular,*

$$R_G^* = R_{\text{DF}}^*. \quad (13)$$

Proof. Detailed proof is provided in Appendix A.3. \square

Theorem 1 indicates that if the channel gain of the source-relay link is sufficiently large, compared to the other two links, the generalized DF-CF with Gaussian signals

does not yield rates higher than those achieved by the DF scheme. In this region DF outperforms CF.

Remark 2. *By comparing $R_1 \leq \mathcal{C}(\alpha_1\gamma_1)$ in (64) with $R_2 \leq \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\gamma_0\gamma_2})$ in (65) in Appendix A.3, it can be seen that when $\gamma_0(1 + \gamma_2) < \gamma_1 < \gamma_0 + \gamma_2$, $R_1 < R_2$; i.e., R_1 is the constraining rate. In this case, β_0 can be set to be less than 1 without reducing the achievable rate of the generalized DF-CF scheme. Setting $\beta_0 < 1$ implies that the relay uses a fraction $\beta_1 > 0$ of its power to transmit its estimation of the received signal. Setting $\alpha_1 = 1$ and restricting R_2 to be greater than or equal to R_1 yields $\beta_1 \leq \frac{1+\gamma_0}{1+\gamma_1} \left(1 - \frac{\gamma_1-\gamma_0}{\gamma_2}\right)$. \square*

This remark implies that the generalized DF-CF can be beneficial in scenarios that are more general than the considered three-node one. Drawing insight from this remark, we note that for the generalized DF-CF scheme with $\beta_1 > 0$ to yield the same rate as the DF scheme, the source encodes the message and the relay successfully decodes it as in the DF scheme. The relay uses the decoded message and an estimation of the received signal to synthesize its transmitted codeword. However, the estimation of the received signal will not provide the destination with additional information about the codeword transmitted by the source, and any additional information it may contain will depend on the noise at the relay. Because this noise and the noise at the destination are independent, the additional information potentially contained in the estimation portion of the signal at the destination does not facilitate decoding. This implies that, in this SNR region, the generalized DF-CF scheme may be advantageous than the DF and CF schemes considered in [26] for Gaussian channels when the relay and destination noises are correlated.

2.4.2 The Case of $\gamma_1 \leq \gamma_0$

When $\gamma_1 \leq \gamma_0$, the channel gain of the source-to-relay channel is less than or equal to that of the source-to-destination channel, which corresponds to the case of $a \leq 1$ in Fig. 1.

To arrive at the main result in this SNR region, we first provide the following two lemmas.

Lemma 2. *For the Gaussian channel shown in Fig. 1, the maximum rate that the generalized DF-CF scheme achieves using Gaussian signals is attained when the constraint in (6b) is satisfied with equality, that is, in that case,*

$$I(\hat{Y}_1; Y_1 | X_2, U, Y) = I(X_2; Y | V). \quad (14)$$

Proof. We consider the channel in Fig. 1 for the case of the Gaussian signals and Gaussian channels. To prove this lemma, consider the following three cases of α_2 and γ' :

1. $\alpha_2 = 0$, arbitrary γ' ;
2. $\alpha_2 > 0, \gamma' < \infty$; and
3. $\alpha_2 > 0, \gamma' \rightarrow \infty$.

We will show that in each case if the maximum rate of the generalized DF-CF scheme is attained, the constraint on γ' in (6b) is satisfied with equality. See details in Appendix A.4.

□

Lemma 3. *Rate R_1 and R_2 in (3) (cf. Theorem 7 in [2]) satisfy*

$$R_1 = R_2 + \Delta, \quad (15)$$

where $\Delta = -I(X_2, U; Y) + I(X_2; Y|V) + I(U; Y_1|X_2, V)$.

Proof. Invoking Lemma 2 in the expression of R_2 in Theorem 7 in [2] yields

$$\begin{aligned}
R_2 &= I(X_1, X_2; Y) - I(\hat{Y}_1; Y_1|X_2, X_1, U, Y) \\
&= I(X_1, X_2; Y) - I(\hat{Y}_1; Y_1|X_2, X_1, U, Y) - I(X_2; Y|V) + I(\hat{Y}_1; Y_1|X_2, U, Y) \\
&= I(X_1; Y|X_2) + I(X_1; \hat{Y}_1|X_2, U, Y) + I(V; Y) \\
&= I(V; Y) + I(U; Y|X_2, V) + I(X_1; Y, \hat{Y}_1|X_2, U), \tag{16}
\end{aligned}$$

where the second equality follows from Lemma 2. The third and fourth equalities follows from using the chain rule and the fact that, by construction, (V, X_2, Y) form the Markov chain $V \rightarrow X_2 \rightarrow Y$ and that, conditional on (X_2, U) , (X_1, Y_1, \hat{Y}_1) form the Markov chain $X_1 \rightarrow Y_1 \rightarrow \hat{Y}_1$, and (U, X_1, Y) form the Markov chain $U \rightarrow X_1 \rightarrow Y$.

Substituting for $I(X_1; Y, \hat{Y}_1|X_2, U)$ from (16) in the first term of R_1 , in (3a) yields

$$\begin{aligned}
R_1 &= R_2 - I(V; Y) - I(U; Y|X_2, V) + I(U; Y_1|X_2, V) \\
&= R_2 - I(X_2, U, V; Y) + I(X_2; Y|V) + I(U; Y_1|X_2, V) \\
&= R_2 - I(X_2, U; Y) + I(X_2; Y|V) + I(U; Y_1|X_2, V) \tag{17} \\
&= R_2 + \Delta,
\end{aligned}$$

where $\Delta \triangleq -I(X_2, U; Y) + I(X_2; Y|V) + I(U; Y_1|X_2, V)$. In writing (17) we have used the fact that, conditioned on (X_2, U) , Y is independent of V . \square

The main result for the SNR region $\gamma_1 \leq \gamma_0$ is provided in the following theorem.

Theorem 2. *When the channel gains in Fig. 1 satisfy $\gamma_1 \leq \gamma_0$, the generalized DF-CF*

scheme using Gaussian signals reduces to the corresponding CF scheme. In particular,

$$R_G^* = R_{CF}^*. \quad (18)$$

Proof. Detailed proof is provided in Appendix A.5. \square

Theorem 2 suggests that when the signals are Gaussian and the channel gain of the source-to-relay link is less than or equal to that of the source-destination link, the generalized DF-CF scheme does not provide higher rate than the CF scheme. In this region, CF outperforms DF. To draw further insight into this theorem, we note that when, $\gamma_1 \leq \gamma_0$, the rate of the DF scheme is constrained by the rate that can be reliably communicated over the source-relay link. Now, in the generalized DF-CF scheme, the partial codewords $\{\mathbf{u}\} = \mathcal{U}$ contain the messages to be decoded by the relay. However, since the source-relay link is weaker than the source destination-link, the rate of \mathcal{U} must be sufficiently small to ensure its successful decoding. In fact, Theorem 2 asserts that, when the signals are Gaussian, this rate must be exactly zero.

2.4.3 The Case of SNRs Yielding $R_{DF}^* = R_{CF}^*$

In this section we will show that in the SNR region at which $R_{DF}^* = R_{CF}^*$, the rate yielded by the generalized DF-CF scheme is strictly higher than the rate yielded by either scheme. We begin by stating the following lemma.

Lemma 4. *For the Gaussian channel shown in Figure 1, the DF scheme and the CF scheme using Gaussian signals yield the same maximum achievable rate if and only if*

$$\gamma_1(1 + \gamma_1) = \gamma_0(1 + \gamma_0 + \gamma_2) \quad (19)$$

Proof. To prove the direct part of this lemma, we substitute $\gamma_1(1 + \gamma_1) = \gamma_0(1 + \gamma_0 + \gamma_2)$

in (10). Simplifying yields $R_{\text{CF}}^* = \mathcal{C}(\gamma_1)$. In this SNR region, it can be seen that $\gamma_1 < \gamma_0 + \gamma_2$. Using this in (9), we have $R_{\text{DF},1} < R_{\text{DF},2}$, which implies that $R_{\text{DF}}^* = R_{\text{DF},1}^* = \mathcal{C}(\gamma_1)$. This completes the proof of the direct part.

To prove the converse, we need to show that if $R_{\text{DF}}^* = R_{\text{CF}}^*$ then $\gamma_1(1 + \gamma_1) = \gamma_0(1 + \gamma_0 + \gamma_2)$. We begin by providing an upper bound on the maximum achievable rate of the CF scheme using Gaussian signals, R_{CF}^* . From (10), we have, for any $\rho_0 \geq 0$,

$$\begin{aligned} R_{\text{CF}}^* &= \mathcal{C}\left(\gamma_0 + \frac{\gamma_1\gamma_2}{1 + \gamma_0 + \gamma_1 + \gamma_2}\right) \\ &< \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\rho_0\gamma_0\gamma_2}). \end{aligned}$$

Comparing this bound with (9), it can be seen that $R_{\text{CF}}^* < R_{\text{DF},2}$. This implies that if the maximum achievable rate of the DF scheme is given by $R_{\text{DF}}^* = R_{\text{DF},2} \leq R_{\text{DF},1}$, then $R_{\text{CF}}^* < R_{\text{DF}}^*$. Hence for R_{DF}^* to be equal to R_{CF}^* , $R_{\text{DF},1}$ must be the constraining rate of the DF scheme, i.e., $R_{\text{DF}}^* = R_{\text{DF},1} < R_{\text{DF},2}$. However, from (9), this implies that $\gamma_1 < \gamma_0 + \gamma_2$. Otherwise, if this condition is not satisfied, the maximum rate of the DF scheme is achieved with $\rho_i \geq 0$, $i = 0, 1$, such that $\sum_{i=0}^1 \rho_i = 1$ and $R_{\text{DF},1} = R_{\text{DF},2}$.

From (9), it can be seen that, for $\gamma_1 < \gamma_0 + \gamma_2$, $R_{\text{DF}}^* = R_{\text{DF},1} = \mathcal{C}(\gamma_1)$. Equating this rate with R_{CF}^* in (10) yields $\gamma_1(1 + \gamma_1) = \gamma_0(1 + \gamma_0 + \gamma_2)$, which completes the proof of the lemma. \square

An important implication of this lemma is given in the following corollary.

Corollary 1. *When $\gamma_1(1 + \gamma_1) > \gamma_0(1 + \gamma_0 + \gamma_2)$, $R_{\text{DF}}^* > R_{\text{CF}}^*$ and when $\gamma_1(1 + \gamma_1) < \gamma_0(1 + \gamma_0 + \gamma_2)$, $R_{\text{DF}}^* < R_{\text{CF}}^*$.*

Proof. The SNR region $\gamma_1(1 + \gamma_1) > \gamma_0(1 + \gamma_0 + \gamma_2)$ can be divided into two non-overlapping sub-regions: $\gamma_1 \geq \gamma_0 + \gamma_2$ and $\gamma_0 + \gamma_2 > \gamma_1 > \frac{\gamma_0(1 + \gamma_0 + \gamma_2)}{1 + \gamma_1}$.

For SNRs satisfying $\gamma_1 \geq \gamma_0 + \gamma_2$, it can be seen from (9) that $R_{\text{DF}}^* = R_{\text{DF},1} = R_{\text{DF},2} > \mathcal{C}(\gamma_0 + \gamma_2)$. Now, the expression of R_{CF}^* in (10) yields $R_{\text{CF}}^* < \mathcal{C}(\gamma_0 + \gamma_2)$. Hence, in this SNR region, $R_{\text{DF}}^* > R_{\text{CF}}^*$.

We now consider the SNR region in which $\gamma_0 + \gamma_2 > \gamma_1 > \frac{\gamma_0(1+\gamma_0+\gamma_2)}{1+\gamma_1}$. In this region, we have from (9) that $R_{\text{DF}}^* = \mathcal{C}(\gamma_1) = R_{\text{DF},1} < R_{\text{DF},2}$. Using the fact that in this region $\gamma_1 > \frac{\gamma_0(1+\gamma_0+\gamma_2)}{1+\gamma_1}$ in (10) yields $R_{\text{DF}}^* > R_{\text{CF}}^*$, which completes the proof of the first statement.

For the second statement of the corollary, it can be verified that the condition that $\gamma_1(1+\gamma_1) < \gamma_0(1+\gamma_0+\gamma_2)$ is equivalent to the condition that $\gamma_1 < \gamma_0 + \frac{\gamma_1\gamma_2}{1+\gamma_0+\gamma_1+\gamma_2} < \gamma_0 + \gamma_2$. Using these inequalities to bound R_{CF}^* in (10) and noting that in this region $R_{\text{DF},1} < R_{\text{DF},2}$, it follows that $R_{\text{DF}}^* < R_{\text{CF}}^*$, which completes the proof of the second statement. \square

Corollary 1 provides an SNR threshold below which DF is more advantageous than CF, and above which CF is more advantageous than DF. This result will be used to develop a DF/CF switching scheme which will be shown to yield significant rate advantages over individual schemes under quasi-static Rayleigh fading channel conditions.

Using Lemma 4, we have the following result.

Theorem 3. *Consider the Gaussian channel in Figure 1 and suppose that the channel SNRs are finite and bounded away from zero, i.e., $0 < \gamma_0, \gamma_1, \gamma_2 < \infty$. If $\gamma_i, i = 0, 1, 2$, satisfy (19), then there exist power partitions, $\{\alpha_i\}_{i=0}^2$ and $\{\beta_j\}_{j=0}^1$, and normalized estimation noise, γ' , such that the generalized DF-CF using Gaussian signals provides a higher rate than the corresponding DF and CF schemes.*

Proof. To prove this theorem, we will use contradiction. First, we will assume that $R_{\text{G}}^* = R_{\text{CF}}^* = R_{\text{DF}}^*$, and then we will show that this R_{G}^* and the corresponding power partitions and γ' do not satisfy the KKT necessary optimality conditions in (8).

First, we recall the result of Remark 1, indicating that the generalized DF-CF scheme reduces to the CF scheme when $\alpha_2 = 1$ and $\beta_1 = 1$. Substituting from (19) in (11) yields

$$\gamma' = \frac{\gamma_0}{\gamma_1 - \gamma_0}. \quad (20)$$

Note that the SNR condition in Lemma 4 implies that $\gamma_1 > \gamma_0$, and hence that $\gamma' > 0$.

Next, we will examine the KKT system (8). We note that if $\alpha_2 = 1$ and $\beta_1 = 1$ are optimal, then for the inequalities in (8f) with $i = 4, 7$, we must have $\lambda_4 = \lambda_7 = 0$.

Substituting for γ' from (20) in (8c), the coefficient of λ_1 vanishes and using $\lambda_4 = 0$ yields $\lambda_3 \frac{\gamma_1}{(1+\gamma_0)^2} + \lambda_6 = 0$, which implies that $\lambda_3 = \lambda_6 = 0$. Using this result and the fact that $\gamma_1 > \gamma_0$ in (8e) yields $\lambda_1 \gamma_1 = \lambda_2 (1 + \gamma_0 + \gamma_2)$.

Now, invoking $\lambda_1 + \lambda_2 = 1$, yields $\lambda_1, \lambda_2 > 0$; both λ_1 and λ_2 are finite and bounded away from 0. Substituting for λ_3, λ_7 and γ' in (8d) yields $-\lambda_2 \sqrt{\frac{\alpha_0}{\beta_0}} \frac{\gamma_0 \sqrt{\gamma_0 \gamma_2}}{\gamma_1} = \lambda_8$. Since $\lambda_2 > 0$, for this equality to be satisfied, we must have $\lambda_8 = 0$, and

$$\sqrt{\frac{\alpha_0}{\beta_0}} = 0. \quad (21)$$

Substituting for λ_3, λ_6 and γ' in (8b) yields

$$\lambda_1 \gamma_1 = \lambda_2 \sqrt{\frac{\beta_0}{\alpha_0}} \frac{\gamma_0 \sqrt{\gamma_0 \gamma_2}}{\gamma_1} + \lambda_5. \quad (22)$$

Now, the left hand side is finite. However, using (21), and the fact that $\lambda_2 > 0$ and $\lambda_5 \geq 0$ implies that the right hand side is infinite. Hence, we conclude that the equality in (22) does not hold, which indicates that, when the SNRs satisfy the condition in Lemma 4, the power partitions $\alpha_2 = 1$ and $\beta_1 = 1$ do not satisfy the KKT conditions necessary for optimality. In other words, reducing the generalized DF-CF scheme to the CF scheme by setting $\alpha_2 = \beta_1 = 1$, is suboptimal; i.e., $R_G^* > R_{CF}^*$.

To complete the proof of the theorem, we note that at the SNRs satisfying

Lemma 4, $R_{\text{DF}}^* = R_{\text{CF}}^*$, which implies that, in this region, $R_{\text{G}}^* > R_{\text{DF}}^*$. \square

Theorem 3 shows that for SNRs satisfying the condition of Lemma 4, the generalized DF-CF using Gaussian signals yields strictly higher rates than the corresponding DF and CF schemes.

So far we have restricted our attention to strictly positive finite SNRs. In the following remarks we will consider two extreme cases, namely, $\gamma_2 \rightarrow \infty$ and $\gamma_0 \rightarrow 0$.

Remark 3. *When $\gamma_2 \rightarrow \infty$ and $0 < \gamma_0, \gamma_1 < \infty$, the generalized DF-CF scheme using Gaussian signals reduces to the corresponding CF scheme and achieves the capacity of the relay channel; i.e., $R_{\text{G}}^* = R_{\text{CF}}^* = C$.* \square

The proof of this remark uses the fact that the Gaussian distribution maximizes the cut-set bound [20] for the considered Gaussian channel and that this bound is achieved by the CF scheme using Gaussian signals when $\gamma_2 \rightarrow \infty$. A similar observation has been independently made in [21]. The fact that CF is a special case of the generalized CF-DF yields the result.

Remark 4. *When $\gamma_0 \rightarrow 0$, $0 < \gamma_1$ and $\gamma_2 < \infty$, the generalized DF-CF scheme reduces to the DF scheme, and achieves the capacity of the channel in Fig. 1. In particular, $R_{\text{G}}^* = R_{\text{DF}}^* = C$.* \square

To prove this remark, we invoke the fact that the cut-set bound is maximized by the Gaussian distribution and that Gaussian signals maximize the rate achieved by the DF scheme for the considered relay channel. Direct substitution for $\gamma_0 \rightarrow 0$ in the corresponding rate expressions shows that the DF scheme meets the cut-set bound and hence achieves capacity. The fact that DF is a special case of the generalized CF-DF scheme yields the result.

2.4.4 An Upper Bound on the Gain of the Generalized DF-CF

We now provide an upper bound on the gain of the generalized DF-CF using Gaussian signals over the corresponding DF and CF schemes. In particular, we prove the following theorem.

Theorem 4. *Using Gaussian signals, the maximum rate achieved by the generalized DF-CF scheme is at most 0.5 bpcu higher than that achieved by the individual DF and CF schemes.*

Proof. Detailed proof is provided in Appendix A.6. □

This theorem implies that, despite the inherent coding and decoding complexity of the generalized DF-CF, its rate advantage over DF and CF is relatively small. This indicates that the practical benefit of the analysis of this scheme is to enable appropriate selection of either DF or CF relaying depending on the given channel realization.

2.5 Numerical Results

In this section, we will confirm the analysis of Section 2.4 by numerical examples. We will also provide an example in which this analysis is exploited to adapt the relay operation mode to the channel gains.

The maximum DF rate, R_{DF}^* , is obtained using (9), by setting $\rho_0^* = 0$, or such that $R_{\text{DF},1} = R_{\text{DF},2}$, depending on the values of γ_i , $i = 0, 1, 2$. The maximum CF rate, R_{CF}^* , is evaluated using (10), and the maximum generalized DF-CF rate, R_G^* , is evaluated by using the KKT conditions corresponding to (6) to reduce the search

for γ^* and the optimal power partitions to two parameters, viz., $\alpha_0, \beta_0 \in [0, 1]$. The cut-set bound is calculated using the expression in Theorem 4 in [2].

In Fig. 2, the SNR of the source-destination link is set to be $\gamma_0 = 5$ dB and that of the relay-destination link is set to be $\gamma_2 = 5.5$ dB, which yields $\gamma_0(1 + \gamma_2) = 11.5$ dB.

From Fig. 2(a) it can be seen that, in agreement with Theorem 1, for $\gamma_1 > 11.5$ dB, the maximum achievable rate of the generalized DF-CF scheme, R_G^* , coincides with the maximum achievable rate of the DF scheme, R_{DF}^* . Similarly, in agreement with Theorem 2, for $\gamma_1 \leq 5$ dB, R_G^* coincides with the maximum achievable rate of the CF scheme, R_{CF}^* .

To investigate the performance of the generalized DF-CF scheme when $\gamma_0 < \gamma_1 \leq \gamma_0(1 + \gamma_2)$, in Figure 2(b), we plot a magnified version of the region spanned by $\gamma_1 \in [6.3, 6.6]$. As asserted by Theorem 3, it can be seen that, when $\gamma_1 = 6.496377$ dB, $R_{DF}^* = R_{CF}^* = 1.22486$ bpcu. At this SNR, the generalized DF-CF scheme yields a rate advantage $\Delta_G = R_G^* - R_{DF}^* = 2.62 \times 10^{-3}$ bpcu. This rate advantage is obtained by setting $\alpha_0 = 0.004$, $\alpha_1 = 0.762621$, $\alpha_2 = 1 - \alpha_1 - \alpha_0$, $\beta_0 = 0.179$, and $\beta_1 = 1 - \beta_0$. In agreement with Theorem 4, the gain offered by the generalized DF-CF scheme is strictly less than 0.5 bpcu. Similar to the numerical results reported for the generalization in [39], the rate advantage of the generalization considered herein is marginal. However, this does not preclude the possibility that the generalized DF-CF might yield higher rate gains for more general networks, or when the relay and destination noises are correlated; cf. Remark 2.

In Fig. 3 we verify Remarks 3 and 4. In Figure 3(a), we set $\gamma_0 = 1$ dB and $\gamma_1 = 5$ dB. In accordance with Remark 3, it can be seen that as γ_2 becomes sufficiently large, R_G^* coincides with R_{CF}^* and with the cut-set bound. Figure 3(b) is obtained by setting $\gamma_1 = 10$ dB and $\gamma_2 = 5$ dB. As stated in Remark 4, as γ_0 becomes sufficiently small, R_G^* coincides with R_{DF}^* and with the cut-set bound.

Finally, we provide a numerical example to expose the practical implications of

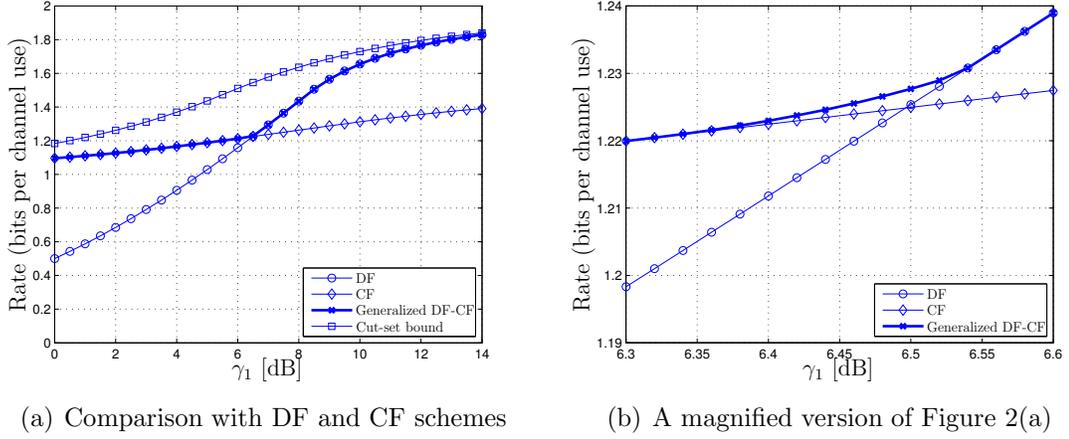
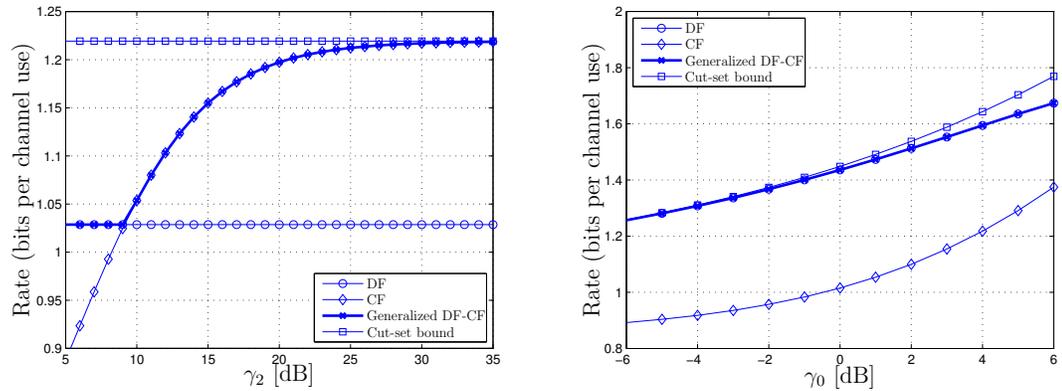


Figure 2: Maximum achievable rate of the generalized DF-CF scheme.

the analysis of the generalized DF-CF scheme. To do so, we consider a quasi-static frequency-flat Rayleigh fading relay channel in which the channel coefficients take on random values, but are held constant for the entire signalling duration. The coefficients are drawn from the standard zero mean Gaussian distribution with the variance adjusted to yield average SNRs of $\bar{\gamma}_0 = E\{\gamma_0\} = 5$ dB, $\bar{\gamma}_2 = E\{\gamma_2\} = 5.5$ dB and $\bar{\gamma}_1 = E\{\gamma_1\}$ ranging between 0 and 20 dB. We assume that the instantaneous channel gains, and hence $\{\gamma_i\}_{i=0}^2$, are available at source, relay and destination, and we compare the following three schemes: the standard DF and CF schemes and a DF/CF switching scheme. In the standard schemes the relay uses either DF or CF relaying regardless of the channel realization, whereas in the switching scheme, the relay uses Corollary 1 to select its operation mode depending on the current realization of $\{\gamma_i\}_{i=0}^2$. From Figure 4, it can be seen that enabling the relay to switch between DF and CF operating modes yields a significant gain over either scheme. For instance, at a rate of 1.1 bpcu, the gain of the DF/CF switching scheme over the standard DF and CF schemes is about 4 dB.



(a) Generalized DF-CF reduces to CF and (b) Generalized DF-CF reduces to DF and achieves capacity

Figure 3: Asymptotic optimality of the generalized DF-CF scheme

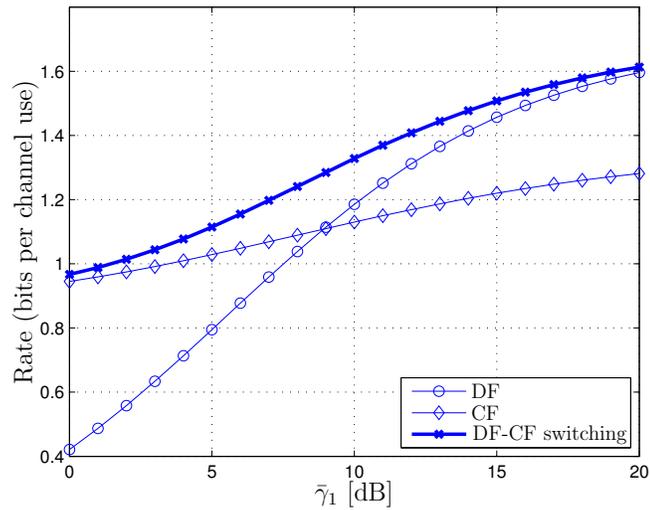


Figure 4: Maximum achievable rate in quasi-static Rayleigh fading channel.

Chapter 3

A New Decoding Procedure to Support the N -to-1 Mapping in Compress-and-Forward Relaying

In this chapter, a forward decoding procedure is developed for the compress-and-forward relaying scheme. This procedure uses a layered framework and is based on exploiting a feature of the N -to-1 mapping inherent in the underlying Wyner-Ziv binning. It is shown that exploiting this mapping enables the relaxation of the constraint on the rate of the relay codewords representing the bin indices. The work in this chapter was published, in part, in [14]. A more complete version is currently under review for potential publication in the IEEE Transactions on Information Theory, see [15].

3.1 Introduction

Compress-and-forward (CF) [2] is a classical relaying scheme for communicating over relay channels. In conventional CF, the source transmits a new codeword in each time block. The relay uses a pre-designed codebook to generate descriptions of its received signal. Using Wyner-Ziv binning [8], the codewords in the codebook of

this description are randomly partitioned into non-overlapping bins, a process that results in an N -to-1 mapping from the description codewords to the bin indices, where $N \geq 1$. In each block, the CF relay provides a description of its received signal and sends its bin index in the next block to facilitate decoding at the receiver. Decoding at the CF receiver comprises three steps [2]:

1. decoding the bin index transmitted from the relay;
2. using the received signal as side information [8] to decode the relay description codeword in the bin; and
3. recovering the transmitted codeword from the source with the facilitation of the relay description and its corresponding bin index.

The last two steps can also be performed jointly [41].

CF relaying was originally proposed for the three-node relay channel, but was later extended to channels with multiple relays [20, 42, 43]. CF is also known to be capacity achieving for various classes of relay channels [31, 34, 44–46].

Akin to CF is the noisy network coding (NNC) scheme provided in [9]. The philosophy that underlies this scheme resembles, to some extent, that of CF. However, there are three differences between the CF scheme and the NNC one.

- First, in contrast with random binning used in CF, in NNC the relay transmits a codeword that bears a 1-to-1 correspondence with the description codeword.
- Second, in NNC the source uses repetitive transmission, wherein one long message is encoded over a large number of blocks. This is in contrast with conventional CF, wherein a new short message is transmitted by the source in each time block.

- Finally, in NNC, the received signals in all time blocks are concatenated and decoded jointly; in CF the decoding is performed on a forward block-by-block basis.

Repetitive encoding of long messages over a large number of blocks incurs significant delay, which renders short messaging more desirable. Variations of the original NNC that use short message encoding (SNNC) were proposed in [10, 47–49]. Despite their differences, both NNC and SNNC decoding use the inherent 1-to-1 mapping between the description codewords at the relay and its transmitted codewords. This 1-to-1 mapping can be seen as a special case of the general Wyner-Ziv binning with equal rate of the Wyner-Ziv codes and the description codewords at the relay.

In [47], SNNC was studied for the standard three-node relay channel when forward and backward decoding are used. In [10] and [48], it was shown that SNNC yields the same rate region as NNC when either backward decoding or joint decoding with concatenated blocks is used in the multimessage network considered therein. It is worth noting that although short messages are encoded at the source, either backward decoding or joint decoding with concatenated blocks decoding still causes long delay since both decoding procedures require the reception of all the transmitted codewords from the source.

Forward decoding that uses the SNNC codebook structure was investigated for the multimessage network in [49]. This decoding is based on ordered partitions of all the nodes in the network, thereby resulting in a set of constraints on the achievable rate. Using a geometric approach, it was shown that, in the multimessage network considered in [49], there exist ordered partitions that yield a rate region that coincides with the one achieved by NNC.

In the standard three-node relay channel, conventional CF, NNC and SNNC achieve the same rate. However, in more general multimessage networks considered in [9, 10, 49], NNC and SNNC achieve rate regions larger than that achieved by

conventional CF.

In this chapter, we consider a multmessage network similar to the one considered in [9, 10, 49]. For this network, we develop a decoding procedure based on the conventional CF codebook structure. The proposed procedure uses multiple layers of the joint typicality sets, which will be defined in Section 3.2, for decoding. At each layer, a subset of the transmitted codewords and the codewords of the relay descriptions are uniquely decoded. The codebooks that contain these uniquely decoded codewords are used to construct the joint typicality set at the next decoding layer. Such sequential construction of the joint typicality set at each layer provides a hierarchical structure whereby the joint typicality set at one layer is a proper subset of that at the layer below it. This construction provides in effect a systematic method for obtaining the ordered partitions conceived in [49]. In addition to the layered framework, the proposed decoding procedure exploits the N -to-1 mapping that underlies the Wyner-Ziv binning. We show that this procedure is able to achieve the same rate region as that achieved by SNNC in the multmessage network. However, because of the N -to-1 mapping, this procedure enables the relaxation of the constraint on the rate of the bin indices. We will show that in the considered multmessage network, our decoding procedure allows the rate of the relay bin indices to be strictly lower than the rate of the relay description of its received signal and still achieves the same rate regions as NNC and SNNC. This result implies that, although in general, the N -to-1 mapping inherent in the Wyner-Ziv binning introduces ambiguity in conveying the relay description to the receivers, certain level of ambiguity does not cause additional rate loss when side information is available at the receivers.

Notation: In this chapter, a sequence of random variables of node index d_k is denoted as \mathbf{X}_k . A k -tuple of random variables is denoted as $X_{\mathcal{A}} \triangleq (X_k : d_k \in \mathcal{A})$. The sum rate of the codebooks of a set of nodes \mathcal{A} is denoted by $R_{\mathcal{A}} \triangleq \sum_{d_k \in \mathcal{A}} R_{d_k}$.

3.2 Preliminaries

We first provide definitions of typical sequence and jointly typical sequences according to [18].

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

1. For all $a \in \mathcal{X}$ with $p(a) > 0$, we have

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

2. For all $a \in \mathcal{X}$ with $p(a) = 0$, $N(a|\mathbf{x}) = 0$.

$N(a|\mathbf{x})$ is the number of occurrences of a in the sequence \mathbf{x} .

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

1. For all $(a, b) \in \mathcal{X} \times \mathcal{Y}$ with $p(a, b) > 0$, we have

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

2. For all $(a, b) \in \mathcal{X} \times \mathcal{Y}$ with $p(a, b) = 0$, $N(a, b|\mathbf{x}, \mathbf{y}) = 0$.

$N(a, b|\mathbf{x}, \mathbf{y})$ is the number of occurrences of (a, b) in the pair of sequences (\mathbf{x}, \mathbf{y}) .

The set of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$ such that (\mathbf{x}, \mathbf{y}) is strongly typical is called the strong joint typicality set, or referred to as joint typicality set in this thesis, and is denoted by $\mathcal{A}_\epsilon^{(n)}$.

Next, we provide the following three important lemmas according to [41].

Conditional Typicality Lemma: Let $(X, Y) \sim p(x, y)$. Suppose that $\mathbf{x} \in \mathcal{A}_{\epsilon'}^{(n)}$ and $\mathbf{Y} \sim p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$. Then, for every $\epsilon' < \epsilon$,

$$\lim_{n \rightarrow \infty} P\{\mathbf{x}, \mathbf{Y} \in \mathcal{A}_{\epsilon'}^{(n)}\} = 1.$$

Joint Typicality Lemma: Let $(X, Y, Z) \sim p(x, y, z)$ and $\epsilon' < \epsilon$. Then there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \rightarrow 0$ such that the following statements hold:

1. If $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ is a pair of arbitrary sequences and $\tilde{\mathbf{Z}} \sim \prod_{i=1}^n p_{Z|X}(z_i|\tilde{x}_i)$, then

$$P\{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{Z}}) \in \mathcal{A}_{\epsilon'}^{(n)}\} \leq 2^{I(Y;Z|X) - \delta(\epsilon)}.$$

2. If $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in \mathcal{A}_{\epsilon'}^{(n)}$ and $\tilde{\mathbf{Z}} \sim \prod_{i=1}^n p_{Z|X}(z_i|\tilde{x}_i)$, then

$$P\{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{Z}}) \in \mathcal{A}_{\epsilon'}^{(n)}\} \geq 2^{I(Y;Z|X) + \delta(\epsilon)}.$$

Covering Lemma: Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$ and $\epsilon' < \epsilon$. Let $(\mathbf{U}, \mathbf{X}) \sim p(\mathbf{u}, \mathbf{x})$ be a pair of random sequences with $\lim_{n \rightarrow \infty} P\{(\mathbf{U}, \mathbf{X}) \in \mathcal{A}_{\epsilon'}^{(n)}\} = 1$, and let $\hat{\mathbf{X}}(m), m \in \mathcal{A}$, where $|\mathcal{A}| \geq 2^{nR}$, be random sequences, conditionally independent of each other and of \mathbf{X} given \mathbf{U} , each distributed according to $\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$. Then there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \rightarrow 0$ such that

$$\lim_{n \rightarrow \infty} P\{(\mathbf{U}, \mathbf{X}, \hat{\mathbf{X}}(m)) \notin \mathcal{A}_{\epsilon'}^{(n)}, \text{ for all } m \in \mathcal{A}\} = 0,$$

if $R > I(X; \hat{X}|U) + \delta(\epsilon)$.

Now we begin the analysis by reviewing the achievable rate of SNNC [10] (with 1-to-1 mapping) and conventional CF (with N -to-1 mapping) [41] in the standard three-node relay channel [2] shown in Fig. 5. Our goal in this chapter is to gain insight into the relay transmission rate and its impact to the achievable rate.

For both SNNC and conventional CF, the source sends \mathbf{X}_1 in each block and the relay and the receiver receive \mathbf{Y}_2 and \mathbf{Y}_D , respectively. Upon receiving \mathbf{Y}_2 , the relay obtains a description, $\hat{\mathbf{Y}}_2$, of its received signal, which is mapped to \mathbf{X}_2 using Wyner-Ziv binning. The relay then sends \mathbf{X}_2 in the next block to the receiver to facilitate decoding. The difference between SNNC and conventional CF in the codebook structure is the relationship between $\hat{\mathbf{Y}}_2$ and \mathbf{X}_2 . In SNNC, each $\hat{\mathbf{y}}_2$ is mapped to a distinct \mathbf{x}_2 , whereas in conventional CF, potentially multiple $\hat{\mathbf{y}}_2$ are assigned to one bin index which is mapped to an \mathbf{x}_2 .

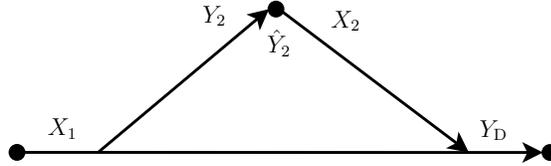


Figure 5: Standard three-node relay channel (SNNC or CF).

Let R_1, \check{R}_2 and \hat{R}_2 be the rates of the codewords that represent the source message, the relay bin index and the relay description of its received signal, respectively. Consider the probability mass functions (pmfs) of the form

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

In SNNC [10, 49], $\check{R}_2 = \hat{R}_2$, and the following rate can be achieved:

$$R_1 \leq I(X_1; \hat{Y}_2, Y_D|X_2), \quad (23a)$$

$$R_1 \leq I(X_1, X_2; Y_D) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_D). \quad (23b)$$

In conventional CF, $\check{R}_2 \leq \hat{R}_2$, and the rate satisfying (23a) and the following

constraints is achievable, cf. [41, Sect. 16.7]:

$$R_1 \leq I(X_1; Y_D | X_2) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_D) + \check{R}_2, \quad (24a)$$

$$\check{R}_2 \leq I(X_2; Y_D). \quad (24b)$$

Choosing $\check{R}_2^* = I(X_2; Y_D)$ in (24b) maximizes R_1 and yields the same rate expression as SNNC.

For the three-node relay channel, it was shown in [39] and [41, Remark 16.3] that the achievable rate is maximized when the two constraints in (23) are equal, which yields

$$I(X_2; Y_D) = I(\hat{Y}_2; Y_2 | X_2, Y_D). \quad (25)$$

It can be seen that when (25) is satisfied, CF is able to achieve the same rate as SNNC and yields $\check{R}_2^* = I(X_2; Y_D) = I(\hat{Y}_2; Y_2 | X_2, Y_D) \leq I(\hat{Y}_2; Y_2 | X_2) \leq \hat{R}_2$, which contrasts the condition in SNNC, wherein $\check{R}_2 = \hat{R}_2$.

In the next chapter, it will be shown that the fact that a modified CF decoding procedure achieves the same rate as SNNC but uses a lower rate on the bin indices holds in more general multmessage networks.

3.3 A Layered Forward Decoding Procedure for Multimessage Network

In a relay network with multiple receivers, when SNNC (with 1-to-1 mapping) is used, the rate of the codewords representing the bin indices at the relay is an intermediate parameter and can be eliminated from the expressions of the achievable rate region. In contrast, when conventional CF (with N -to-1 mapping) is used in the presence of multiple receivers, the rate of the codewords representing the bin indices at the

relay cannot be readily eliminated. The decoding of the relay transmitted codewords at each receiver imposes a constraint on the rate of the bin indices. This constraint appears in the achievable rate expressions and induces a rate loss in comparison with SNNC.

In this chapter, we analyze a CF forward decoding procedure in which two component strategies are combined: the N -to-1 mapping characteristic of conventional CF [2] and the sliding-window decoding of the relay messages characteristic of SNNC, see e.g., [49]. Combining these components results in a new decoding procedure that subsumes conventional CF and SNNC, and that therefore enables potentially higher rates to be achieved. Indeed, subsequent developments in Theorems 5 and 6 show that combining these components yields the same rate expressions as NNC and SNNC but with a more relaxed constraint on rate of the bin indices at the relays. The advantage of the relaxed constraint on rate of the bin indices will be illustrated in detail in the next chapter.

Consider the multimessage network shown in Fig. 6. The network contains a set of nodes $\mathcal{N} = \{1, 2, \dots, N\}$, each of which acts as a source, a receiver and a relay. As a source, node $d_k \in \mathcal{N}$ sends an independent common message through the transmission of \mathbf{X}_k to the set of its destinations $\mathcal{D}_{d_k} \subseteq \mathcal{N}$. The set of nodes that wish to send messages to d_k is denoted by $\mathcal{S}_{d_k} \subseteq \mathcal{N}$. As a relay to assist the transmission of other nodes, d_k provides a description, $\hat{\mathbf{Y}}_k$, of its received signal \mathbf{Y}_{d_k} , and facilitates the decoding at other nodes through the transmission of \mathbf{X}_k . Cooperation between nodes in this network is based on the facilitation provided by the encoding procedure at each node and the use of received signal as side information in the decoding procedure at each receiver.

In recovering the messages from the nodes in \mathcal{S}_{d_k} at d_k , the receiver can treat the information of the messages from the nodes in $\mathcal{N} \setminus \mathcal{S}_{d_k}$ as interference. Using this approach and SNNC (with 1-to-1 mapping), expressions for the achievable regions

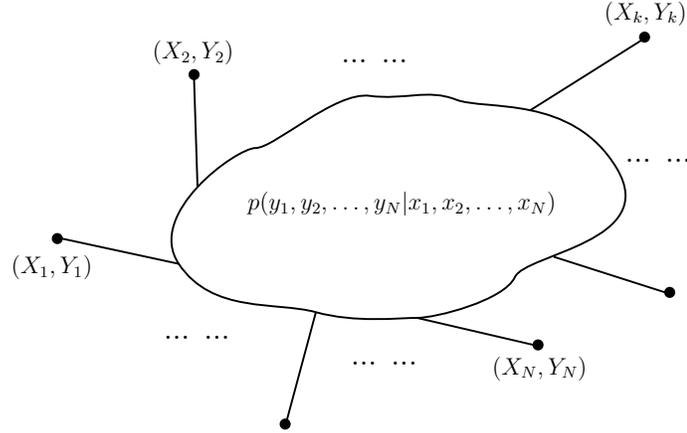


Figure 6: An N -node multmessage network.

are provided in [49]. In Theorem 5 herein, this result is extended to the codebook structure that bears the general N -to-1 mapping. In contrast, in Theorem 6, an achievable rate region is provided for the case when the receiver at d_k treats the interference from the nodes in $\mathcal{N} \setminus \mathcal{S}_{d_k}$ as noise instead of attempting to decode it.

Theorem 5. *Let $(\times_{k=1}^N \mathcal{X}_k, p(y^N | x^N), \times_{k=1}^N \mathcal{Y}_{d_k})$ be the general discrete memoryless multmessage network, a rate tuple (R_1, \dots, R_N) is achievable if:*

$$R_{\mathcal{S}} \leq \min_{d_k \in \mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}}} I(X_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}, \quad (26)$$

for all subsets $\mathcal{S} \subset \mathcal{N}$ and $\check{\mathcal{S}} \subseteq \mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$, such that $\mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}} \neq \emptyset$; the set $\mathcal{D}_{\mathcal{S}} \triangleq \cup_{d_l \in \mathcal{S}} \mathcal{D}_{d_l}$, $\check{\mathcal{S}}^c \triangleq \mathcal{S}^c \setminus \check{\mathcal{S}}$.

The implications of the inequalities in (26) can be inferred by considering a three node network. In such a network, there are two possibilities for \mathcal{S} : $\mathcal{S} = \{1\}$ and $\mathcal{S} = \{1, 2\}$. For the first possibility, there are two cases for $\check{\mathcal{S}}$: $\check{\mathcal{S}} = \emptyset$ and $\check{\mathcal{S}} = \{2\}$.

Using these possibilities in (26) yields the following bounds on the source rate, R_1 :

$$\text{for } \mathcal{S} = \{1\}, \mathcal{S}^c = \{2\}, \check{\mathcal{S}} = \emptyset, \check{\mathcal{S}}^c = \{2\}, R_1 \leq I(X_1; \hat{Y}_2, Y_D | X_2); \quad (27a)$$

$$\text{for } \mathcal{S} = \{1, 2\}, \mathcal{S}^c = \emptyset, \check{\mathcal{S}} = \check{\mathcal{S}}^c = \emptyset, \quad R_1 \leq I(X_1, X_2; Y_D) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_D); \quad (27b)$$

$$\text{for } \mathcal{S} = \{1\}, \mathcal{S}^c = \{2\}, \check{\mathcal{S}} = \{2\}, \check{\mathcal{S}}^c = \emptyset, R_1 \leq I(X_1; Y_D | X_2) - I(\hat{Y}_2; Y_2 | X_1, X_2, Y_D) + \check{R}_2. \quad (27c)$$

The first two bounds are identical to the standard CF bounds, cf. [41, Sect. 16.7] and (1a) and (1b). For the third bound we note that choosing $\check{R}_2 \geq I(\hat{Y}_2; Y_2 | X_2, Y_D)$ reduces (27c) to (27a), and choosing $\check{R}_2 \geq I(X_2; Y_D)$ reduces (27c) to (27b). This implies that for this scenario, (27c) is redundant and subsequently, the proposed scheme does not yield an advantage beyond conventional CF in the three node network.

The expression on the right hand side of (26) can be regarded as a generalization of the rate achieved by NNC and SNNC. In particular, as shown in Remark 6, when R_{ξ} satisfies the conditions in (29), the rate expression in Theorem 1 reduces to the one achieved by NNC and SNNC. As such, using the proposed signalling strategy in particular networks can in general yield a rate region that includes the rate region that can be achieved by the NNC and SNNC schemes; the additional advantage of the proposed scheme follows from exploiting the N -to-1 mapping as elucidated in detail in the proof of Theorem 5 and the examples in Sect. 4.1.1 and 4.1.2.

Proof. The decoding procedure uses strong joint typicality and features a layered framework. For layer 1 in this framework, the receiver considers the codebooks of all the nodes in the network and constructs the set containing the codewords that are jointly typical with the signal received in a particular block. For layer 2, the receiver considers only those codebooks that correspond to exactly one codeword in the joint typicality set in layer 1, and subsequently constructs the set containing the codewords that are jointly typical with the signal received in the following block. Hence, the codebooks considered at layer 2 are only a subset of those at layer 1. Subsequent layers

are constructed in a similar way until all the relay description codewords are uniquely decoded (further discussions on the number of decoding layers will be provided at the end of Sect. 3.3). Using the joint typicality sets at all the layers jointly, the receiver recovers the source messages. The details are provided in Appendix B.1. \square

Now, we make three remarks.

Remark 5. *When $\mathcal{A}_{\ell_m} = \mathcal{A}_{\ell_z}$, we have $\check{\mathcal{S}} = \mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z} = \emptyset$ and $\check{\mathcal{S}}^c = \mathcal{S}^c$. The result in Theorem 5 reduces to the following simplified form:*

$$R(\mathcal{S}) \leq I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}), \quad (28)$$

for $\mathcal{S} \cap \mathcal{S}_{d_k} \neq \emptyset$.

The simplified form of the achievable rate region coincides with that of NNC and SNNC.

Remark 6. *The rate region described in Theorem 5 reduces to the simplified form in (28) when $\check{R}_{\check{\mathcal{S}}}$ satisfies the following condition:*

$$\check{R}_{\check{\mathcal{S}}} \geq \min\{I(X_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) - R_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}, I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k})\}. \quad (29)$$

Proof. See details in Appendix B.2. \square

Remark 7. *Theorem 5 shows that the proposed scheme achieves the same rate region as NNC and SNNC but with reduced the rates of codewords representing the bin indices. This reduction will be shown to result in a rate advantage in the next chapter.*

Proof. To expose the reduction in the rate of the bin indices, it suffices to show that

the right hand side of (29) is upper bounded by $\hat{R}_{\check{\mathcal{S}}}$. We have

$$\begin{aligned}
& I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&= \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_{\check{\mathcal{S}}}|X_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&= \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i|X_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&\leq \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i|X_i) \\
&\leq \hat{R}_{\check{\mathcal{S}}},
\end{aligned}$$

where we have used $\check{\mathcal{S}} \subseteq \mathcal{S}^c$. Clearly, the lower bound on $\check{R}_{\check{\mathcal{S}}}$ is lower than $\hat{R}_{\check{\mathcal{S}}}$. \square

We now consider the case in which the transmitted codewords from the nodes in $\mathcal{N} \setminus \mathcal{S}_{d_k}$ can only provide information about the bin indices to facilitate decoding at d_k , and the receiver at d_k treats the information representing the message indices in $\mathbf{X}_{\mathcal{N} \setminus \mathcal{S}_{d_k}}$ as noise. Using this approach, we provide the following theorem.

Theorem 6. *Let $(\times_{k=1}^N \mathcal{X}_k, p(y^N|x^N), \times_{k=1}^N \mathcal{Y}_{d_k})$ be the general discrete memoryless multmessage network, a rate tuple (R_1, \dots, R_N) is achievable if*

$$R_{\mathcal{T}} \leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c}|X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}, \quad (30)$$

for all subsets $\mathcal{S} \subset \mathcal{N}$, $\check{\mathcal{S}} \subseteq \mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$ and $\mathcal{T} \subseteq \mathcal{S}_{d_k}$, where $\mathcal{T} \neq \emptyset$; the subset $\check{\mathcal{S}}^c = \mathcal{S}^c \setminus \check{\mathcal{S}}$ and $\mathcal{T}^c = \mathcal{S}_{d_k} \setminus \mathcal{T}$.

Similar to the case of Theorem 5, we note that the expression on the right hand side of (30) can be regarded as a generalization of the rate achieved by NNC. In particular, when $R_{\check{\mathcal{S}}}$ satisfies the criteria in (32), the right hand side of (30) reduces to (11), cf. Remark 9 and [9, Theorem 3]. Hence, in general, the signalling strategy that underlies Theorem 6 offers the potential of achieving rate regions that include

those achieved by the corresponding NNC scheme; the advantage of that signalling strategy follows from exploiting the N -to-1 mapping as elucidated in detail in the proof of Theorem 6 and the example in Sect. 4.1.2.

Proof. Using the same philosophy as that used in the proof of Theorem 5, the decoding procedure herein also features a layered framework. Details are provided in Appendix B.3. \square

Next, we make three remarks.

Remark 8. *When $\mathcal{A}_{\ell,s} = \mathcal{A}_{\ell,z}$, we have $\check{\mathcal{S}} = \mathcal{A}_{\ell,s} \setminus \mathcal{A}_{\ell,z} = \emptyset$ and $\check{\mathcal{S}}^c = \mathcal{S}^c$. The result in Theorem 6 reduces to the following simplified form:*

$$R_{\mathcal{T}} \leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}). \quad (31)$$

for $\mathcal{T} \neq \emptyset$.

\square

The simplified form of the achievable rate region coincides with Theorem 3 in [9].

Remark 9. *The rate region described in Theorem 6 reduces to the simplified form in (31) when $\check{R}_{\check{\mathcal{S}}}$ satisfies the following condition:*

$$\check{R}_{\check{\mathcal{S}}} \geq \min\{I(U_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\check{\mathcal{T}}^c}, U_{\check{\mathcal{S}}^c}), I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\check{\mathcal{T}}^c}, U_{\check{\mathcal{S}}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k})\}. \quad (32)$$

Proof. Detailed proof is available in Appendix B.4. \square

Remark 10. *Theorem 6 shows that the proposed scheme achieves the same rate region as NNC but with reduced bin indices rates. This reduction will be shown to result in a rate advantage in Section 4.1.2.*

Proof. To expose the reduction in the rate of the bin indices, it suffices to show that the right hand side of (32) is upper bounded by $\hat{R}_{\check{\mathcal{S}}}$. Towards that end, we write

$$\begin{aligned}
& I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&= \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_{\check{\mathcal{S}}}|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&= \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i|X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\mathcal{K}_i}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\
&\leq \sum_{i \in \check{\mathcal{S}}} I(\hat{Y}_i; Y_i|U_i) \\
&\leq \hat{R}_{\check{\mathcal{S}}},
\end{aligned}$$

where we have used $\check{\mathcal{S}} \subseteq \mathcal{S}^c$. Clearly, the lower bound on $\check{R}_{\check{\mathcal{S}}}$ is lower than $\hat{R}_{\check{\mathcal{S}}}$. \square

Since the 1-to-1 mapping satisfies the condition in (32), the achievable rate region of the simplified form in (31) can also be obtained by the use of short message encoding with 1-to-1 mapping. This observation extends the results provided in [10] and [49] for SNNC.

To obtain an upper bound on the maximum number of decoding layers, we recall that the decoding procedure described in the proofs of Theorem 5 and 6 progresses from one layer to the next depending on the outcome of a joint-typicality test. In particular, if at a given layer i , multiple description codewords of a particular node are found to be in the joint-typicality set, the decoding procedure excludes the codebooks corresponding to the multiple description codewords and progresses to the next layer, i.e., layer $i + 1$. The decoding procedure stops once it reaches a layer with no multiple description codewords of any node in the joint-typicality set. From this procedure, it can be seen that the number of candidate description codebooks examined in each layer decreases strictly from one layer to the next, which implies that the maximum number of layers, \mathcal{L} , is upper bounded by $N - 1$. Hence, the decoding delay in the

network can be upper bounded by $N - 1$ blocks. In block 1, each node uses a message index and a known bin index in encoding. From block 2, each nodes encodes according to the procedure provided in the proofs of the theorems. To end the transmission, each node continues to encode and transmit for $N - 2$ blocks using a known message index and the bin index of the description of its received signal in the previous block.

At last, we note that for the cooperative multmessage network shown in Fig. 6, the results of using N -to-1 mapping in Theorem 5 and 6 do not yield rate gain in comparison with their simplified forms which can also be achieved by SNNC (with 1-to-1 mapping). However, our goal is not to show the rate advantage of the N -to-1 mapping in this network, but rather to show its advantage in the network instances considered in the next chapter.

3.4 Conclusion

In this chapter, we provided a layered forward decoding procedure that enables exploiting the N -to-1 mapping that underlies CF relaying. This procedure relaxes the rate constraint on the bin indices, and is subsequently able to yield a rate advantage over CF-based schemes that use 1-to-1 mapping.

Chapter 4

Exploiting the N -to-1 Mapping in Compress-and-Forward Relaying

In this chapter, we show that exploiting this mapping enables the relaxation of the constraint on the rate of the relay codewords representing the bin indices. For the cooperative multmessage network, the proposed procedure achieves the same rate region as the short-message noisy network coding (SNNC) scheme. However, this procedure is more advantageous for other networks including the two networks presented herein. The first network is a relay chain one with two destinations, whereas the second network is a partially cooperative multmessage one with three destinations. In both networks, side information is available to a subset of the decoding nodes, but not to the rest of the nodes, and in both cases, the network benefits from the relaxation of the rate of the CF bin indices. This relaxation results in rate regions larger than those achieved by conventional CF and SNNC. The work in this chapter is currently under review for the potential publication in the IEEE Transactions on Information Theory, see [15].

4.1 Introduction

Because of the N -to-1 mapping, the procedure provided in Chapter 3 enables the relaxation of the constraint on the rate of the bin indices, which will be shown to be beneficial in certain cases herein. In particular, we will provide hereinafter instances in which the N -to-1 mapping yields a rate advantage. A key feature of those instances is that side information from the source is available only to subset of the nodes. This can be caused, for instance, when the link between the source and the receiving node is broken. The lack of side information may induce rate loss when the relay description is used to recover the CF bin indices. In such situations it is more beneficial for the receiving node to recover the CF bin indices directly without side information, which enables the node to take advantage of the relaxed constraint on the rate of the CF bin indices. To investigate the gain of the relaxation of the rate constraint on the CF bin indices, we consider two networks.

In the first network, a source broadcasts a common message to two destinations, which are assisted by a chain of two cascaded relays, cf. [29]. In this chain, the first relay receives signals from the source, uses the CF strategy and forwards the bin indices of its description codewords to the second relay and the two destinations. The second relay only receives signals from the first relay and uses the decode-and-forward (DF) strategy to assist the destinations in recovering the bin indices transmitted by the first relay. Since the DF relay does not have a direct link from the source, it does not have access to side information. To assist decoding at the receiver, the DF relay recovers the bin index transmitted by the CF relay without side information. This implies that a lower transmission rate of the CF relay is beneficial to the DF relay and hence a rate advantage can be observed.

In the second network, a source S sends common messages to two receivers D_1 and D_2 with the assistance of a relay R . The relay R cooperates with S in the transmission

to D_1 and D_2 , and also sends independent messages to its own destination D_3 . Unlike D_1 and D_2 , destination D_3 does not have direct link from S and hence the link between R and D_3 is a standard point-to-point one. Without the signal from S , the cooperation between S and R is not available to D_3 . Therefore, the network is only partially cooperative. Since the direct links from S to D_1 and D_2 provide side information to D_1 and D_2 , the results in Chapter 3 imply that using side information, N -to-1 mapping and 1-to-1 can achieve the same rate for D_1 and D_2 . In fact, the relay transmission rate (the rate of the bin indices) can be strictly lower than the rate of the relay description without inducing additional rate loss. However, the situation for D_3 is different. Since D_3 only has a direct link from R , it does not have access to side information. In this case the partial signal from R containing the relay bin index can be seen as either noise or interference at D_3 . We will show that a lower relay transmission rate (lower rate of the relay bin indices) by using N -to-1 mapping is able to yield rate gains.

In the analysis of the different relay strategies in this chapter, we focus on the comparison between SNNC and CF with the decoding procedure provided in Chapter 3. We note that this does not necessarily restrict the advantage of using the new decoding procedure over other strategies. However, detailed analysis in comparison with those strategies is out of the scope of this thesis.

We now analyze the achievable rates of these two networks.

4.1.1 The Achievable Rate of A Broadcast Relay Chain Network

We now describe the two-destination relay network shown in Fig. 7. In this network, a source S wishes to send a common message to two destinations D_1 and D_2 through the transmission of \mathbf{X}_1 with the assistance of a CF relay R_1 and a DF relay

R_2 in a chain. The received signals at relay R_1 and R_2 are denoted by $\mathbf{Y}_i, i = 2, 3$, respectively, and the received signal \mathbf{Y}_3 is independent of \mathbf{X}_1 . The received signals at D_1 and D_2 are denoted by $\mathbf{Y}_{D_i}, i = 1$ and 2 , respectively.

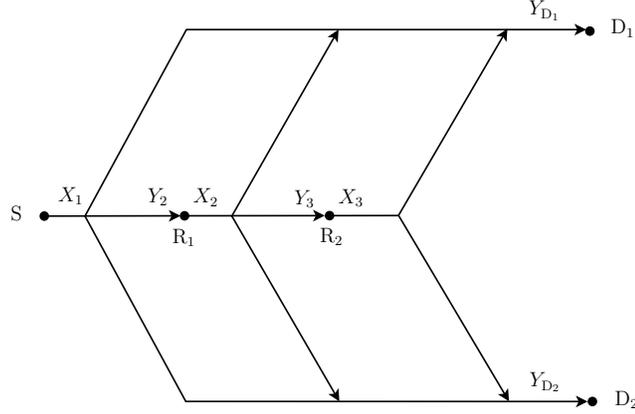


Figure 7: A two-destination broadcast relay chain network.

Without R_2 , the network reduces to a broadcast relay channel, cf. [50]. In that case, using N -to-1 mapping yields the same achievable rate R_1 as the. However, since \mathbf{Y}_3 is independent of \mathbf{X}_1 , correct decoding at R_2 imposes a constraint on the rate of the CF bin indices.

We evaluate three relaying strategies in the analysis of the achievable rate of this network. In all the considered strategies, node S uses the standard CF codebook structure. The relay R_2 uses the standard DF codebook structure and procedure. The difference between these three strategies lies in the way that the relay R_1 operates. In particular, in

- Strategy 1, the decoding procedure combines DF decoding and the decoding procedure of Theorem 5;
- Strategy 2, the decoding procedure combines DF decoding and the decoding procedure of SNNC; and in

- Strategy 3, the decoding procedure combines DF decoding and the decoding procedure of conventional CF.

The detailed procedures are provided in Section 4.1.1.1.

Next, we provide the achievable rate expressions for these strategies for the discrete memoryless case and the Gaussian case.

4.1.1.1 The discrete memoryless case

The achievable rate corresponding to the above relaying strategies for the discrete memoryless case are provided in the following corollary.

Corollary 2. *For the discrete memoryless network in Fig. 7, $(\mathcal{X}_1, p(y_2|x_1)p(y_3|x_2)p(y_{D_1}, y_{D_2}|x_1, x_2, x_3), \mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_{D_1} \times \mathcal{Y}_{D_2})$, the rate R_1 is achievable, where*

- for Strategy 1,

$$R_1 \leq \sup_{i=1,2} \min \{ I(X_1; \hat{Y}_2, Y_{D_i} | X_2, X_3), I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) + \min \{ I(X_2; Y_3 | X_3), I(X_2, X_3; Y_{D_i}) \} \}; \quad (33)$$

- for Strategy 2,

$$R_1 \leq \sup_{i=1,2} \min \{ I(X_1; \hat{Y}_2, Y_{D_i} | X_2, X_3), I(X_1, X_2, X_3; Y_{D_i}) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}), \quad (34)$$

subject to

$$I(\hat{Y}_2; Y_2 | X_2, X_3) \leq I(X_2; Y_3 | X_3);$$

- for Strategy 3,

$$\begin{aligned}
R_1 \leq \sup \min_{i=1,2} \{ & \min\{I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3), \\
& I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) \\
& + \min\{I(X_2; Y_3|X_3), \min_{i=1,2} I(X_2, X_3; Y_{D_i})\}\}, \quad (35)
\end{aligned}$$

where the supreme is taken over the pmfs of the form

$$\begin{aligned}
p(x_1, x_2, x_3, y_2, y_3, y_{D_1}, y_{D_2}, \hat{y}_2) = & p(x_1)p(x_2|x_3)p(x_3) \\
& p(y_2|x_1)p(y_3|x_2)p(y_{D_1}, y_{D_2}|x_1, x_2, x_3)p(\hat{y}_2|x_2, x_3, y_2).
\end{aligned}$$

□

Proof. See details in Appendix C.1. □

We note that Strategies 2 and 3 can be regarded as special cases of Strategy 1. Hence, Strategy 1 offers the potential of yielding a higher achievable rate than Strategies 2 and 3. We will next show that this is actually the case when the network is Gaussian.

4.1.1.2 The Gaussian case

Now, we compare the considered strategies for the Gaussian network depicted in Fig. 8. The network shown in this figure, is composed of scalar channel coefficients, independent additive Gaussian noises and Gaussian codebooks with average power constraints. As shown in Fig. 8, the transmitted signals from nodes S, R₁ and R₂ are denoted by $X_i \sim \mathcal{N}(0, P_i)$, where P_i is the average transmit power, $i = 1, 2, 3$, respectively. The gain of the S-to-R₁ and S-to-D_{*i*} links are denoted by a_{SR} and a_{SD_i} , $i = 1, 2$. The gain of the R₁-to-R₂ and R₁-to-D_{*i*} links are denoted by a_{RR} and

a_{RD_i} , $i = 1, 2$, respectively. The gain of the R₂-to-D_{*i*} link is denoted by a_{RS_i} , $i = 1, 2$. The independent additive noises on the S-to-R₁, and R₁-to-R₂ links are denoted by Z_2 and Z_3 , respectively, and that at the receiver D_{*i*} is denoted by Z_{D_i} , $i = 1, 2$. All noises are Gaussian distributed with zero mean and unit variance. Using this notation, the received signals at R₁, R₂ and D_{*i*}, $i = 1, 2$, can be expressed as

$$\begin{aligned}
 Y_2 &= a_{SR}X_1 + Z_2, \\
 Y_3 &= a_{RR}X_2 + Z_3, \\
 Y_{D_1} &= a_{SD_1}X_1 + a_{RD_1}X_2 + a_{RS_1}X_3 + Z_{D_1}, \\
 Y_{D_2} &= a_{SD_2}X_1 + a_{RD_2}X_2 + a_{RS_2}X_3 + Z_{D_2}.
 \end{aligned} \tag{36}$$

Denoting the description of the received signal at the R₁ by $\hat{Y}_2 = Y_2 + N'$, where $N' \sim \mathcal{N}(0, N')$ [21], we now define following SNRs:

$$\begin{aligned}
 \gamma_{SR} &= a_{SR}^2 P_1, & \gamma_{RR} &= a_{RR}^2 P_2, \\
 \gamma_{SD_i} &= a_{SD_i}^2 P_1, & \gamma_{RD_i} &= a_{RD_i}^2 P_2 \\
 \gamma_{RS_i} &= a_{RS_i}^2 P_3, & \gamma' &= N', & i &= 1, 2.
 \end{aligned}$$

Let $\mathcal{C}(x) = \frac{1}{2} \log_2(1 + x)$ and let ρ be the correlation coefficient between X_2 and

X_3 , i.e., $\rho = \frac{\mathbb{E}(X_2 X_3)}{\sqrt{P_2 P_3}}$. Using the technique in [13], we have

$$\begin{aligned}
I(X_1; \hat{Y}_2, Y_{D_i} | X_2, X_3) &= \mathcal{C}\left(\frac{\gamma_{SR}}{1 + \gamma'} + \gamma_{SD_i}\right), \\
I(X_1; Y_{D_i} | X_2, X_3) &= \mathcal{C}(\gamma_{SD_i}), \\
I(X_2, X_3; Y_{D_i}) &= \mathcal{C}\left(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1 + \gamma_{SD_i}}\right), \\
I(X_1, X_2, X_3; Y_{D_i}) &= \mathcal{C}(\gamma_{SD_i} + \gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}), \\
I(X_2; Y_3 | X_3) &= \mathcal{C}((1 - \rho^2)\gamma_{RR}), \\
I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) &= \mathcal{C}(1/\gamma'), \\
I(\hat{Y}_2; Y_2 | X_2, X_3) &= \mathcal{C}\left(\frac{1 + \gamma_{SR}}{\gamma'}\right).
\end{aligned}$$

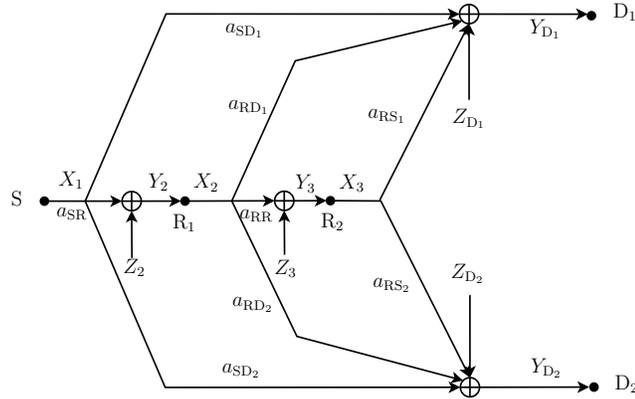


Figure 8: A two-destination Gaussian broadcast relay network with a relay chain, where Z' is an independent additive noise and Z_2, Z_3, Z_{D_1} and $Z_{D_2} \sim \mathcal{N}(0, 1)$.

Using these results in Corollary 2, we have the following proposition.

Proposition 2. For the network shown in Fig. 8, the rate R_1 is achievable, where

- for Strategy 1:

$$\begin{aligned}
R_1 \leq \max_{\rho, \gamma'} \min_{i=1,2} \min \left\{ \mathcal{C}\left(\frac{\gamma_{SR}}{1 + \gamma'} + \gamma_{SD_i}\right), \mathcal{C}(\gamma_{SD_i}) - \mathcal{C}\left(\frac{1}{\gamma'}\right) \right. \\
\left. + \min \left\{ \mathcal{C}\left(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1 + \gamma_{SD_i}}\right), \mathcal{C}((1 - \rho^2)\gamma_{RR}) \right\} \right\};
\end{aligned}$$

- for Strategy 2:

$$R_1 \leq \max_{\rho, \gamma'} \min_{i=1,2} \min \left\{ \mathcal{C} \left(\frac{\gamma_{SR}}{1 + \gamma'} + \gamma_{SD_i} \right), \right. \\ \left. \mathcal{C}(\gamma_{SD_i} + \gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}) - \mathcal{C}(1/\gamma') \right\},$$

subject to

$$\mathcal{C} \left(\frac{1 + \gamma_{SR}}{\gamma'} \right) \leq \mathcal{C}((1 - \rho^2)\gamma_{RR});$$

- for Strategy 3:

$$R_1 \leq \max_{\rho, \gamma'} \min_{i=1,2} \left\{ \min \left\{ \mathcal{C} \left(\frac{\gamma_{SR}}{1 + \gamma'} + \gamma_{SD_i} \right), \mathcal{C}(\gamma_{SD_i}) - \mathcal{C}(1/\gamma') \right. \right. \\ \left. \left. + \min \left\{ \mathcal{C}((1 - \rho^2)\gamma_{RR}), \min_{i=1,2} \left(\frac{\gamma_{RD_i} + \gamma_{RS_i} + 2\rho\sqrt{\gamma_{RD_i}\gamma_{RS_i}}}{1 + \gamma_{SD_i}} \right) \right\} \right\} \right\},$$

where $\rho \in [-1, 1]$, $\gamma' \geq 0$.

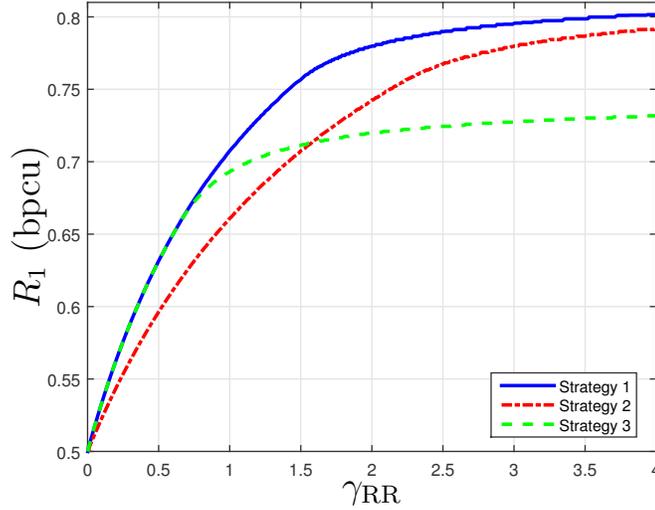


Figure 9: Achievable rates of a two-destination Gaussian broadcast relay chain network in Fig. 8, $\gamma_{SR} = 2$, $\gamma_{SD_1} = 1$, $\gamma_{SD_2} = 2$, $\gamma_{RD_1} = 2$, $\gamma_{RD_2} = 1$, $\gamma_{RS_1} = 2$ and $\gamma_{RS_2} = 1$.

To illustrate the advantage of Strategy 1, in Fig. 9 we plot the achievable rates

provided in Proposition 2 for the three strategies when $\gamma_{SR} = 2$, $\gamma_{SD_1} = 1$, $\gamma_{SD_2} = 2$, $\gamma_{RD_1} = 2$, $\gamma_{RD_2} = 1$, $\gamma_{RS_1} = 2$, $\gamma_{RS_2} = 1$ and $0 \leq \gamma_{RR} \leq 4$.

From Fig. 9 it can be seen that when $\gamma_{RR} \leq 0.7$, Strategies 1 and 3 achieve the same rate, which is higher than that achieved by Strategy 2. For $\gamma_{RR} > 0.7$, Strategy 1 achieves a higher rate than both Strategies 2 and 3. Fig. 9 also shows that for $\gamma_{RR} \leq 1.6$, Strategy 3 achieves a higher rate than Strategy 2, and for $\gamma_{RR} > 1.6$, Strategy 2 achieves a higher rate than Strategy 3. Note that if γ_{RR} is sufficiently large, the constraint on γ' in the achievable rate of Strategy 2 becomes inactive. In that case, Strategies 1 and 2 yield the same rate.

4.1.2 Achievable Rate Region of A Partially Cooperative Multimessage Network

In this section, we consider an example that belongs to a class of multimessage network. In this class, each source node wishes to send an independent message to its destinations with the assistance of relay nodes. Each relay node has its own independent message and wishes to send it to its own destinations through the direct link without the assistance of other nodes. The destinations of the relay nodes do not have direct links from other nodes. The set of destinations of the source nodes and the set of destinations of the relay nodes are disjoint. Each destination recovers its intended messages without collaboration.

Fig. 10 shows an example of such network. In this example, source S wishes to send a common message to two receivers D_1 and D_2 , with the assistance of node R. In addition, node R has its own independent message and wishes to send it to a third destination, D_3 . Since the only link to D_3 is the R-to- D_3 one, node R sends its message to D_3 without being assisted by S. Hence, the network is only a partially cooperative one.

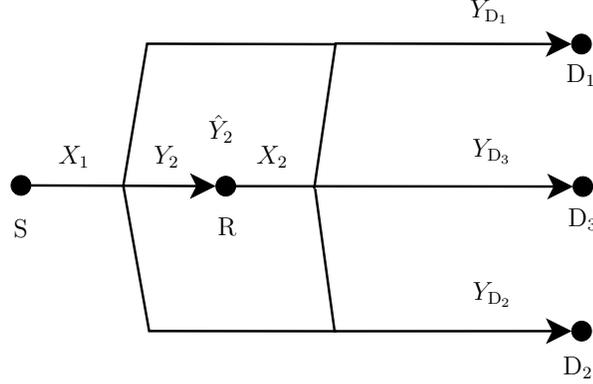


Figure 10: A two-message three-receiver partially cooperative network.

For the network in Fig. 10, we consider a relaying scheme in which node R facilitates decoding at D_1 and D_2 by transmitting the bin index of a description of its received signal. Let \mathbf{X}_1 and \mathbf{X}_2 be the codewords sent by S and R, respectively. Let \mathbf{Y}_2 and \mathbf{Y}_{D_i} be the received signal at R and $D_i, i = 1, 2, 3$, respectively, and let $\hat{\mathbf{Y}}_2$ be the codewords corresponding to the description of R of its received signal. We use R_1, R_2, \check{R}_2 and \hat{R}_2 to denote the rate of \mathcal{X}_1 , the rate of the independent message sent from R to D_3 , the rate of the relay bin indices and the rate of the description codebook at the relay, respectively.

Without D_3 , the network reduces to a broadcast relay channel, cf. [50]. In that case, Theorems 5 and 6 imply that using either the N -to-1 or the 1-to-1 mapping yields the same achievable rate, R_1 . However, the presence of D_3 and the fact that its received signal, \mathbf{Y}_{D_3} , does not contain information about \mathbf{X}_1 implies that these rates are not necessarily identical. To explore this possibility, we consider the following decoding strategies:

1. Use SNNC codebook structure at S and R. Use SNNC decoding procedure (Theorem 5 with $\check{R}_2 = \hat{R}_2$) at $D_i, i = 1, 2, 3$. The corresponding achievable rate region is denoted by \mathcal{R}_1 .

2. Use SNNC codebook structure at S and R. Only decodes desired codewords at $D_i, i = 1, 2, 3$, and treat the undesired signal as noise (Theorem 6 with $\check{R}_2 = \hat{R}_2$). The corresponding achievable rate region is denoted by \mathcal{R}_2 .
3. Use CF codebook structure at S and R. Use the decoding procedure in Theorem 5 at D_1 and D_2 , and directly recover the intended message from R at D_3 . The corresponding achievable rate region is denoted by \mathcal{R}_3 .
4. Use CF codebook structure at S and R. Use the decoding procedure in Theorem 6 at D_1 and D_2 , and directly recover the intended message from R at D_3 . The corresponding achievable rate region is denoted by \mathcal{R}_4 .

Next, we provide the achievable rate expressions for these strategies for the discrete memoryless case and the Gaussian case.

4.1.2.1 The discrete memoryless case

In the following corollary we provide expressions for the rate regions that can be achieved by each of the above strategies in the discrete memoryless case:

Corollary 3. *For the discrete memoryless network in Fig. 10,*

$$(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_{D_1}, y_{D_2} | x_1, x_2) p(y_{D_3} | x_2), \mathcal{Y}_2 \times \mathcal{Y}_{D_1} \times \mathcal{Y}_{D_2} \times \mathcal{Y}_{D_3}),$$

consider fixed pmf of the form:

- for Strategy 1 and 3,

$$p(x_1, x_2, \hat{y}_2, y_2, y_{D_1}, y_{D_2}, y_{D_3}) = p(x_1) p(x_2) p(\hat{y}_2 | x_2, y_2) p(y_2, y_{D_1}, y_{D_2} | x_1, x_2) p(y_{D_3} | x_2).$$

- for Strategy 2 and 4,

$$\begin{aligned} p(x_1, x_2, u, \hat{y}_2, y_2, y_{D_3}, y_{D_3}) \\ = p(x_1)p(x_2|u)p(u)p(\hat{y}_2|u, y_2)p(y_2, y_{D_1}, y_{D_2}|x_1, x_2)p(y_{D_3}|x_2). \end{aligned}$$

Using Strategy 1, the rate pair (R_1, R_2) is achievable, where

$$R_1 + R_2 \leq \min_{i=1,2} \{I(X_1, X_2; Y_{D_i}) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{D_i})\}, \quad (40a)$$

$$R_1 \leq \min_{i=1,2} I(X_1; \hat{Y}_2, Y_{D_i}|X_2), \quad (40b)$$

$$R_1 + R_2 \leq I(X_2; Y_{D_3}) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{D_3}). \quad (40c)$$

Using Strategy 2, the rate pair (R_1, R_2) is achievable, where

$$R_1 \leq \min_{i=1,2} \{I(X_1, U_2; Y_{D_i}) - I(\hat{Y}_2; Y_2|X_1, U_2, Y_{D_i})\}, \quad (41a)$$

$$R_1 \leq \min_{i=1,2} I(X_1; \hat{Y}_2, Y_{D_i}|U_2), \quad (41b)$$

$$R_2 \leq \min\{I(X_2; Y_{D_3}|U), I(X_2; Y_{D_3}) - I(\hat{Y}_2; Y_2|U_2, Y_{D_3})\}. \quad (41c)$$

Using Strategy 3, the rate pair (R_1, R_2) is achievable, if the sum rate and R_1 satisfy (40a) and (40b), respectively, and the following constraints on R_1 and R_2 are satisfied:

$$R_1 \leq I(X_1; Y_{D_i}|X_2) - I(\hat{Y}_2; Y_2|X_1, X_2, Y_{D_i}) + \check{R}_2, \quad (42a)$$

$$R_2 \leq I(X_2; Y_{D_3}) - \check{R}_2. \quad (42b)$$

Using Strategy 4, the rate pair (R_1, R_2) is achievable, if R_1 satisfies (41a)

and (41b), and the following constraints on R_1 and R_2 are satisfied:

$$R_1 \leq I(X_1; Y_{D_i} | U_2) - I(\hat{Y}_2; Y_2 | X_1, U_2, Y_{D_i}) + \check{R}_2, \quad (43a)$$

$$R_2 \leq \min\{I(X_2; Y_{D_3} | U_2), I(X_2; Y_{D_3}) - \check{R}_2\}. \quad (43b)$$

□

Next, we will particularize the network in Fig. 10 to the Gaussian case.

4.1.2.2 The Gaussian case

Now we consider the case that each link in Fig. 10 is an additive white Gaussian channel with i.i.d. zero mean unit variance Gaussian noises Z_2 at R, and Z_{D_i} and at D_i , $i = 1, 2, 3$, respectively. Nodes S and R are assumed to use Gaussian codebooks with average transmit power constraints. This case is shown in Fig. 11. For constructing the codebook of node R in Strategies 2 and 4, we use $\alpha_0 \in [0, 1]$ to represent the fraction of power that R allocates to transmit the bin index and $\alpha_1 = 1 - \alpha_0$ to represent the fraction of power that R allocates to transmit its own message index. The SNRs of the S-to-R, S-to- D_i and R-to- D_i links are denoted by γ_{SR} , γ_{SD_i} and γ_{RD_i} , $i = 1, 2, 3$, respectively. The variance of the additional noise in the relay description of its received signal is denoted by γ' [21]. Using these notations and a technique similar to the one in [13], Corollary 3 can be readily used to obtain expressions for achievable rates on this network. These expressions are recorded in the following proposition.

Proposition 3. *For the Gaussian network shown in Fig. 11, the rate pair (R_1, R_2) is achievable, where for*

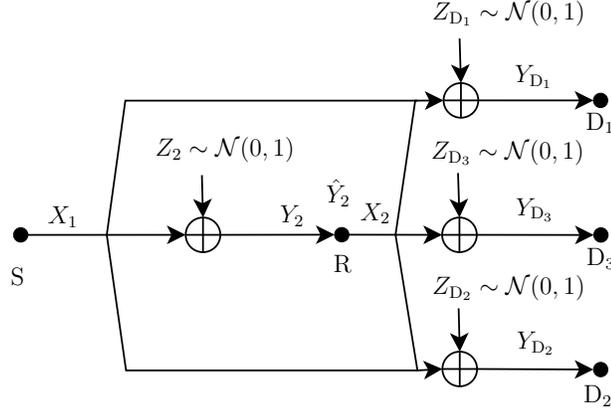


Figure 11: Gaussian particularization of the network in Fig. 10.

- *Strategy 1:*

$$R_1 + R_2 \leq \min_{i=1,2} \mathcal{C}(\gamma_{SD_i} + \gamma_{RD_i}) - \mathcal{C}(1/\gamma'), \quad (44a)$$

$$R_1 \leq \min_{i=1,2} \mathcal{C}\left(\frac{\gamma_{SR}}{1 + \gamma'} + \gamma_{SD_i}\right), \quad (44b)$$

$$R_1 + R_2 \leq \mathcal{C}(\gamma_{RD_3}) - \mathcal{C}(1/\gamma'); \quad (44c)$$

- *Strategy 2:*

$$R_1 \leq \min_{i=1,2} \mathcal{C}\left(\frac{\gamma_{SD_i} + \alpha_0 \gamma_{RD_i}}{1 + (1 - \alpha_0) \gamma_{RD_i}}\right) - \mathcal{C}(1/\gamma'), \quad (45a)$$

$$R_1 \leq \mathcal{C}\left(\frac{\gamma_{SR}}{1 + \gamma'} + \frac{\gamma_{SD_i}}{1 + (1 - \alpha_0) \gamma_{RD_i}}\right), \quad (45b)$$

$$R_2 \leq \min\left\{\mathcal{C}((1 - \alpha_0) \gamma_{RD_3}), \mathcal{C}(\gamma_{RD_3}) - \mathcal{C}\left(\frac{1 + \gamma_{SR}}{\gamma'}\right)\right\}; \quad (45c)$$

- *Strategy 3:*

$$R_1, R_2 \text{ satisfy (44a) and (44b),} \quad (46a)$$

$$R_1 \leq \min_{i=1,2} \mathcal{C}(\gamma_{SD_i}) - \mathcal{C}(1/\gamma') + \check{R}_2, \quad (46b)$$

$$R_2 \leq \mathcal{C}(\gamma_{RD_3}) - \check{R}_2, \quad (46c)$$

$$\check{R}_2 \leq \mathcal{C}\left(\frac{1 + \gamma_{SR}}{\gamma'}\right); \quad (46d)$$

- *Strategy 4:*

$$R_1, R_2 \text{ satisfy (45a), (45b) and (46d),} \quad (47a)$$

$$R_1 \leq \min_{i=1,2} \mathcal{C}\left(\frac{\gamma_{SD_i}}{1 + (1 - \alpha_0)\gamma_{RD_i}}\right) - \mathcal{C}(1/\gamma') + \check{R}_2, \quad (47b)$$

$$R_2 \leq \min\{\mathcal{C}((1 - \alpha_0)\gamma_{RD_3}), \mathcal{C}(\gamma_{RD_3}) - \check{R}_2\}; \quad (47c)$$

where $\alpha_0 \in [0, 1]$.

We now compare the rate expressions of Strategies 1 and 3. From Remark 6, constraint (46b) on R_1 in Strategy 3 coincides with (44b) in Strategy 1 when $\check{R}_2 = \max_{i=1,2} \mathcal{C}\left(\frac{1}{\gamma'} + \frac{\gamma_{SR}}{\gamma'(1+\gamma_{SD_i})}\right)$. Without loss of generality, assume $\gamma_{SD_1} \leq \gamma_{SD_2}$. Then let $\check{R}_2 = \mathcal{C}\left(\frac{1}{\gamma'} + \frac{\gamma_{SR}}{\gamma'(1+\gamma_{SD_1})}\right)$ and the constraint on R_1 from the decoding at D_1 is tighter than that at D_2 in both strategies according to (44b). In this case, from (46b) and (46c), the constraint on the sum rate from the decoding at D_1 and D_3 is tighter than that at D_2 and D_3 in Strategy 3. It can be seen that the sum rate constraint from (46b) and (46c) in Strategy 3, $\mathcal{C}(\gamma_{SD_1}) + \mathcal{C}(\gamma_{RD_3}) - \mathcal{C}(1/\gamma')$, is more relaxed than that in (44c) in Strategy 1. Note that (44a) is a common sum rate constraint in both strategies. Hence the sum rate constraint in Strategy 3 is more relaxed than that

in Strategy 1 in general. This implies that in general, when (46c) is satisfied, the constraint on R_2 in Strategy 3 is more relaxed than that in Strategy 1 for the same constraint on R_1 between the two strategies, which yields $\mathcal{R}_1 \subseteq \mathcal{R}_3$.

To compare the rate expressions of Strategies 2 and 4, it can be shown that from Remark 9, constraint (47b) reduces to (45a) when setting $\tilde{R}_2 = \max_{i=1,2} \mathcal{C}\left(\frac{1}{\gamma} + \frac{\gamma_{\text{SR}}(1+(1-\alpha_0)\gamma_{\text{RD}_i})}{\gamma'(1+(1-\alpha_0)\gamma_{\text{RD}_i} + \gamma_{\text{SD}_i})}\right) \leq \mathcal{C}\left(\frac{1}{\gamma} + \frac{\gamma_{\text{SR}}}{\gamma'}\right)$. Using this in (47c) implies that Strategy 4 yields a more relaxed constraint on R_2 in comparison with the constraint in (45c) for Strategy 2. Hence, we have $\mathcal{R}_2 \subseteq \mathcal{R}_4$.

Fig. 12 provides the achievable rate regions of each strategy for an SNR instance in which $\gamma_{\text{SR}} = 2$, $\gamma_{\text{SD}_1} = 1$, $\gamma_{\text{RD}_1} = 4$, $\gamma_{\text{SD}_2} = 2$, $\gamma_{\text{RD}_2} = 1$, $\gamma_{\text{RD}_3} = 1$. For this instance, it can be seen from Fig. 12 that $\mathcal{R}_1 \subset \mathcal{R}_2 \subset \mathcal{R}_4 \subset \mathcal{R}_3$. Strategy 1 in which D_1 and D_2 treat undesired signal from R as interference, in comparison with Strategy 2 in which D_1 and D_2 treat undesired signal from R as noise, yields smaller achievable rate region. Interestingly, the same decoding does not induce rate loss in Strategy 3 in comparison with Strategy 4. Under this SNR condition, it is beneficial to use Strategy 3 which yields a larger achievable rate region.

4.2 Conclusion

In this chapter, we illustrate the advantage of the procedure provided in Chapter 3. We considered two networks, a two-destination broadcast relay chain network and a partially cooperative multmessage network. In both networks, side information is only available to a subset of the receiving nodes, but not to the rest of the receiving nodes. Our findings are confirmed by numerical evaluation of Gaussian instances of these networks.

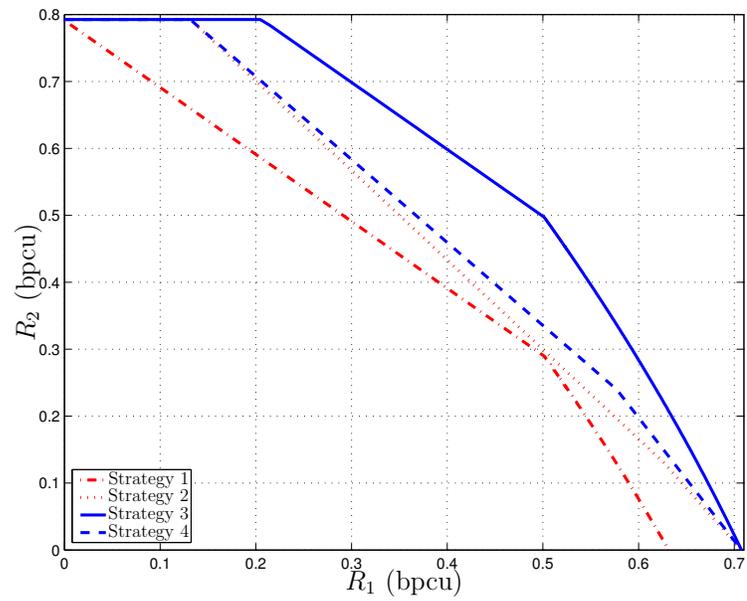


Figure 12: Rate region for the network in Figure 11 ($\gamma_{\text{SR}} = 2$, $\gamma_{\text{SD}_1} = 1$, $\gamma_{\text{RD}_1} = 4$, $\gamma_{\text{SD}_2} = 2$, $\gamma_{\text{RD}_2} = 1$, $\gamma_{\text{RD}_3} = 1$).

Chapter 5

Military Application of the Advanced Relaying Strategy

In this chapter, we study a friendly eavesdropper who assists communication in a broadcast scenario in which one transmitter wishes to send a common message to two receivers in the presence of a malicious jammer. The jammer attempts to disrupt communication by transmitting a high power Gaussian signal, whereas the friendly eavesdropper ‘hears’ the jammer’s transmission and sends an assisting signal to the destinations over an orthogonal channel in order to help them alleviate the jammer’s impact. We derive an expression for capacity, i.e., the maximum data rate that can be reliably communicated from the transmitter to the receivers, and we show that it is optimal for the friendly eavesdropper to send a Gaussian description of the jamming signal with the help of a scheme based on a modified compress-and-forward relaying that uses the decoding procedure provided in the proof of Theorem 5. The work in this chapter is to be presented in the IEEE Information Theory Workshop in Jeju, Korea, November, 2015, see [16].

5.1 Introduction

In various military applications a transmitter may wish to send a common message to multiple receivers in the presence of an antagonistic jammer. This situation arises, for instance in the scenario illustrated in Figure 13 when an unmanned drone wishes to send a description of the battlefield to ground troops, and the adversary attempts to disrupt communication. The impact of the jammer can be partially alleviated by an ally agent in the geographic proximity of the jammer which acts as a friendly eavesdropper that ‘hears’ the jammer’s signal and sends a description thereof to multiple receivers over an orthogonal channel. Neither the optimal signalling strategy of the eavesdropper nor a quantification of its utility is available, and the focus of this chapter is to investigate these aspects.



Figure 13: An illustration of the considered communication system.

The impact of jamming has been considered in various communication scenarios. For instance, cases in which the jammer sends a correlated version of the transmitter’s signal were considered in [51] from a mean-squared error perspective and in [52] from a capacity perspective. The case in which the jammer’s signal are not correlated with the transmitter’s signal was considered in [53]. It was shown therein that under

individual average power constraints the transmitter signal that enables the highest data rate to be communicated is Gaussian distributed and the jammer's signal that minimizes the communicated rate is also Gaussian distributed. Applications of Gaussian jamming and counter jamming were studied in [54–56]. For instance, service disruption due to the injection of malicious signals into an all-optical-network was considered in [54], whereas the effect of multiple antenna jamming and the potential of counter jamming in multi-carrier direct-sequence spread-spectrum systems were considered in [55] and [56], respectively. Other instances of communication scenarios in the presence of jamming can be found in [57–59].

In this chapter, we consider the situation in which a friendly eavesdropper assists communication in a broadcast scenario [60]. In this scenario one transmitter wishes to send a common message to two receivers in the presence of a malicious jammer that sends a zero mean Gaussian signal; the jammer and the transmitter's signals are uncorrelated. The received power of the jammer's signal is much higher than that of the receivers' background noises, which are therefore assumed to be negligible, see e.g., [56]. A friendly eavesdropper is able to pick the jammer's signal and attempts to assist the receivers by sending a description of the jammer's signal on an orthogonal channel. The channel between the eavesdropper and the receivers can be modelled as another Gaussian broadcast one. In fact, it is the noises on the links between the eavesdropper and the receivers that render rate-efficient communication challenging; without these noises the eavesdropper can simply forward its observation to the receivers in order to eliminate the jammer's signal. The eavesdropper has a maximum power budget which induces a constraint on its maximum transmission rate. The structure of the jammer's signal is unknown to the eavesdropper and the receivers.

To ensure causality, the eavesdropper's transmitted signal lags its received signal by one block. This implies that the jammer's, and subsequently the receivers' signals, are statistically independent of the eavesdropper transmitted signal. To analyze the

maximum data rate that can be communicated between the transmitter and the receivers, we conceive the role of the friendly eavesdropper as that of a standard relay, but with the exception that the relay (eavesdropper) in this case has no access to the transmitter's signal. Hence, the channel between the transmitter and the receivers resembles a broadcast relay channel with strictly causal side information at the relay, but with the key difference with this broadcast scenario [61] being that the eavesdropper does not have access to the transmitter's codebooks, and the key difference with relaying schemes with strictly causal side information being that these schemes do not consider a broadcasting scenario [62]. A counterpart of the scenario considered herein is the one in [63]. Therein the eavesdropper was malicious and a friendly jammer (relay) forwarded noise to the eavesdropper to confuse it.

To derive an expression for the capacity of the channel considered herein, we begin by deriving an expression for the corresponding cut-set upper bound [18]. We then show that this bound can be achieved by a signalling strategy in which the friendly eavesdropper uses a scheme based on compress-and-forward (CF) [2] to send a description of the Gaussian jamming signal to the receivers. To decode the eavesdropper's signal and to subsequently use it to alleviate the effect of jamming, the receivers use the decoding procedure provided in the proof of Theorem 5 (also cf. [38]) rather than the standard CF one.

It is worth noting that although other relaying techniques might be able to achieve the capacity of the channel considered herein, neither amplify-and-forward [21] nor decode-and-forward [2] does: amplify-and-forward yields a strictly lower rate, as will be shown below, and decode-and-forward can be readily excluded because the jammer does not cooperate with the eavesdropper.

5.2 Channel Model

Consider the channel model depicted in Figure 14. In this figure, the transmitter sends a common signal to two receivers that cannot collaborate. A malicious jammer attempts to disrupt communication by sending an independent Gaussian signal. The jammer's signal is 'heard' by a friendly eavesdropper, which attempts to assist the receivers by sending a description of the jammer's signal over an orthogonal channel.

Let the signal sent by the transmitter be denoted by X and let the signal sent by the eavesdropper on an orthogonal channel be denoted by X_e . The jamming signal is denoted by $J \sim \mathcal{N}(0, P_J)$. The jammer-eavesdropper channel gain is normalized to 1, the transmitter-receiver i channel gain is denoted by a_i and the jammer-receiver i channel gain is denoted by b_i , $i = 1, 2$. The received signal of the eavesdropper is denoted by Y_e . Each receiver i , receives two orthogonal signals: one from the transmitter contaminated by the jammer's signal, which is denoted by Y_i , and one from the friendly eavesdropper contaminated by additive Gaussian noise, which is denoted by $Y_{s,i}$, $i = 1, 2$. We denote the additive Gaussian noise component of $Y_{s,i}$ by $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, 2$. The friendly eavesdropper transmits its signal at a rate $R_e \leq \sup_{p(x_e)} \max_{i=1,2} I(X_e; Y_{s,i})$. Using this notation, the received signals at the eavesdropper and receiver i , $i = 1, 2$, can be expressed as

$$\begin{aligned} Y_e &= J, \\ Y_1 &= a_1 X + b_1 J, \quad Y_2 = a_2 X + b_2 J, \\ Y_{s,1} &= X_e + Z_1, \quad Y_{s,2} = X_e + Z_2. \end{aligned} \tag{48}$$

The average power of the jammer's Gaussian signal is given by $E(J^2) = P_J$, whereas the transmitter and the friendly eavesdropper each is subject to its individual average

transmit power constraint, $E(X^2) \leq P$, and $E(X_e^2) \leq P_e$, respectively. In the next

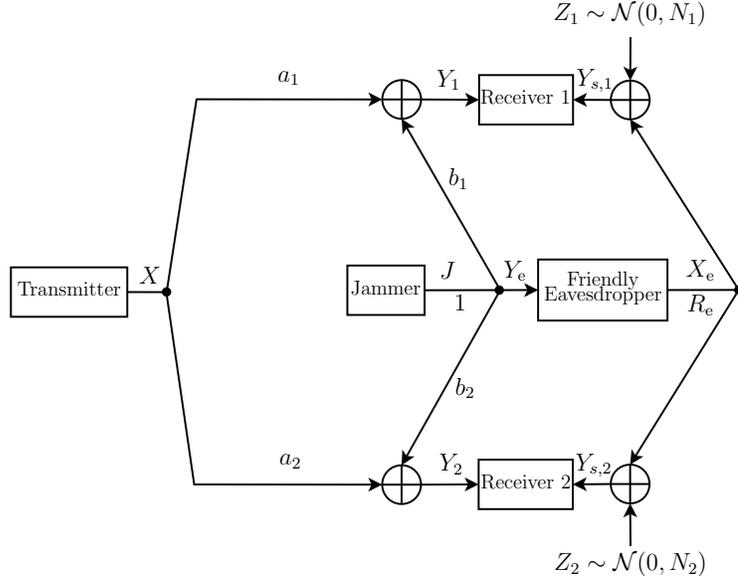


Figure 14: A broadcast channel in the presence of a Gaussian jammer with a friendly eavesdropper, $J \sim \mathcal{N}(0, P_J)$.

section, we will show that a modified CF scheme that uses the decoding procedure provided in the proof of Theorem 5 can be used to achieve the capacity of the channel in Figure 14.

5.3 Capacity Results

Let the maximum signal-to-jamming power ratio (SJR) of the transmitter-receiver i channel be denoted by $\gamma_i = \frac{a_i^2 P}{b_i^2 P_J}$, $i = 1, 2$, and let the maximum SNR of the eavesdropper-receiver i channel be denoted by $\gamma_{e,i} = \frac{P_e}{N_i}$, $i = 1, 2$. Let $\mathcal{C}(x) \triangleq \frac{1}{2} \log_2(1 + x)$. Our main result is recorded in the following theorem.

Theorem 7. *The capacity of the channel in Figure 14 can be achieved when the eavesdropper sends a Gaussian description of the jammer's signal at a rate $R_e =$*

$\max\{\mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_{e,2})\}$. This capacity is given by

$$C = \min\{\mathcal{C}(\gamma_1) + \mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_2) + \mathcal{C}(\gamma_{e,2})\}. \quad (49)$$

Proof. To prove the converse, we first show that choosing X and X_e to be Gaussian-distributed with average powers P and P_e , respectively, maximizes the cut-set upper bound [41, Sect. 18.1].

To complete the proof of the theorem, we show that the cut-set upper bound can be achieved when the eavesdropper uses a strategy that resembles standard CF relaying, but with the decoding procedure provided in the proof of Theorem 5. In particular, the eavesdropper uses two Gaussian codebooks $\hat{\mathcal{Y}}_e$ and \mathcal{X}_e with the powers and rates described in Appendix D.1. Upon receiving the jammer's signal, the eavesdropper finds a codeword in $\hat{\mathcal{Y}}_e$ that is jointly typical with it. The eavesdropper uses Wyner-Ziv binning [8] to determine the codeword to be transmitted in the next block from \mathcal{X}_e . Instead of using standard CF decoding, the receivers use the decoding procedure provided in the proof of Theorem 5 (cf. [38]) to recover the eavesdropper's message. In the decoding procedure, each receiver uses its knowledge of the codebooks $\hat{\mathcal{Y}}_e$ and \mathcal{X} as side information to recover the message from the eavesdropper and subsequently the message from the transmitter. (In standard CF $\hat{\mathcal{Y}}_e$ and \mathcal{X} are not used in recovering the message from the eavesdropper.)

Detailed proof is provided below.

Proof of Converse: Using the cut-set bound, the rate, R , of the common message to both receivers can be upper bounded by

$$R \leq \min_{i=1,2} I(X, X_e; Y_i, Y_{s,i}). \quad (50)$$

From (50), it follows that, for $i = 1, 2$,

$$\begin{aligned} R &\leq I(X, X_e; Y_i, Y_{s,i}) \\ &= h(Y_i, Y_{s,i}) - h(Y_i, Y_{s,i} | X, X_e) \\ &= h(a_i X + J, X_e + Z_i) - h(J, Z_i). \end{aligned}$$

Since J and Z_i are independent, and X and X_e are independent of J and Z_i , choosing X and X_e to be independent maximizes $h(a_i X + J, X_e + Z_i)$, and hence maximizes $I(X, X_e; Y_i, Y_{s,i})$. The independence of (X, Y_i) and $(X_e, Y_{s,i})$ implies that

$$R \leq I(X; Y_i) + I(X_e; Y_{s,i}), \quad i = 1, 2. \quad (51)$$

Since $Y_{s,i}$ and Y_i are received on orthogonal channels, and X and X_e must satisfy their respective average power constraints, it can be readily seen that choosing X and X_e to be independent Gaussian random variables maximizes the right hand side of (51), whence $R \leq \mathcal{C}(\gamma_i) + \mathcal{C}(\gamma_{e,i})$.

Proof of Achievability: The proof of achievability resembles that of Theorem 5 in Section. 3.3. See detailed proof in Appendix D.1.

□

It is worth noting that in proving Theorem 7 we assumed that both receivers use the same decoding procedure. However, the statement of the theorem holds if the receiver with less noise power on the eavesdropper link uses standard CF decoding to recover the eavesdropper's message; standard CF decoding is more straightforward.

In contrast to CF, NNC and SNNC, hash-and-forward (HF) [31] is a relaying scheme in which the relay partitions the received sequences, rather than the description thereof, into non-overlapping bins and transmits the bin index (hash index) to

the receiver to assist decoding. We note that similar to Wyner-Ziv binning, hashing results in N -to-1 mapping. It has been shown in [31] and [45] that HF and conventional CF are capacity achieving for a class of deterministic relay channels. In that class, the relay received signal is a function of the source transmitted signal and the received signal at the destination, and the relay-to-destination link is a noiseless link with capacity R_0 . It can be shown that HF can also achieve the capacity of the network considered herein. However, HF requires that the relay has the knowledge of its all possible channel output sequences in order to construct random hashing (binning), which makes it unsuitable in dealing with the jamming channel since the jammer's signal is not known.

So far, we have shown that the capacity of the channel described in Section 5.2 can be achieved when the eavesdropper sends a Gaussian description of the jammer's signal at a rate $R_e = \max\{\mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_{e,2})\}$. Since this rate is higher than the capacity of the link between the eavesdropper and the receiver with the higher noise, this receiver will not be able to recover the eavesdropper message if it uses standard CF decoding [62], but will be able to recover it if it uses the decoding procedure provided in the proof of Theorem 5 which incorporates \mathcal{X} .

5.4 Comparison with Other Eavesdropping Signalling Schemes

We now compare the rates that can be achieved in the absence of the friendly eavesdropper, and when this eavesdropper uses either CF with standard decoding or AF relaying.

5.4.1 No Eavesdropper Case

In the absence of the eavesdropper, the channel capacity can be readily seen to be

$$C_{\text{No Eavesdropper}} = \min_{i=1,2} \mathcal{C}(\gamma_i). \quad (52)$$

Hence, using $\Delta \triangleq |\mathcal{C}(\gamma_1) - \mathcal{C}(\gamma_2)|$, the rate gain provided by the friendly eavesdropper can be expressed as

$$\begin{aligned} \min\{\Delta + \mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_{e,2})\}, & \quad \gamma_1 \geq \gamma_2, \\ \min\{\mathcal{C}(\gamma_{e,1}), \Delta + \mathcal{C}(\gamma_{e,2})\}, & \quad \gamma_1 < \gamma_2. \end{aligned}$$

□

It is of interest to note that when the transmitter (the drone) is sufficiently far from the receivers (ground troops), $\Delta \approx 0$ and the advantage of having the eavesdropper is approximately $\min\{\mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_{e,2})\}$. This is in contrast with the eavesdropper rate, which is given by $\max\{\mathcal{C}(\gamma_{e,1}), \mathcal{C}(\gamma_{e,2})\}$.

5.4.2 CF With Standard Decoding Case

Using standard CF decoding [62] at both receivers to recover the eavesdropper's message without using $\hat{\mathcal{Y}}_e$ and \mathcal{X} induces a constraint on the eavesdropper's transmission rate, R_e . In particular, $R_e \leq \min_{i=1,2} I(X_e; Y_{s,i})$. Using the standard approach, it can be verified that CF with Gaussian codebooks and standard decoding achieves the following rate:

$$R_{\text{CF}} \leq \min_{i=1,2} \mathcal{C}(\gamma_i) + \min_{i=1,2} \mathcal{C}(\gamma_{e,i}), \quad (53)$$

which is generally less than capacity, cf. Section 5.5.

5.4.3 Amplify-and-Forward Case

When the eavesdropper uses non-regenerative AF relaying, the optimal receivers' strategy can be readily seen to be using the signal received from the eavesdropper to partially cancel the jamming signal. The maximum rate that can be achieved by this scheme is given by

$$R_{\text{AF}} = \min_{i=1,2} \{ \mathcal{C}(\gamma_i(1 + \gamma_{e,i})) \}. \quad (54)$$

Proof. See details in Appendix D.2. □

We observe that, although the eavesdropper receives a noiseless version of the jamming signal, the rate achieved by this scheme is strictly below capacity.

5.5 Numerical Comparison

To illustrate the advantage of CF with the decoding procedure provided in the proof of Theorem 5, in Figure 15 we compare the rates achieved by the schemes outlined in Section 5.4 with the capacity expression provided in Theorem 7. In particular, the capacity expression (49) is compared with: 1) the capacity of the broadcast channel in the absence of the friendly eavesdropper, cf. (52); 2) the rate achieved when the eavesdropper uses CF, but the receivers use standard decoders to recover the eavesdropper message, cf. (53); and 3) the rate achieved when the eavesdropper uses AF, cf. (54). We consider instances in which the SJR of receiver 1, γ_1 , varies from 0 to 4.5 and the SJR of receiver 2 is $\gamma_2 = 4$, and the SNRs of the eavesdropper to the receivers channels are $\gamma_{e,1} = 3$ and $\gamma_{e,2} = 2$. From this figure, it can be seen that the achievable rate of CF with standard decoding is strictly below capacity in the high jamming regime, e.g., when $\gamma_1 < 4$. When the jamming signal power is relatively low, e.g., when $\gamma_1 \geq 4$, CF with standard decoding also achieves capacity. In contrast, the rate achieved by AF is strictly below capacity. The gap

between the rate achieved when the eavesdropper uses AF and when it uses CF with the decoding procedure provided in the proof of Theorem 5 (capacity) is reduced as γ_1 increases from 0 to 3. For $\gamma_1 \geq 3$, this gap is constant and strictly greater than zero. Without the friendly eavesdropper, the capacity is strictly below the rate achieved when the eavesdropper uses CF with either standard decoding or the one provided in the proof of Theorem 5.

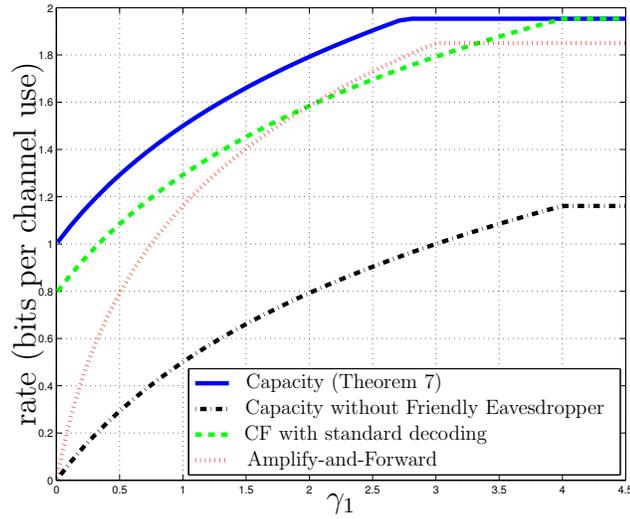


Figure 15: Comparison between capacity (Theorem 7) and the rates achievable using the schemes in Section 5.4, for $\gamma_2 = 4$, $\gamma_{e,1} = 3$, $\gamma_{e,2} = 2$.

5.6 Conclusion

We considered a two-receiver broadcast channel with a Gaussian jammer and a friendly eavesdropper. The friendly eavesdropper ‘hears’ the jamming signal and sends a description thereof to the receivers to help them to reduce the jammer’s impact. We showed that the capacity of this channel can be achieved when the eavesdropper uses CF relaying with Gaussian codebooks and the receivers use the decoding procedure provided in the proof of Theorem 5 to recover the eavesdropper’s

message. Our results are confirmed by analytical and numerical comparisons.

Chapter 6

Conclusions and Future Work

6.1 Conclusions and Contributions

The work described in this thesis has been concerned with the signaling structures and procedures of various relaying schemes to achieve higher rate in relay channels and relay networks. In this final chapter, we will conclude by describing the progress made towards this goal. We will also suggest some future research directions that could provide next steps along the path to achieve higher rate in the relay channels.

The conclusions and contributions of this thesis can be summarized as follows:

- We showed that when $\gamma_0(1 + \gamma_2) < \gamma_1$, generalized DF-CF with Gaussian codebooks reduces to the underlying DF; when $\gamma_1 \leq \gamma_0$, generalized DF-CF reduces to the underlying CF.
- We proved that there exist SNR regions in which the generalized DF-CF scheme is guaranteed to yield higher achievable rates than the underlying DF and CF schemes; however, the rate advantage of generalized DF-CF is within 0.5 bpcu.
- In light of the generalization of DF and CF, we showed that switching between DF and CF is able to achieve higher rates than both fixed DF and CF.

- We proposed a new decoding procedure for CF and showed that using this procedure, the rate constraint on the bin indices can be relaxed, and it is able to achieve the same rate as NNC and SNNC in the multimessage network.
- We demonstrated the rate advantage of the proposed decoding procedure in a broadcast relay chain network and a partial cooperative network when side information is available to only a subset of the receiving nodes in the networks.
- We introduced a “friendly” eavesdropper in the military communication network in the presence of a malicious jammer and studied the maximum reliable communication rate of this network. We provided the cut-set bound of this network and showed that it can be achieved using CF with the proposed decoding procedure.

6.2 Future Work

The results presented in this thesis have demonstrated the rate advantage of several CF related relaying schemes, namely, the generalization of DF and CF [2, Theorem 7], and the proposed decoding procedure for CF. The investigation of the CF related relaying schemes can be extended in a number of ways:

- In the multimessage network considered in Chapter 3, the proofs of the achievability of Theorems 5 and 6 adopt a network coding philosophy whereby each node encodes its own message and the description of its received signal, which mirrors the one used in [9]. However, for the case that all codebooks, including the description codebooks of all the nodes, are exposed to each node in the network, certain improvement would merit further investigation.

It can be shown that the recovery of other nodes’ description codewords of the received signals does not induce rate loss at the decoding node. Using this

observation, the decoding node can provide auxiliary information of the other nodes' description and send it into the network. We conjecture that there may exist conditions under which some other destinations will be able to benefit from such auxiliary information.

- In Chapter 5, we introduced a “friendly eavesdropper” to a communication network in the presence of a malicious jammer. There is a plethora of possible areas in which the schemes to achieve higher rate is of great interest for further investigation. Two immediate cases are provided below.
 - Case 1: A “friendly” eavesdropper in the vicinity of the source is able to decode the source message and provide a description of the jammer’s signal to the receiver. In this case, it might be beneficial to explore such a cooperative strategy that the eavesdropper cooperates with the source using DF and describes the jammer’s signal using CF. The eavesdropper may superimpose the CF codeword on the DF codeword. Hence, the eavesdropper is able to both assist and facilitate the decoding at the receivers. The codebook structure and the procedure are, to some extent, similar to the generalized DF-CF in [2, Theorem 7], which is analyzed in Chapter 2 in this thesis. In contrast to that generalization, here the signal from the source and the signal from the jammer are not superimposed. We expect that a decoding procedure at the receivers that on one hand exploits the assistance of the DF strategy provided by the eavesdropper and on the other hand alleviates the impact of the jammer’s signal using the CF strategy provided by the eavesdropper will be able to take advantage of the presence of the eavesdropper.
 - Case 2: A “friendly” eavesdropper, in the vicinity of the jammer, picks the jammer’s signal and embeds the bin index of the description of the

jammer’s signal in a host content and transmits it. This resembles, to some extent, watermarking. The use of such watermarking in the relay broadcast channel can be used as a practical provision to provide information security to the eavesdropper’s signal. To analyze the optimal signaling strategy for the eavesdropper in this case, the host content can be modeled as non-causal channel state information available at the encoder (eavesdropper). In this case, integrating the decoding procedure proposed in Chapter 3 into the framework of relay on dirty-paper may be optimal under certain conditions.

6.3 Thesis Summary

In this thesis, we investigated generalized DF-CF in the standard three-node relay channel in Gaussian settings in Chapter 2. We showed that in various SNR regions, this generalization reduces to its underlying individual DF and CF. It was proved that there exists SNR regions in which this generalization is guaranteed to yield higher rate than both DF and CF. However, such gain is within 0.5 bpcu. In our investigation of the CF-based relaying schemes in Chapter 3, we proposed a decoding procedure that uses N -to-1 mapping instead of 1-to-1 mapping and showed that it allows CF to relax the rate constraint on the bin indices. The rate advantage of exploiting the N -to-1 mapping is shown in two network examples in Chapter 4. In Chapter 5, we introduced a “friendly” eavesdropper in a broadcast relay channel in the presence of a Gaussian jammer. We showed that the proposed procedure is capacity achieving for that channel. We believe that the work done in this thesis can be extended in various ways.

Appendices

A.1 Proof of Proposition 1

To prove the proposition, we use the standard definitions for entropy and mutual information (cf. [18]) and use (1) and (5).

Note that the left hand side of (3b) yields

$$\begin{aligned}
& I(\hat{Y}_1; Y_1 | Y, X_2, U) \\
&= h(\hat{Y}_1 | Y, X_2, U) - h(\hat{Y}_1 | Y_1, Y, X_2, U) \\
&= h(Y_1 + Z_1 + Z' | X_1 + bX_2 + Z, X_2, U) - h(Y_1 + Z_1 + Z' | X_1 + bX_2 + Z, X_2, U, Y_1) \\
&= h(aX_{12} + Z_1 + Z' | X_{12} + Z) - h(Z') \\
&= \frac{1}{2} \log_2 \left(\alpha_2 a^2 P_{X_1} + N + N' - \frac{\alpha_2^2 a^2 P_{X_1}^2}{\alpha_2 P_{X_1} + N} \right) - \frac{1}{2} \log_2 N' \\
&= \frac{1}{2} \log_2 \left(1 + \frac{1}{\gamma'} + \frac{\alpha_2 \gamma_1}{\gamma' (1 + \alpha_2 \gamma_0)} \right). \tag{55}
\end{aligned}$$

The right hand side of (3b) yields

$$\begin{aligned}
& I(X_2; Y | V) \\
&= h(Y | V) - h(Y | X_2, V) \\
&= h(X_1 + bX_2 + Z | V) - h(X_1 + bX_2 + Z | X_2, V) \\
&= h(X_{11} + X_{12} + b\beta_1 X_{21} + Z) - h(X_{11} + X_{12} + Z) \\
&= \frac{1}{2} \log_2 \frac{(\alpha_1 + \alpha_2) P_{X_1} + \beta_1 b^2 P_{X_2} + N}{(\alpha_1 + \alpha_2) P_{X_1} + N} \\
&= \frac{1}{2} \log_2 \left(1 + \frac{\beta_1 \gamma_2}{1 + (\alpha_1 + \alpha_2) \gamma_0} \right). \tag{56}
\end{aligned}$$

Substituting (55) and (56) into (3b), we have

$$\frac{1}{\gamma'} + \frac{\alpha_2 \gamma_1}{\gamma' (1 + \alpha_2 \gamma_0)} \leq \frac{\beta_1 \gamma_2}{1 + (\alpha_1 + \alpha_2) \gamma_0},$$

which, after simplification, yields (6b).

From the first term inside the minimization of (3a) we have

$$\begin{aligned}
R_1 &= I(X_1; Y, \hat{Y}_1 | X_2, U) + I(U; Y_1 | X_2, V) \\
&= h(Y, \hat{Y}_1 | X_2, U) - h(Y, \hat{Y}_1 | Y, X_2, U) \\
&= h(X_1 + bX_2 + Z, aX_1 + Z_1 + Z' | X_2, U) \\
&\quad - h(X_1 + bX_2 + Z, aX_1 + Z_1 + Z' | X_1, X_2, U) \\
&\quad + h(aX_1 + Z_1 | X_2, V) - h(aX_1 + Z_1 | X_2, U, V) \\
&= h(X_{12} + Z, aX_{12} + Z_1 + Z') - h(Z, Z_1 + Z') \\
&\quad + h(a(X_{11} + X_{12}) + Z_1) - h(aX_{12} + Z_1) \\
&= \frac{1}{2} \log_2 \left(\frac{(\alpha_2 P_{X_1}^2 + N)(a^2 \alpha_2 P_{X_1}^2 + N + N') - a^2 \alpha_2^2 P_{X_1}^4}{N(N + N')} \right. \\
&\quad \left. \cdot \frac{a^2(\alpha_1 + \alpha_2)P_{X_1}^2 + N}{a^2 \alpha_2 P_{X_1}^2 + N} \right) \\
&= \frac{1}{2} \log_2 \left(\left(1 + \frac{\alpha_2 P_{X_1}^2}{N} + \frac{a^2 \alpha_2 P_{X_1}^2}{N + N'} \right) \left(1 + \frac{a^2 \alpha_1 P_{X_1}^2}{a^2 \alpha_2 P_{X_1}^2 + N} \right) \right) \\
&= \frac{1}{2} \log_2 \left(\left(1 + \alpha_2 \gamma_0 + \frac{\alpha_2 \gamma_1}{1 + \gamma'} \right) \left(1 + \frac{\alpha_1 \gamma_0}{1 + \alpha_2 \gamma_0} \right) \right),
\end{aligned}$$

which yields (6c).

From the second term inside the minimization of (3a) we have

$$\begin{aligned}
R_1 &= I(X_1, X_2; Y) - I(\hat{Y}_1; Y_1 | X_2, X_1, U, Y) \\
&= h(Y) - h(Y | X_1, X_2) - h(\hat{Y}_1 | X_2, X_1, U, Y) + h(\hat{Y}_1 | Y_1, X_2, X_1, U, Y) \\
&= h(X_1 + bX_2 + Z) - h(Z) \\
&\quad - h(aX_1 + Z_1 + Z' | X_1, X_2, U, Y) + h(aX_1 + Z_1 + Z' | X_1, X_2, U, Y, Y_1) \\
&= h(X_1 + bX_2 + Z) - h(Z) - h(Z_1 + Z') + h(Z') \\
&= \frac{1}{2} \log_2 \left(\frac{a^2 P_{X_1} + b^2 P_{X_2} + 2\sqrt{b^2 \alpha_0 P_{X_1} \beta_0 P_{X_2}} + N}{N} \right) - \frac{1}{2} \log_2 \left(\frac{N + N'}{N'} \right) \\
&= \frac{1}{2} \log_2 (1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}) - \frac{1}{2} \log_2 \left(1 + \frac{1}{\gamma'} \right),
\end{aligned}$$

which yields (6d).

Constraint (6e) and (6f) follow from the definition of α_i and β_j for $i = 0, 1, 2$ and $j = 0, 1$, respectively.

A.2 Proof of the achievable rate expressions of DF and CF in the Gaussian channel

The achievable of the DF relaying scheme (cf. [2, Theorem 1]) by

$$R_{\text{DF}} \leq \sup \{ \min \{ I(X_1; Y_1 | X_2), I(X_1, X_2; Y) \} \}.$$

Using (1), (2) and $\sqrt{\rho_0} = E\{X_1 X_2\}$, $\rho_0 + \rho_1 = 1$, we have

$$\begin{aligned}
I(X_1; Y_1 | X_2) &= h(Y_1 | X_2) - h(Y_1 | X_1, X_2) \\
&= h(a\sqrt{\rho_1}X_1 + Z_1) - h(Z_1) \\
&= \frac{1}{2} \log_2(1 + \rho_1 \gamma_1) \\
&= \mathcal{C}(\rho_1 \gamma_1).
\end{aligned} \tag{57}$$

$$\begin{aligned}
I(X_1, X_2; Y) &= h(Y) - h(Y | X_1, X_2) \\
&= h(X_1 + bX_2 + Z) - h(Z) \\
&= \frac{1}{2} \log_2(1 + \gamma_0 + \gamma_2 + 2\sqrt{\rho_0 \gamma_0 \gamma_2}) \\
&= \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\rho_0 \gamma_0 \gamma_2}).
\end{aligned} \tag{58}$$

Using (57) and (58), we have

$$R_{\text{DF}} \leq \max_{\rho_1} \min\{\mathcal{C}(\rho_1 \gamma_1), \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\rho_0 \gamma_0 \gamma_2})\}$$

This completes the proof for DF.

The achievable rate expression of the CF scheme in [41] is given by

$$R_{\text{CF}} \leq \sup \min\{I(X_1; \hat{Y}_1, Y | X_2), I(X_1, X_2; Y) - I(\hat{Y}_1; Y_1 | X_1, X_2, Y)\}.$$

Using (1), (2), $\hat{Y}_1 = Y_1 + X_2 + Z_1 + Z'$, where $Z' \sim \mathcal{N}(0, N')$, and $\gamma' = \frac{N'}{N}$, we have

$$\begin{aligned}
& I(X_1; \hat{Y}_1, Y | X_2) \\
&= h(\hat{Y}_1, Y | X_2) - h(\hat{Y}_1, Y | X_1, X_2) \\
&= h(aX_1 + Z_1 + Z', X_1 + bX_2 + Z) - h(Z_1 + Z', Z) \\
&= \frac{1}{2} \log_2 \left(1 + \gamma_0 + \frac{\gamma_1}{1 + \gamma'} \right) \\
&= \mathcal{C} \left(\gamma_0 + \frac{\gamma_1}{1 + \gamma'} \right). \tag{59}
\end{aligned}$$

$$\begin{aligned}
& I(X_1, X_2; Y) - I(I(\hat{Y}_1; Y_1 | X_1, X_2, Y)) \\
&= h(Y) - h(Y | X_1, X_2) - h(I(\hat{Y}_1 | X_1, X_2, Y)) + h((I(\hat{Y}_1; Y_1 | Y_1, X_1, X_2, Y))) \\
&= \frac{1}{2} \log_2(1 + \gamma_0 + \gamma_2) - \frac{1}{2} \log_2(1 + 1/\gamma') \\
&= \mathcal{C}(\gamma_0 + \gamma_2) - \mathcal{C}(1/\gamma'). \tag{60}
\end{aligned}$$

Using (59) and (60), we have

$$R_{\text{CF}} \leq \max_{\gamma'} \min \left\{ \mathcal{C} \left(\gamma_0 + \frac{\gamma_1}{1 + \gamma'} \right), \mathcal{C}(\gamma_0 + \gamma_2) - \mathcal{C}(1/\gamma') \right\}$$

Since the two terms inside the minimization are monotonically increasing and decreasing, respectively, over γ' , equalizing the two terms yields that the optimal value of γ' is given by

$$\gamma'_{\text{CF}}^* = \frac{1 + \gamma_0 + \gamma_1}{\gamma_2}$$

and

$$R_{\text{CF}}^* = \mathcal{C} \left(\gamma_0 + \frac{\gamma_1 \gamma_2}{1 + \gamma_0 + \gamma_1 + \gamma_2} \right).$$

This completes the proof for CF.

A.3 Proof of Theorem 1

To prove Theorem 1, we consider the equality in (8c). It can be seen from Lemma 1 that, because $\gamma_1 > \gamma_0(1 + \gamma_2)$, the coefficient of λ_1 in this equality and the coefficient of λ_3 are strictly positive. There are two distinct possibilities, namely, $\lambda_1 = 0$ and $\lambda_1 > 0$.

Case 1 ($\lambda_1 = 0$). In this case, invoking (8a) yields $\lambda_2 = 1$. Substituting $\lambda_1 = 0$ and $\lambda_2 = 1$ in (8e) yields

$$\frac{1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\beta_0\gamma_0\gamma_2}}{(1 + \gamma')^2} = -\lambda_3 \frac{\gamma_2(1 - \beta_0)}{1 + \gamma_0(1 - \alpha_0)}. \quad (61)$$

This equality is satisfied if and only if both sides of this equation are zero. Hence, γ' must be infinite and $\lambda_3 = 0$ or $\beta_0 = 1$.

We will show that $\lambda_6 > 0$. To do so, we consider (8b) and use $\gamma' = \infty$, $\lambda_1 = 0$ and $\lambda_2 = 1$ to write

$$\sqrt{\gamma_0\gamma_2}\sqrt{\frac{\beta_0}{\alpha_0}} + \lambda_3 \frac{\gamma_0\gamma_2\gamma'(1 - \beta_0)}{(1 + (1 - \alpha_0)\gamma_0)^2} + \lambda_5 = \lambda_6. \quad (62)$$

For any non-negative λ_3 , if $0 < \beta_0 \leq 1$, the left hand side of (62) is strictly positive, which implies that $\lambda_6 > 0$. It remains to consider the case of $\beta_0 = 0$. In this case the first term on the left hand side of (62) vanishes and we are left with two possibilities: either $\lambda_3\gamma' > 0$, which yields $\lambda_6 > 0$, or $\lambda_3\gamma' = 0$. In the latter case, we consider (8d), from which we have

$$\sqrt{\gamma_0\gamma_2}\sqrt{\frac{\alpha_0}{\beta_0}} + \lambda_8 = \lambda_7. \quad (63)$$

Since $\beta_0 = 0$, we have from the last equality in (8f) with $i = 7$ that $\lambda_7 = 0$. Using

this observation and the fact that $\beta_0 = 0$ implies that $\sqrt{\frac{\alpha_0}{\beta_0}} \rightarrow \infty$ in (63) yields a contradiction¹, i.e., the KKT system cannot be satisfied with $\lambda_1 = 0$, $\beta_0 = 0$ and $\lambda_3\gamma' = 0$.

Hence, for the KKT system in (8) to be satisfied with $\lambda_1 = 0$ when $\gamma_1 > \gamma_0(1 + \gamma_2)$, we must have $\lambda_6 > 0$. Using this in (8c) yields $\lambda_4 > 0$, which using (8f) yields the following optimal power partitions: $\alpha_0 = 1$, $\alpha_1 = 0$, $\alpha_2 = 0$. Using these partitions together with $\gamma' \rightarrow \infty$, yields $R_1 = 0$ in (6c). Hence, the solution of the KKT system with $\lambda_1 = 0$ does not correspond to the maximum rate of the generalized DF-CF scheme.

Case 2 ($\lambda_1 > 0$). Using Lemma 1, it can be seen that the first term in (8c) is strictly positive, which implies that $\lambda_4 > 0$ and thus $\alpha_2 = 0$. Substituting in (6c) yields

$$R_1 = \mathcal{C}(\alpha_1\gamma_1), \quad (64)$$

which does not depend on β_0 and γ' . Hence, choosing β_0 to maximize R_2 in (6d) does not reduce R_1 . From (6d) it can be seen that R_2 is monotonically increasing in both β_0 and γ' . However, from (6b), γ' is bounded by a monotonically increasing function of β_0 . Hence, maximizing β_0 directly maximizes the positive term of R_2 and minimizes the negative term by maximizing the lower bound on γ' . In particular, R_2 is maximized by setting $\beta_0 = 1$, which yields

$$R_2 = \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\alpha_0\gamma_0\gamma_2}). \quad (65)$$

Hence, we have shown that the optimal solution of (7) corresponds to Case 2 in which $\alpha_2 = 0$ and $\beta_1 = 1 - \beta_0 = 0$. Using these settings in Proposition 1 yields the statement of the theorem.

¹Here we assume that $\sqrt{\frac{\beta_0}{\alpha_0}} = 0$. Otherwise, λ_6 would be strictly positive as claimed.

A.4 Proof of Lemma 2

From the complementarity slackness conditions in (8f), it can be seen that if λ_3 were strictly positive, the constraint on γ' must be active and the lemma is automatically proved. Hence, to complete the proof of this lemma, it remains to consider the case of $\lambda_3 = 0$, which is what we do in the forthcoming analysis.

Case 1 [$\alpha_2 = 0$ and $\gamma' \geq 0$] In this case, substituting in (6c) and (6d) yields

$$\begin{aligned} R_1 &= \mathcal{C}(\alpha_1 \gamma_1), \quad \text{and} \\ R_2 &= \frac{1}{2} \log_2 \left((1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}) \frac{\gamma'}{1 + \gamma'} \right). \end{aligned}$$

Following an argument analogous to the one used in Case 2 in the proof of Theorem 1, it can be seen that in this case, the generalized DF-CF reduces to DF and $\gamma' \rightarrow \infty$ which satisfies the constraint in (6b) with equality.

Case 2 [$\alpha_2 > 0$ and $\gamma' < \infty$] Using $\lambda_3 = 0$ and that γ' is finite in (8e) and (8d) yields

$$\lambda_1 \frac{\alpha_2 \gamma_1 (1 + (1 - \alpha_0) \gamma_1)}{1 + \alpha_2 \gamma_1} = \lambda_2 (1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}), \quad (66)$$

and

$$\lambda_2 \sqrt{\frac{\alpha_0 \gamma_0 \gamma_2}{\beta_0}} \frac{\gamma'}{1 + \gamma'} + \lambda_8 = \lambda_7. \quad (67)$$

Now, since, from (8a), $\lambda_1 + \lambda_2 = 1$, it follows from (66) that, in this case, $\lambda_1 > 0$ and $\lambda_2 > 0$.

For $\gamma_2 < \infty$, it can be seen from (6b) that $\gamma' > 0$. Using this observation in (67), we identify two possible cases: $\sqrt{\frac{\alpha_0}{\beta_0}} > 0$ and $\sqrt{\frac{\alpha_0}{\beta_0}} = 0$.

We will show that the first case yields a contradiction. In particular, suppose

that $\sqrt{\frac{\alpha_0}{\beta_0}} > 0$. In this case, it is immediate that the left hand side of (67) is strictly positive and that $\lambda_7 > 0$. This implies that $\beta_1 = 0$ (cf. (8f)). Using this in (6b) yields $\gamma' \geq \infty$, which contradicts the condition $\gamma' < \infty$.

For the second case, we have $\sqrt{\frac{\alpha_0}{\beta_0}} = 0$. Substituting for $\lambda_3 = 0$ in (8b) yields

$$\lambda_1 \left(1 + \frac{\alpha_2 \gamma_1}{1 + \gamma'} + \alpha_2 \gamma_0 \right) \frac{\gamma_1}{1 + \alpha_2 \gamma_1} + \lambda_6 = \lambda_2 \sqrt{\frac{\beta_0}{\alpha_0}} \sqrt{\gamma_0 \gamma_2} \frac{\gamma'}{1 + \gamma'} + \lambda_5. \quad (68)$$

Since $0 < \lambda_i, \gamma_j, \gamma' < \infty$, $i = 1, 2$, $j = 0, 1, 2$, it can be seen that, because $\sqrt{\frac{\alpha_0}{\beta_0}} = 0$, the right hand side of (68) is positive and infinite. Hence, it can be seen that $\lambda_6 = \infty$. This implies that $\alpha_0 + \alpha_2 = 1$. Invoking the fact that in this case $\alpha_0 = 0$, we have $\alpha_2 = 1$. Using this in (6c) and (6d) yields $R_1 = \mathcal{C}\left(\gamma_0 + \frac{\gamma_1}{1 + \gamma'}\right)$, and $R_2 = \mathcal{C}\left(\gamma_0 + \gamma_2\right) - \mathcal{C}\left(\frac{1}{\gamma'}\right)$. Optimizing $\min\{R_1, R_2\}$ over γ' satisfying (6b), yields $\beta_1 = 1$ and the constraint holds with equality.

Case 3 [$\alpha_2 > 0$ and $\gamma' \rightarrow \infty$] In this case, we consider the SNR regions: $\gamma_0 < \gamma_1$ and $\gamma_0 \geq \gamma_1$.

a) The case of $\gamma_0 < \gamma_1$: Using $\lambda_3 = 0$ and $\gamma' \rightarrow \infty$ in (8c) yields

$$\lambda_1 \frac{(1 + (1 - \alpha_0)\gamma_1)(\gamma_1 - \gamma_0)}{(1 + \alpha_2 \gamma_1)^2} - \lambda_4 + \lambda_6 = 0.$$

When $\gamma_0 < \gamma_1$, the coefficient of λ_1 is strictly positive. Following the proof of Theorem 1 it can be shown that the optimization leads to $\alpha_2 = 0$, which contradicts the condition $\alpha_2 > 0$.

b) The case of $\gamma_0 \geq \gamma_1$: In this case, we rewrite the rate expressions in Proposition 1

as

$$\begin{aligned} R_1 &= \frac{(1 + \alpha_2 \gamma_0)(1 + (\alpha_1 + \alpha_2) \gamma_1)}{1 + \alpha_2 \gamma_1}, \quad \text{and} \\ R_2 &= \mathcal{C}\left(1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}\right). \end{aligned} \quad (69)$$

Since $\gamma_0 \geq \gamma_1$, it can be seen from (69) that R_1 is maximized when $\alpha_2 = 1$. However, in this case $R_1 = \mathcal{C}(\gamma_0) < \mathcal{C}\left(\gamma_0 + \frac{\gamma_1 \gamma_2}{1 + \gamma_0 + \gamma_1 + \gamma_2}\right) = R_{\text{CF}}^*$, where the last equality follows from (10). This implies that the generalized DF-CF scheme achieves a rate that is strictly less than the rate achieved by the CF scheme. This contradicts $R_{\text{CF}}^* \leq R_{\text{G}}^*$, which follows from that the CF scheme is a special case of the generalized DF-CF.

Hence, when $\gamma_i < \infty$, $i = 0, 1, 2$, setting $\alpha_2 > 0$ and $\gamma' \rightarrow \infty$ does not allow the generalized DF-CF to attain its maximum rate.

Finally, gathering our results from cases 1, 2 and 3 and the case of $\lambda_3 > 0$, it can be seen that for all relevant instances of α_2 and γ' , the maximum rate that the generalized DF-CF scheme achieves using Gaussian signals is attained when the constraint in (6b) is satisfied with equality.

We note that another proof that uses a different approach has been obtained independently in Lemma 2 and Remark 5 of [39].

A.5 Proof of Theorem 2

To proceed with the proof of Theorem 2, we invoke the construction in Section 2.3.2 into (15) to show that when $\gamma_1 \leq \gamma_0$, $\Delta \leq 0$; i.e., in that region, R_1 is the constraining rate.

Invoking the construction in Section 2.3.2, it is straightforward to show that

$$I(X_2, U; Y) = \frac{1}{2} \log_2 \left(\frac{1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}}{1 + \alpha_2 \gamma_0} \right), \quad (70a)$$

$$I(X_2, Y|V) = \frac{1}{2} \log_2 \left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_0 + \beta_1 \gamma_2}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right), \quad (70b)$$

$$I(U, Y_1|X_2, V) = \frac{1}{2} \log_2 \left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + \alpha_2 \gamma_1} \right). \quad (70c)$$

Substituting these expressions into the expression of Δ in Lemma 3 yields

$$\Delta = \frac{1}{2} \log_2 \left(\left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right) \left(\frac{1 + \alpha_2 \gamma_0}{1 + \alpha_2 \gamma_1} \right) \left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_0 + \beta_1 \gamma_2}{1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}} \right) \right). \quad (71)$$

Noting that $\alpha_1 + \alpha_2 \leq 1$ and $\beta_1 \leq 1$, it is immediate that $\frac{1 + (\alpha_1 + \alpha_2)\gamma_0 + \beta_1 \gamma_2}{1 + \gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}} \leq 1$. Furthermore, it can be shown that, for $\gamma_1 \leq \gamma_0$, $\left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right) \left(\frac{1 + \alpha_2 \gamma_0}{1 + \alpha_2 \gamma_1} \right) \leq 1$, which implies that $\Delta \leq 0$. Using this in (15), we have that, when $\gamma_1 \leq \gamma_0$, R_1 is the constraining rate and to maximize it we substitute for R_2 from (6d) and for Δ from (71) into (15), which yields

$$R_1 = \frac{1}{2} \log_2 \left(\left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right) \left(\frac{1 + \alpha_2 \gamma_0}{1 + \alpha_2 \gamma_1} \right) (1 + \gamma_0(\alpha_1 + \alpha_2) + \gamma_2 \beta_1) \left(\frac{\gamma'}{1 + \gamma'} \right) \right). \quad (72)$$

Invoking Lemma 2 and substituting for γ' from (6b) in (A.5) yields

$$\begin{aligned} R_1 = & \frac{1}{2} \log_2 \left(\left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right) \left(\frac{1 + \alpha_2 \gamma_0}{1 + \alpha_2 \gamma_1} \right) \right) \\ & + \frac{1}{2} \log_2 \left((1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_0) \right) \\ & + \frac{1}{2} \log_2 \left(\frac{(1 + (\alpha_1 + \alpha_2)\gamma_0) + \beta_1 \gamma_2}{\beta_1 \gamma_2 (1 + \alpha_2 \gamma_0) + (1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_0)} \right). \end{aligned}$$

The first two terms are independent of $\{\beta_j\}_{j=0}^1$, and the last term can be shown to be monotonically increasing in β_1 . Hence, the optimal value of β_1 in this SNR range

is 1. Using this result, we rewrite R_1 as

$$R_1 = \frac{1}{2} \log_2 \left(\left(\frac{1 + (\alpha_1 + \alpha_2)\gamma_1}{1 + (\alpha_1 + \alpha_2)\gamma_0} \right) \left(\frac{1 + \alpha_2\gamma_0}{1 + \alpha_2\gamma_1} \right) \right) + \frac{1}{2} \log_2 \left((1 + (\alpha_1 + \alpha_2)\gamma_0 + \gamma_2) \right) \\ + \frac{1}{2} \log_2 \left(\frac{1 + \alpha_2(\gamma_0 + \gamma_1)}{\frac{\gamma_2(1 + \alpha_2\gamma_0)}{1 + (\alpha_1 + \alpha_2)\gamma_0} + (1 + \alpha_2(\gamma_0 + \gamma_1))} \right).$$

Since $\gamma_1 \leq \gamma_0$, it can be shown that the first term is upper bounded by 1 and the bound is reached when $\alpha_1 = 0$. The second term is maximized when $\alpha_1 + \alpha_2 = 1$. For the last term, the denominator is minimized when $\alpha_1 + \alpha_2 = 1$. The resulting expression is monotonically increasing in α_2 and is maximized when $\alpha_2 = 1$. Hence, the rate maximizing power partitions for $\gamma_1 \leq \gamma_0$ are $\alpha_2 = 1$ and $\beta_1 = 1$. From Proposition 1, such power partitions reduce the generalized DF-CF scheme to the CF scheme, which completes the proof of Theorem 2.

A.6 Proof of Theorem 4

To prove Theorem 4, let $\Delta_G = R_G^* - \max\{R_{\text{DF}}^*, R_{\text{CF}}^*\}$ and consider the cases of $\gamma_1 \leq \gamma_0$ and $\gamma_1 > \gamma_0$.

From Theorem 2, we have that when $\gamma_1 \leq \gamma_0$, $R_G^* = R_{\text{CF}}^* \geq R_{\text{DF}}^*$, and hence in that region, $\Delta_G = 0$.

We now consider the SNR region $\gamma_1 > \gamma_0$. From the definition of Δ_G , we can write

$$\begin{aligned} \Delta_G &\leq R_G^* - R_{\text{DF}}^* \\ &= \max_{\mathcal{S}} \min\{R_1, R_2\} - R_{\text{DF}}^* \\ &\leq \max_{\mathcal{S}} R_1 - R_{\text{DF}}^*, \end{aligned} \tag{73}$$

where \mathcal{S} is the feasible set of the problem in (6), and R_1 and R_2 are defined in (6c) and (6d), respectively. Using (9) and the fact that $\rho_i^* \in [0, 1]$, $i = 0, 1$, and $\rho_0^* + \rho_1^* = 1$,

we have

$$R_{\text{DF}}^* = \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\rho_0^* \gamma_0 \gamma_2}) = \mathcal{C}(\rho_1^* \gamma_1). \quad (74)$$

Using the monotonicity of the expression of R_2 in (6d) in γ' , we can write

$$R_2 \leq \mathcal{C}(\gamma_0 + \gamma_2 + 2\sqrt{\alpha_0 \beta_0 \gamma_0 \gamma_2}). \quad (75)$$

Since $R_G^* \geq R_{\text{DF}}^*$, a comparison of (74) and (75) reveals that $\alpha_0 \beta_0 \geq \rho_0^*$. But, since $\beta_0 \in [0, 1]$, we must have $\alpha_0 \geq \rho_0^*$. Hence, invoking $\sum_{i=0}^2 \alpha_i = \sum_{i=0}^1 \rho_i^* = 1$ yields

$$\alpha_1 + \alpha_2 \leq \rho_1^*. \quad (76)$$

We will use (76) to obtain a bound on the Δ_G . In particular, from (73) and (74), we can write

$$\begin{aligned} \Delta_G &\leq \max_{\alpha_1, \alpha_2 \in [0, 1]} \left\{ \frac{1}{2} \log_2 \left(\frac{(1 + \alpha_2(\gamma_0 + \gamma_1))(1 + (\alpha_1 + \alpha_2)\gamma_1)}{(1 + \alpha_2\gamma_1)(1 + \rho_1^*\gamma_1)} \right) \right\} \\ &\leq \max_{\alpha_2 \in [0, 1]} \left\{ \frac{1}{2} \log_2 \left(1 + \frac{\alpha_2 \gamma_0}{1 + \alpha_2 \gamma_1} \right) \right\} \end{aligned} \quad (77)$$

$$\leq \frac{1}{2} \log_2 \left(1 + \frac{\gamma_0}{1 + \gamma_1} \right) \quad (78)$$

$$< 0.5 \text{ bpcu},$$

where (77) follows from invoking (76), (78) follows from the monotonicity of the rational expression in α_2 , and the last inequality follows from using the fact that $\gamma_1 > \gamma_0$.

B.1 Proof of Theorem 5

Codebook Generation For node d_k , generate $2^{n(R_k + \tilde{R}_k)}$ independent identically distributed (i.i.d.) $\mathbf{x}_k(m_k, s_k)$, each according to distribution $p(\mathbf{x}_k) = \prod_{i=1}^n p(x_{ki})$, $m_k \in [1 : 2^{nR_k}]$, $s_k \in [1 : 2^{n\tilde{R}_k}]$; for each $\mathbf{x}_k(m_k, s_k)$, generate $2^{n\tilde{R}_k}$ i.i.d. $\hat{\mathbf{y}}_k(z_k | m_k, s_k)$, each according to distribution $p(\hat{\mathbf{y}}_k | \mathbf{x}_k) = \prod_{i=1}^n p(\hat{y}_{ki} | x_{ki})$, $z_k \in [1 : 2^{n\tilde{R}_k}]$.

Random Binning For node d_k , randomly partition the set $\{1, \dots, 2^{n\tilde{R}_k}\}$ into $2^{n\tilde{R}_k}$ bins. Let $s_k = \mathcal{B}_k(z_k)$ denote the N -to-1 mapping corresponding to random binning at d_k .

Encoding Let b be the current block. At the end of block b , node d_k

- finds z_k such that $(\hat{\mathbf{y}}_k(z_k | m_{k,b}, s_{k,b}), \mathbf{x}_k(m_{k,b}, s_{k,b}), \mathbf{y}_{d_k}(b))$ are jointly ϵ -typical.

By the covering lemma in Section 3.2 (cf. [41]), such z_k exists z_k as $n \rightarrow \infty$ if

$$\hat{R}_k \geq I(\hat{Y}_k; Y_{d_k} | X_k). \quad (79)$$

If more than one such z_k exist, choose the smallest z_k and let $z_{k,b} = z_k$;

- determines $s_k = \mathcal{B}_k(z_{k,b})$ and lets $s_{k,b+1} = s_k$;

Codewords $\mathbf{x}_k(m_{k,b}, s_{k,b})$ are sent in block b from all $d_k \in \mathcal{N}$.

Decoding Procedure We provide a layered forward decoding procedure for d_k .

Let i, j denote the decoding layer and the block number of the received signal used at layer i , respectively. Let ℓ_{d_k} be the layer at which the decoding at d_k ends.

We have

$$j = b - \ell_{d_k} + i. \quad (80)$$

Using this relationship, we drop the block number j from the expressions in the analysis of the decoding procedure when it is clear. Furthermore, in the analysis of the decoding at d_k , node identity d_k is omitted from the subscript in ℓ_{d_k} ($\ell \triangleq \ell_{d_k}$) and various sets of nodes when it is clear.

Let \mathcal{L} be the maximum number of the decoding layers in the network, i.e., $\mathcal{L} \triangleq \max_{d_k \in \mathcal{N}} \ell_{d_k}$. At any $d_k \in \mathcal{N}$,

$$1 \leq i \leq \ell \leq \mathcal{L} \leq N - 1. \quad (81)$$

Consider the receiver at d_k . Assuming that in block $b \geq \ell$, $s_{\mathcal{N}, b-\ell}$ has been uniquely recovered, the receiver at d_k

- constructs the following jointly typical set:

$$(\mathbf{x}_{\mathcal{A}_{m,i}}, \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i}}, \mathbf{y}_{d_k}(b - \ell + i)). \quad (82)$$

(The sets of nodes $\mathcal{A}_{m,i}$ and $\mathcal{A}_{z(J),i}$ for $i \leq \ell$ will be made clear below.)

- forms the set $\mathcal{A}_{m,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{x}_{d_l}(\hat{m}_{d_l})$ in the jointly ϵ -typical set in (82), and $d_l \in \mathcal{A}_{m,i}$.
- forms the set $\mathcal{A}_{z(J^c),i+1}$: which contains the node d_l , $d_l \in \mathcal{A}_{m,i+1} \cap \mathcal{A}_{z(J),j}^c$, for all $j < i$; (The set $\mathcal{A}_{z(J),j}^c$ is defined below.)
- forms the set $\mathcal{A}_{z(J),i+1}^c$: for each node d_l in this set, there exist multiple $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l} | \hat{m}_{d_l})$ in the jointly ϵ -typical set in (82), and $d_l \in \mathcal{A}_{m,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- forms the set $\mathcal{A}_{z(J),i+1}$: for each node d_l in this set, there exists a unique $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l} | \hat{m}_{d_l})$ in the jointly ϵ -typical set in (82), and $d_l \in \mathcal{A}_{m,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- proceeds to layer $i + 1$ if $\mathcal{A}_{z(J),i+1}^c \neq \emptyset$;

- ends at layer i when $\mathcal{A}_{z(J),i+1}^c = \emptyset$.

Using the jointly ϵ -typical sets in ℓ layers jointly, the receiver at d_k declares that $m_{\mathcal{S}_{d_k}} = \hat{m}_{\mathcal{S}_{d_k}}$ was sent in block $b - \ell$; and that $s_{\mathcal{N}} = \hat{s}_{\mathcal{N}}$ was sent in block $b - \ell + 1$.

Let $\mathcal{A}_{\ell_m} \triangleq \mathcal{A}_{m,\ell+1} = \mathcal{A}_{z(J),\ell+1} \cup \mathcal{A}_{z(J^c),\ell+1}$ and $\mathcal{A}_{\ell_z} \triangleq \mathcal{A}_{z(J),\ell+1}$. Furthermore, let $\mathcal{A}_{m,1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1}^c \triangleq \emptyset$ and $\mathcal{A}_{z(J^c),1} \triangleq \emptyset$. In Lemmas 5 and 6, we provide useful properties of the above sets.

Lemma 5. *By definition, the sets formed by d_k in the decoding procedure have the following properties, for $i \leq \ell$,*

1. $\mathcal{A}_{m,i+1}^c = \mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}$;
2. $\mathcal{A}_{z(J),i}$, $\mathcal{A}_{z(J),i}^c$ and $\mathcal{A}_{z(J^c),i}$ are disjoint, $\mathcal{A}_{z(J),i} \cup \mathcal{A}_{z(J),i}^c \cup \mathcal{A}_{z(J^c),i} = \mathcal{A}_{m,i}$;
3. $\mathcal{A}_{\ell_m} \subseteq \mathcal{A}_{m,i} \subseteq \mathcal{A}_{m,j}$, for $i > j$;
4. $(\mathcal{A}_{z(J),i} \cup \mathcal{A}_{z(J),i}^c) = \mathcal{A}_{z(J),i-1} \subseteq \mathcal{A}_{z(J),j}$, for $i > j$;
5. $\mathcal{A}_{z(J),i} \cap \mathcal{A}_{z(J^c),i+1} = \emptyset$;
6. $\mathcal{A}_{\ell_z} \subseteq \mathcal{A}_{z(J),i+1} \subseteq \mathcal{A}_{z(J),i}$;
7. For the receiver at $\forall d_k$, $d_k \in \mathcal{D}_{\mathcal{S}_{d_k}}$ and $d_k \in \mathcal{A}_{m,i}$, $d_k \in \mathcal{A}_{z(J),i}$. Hence, $d_k \in \mathcal{A}_{\ell_m}$, $d_k \in \mathcal{A}_{\ell_z}$.
8. $\mathcal{A}_{z(J),\ell+1}^c = \emptyset$.

Lemma 6. *For the sets formed by the receiver, define $\mathcal{A}_{z(J),i \setminus i+1} \triangleq ((\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}^c$. The following equality holds:*

$$\mathcal{A}_{z(J),i \setminus i+1} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}. \quad (83)$$

Proof. To prove the lemma, consider

$$\begin{aligned}
& \mathcal{A}_{z(J),i|i+1} \\
&= (\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}^c \\
&\stackrel{(a)}{=} (\mathcal{A}_{m,i+1}^c \setminus (\mathcal{A}_{z(J),i}^c \cup \mathcal{A}_{z(J^c),i})) \cup \mathcal{A}_{z(J),i+1}^c \\
&= (\mathcal{A}_{m,i+1}^c \setminus (\mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J),i})) \cup \mathcal{A}_{z(J),i+1}^c \\
&\stackrel{(b)}{=} (\mathcal{A}_{m,i+1}^c \cap \mathcal{A}_{z(J),i}) \cup (\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{m,i}) \cup \mathcal{A}_{z(J),i+1}^c \\
&= (\mathcal{A}_{m,i+1}^c \cap \mathcal{A}_{z(J),i}) \cup \mathcal{A}_{z(J),i+1}^c \\
&\stackrel{(c)}{=} (\mathcal{A}_{m,i+1}^c \cup \mathcal{A}_{z(J),i+1}^c) \cap (\mathcal{A}_{z(J),i} \cup \mathcal{A}_{z(J),i+1}^c) \\
&\stackrel{(d)}{=} (\mathcal{A}_{m,i+1}^c \cup \mathcal{A}_{z(J),i+1}^c) \cap \mathcal{A}_{z(J),i} \\
&\stackrel{(e)}{=} ((\mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}) \cup \mathcal{A}_{z(J),i+1}^c) \cap \mathcal{A}_{z(J),i} \\
&= (\mathcal{A}_{m,i} \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1})) \cap \mathcal{A}_{z(J),i} \\
&\stackrel{(f)}{=} (\mathcal{A}_{z(J),i} \cap \mathcal{A}_{m,i}) \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1}) \\
&= \mathcal{A}_{z(J),i} \setminus (\mathcal{A}_{z(J^c),i+1} \cup \mathcal{A}_{z(J),i+1}) \\
&= \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J^c),i+1} \setminus \mathcal{A}_{z(J),i+1} \\
&\stackrel{(g)}{=} \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}.
\end{aligned}$$

The proof uses the properties in Lemma 5. In particular, (a) follows from Property 2; (b) follows Property 1 such that $\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{m,i} = \emptyset$; (c) follows from Property 4; (d) follows from Property 1; (e) follows from Property 2 and 3; (f) follows from Property 2; (g) follows from Property 5. \square

Analysis of the Probability of Error Without loss of generality, assume that for any node $d_k \in \mathcal{N}$, $m_l = 1$, $s_l = 1$ were transmitted and $z_l = 1$ was selected in block $b - \ell, b - \ell + 1, \dots, b$.

We begin the analysis of the probability of error by providing the following lemma,

which applies to any decoding layer i at any given decoding node d .

Lemma 7. *Let \mathcal{U} and \mathcal{V} be all the nodes whose codewords $\mathbf{X}_{\mathcal{U}}$ and $\hat{\mathbf{Y}}_{\mathcal{V}}$ are considered at a given layer of the decoding procedure. Let the codewords of $\mathcal{M} \subseteq \mathcal{U}$ and $\mathcal{Z} \subseteq \mathcal{V}$ be $\mathbf{X}_{\mathcal{M}} \subseteq \mathbf{X}_{\mathcal{U}}$ and $\hat{\mathbf{Y}}_{\mathcal{Z}} \subseteq \hat{\mathbf{Y}}_{\mathcal{V}}$, respectively, where for each node in \mathcal{M} and each node in \mathcal{Z} there are multiple codewords that lie in the jointly typicality set at the considered layer. For any sets $\mathcal{G} \subseteq \mathcal{M}$ and $\mathcal{F} \subseteq \mathcal{Z}$,*

$$\begin{aligned} P((\mathbf{X}_{\mathcal{G}}((m, s) \neq (1, 1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_{\epsilon}^{(n)}) \\ \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{M}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}, \end{aligned} \quad (84)$$

and

$$P((\mathbf{x}_{\mathcal{G}}(1, 1), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_{\epsilon}^{(n)}) \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{M}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}, \quad (85)$$

where

$$\mathcal{I}_0 = I(X_{\mathcal{M}}; X_{\mathcal{U} \setminus \mathcal{M}}, \hat{Y}_{\mathcal{V} \setminus \mathcal{Z}}, Y_d) + \sum_{i \in \mathcal{Z}} I(\hat{Y}_i; X_{\mathcal{U}}, \hat{Y}_{\mathcal{V} \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i), \quad (86)$$

where $\hat{Y}_{\mathcal{K}_i} \triangleq \hat{Y}_{\{d, i' \in \mathcal{Z}: i' < i\}}$, for $i \in \mathcal{Z}$.

Proof. Using joint typicality lemma in Section 3.2 (cf. [41, Sect. 2.5.1]), the first statement in the lemma can be readily obtained.

To prove the second the statement in the lemma, consider that the probability $P(\mathbf{X}_{\mathcal{G}}((m, s) \neq (1, 1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_{\epsilon}^{(n)}$ can be upper bounded by

$$2^{n(R_{\mathcal{G}} + \check{R}_{\mathcal{G}} + \hat{R}_{\mathcal{F}} - \mathcal{I}')} \text{, where} \quad (87a)$$

$$\mathcal{I}' = I(X_{\mathcal{G}}; X_{\mathcal{U} \setminus \mathcal{M}}, \hat{Y}_{\mathcal{V} \setminus \mathcal{Z}}, Y_d) + \sum_{i \in \mathcal{F}} I(\hat{Y}_i; X_{\mathcal{U} \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{\mathcal{V} \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i). \quad (87b)$$

Note that if $R_{\mathcal{G}} + \check{R}_{\mathcal{G}} + \hat{R}_{\mathcal{F}} \leq \mathcal{I}'$, the probability $P(\mathbf{X}_{\mathcal{G}}((m, s) \neq (1, 1)), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq$

1), $\mathbf{x}_{U \setminus \mathcal{M}}, \hat{\mathbf{y}}_{V \setminus \mathcal{Z}}, \mathbf{Y}_d \in \mathcal{A}_\epsilon^{(n)}$) vanishes as $n \rightarrow 0$, which contradicts the definition of \mathcal{G} and \mathcal{F} . Hence (87a) and (87b) provide

$$R_{\mathcal{G}} + \check{R}_{\mathcal{G}} + \hat{R}_{\mathcal{F}} > \mathcal{I}'. \quad (88)$$

Now, the probability $P((\mathbf{x}_{\mathcal{G}}((1, 1)), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{U \setminus \mathcal{M}}, \hat{\mathbf{y}}_{V \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)})$ can be upper bounded by

$$P((\mathbf{x}_{\mathcal{G}}((1, 1)), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{U \setminus \mathcal{M}}, \hat{\mathbf{y}}_{V \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \leq 2^{n(R_{\mathcal{M} \setminus \mathcal{G}} + \check{R}_{\mathcal{M} \setminus \mathcal{G}} + \hat{R}_{\mathcal{Z} \setminus \mathcal{F}} - \mathcal{I}_1)}, \quad (89)$$

where

$$\begin{aligned} \mathcal{I}_1 &= I(X_{\mathcal{M} \setminus \mathcal{G}}; X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z} \cup \mathcal{F}}, Y_d) + \sum_{i \in \mathcal{Z} \setminus \mathcal{F}} I(\hat{Y}_i; X_U, \hat{Y}_{V \setminus \mathcal{Z} \cup \mathcal{F}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) \\ &= I(X_{\mathcal{M}}; X_{U \setminus \mathcal{M}}, \hat{Y}_{V \setminus \mathcal{Z}}, Y_d) - I(X_{\mathcal{G}}; \hat{Y}_{V \setminus \mathcal{Z}}, Y_d | X_{U \setminus \mathcal{M}}) + I(X_{\mathcal{M} \setminus \mathcal{G}}; \hat{Y}_{\mathcal{F}} | X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z}}, Y_d) \\ &\quad + \sum_{i \in \mathcal{Z}} I(\hat{Y}_i; X_U, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) - \sum_{i \in \mathcal{F}} I(\hat{Y}_i; X_U, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) \\ &= I(X_{\mathcal{M}}; X_{U \setminus \mathcal{M}}, \hat{Y}_{V \setminus \mathcal{Z}}, Y_d) + \sum_{i \in \mathcal{Z}} I(\hat{Y}_i; X_U, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) - I(X_{\mathcal{G}}; \hat{Y}_{V \setminus \mathcal{Z}}, Y_d | X_{U \setminus \mathcal{M}}) \\ &\quad + I(X_{\mathcal{M} \setminus \mathcal{G}}; \hat{Y}_{\mathcal{F}} | X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z}}, Y_d) - \sum_{i \in \mathcal{F}} I(\hat{Y}_i; X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) \\ &\quad - \sum_{i \in \mathcal{F}} I(\hat{Y}_i; X_{\mathcal{M} \setminus \mathcal{G}} | X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d) \\ &= I(X_{\mathcal{M}}; X_{U \setminus \mathcal{M}}, \hat{Y}_{V \setminus \mathcal{Z}}, Y_d) + \sum_{i \in \mathcal{Z}} I(\hat{Y}_i; X_U, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) \\ &\quad - I(X_{\mathcal{G}}; \hat{Y}_{V \setminus \mathcal{Z}}, Y_d | X_{U \setminus \mathcal{M}}) - \sum_{i \in \mathcal{F}} I(\hat{Y}_i; X_{U \setminus \mathcal{M} \cup \mathcal{G}}, \hat{Y}_{V \setminus \mathcal{Z}}, \hat{Y}_{\mathcal{K}_i}, Y_d | X_i) \\ &= I_0 - \mathcal{I}' \end{aligned} \quad (90)$$

$$\geq I_0 - R_{\mathcal{G}} - \check{R}_{\mathcal{G}} - R_{\mathcal{F}}, \quad (91)$$

where (90) follows from (86) and (87b), and (91) follows from (88). Substituting (91) in (89) yields the result of the second statement in the lemma. which completes the proof of the lemma. \square

We note that by definition, the following relationship holds between the sets in Lemma 7 and the sets defined in the decoding procedure:

$$\mathcal{U} = \mathcal{A}_{m,i}, \quad \mathcal{V} = \mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J^c),i+1}, \quad (92a)$$

$$\mathcal{M} = \mathcal{A}_{m,i+1}^c, \quad \mathcal{Z} = \mathcal{A}_{z(J),i+1}^c. \quad (92b)$$

Using Properties 4 and 5 in Lemma 5, set \mathcal{Z} can also be written as

$$\mathcal{Z} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J^c),i+1}. \quad (93)$$

Furthermore, using Property 2 in Lemma 5, we have

$$\mathcal{V} \setminus \mathcal{Z} = (\mathcal{A}_{m,i} \setminus \mathcal{A}_{z(J^c),i+1}) \setminus \mathcal{A}_{z(J),i+1}^c = \mathcal{A}_{z(J),i+1}. \quad (94)$$

Now we analyze the probability of error. Define the following events for layer i in the decoding procedure:

$$\begin{aligned} \mathcal{E}_{i,1} &= \{(\mathbf{x}_{\mathcal{A}_{m,i}}(1, 1), \hat{\mathbf{y}}_{z(J),i}(1|1, 1), \mathbf{Y}_{d_k}(b-\ell+i)) \notin \mathcal{A}_\epsilon^{(n)}\}; \\ \mathcal{E}_{i,2} &= \{(\mathbf{X}_{\mathcal{A}_{m,i+1}^c}(\hat{m}, \hat{s}), \mathbf{x}_{\mathcal{A}_{m,i+1}}(1, 1), \\ &\quad \hat{\mathbf{Y}}_{\mathcal{A}_{z(J),i+1}^c}(\hat{z}|1, 1), \hat{\mathbf{Y}}_{(\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}(\hat{z}|\hat{m}, \hat{s}), \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i+1}}(1|1, 1), \\ &\quad \mathbf{Y}_{d_k}(b-\ell+i)) \in \mathcal{A}_\epsilon^{(n)}, \text{ for some } \hat{m}, \hat{z}, \hat{s} \in \mathcal{B}(\hat{z})\}. \end{aligned}$$

The receiver at d_k makes an error if any event in $\mathcal{E} \triangleq (\cup_i \mathcal{E}_{i,1}) \cup (\cap_i \mathcal{E}_{i,2})$ occurs for some $\hat{m} \neq 1, \hat{s} \neq 1$. Using the union bound, the probability of error is given by

$P(\mathcal{E}) = P((\cup_i \mathcal{E}_{i,1}) \cup (\cap_i \mathcal{E}_{i,2})) = P(\cup_i \mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2}) \leq \sum_i P(\mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2})$. By the conditional typicality lemma in Section 3.2 (cf. [41, Sect. 2.5]), $P(\mathcal{E}_{i,1}) \rightarrow 0$ as $n \rightarrow \infty$. Now we upper bound $P(\cap_i \mathcal{E}_{i,2})$. Consider the probability $P(\mathcal{E}_{i,2})$ for the case that $(\hat{m}, \hat{s}) \neq (1, 1)$ for $\mathbf{x}_l(\hat{m}, \hat{s}), d_l \in \mathcal{A}_{m,i+1}^c$, which can be bounded by

$$P(\mathcal{E}_{i,2}) \leq \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{m,i}}, z_{\mathcal{A}_{z,i}}} 2^{-n\beta_i}, \text{ where} \quad (95a)$$

$$\beta_i = \sum_{j=1}^3 \beta_{i,j}, \text{ and} \quad (95b)$$

$$\begin{aligned} \beta_{i,1} &= I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}), \\ \beta_{i,2} &= \sum_{i \in (\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i), \\ \beta_{i,3} &= \sum_{i \in \mathcal{A}_{z(J),i+1}^c} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{((\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i). \end{aligned}$$

Note that using chain rule and Lemma 6, we can rewrite β_i in (95b) as

$$\beta_i = I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}) + \sum_{i \in \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i),$$

which can be shown, by substituting the sets using (92), (93) and (94), to have same the form as (86) in Lemma 7.

For node $d_l \in \mathcal{G} \subseteq \mathcal{M} \triangleq \mathcal{A}_{m,i+1}^c$ and $d_k \in \mathcal{F} \subseteq \mathcal{Z} \triangleq \mathcal{A}_{z(J),i+1}^c$, multiple $\mathbf{x}_l(\hat{m}, \hat{s})$ and $\hat{\mathbf{y}}_k(\hat{z}|\hat{m}, \hat{s})$ are found in the joint typicality set, respectively, at layer i . For the case $(\hat{m}, \hat{s}) \neq (1, 1)$ and $\hat{z} \neq 1$, the probability of the event is given by $P(\mathcal{E}_{i,2})$, for which an upper bounded is provided in (95). For the case $(\hat{m}, \hat{s}) = (1, 1)$ and the case $\hat{z} = 1$, substituting sets in Lemma 7 using (92) provides that the probability of these cases can also be upper bounded by (95).

Hence, we can upper bound the probability $P(\cap_i \mathcal{E}_{i,2})$ as

$$\begin{aligned}
& P\left(\bigcap_{i=1}^{\ell} \mathcal{E}_{i,2}\right) \\
& \leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{m,i}}, z_{\mathcal{A}_{z,i}}} P(\mathcal{E}_{i,2}) \\
& \leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i+1}^c}, s_{\mathcal{A}_{m,i+1}^c}, z_{\mathcal{A}_{z(J),i+1}}} 2^{-n\beta_i} \\
& \leq \prod_{i=1}^{\ell-1} \prod_{l \in m_{\mathcal{A}_{m,i+1}^c}} 2^{nR_l} \prod_{l \in \mathcal{A}_{z(J),i+1}} 2^{n\hat{R}_l} \cdot 2^{-n \sum_{j=1}^3 \beta_{i,j}} \\
& \quad \prod_{l \in m_{\mathcal{A}_{\ell_m}^c}} 2^{nR_l} \prod_{l \in \mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}} 2^{-n\check{R}_l} \prod_{l \in \mathcal{A}_{z(J),\ell-1} \setminus \ell} 2^{n\hat{R}_l} \cdot 2^{-n \sum_{j=1}^3 \beta_{i,j}} \\
& = 2^{nR_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}} \cdot 2^{-n\check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}}} \cdot 2^{n \sum_{i=1}^{\ell} (\hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} - \sum_{j=1}^3 \beta_{i,j})}. \tag{96}
\end{aligned}$$

Using chain rule and the definition of \mathcal{A}_{ℓ_z} , we can expand $\beta_{i,1}$ as

$$\begin{aligned}
\beta_{i,1} & = I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}) + I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J),\ell+1}} | X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}) \\
& = I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}) + I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \tag{97}
\end{aligned}$$

We define

$$\begin{aligned}
\beta_{i,1,1} & = I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}), \\
\beta_{i,1,2} & = I(X_{\mathcal{A}_{m,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).
\end{aligned}$$

Now, consider the summation of $\beta_{i,1,1}$ from two consecutive decoding layers ℓ and

$\ell - 1$:

$$\begin{aligned}
& (X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}) + I(X_{\mathcal{A}_{m,\ell}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,\ell-1} \setminus \mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}})
\end{aligned}$$

where the first equality follows the chain rule and in the last equality we have used

$$(\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}) \cup \mathcal{A}_{m,\ell}^c = \mathcal{A}_{m,\ell-1} \setminus \mathcal{A}_{m,\ell+1}$$

by Property 1 and 3 in Lemma 5.

Using the same technique iteratively, it can be shown that

$$\sum_{i=\hat{i}}^{\ell} \beta_{i,1,1} = I(X_{\mathcal{A}_{m,\hat{i}} \setminus \mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}).$$

Hence, $\sum_{i=1}^{\ell} \beta_{i,1,1}$ yields

$$\begin{aligned}
\sum_{i=1}^{\ell} \beta_{i,1,1} &= I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{m,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}) \\
&= I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}). \tag{98}
\end{aligned}$$

Next, consider the summation of $\beta_{i,2}$ and $\beta_{i,3}$ in (96) for all possible \hat{z} . We define

$\beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}}$, where

$$\begin{aligned}
& \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \\
&= \sum_{i \in (\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} (I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z,i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i) - \hat{R}_i) \\
&\quad + \sum_{i \in \mathcal{A}_{z,i+1}^c} (-\hat{R}_i + I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}), \hat{Y}_{((\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}} | X_i)) \\
&\leq \sum_{i \in (\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} -I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}) \\
&\quad - \sum_{i \in \mathcal{A}_{z,i+1}^c} I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \hat{Y}_{((\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}})) \\
&= -I(\hat{Y}_{(\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}; Y_{(\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} | X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k})) \\
&\quad - I(\hat{Y}_{\mathcal{A}_{z(J),i+1}^c}; Y_{\mathcal{A}_{z(J),i+1}^c} | X_{\mathcal{A}_{m,i}}, \hat{Y}_{((\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}}, Y_{d_k})) \\
&= -I(\hat{Y}_{(\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}^c}; Y_{(\mathcal{A}_{m,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}, Y_{\mathcal{A}_{z(J),i+1}^c} | X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k})) \\
&= -I(\hat{Y}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}}; Y_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} | X_{\mathcal{A}_{m,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k})),
\end{aligned}$$

where in the first inequality we have used (79), the second and third equality follow the chain rule, and in the last equality we have used (83) from Lemma 6.

By definition, $\mathcal{A}_{z(J),\ell+1} = \mathcal{A}_{\ell_z}$, hence $\mathcal{A}_{z(J),\ell+1} \setminus \mathcal{A}_{\ell_z} = \emptyset$ and $\beta_{\ell,1,2} = 0$. Now, for

layer ℓ and $\ell - 1$, we calculate

$$\begin{aligned}
& \beta_{\ell,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{z(J),\ell+1}} + \beta_{\ell-1,1,2} + \beta_{\ell-1,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} \\
& \leq -I(\hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + I(X_{\mathcal{A}_{m,\ell}^c}; \hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& \quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
& = -I(\hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& \quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
& = -I(\hat{Y}_{(\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}) \cup (\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z})}; Y_{(\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}) \cup (\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z})} | X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& = -I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}), \tag{99}
\end{aligned}$$

where the first and second equality follow the chain rule, and in the last equality we have used

$$(\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}) \cup (\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}) = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{\ell_z} \tag{100}$$

for $\forall i \leq \ell$, which is a direct consequence from Property 6 in Lemma 5.

Using this technique iteratively, we obtain

$$\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \leq -I(\hat{Y}_{\mathcal{A}_{z(J),\tilde{i}} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\tilde{i}} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\tilde{i}}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).$$

Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \leq -I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \tag{101}$$

Using (98) and (101), the probability in (96) can be upper bounded by

$$\begin{aligned} P(\cap_i \mathcal{E}_{i,2}) &\leq 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}})} \cdot 2^{-n(\sum_{i=1}^{\ell} \beta_{i,1,1} + \sum_{i=1}^{\ell} (\beta_{i,1,2} + \beta_{i,2U3}))} \\ &= 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \mathcal{I}(\mathcal{A}))}, \end{aligned}$$

where $\mathcal{I}(\mathcal{A})$ is given by

$$I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}) - I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}}.$$

Using $\mathcal{A}_{m,1} = \mathcal{A}_{z(J),1}$ by definition, $\mathcal{I}(\mathcal{A})$ can be written as

$$I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}) - I(\hat{Y}_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}}.$$

Let $\mathcal{S} \triangleq \mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}$, $\mathcal{S}^c \triangleq \mathcal{A}_{\ell_m}$, hence $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Let $\check{\mathcal{S}} \triangleq \mathcal{A}_{\ell_m} \setminus \mathcal{A}_{\ell_z}$, $\check{\mathcal{S}}^c \triangleq \mathcal{A}_{\ell_z}$, hence $\check{\mathcal{S}}^c = \mathcal{A}_{\ell_m} \setminus \check{\mathcal{S}}$. The above result yields that $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$, if

$$R_{\mathcal{S}} \leq I(X_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}.$$

Using the technique similar to the one in [10], it can be shown that for $\mathcal{S} \cap \mathcal{S}_{d_k} = \emptyset$, the constraints in the above inequality can be dropped and $d_k \in \mathcal{D}_{\mathcal{S}}$. By Property 7 in Lemma 5, $d_k \in \mathcal{A}_{\ell_m} = \mathcal{S}^c$. Hence $d_k \in \mathcal{S}^c \cap \mathcal{D}_{\mathcal{S}}$. This completes the proof.

B.2 Proof of Remark 6

To prove the result in Remark 6, first we note that when

$$\check{R}_{\check{\mathcal{S}}} \geq I(X_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) - R_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}, \quad (102)$$

the probability of error tends to 0 as $n \rightarrow \infty$ if the rate expression in (26) satisfies

$$\begin{aligned} R_{\mathcal{N} \setminus \check{\mathcal{S}}^c} &\leq I(X_{\mathcal{S} \cup (\mathcal{S}^c \setminus \check{\mathcal{S}}^c)}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\check{\mathcal{S}}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where we have used

$$\mathcal{S} \cup (\mathcal{S}^c \setminus \check{\mathcal{S}}^c) = \mathcal{N} \setminus \check{\mathcal{S}}^c$$

since $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c$ and $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Redefine $\mathcal{S} \triangleq \mathcal{N} \setminus \check{\mathcal{S}}^c$ and $\mathcal{S}^c \triangleq \check{\mathcal{S}}^c$, we have

$$R_{\mathcal{S}} \leq I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}),$$

which is the simplified form of the rate expression (28).

Next, in general the rate expression in (26) can be modified to the following form:

$$\begin{aligned} R_{\mathcal{S}} &\leq I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &= I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) + I(Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} - I(\hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - (\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} - I(\hat{Y}_{\check{\mathcal{S}}}^c; Y_{\check{\mathcal{S}}}^c | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}). \end{aligned} \tag{103}$$

where the first two steps follow the chain rule; in the last equality we have also used chain rule and $\mathcal{S}^c \setminus \check{\mathcal{S}}^c = \check{\mathcal{S}}$ and $(\mathcal{N} \setminus \check{\mathcal{S}}^c) \setminus (\mathcal{S}^c \setminus \check{\mathcal{S}}^c) = \mathcal{S}$, all by definition.

It can be seen that if

$$\check{R}_{\check{\mathcal{S}}} \geq I(\hat{Y}_{\check{\mathcal{S}}}^c; Y_{\check{\mathcal{S}}}^c | X_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \tag{104}$$

the rate expression in (103) reduces to its simplified form (28).

B.3 Proof of Theorem 6

Codebook Generation For node d_k , generate $2^{n\hat{R}_k}$ i.i.d. $\mathbf{u}_k(s_k)$, each according to distribution $p(\mathbf{u}_k) = \prod_{i=1}^n p(u_{ki})$, $s_k \in [1 : 2^{n\hat{R}_k}]$; for each $\mathbf{u}_k(s_k)$, generate 2^{nR_k} i.i.d. $\mathbf{x}_k(m_k|s_k)$, each according to distribution $p(\mathbf{x}_k|\mathbf{u}_k) = \prod_{i=1}^n p(x_{ki}|s_{ki})$, $m_k = [1 : 2^{nR_k}]$; for each $\mathbf{u}_k(s_k)$, generate $2^{n\hat{R}_k}$ i.i.d. $\hat{\mathbf{y}}_k(z_k|s_k)$, each according to distribution $p(\hat{\mathbf{y}}_k|\mathbf{u}_k) = \prod_{i=1}^n p(\hat{y}_{ki}|s_{ki})$, $z_k \in [1 : 2^{n\hat{R}_k}]$.

Random Binning For node d_k , randomly partition the set $\{1, \dots, 2^{n\hat{R}_k}\}$ into $2^{n\hat{R}_k}$ bins. Let $s_k = \mathcal{B}_k(z_k)$ denote the N -to-1 mapping at d_k as the result of binning.

Encoding Let b be the current block. At the end of block b , node d_k

- finds an index z_k such that $(\hat{\mathbf{y}}_k(z_k|s_{k,b}), \mathbf{u}_k(s_{k,b}), \mathbf{y}_{d_k}(b))$ are jointly typical. By the covering lemma in Section 3.2 (cf. [41]), such a z_k exists as $n \rightarrow \infty$ if

$$\hat{R}_k \geq I(\hat{Y}_k; Y_{d_k} | U_k). \quad (105)$$

If more than one such z_k exist, choose the lowest z_k and let $z_{k,b} = z_k$;

- determines $s_k = \mathcal{B}_k(z_{k,b})$ and lets $s_{k,b+1} = s_k$;

Codewords $\mathbf{x}_k(m_{k,b}|s_{k,b})$ are sent in block b from all $d_k \in \mathcal{N}$.

Decoding Procedure Similar to the procedure provided in the proof of Theorem 5, the procedure herein employs the layered forward decoding strategy and the relationships in (80) and (81) hold. The main difference between the two procedures lies in the sets formed at the receiver at each decoding layer.

Consider the receiver at d_k . Assuming that in block b , $s_{\mathcal{N},b-\ell}$ has been uniquely recovered, where $\ell \triangleq \ell_{d_k}$, the receiver at d_k

- constructs the following jointly ϵ -typical set:

$$(\mathbf{x}_{\mathcal{A}_{m,i}}, \mathbf{u}_{\mathcal{A}_{s,i}}, \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i}}, \mathbf{y}_{d_k}(b - \ell + i)). \quad (106)$$

(The definition of the sets of nodes $\mathcal{A}_{m,i}$, $\mathcal{A}_{s,i}$ and $\mathcal{A}_{z(J),i}$ for $i \leq \ell$ is given below.)

- forms the set $\mathcal{A}_{s,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{u}_{d_l}(\hat{s}_{d_l})$ in the jointly ϵ -typical set in (106), and $d_l \in \mathcal{A}_{s,i}$;
- forms the set $\mathcal{A}_{m,i+1}$: for each node d_l in this set, there exists a unique $\mathbf{x}_{d_l}(\hat{m}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ -typical set in (106), and $d_l \in \mathcal{A}_{m,i} \cap \mathcal{A}_{s,i+1}$;
- forms set $\mathcal{A}_{z(J^c),i+1}$, which contains the node $d_l \in \mathcal{A}_{s,i+1} \cap \mathcal{A}_{z(J),j}^c$, for all $j < i$;
(The set $\mathcal{A}_{z(J),j}^c$ is defined below.)
- forms the set $\mathcal{A}_{z(J),i+1}^c$: for each node d_l in this set, there exist multiple $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ -typical set in (106), and $d_l \in \mathcal{A}_{s,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- forms the set $\mathcal{A}_{z(J),i+1}$: for each node d_l in this set, there exists a unique $\hat{\mathbf{y}}_{d_l}(\hat{z}_{d_l}|\hat{s}_{d_l})$ in the jointly ϵ -typical set in (106), and $d_l \in \mathcal{A}_{s,i+1} \setminus \mathcal{A}_{z(J^c),i+1}$;
- proceeds to layer $i + 1$ if $\mathcal{A}_{z(J),i+1}^c \neq \emptyset$.
- ends at layer i when $\mathcal{A}_{z(J),i+1} = \emptyset$.

Using the jointly ϵ -typical sets in ℓ layers jointly, the receiver at d_k declares that $m_{\mathcal{S}_{d_k}} = \hat{m}_{\mathcal{S}_{d_k}}$ were sent in block $b - \ell$; and that $s_{\mathcal{N}} = \hat{s}_{\mathcal{N}}$ were sent in block $b - \ell + 1$.

Let $\mathcal{A}_{\ell_m} \triangleq \mathcal{A}_{m,\ell+1}$, $\mathcal{A}_{\ell_s} \triangleq \mathcal{A}_{s,\ell+1} = \mathcal{A}_{z(J),\ell+1} \cup \mathcal{A}_{z(J^c),\ell+1}$ and $\mathcal{A}_{\ell_z} \triangleq \mathcal{A}_{z(J),\ell+1}$. Furthermore, let $\mathcal{A}_{m,1} \triangleq \mathcal{S}_{d_k}$, $\mathcal{A}_{s,1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1} \triangleq \mathcal{N}$, $\mathcal{A}_{z(J),1}^c \triangleq \emptyset$ and $\mathcal{A}_{z(J^c),1} \triangleq \emptyset$. In Lemmas 8 and 9 we provide the counterparts of Lemmas 5 and 6 for the decoding procedure of Theorem 6.

Lemma 8. *The sets formed by the receiver at $d_k \in \mathcal{N}$ in the decoding procedure have the following properties, for $i \leq \ell$,*

1. $\mathcal{A}_{m,i+1}^c = \mathcal{A}_{m,i} \setminus \mathcal{A}_{m,i+1}$, $\mathcal{A}_{s,i+1}^c = \mathcal{A}_{s,i} \setminus \mathcal{A}_{s,i+1}$;
2. $\mathcal{A}_{z(J),i}$, $\mathcal{A}_{z(J),i}^c$ and $\mathcal{A}_{z(J^c),i}$ are disjoint, $\mathcal{A}_{z(J),i} \cup \mathcal{A}_{z(J),i}^c \cup \mathcal{A}_{z(J^c),i} = \mathcal{A}_{s,i}$;
3. $\mathcal{A}_{\ell_m} \subseteq \mathcal{A}_{m,i} \subseteq \mathcal{A}_{m,j}$, $\mathcal{A}_{\ell_s} \subseteq \mathcal{A}_{s,i} \subseteq \mathcal{A}_{s,j}$, for $i > j$;
4. $(\mathcal{A}_{z(J),i} \cup \mathcal{A}_{z(J),i}^c) = \mathcal{A}_{z(J),i-1} \subseteq \mathcal{A}_{z(J),j}$, for $i > j$;
5. $\mathcal{A}_{z(J),i} \cap \mathcal{A}_{z(J^c),i+1} = \emptyset$;
6. $\mathcal{A}_{\ell_z} \subseteq \mathcal{A}_{z(J),i+1} \subseteq \mathcal{A}_{z(J),i}$;
7. For the receiver at $\forall d_k$, $d_k \in \mathcal{D}_{\mathcal{S}_{d_k}}$ and $d_k \in \mathcal{A}_{m,i}$, $d_k \in \mathcal{A}_{z(J),i}$. Hence, $d_k \in \mathcal{A}_{\ell_m}$, $d_k \in \mathcal{A}_{\ell_z}$.
8. $\mathcal{A}_{z(J),\ell+1}^c = \emptyset$;
9. $\mathcal{A}_{z(J),i} \subseteq \mathcal{A}_{s,i}$, $\mathcal{A}_{m,i} \subseteq \mathcal{A}_{s,i}$.

Lemma 9. *For the sets formed by the receiver, define $\mathcal{A}_{z(J),i \setminus i+1} \triangleq ((\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}^c$. The following equality holds:*

$$\mathcal{A}_{z(J),i \setminus i+1} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}. \quad (107)$$

Proof. Replacing $\mathcal{A}_{m,i+1}^c$ and $\mathcal{A}_{m,i}$ by $\mathcal{A}_{s,i+1}^c$ and $\mathcal{A}_{s,i}$, respectively, the lemma can be proved using a technique similar to the one in the proof of Lemma 6. Details are omitted. \square

Analysis of the Probability of Error Without loss of generality, assume that for any node $d_k \in \mathcal{N}$, $m_l = 1$, $s_l = 1$ were transmitted and $z_l = 1$ was selected in block $b - \ell, b - \ell + 1, \dots, b$.

We begin the analysis of the probability of error by providing the following lemma, which is a counterpart of Lemma 7 and applies to any decoding layer at any decoding node d .

Lemma 10. *Let \mathcal{U}, \mathcal{W} and \mathcal{V} be all the nodes whose codewords $\mathbf{X}_{\mathcal{U}}, \mathbf{U}_{\mathcal{W}}$ and $\hat{\mathbf{Y}}_{\mathcal{V}}$ are considered at a given layer of the decoding procedure. Let the codewords of $\mathcal{M} \subseteq \mathcal{U}, \mathcal{J} \subseteq \mathcal{W}$ and $\mathcal{Z} \subseteq \mathcal{V}$ be $\mathbf{X}_{\mathcal{M}} \subseteq \mathbf{X}_{\mathcal{U}}, \mathbf{U}_{\mathcal{J}} \subseteq \mathbf{U}_{\mathcal{W}}$ and $\hat{\mathbf{Y}}_{\mathcal{Z}} \subseteq \hat{\mathbf{Y}}_{\mathcal{V}}$, respectively, where for each node in \mathcal{M} , each node in \mathcal{W} and each node in \mathcal{Z} there are multiple codewords that lie in the jointly typicality set at the considered layer. For any sets $\mathcal{G} \subseteq \mathcal{M}, \mathcal{H} \subseteq \mathcal{J}$ and $\mathcal{F} \subseteq \mathcal{Z}$,*

$$P((\mathbf{X}_{\mathcal{G}}(m \neq 1), \mathbf{U}_{\mathcal{H}}(s \neq 1), \hat{\mathbf{Y}}_{\mathcal{F}}(z \neq 1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \mathbf{U}_{\mathcal{J} \setminus \mathcal{H}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \mathbf{u}_{\mathcal{W} \setminus \mathcal{J}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{J}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}; \quad (108)$$

and

$$P((\mathbf{x}_{\mathcal{G}}(1), \mathbf{u}_{\mathcal{H}}(1), \hat{\mathbf{y}}_{\mathcal{F}}(1), \mathbf{X}_{\mathcal{M} \setminus \mathcal{G}}, \mathbf{U}_{\mathcal{J} \setminus \mathcal{H}}, \hat{\mathbf{Y}}_{\mathcal{Z} \setminus \mathcal{F}}, \mathbf{x}_{\mathcal{U} \setminus \mathcal{M}}, \mathbf{u}_{\mathcal{W} \setminus \mathcal{J}}, \hat{\mathbf{y}}_{\mathcal{V} \setminus \mathcal{Z}}, \mathbf{Y}_d) \in \mathcal{A}_\epsilon^{(n)}) \leq 2^{n(R_{\mathcal{M}} + \check{R}_{\mathcal{J}} + \hat{R}_{\mathcal{Z}} - \mathcal{I}_0)}, \quad (109)$$

where

$$\mathcal{I}_0 = I(X_{\mathcal{M}}, U_{\mathcal{J}}; \hat{\mathbf{Y}}_{\mathcal{V} \setminus \mathcal{Z}}, Y_d | X_{\mathcal{U} \setminus \mathcal{M}}, U_{\mathcal{W} \setminus \mathcal{J}}) + \sum_{i \in \mathcal{Z}} I(\hat{\mathbf{Y}}_i; X_{\mathcal{U}}, U_{\mathcal{W}}, \hat{\mathbf{Y}}_{\mathcal{V} \setminus \mathcal{Z}}, \hat{\mathbf{Y}}_{\mathcal{K}_i}, Y_d | U_i), \quad (110)$$

Proof. Using the following substitution in Lemma 7,

$$(X_{\mathcal{M}}, U_{\mathcal{J}}) \triangleq X_{\mathcal{M}} \text{ and } (X_{\mathcal{U}}, U_{\mathcal{W}}) \triangleq X_{\mathcal{U}},$$

the result in Lemma 10 can be readily obtained. \square

We note that by definition, the following relationship holds between the sets in Lemma 10 and the sets defined in the decoding procedure for Theorem 6:

$$\mathcal{U} = \mathcal{A}_{m,i}, \quad \mathcal{W} = \mathcal{A}_{s,i}, \quad \mathcal{V} = \mathcal{A}_{s,i} \setminus \mathcal{A}_{z(J^c),i+1}, \quad (111a)$$

$$\mathcal{M} = \mathcal{A}_{m,i+1}^c, \quad \mathcal{J} = \mathcal{A}_{s,i+1}^c, \quad \mathcal{Z} = \mathcal{A}_{z(J),i+1}^c. \quad (111b)$$

Using Properties 4 and 5 in Lemma 5, set \mathcal{Z} can also be written as

$$\mathcal{Z} = \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J^c),i+1}. \quad (112)$$

Furthermore, using Property 2 in Lemma 5, we have

$$\mathcal{V} \setminus \mathcal{Z} = (\mathcal{A}_{s,i} \setminus \mathcal{A}_{z(J^c),i+1}) \setminus \mathcal{A}_{z(J),i+1}^c = \mathcal{A}_{z(J),i+1}. \quad (113)$$

Now we analyze the probability of error. Define the following events for layer i in the decoding procedure:

$$\begin{aligned} \mathcal{E}_{i,1} &= \{(\mathbf{x}_{\mathcal{A}_{m,i}}(1|1), \mathbf{u}_{\mathcal{A}_{s,i}}(1), \hat{\mathbf{y}}_{z(J),i}(1|1), \mathbf{Y}_{d_k}(b-\ell+i)) \notin \mathcal{A}_\epsilon^{(n)}\}; \\ \mathcal{E}_{i,2} &= \{(\mathbf{U}_{\mathcal{A}_{s,i+1}^c}(\hat{s}), \mathbf{u}_{\mathcal{A}_{s,i+1}}(1), \\ &\quad \mathbf{X}_{\mathcal{A}_{m,i+1}^c}(\hat{m}|\hat{s}), \mathbf{X}_{\mathcal{A}_{m,i+1}^c}(\hat{m}|1), \mathbf{x}_{\mathcal{A}_{m,i+1}}(1|1), \\ &\quad \hat{\mathbf{Y}}_{\mathcal{A}_{z(J),i+1}^c}(\hat{z}|1), \hat{\mathbf{Y}}_{(\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}(\hat{z}|\hat{s}), \hat{\mathbf{y}}_{\mathcal{A}_{z(J),i+1}}(1|1), \\ &\quad \mathbf{Y}_{d_k}(b-\ell+i)) \in \mathcal{A}_\epsilon^{(n)}, \\ &\quad \text{for some } \hat{m}, \hat{z} \text{ and } \hat{s} \in \mathcal{B}(\hat{z})\}. \end{aligned}$$

The receiver at d_k makes an error if any event in $\mathcal{E} \triangleq (\cup_i \mathcal{E}_{i,1}) \cup (\cap_i \mathcal{E}_{i,2})$ for some $\hat{m} \neq 1, \hat{s} \neq 1$ occurs. Using the union bound, the probability of error is given by $P(\mathcal{E}) = P((\cup_i \mathcal{E}_{i,1}) \cup (\cap_i \mathcal{E}_{i,2})) = P(\cup_i \mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2}) \leq \sum_i P(\mathcal{E}_{i,1}) + P(\cap_i \mathcal{E}_{i,2})$. By

the conditional typicality lemma in Section 3.2 (cf. [41, Sect. 2.5]), $P(\mathcal{E}_{i,1}) \rightarrow 0$ as $n \rightarrow \infty$. Now we upper bound $P(\cap_i \mathcal{E}_{i,2})$. Consider the probability $P(\mathcal{E}_{i,2})$ for the case that $\hat{m} \neq 1$ for $\mathbf{x}_l(\hat{m}|\hat{s}), d_l \in \mathcal{A}_{m,i+1}^c$ and $\hat{s} \neq 1$ for $\mathbf{u}_l(\hat{s}), d_l \in \mathcal{A}_{s,i+1}^c$, which can be bounded by

$$P(\mathcal{E}_{i,2}) \leq \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{s,i}}, z_{\mathcal{A}_{z,i}}} 2^{-n\beta_i}, \text{ where} \quad (114a)$$

$$\beta_i = \sum_{j=1}^3 \beta_{i,j}, \text{ and} \quad (114b)$$

$$\begin{aligned} \beta_{i,1} &= I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}), \\ \beta_{i,2} &= \sum_{i \in (\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i), \\ \beta_{i,3} &= \sum_{i \in \mathcal{A}_{z(J),i+1}^c} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{((\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i). \end{aligned}$$

Note that using chain rule and Lemma 6, we can rewrite β_i in (114b) as

$$\begin{aligned} \beta_i &= I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\ &\quad + \sum_{i \in \mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | X_i), \end{aligned}$$

which can be shown, by substituting the sets using (111), (112) and (113), to have same the form as (110) in Lemma 10.

For node $d_l \in \mathcal{G} \subseteq \mathcal{M} \triangleq \mathcal{A}_{m,i+1}^c, d_h \in \mathcal{H} \subseteq \mathcal{J} \triangleq \mathcal{A}_{s,i+1}^c$ and $d_k \in \mathcal{F} \subseteq \mathcal{Z} \triangleq \mathcal{A}_{z(J),i+1}^c$, multiple $\mathbf{x}_l(\hat{m}|\hat{s}), \mathbf{u}(\hat{s})$ and $\hat{\mathbf{y}}_k(\hat{z}|\hat{s})$ are found in the joint typicality set, respectively, at layer i . For the case $\hat{m} \neq 1, \hat{s} \neq 1$ and $\hat{z} \neq 1$, the probability of the event is given by $P(\mathcal{E}_{i,2})$, for which an upper bounded is provided in (114). For the case $\hat{m} = 1, \hat{s} = 1$ and the case $\hat{z} = 1$, substituting sets in Lemma 10 using (111) provides that the probability of these cases can also be upper bounded

by (95).

Now consider the probability of event $\mathcal{E}_{i,2}$ over the ℓ layers:

$$\begin{aligned}
& P\left(\bigcap_{i=1}^{\ell} \mathcal{E}_{i,2}\right) \\
& \leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i}}, s_{\mathcal{A}_{s,i}}, z_{\mathcal{A}_{z,i}}} P(\mathcal{E}_{i,2}) \\
& \leq \prod_{i=1}^{\ell} \sum_{m_{\mathcal{A}_{m,i+1}^c}, s_{\mathcal{A}_{s,i+1}^c}, z_{\mathcal{A}_{z(J),i\setminus i+1}}} 2^{-n\beta_i} \\
& \leq \prod_{i=1}^{\ell-1} \prod_{l \in m_{\mathcal{A}_{m,i+1}^c}} 2^{nR_l} \prod_{l \in \mathcal{A}_{z(J),i\setminus i+1}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^3 \beta_{i,j}} \\
& \quad \prod_{l \in m_{\mathcal{A}_{\ell_m}^c}} 2^{nR_l} \prod_{l \in \mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}} 2^{-n\check{R}_l} \prod_{l \in \mathcal{A}_{z(J),\ell-1\setminus \ell}} 2^{n\hat{R}_l} \cdot 2^{-n\sum_{j=1}^3 \beta_{i,j}} \\
& = 2^{nR_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}} \cdot 2^{-n\check{R}_{\mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}}} \\
& \quad \cdot 2^{n\sum_{i=1}^{\ell} (\hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} - \sum_{j=1}^3 \beta_{i,j})}. \tag{115}
\end{aligned}$$

Using chain rule and the definition of \mathcal{A}_{ℓ_z} , we can expand $\beta_{i,1}$ as

$$\begin{aligned}
& \beta_{i,1} \\
& = I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\
& \quad + I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{z(J),\ell+1}} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell+1}}) \\
& = I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\
& \quad + I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \tag{116}
\end{aligned}$$

We define

$$\begin{aligned}\beta_{i,1,1} &= I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}) \\ \beta_{i,1,2} &= I(X_{\mathcal{A}_{m,i+1}^c}, U_{\mathcal{A}_{s,i+1}^c}; \hat{Y}_{\mathcal{A}_{z(J),i+1} \setminus \mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,i+1}}, U_{\mathcal{A}_{s,i+1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).\end{aligned}$$

Now, consider the summation of $\beta_{i,1,1}$ from two consecutive decoding layers ℓ and $\ell - 1$:

$$\begin{aligned}& (X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell} \setminus \mathcal{A}_{s,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\ & \quad + I(X_{\mathcal{A}_{m,\ell}^c}, U_{\mathcal{A}_{s,\ell}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell} \setminus \mathcal{A}_{s,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\ &= I(X_{\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}}, X_{\mathcal{A}_{m,\ell}^c}, U_{\mathcal{A}_{s,\ell} \setminus \mathcal{A}_{s,\ell+1}}, U_{\mathcal{A}_{s,\ell}^c}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}}) \\ &= I(X_{\mathcal{A}_{m,\ell-1} \setminus \mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell-1} \setminus \mathcal{A}_{s,\ell+1}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{m,\ell+1}}, U_{\mathcal{A}_{s,\ell+1}})\end{aligned}$$

where the first equality follows the chain rule and in the last equality we have used

$$\begin{aligned}(\mathcal{A}_{m,\ell} \setminus \mathcal{A}_{m,\ell+1}) \cup \mathcal{A}_{m,\ell}^c &= \mathcal{A}_{m,\ell-1} \setminus \mathcal{A}_{m,\ell+1}, \\ (\mathcal{A}_{s,\ell} \setminus \mathcal{A}_{s,\ell+1}) \cup \mathcal{A}_{s,\ell}^c &= \mathcal{A}_{s,\ell-1} \setminus \mathcal{A}_{s,\ell+1}\end{aligned}$$

by Property 1 and 3 in Lemma 8.

Using this technique iteratively, we obtain

$$\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,1} = I(X_{\mathcal{A}_{m,\tilde{i}} \setminus \mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,\tilde{i}} \setminus \mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}).$$

Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,1} = I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}). \quad (117)$$

Next, consider the summation of $\beta_{i,2}$ and $\beta_{i,3}$ in (115) for all possible \hat{z} . We define

$$\beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}},$$

where

$$\begin{aligned} & \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \\ &= \sum_{i \in (\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z,i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k} | U_i) - \hat{R}_i \\ & \quad + \sum_{i \in \mathcal{A}_{z,i+1}^c} -\hat{R}_i + I(\hat{Y}_i; X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \hat{Y}_{((\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}} | U_i) \\ &\leq \sum_{i \in (\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} -I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}) \\ & \quad - \sum_{i \in \mathcal{A}_{z,i+1}^c} I(\hat{Y}_i; Y_i | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{K}_i}, Y_{d_k}, \hat{Y}_{((\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}}) \\ &= -I(\hat{Y}_{(\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}; Y_{(\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\ & \quad - I(\hat{Y}_{\mathcal{A}_{z(J),i+1}^c}; Y_{\mathcal{A}_{z(J),i+1}^c} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{((\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}) \cup \mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\ &= -I(\hat{Y}_{(\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}; \hat{Y}_{\mathcal{A}_{z(J),i+1}^c}; Y_{(\mathcal{A}_{s,i+1}^c \setminus \mathcal{A}_{z(J),i}^c) \setminus \mathcal{A}_{z(J^c),i}}, Y_{\mathcal{A}_{z(J),i+1}^c} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}) \\ &= -I(\hat{Y}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}}; Y_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} | X_{\mathcal{A}_{m,i}}, U_{\mathcal{A}_{s,i}}, \hat{Y}_{\mathcal{A}_{z(J),i+1}}, Y_{d_k}), \end{aligned}$$

where in the first inequality we have used (105), the second and third equality follow the chain rule, and in the last equality we have used (107) from Lemma 9.

By definition, $\mathcal{A}_{z(J),\ell+1} = \mathcal{A}_{\ell_z}$, hence $\mathcal{A}_{z(J),\ell+1} \setminus \mathcal{A}_{\ell_z} = \emptyset$ and $\beta_{\ell,1,2} = 0$. Now, for

layer ℓ and $\ell - 1$, we calculate

$$\begin{aligned}
& \beta_{\ell,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{z(J),\ell+1}} + \beta_{\ell-1,1,2} + \beta_{\ell-1,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} \\
& \leq -I(\hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, U_{\mathcal{A}_{s,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& \quad + I(X_{\mathcal{A}_{m,\ell}^c}, U_{\mathcal{A}_{s,\ell}^c}; \hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell}}, U_{\mathcal{A}_{s,\ell}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& \quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
& = -I(\hat{Y}_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& \quad - I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{z(J),\ell}}, Y_{d_k}) \\
& = -I(\hat{Y}_{(\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}) \cup (\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z})}; Y_{(\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{z(J),\ell}) \cup (\mathcal{A}_{z(J),\ell} \setminus \mathcal{A}_{\ell_z})} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) \\
& = -I(\hat{Y}_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),\ell-1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\ell-1}}, U_{\mathcal{A}_{s,\ell-1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}), \tag{118}
\end{aligned}$$

where the first and second equality follow the chain rule, and in the last equality we have used a property similar to (100), which is a direct consequence of Property 6 of Lemma 8.

Using this technique iteratively, we obtain

$$\sum_{i=\tilde{i}}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} \leq -I(\hat{Y}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,\tilde{i}}}, U_{\mathcal{A}_{s,\tilde{i}}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}).$$

Consider the summation over all ℓ layers, we have

$$\sum_{i=1}^{\ell} \beta_{i,1,2} + \beta_{i,2\cup 3} - \hat{R}_{\mathcal{A}_{z(J),i} \setminus \mathcal{A}_{z(J),i+1}} = -I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}). \tag{119}$$

Using (117) and (119), the probability in (115) can be upper bounded by

$$P(\cap_i \mathcal{E}_{i,2}) \leq 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \check{R}_{\mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}})} \cdot 2^{-n(\sum_{i=1}^{\ell} \beta_{i,1,1} + \sum_{i=1}^{\ell} (\beta_{i,1,2} + \beta_{i,2\cup 3}))} = 2^{n(R_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}} - \mathcal{I}(\mathcal{A}))},$$

where $\mathcal{I}(\mathcal{A})$ is given by

$$\begin{aligned} & I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}) \\ & - I(\hat{Y}_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{z(J),1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}}. \end{aligned}$$

Using $\mathcal{A}_{s,1} = \mathcal{A}_{z(J),1}$ by definition, $\mathcal{I}(\mathcal{A})$ can be written as

$$\begin{aligned} & I(X_{\mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_s}}; \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k} | X_{\mathcal{A}_{\ell_m}}, U_{\mathcal{A}_{\ell_s}}) \\ & - I(\hat{Y}_{\mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_z}}; Y_{\mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_z}} | X_{\mathcal{A}_{m,1}}, U_{\mathcal{A}_{s,1}}, \hat{Y}_{\mathcal{A}_{\ell_z}}, Y_{d_k}) + \check{R}_{\mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}}. \end{aligned}$$

Let $\mathcal{T} \triangleq \mathcal{A}_{m,1} \setminus \mathcal{A}_{\ell_m}$, $\mathcal{T}^c \triangleq \mathcal{A}_{\ell_m}$. Let $\mathcal{S} \triangleq \mathcal{A}_{s,1} \setminus \mathcal{A}_{\ell_s}$, $\mathcal{S}^c \triangleq \mathcal{A}_{\ell_s}$, hence $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Let $\check{\mathcal{S}} \triangleq \mathcal{A}_{\ell_s} \setminus \mathcal{A}_{\ell_z}$, $\check{\mathcal{S}}^c \triangleq \mathcal{A}_{\ell_z}$, hence $\check{\mathcal{S}}^c = \mathcal{A}_{\ell_s} \setminus \check{\mathcal{S}}$. The above result yields that $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$, if

$$R_{\mathcal{T}} \leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}}.$$

Assume that the constraint is violated for $\mathcal{T} = \emptyset$, we have

$$I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) > I(U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{S}^c}) + \check{R}_{\check{\mathcal{S}}}. \quad (120)$$

In this case, the constraint on $R_{\mathcal{T}}$ can be bounded by:

$$\begin{aligned} R_{\mathcal{T}} & \leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ & = I(X_{\mathcal{T}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) + I(U_{\mathcal{S}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{S}^c}) \\ & \quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ & < I(X_{\mathcal{T}}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}), \end{aligned}$$

where the last inequality follows from (120). This implies that the receiver can treat

the signal from the nodes in $\mathcal{S} \setminus \mathcal{T}$ as noise. Hence, the constraint for the case $\mathcal{T} = \emptyset$ can be dropped. This completes the proof.

B.4 Proof of Remark 9

To prove the result in Remark 9, first we note that when

$$\check{R}_{\check{\mathcal{S}}} \geq I(U_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\check{\mathcal{S}}^c}), \quad (121)$$

the rate expression in (30) yields

$$\begin{aligned} R_{\mathcal{T}} &\leq I(X_{\mathcal{T}}, U_{\mathcal{S} \cup \mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\check{\mathcal{S}}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{T}}, U_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\check{\mathcal{S}}^c}) - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where we have used $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c$ and $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. Redefine $\mathcal{S} \triangleq \mathcal{N} \setminus \check{\mathcal{S}}^c$ and $\mathcal{S}^c \triangleq \check{\mathcal{S}}^c$ yields the simplified form of the rate expression (31).

Next, in general, the rate expression in (30) can be modified to the following form:

$$\begin{aligned} R_{\mathcal{T}} &\leq I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) - I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &= I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}) + I(Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; \hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &\quad - I(\hat{Y}_{\mathcal{N} \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{N} \setminus \check{\mathcal{S}}^c} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) + \check{R}_{\check{\mathcal{S}}} \\ &\quad - I(\hat{Y}_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c}; Y_{\mathcal{S}^c \setminus \check{\mathcal{S}}^c} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}) \\ &= I(X_{\mathcal{T}}, U_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{d_k} | X_{\mathcal{S}^c}) - (\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{S}_{d_k}}, U_{\mathcal{N}}, \hat{Y}_{\mathcal{S}^c}, Y_{d_k}) \\ &\quad + \check{R}_{\check{\mathcal{S}}} - I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \end{aligned}$$

where in the last equality we have used $\check{\mathcal{S}}^c = \mathcal{S}^c \setminus \check{\mathcal{S}}$ and $(\mathcal{N} \setminus \check{\mathcal{S}}^c) \setminus (\mathcal{S}^c \setminus \check{\mathcal{S}}^c) = \mathcal{S}$ due

to $\check{\mathcal{S}}^c \subseteq \mathcal{S}^c \subseteq \mathcal{N}$ and $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$. It can be seen that if

$$\check{R}_{\check{\mathcal{S}}} \geq I(\hat{Y}_{\check{\mathcal{S}}}; Y_{\check{\mathcal{S}}} | X_{\mathcal{T}^c}, U_{\mathcal{S}^c}, \hat{Y}_{\check{\mathcal{S}}^c}, Y_{d_k}), \quad (122)$$

the rate expression reduces to its simplified form (28).

Conditions (121) and (122) together provide the desired result.

C.1 Proof of Corollary 2

Codebook generation and the encoding procedure incorporates those in Theorem 5 and the standard DF [2]. The detailed procedures are provided herein for completeness.

Codebook Generation

- Generate 2^{nR_1} i.i.d $\mathbf{x}_1(m_1)$, each according to distribution $p(\mathbf{x}_1) = \prod_{i=1}^n p(x_{1i})$, $m_1 \in [1 : 2^{nR_1}]$.
- Generate $2^{n\check{R}_3}$ i.i.d $\mathbf{x}_3(s_3)$, each according to distribution $p(\mathbf{x}_3) = \prod_{i=1}^n p(x_{3i})$, $s_3 \in [1 : 2^{n\check{R}_3}]$.
- For each $\mathbf{x}_3(s_3)$, generate $2^{n\check{R}_2}$ i.i.d $\mathbf{x}_2(s_2)$, each according to distribution $p(\mathbf{x}_2|\mathbf{x}_3) = \prod_{i=1}^n p(x_{2i}|x_{3i})$, $s_2 \in [1 : 2^{n\check{R}_2}]$.
- For each $(\mathbf{x}_2(s_2), \mathbf{x}_3(s_3))$ pair, generate $2^{n\hat{R}_2}$ i.i.d $\hat{\mathbf{y}}_2(z_2)$, each according to distribution $p(\hat{\mathbf{y}}_2|\mathbf{x}_2, \mathbf{x}_3) = \prod_{i=1}^n p(\hat{y}_{2i}|x_{2i}, x_{3i})$, $z_2 \in [1 : 2^{n\hat{R}_2}]$.

Random Binning

- Randomly partition the set $\{1, 2, \dots, 2^{n\hat{R}_2}\}$ into $2^{n\check{R}_2}$ bins. Let $s_2 = \mathcal{B}_2(z_2)$ denote the N -to-1 mapping as the results of binning.
- Randomly partition the set $\{1, 2, \dots, 2^{n\check{R}_2}\}$ into $2^{n\check{R}_3}$ bins. Let $s_3 = \mathcal{B}_3(s_2)$ denote the N -to-1 mapping as the results of binning.

Encoding In block j ,

- source S encodes $\mathbf{x}_1(s_{1,j+1})$;

- relay R_1 finds an index z such that $(\mathbf{x}_2(s_j), \mathbf{x}_3(s_j), \hat{\mathbf{y}}_2(z|s_{2,j}, s_{3,j}), \mathbf{y}_2(j))$ are jointly typical. Such a z exists as $n \rightarrow \infty$ if

$$\hat{R}_2 \geq I(\hat{Y}_2; Y_2 | X_2, X_3). \quad (123)$$

If there exist more than one such z , choose the smallest one and let $z_{2,j} = z$;

- relay R_1 finds the bin index $s_{2,j+1} = \mathcal{B}_2(z_{2,j})$ and $s_{3,j+2} = \mathcal{B}_2(s_{2,j+1})$.
- relay R_2 finds a unique index s_2 such that $(\mathbf{x}_2(s_2), \mathbf{x}_3(s_{3,j}), \mathbf{y}_3(j))$ are jointly ϵ -typical. The probability of error tends to 0 as $n \rightarrow \infty$ if

$$\check{R}_2 \leq I(X_2; Y_3 | X_3), \quad (124)$$

and obtains $s_{3,j+1} = \mathcal{B}_2(s_2)$;

- codewords $\mathbf{x}_1(m_{1,j})$, $\mathbf{x}_2(s_{2,j})$ and $\mathbf{x}_3(s_{3,j})$ are sent into the channel.

Decoding and Probability of Error Now, we upper bound the probability of erroneous decoding at the destinations and analyze the constraints on the rate of bin indices \check{R}_2 and \check{R}_3 .

- Using Strategy 1, the decoding procedure at the destinations $D_i, i = 1, 2$, partially follows the one in Theorem 5. In addition, at each decoding layer, the codeword transmitted by the relay R_2 that represents the bin index of s_2 , i.e., $\mathbf{x}_3(s_3)$ where $s_3 = \mathcal{B}_2(s_2)$ for each s_2 , must be jointly typical with the received signal at the next layer. Hence, using Theorem 5 and (123), when the following constraints are satisfied, the probability of error tends

to 0 as $n \rightarrow \infty$.

$$R_1 \leq I(X_1; \hat{Y}_2, Y_{D_i} | X_2, X_3), \quad (125a)$$

$$R_1 \leq I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) + \check{R}_2, \quad (125b)$$

$$R_1 \leq I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) + I(X_2; Y_{D_i} | X_3) + \check{R}_3, \quad (125c)$$

$$R_1 \leq I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) + I(X_2, X_3; Y_{D_i}). \quad (125d)$$

Now we simplify these constraints. First, it can be seen that when

$$\check{R}_3 \geq \check{R}_2 - \min_{i=1,2} I(X_2; Y_{D_i} | X_3), \quad (126)$$

the constraint (125c) is more relaxed than (125b). We note that random binning imposes $\check{R}_3 \leq \check{R}_2$, which is satisfied under condition (126). Hence (125c) can be dropped without inducing additional constraint.

Next, we consider the constraint (125b) in two cases.

* case 1: $I(\hat{Y}_2; Y_2 | X_2, X_3, Y_{D_i}) \leq I(X_2; Y_3 | X_3)$. In this case, choosing $I(\hat{Y}_2; Y_2 | X_2, X_3, Y_{D_i}) \leq \check{R}_2 \leq I(X_2; Y_3 | X_3)$ renders the constraint (125b) to be more relaxed than (125a).

* case 2: $I(\hat{Y}_2; Y_2 | X_2, X_3, Y_{D_i}) > I(X_2; Y_3 | X_3)$. In this case, choosing $\check{R}_2 \leq I(X_2; Y_3 | X_3) < I(\hat{Y}_2; Y_2 | X_2, X_3, Y_{D_i})$ renders the constraint (125b) to be more relaxed than

$$R_1 \leq I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) + I(X_2; Y_3 | X_3). \quad (127)$$

Constraint (125d) and (127) provide

$$R_1 \leq I(X_1; Y_{D_i} | X_2, X_3) - I(\hat{Y}_2; Y_2 | X_1, X_2, X_3, Y_{D_i}) \\ + \min\{I(X_2; Y_3 | X_3), I(X_2, X_3; Y_{D_i})\}. \quad (128)$$

In both cases, constraint (124) and the condition $\check{R}_2 \leq \hat{R}_2$ which is imposed by random binning are satisfied. Therefore, constraint (125b) can be dropped, and (125a) and (128) yields (33) in the corollary.

- Using Strategy 2, the 1-to-1 mapping implies $\hat{R}_2 = \check{R}_2$, and hence (123) and (124) impose the following constraint:

$$I(\hat{Y}_2; Y_2 | X_2, X_3) \leq \hat{R}_2 = \check{R}_2 \leq I(X_2; Y_3 | X_3). \quad (129)$$

However, for the case $\check{R}_2 = \hat{R}_2$, constraint (125b) is more relaxed than (125a). Furthermore, constraint (125c) can be dropped when selecting

$$\check{R}_3 \geq \max_{i=1,2} \min\{I(X_3; Y_{D_i}), \check{R}_2 - I(\check{Y}_2; Y_{D_i} | X_2, X_3, Y_{D_i}) - I(X_2; Y_{D_i} | X_3)\}. \quad (130)$$

Constraints (125a), (125d) and (129) yield (34).

- Using Strategy 3, the recovery of s_2 at the destination D_i implies that

$$\check{R}_2 \leq \min_{i=1,2} \{I(X_2; Y_{D_i} | X_3) + \min\{\check{R}_3, I(X_3; Y_{D_i})\}\}, \quad (131)$$

in which the rate \check{R}_3 can be chosen as

$$\check{R}_3 \geq I(X_3; Y_{D_i}), \quad (132)$$

such that the right hand side of (131) is maximized, and we obtain

$$\check{R}_2 \leq \min_{i=1,2} I(X_2, X_3; Y_{D_i}). \quad (133)$$

If (124) is not binding, it can be seen from (132) and (133) that $\check{R}_3 \leq \check{R}_2$ can be satisfied. On the other hand, if (124) is binding, we can choose $\check{R}_3 \leq \check{R}_2 = I(X_2; Y_3|X_3)$. In both cases, the constraint $\check{R}_3 \leq \check{R}_2$ imposed by random binning can be satisfied and we have

$$\check{R}_2 \leq \min\{I(X_2; Y_3|X_3), \min_{i=1,2}\{I(X_2, X_3; Y_{D_i})\}\} \quad (134)$$

Next, the destinations use s_2 to recover m_1 . The probability of erroneous decoding tends to 0 as $n \rightarrow \infty$ if

$$R_1 \leq I(X_1; \hat{Y}_2, Y_{D_i}|X_2, X_3) \quad (135)$$

$$R_1 \leq I(X_1; Y_{D_i}|X_2, X_3) - I(\hat{Y}_2; Y_2|X_1, X_2, X_3, Y_{D_i}) + \check{R}_2. \quad (136)$$

Hence, when choosing

$$\check{R}_2 = \min\{I(\hat{Y}_2; Y_2|X_2, X_3, Y_{D_i}), \min\{I(X_2; Y_3|X_3), \min_{i=1,2}\{I(X_2, X_3; Y_{D_i})\}\}\},$$

which satisfies the constraint $\check{R}_2 \leq \hat{R}_2$ imposed by random binning. The constraints (135) and (136) yield (35).

This completes the proof of the corollary.

D.1 Analysis of Probability of Error for Theorem 7

Throughout the proof we will assume, without loss of generality, that $N_1 \leq N_2$.

Codebook Generation Generate 2^{nR} i.i.d. $\mathbf{x}(m)$ following $p(\mathbf{x}) = \prod_{i=1}^n p(x_i)$, $m \in [1, 2^{nR}]$. Generate $2^{nI(X_e; Y_{s,1})}$ i.i.d. $\mathbf{x}_e(s)$ following $p(\mathbf{x}_e) = \prod_{i=1}^n p(x_{ei})$, and $s \in [1, 2^{nI(X_e; Y_{s,1})}]$. For each $\mathbf{x}_e(s)$, generate $2^{n\hat{R}_e}$ i.i.d. $\hat{\mathbf{y}}_e(z|s)$ following $p(\hat{\mathbf{y}}_e|\mathbf{x}_e) = \prod_{i=1}^n p(\hat{y}_{ei}|x_{ei})$, $z \in [1, 2^{n\hat{R}_e}]$.

Random Binning The set $\{1, \dots, 2^{n\hat{R}_e}\}$ is randomly binned in $2^{nI(X_e; Y_{s,1})}$ cells. Denote the mapping by $s = \mathcal{B}(z)$.

Encoding In block b , the eavesdropper finds an index z_b such that $(\mathbf{x}_e(s_b), \hat{\mathbf{y}}_e(z_b|s_b), \mathbf{y}_e(b)) \in \mathcal{A}_\epsilon^{(n)}$. From the covering lemma in Section 3.2 (cf. [41]), such z_b exists if n is sufficiently large and

$$\hat{R}_e \geq I(\hat{Y}_e; Y_e|X_e). \quad (137)$$

If more than one z is found, choose the smallest z and let $s_{b+1} = \mathcal{B}(z_b)$. Index m_b and s_b are transmitted by the transmitter and the eavesdropper, respectively.

Decoding Assume that at the end of block b , receiver 1 and receiver 2 have correctly decoded m_{b-1} and s_{b-1} .

1. Decoding s_b : Receiver i , $i = 1, 2$, does two steps:

(a) The receiver determines two sets, $\mathcal{S}_z^{(b-1)}$ and $\mathcal{S}_s^{(b)}$:

- The set $\mathcal{S}_z^{(b-1)}$ contains the indices \hat{z} for which $(\mathbf{x}(m_{b-1}), \mathbf{x}_e(s_{b-1}), \hat{\mathbf{y}}_e(\hat{z}|s_{b-1}), \mathbf{y}_{s,i}(b-1), \mathbf{y}_i(b-1)) \in \mathcal{A}_\epsilon^{(n)}$.
- The receiver determines the set $\mathcal{S}_s^{(b)}$ which contains the indices $\hat{s} = \mathcal{B}(\hat{z})$ for each $\hat{z} \in \mathcal{S}_z^{(b-1)}$.

(b) The receiver declares that $s_b = \hat{s}$ was sent in block b if there exists a unique index $\hat{s} \in \mathcal{S}_s^{(b)}$ such that $(\mathbf{x}(\hat{m}), \mathbf{x}_e(\hat{s}), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_e^{(n)}$ for some \hat{m} .

2. Recovering m_b : Using the index s_b obtained in Step 1, receiver i constructs a set $\mathcal{S}'_{z,i} = \{\hat{z} | (\mathbf{x}_e(s_b), \hat{\mathbf{y}}_e(\hat{z}|s_b), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_e^{(n)}, (\mathbf{x}_e(\hat{s}), \mathbf{y}_{s,i}(b+1), \mathbf{y}_i(b+1)) \in \mathcal{A}_e^{(n)}, \hat{s} = \mathcal{B}(\hat{z})\}$. The receiver declares that $m_b = \hat{m}$ was sent in block b if, for some $\hat{z} \in \mathcal{S}'_{z,i}$, there is a unique \hat{m} such that $(\mathbf{x}(\hat{m}), \hat{\mathbf{y}}_e(\hat{z}|s_b), \mathbf{x}_e(s_b), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_e^{(n)}$.

Next we analyze the probability of error.

Without loss of generality, assume that the index pair $(m, s) = (1, 1)$ is transmitted in block b and block $b+1$. We define the following error events for the recovery of s_b .

$$\begin{aligned} \mathcal{E}_{s,i} &= \{(\mathbf{x}(1), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(1|1), \mathbf{y}_{s,i}(b-1), \mathbf{y}_i(b-1)) \notin \mathcal{A}_e^{(n)} \\ &\quad \cup (\mathbf{x}(1), \mathbf{x}_e(1), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \notin \mathcal{A}_e^{(n)}, \text{ for some } z \neq 1\}; \\ \mathcal{E}_{s,2} &= \{(\mathbf{x}(1), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(z|1), \mathbf{y}_{s,i}(b-1), \mathbf{y}_i(b-1)) \in \mathcal{A}_e^{(n)} \\ &\quad \cap (\mathbf{x}(1), \mathbf{x}_e(s), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_e^{(n)}, \text{ for some } z \neq 1, s = \mathcal{B}(z) \neq 1\}; \\ \mathcal{E}_{s,3} &= \{(\mathbf{x}(1), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(z|1), \mathbf{y}_{s,i}(b-1), \mathbf{y}_i(b-1)) \in \mathcal{A}_e^{(n)} \\ &\quad \cap (\mathbf{x}(m), \mathbf{x}_e(s), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_e^{(n)}, \text{ for some } z \neq 1, s = \mathcal{B}(z) \neq 1, m \neq 1\}. \end{aligned}$$

The receiver makes an error if any events in $\mathcal{E}_s = \cup_{j=1}^3 \mathcal{E}_{s,j}$ occurs. Using the union bound, we have $P(\mathcal{E}_s) = P(\cup_{j=1}^3 \mathcal{E}_{s,j}) \leq \sum_{j=1}^3 P(\mathcal{E}_{s,j})$.

Let $\hat{Y}_e = J + Z'$, where $Z' \sim \mathcal{N}(0, N')$. Define $\gamma' = \frac{N'}{P_J}$. Using $Y_e = J$ and $Y_i = a_i X + b_i J$, we have

$$I(\hat{Y}_e; Y_e | X, X_e, Y_i, Y_{s,i}) = h(Z') - h(Z') = 0. \quad (138)$$

Since (138) holds for any value of γ' , it can be arbitrarily chosen.

Now we upper bound $P(\mathcal{E}_{s,j}), j = 1, 2, 3$. By the conditional joint typicality lemma Section 3.2 (cf. [41, Sect. 2.5]), $P(\mathcal{E}_{s,1}) \rightarrow 0$ as $n \rightarrow \infty$.

The probability of $\mathcal{E}_{s,2}$ can be upper bounded by

$$P(\mathcal{E}_{s,2}) \leq 2^{n(\hat{R} - I(\hat{Y}_e; X, Y_i, Y_{s,i} | X_e) - I(X_e; Y_i, Y_{s,i} | X))}.$$

Because of (138), we have $I(X_e; Y_i, Y_{s,i} | X) \geq I(\hat{Y}_i; Y_e | X, X_e, Y_i, Y_{s,i})$. Hence using (137), $P(\mathcal{E}_{s,2}) \rightarrow 0$ as $n \rightarrow \infty$.

The probability of $\mathcal{E}_{s,3}$ can be upper bounded by

$$P(\mathcal{E}_{s,3}) \leq 2^{n(R + \hat{R} - I(\hat{Y}_e; X, Y_i, Y_{s,i} | X_e) - I(X, X_e; Y_i, Y_{s,i}))}.$$

Using (137), we have $P(\mathcal{E}_{s,3}) \rightarrow 0$ as $n \rightarrow \infty$ if

$$R \leq I(X, X_e; Y_i, Y_{s,i}) - I(\hat{Y}_e; Y_e | X, X_e, Y_i, Y_{s,i}),$$

which using (138) yields

$$R \leq I(X, X_e; Y_i, Y_{s,i}) = I(X; Y_i) + I(X_e; Y_{s,i}). \quad (139)$$

Thus, when (139) is satisfied, $P(\mathcal{E}_s)$ tends to 0 as $n \rightarrow \infty$.

To analyze the probability of error for the recovery of m_b , we define the following error events for $i = 1, 2$, respectively:

$$\begin{aligned}\mathcal{E}_{m,1} &= \{(\mathbf{x}(1), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(1|1), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \notin \mathcal{A}_\epsilon^{(n)} \\ &\quad \cup (\mathbf{x}_e(1), \mathbf{y}_{s,i}(b+1), \mathbf{y}_i(b+1)) \notin \mathcal{A}_\epsilon^{(n)}\}; \\ \mathcal{E}_{m,2} &= \{(\mathbf{x}(m), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(1|1), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_\epsilon^{(n)}\}; \\ \mathcal{E}_{m,3} &= \{(\mathbf{x}(m), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(z|1), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_\epsilon^{(n)} \\ &\quad \cap (\mathbf{x}_e(1), \mathbf{y}_{s,i}(b+1), \mathbf{y}_i(b+1)) \in \mathcal{A}_\epsilon^{(n)}, \text{ for } m \neq 1, z \neq 1\}; \\ \mathcal{E}_{m,4} &= \{(\mathbf{x}(m), \mathbf{x}_e(1), \hat{\mathbf{y}}_e(z|1), \mathbf{y}_{s,i}(b), \mathbf{y}_i(b)) \in \mathcal{A}_\epsilon^{(n)} \\ &\quad \cap (\mathbf{x}_e(s), \mathbf{y}_{s,i}(b+1), \mathbf{y}_i(b+1)) \in \mathcal{A}_\epsilon^{(n)}, \text{ for } m \neq 1, z \neq 1, s = \mathcal{B}(z) \neq 1\}.\end{aligned}$$

The receiver makes an error if any events in $\mathcal{E}_m = \cup_{j=1}^4 \mathcal{E}_{m,j}$ occurs. Hence, $P(\mathcal{E}_m) \leq \sum_{j=1}^4 P(\mathcal{E}_{m,j})$.

Now we bound $P(\mathcal{E}_{m,j})$, $j = 1, 2, 3, 4$. By the conditional joint typicality lemma in Section 3.2 (cf. [41, Sect. 2.5]), $P(\mathcal{E}_{m,1}) \rightarrow 0$ as $n \rightarrow \infty$.

For $\mathcal{E}_{m,2}$, we have $P(\mathcal{E}_{m,2}) \leq 2^{n(R-I(X;\hat{Y}_e, Y_i, Y_{s,i}|X_e))}$. Hence, $P(\mathcal{E}_{m,2}) \rightarrow 0$ as $n \rightarrow \infty$ if for $i = 1, 2$,

$$R \leq I(X; \hat{Y}_e, Y_i, Y_{s,i}|X_e). \quad (140)$$

The probability of $\mathcal{E}_{m,3}$ can be bounded by

$$P(\mathcal{E}_{m,3}) \leq 2^{n(R+\hat{R}_e-I(X;Y_{s,1})-I(X;Y_{s,i}, Y_i|X_e)-I(\hat{Y}_e; X, Y_{s,i}, Y_i|X_e))}.$$

Using (137), yields $P(\mathcal{E}_{m,3}) \rightarrow 0$ as $n \rightarrow \infty$ if

$$R \leq I(X; Y_{s,i}, Y_i|X_e) - I(\hat{Y}_e; Y_e|X, X_e, Y_{s,i}, Y_i) + I(X; Y_{s,1}).$$

Using (138), the latter condition is satisfied for $i = 1, 2$, when

$$R \leq I(X; Y_{s,i}, Y_i | X_e) + I(X; Y_{s,1}) = I(X; Y_i) + I(X_e; Y_{s,1}). \quad (141)$$

The probability of $\mathcal{E}_{m,4}$ can be bounded by

$$P(\mathcal{E}_{m,4}) \leq 2^{n(R + \hat{R}_e - I(X; Y_{s,i}, Y_i | X_e) - I(\hat{Y}_e; X, Y_i, Y_{s,i} | X_e) - I(X_e; Y_i, Y_{s,i}))}.$$

Using (137), we have $P(\mathcal{E}_{m,4}) \rightarrow 0$ as $n \rightarrow \infty$ if

$$R \leq I(X, X_e; Y_i, Y_{s,i}) - I(\hat{Y}_e; Y_e | X, X_e, Y_i, Y_{s,i}).$$

Using (138), the latter condition yields the same constraint as (139).

Now we analyze the constraints in (139) and (141). We note that $I(X_e; Y_{s,1}) \geq I(X_e; Y_{s,2})$ since $N_1 \leq N_2$. Hence, the constraint in (139) is tighter than that in (141) for $i = 2$. Using this observation, the constraints in (141) can be dropped.

Using Gaussian codebooks (139) yields $R \leq \mathcal{C}(\gamma_i) + \mathcal{C}(\gamma_{e,i})$.

Next, consider (140). We have

$$R \leq I(X; \hat{Y}_e, Y_i, Y_{s,i} | X_e) = \mathcal{C}(\gamma_i(1 + 1/\gamma')). \quad (142)$$

It can be shown that, when $\gamma' \geq 0$ such that

$$\gamma' \leq \min_{i=1,2} \frac{\gamma_i}{(1 + \gamma_i)\gamma_{e,i}}$$

is satisfied, the right hand side of (142) is larger than the smaller of the two arguments of the minimization of (49). This choice of γ' renders (142) redundant and completes the proof.

D.2 Proof of the Achievable Rate in Section 5.4.3

Let the signal transmitted by eavesdropper be denoted by cJ , where c is the gain of the amplifier. Hence,

$$c^2 = P_e/P_J. \quad (143)$$

At receiver i , the received signal from the eavesdropper can be expressed as $Y_{s,i} = cJ + Z_i$. Receiver i linearly combines the received signal Y_i and $Y_{s,i}$ to recover the message from the transmitter. The combined signal can be expressed as $Y_i + \alpha Y_{s,i} = a_i X + b_i J - \alpha(cJ + Z_i)$, where α is the combining weight to be optimized. The maximum rate that can be achieved by AF is given by

$$R_{\text{AF}} = \min_{i=1,2} \max_{\alpha} \mathcal{C}\left(\frac{a_i^2 P}{P_{J,Z_i}}\right), \quad (144)$$

where $P_{J,Z_i} = (b_i - \alpha c)^2 P_J + \alpha^2 N_i$ is the jamming and noise power.

Optimizing α yields

$$P_{J,Z_i}^* = \frac{b_i^2 P_J N_i}{c^2 P_J + N_i}, \quad i = 1, 2.$$

Using this result and (143) in (144) yields

$$R_{\text{AF}} \leq \mathcal{C}\left(\frac{a_i^2 P(P_e + N)}{b_i^2 P_J N_i}\right) = \mathcal{C}(\gamma_i(1 + \gamma_{e,i})), \quad i = 1, 2,$$

which completes the proof.

List of References

- [1] E. C. van der Meulen, “Three-terminal communication channels,” *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.
- [2] T. M. Cover and A. A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 572–584, Sept. 1979.
- [3] A. Nosratinia and A. Hedayat, “Cooperative communication in wireless networks,” *IEEE Commun. Mag.*, vol. 42, pp. 74–80, Oct. 2004.
- [4] G. Kramer, I. Marić, and R. D. Yates, “Cooperative communications,” *Found. Trend. Commun. Inf. Theory*, vol. 1, no. 3–4, pp. 271–425, 2006.
- [5] H.-F. Chong, M. Motani, and H. Krishna, “Generalized backward decoding strategies for the relay channel,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 394–401, Jan. 2007.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [7] J. G. Proakis and M. Salehi, *Digital Communications*. New York: McGraw-Hill, 5th ed., 2008.
- [8] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, pp. 1–10, Jan. 1976.
- [9] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, “Noisy network coding,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 3132–3152, May 2011.
- [10] J. Hou and G. Kramer, “Short message noisy network coding for multiple sources,” in *Proc. IEEE Int. Symp. Inf. Theory*, (Cambridge, MA, USA), pp. 1677–1681, July 2012.

- [11] P. Zhong and M. Vu, “Combined decode-forward and layered noisy network coding schemes for relay channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, (Cambridge, MA, USA), pp. 1341–1345, July 2012.
- [12] K. Luo, R. H. Gohary, and H. Yanikomeroglu, “On the generalization of decode-and-forward and compress-and-forward for Gaussian relay channel,” in *Proceedings of IEEE Info. Theory Workshop (ITW)*, (Paraty, Brazil), pp. 623 – 627, Oct. 2011.
- [13] K. Luo, R. H. Gohary, and H. Yanikomeroglu, “Analysis of the generalized DF-CF for Gaussian relay channels: decode or compress?,” *IEEE Trans. Commun.*, vol. 61, pp. 1810–1821, May 2013.
- [14] K. Luo, R. H. Gohary, and H. Yanikomeroglu, “A decoding procedure for compress-and-forward and quantize-and-forward relaying,” in *Proc. 50th Allerton Conf. Commun, Control Comput.*, (Monticello, IL), Oct. 2012.
- [15] K. Luo, R. H. Gohary, and H. Yanikomeroglu, “Exploiting the N -to-1 mapping in compress-and-forward relaying,” under review in *IEEE Trans. Inform. Theory* (submission: 12 July 2014, 1st results: 13 March 2015, 1st revision: 11 June 2015).
- [16] K. Luo, R. H. Gohary, and H. Yanikomeroglu, “The capacity of a broadcast channel with Gaussian jamming and a friendly eavesdropper,” (to be presented in *IEEE Info. Theory Workshop (ITW)*, Jeju, Korea), Oct. 2015.
- [17] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband radio,” *IEEE Commun. Mag.*, vol. 42, pp. 80–89, Sept. 2004.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [19] T. J. Oechtering and H. Boche, “Bidirectional regenerative half-duplex relaying using relay selection,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1879–1888, May 2008.
- [20] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Sept. 2005.

- [21] A. A. El Gamal, M. Mohseni, and S. Zahedi, “Bounds on capacity and minimum energy-per-bit for AWGN relay channels,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1545–1561, Apr. 2006.
- [22] X. Cai, Y. Yao, and G. B. Giannakis, “Achievable rates in low-power relay links over fading channels,” *IEEE Trans. Commun.*, vol. 53, pp. 184–194, Jan. 2005.
- [23] W. Su and X. Lui, “Optimum selection relaying protocols in cooperative wireless networks,” *IEEE Trans. Commun.*, vol. 58, pp. 52–57, Jan. 2010.
- [24] L. Zhou and W. Yu, “Capacity of the Gaussian relay channel with correlated noises to within a constant gap,” *IEEE Commun. Lett.*, vol. 16, pp. 2–5, Jan. 2012.
- [25] N. Ksairi, P. Ciblat, P. Bianchi, and W. Hachem, “Performance analysis over slow fading channels of a half-duplex single-relay protocol: Decode or quantize and forward,” *IEEE Trans. Commun.*, vol. 60, pp. 2009–2016, July 2012.
- [26] L. Zhang, J. Jiang, A. J. Goldsmith, and S. Cui, “Study of Gaussian relay channels with correlated noises,” *IEEE Trans. Commun.*, vol. 59, pp. 863–876, Mar. 2011.
- [27] A. Høst-Madsen and J. Zhang, “Capacity bounds and power allocation for wireless relay channels,” *IEEE Trans. Inf. Theory*, vol. 51, pp. 2020–2040, June 2005.
- [28] B. Wang, J. Zhang, and A. Høst-Madsen, “On the capacity of MIMO relay channels,” *IEEE Trans. Inf. Theory*, vol. 51, pp. 20–43, Jan. 2005.
- [29] P. Razaghi and W. Yu, “Parity forwarding for multiple-relay networks,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 158–173, Jan. 2009.
- [30] A. E. Gamal and M. Aref, “The capacity of the semideterministic relay channel,” *IEEE Trans. Inf. Theory*, vol. 28, pp. 536–536, May 1982.
- [31] C. Thomas M and Y.-H. Kim, “Capacity of a class of deterministic relay channels,” *Proc. IEEE Int. Symp. Inf. Theory*, pp. 591–595, June 2007.
- [32] A. Høst-Madsen, “Deterministic capacity of networks,” *Proc. IEEE Inf. Theory Wkshp.*, pp. 601–606, Sept. 2007.
- [33] J. Cui, “The capacity of Gaussian orthogonal multiple-access relay channel,” *IEEE Commun. Lett.*, vol. 15, pp. 365–367, Apr. 2011.

- [34] M. Aleksic, P. Razaghi, and W. Yu, “Capacity of a class of modulo-sum relay channels,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 921–930, Mar. 2009.
- [35] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Approximate capacity of Gaussian relay networks,” *Proc. IEEE Int. Symp. Inf. Theory*, pp. 474–478, July 2008.
- [36] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 1872–1905, Apr. 2011.
- [37] P. Rost and G. Fettweis, “Analysis of a mixed strategy for multiple relay networks,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 174–189, Jan. 2009.
- [38] S. R. Bhaskaran, “Forward decoding over a relay channel,” *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2673–2677, July 2008.
- [39] H.-F. Chong and M. Motani, “On achievable rates for the general relay channel,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 1249–1266, Mar. 2011.
- [40] D. P. Bertsekas, *Nonlinear Programming*. Nashua, NH: Athena Scientific, 2nd ed., 1999.
- [41] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2012.
- [42] M. H. Yassaee and M. R. Aref, “Generalized compress-and-forward strategy for relay,” *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2683–2687, July 2008.
- [43] X. Wu and L.-L. Xie, “On the optimal compressions in the compress-and-forward relay schemes,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 2613–2628, May 2013.
- [44] Y.-H. Kim, “Coding techniques for primitive relay channels,” in *Proc. 45th Allerton Conf. Commun., Control Comput.*, (Monticello, IL), pp. 129–135, Sept. 2007.
- [45] Y.-H. Kim, “Capacity of a class of deterministic relay channels,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 1328–1329, Mar. 2008.
- [46] H.-F. Chong and M. Motani, “The capacity of several new classes of semi-deterministic relay channels,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 6397–6404, Oct. 2011.
- [47] G. Kramer and H. Jie, “Short-message quantize-forward network coding,” in *Multi-Carrier Sys. Sols. (MC-SS), Int. Wkshp.*, (Munich), pp. 1–3, May 2011.

- [48] J. Hou and G. Kramer, “Short message noisy network coding,” Aug. 2013. Available at: <http://arxiv-web3.library.cornell.edu/abs/1304.1692v2>.
- [49] M. H. Yassaee and M. R. Aref, “Slepian-Wolf coding over cooperative relay networks,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 3462–3482, June 2011.
- [50] Y. Liang and G. Kramer, “Rate regions for relay broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 3517–3535, Oct. 2007.
- [51] T. Basar, “The Gaussian test channel with an intelligent jammer,” *IEEE Trans. Inf. Theory*, vol. 29, pp. 152–157, Jan. 1983.
- [52] M. Médard, “Capacity of correlated jamming channels,” in *Proc. 35th Allerton Conf. Commun., Control Comput.*, (Monticello, IL), pp. 1043–1052, Sept. 1997. Also available at: <http://www.mit.edu/~medard/pubs.html>.
- [53] R. J. McEliece and W. E. Stark, “An information theoretic study of communication in the presence of jamming,” in *Proc. IEEE Int. Conf. Commun.*, (Denver, CO, USA), pp. 45.3.1–45.3.5, June 1981.
- [54] M. Médard, D. Marquis, R. A. Barry, and S. J. Finn, “Security issues on all-optical-networks,” *IEEE Network*, vol. 47, pp. 42–48, May 1997.
- [55] E. A. Jorswieck, H. Boche, and M. Weckerle, “Optimal transmitter and jamming strategies in Gaussian MIMO channels,” in *Proc. IEEE Vehic. Technol. Conf.*, (Stockholm, Sweden), pp. 978–982, May 2005.
- [56] K. Cheun, K. Choi, H. Lim, and K. Lee, “Antijamming performance of a multicarrier direct-sequence spread-spectrum system,” *IEEE Trans. Commun.*, vol. 47, pp. 1781–1784, Dec. 1999.
- [57] A. Kashyap, T. Basar, and R. Srikant, “Correlated jamming on MIMO Gaussian fading channels,” *IEEE Trans. Inf. Theory*, vol. 50, pp. 2119–2123, Sept. 2004.
- [58] E. Tekin and A. Yener, “The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 2735–2751, June 2008.
- [59] S. Shafie and S. Ulukus, “Mutual information games in multiuser channels with correlated jamming,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 4598–4607, Oct. 2009.

- [60] T. M. Cover, “Broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 2–14, Jan. 1972.
- [61] Y. Liang and V. V. Veeravalli, “Cooperative relay broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 900–928, Mar. 2007.
- [62] M. N. Khormuji, A. A. Zaidi, and M. Skoglund, “Interference management using nonlinear relaying,” *IEEE Trans. Commun.*, vol. 58, pp. 1924–1930, July 2010.
- [63] L. Lai and H. E. Gamal, “The relay-eavesdropper channel: Cooperation for secrecy,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 4005–4019, Sept. 2008.