KTH Electrical Engineering

Cooperative Strategies in Multi-Terminal Wireless Relay Networks

JINFENG DU

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2012

TRITA-EE 2012:044
ISSN 1653-5146
ISBN 978-91-7501-512-5

KTH Royal Institute of Technology
School of Electrical Engineering
Communication Theory Laboratory
SE-100 44 Stockholm, SWEDEN

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation fredag den 9 november 2012 klockan 13.15 i hörsal Q1, Osquldas väg 4, Stockholm.

Tryck: Universitetsservice US AB

To Feifan & Britta.

Abstract

Smart phones and tablet computers have greatly boosted the demand for services via wireless access points, keeping constant pressure on the network providers to deliver vast amounts of data over the wireless infrastructure. To enlarge coverage and enhance throughput, relaying has been adopted in the new generation of wireless communication systems, such as in the Long-Term Evolution Advanced standard, and will continue to play an important role in the next generation wireless infrastructure. Depending on functionality, relaying can be characterizing into three main categories: amplify-and-forward (AF), compression-and-forward (CF), and decode-and-forward (DF). In this thesis, we investigate different cooperative strategies in wireless networks when relaying is in use.

We first investigate the capacity outer and inner bounds for a wireless multicast relay network where two sources, connected by error-free backhaul, multicast to two destinations with the help of a full-duplex relay node. For high-rate backhaul scenarios, we find the exact cut-set bound of the capacity region by extending the proof of the converse for the Gaussian relay channel. For low-rate backhaul scenarios, we present two genie-aided outer bounds by extending the previous proof and introducing two lemmas on conditional (co-)variance. Our inner bounds are derived from various cooperative strategies by combining DF/CF/AF relaying with network coding schemes. We also extend the noisy network coding scheme and the short-message noisy network coding approach to correlated sources. For low-rate backhaul, we propose a new coding scheme, partial-decode-and-forward based linear network coding. We derive the achievable rate regions for these schemes and measure the performance in term of achievable rates over Gaussian channels. By numerical investigation we observe significant gains over benchmark schemes and demonstrate that the gap between upper and lower bounds is in general not large. We also show that for high-rate backhaul, the cut-set bound can be achieved when the signal-to-noise ratios lie in the sphere defined by the source-relay and relay-destination channel gains.

For wireless networks with independent noise, we propose a simple framework to get capacity outer and inner bounds based on the "one-shot" bounding models. We first extend the models for two-user broadcast channels to many-user scenarios and then establish the gap between upper and lower bounding models. For networks with coupled links, we propose a channel decoupling method which can decompose the network into overlapping multiple-access channels and broadcast channels. We then apply the one-shot models and create an upper bounding network with only bit-pipe connections. When developing the lower bounding network, we propose a two-step update of these models for each coupled broadcast and multiple-access channels. We demonstrate by some examples that the resulting upper bound is in general very good and the gap between the upper and lower bounds is usually not large.

For relay-aided downlink scenarios, we propose a cooperation scheme by cancelling interference at the transmitter. It is indeed a symbol-by-symbol approach to one-dimension dirty paper coding (DPC). For finite-alphabet

signaling and interference, we derive the optimal (in terms of maximum mutual information) modulator under a given power constraint. A sub-optimal modulator is also proposed by formulating an optimization problem that maximizes the minimum distance of the signal constellation, and this non-convex optimization problem is approximately solved by semi-definite relaxation. Bit-level simulation shows that the optimal and sub-optimal modulators can achieve significant gains over the Tomlinson-Harashima precoder (THP) benchmark and over non-DPC reference schemes, especially when the power of the interference is larger than the power of the noise.

Acknowledgments

The work presented in this thesis was conducted during my Ph.D. study at Communication Theory Lab, Royal Institute of Technology (KTH), from 2009 to 2012. I would like to take this opportunity to thank many individuals whose support and inspiration make this thesis possible.

First and foremost, I would like to thank my supervisor Prof. Mikael Skoglund, for giving me this precious opportunity to pursuit my Ph.D. in telecommunications and for sharing your vision and expertise. Your continuous encouragement and trust during the past few years have helped me to conquer hard problems, and the almost unbounded freedom to explore research topics makes my Ph.D. study more joyful. I'm grateful to my co-advisor Prof. Ming Xiao for your support and encouragement, and for treating me as an independent researcher from day one. I'm as grateful for the privilege of discussing with you whatever I encountered whenever possible.

I would like to express my gratitude to Prof. Muriel Médard for welcoming and mentoring me as a visiting research student in your group at Massachusetts Institute of Technology (MIT), during the second half of 2011. Your insights and wisdom have inspired me to think further and dream bigger in research. I'm also grateful for introducing me to network equivalent theory and for many of our fruitful discussions, which have substantially improved the quality of this thesis. I would also like to thank Prof. Shlomo Shamai (Shitz) for sharing your insights on noisy network coding and relaying in general, which enriches the content of this thesis. I'm especially grateful for your suggestions on future work that will lead to more collaborations in the near future.

Many thanks give to my collaborator and master thesis supervisor Prof. Erik G. Larsson: your hands-on training and close supervision benefit me long after our collaboration. My former supervisor Docent Svante Signell, who introduced me to generalized multi-carrier communication and guided me through my Licentiate degree, deserves special acknowledgment. I would also like to acknowledgment my collaborators Dr. Pei Xiao, Jawad Manssour, Dr. Jinsong Wu, and Dr. Qingchun Chen, whose contributions have not been included in this thesis, but help to shape my way towards an independent researcher.

The amazing creative environment and social atmosphere on the "4th floor"–Communication Theory and Signal Processing laboratories–benefit my (and many others') research in many aspects, thanks to all the previous and present colleagues. I have enjoyed very much the constructive discussions in our internal seminars, which have greatly broadened my knowledge base and research interest, and deepened my understanding in many interesting topics. I'm truly grateful to all the discussions we have. I'm indebted to Ricardo Blasco Serrano, Hieu T. Do, Kittipong Kittichokechai, and Efthymios Stathakis, who have helped proofreading part of the thesis. I would like to thank Prof. Mats Bengtsson for quality review of the thesis and all your constructive comments. I would also like to thank Annika Augustsson, Iréne Kindblom, Tove Schwartz, and Raine Tiivel for taking care of administrative issues.

I would like to thank Prof. Lizhong Zheng for taking time to act as opponent for this thesis, and also Prof. Petar Popovski, Prof. Olav Tirkkonen, and Prof. Mikael Johansson for participating in the evaluation committee.

I would like to give special thanks to all my Chinese colleagues at KTH and friends in Sweden. Without you, my life would not be as colorful as what it is today.

My deepest gratitude goes to my parents and beloved family members whose endless love and support is always there whenever I need. I would like to thank my wife Feifan, for your love, your care, all the joy you have created for me, for your sacrifice when creating and bringing up our little angel Britta, and for making anywhere we stay together my home.

Jinfeng Du
Stockholm, October 2012

Contents

Chapter 1

Introduction

In this chapter we will first briefly describe the motivation behind the work presented in this thesis and discuss situations in practice where the results can be applied. We will also present a review of the recent technical progress in related fields. We aim to provide a general understanding of the practical problems this thesis intends to solve, why they are important, and how it will be possible to benefit from these research results in the future. The outline of this thesis with a summary of the main contributions will be presented at the end of this chapter together with a list of notation that will be frequently used in this thesis.

1.1 Motivation

The society we are now living in becomes more and more connected by and dependent on the wireless communication infrastructure. The mobile phone is nowadays not only a telephone, but also a convenient and almost[1] all-time-available access point to our social networks, public services, and even some consumer products. Currently in Sweden, people can easily declare tax, report sickness/parental leave, buy bus/train tickets, pay parking fee, and access many other services via the traditional short message service (SMS). Smart phones and tablet computers equipped with greatly enhanced functionality and explosively growing number of small software (so called *application*) have dramatically improved both the quality and quantity of services accessible via wireless connections. For example, customers of the Skandinaviska Enskilda Banken (SEB) in Stockholm can now check the waiting time in nearby SEB branches so that they can plan their journey while walking down the street. Such location based services and products, as well as personalized entertainment contents and user generated multimedia materials, have become more and more popular among smart phone users, and all of them require data transmission to and/or from access points via wireless connection. The demand for services via

[1]The vision of "anytime, anywhere" connection depends very much on the stability and robustness of the wireless communication infrastructure.

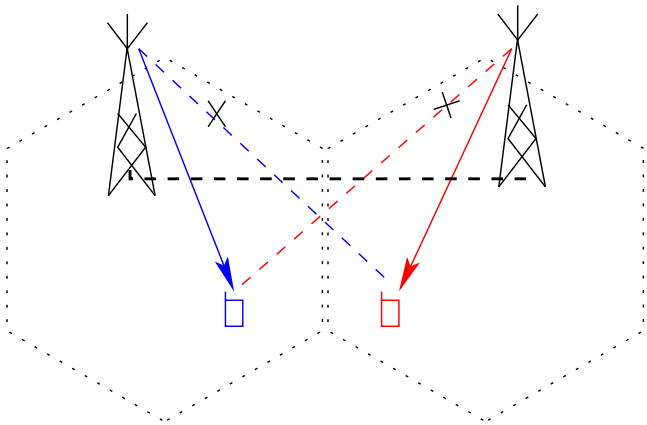


Figure 1.1. In current wireless communication systems, interference signals are deliberately suppressed by transmission via non-overlapping wireless channels, e.g., at different time/frequency-band/direction. Access points or base stations connected via backhaul (fiber or microwave) exchange controlling messages to coordinate the resource allocation which facilitates interference suppression.

wireless access points has been and will continue to be the main drive that keeps constant pressure on the wireless network providers to deliver vast amount of data over the wireless infrastructure, which in turn requires a more efficient usage of the valuable resources, namely radio bandwidth (i.e., spectrum) and energy.

Due to the broadcast nature of wireless transmission, signals dedicated to one user will be overheard by its neighbors. When users do not cooperate, as is usually the case in current systems, such overheard signals degrade the quality of the desired signal and therefore are treated as interference. Ever since the birth of wireless communication about one hundred years ago, numerous research efforts have been devoted to formulate a virtual point-to-point connection between source and destination nodes by suppressing the interference. As illustrated in Figure 1.1, interference signals originating from parallel transmission in the neighborhood can be deliberately suppressed by scheduling such parallel transmission at different time slots, frequency bands, spacial direction, or with different (preferably orthogonal) digital sequences. The spectrum and energy efficiency of such point-to-point wireless connection has been constantly improved via new innovations in antenna design, signal processing, and modulation and coding design. As the throughput of the point-to-point wireless connection is approaching its theoretical limit, it becomes harder and harder to meet the ever growing data rate requirement by further improving the spectrum and energy efficiency.

To overcome such difficulties, the broadcast nature of wireless transmission has to be taken into consideration during the design and innovation of wireless communication techniques. The overheard signal, although appears destructive to one user, might be helpful for another user nearby if the two users are allowed to co-

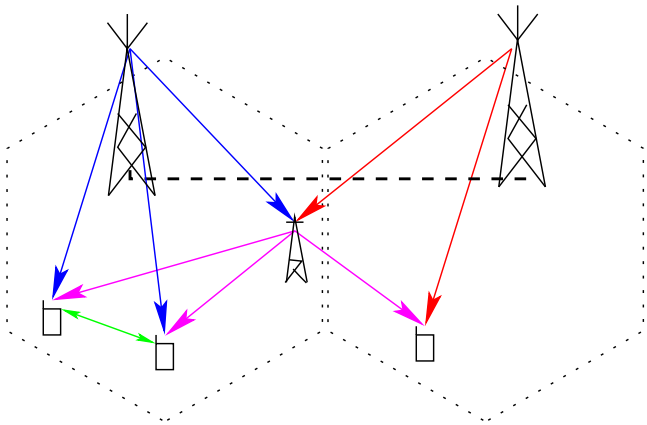


Figure 1.2. In future wireless communication systems, cooperation among transmitters and/or among receivers will be widely adopted with the assistance of dedicated relay nodes.

operate. When users can cooperate, the destructive interference signal becomes a valuable resource and therefore can be utilized to assist the decoding of desired signals, leading to higher energy efficiency. Besides, cooperation allows parallel transmission over the same channel and hence has the potential to greatly increase the spectrum efficiency. Such communication scheme is named cooperative communication to differentiate from the traditional point-to-point communication scheme. The cooperation can be carried out among source nodes, among destinations, and with aid from dedicated relay nodes, as illustrated in Figure 1.2. The cooperation among wireless access points (base stations) can be realized via the widely deployed backhaul connection, either fiber or microwave, and the cooperation among user terminals can be achieved via device-to-device communication channels. Although dedicated relay nodes, known as repeaters, have been introduced to assist long distance wireless transmission around one hundred years ago shortly after the invention of triode vacuum tube, relays with more advanced functionality were not considered for commercial deployment until several years ago. Dedicated relay nodes have been adopted in the next generation wireless communication systems, such as in the Long-Term Evolution Advanced (LTE-Advanced) standard, which are expected to come into commercial deployment within a few years. Relay nodes with advanced functionality will continue to play an important role in the future communication systems.

The results present in this thesis will provide better understanding of various cooperative communication strategies that are proposed/investigated in our research work, by quantifying their theoretical performance limits and highlighting the principles and insights which will guide the design and implementation of future cooperative communication systems.

1.2 Background and Problem Formulation

In this thesis, we investigate different cooperative communication strategies in wireless networks when relaying is in use. We focus on the fundamental limits of these cooperation schemes to gain insights on the design and implementation of such cooperation schemes in future wireless communication systems.

1.2.1 A Brief Review of Related Work

In this section we provide a brief review of previous work that relates to the main building blocks in our proposed schemes, namely relaying, network coding, and source cooperation. The related work of cancelling interference at transmitter will be presented in Chapter 7.

Relaying Techniques

Relaying-based cooperative communication techniques have the potential to boost both the communication range and data rate. A full understanding of such systems, even for the original three-node relay network [vdM71], is however not yet available. In the last 30 years, numerous research efforts have been devoted to the relay networks. Capacity bounds and various cooperative strategies for three-node relaying networks (source-relay-sink, or two cooperative sources and one sink) have been studied in [CE79], where two fundamental relaying schemes, decode-and-forward (DF) and compress-and-forward (CF), are formally introduced and characterized and capacity results have been established for degraded and reversely degraded relay channels. Upper and lower bounds on the outage capacity for the three-node relay channel in fading have been studied in [HMZ05]. Various encoding schemes have been investigated for multiple-access relay channels (MARC) [KGG05, KvW00] involving multiple sources and a single destination, and for broadcast relay channels (BRC) [KGG05, LK07] where a single source transmits messages to multiple destinations. Three decoding protocols, namely *forward decoding* [CE79], *backward decoding* [Car82], and *sliding-window decoding* [Wil82], have been summarized and extended to multiple-source or multiple-relay scenarios in [KGG05]. Recent results on capacity bounds for multiple-source multiple-destination relay networks, [SE07, GSG+09, GSG+10, AH09, TY11, ZY11] and references therein, have provided valuable insight into the benefits of cooperative relaying, either half-duplex (a relay that cannot transmit and receive simultaneously) or full-duplex (can support simultaneous reception and transmission), and demonstrated various tools to bound the capacity region.

Network Coding

The concept of network coding, which essentially means to combine multiple messages together, was first formally introduced and characterized in [ACLY00]. A

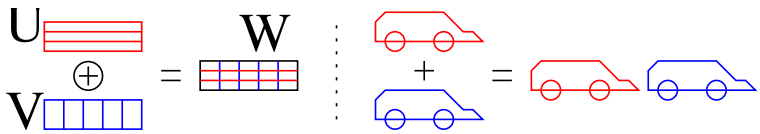


Figure 1.3. Illustration of the main ingredient of network coding: information flows (message blocks) can be *mixed* into one without increasing the size as contrast to the commodity flow.

simple illustration of the main idea behind network coding can be found in Figure 1.3. Unlike commodity flow where an operation at any intermittent node can not affect the volume of the flow that passes through, the information flow can actually be combined efficiently. It has been proved in [ACLY00] that network coding can achieve the max-flow min-cut bound in single-source multicast networks. It is further proved in [LYC03] that linear coding is sufficient to achieve the optimality of network coding in the single-source multicast setup. An algebraic approach [KM03] has been introduced into the network coding framework which greatly simplifies the analysis of data network capacity. Necessary and sufficient conditions for the feasibility of given transmission tasks over a given network has been established in the case that the network only permits linear coding. A distributed random linear network coding approach has been introduced in [HMK+06] for general multicast networks and shown to be robust to network changes or link failures.

As different messages mix up at the relay node by nature in wireless networks, various network coding approaches can be introduced at the relay to boost system capacity. For instance, as demonstrated in [KRH+08], one may first receive individual messages separately, combine them and then transmit based on knowledge overheard from neighboring transmission. One can also schedule the parallel transmission carefully such that the overheard signals can be used directly for network coding, as demonstrated in [KGK07, KMG+07] where amplify-and-forward (AF) relaying has been utilized. Such a scheme, coined as analog network coding (ANC), has been proven to be asymptotically optimal [MGM12] in multihop relay networks. Apart from AF relaying, more advanced relaying functionality can be utilized to carry out NC operation directly based on the received signal. For example, when quantization is performed by the relay, a quantize-map-and-forward (QMF) scheme has been proposed in [ADT11] for unicast networks. With symbol-by-symbol scalar quantization, QMF has been proved to be approximately optimal (within a constant gap to the cut-set bound). The principle of noisy network coding (NNC) [LKEC11], which can be regarded as an extension of QMF with vector quantization, can be easily extended to multiple-source and/or multiple-relay networks. In [WNPS10] joint NC and physical layer coding is performed via lattice coding for the bi-directional relay channel. Linear network coding and lattice codes with decode-and-forward relaying are investigated in [GSG+10]. One may also decode a linear combination of the transmitted messages directly from the mixed signal and forward the combination itself together with the corresponding coefficients, as demonstrated in [NG11]

under the name of compute-and-forward where structure codes are utilized such that the linear combination of messages is still a valid message.

Source Cooperation

Apart from introducing dedicated relay nodes to help the transmission, one can also utilize cooperative strategies among sources [Wil83, DMT06, MYK07, NJGM07, BLW08, SGP$^+$09] and/or among destinations [LTW04, NJGM07, SGP$^+$09] with the help of orthogonal conferencing channels. Willems [Wil83] introduced source-conferencing for the discrete memoryless multiple-access channel (DM-MAC) and characterized the capacity region. Bross et. al [BLW08] extended the coding scheme to the Gaussian setting and proposed a new converse. Coding schemes and capacity regions for the compound MAC with conferencing encoders have been studied in [MYK07, SGP$^+$09]. Interference channels with unidirectional conferencing encoders are investigated in [DMT06, MYK07]. Capacity bounds within a constant gap for interference channels with limited source cooperation have been characterized in [WT11] for out-of-band source-conferencing and in [PV11] for in-band cooperation channels. Diversity gains by source cooperation in fading channels with full/partial channel state information (CSI) have been studied in [SEA03a, SEA03b, LTW04, HM06, NJGM07]. The trade-off between sharing message and local CSI among source nodes through finite-rate backhaul has been studied in [Ray06, WBBJ11, ZG11].

1.2.2 Problem Formulation

Capacity Bounds for Multiple Multicast Relay Networks

We focus on a relay-aided two-source two-destination multicast network with backhaul support, as shown in Figure 1.4. Source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$ multicast their individual message W_1 at rate R_1 and W_2 at rate R_2, respectively, to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$, with the help of a relay $\mathcal{R}$. The nodes $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{R}$ use the same channel resource (i.e. co-channel transmission) and transmitted signals mix at all the receiving terminals and are subjected to Gaussian noise. In addition, the source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$ are connected by orthogonal limited-rate error-free conferencing links (corresponding to the presence of a backhaul) with capacities C_{12} and C_{21}, respectively.

The model in Figure 1.4 is generic and interesting since it is a combination of relaying, MARC, BRC, source cooperation, and network coding. It covers a class of different building blocks and can be extended to more general networks, by tuning the channel gains g_{ij} and C_{12}, C_{21} within the range $[0, \infty)$. It can be applied, for example, to cellular downlink scenarios where two base stations, connected through the (fiber or microwave) backhaul, multicast multimedia content to two mobile terminals, one in each cell, with the help of a dedicated relay deployed at the common cell boundary. Since the base stations are connected through the (fiber

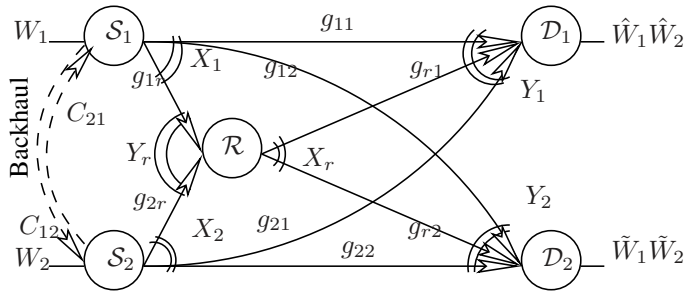


Figure 1.4. Source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with orthogonal and error-free backhaul (with rate C_{12} and C_{12} bits per channel use), multicast information W_1 at rate R_1 and W_2 at rate R_2 respectively to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ through Gaussian channels, with aid from a full-duplex relay $\mathcal{R}$.

or microwave) backhaul, more general network coding schemes can be used at the relay to cooperate with the sources' transmission.

We are interested in the maximum achievable rates supported by such systems. The meaning of "achievable" can be explained as follows: given a rate pair (R_1, R_2), when $\mathcal{S}_1$ transmits at rate R_1 and $\mathcal{S}_2$ transmits at rate R_2 using a cooperative transmission strategy, if it is possible that the destination nodes $\mathcal{D}_1$ and $\mathcal{D}_2$ can decode the messages with an error probability that can be made arbitrarily small, then we say that the rate pair (R_1, R_2) is achievable. An achievable rate region of a cooperative strategy is defined to be the set of all the achievable rate pairs supported by the strategy. The capacity region of a system is defined as the union of all the achievable rate regions, and therefore it has the following two properties: all the rate pairs inside the region are achievable, and no rate pair outside the region is achievable.

We aim at evaluating the theoretical limits of the capacity region for the system shown in Figure 1.4. We will propose various cooperative strategies where source cooperation and network coding are designed jointly with the relaying. The achievable rate regions of the corresponding strategies will be characterized and serve as the inner bounds of the capacity region. We will also setup outer bounds for the capacity region.

Cooperation by Cancelling Interference at Transmitter

The bounds on capacity are in general established by coding over an infinite number of dimensions. To obtain an understanding of what one can achieve in small (or a single) dimensions of signals and at low complexity, we consider a communication network where the base station transmits information symbols ω_1 and ω_2 to user 1 and user 2, respectively, with the aid of a half-duplex relay. As illustrated in Figure 1.5, the relay is dedicated to assist user 1 (the weaker/more distant user) whose direct link with the source fails. The base station transmits x_1 (signal for ω_1)

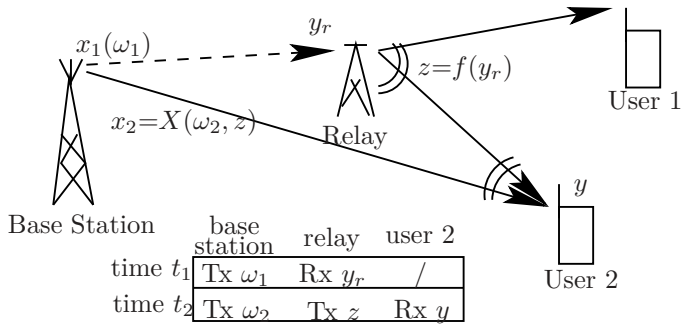


Figure 1.5. The base station transmits ω_1 to user 1 during time slot t_1 and ω_2 to user 2 during time slot t_2. The relaying signal $z{=}f(y_r)$ dedicated for user 1 appears as "interference" for user 2. With non-causal knowledge of z, the base station can design a DPC modulator $x_2 = X(\omega_2, z)$ given the information symbol ω_2 and the interference z.

during time slot t_1 and x_2 (signal for ω_2) during t_2. The relay listens to the base station during t_1 and transmits $z = f(y_r)$ during t_2, where y_r is the received signal at the relay during t_1 and $f(\cdot)$ is a relay mapping function. The relaying signal z, which is useful for user 1, appears as interference for user 2. Assuming that the relaying function $f(\cdot)$ is known at the base station and that the source-relay link is good, the "interference" z will be known non-causally at the base station with high probability, effectively resulting in the Costa problem (also known as dirty paper coding after [Cos83]). We will propose a symbol-by-symbol scheme for cancelling the interference known at the transmitter in the relay-aided downlink channel.

1.3 Thesis Outline and Contributions

Chapter 2 will first introduce the system model of the work and provide justifications for the assumptions made, followed by a brief description of fundamental tools that will serve as cornerstones for our design and analysis of cooperative communication strategies. For wireless multiple multicast relay networks with backhaul support between source nodes, Chapter 3 focuses on the cut-set bound based capacity outer bounds, and Chapter 4 describes various cooperative NC strategies based on a DF relay. Chapter 5 investigates cooperative strategies when relay with compression or amplification functionality is utilized. Chapter 6 proposes general bounding models that can construct in an efficient way noiseless bounding networks for noisy networks with independent noise. A one dimensional low complexity cooperation scheme by cancelling interference at transmitter in a relay aided downlink broadcast channel is presented in Chapter 7. Chapter 8 concludes this thesis.

1.3.1 Capacity Outer Bounds for Multicast Relay Networks

Chapter 3 investigates capacity outer bounds for the wireless multicast relay network with backhaul between source nodes, as shown in Figure 1.4. For the scenario when the source nodes can fully cooperate, i.e., with high-rate backhaul ($C_{12} \geq R_1$, $C_{21} \geq R_2$), we presented the exact cut-set bound by extending the proof of the converse for the Gaussian relay channel as stated in [CE79]. For low-rate backhaul ($0 \leq C_{12} < R_1$, $0 \leq C_{21} < R_2$), we present two genie-aided outer bounds by extending the previous proof and introducing two lemmas on conditional (co-)variance.

The results on outer bounds have been published in the following papers:

[DXS11a] J. Du, M. Xiao, and M. Skoglund, "Capacity bounds for backhaul-supported wireless multicast relay networks with cross-links," in *Proceedings IEEE International Conference on Communications (ICC)*, Jun. 2011.

[DXS11b] J. Du, M. Xiao, and M. Skoglund, "Cooperative strategies for relay-aided multi-cell wireless networks with backhaul," *IEEE Transactions on Communications*, vol. 59, pp. 2502–2514, Sep. 2011.

[DXSM] J. Du, M. Xiao, M. Skoglund, and M. Médard, "Wireless multicast relay networks with limited-rate source-conferencing," *IEEE Journal on Selected Areas in Communications*, special issue on Theories and Methods for Advanced Wireless Relays. To appear.

1.3.2 Capacity Inner Bounds by Cooperative Relaying Strategies

Chapter 4 investigates DF relaying based cooperative strategies based on different network coding schemes, namely, finite field network coding, linear network coding, lattice coding. We derive the achievable rate regions for these schemes and show that for high-rate backhaul, the cut-set bound can be achieved when the signal-to-noise ratios lie in the sphere defined by the source-relay and relay-destination channel gains. For low-rate backhaul scenarios, we propose a new coding scheme, partial-decode-and-forward based linear network coding, which is essentially a hybrid scheme utilizing rate-splitting and messages exchange at the source nodes, partial decoding and linear network coding at the relay, and joint decoding at each destination.

Chapter 5 focuses on non-decoding relaying based cooperation schemes. We extend the noisy network coding (NNC) scheme to the scenario with partial source cooperation. We also demonstrate that by using short-message NNC (SNNC) with rate splitting, message exchange via backhaul, and superposition coding at source nodes, SNNC can achieve a strictly larger rate region than NNC with compression forwarding, as long as the destination nodes in SNNC scheme have the option

to treat relaying signals from relay nodes as noise. A low-complexity alternative scheme, AF based ANC, is also investigated and shown to benefit greatly from message exchange via backhaul and can even outperform NNC when the coherent combining gain is dominant.

Significant parts of this work have already been published in [DXS11b, DXSM], and in the following two papers:

[DXS10b] J. Du, M. Xiao, and M. Skoglund, "Cooperative strategies for relay-aided multi-cell wireless networks with backhaul," in *Proceedings IEEE Information Theory Workshop (ITW)*, Aug. 2010.

[DXSS12] J. Du, M. Xiao, M. Skoglund, and S. Shamai (Shitz), "Short-message noisy network coding with partial source cooperation," in *Proceedings IEEE Information Theory Workshop (ITW)*, Sep. 2012.

1.3.3 General Bounding Models for Networks with Independent Noise

In Chapter 6 we propose a simple framework to get capacity outer and inner bounds for wireless networks with independent noise. We first extend the "one-shot" bounding tools proposed in [CME11] for the two-user broadcast channel to many-user scenarios and then establish the gap between upper and lower bounding models. For networks with coupled multiple-access and broadcast channels, we propose a channel decoupling method which can decompose the network into overlapping multiple-access channels and broadcast channels. We then apply the one-shot upper bounding blocks and create an upper bounding network with only bit-pipe connections, on which the cut-set bound can be easily calculated. This will serve as a natural upper bound for the original network. When developing the lower bounding network, we propose an update of these lower bounding models for each coupled broadcast and multiple-access channels. We demonstrate by some examples that the resulting upper bound is in general very good and the gap between the upper and lower bounds is usually not large.

1.3.4 Cooperation by Cancelling Interference at Transmitter

In Chapter 7 we propose a practical symbol-by-symbol scheme for cancellation of interference known at the transmitter in a relay-aided downlink channel. For finite-alphabet signaling and interference, we derive the optimal (in terms of maximum mutual information) modulator under a given power constraint. A sub-optimal modulator is also proposed by formulating an optimization problem that maximizes the minimum distance of the signal constellation, and this non-convex optimization problem is approximately solved by semi-definite relaxation. For the case of binary signaling with binary interference, we obtain a closed-form solution for

the sub-optimal modulator, which only suffers little performance degradation compared to the optimal modulator in the region of interest. For more general signal constellations and more general interference distributions, we propose an optimized Tomlinson-Harashima precoder (THP), which uniformly outperforms conventional THP with heuristic parameters.

Majority of the contents have been published in the following papers:

[DLS06] J. Du, E. G. Larsson, and M. Skoglund, "Costa precoding in one dimension," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006.

[DLXS11] J. Du, E. G. Larsson, M. Xiao, and M. Skoglund, "Optimal symbol-by-symbol Costa precoding for a relay-aided downlink channel," *IEEE Transactions on Communications*, vol. 59, pp. 2274–2284, Aug. 2011.

1.3.5 Contributions Outside the Thesis

We propose in [DXS10a] several capacity outer bounds for the wireless multicast relay network as shown in Figure 1.4 but without cross-links. The results presented in [DXS10a] have been overtaken by the new results presented in [DXS11b] and are therefore not included in this thesis. In [DS09] we have proposed a novel preamble-based channel estimation method for the OFDM/OQAM multi-carrier system based on the structure of self-interference. In [DXWC12] we have proposed blind channel estimation methods for multiple-antenna isotropic orthogonal transform algorithm (IOTA) based multi-carrier systems. The contribution in [DS09, DXWC12] is not inline with the rest material presented in this thesis and therefore not included.

[DS09] J. Du and S. Signell, "Novel preamble-based channel estimation for OFDM/OQAM systems," in *Proceedings IEEE International Conference on Communications (ICC)*, Jun. 2009.

[DXS10a] J. Du, M. Xiao, and M. Skoglund, "Capacity bounds for relay-aided wireless multiple multicast with backhaul," in *Proceedings International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2010.

[DXWC12] J. Du, P. Xiao, J. Wu, and Q. Chen, "Design of isotropic orthogonal transform algorithm-based multicarrier systems with blind channel estimation," *IET Communications*, 2012, accepted for publication. DOI: 10.1049/iet-com.2012.0029

1.4 Notation and Acronyms

Notation

X	Real-valued random variable
x	A realization of the random variable X
$\mathcal{X}$	The set of alphabet (or the support set[2]) of X
$\|\mathcal{X}\|$	Cardinality of a set $\mathcal{X}$
$X^{(n)}$	A vector of length n whose elements are realizations of X
$p(x)$	Probability density/mass function of X
$p(x,y)$	Joint probability density/mass function of (X,Y)
$p(x\|y)$	Conditional probability density/mass function of X given Y
$H(X)$	Entropy of X with a discrete alphabet
$H(X,Y)$	Joint entropy of X and Y
$H(X\|Y)$	Conditional entropy of X given Y
$h(X)$	Differential entropy of X with a continuous-valued alphabet
$I(X;Y)$	Mutual information between X and Y
$I(X;Y\|Z)$	Conditional mutual information between X and Y given Z
$E[X]$	Expected value of X
$E[X\|Y]$	Conditional expectation of X given Y
$\mathrm{Var}(X)$	Variance of X
$\mathrm{Var}(X\|Y)$	Conditional variance of X given Y
$\mathrm{Cov}(X,Y)$	Co-variance between X and Y
$\mathrm{Cov}(X,Y\|Z)$	Conditional co-variance of X and Y given Z
X-Y-Z	Markov chain, i.e., $p(xz\|y) = p(x\|y)p(z\|y)$
$\log(\cdot)$	Logarithm operator of base 2, unless stated otherwise
$\mathcal{N}(\mu,\sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{C}(x)$	Gaussian capacity function with $\mathcal{C}(x) = \max\{\frac{1}{2}\log(1+x), 0\}$
$\mathrm{Re}\{\cdot\}$	Take the real part of a complex number
$\|a\|$	Absolute value of a number a
$\sum_{i=1}^{n}$	Summation of items from index $i=1$ up to $i=n$
$\prod_{i=1}^{n}$	Product of items from index $i=1$ up to $i=n$
$N!$	Factorial of the integer N
$(\cdot)^*$	Complex conjugate of a complex number/vector/matrix
$(\cdot)^T$	Matrix/vector transpose
$\boldsymbol{a}$	Vector $\boldsymbol{a}$ (bold small letter)
$\boldsymbol{A}$	Matrix $\boldsymbol{A}$ (bold capital letter)
$\mathrm{Tr}(\boldsymbol{A})$	Trace of matrix $\boldsymbol{A}$
$\|\boldsymbol{A}\|$	Determinate of matrix $\boldsymbol{A}$
$\mathrm{diag}(\boldsymbol{A})$	A vector generated by the diagonal elements of matrix $\boldsymbol{A}$
$\mathrm{diag}(\boldsymbol{a})$	A diagonal matrix generated from vector $\boldsymbol{a}$

[2]Strictly speaking, the support set is normally a subset of the alphabet set due to the possible existence of dummy elements (with probability 0), which will not contribute to information quantity and therefore can be neglected.

Acronyms

AF	Amplify-and-forward
ANC	Analog network coding
AWGN	Additive white Gaussian noise
BC	Broadcast channel
BER	Bit error rate
bpcu	Bit per channel use
BPSK	Binary phase shift keying
BRC	Broadcast relay channel
CF	Compress-and-forward
CSI	Channel state information
CSIT	Channel state information at transmitter
DF	Decode-and-forward
DPC	Dirty paper coding
FNC	Finite-field network coding
IC	Interference channel
IEEE	Institute of electrical and electronics engineers
IFRC	Interference relay channel
i.i.d	Independent and identically distributed
LNC	Linear network coding
LTE	Long-term evolution
MAC	Multiple-access channel
MAP	Maximum a posteriori
MARC	Multiple-access relay channel
MIMO	Multiple-input multiple-output
ML	Maximum likelihood
MMSE	Minimum mean square error
M-QAM	M-ary quadrature amplitude modulation
NBF	Network coding based beamforming
NC	Network coding
NNC	Noisy network coding
OFDM	Orthogonal frequency division multiplexing
pDF	Partial-decode-and-forward
pdf	Probability density function
pmf	Probability mass function
QCQP	Quadratically-constrained quadratic program
QMF	Quantize-map-and-forward
SDR	Semi-definite relaxation
SINR	Signal to interference plus noise ratio
SNNC	Short-message noisy network coding
SNR	Signal to noise ratio
THP	Tomlinson-Harashima precoder/precoding (THP)
WLAN	Wireless local area network

Chapter 2

System Model and Fundamental Tools

In this chapter, we will present the system model and justifications for some assumptions associated with the model, as well as some principal definitions. We will also present a brief description of fundamental tools in cooperative communication. The description aims to provide a brief yet clear introduction of the main building blocks and highlight the intuition behind, rather than a thorough and rigorous technical review. We refer to [CT06, EK11] for more rigorous treatment of basic definitions and concepts.

2.1 System Model and Justifications

2.1.1 System Model

As stated in Chapter 1, we focus on wireless communication scenarios where the transmitted signal can be received by nodes located in nearby areas (i.e., broadcast channel), and the received signal at one node is normally composed of inputs from neighboring transmitting nodes (i.e., multiple-access channel). Each receiving node suffers from an independent additive white Gaussian noise (AWGN) and each transmitting node has an average power constraint[1].

More specifically, the system shown in Figure 1.4 can be described as follows

$$\begin{aligned} Y_1^{(n)} &= g_{11}X_1^{(n)} + g_{21}X_2^{(n)} \quad + g_{r1}X_r^{(n)} + Z_1^{(n)}, \\ Y_2^{(n)} &= g_{12}X_1^{(n)} + g_{22}X_2^{(n)} \quad + g_{r2}X_r^{(n)} + Z_2^{(n)}, \\ Y_r^{(n)} &= g_{1r}X_1^{(n)} + g_{2r}X_2^{(n)} \quad\quad\quad + Z_r^{(n)}, \end{aligned} \tag{2.1}$$

[1]There are works that consider peak power constraints, which may lead to different results and conclusions when applied to our framework. We will only focus on average power constraints in this thesis.

where $g_{ik} \geq 0$, $i, k = 1, 2, r$ are the individual channel gains, $X_i^{(n)}$, $Y_i^{(n)}$, $Z_i^{(n)}$, $i = 1, 2, r$ are n-dimensional vectors for the transmitted signals, received signals, and additive noise, respectively. All the transmitted signals are subject to average power constraints, i.e.,

$$\frac{1}{n} \sum_{k=1}^{n} X_{i,k}^2 \leq P_i, \text{ for } i = 1, 2, r. \tag{2.2}$$

The system shown in Figure 1.5 can be described as follows

$$\begin{aligned} y_{r,t_1} &= x_1 + n_r, \\ y_{t_1} &= x_1 + n_1, \\ z &= f(y_{r,t_1}), \\ y_{t_2} &= x_2 + z + n, \end{aligned} \tag{2.3}$$

where x_1, x_2, and z are also subject to average power constraints.

Note that the channel coefficients in (2.1) and (2.3) are assumed to be positive scalars, even though the wireless channel is normally both time and frequency dispersive. Therefore (2.1) and (2.3) implicitly assume perfect channel state information at transmitter (CSIT) and simultaneous perfect synchronization at all receivers. This assumption, although widely adopted in information-theoretic work without sufficient justification, is optimistic in practice. Therefore, the results we obtain based on the above assumptions will in general serve as upper bounds on any practical performance, and can be directly extended in a similar way as in [HMZ05] to scenarios where constructive (co-phase) addition is not available.

In the following, we will present some practical schemes that to some extent can justify the above assumptions.

2.1.2 Justification for Perfect Synchronization

The wireless channel is far more complex than a positive scalar coefficient as indicated in (2.1). As the radio wave propagates in the air, it is reflected by buildings and the ground, diffused by small particles, and impeded by large obstacles, which create multiple copies of the original waveform experiencing different time-delay/phase-distortion/amplitude-attenuation.

Simultaneous perfect time synchronization at all receiving nodes seems unrealistic at the first glance due to the different wave propagation delays among transmitter-receiver pairs [ZMM08]. However, it can actually be realized under the framework of orthogonal frequency division multiplexing (OFDM) system which is the cornerstone of the 4th generation (4G) wireless communication systems. In OFDM systems, multi-path signal components with different delays will be translated into a single complex-valued channel coefficient in the frequency domain as long as the length of the cyclic-prefix (CP) is larger than the maximum delay spread. As each transmitted signal will contribute an additive component to the complex-

valued coefficient, its phase can be compensated before transmission as long as the phase of each complex component can be accurately estimated.

Although perfect phase synchronization (thus constructive addition) might be realized at one receiving node, it is impossible to realize simultaneous perfect phase alignment at multiple receiving nodes due to the random phase shift caused by independent reflection and diffusion. Therefore a trade-off among different receiving nodes with respect to phase alignment has to be sought after to maximize the expected system benefit (to be specified by the designer). We will demonstrate this trade-off based on the system described by (2.1).

Let $h_{ik} = a_{ik}e^{j\phi_{ik}}$, $i, k = 1, 2, r$ be the additive complex-valued channel component in the frequency domain from transmitting node i to receiving node k, with $a_{ik} \geq 0$ be the amplitude and ϕ_{ik} be the phase, and we assume that h_{ik} is known perfectly at transmitting node i. Let θ_i be the phase compensation to be carried out by node i, then the complex-valued channel coefficient in frequency domain at receiving node $\mathcal{D}_1$ can be written as

$$\begin{aligned} H_1 &= e^{j\theta_1}h_{11} + e^{j\theta_2}h_{21} + e^{j\theta_r}h_{r1} \\ &= a_{11}e^{j(\theta_1+\phi_{11})} + a_{21}e^{j(\theta_2+\phi_{21})} + a_{r1}e^{j(\theta_r+\phi_{r1})}. \end{aligned} \tag{2.4}$$

Similarly, we have

$$H_2 = a_{12}e^{j(\theta_1+\phi_{12})} + a_{22}e^{j(\theta_2+\phi_{22})} + a_{r2}e^{j(\theta_r+\phi_{r2})}, \tag{2.5}$$

$$H_r = a_{1r}e^{j(\theta_1+\phi_{1r})} + a_{2r}e^{j(\theta_2+\phi_{2r})}. \tag{2.6}$$

Simultaneous perfect phase alignment shall require the following equations (mod 2π is omitted for brevity) to hold simultaneously

$$\begin{aligned} &\theta_1 + \phi_{11} = \theta_2 + \phi_{21}, && \theta_1 + \phi_{11} = \theta_r + \phi_{r1}, \\ &\theta_1 + \phi_{12} = \theta_2 + \phi_{22}, && \theta_1 + \phi_{12} = \theta_r + \phi_{r2}, \\ &\theta_1 + \phi_{1r} = \theta_2 + \phi_{2r}. \end{aligned}$$

Since ϕ_{ik} are independent, the above equations are overdetermined and therefore have no solution, i.e., simultaneous perfect phase alignment at all receiving nodes cannot be achieved. However, one can still perform phase compensation and achieve imperfect phase alignment. One can find a trade-off among the receiving nodes with respect to $(\theta_1, \theta_2, \theta_r)$ as follows

$$(\theta_1, \theta_2, \theta_r) = \arg\max f(H_1, H_2, H_r), \tag{2.7}$$

where $f(H_1, H_2, H_r)$ is the utility function representing the expected system benefit specified by the system designer. For example, one may set

$$f(H_1, H_2, H_r) = \lambda_1|H_1|^2 + \lambda_2|H_2|^2 + \lambda_r|H_r|^2,$$

where λ_i, $i = 1, 2, r$ are the weighting coefficients. How to determine the utility function based on the performance metrics is interesting but out of the scope of this thesis, and therefore we will assume the utility function is determined and the optimization problem (2.7) can be efficiently solved.

2.1.3 Justification for Perfect CSIT

The perfect CSIT assumption for a point-to-point communication link can be justified in the case of time-division duplexing (TDD) transmission by channel reciprocity. Assuming perfect channel estimation during the receiving stage, the channel state information during the transmitting stage can be well predicted as long as the time interval between reception and transmission stages is not large, and the channel state does not vary fast. For multiple-terminal communication networks, however, it is not that straightforward. Below we will demonstrate the proposed scheme[2] based on the system described by (2.1) as follows.

1. Each of the reception nodes ($\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{R}$) broadcasts a training sequence for channel estimation, one by one in a predetermined order, and each of the transmission nodes ($\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{R}$) estimates the corresponding channel coefficients in the frequency domain.

   Note that, after step 1, $\mathcal{S}_1$ has a good prediction of the additive complex-valued channel components h_{11}, h_{12}, and h_{1r} based on the property of channel reciprocity. Similarly, $\mathcal{S}_2$ knows h_{21}, h_{22}, and h_{2r}, and $\mathcal{R}$ knows h_{r1} and h_{r2}.

2. After determining the optimal phase compensation $(\theta_1, \theta_2, \theta_r)$ by solving the optimization problem (2.7), $\mathcal{S}_1$ compensates the phase and transmits $e^{j\theta_1}X_1$ and $\mathcal{S}_2$ transmits $e^{j\theta_2}X_2$. It is similar for the transmission at $\mathcal{R}$.

3. Signals mix up at the receiving nodes in an approximately coherent manner, and the connection between g_{ik} in (2.1) and h_{ik} in (2.4)–(2.6) can be set as

$$g_{ik} = \mathrm{Re}\left\{ e^{j\theta_i} h_{ik} \frac{H_k^*}{|H_k|} \right\},$$

   where $\mathrm{Re}\{\cdot\}$ takes the real part, H_k^* is the complex conjugate of H_k and $|H_k|$ is its absolute value.

2.1.4 Justification for Full-Duplex Relaying

Note that (2.3) assumes half-duplex relaying while (2.1) assumes full-duplex relaying with a block delay, which can be justified by equipping the relay node with two radio frequency (RF) front-ends with a shared cache memory. The signal received by one RF front-end will be processed and then saved to the shared cache memory; in the next block the saved information will be read and processed and then sent out via the other RF front-end. Here we assume that the self-interference, i.e. the signal transmission from transmit RF front-end to the receive RF front-end, can be perfectly cancelled since the transmitted signal by the relay is perfectly known by the relay node itself. The effective channel between the transmit RF front-end to

[2]This scheme is the result of an interesting discussion with Peter Larsson.

the receive RF front-end can be well estimated via training sequences. One example of full-duplex communication implementation can be found in [ADSS12] where one extra transmit RF front-end is used to assist the self-interference cancellation at the receive RF front-end.

2.2 Principal Definitions

2.2.1 Channel Coding, Achievable Rate, and Capacity

A point-to-point channel can be modeled/defined by a conditional probability distribution $p(y|x)$ with channel input $X \in \mathcal{X}$ and channel output $Y \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabet sets consisting of discrete- or continuous-valued elements. A communication task through such a point-to-point channel is to convey some message W by transmitting a sequence of elements selected from $\mathcal{X}$ through the channel, such that after receiving an output sequence, one can guess a message $\hat{W}$ that is most likely to be the one transmitted. If the guess is correct, i.e. $\hat{W} = W$, then we can declare that the communication task is successful, otherwise we declare an error. In the following we will define some important concepts related to point-to-point communication.

Definition 2.1. ***Channel Coding Scheme***
A channel coding scheme is based on a codebook and the associated encoder and decoder.
A codebook is a set of sequences $\{x^{(n)}\}$ with cardinality[3] 2^{nR}, and each sequence in the set is associated with a unique index out of $\{1, 2, \ldots, 2^{nR}\}$. The set of sequences and their associated indices are known at both the encoder and the decoder before the transmission.
An encoder $f(\cdot)$ will select a sequence $X^{(n)}$ from the codebook based on the message W to be transmitted.
A decoder $g(\cdot)$ will select a message $\hat{W}$ based on the channel output sequence $Y^{(n)}$.

Definition 2.2. ***Achievable Rate***
For a message $W \in \{1, 2, \ldots, 2^{nR}\}$ to be transmitted via the channel using the coding scheme defined in Definition 2.1, if the probability of error can be made arbitrarily small when n is sufficiently large, we say that communication at rate $\frac{nR}{n} = R$ bits per channel use (bpcu) is achievable. The rate R is called an achievable rate for the coding scheme through the channel.

Definition 2.3. ***Capacity***
The capacity of a point-to-point channel is the supremum of all the achievable rates that can be realized by any possible coding scheme.

[3]For a real-valued R, 2^{nR} in general is not necessarily an integer, which leads to a notation $\lceil 2^{nR} \rceil$ to ensure the cardinality to be an integer. Here we simply treat 2^{nR} as an integer to simplify the notation.

Similarly, for multiple terminal communication which involves more than one input and/or output alphabet, we can define the coding scheme (also referred to cooperative strategy since it includes interaction among multiple encoding/decoding nodes), the achievable rate region, and the capacity region.

2.2.2 Entropy and Mutual Information

Following the notation in [CT06], we will introduce the most important quantities which will frequently appear in the rest of this thesis.

Definition 2.4. ***Entropy***
For a discrete random variable $X \in \mathcal{X}$ with probability mass function (pmf) $p(x)$, the entropy of X is defined as

$$H(X) = -\sum_{x} p(x) \log(p(x)).$$

Unless otherwise stated, the base of the logarithm operator $\log(\cdot)$ is set to 2 and therefore the resulting information is measured in bits.

Definition 2.5. ***Conditional Entropy***
For discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint pmf $p(x, y)$, the conditional entropy of X given Y

$$\begin{aligned} H(X|Y) &= \sum_{y} p(y) H(X|y) = \sum_{y} p(y) \left(-\sum_{x} p(x|y) \log(p(x|y)) \right) \\ &= -\sum_{x,y} p(x, y) \log(p(x|y)), \end{aligned}$$

where $p(x|y)$ is the conditional probability of $X = x$ given observation $Y = y$, $p(y)$ is the marginal distribution of $p(x, y)$, and $H(X|y)$ is the entropy of X given observation $Y = y$.

Definition 2.6. ***Joint Entropy***
For discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint pmf $p(x, y)$, the joint entropy of X and Y is defined as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log(p(x, y)).$$

From the observer's point of view, $H(X)$ is the *amount of uncertainty* (i.e. information unknown to the observer) contained in X, and $H(X|Y)$ is the amount of uncertainty left in X after observing Y. The difference between $H(X)$ and $H(X|Y)$ is the *information* of X obtained by observing Y, and this quantity is defined as the mutual information between X and Y.

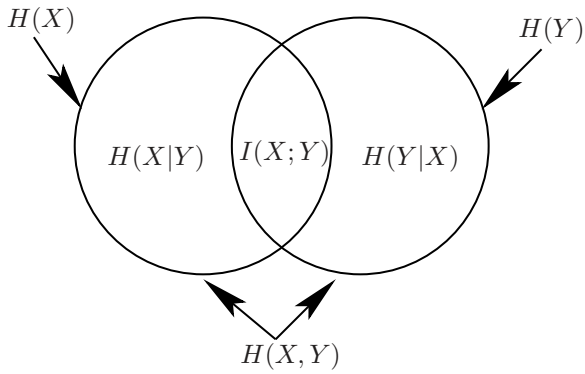


Figure 2.1. Relationship between entropy and mutual information. The two circles correspond to entropy $H(X)$ and $H(Y)$, the mutual information $I(X;Y)$ corresponds to the intersection, and the joint entropy $H(X,Y)$ corresponds to the union.

Definition 2.7. ***Mutual Information***
For discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint pmf $p(x,y)$, the mutual information between X and Y is defined as

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = \sum_{x,y} p(x,y)\log(p(x|y)) - \sum_{x,y} p(x,y)\log(p(x)) \\ &= \sum_{x,y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right). \end{aligned}$$

It is straightforward to verify the following relationship (which is also demonstrated by the famous diagram shown in Figure 2.1)

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y).$$

Similarly, for continuous-valued random variables, we can define differential entropy, joint/conditional differential entropy, and mutual information by replacing the summation with integration and the pmf with corresponding probability density functions (pdf), e.g.,

$$\begin{aligned} h(X) &= -\int_{p(x)>0} p(x)\log(p(x))dx, \\ I(X;Y) &= \iint_{p(x,y)>0} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)dxdy. \end{aligned} \tag{2.8}$$

2.2.3 Cut-set Bound

As shown in Figure 2.2, given a network consisting of m nodes with the channel transition probability $p(y_1,\ldots,y_m|x_1,\ldots,x_m)$, where X_i and Y_i are the transmit-

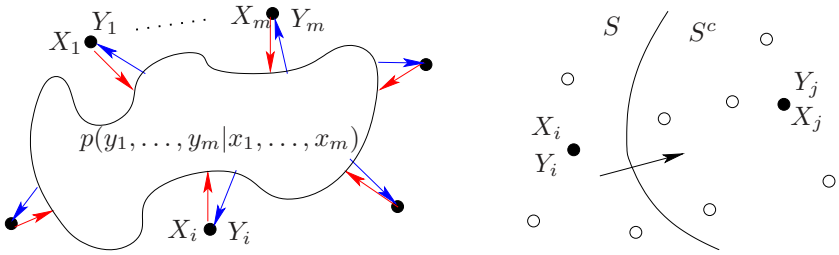


Figure 2.2. On the left is a network consisting of m nodes with the channel transition probability $p(y_1, \ldots, y_m | x_1, \ldots, x_m)$, where X_i and Y_i are the transmitted signal and the received signal by node i, respectively. A cut (S, S^c) that separates the source-destination pair (i, j) is illustrated on the right.

ted signal and the received signal by node i, respectively, we can define various communication tasks over this network. Denoting $R^{(i\to j)}$ as the rate of information transportation from node i to node j, we are interested in the achievable rate region that can be supported by the network. We can partition the set of nodes $T = \{1, 2, \ldots, m\}$ into two disjoint sets, S and its complement S^c, i.e. $S \cap S^c = \emptyset$ and $S \cup S^c = T$, where S contains some source nodes and S^c contains corresponding destination nodes. Such a partition (S, S^c) is called a cut of the network with respect to the corresponding source-destination pairs, as illustrated in Figure 2.2.

By the cut-set theorem [CT06], the accumulated maximum achievable rate from the source nodes contained in S to any of the destinations contained in S^c can be no larger than the minimum of the conditional mutual information flows across all possible cuts (S, S^c), maximized over a joint distribution for the transmit signals. A formal description can be found in [CT06, p.589] and cited below for completeness.

Theorem 2.1. *Denoting $X(S) = \{X_i : i \in S\}$, if $\{R^{(i\to j)}\}$ are achievable, there exists $p(x_1, \ldots, x_m)$ such that*

$$\sum_{i \in S, j \in S^c} R^{(i\to j)} \leq I(X(S); Y(S^c) | X(S^c))$$

for all $S \subset \{1, 2, \ldots, m\}$.

This outer bound is called the cut-set bound, and it has a simple interpretation. The information flow from S to S^c can be no larger than the rate supported by the multiple-input multiple-output (MIMO) channel defined by the cut, i.e., when all the transmitting nodes in S can cooperate and all the receiving nodes in S^c can also cooperate. The cut-set bound is tight in most of the scenarios[4] where the capacity is known, and therefore will be utilized in our work when constructing capacity upper/outer bounds.

[4]The cut-set bound is reported not tight in several scenarios, for example, in [TY11] for a Gaussian interference relay channel where an upper bound tighter than the cut-set bound has been established.

In Theorem 2.1 we assume that all messages with corresponding rate $R^{(i\to j)}$ are independent. For multicast transmission where the same message is transmitted to multiple destinations, we will denote $R^{(i\to D)}$ as the multicast rate from node i to all the receivers in the set D. Then for all $j \in D$, we should replace $R^{(i\to j)}$ by $R^{(i\to D)}$ when constructing the cut-set bound.

2.3 Fundamental Tools

In this section we will introduce some fundamental tools that serve as cornerstones for the cooperative strategies developed in this thesis.

2.3.1 Typical Sequence and Joint Typicality

A sequence $x^{(n)}$ of length n is said to be a typical sequence with respect to a predefined pmf/pdf $p(x)$ if its empirical distribution $\tilde{p}(x)$ is within a predefined distance from the true distribution $p(x)$. Therefore in the definition related to typicality, it is the set of elements rather than the order in a sequence that matters.

A formal definition of typical sequence used in this thesis is listed below.

Definition 2.8. ***Typical Sequence (ϵ-typicality)***
A sequence $x^{(n)}$ is a typical sequence with respect to a pmf/pdf $p(x)$ and a positive constant ϵ if

$$\left| -\frac{1}{n}\log\left(p(x^{(n)})\right) - H(X) \right| < \epsilon, \tag{2.9}$$

where $p(x^{(n)})$ is the probability of the sequence $x^{(n)}$ and $H(X)$ (or $h(X)$) is the (differential) entropy of X.

Note that the empirical entropy $-\frac{1}{n}\log\left(p(x^{(n)})\right)$ converges in probability to the true entropy $H(X)$ (or $h(X)$) as n goes to infinity if all the elements of $x^{(n)}$ are independently and identically distributed according to $p(x)$. There are different ways to measure the distance between two distributions, which leads to a few variations of the definition of typical sequences [Kra07], normally known as letter-typicality, where the empirical distribution of elements in a sequence is within a constant gap (either additive or multiplicative) from the true distribution. The Definition 2.8 is also referred as entropy-typicality in contrast to letter-typicality. For continuous-valued random variables, only entropy-typicality applies. For discrete random variables, both definitions are useful and letter-typicality normally gives stronger results. In the rest of this thesis, we will *not* differentiate these definitions unless stated otherwise.

Denoting $T_\epsilon^{(n)}(X)$ as the set of all typical sequences of length n with respect to distribution $p(x)$ and ϵ, then we can conclude from (2.9) that for any $x^{(n)} \in T_\epsilon^{(n)}(X)$ we have $p(x^{(n)}) \approx 2^{-nH(X)}$, or more precisely,

$$2^{-n(H(X)+\epsilon)} < p(x^{(n)}) < 2^{-n(H(X)-\epsilon)}, \text{ for all } x^{(n)} \in T_\epsilon^{(n)}(X).$$

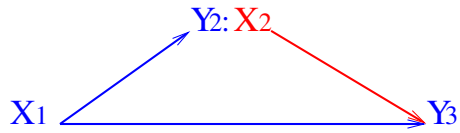


Figure 2.3. A source-relay-destination (S-R-D) network with source node broadcasting X_1 and relay nodes transmitting X_2. Relay receives Y_2 and the destination receives Y_3, with the channel transition probability $p(y_2, y_3|x_1, x_2)$.

Definition 2.9. ***Jointly Typical Sequences***
A pair of sequences $(x^{(n)}, y^{(n)})$ *are jointly typical with respect to a pmf/pdf* $p(x, y)$ *and a positive constant* ϵ*, i.e.* $(x^{(n)}, y^{(n)}) \in T_\epsilon^{(n)}(XY)$*, if*

$$\left|-\frac{1}{n}\log\left(p(x^{(n)})\right) - H(X)\right| < \epsilon,$$
$$\left|-\frac{1}{n}\log\left(p(y^{(n)})\right) - H(Y)\right| < \epsilon,$$
$$\left|-\frac{1}{n}\log\left(p(x^{(n)}, p(y^{(n)})\right) - H(X,Y)\right| < \epsilon.$$

Therefore $(x^{(n)}, y^{(n)}) \in T_\epsilon^{(n)}(XY)$ implies $x^{(n)} \in T_\epsilon^{(n)}(X)$ and $y^{(n)} \in T_\epsilon^{(n)}(Y)$.

Joint typicality will be frequently used throughout this thesis as a key component in designing encoding and decoding schemes, where code books are generated randomly in a memoryless fashion: we create a codebook consisting of 2^{nR} randomly and independently generated codewords $\{x^{(n)}\}$, each of length n, according to the distribution $\prod_{i=1}^{n} p(x_i)$. We assign a codeword $x^{(n)}$ to a message $W \in \{1, 2, \ldots, 2^{nR}\}$ uniformly at random, and associate them via an encoding function $x^{(n)}(W)$, omitting their explicit relationship where appropriate.

2.3.2 Relaying

Two fundamental relaying schemes, decode-and-forward (DF) and compress-and-forward (CF), are first introduced and characterized in [CE79] for the classical source-relay-destination (S-R-D) network as shown in Figure 2.3. In both DF and CF schemes, block Markov encoding is used: message W is evenly divided into B blocks $W_1, \ldots, W_B$, each with nR bits, and the transmission is completed over $B + 1$ blocks, where the transmission in each block (except for the first or the last block) only depends on the current and some previous message blocks. As each transmission is over n channel uses, the overall rate is $\frac{BnR}{(B+1)n}$ bits per channel use, which converges to R when B goes to infinity.

Table 2.1. Illustration of the encoding and decoding process for decode-and-forward (DF) relaying in the S-R-D network. $W_0 = W_{B+1} = 1$ by convention.

$t =$	1	2	$\cdots$	$B+1$
$\mathcal{S}$ TX	$X_{1,1}^{(n)}(W_1, W_0)$	$X_{1,2}^{(n)}(W_2, W_1)$	$\cdots$	$X_{1,B+1}^{(n)}(W_{B+1}, W_B)$
$\mathcal{R}$ RX	W_1	W_2	$\cdots$	/
$\mathcal{R}$ TX	$X_{2,1}^{(n)}(W_0)$	$X_{2,2}^{(n)}(W_1)$	$\cdots$	$X_{2,B+1}^{(n)}(W_B)$
$\mathcal{D}$ RX	/	W_1	$\cdots$	W_B

Decode-and-Forward (DF) Relaying

When DF relaying is utilized, the source-relay link quality is assumed to be better than the source-destination link quality and therefore successful decoding of the transmitted message can be realized at the relay node. Then the transmission at the relay node will cooperate with the source's transmission to help the decoding at the destination node. DF was first proposed in [CE79] where block Markov encoding, superposition coding, and Slepian–Wolf binning [SW73] are used simultaneously for source-relay cooperation and *forward successive decoding* is performed at both the relay and the destination nodes. DF relaying can also be realized by block Markov encoding and superposition coding without binning, if the destination performs *backward decoding* [Car82] or *sliding-window decoding* [Wil82] instead. We will explain the DF scheme by the latter approach with *backward decoding*. A more thorough review of the DF relaying techniques can be found in [KGG05].

The random codebooks are generated as follows. Fix a distribution $p(x_1, x_2)$. For each block $t = 1, \ldots, B+1$, we generate a codebook $\{x_{2,t}^{(n)}\}$ consisting of 2^{nR} codewords generated independently according to $p(x_2)$ and assign the coding index uniformly at random as we previously defined at the end of Section 2.3.1. For each codeword $x_{2,t}^{(n)}(u)$, $u \in \{1, \ldots, 2^{nR}\}$, we generate independently 2^{nR} codewords $\{x_{1,t}^{(n)}(v, u) : v = 1, \ldots, 2^{nR}\}$, with the ith element of $x_{1,t}^{(n)}(v, u)$ generated according to $p(x_1|x_{2,t,i}(u))$, where $x_{2,t,i}(u)$ is the ith element of $x_{2,t}^{(n)}(u)$.

The encoding and decoding process is illustrated in Table 2.1. During block t, the source transmits $X_{1,t}^{(n)}(W_t, W_{t-1})$ where $W_0 = W_{B+1} = 1$ by convention (i.e., no information). Assuming W_{t-1} has been successfully decoded at the relay node after block $t-1$, W_t can be reliably recovered from $Y_{2,t}^{(n)}$, which is a noisy version of $X_{1,t}^{(n)}(W_t, W_{t-1})$, if the information rate R is no larger than the mutual information supported by the source-relay link, i.e.,

$$R < I(X_1; Y_2).$$

During block $t+1$, relay will transmit $X_{2,t+1}^{(n)}(W_t)$. Decoding at the destination is carried out when all the transmission has been finished (*backward decoding*): After block $B+1$, the received signal $Y_{3,B+1}^{(n)}$ only depends on $X_{1,B+1}^{(n)}(1, W_B)$ and

Table 2.2. Illustration of the encoding and decoding process for compress-and-forward (CF) relaying in the S-R-D network. $s_1 = 1$ by convention.

$t=$	1	2	$\cdots$	B	$B+1$
$\mathcal{S}$ TX	$X_{1,1}^{(n)}(W_1)$	$X_{1,2}^{(n)}(W_2)$	$\cdots$	$X_{1,B}^{(n)}(W_B)$	$X_{1,B+1}^{(n)}(1)$
$\mathcal{R}$ TX	$X_{2,1}^{(n)}(s_1)$	$X_{2,2}^{(n)}(s_2)$	$\cdots$	$X_{2,B}^{(n)}(s_B)$	$X_{2,B+1}^{(n)}(s_{B+1})$
$\mathcal{R}$ RX	$\hat{Y}_{2,1}^{(n)}(z_1\|s_1)$	$\hat{Y}_{2,2}^{(n)}(z_2\|s_2)$	$\cdots$	$\hat{Y}_{2,B}^{(n)}(z_B\|s_B)$	/
binning	$s_2 = h(z_1)$	$s_3 = h(z_2)$	$\cdots$	$s_{B+1} = h(z_B)$	/
$\mathcal{D}$ RX	/	$s_2 : z_1 : W_1$	$\cdots$	$s_B : z_{B-1} : W_{B-1}$	$s_{B+1} : z_B : W_B$

$X_{2,B+1}^{(n)}(W_B)$, and therefore successful decoding of W_B can be realized if

$$R < I(X_1X_2; Y_3).$$

Then we can advance the decoding process for W_{B-1}, W_{B-2}, ..., W_1 sequentially as long as the above rate constraint is held. The achievable rate for DF relaying is therefore

$$R < \max_{p(x_1,x_2)} \min\{I(X_1; Y_2),\ I(X_1X_2; Y_3)\}, \tag{2.10}$$

where the maximization is taken over all possible distributions $p(x_1, x_2)$.

Compress-and-Forward (CF) Relaying

The standard compress-and-forward (CF) relaying strategy [CE79] provides the destination node(s) with a noisy yet structured observation (compression) of the received signal at the relay node via the use of an independent codebook. For each block t, we generate independently two random codebooks $\{x_{1,t}^{(n)}\}$ of size 2^{nR} according to $p(x_1)$ and $\{x_{2,t}^{(n)}\}$ of size 2^{nR_2} according to $p(x_2)$. For each codeword $x_{2,t}^{(n)}(v)$, $v = 1, \ldots, 2^{nR_2}$, we generate $2^{nR_2'}$ codewords $\{\hat{y}_{2,t}^{(n)}(w|v) : w = 1, \ldots, 2^{nR_2'}\}$, with the ith element of $\hat{y}_{2,t}^{(n)}(w|v)$ generated according to $p(\hat{y}_2|x_{2,t,i}(v))$.

The encoding and decoding process of the CF strategy is illustrated in Table 2.2. During block t, the source transmits $X_{1,t}^{(n)}(W_t)$ and the relay transmits $X_{2,t}^{(n)}(s_t)$, where $s_t \in \{1, \ldots, 2^{nR_2}\}$ is the bin index determined at the end of block $t-1$, with $s_1 = 1$ by convention. The relay performs compression based on Wyner–Ziv binning [WZ76]: upon receiving $Y_{2,t}$, the relay will find an index $z_t \in \{1, \ldots, 2^{nR_2'}\}$ such that $(\hat{Y}_{2,t}^{(n)}(z_t|s_t), Y_{2,t}^{(n)}, X_{2,t}^{(n)}(s_t))$ are jointly typical, and this happens with high probability if

$$R_2' > I(Y_2; \hat{Y}_2|X_2);$$

z_t is then used to find the bin index via $s_{t+1} = h(z_t)$ where $h(\cdot)$ is a deterministic many-to-one mapping function indicates the binning process. The destination performs block-by-block *forward successive decoding*. It first determines the bin index

$s_t \in \{1, \ldots, 2^{nR_2}\}$ such that $(X_{2,t}^{(n)}(s_t), Y_{3,t}^{(n)})$ are jointly typical, and this can be done reliably if

$$R_2 < I(X_2; Y_3).$$

The destination then determines z_{t-1} via binning: it tries to find $z_{t-1} \in \{1, \ldots, 2^{nR_2'}\}$ such that $(\hat{Y}_{2,t-1}^{(n)}(z_{t-1}|s_{t-1}), Y_{3,t-1}^{(n)}, X_{2,t-1}^{(n)}(s_{t-1}))$ are jointly typical and $s_t = h(z_{t-1})$, and this happens with high probability if

$$R_2' < I(\hat{Y}_2; Y_3|X_2) + R_2.$$

The above decoding via binning process can also be interpreted like this: the bin index s_t provides a blurred description of the possible compression index $z_{t-1} \in \{z : h(z) = s_t\}$, and its previous observation $Y_{3,t-1}^{(n)}$ and relay's transmission $X_{2,t-1}^{(n)}(s_{t-1})$ help to resolve the residual ambiguity about z_{t-1}.

Based on $\hat{Y}_{2,t-1}^{(n)}(z_{t-1}|s_{t-1})$ and $Y_{3,t-1}^{(n)}$, the destination finally decodes the message W_{t-1} reliably via joint typicality decoding, i.e., find $W_{t-1} \in \{1, \ldots, 2^{nR}\}$ such that $(X_{1,t-1}^{(n)}(W_{t-1}), \hat{Y}_{2,t-1}^{(n)}(z_{t-1}|s_{t-1}), Y_{3,t-1}^{(n)}, X_{2,t-1}^{(n)}(s_{t-1}))$ are jointly typical, if

$$R < I(X_1; Y_3\hat{Y}_2|X_2).$$

After cancelling intermediate variables R_2 and R_2', the rate achieved by CF in the classical 3-node S-R-D relay network can be written as [CE79]

$$R < \max I(X_1; Y_3\hat{Y}_2|X_2)$$

subject to

$$I(Y_2; \hat{Y}_2|X_2Y_3) \leq I(X_2; Y_3),$$

where the maximization is over all distributions that can be factorized as

$$p(x_1)p(x_2)p(\hat{y}_2|y_2, x_2)p(y_2, y_3|x_1, x_2).$$

The achievable rate of CF can be written in another equivalent format [EMZ06],

$$R < \max \min[I(X_1; Y_3\hat{Y}_2|X_2), I(X_1X_2; Y_3) - I(Y_2; \hat{Y}_2|X_1X_2Y_3)]. \tag{2.11}$$

2.3.3 Network Coding

The concept of network coding (NC) was first formally introduced and characterized in the seminal paper [ACLY00] where multiple messages are combined into a new message at intermediate nodes and forwarded. One option of such combination is to employ linear operations over finite fields. Given k messages (packets in data network) each of length L bits[5], the linear operation can be carried out over finite field GF(2^q), where q is a positive integer. For each message, we group q consecutive

[5]We can append zeros at the end of shorter messages to ensure the same length.

bits together and treat them as a symbol in GF(2^q), and the resulting vector of symbols will be utilized for encoding. Denote $\boldsymbol{v}_i$, $i = 1, \ldots, k$ as the vectors of symbols, and g_i as the coefficient taken from GF(2^q) for $\boldsymbol{v}_i$, then the network coded vector can be written as

$$\boldsymbol{u} = \sum_{i=1}^{k} g_i \boldsymbol{v}_i,$$

where the summation is carried out in GF(2^q). For the case $q = 1$, the summation is simply the XOR operation. The resulting vector $\boldsymbol{u}$ is then translated back to a vector of bits (NC coded message) and forwarded together with the combination coefficients $[g_1, \ldots, g_k]$. The overhead of transmitting the NC coefficients is negligible if the length of each message is much larger than the number of messages to be transmitted, i.e., $L \gg k$, which is very common in data networks when considering the size of a packet and the small group of packets that are transmitted together.

The destination nodes, upon receiving sufficient number of linearly independent combinations of the transmitted messages, can recover the original messages by inverting the coefficient matrix. These coefficients can be deliberately chosen to minimize the number of received NC messages for successful decoding, but such design becomes prohibitively difficult when the size of the network becomes large. Fortunately, as suggested in [HMK$^+$06], if these coefficients are determined in a random and distributed fashion, successful decoding can be realized with high probability if the field size 2^q is sufficiently large.

In wireless communications where signals from neighboring transmitters are combined together at the receiver by nature, network coding can also be done in the analog domain, i.e., directly based on the received signal. Since the mapping from multiple messages onto one message of the same length implicitly requires compression, coding with structures has to be used when transmitting individual messages. For example, nested lattice codes can be used for this purpose [WNPS10]. Such compression can also be done in an implicit manner. For example, the received signal at intermediate nodes can be quantized and then re-encoded [ADT11, LKEC11] or simply be scaled [KGK07] before forwarding to subsequent nodes.

2.4 Summary

In this chapter, we have presented the system models and provided justifications for the assumptions on perfect synchronization, perfect CSIT, and full-duplex relaying. We have also presented some principal definitions that will be frequently used in this thesis. We have also briefly reviewed some fundamental tools namely, the joint typicality, relaying, and network coding, that will serve as cornerstone in subsequent chapters when discussing cooperative strategies.

Chapter 3

Capacity Outer Bounds

In this chapter, we investigate capacity outer bounds for the wireless multicast relay network shown in Figure 1.4, where two sources simultaneously multicast to two destinations with the help of a shared full-duplex relay node. The two sources and the relay use the same channel resource (i.e. co-channel transmission). Furthermore, we assume that the two source nodes are connected by orthogonal error-free conferencing links. As illustrated in Chapter 1, in connection to Figure 1.1, such a connection can be realized for example by utilizing the existing fiber or microwave backhaul that connects the base stations in a cellular networks. Obviously, if a connection exists between the base stations, it will be possible to use it to improve performance. This will be illustrated in quantitative terms here and subsequently in Chapters 4 and 5.

As shown in Figure 1.4, we assume that all the individual channel gains $g_{ij} \geq 0$, $i, j = 1, 2, r$ are time-invariant and known to every node in the network. The scenario of only local/partial CSI, requiring a trade-off between message and CSI exchange as demonstrated in [ZG11, Ray06, WBBJ11], is left to future work. Given an average transmit power constraint P_i, fixed channel gain g_{ij}, and noise power N_j, we can characterize the transmission link by only its signal-to-noise ratio (SNR) $\gamma_{ij} = P_i g_{ij}^2 / N_j$, as illustrated in Figure 3.1, without distinguishing the origin of various SNR contributors. The system shown in Figure 3.1 can be modelled as follows

$$\begin{aligned} Y_1^{(n)} &= \sqrt{\gamma_{11}} X_1^{(n)} + \sqrt{\gamma_{21}} X_2^{(n)} \quad + \sqrt{\gamma_{r1}} X_r^{(n)} + Z_1^{(n)}, \\ Y_2^{(n)} &= \sqrt{\gamma_{12}} X_1^{(n)} + \sqrt{\gamma_{22}} X_2^{(n)} \quad + \sqrt{\gamma_{r2}} X_r^{(n)} + Z_2^{(n)}, \\ Y_r^{(n)} &= \sqrt{\gamma_{1r}} X_1^{(n)} + \sqrt{\gamma_{2r}} X_2^{(n)} \quad + Z_r^{(n)}, \end{aligned} \tag{3.1}$$

where the noise components $Z_{i,k}$, $i = 1, 2, r$ and $k = 1, ..., n$ are i.i.d. Gaussian with zero-mean unit-variance, and all the transmitted signals $X_i^{(n)}$ are subject to average unit-power constraints, i.e.,

$$\frac{1}{n} \sum_{k=1}^{n} X_{i,k}^2 \leq 1. \tag{3.2}$$

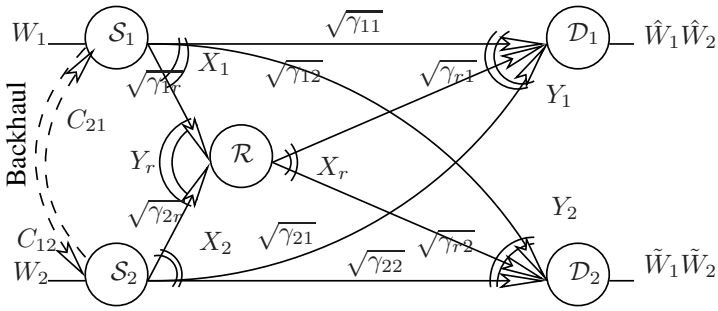


Figure 3.1. Two source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with orthogonal and error-free backhaul (with rate C_{12} and C_{12} bits per channel use), multicast information W_1 at rate R_1 and W_2 at rate R_2 respectively to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ through Gaussian channels, with aid from a full-duplex relay $\mathcal{R}$. $\gamma_{ij}\geq 0, i, j=1, 2, r$ are the effective link SNR.

3.1 Cut-set Bounds

Proposition 3.1. *The cut-set bound for the multicast network in Figure 3.1 can be characterized by*

$$
\begin{aligned}
C_{cut\text{-}set} = \bigcup_{p(x_1,x_2,x_r)} \Big\{ & (R_1, R_2) : R_1 \geq 0,\ R_2 \geq 0, \qquad (3.3)\\
& R_1 \leq C_{12} + \frac{1}{n} \min_{d\in\{1,2\}} \Big\{ I(X_1^{(n)}; Y_d^{(n)} Y_r^{(n)} | X_2^{(n)} X_r^{(n)} X_s^{(n)}),\\
& \qquad\qquad I(X_1^{(n)} X_r^{(n)}; Y_d^{(n)} | X_2^{(n)} X_s^{(n)}) \Big\} + \epsilon_n,\\
& R_2 \leq C_{21} + \frac{1}{n} \min_{d\in\{1,2\}} \Big\{ I(X_2^{(n)}; Y_d^{(n)} Y_r^{(n)} | X_1^{(n)} X_r^{(n)} X_s^{(n)}),\\
& \qquad\qquad I(X_2^{(n)} X_r^{(n)}; Y_d^{(n)} | X_1^{(n)} X_s^{(n)}) \Big\} + \epsilon_n,\\
& R_1 + R_2 \leq \frac{1}{n} \min_{d\in\{1,2\}} \Big\{ I(X_1^{(n)} X_2^{(n)}; Y_d^{(n)} Y_r^{(n)} | X_r^{(n)}),\\
& \qquad\qquad I(X_1^{(n)} X_2^{(n)} X_r^{(n)}; Y_d^{(n)}) \Big\} + \epsilon_n,\\
& R_1 + R_2 \leq C_{12} + C_{21} + \frac{1}{n} \min_{d\in\{1,2\}} \Big\{ I(X_1^{(n)} X_2^{(n)}; Y_d^{(n)} Y_r^{(n)} | X_r^{(n)} X_s^{(n)}),\\
& \qquad\qquad I(X_1^{(n)} X_2^{(n)} X_r^{(n)}; Y_d^{(n)} | X_s^{(n)}) \Big\} + \epsilon_n \Big\},
\end{aligned}
$$

where $X_s^{(n)}$ represent symbols transmitted via the conferencing links, $X_1^{(n)}$, $X_2^{(n)}$ and $X_r^{(n)}$ are subject to the average power constraint (3.2)*, $\epsilon_n \to 0$ as $n\to\infty$, and*

the joint probability is partitioned as

$$\prod p(x_s, x_r)p(x_1|x_s, x_r)p(x_2|x_s, x_r)p(y_r|x_1, x_2)p(y_1|x_1, x_2, x_r)p(y_2|x_1, x_2, x_r).$$

Note that the rate constraints in (3.3) are for multicast transmission and therefore the minimization is also taken over all the destination nodes.

Proof. By Theorem 2.1, the maximum achievable rate from the source nodes to any of the destinations can be no larger than the minimum of the (conditional) mutual information flows across all possible cuts, maximized over a joint distribution for the transmitted signals. We can therefore apply Theorem 2.1 directly to the multicast network in Figure 3.1 and evaluate all the possible cuts. By taking into account the power constraint [Wil83] and the potential correlation between X_1, X_2 and X_r, we can get (3.3). □

For the multicast relay network with high-rate backhaul, we find the exact cut-set bound of the capacity region by extending the proof of the converse developed by Cover and El Gamal [CE79] for the Gaussian relay channel. For low-rate backhaul scenarios, we present two genie-aided outer bounds by extending the previous proof and introducing two lemmas on conditional (co-)variance.

3.2 Exact Cut-set Bounds for High-rate Backhaul

Define a rate region (R_1, R_2) such that $R_1 \geq 0$, $R_2 \geq 0$, and

$$\begin{aligned} R_1 + R_2 \leq C_0 = \sup_{0\leq\alpha_1,\alpha_2,\rho\leq 1} \min\Big\{ & \\ & \mathcal{C}\left((\gamma_{11}+\gamma_{1r})\alpha_1+(1-\rho^2)\alpha_1\alpha_2(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\right. \\ & \quad \left.+(\gamma_{21}+\gamma_{2r})\alpha_2 + 2\rho\sqrt{\alpha_1\alpha_2}(\sqrt{\gamma_{11}\gamma_{21}}+\sqrt{\gamma_{1r}\gamma_{2r}})\right), \\ & \mathcal{C}\left((\gamma_{12}+\gamma_{1r})\alpha_1+(1-\rho^2)\alpha_1\alpha_2(\sqrt{\gamma_{12}\gamma_{2r}}-\sqrt{\gamma_{22}\gamma_{1r}})^2\right. \\ & \quad \left.+(\gamma_{22}+\gamma_{2r})\alpha_2 + 2\rho\sqrt{\alpha_1\alpha_2}(\sqrt{\gamma_{12}\gamma_{22}}+\sqrt{\gamma_{1r}\gamma_{2r}})\right), \\ & \mathcal{C}\Big(\gamma_{11} + \gamma_{21} + \gamma_{r1} + 2\sqrt{\bar{\alpha}_1\gamma_{11}\gamma_{r1}} + 2\sqrt{\bar{\alpha}_2\gamma_{21}\gamma_{r1}} \\ & \qquad +2(\rho\sqrt{\alpha_1\alpha_2} + \sqrt{\bar{\alpha}_1\bar{\alpha}_2})\sqrt{\gamma_{11}\gamma_{21}}\Big), \\ & \mathcal{C}\Big(\gamma_{12} + \gamma_{22} + \gamma_{r2} + 2\sqrt{\bar{\alpha}_1\gamma_{12}\gamma_{r2}} + 2\sqrt{\bar{\alpha}_2\gamma_{22}\gamma_{r2}} \\ & \qquad +2(\rho\sqrt{\alpha_1\alpha_2} + \sqrt{\bar{\alpha}_1\bar{\alpha}_2})\sqrt{\gamma_{12}\gamma_{22}}\Big)\Big\}, \end{aligned} \tag{3.4}$$

where $\bar{\alpha}_1$=1−α_1, $\bar{\alpha}_2$=1−α_2, and $\mathcal{C}(\cdot)$ is the Gaussian capacity function defined as

$$\mathcal{C}(x) = \max\left\{\frac{1}{2}\log(1+x),\ 0\right\}.$$

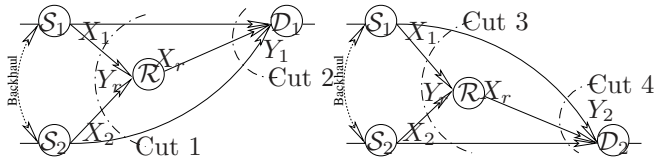


Figure 3.2. The capacity with high-rate backhaul is bounded by the cut-set bound based on the four cuts shown in the figure.

By extending the proof of the converse developed by Cover and El Gamal [CE79] for the Gaussian relay channel, we have characterized the exact cut-set bound for a multicast relay network supported by a high-rate backhaul (i.e., $C_{12}\geq R_1$ and $C_{21}\geq R_2$) as follows.

Theorem 3.1. *For a high-rate backhaul (i.e., $C_{12}\geq R_1$ and $C_{21}\geq R_2$), the cut-set bound $C_{cut\text{-}set}$ in Proposition 3.1 equals to C_0.*

3.2.1 Proof of the Exact Cut-set Bound $C_{\text{cut-set}}$

By Theorem 2.1, the maximum achievable sum rate from the source nodes to any of the destinations can be no larger than the minimum of the mutual information flows across all possible cuts, maximized over a joint distribution for the transmitted signals. For a high-rate backhaul (i.e., $C_{12}\geq R_1$ and $C_{21}\geq R_2$), some of the constraints in (3.3) become redundant and the expression for the upper bound can be reduced to

$$\begin{aligned} R_1+R_2 \leq C_{\text{cut-set}} = \sup_{p(x_1,x_2,x_r)} \min \Bigg\{ & \\ & \frac{1}{n}I(X_1^{(n)},X_2^{(n)};Y_1^{(n)},Y_r^{(n)}|X_r^{(n)}), \frac{1}{n}I(X_1^{(n)},X_2^{(n)},X_r^{(n)};Y_1^{(n)}), \\ & \frac{1}{n}I(X_1^{(n)},X_2^{(n)};Y_2^{(n)},Y_r^{(n)}|X_r^{(n)}), \frac{1}{n}I(X_1^{(n)},X_2^{(n)},X_r^{(n)};Y_2^{(n)}) \Bigg\} + \epsilon_n, \end{aligned} \tag{3.5}$$

where $\epsilon_n\to 0$ as $n\to\infty$, $X_1^{(n)}$, $X_2^{(n)}$ and $X_r^{(n)}$ are potentially correlated. Note that for a high-rate backhaul, only the cut-set bound derived based on four cuts between the two sources and each of the sink are active, as shown in Figure 3.2.

As suggested in [EK10, chp. 17], to find the exact cut-set bound $C_{\text{cut-set}}$, we will first find an upper bound $C_{\text{upp}} \geq C_{\text{cut-set}}$ based on the technique used in [CE79], and then find a lower bound $C_{\text{cut-set, G}} \leq C_{\text{cut-set}}$ by restricting the source distribution to Gaussian, and finally by showing that $C_{\text{cut-set, G}} = C_{\text{upp}}$ we can find the exact cut-set bound $C_{\text{cut-set}} = C_{\text{upp}} = C_{\text{cut-set, G}}$.

The Upper Bound $C_{\mathbf{upp}}$

Following the conventional notation for the differential entropy as defined in (2.8), the mutual information corresponding to cut 2 can be written as

$$\begin{aligned} I(X_1^{(n)}, X_2^{(n)}, X_r^{(n)}; Y_1^{(n)}) &= h(Y_1^{(n)}) - h(Y_1^{(n)}|X_1^{(n)}, X_2^{(n)}, X_r^{(n)}) \\ &= h(Y_1^{(n)}) - h(Z_1^{(n)}) = h(Y_1^{(n)}) - \frac{n}{2}\log(2\pi e). \end{aligned} \tag{3.6}$$

From the maximum entropy lemma [CT06], we get

$$h(Y_1^{(n)}) \leq \sum_{i=1}^{n} h(Y_{1,i}) \leq \sum_{i=1}^{n} \frac{1}{2}\log(2\pi e \mathrm{Var}[Y_{1,i}]), \tag{3.7}$$

where the second equality is achieved when $Y_{1,i}$ is Gaussian distributed. Hence

$$\begin{aligned} \frac{1}{n} I(X_1^{(n)}, X_2^{(n)}, X_r^{(n)}; Y_1^{(n)}) &\leq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log(\mathrm{Var}[Y_{1,i}]) \\ &\leq \frac{1}{2}\log(\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}[Y_{1,i}]), \end{aligned} \tag{3.8}$$

where the last steps follow from Jensen's inequality, with

$$\mathrm{Var}[Y_{1,i}] = 1 + \mathrm{Var}[\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i} + \sqrt{\gamma_{r1}}X_{r,i}]. \tag{3.9}$$

According to the *law of total variance*, for two random variables X and Y on the same probability space, and if the variance of X is finite, then

$$\mathrm{Var}[X] = E(\mathrm{Var}[X|Y]) + \mathrm{Var}[E(X|Y)].$$

We can therefore rewrite (3.9) as

$$\begin{aligned} \mathrm{Var}[Y_{1,i}] &= 1 + \mathrm{Var}[E(\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i} + \sqrt{\gamma_{r1}}X_{r,i}|X_{r,i})] \\ &\quad + E(\mathrm{Var}[\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i} + \sqrt{\gamma_{r1}}X_{r,i}|X_{r,i}]) \\ &= 1 + \mathrm{Var}[\sqrt{\gamma_{11}}E(X_{1,i}|X_{r,i}) + \sqrt{\gamma_{21}}E(X_{2,i}|X_{r,i}) + \sqrt{\gamma_{r1}}X_{r,i}], \\ &\quad + E(\mathrm{Var}[\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i}|X_{r,i}]) \end{aligned} \tag{3.10}$$

where

$$\begin{aligned} &E(\mathrm{Var}[\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i}|X_{r,i}]) \\ &= E(\gamma_{11}\mathrm{Var}[X_{1,i}|X_{r,i}] + \gamma_{21}\mathrm{Var}[X_{2,i}|X_{r,i}] + 2\sqrt{\gamma_{11}\gamma_{21}}\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})) \\ &= \gamma_{11}E(\mathrm{Var}[X_{1,i}|X_{r,i}]) + \gamma_{21}E(\mathrm{Var}[X_{2,i}|X_{r,i}]) + 2\sqrt{\gamma_{11}\gamma_{21}}E(\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})), \end{aligned} \tag{3.11}$$

and

$$\begin{aligned}
&\mathrm{Var}[\sqrt{\gamma_{11}}E(X_{1,i}|X_{r,i}) + \sqrt{\gamma_{21}}E(X_{2,i}|X_{r,i}) + \sqrt{\gamma_{r1}}X_{r,i}] \\
&\quad \le E[(\sqrt{\gamma_{11}}E[X_{1,i}|X_{r,i}] + \sqrt{\gamma_{21}}E[X_{2,i}|X_{r,i}] + \sqrt{\gamma_{r1}}X_{r,i})^2] \\
&\quad = \gamma_{11}E(E^2[X_{1,i}|X_{r,i}]) + 2\sqrt{\gamma_{11}\gamma_{r1}}E(X_{r,i}E[X_{1,i}|X_{r,i}]) \\
&\qquad + \gamma_{21}E(E^2[X_{2,i}|X_{r,i}]) + 2\sqrt{\gamma_{21}\gamma_{r1}}E(X_{r,i}E[X_{2,i}|X_{r,i}]) \\
&\qquad + \gamma_{r1}E(X_{r,i}^2) + 2\sqrt{\gamma_{11}\gamma_{21}}E(E[X_{1,i}|X_{r,i}]E[X_{2,i}|X_{r,i}]).
\end{aligned} \tag{3.12}$$

As in [CE79], define

$$\bar{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{1,i}|X_{r,i})],\ \alpha_1 \in [0,1], \tag{3.13}$$

then we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{r,i})] &= \frac{1}{n}\sum_{i=1}^{n} E[E(X_{1,i}^2|X_{r,i}) - E^2(X_{1,i}|X_{r,i})] \\
&= \frac{1}{n}\sum_{i=1}^{n}(E[X_{1,i}^2] - E[E^2(X_{1,i}|X_{r,i})]) \\
&= \frac{1}{n}\sum_{i=1}^{n} E[X_{1,i}^2] - \frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{1,i}|X_{r,i})] \\
&\le \alpha_1,
\end{aligned} \tag{3.14}$$

where the inequality comes from (3.2). Similarly we have

$$\begin{aligned}
\bar{\alpha}_2 &= \frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{2,i}|X_{r,i})], \\
\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{r,i})] &\le \alpha_2,
\end{aligned} \tag{3.15}$$

where $\alpha_2 \in [0,1]$. On the other hand, as

$$\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i}) = \phi_i\sqrt{\mathrm{Var}(X_{1,i}|X_{r,i})\mathrm{Var}(X_{2,i}|X_{r,i})},$$

where $|\phi_i| \le 1$ is the correlation coefficient, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})] \\
&\le \sqrt{\frac{1}{n}\sum_{i=1}^{n} \phi_i E[\mathrm{Var}(X_{1,i}|X_{r,i})]\frac{1}{n}\sum_{i=1}^{n} \phi_i E[\mathrm{Var}(X_{2,i}|X_{r,i})]} \\
&\le \sqrt{\alpha_1\alpha_2},
\end{aligned} \tag{3.16}$$

where the first inequality is due to the Cauchy–Schwarz inequality and the last step is given by (3.14) and (3.15). Given that $|\phi_i| \leq 1$, we can introduce an auxiliary variable $0 \leq \rho \leq 1$ such that

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})] = \rho\sqrt{\alpha_1\alpha_2}. \tag{3.17}$$

Also, using the Cauchy–Schwarz inequality we get

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E(X_{r,i}E[X_{1,i}|X_{r,i}]) \\
&\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n} E[X_{r,i}^2]\frac{1}{n}\sum_{i=1}^{n} E(E^2[X_{1,i}|X_{r,i}])} \leq \sqrt{\bar{\alpha}_1}, \qquad (3.18)\\
&\frac{1}{n}\sum_{i=1}^{n} E(X_{r,i}E[X_{2,i}|X_{r,i}]) \leq \sqrt{\bar{\alpha}_2}, \qquad (3.19)\\
&\frac{1}{n}\sum_{i=1}^{n} E(E[X_{1,i}|X_{r,i}]E[X_{2,i}|X_{r,i}]) \leq \sqrt{\bar{\alpha}_1\bar{\alpha}_2}. \qquad (3.20)
\end{aligned}$$

Now, substituting (3.10)–(3.20) into (3.8), and applying the same approach also to cut 4, we get

$$\begin{aligned}
\frac{1}{n}I(X_1^{(n)}, X_2^{(n)}, X_r^{(n)}; Y_1^{(n)}) \leq \mathcal{C}(&\gamma_{11} + \gamma_{21} + \gamma_{r1} + 2\sqrt{\bar{\alpha}_1\gamma_{11}\gamma_{r1}} + 2\sqrt{\bar{\alpha}_2\gamma_{21}\gamma_{r1}} \\
&+2(\rho\sqrt{\alpha_1\alpha_2} + \sqrt{\bar{\alpha}_1\bar{\alpha}_2})\sqrt{\gamma_{11}\gamma_{21}}), \qquad (3.21)\\
\frac{1}{n}I(X_1^{(n)}, X_2^{(n)}, X_r^{(n)}; Y_2^{(n)}) \leq \mathcal{C}(&\gamma_{12} + \gamma_{22} + \gamma_{r2} + 2\sqrt{\bar{\alpha}_1\gamma_{12}\gamma_{r2}} + 2\sqrt{\bar{\alpha}_2\gamma_{22}\gamma_{r2}} \\
&+2(\rho\sqrt{\alpha_1\alpha_2} + \sqrt{\bar{\alpha}_1\bar{\alpha}_2})\sqrt{\gamma_{12}\gamma_{22}}). \qquad (3.22)
\end{aligned}$$

For cut 1 we have

$$\begin{aligned}
&I(X_1^{(n)}, X_2^{(n)}; Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) \\
&= h(Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) - h(Y_1^{(n)}, Y_r^{(n)}|X_1^{(n)}, X_2^{(n)}, X_r^{(n)}) \\
&= h(Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) - h(Y_1^{(n)}|X_1^{(n)}, X_2^{(n)}, X_r^{(n)}) - h(Y_r^{(n)}|X_1^{(n)}, X_2^{(n)}, X_r^{(n)}) \\
&= h(Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) - h(Z_1^{(n)}) - h(Z_r^{(n)}) \\
&= h(Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) - n\log(2\pi e) \\
&\leq \frac{1}{2}\sum_{i=1}^{n} \log\left((2\pi e)^2|\boldsymbol{K}_i|\right) - n\log(2\pi e) \\
&= \frac{1}{2}\sum_{i=1}^{n} \log\left(|\boldsymbol{K}_i|\right),
\end{aligned} \tag{3.23}$$

where the second equality in (3.23) comes from the fact that $Y_1^{(n)}$ and $Y_r^{(n)}$ are independent given $(X_1^{(n)}, X_2^{(n)}, X_r^{(n)})$ and the inequality is due to the maximum entropy lemma [CT06], with equality achieved by jointly Gaussian distributed $(Y_{1,i}, Y_{r,i})$ with conditional covariance matrices $\boldsymbol{K}_i$, which are defined by

$$\boldsymbol{K}_i = \begin{bmatrix} E(\mathrm{Var}[Y_{1,i}|X_{r,i}]) & E[\mathrm{Cov}(Y_{1,i}, Y_{r,i}|X_{r,i})] \\ E[\mathrm{Cov}(Y_{1,i}, Y_{r,i}|X_{r,i})] & E(\mathrm{Var}[Y_{r,i}|X_{r,i}]) \end{bmatrix}.$$

Obviously, the covariance matrices $\boldsymbol{K}_i$ are positive semi-definite. Since the function $\log|\boldsymbol{K}|$ is concave [CT98], we can thus bound the throughput of cut 1 from (3.23) as follows

$$\begin{aligned} &\frac{1}{n} I(X_1^{(n)}, X_2^{(n)}; Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) \\ &\le \frac{1}{2}\log\left(\left|\tfrac{1}{n}\textstyle\sum_{i=1}^n \boldsymbol{K}_i\right|\right) \\ &= \frac{1}{2}\log\Big(\tfrac{1}{n}\textstyle\sum_{i=1}^n E(\mathrm{Var}[Y_{1,i}|X_{r,i}])\tfrac{1}{n}\sum_{j=1}^n E(\mathrm{Var}[Y_{r,j}|X_{r,j}]) \\ &\qquad - \left(\tfrac{1}{n}\textstyle\sum_{i=1}^n E[\mathrm{Cov}(Y_{1,i}, Y_{r,i}|X_{r,i})]\right)^2\Big). \end{aligned} \tag{3.24}$$

Furthermore, since

$$\begin{aligned} E(\mathrm{Var}[Y_{1,i}|X_{r,i}]) =& 1 + E(\mathrm{Var}[\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{21}}X_{2,i}|X_{r,i}]), \\ E(\mathrm{Var}[Y_{r,i}|X_{r,i}]) =& 1 + E(\mathrm{Var}[\sqrt{\gamma_{1r}}X_{1,i} + \sqrt{\gamma_{2r}}X_{2,i}|X_{r,i}]), \\ E[\mathrm{Cov}(Y_{1,i}, Y_{r,i}|X_{r,i})] =& \sqrt{\gamma_{11}\gamma_{1r}}E(\mathrm{Var}[X_{1,i}|X_{r,i}]) + \sqrt{\gamma_{21}\gamma_{2r}}E(\mathrm{Var}[X_{2,i}|X_{r,i}]) \\ &+ (\sqrt{\gamma_{11}\gamma_{2r}} + \sqrt{\gamma_{21}\gamma_{1r}})E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})], \end{aligned}$$

by combining with (3.11) and (3.14)–(3.17), we can conclude that

$$\begin{aligned} &\frac{1}{n} I(X_1^{(n)}, X_2^{(n)}; Y_1^{(n)}, Y_r^{(n)}|X_r^{(n)}) \\ &\le \mathcal{C}\left((\gamma_{11}+\gamma_{1r})\alpha_1 + (1-\rho^2)\alpha_1\alpha_2(\sqrt{\gamma_{11}\gamma_{2r}} - \sqrt{\gamma_{21}\gamma_{1r}})^2\right. \\ &\qquad \left. +(\gamma_{21}+\gamma_{2r})\alpha_2 + 2\rho\sqrt{\alpha_1\alpha_2}(\sqrt{\gamma_{11}\gamma_{21}} + \sqrt{\gamma_{1r}\gamma_{2r}})\right). \end{aligned} \tag{3.25}$$

Similarly, we can bound the throughput of cut 3 as follows

$$\begin{aligned} &\frac{1}{n} I(X_1^{(n)}, X_2^{(n)}; Y_2^{(n)}, Y_r^{(n)}|X_r^{(n)}) \\ &\le \mathcal{C}\left((\gamma_{12}+\gamma_{1r})\alpha_1 + (1-\rho^2)\alpha_1\alpha_2(\sqrt{\gamma_{12}\gamma_{2r}} - \sqrt{\gamma_{22}\gamma_{1r}})^2\right. \\ &\qquad \left. +(\gamma_{22}+\gamma_{2r})\alpha_2 + 2\rho\sqrt{\alpha_1\alpha_2}(\sqrt{\gamma_{12}\gamma_{22}} + \sqrt{\gamma_{1r}\gamma_{2r}})\right). \end{aligned} \tag{3.26}$$

By substituting (3.21) (3.22) (3.25) (3.26) into (3.5) and letting $n\to\infty$, comparing the resulting region to (3.4) we can conclude that

$$C_{\text{cut-set}} \le C_{\text{upp}} = C_0.$$

The Lower Bound $C_{\text{cut-set, G}}$

By restricting $p(x_1, x_2, x_r)$ in (3.5) to be Gaussian, we can partition $X_1^{(n)}$, $X_2^{(n)}$ and $X_r^{(n)}$ as follows

$$X_r^{(n)} = U^{(n)}, \tag{3.27a}$$
$$X_1^{(n)} = \sqrt{(1-\rho)\alpha_1}S_1^{(n)} + \sqrt{\rho\alpha_1}V^{(n)} + \sqrt{(1-\alpha_1)}U^{(n)}, \tag{3.27b}$$
$$X_2^{(n)} = \sqrt{(1-\rho)\alpha_2}S_2^{(n)} + \sqrt{\rho\alpha_2}V^{(n)} + \sqrt{(1-\alpha_2)}U^{(n)}, \tag{3.27c}$$

where $S_1^{(n)}, S_2^{(n)}, V^{(n)}, U^{(n)}$ are n-dimensional independent Gaussian random vectors with zero-mean and unit-variance. Auxiliary variables $0\leq\alpha_1, \alpha_2, \rho\leq 1$ are introduced to represent the potential correlation among $X_1^{(n)}, X_2^{(n)}$ and $X_r^{(n)}$ due to cooperation. by substituting (3.27) into (3.1), we can derive from (3.5) that

$$\begin{aligned} C_{\text{cut-set}} \geq C_{\text{cut-set, G}} = \sup_{0\leq\alpha_1,\alpha_2,\rho\leq 1} \min \frac{1}{2n}\sum_{i=1}^{n}\{ \\ \log(\text{Var}[Y_{1,i}]), \log(\text{Var}[Y_{2,i}]), \log\left(|\boldsymbol{K}_{1,i}|\right), \log\left(|\boldsymbol{K}_{2,i}|\right)\} + \epsilon_n, \end{aligned} \tag{3.28}$$

where $\epsilon_n\to 0$ as $n\to\infty$, and for $i = 1, ..., n$, we have

$$\begin{aligned} \text{Var}[Y_{1,i}] &= 1 + \bar{\rho}\alpha_1\gamma_{11} + \bar{\rho}\alpha_2\gamma_{21} + \rho(\sqrt{\alpha_1\gamma_{11}} + \sqrt{\alpha_2\gamma_{21}})^2 \\ &\quad + (\sqrt{\bar{\alpha}_1\gamma_{11}} + \sqrt{\bar{\alpha}_2\gamma_{21}} + \sqrt{\gamma_{r1}})^2, \\ \text{Var}[Y_{2,i}] &= 1 + \bar{\rho}\alpha_1\gamma_{12} + \bar{\rho}\alpha_2\gamma_{22} + \rho(\sqrt{\alpha_1\gamma_{12}} + \sqrt{\alpha_2\gamma_{22}})^2 \\ &\quad + (\sqrt{\bar{\alpha}_1\gamma_{12}} + \sqrt{\bar{\alpha}_2\gamma_{22}} + \sqrt{\gamma_{r2}})^2, \\ |\boldsymbol{K}_{1,i}| &= (1 + \bar{\rho}\alpha_1\gamma_{11} + \bar{\rho}\alpha_2\gamma_{21} + \rho(\sqrt{\alpha_1\gamma_{11}} + \sqrt{\alpha_2\gamma_{21}})^2) \\ &\quad \times (1 + \bar{\rho}\alpha_1\gamma_{1r} + \bar{\rho}\alpha_2\gamma_{2r} + \rho(\sqrt{\alpha_1\gamma_{1r}} + \sqrt{\alpha_2\gamma_{2r}})^2) \\ &\quad -(\sqrt{\gamma_{11}\gamma_{1r}\alpha_1} + \sqrt{\gamma_{21}\gamma_{2r}\alpha_2} + (\sqrt{\gamma_{11}\gamma_{2r}} + \sqrt{\gamma_{21}\gamma_{1r}})\rho\sqrt{\alpha_1\alpha_2})^2, \\ |\boldsymbol{K}_{2,i}| &= (1 + \bar{\rho}\alpha_1\gamma_{12} + \bar{\rho}\alpha_2\gamma_{22} + \rho(\sqrt{\alpha_1\gamma_{12}} + \sqrt{\alpha_2\gamma_{22}})^2) \\ &\quad \times (1 + \bar{\rho}\alpha_1\gamma_{1r} + \bar{\rho}\alpha_2\gamma_{2r} + \rho(\sqrt{\alpha_1\gamma_{1r}} + \sqrt{\alpha_2\gamma_{2r}})^2) \\ &\quad -(\sqrt{\gamma_{12}\gamma_{1r}\alpha_1} + \sqrt{\gamma_{22}\gamma_{2r}\alpha_2} + (\sqrt{\gamma_{12}\gamma_{2r}} + \sqrt{\gamma_{22}\gamma_{1r}})\rho\sqrt{\alpha_1\alpha_2})^2. \end{aligned} \tag{3.29}$$

By substituting (3.29) into (3.28), and letting $n\to\infty$, comparing the resulting region to (3.4) we can conclude that

$$C_{\text{cut-set}} \geq C_{\text{cut-set, G}} = C_0.$$

Recall that $C_{\text{cut-set, G}} \leq C_{\text{cut-set}} \leq C_{\text{upp}}$, we can finally conclude that $C_{\text{cut-set}} = C_0$, i.e., Theorem 3.1 holds.

3.3 Genie-aided Cut-set Bounds for Low-rate Backhaul

By extending the proof of the converse developed by Cover and El Gamal [CE79] for the Gaussian relay channel, we have characterized the exact cut-set bound for high-rate backhaul (i.e., $C_{12}{\geq}R_1$ and $C_{21}{\geq}R_2$) in Section 3.2.1. However, it is difficult to directly apply that result to the low-rate backhaul scenarios (i.e., $C_{12}{<}R_1$ and $C_{21}{<}R_2$) since the transmitted signal at the relay is only partially known to both source nodes owing to the limited-rate conferencing links. Instead, we introduce a genie which tells the two source nodes what exactly the relay is going to transmit, i.e., $X_r^{(n)}$ is known at $\mathcal{S}_1$ and $\mathcal{S}_2$ non-causally. Therefore $X_r^{(n)}$ needs not to be transmitted via the conferencing links, i.e., the conferencing symbols $X_s^{(n)}$ are independent of $X_r^{(n)}$, which indicates that $p(x_r,x_s){=}p(x_r)p(x_s)$ is sufficient for the probability partition in Proposition 3.1. Since $X_1^{(n)}$ is potentially correlated to $X_r^{(n)}$ and $X_s^{(n)}$, we can introduce two independent auxiliary variables $\alpha_1,\rho_1\in[0,1]$ to indicate the dependence of $X_1^{(n)}$ on $X_r^{(n)}$ (via $\bar{\alpha}_1{=}1{-}\alpha_1$) and on $X_s^{(n)}$ (via $\rho_1\alpha_1$). Similarly, $\alpha_2,\rho_2\in[0,1]$ are introduced for $X_2^{(n)}$. Following similar procedures as in [DXS11b, DXS11a], we can bound all the mutual information terms in (3.3) and obtain the following outer bound.

Proposition 3.2. *The cut-set bound $C_{cut\text{-}set}$ in Proposition 3.1 can be outer bounded by*

$$
\begin{aligned}
C_{upp1} = \bigcup_{0\leq\alpha_1,\alpha_2,\rho_1,\rho_2\leq 1} \Big\{ & (R_1,R_2):\ R_1\geq 0,\ R_2\geq 0, \qquad (3.30)\\
& R_1 \leq C_{12} + \min_{d\in\{1,2\}} \left\{ \mathcal{C}\left((\gamma_{1d}+\gamma_{1r})\bar{\rho}_1\alpha_1\right), \mathcal{C}\left(\gamma_{1d}(\bar{\rho}_1\alpha_1+\bar{\alpha}_1)+\gamma_{rd}+2\sqrt{\gamma_{1d}\gamma_{rd}\bar{\alpha}_1}\right)\right\},\\
& R_2 \leq C_{21} + \min_{d\in\{1,2\}} \left\{ \mathcal{C}\left((\gamma_{2d}+\gamma_{2r})\bar{\rho}_2\alpha_2\right), \mathcal{C}\left(\gamma_{2d}(\bar{\rho}_2\alpha_2+\bar{\alpha}_2)+\gamma_{rd}+2\sqrt{\gamma_{2d}\gamma_{rd}\bar{\alpha}_2}\right)\right\},\\
& R_1+R_2 \leq \min_{d\in\{1,2\}} \Big\{ \mathcal{C}((\gamma_{1d}+\gamma_{1r})\alpha_1 + (\sqrt{\gamma_{1d}\gamma_{2r}}-\sqrt{\gamma_{2d}\gamma_{1r}})^2\alpha_1\alpha_2(1-\lambda_d^2\rho_1\rho_2)\\
& \qquad + (\gamma_{2d}+\gamma_{2r})\alpha_2 + 2(\sqrt{\gamma_{1d}\gamma_{2d}}+\sqrt{\gamma_{1r}\gamma_{2r}})\lambda_d\sqrt{\rho_1\rho_2\alpha_1\alpha_2}),\\
& \qquad \mathcal{C}(\gamma_{1d}+\gamma_{2d}+\gamma_{rd}+2\sqrt{\bar{\alpha}_1\gamma_{1d}\gamma_{rd}}+2\sqrt{\bar{\alpha}_2\gamma_{2d}\gamma_{rd}}\\
& \qquad +2\sqrt{\gamma_{1d}\gamma_{2d}}(\sqrt{\rho_1\rho_2\alpha_1\alpha_2}+\sqrt{\bar{\alpha}_1\bar{\alpha}_2}))\Big\},\\
& R_1+R_2 \leq C_{12}+C_{21} + \min_{d\in\{1,2\}} \Big\{ \mathcal{C}\left((\gamma_{1d}+\gamma_{1r})\bar{\rho}_1\alpha_1 + (\gamma_{2d}+\gamma_{2r})\bar{\rho}_2\alpha_2\right.\\
& \qquad \left. +(\sqrt{\gamma_{1d}\gamma_{2r}}-\sqrt{\gamma_{2d}\gamma_{1r}})^2\bar{\rho}_1\bar{\rho}_2\alpha_1\alpha_2\right),\\
& \qquad \mathcal{C}\left(\gamma_{1d}(\bar{\alpha}_1+\bar{\rho}_1\alpha_1)+\gamma_{2d}(\bar{\alpha}_2+\bar{\rho}_2\alpha_2)+\gamma_{rd}\right.\\
& \qquad \left.\left. +2\sqrt{\gamma_{1d}\gamma_{2d}\bar{\alpha}_1\bar{\alpha}_2}+2\sqrt{\gamma_{1d}\gamma_{rd}\bar{\alpha}_1}+2\sqrt{\gamma_{2d}\gamma_{rd}\bar{\alpha}_2}\right)\right\}\Big\},
\end{aligned}
$$

where $\bar{\alpha}_1{=}1{-}\alpha_1$, $\bar{\alpha}_2{=}1{-}\alpha_2$, $\bar{\rho}_1{=}1{-}\rho_1$, $\bar{\rho}_2{=}1{-}\rho_2$, $\lambda_1=\lambda_2=1$ if $\alpha_1\alpha_2\rho_1\rho_2=0$ and

otherwise

$$\lambda_d = \min\left\{1,\ \frac{\sqrt{\gamma_{1d}\gamma_{2d}}+\sqrt{\gamma_{1r}\gamma_{2r}}}{(\sqrt{\gamma_{1d}\gamma_{2r}}-\sqrt{\gamma_{2d}\gamma_{1r}})^2\sqrt{\rho_1\rho_2\alpha_1\alpha_2}}\right\},\ d\in\{1,2\}.$$

Proof. The proof can be found in Appendix 3.5.A. □

3.3.1 Alternative Outer Bound C_{upp4}

By introducing ρ_1, ρ_2 independently, we have

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})] \le \sqrt{\rho_1\rho_2\alpha_1\alpha_2}$$

as stated in (3.40), which leads to a loose outer bound (when $\lambda_1<1$ or $\lambda_2<1$) on the sum rate. If we instead first introduce $\rho \in [0,1]$ such that

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})] = \rho\sqrt{\alpha_1\alpha_2}$$

to get a tighter outer bound on the sum rate, then ρ_1 and ρ_2 become correlated. Therefore, we may first define ρ and ρ_1 independently to get C_{upp2} which is tighter on the sum rate but looser on R_2, and then define ρ and ρ_2 independently to get C_{upp3} which is tighter on the sum rate but looser on R_1, and finally obtain the outer bound C_{upp4} by intersection of C_{upp2} and C_{upp3}.

Proposition 3.3. *The cut-set bound $C_{cut\text{-}set}$ in Proposition 3.1 is outer bounded by*

$$\begin{aligned}
C_{upp2} = \bigcup_{0\le\alpha_1,\alpha_2,\rho,\rho_1\le 1} \Big\{ & (R_1,R_2):\ R_1\ge 0,\ R_2\ge 0, \qquad (3.31)\\
R_1 \le C_{12} + \min_{d\in\{1,2\}} \Big\{ & \mathcal{C}\left((\gamma_{1d}+\gamma_{1r})\bar{\rho}_1\alpha_1\right), \mathcal{C}(\gamma_{1d}(\bar{\rho}_1\alpha_1+\bar{\alpha}_1)+\gamma_{rd}+2\sqrt{\gamma_{1d}\gamma_{rd}\bar{\alpha}_1})\Big\},\\
R_2 \le C_{21} + \min_{d\in\{1,2\}} \Big\{ & \mathcal{C}\left((\gamma_{2d}+\gamma_{2r})(1-\rho^2/\rho_1)\alpha_2\right),\\
& \mathcal{C}(\gamma_{2d}((1-\rho^2/\rho_1)\alpha_2+\bar{\alpha}_2)+\gamma_{rd}+2\sqrt{\gamma_{2d}\gamma_{rd}\bar{\alpha}_2})\Big\},\\
R_1+R_2 \le \min_{d\in\{1,2\}} \Big\{ & \mathcal{C}((\gamma_{1d}+\gamma_{1r})\alpha_1+(\sqrt{\gamma_{1d}\gamma_{2r}}-\sqrt{\gamma_{2d}\gamma_{1r}})^2\alpha_1\alpha_2(1-\rho^2)\\
& +(\gamma_{2d}+\gamma_{2r})\alpha_2+2(\sqrt{\gamma_{1d}\gamma_{2d}}+\sqrt{\gamma_{1r}\gamma_{2r}})\rho\sqrt{\alpha_1\alpha_2}),\\
& \mathcal{C}(\gamma_{1d}+\gamma_{2d}+\gamma_{rd}+2\sqrt{\bar{\alpha}_1\gamma_{1d}\gamma_{rd}}+2\sqrt{\bar{\alpha}_2\gamma_{2d}\gamma_{rd}}\\
& +2\sqrt{\gamma_{1d}\gamma_{2d}}(\rho\sqrt{\alpha_1\alpha_2}+\sqrt{\bar{\alpha}_1\bar{\alpha}_2}))\Big\},
\end{aligned}$$

$$R_1+R_2 \le C_{12}+C_{21}+\min_{d\in\{1,2\}}\Big\{\mathcal{C}((\gamma_{1d}+\gamma_{1r})\bar{\rho}_1\alpha_1+(\gamma_{2d}+\gamma_{2r})(1-\rho^2/\rho_1)\alpha_2 + (\sqrt{\gamma_{1d}\gamma_{2r}}-\sqrt{\gamma_{2d}\gamma_{1r}})^2\alpha_1\alpha_2\bar{\rho}_1(1-\rho^2/\rho_1)), \mathcal{C}(\gamma_{1d}(\bar{\alpha}_1+\bar{\rho}_1\alpha_1)+\gamma_{2d}(\bar{\alpha}_2+(1-\rho^2/\rho_1)\alpha_2)+\gamma_{rd} +2\sqrt{\gamma_{1d}\gamma_{2d}\bar{\alpha}_1\bar{\alpha}_2}+2\sqrt{\gamma_{1d}\gamma_{rd}\bar{\alpha}_1}+2\sqrt{\gamma_{2d}\gamma_{rd}\bar{\alpha}_2})\Big\}\Big\},$$

with $\bar{\alpha}_1{=}1{-}\alpha_1$, $\bar{\alpha}_2{=}1{-}\alpha_2$, $\bar{\rho}_1{=}1{-}\rho_1$, $\rho^2 \le \rho_1$, *and* $\rho^2/\rho_1 = 0$ *for* $\rho = \rho_1 = 0$.

Proof. The proof can be found in Appendix 3.5.B. □

Proposition 3.4. *$C_{cut\text{-}set}$ can be outer bounded by $C_{upp4} = C_{upp2} \cap C_{upp3}$, where C_{upp3} is obtained directly from* (3.31) *by the variable substitution (ρ^2/ρ_1 is treated as a single variable) as follows: $\rho^2/\rho_1 \Leftrightarrow \rho_2$, $1-\rho^2/\rho_1 \Leftrightarrow \bar{\rho}_2$, $\rho_1 \Leftrightarrow \rho^2/\rho_2$, $\bar{\rho}_1 \Leftrightarrow 1-\rho^2/\rho_2$.*

Proof. It is sufficient to prove $C_{\text{cut-set}} \subseteq C_{\text{upp3}}$ by following the same procedure as in Appendix 3.5.B except introducing ρ_2 (instead of ρ_1) such that

$$\bar{\rho}_2\alpha_2 = \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{r,i}X_{s,i})].$$

The supremum operation is over $0{\le}\alpha_1,\alpha_2,\rho,\rho_2{\le}1$ accordingly with $\rho^2{\le}\rho_2$. □

3.4 Summary

In this chapter, we have considered a relay-aided two-source two-sink wireless multicast network with a backhaul link between the source nodes. For the multicast relay network with high-rate backhaul, we find the exact cut-set bound on the capacity region by extending the proof of the converse developed by Cover and El Gamal [CE79] for the Gaussian relay channel. For low-rate backhaul scenarios, we have provided two genie-aided outer bounds by introducing two new lemmas on conditional (co-)variance.

Note that although the original proof of the converse developed by Cover and El Gamal [CE79] and its extensions developed in this chapter seem to be in a standard and uniform format, there are indeed some important tricks behind. In their original proof, Cover and El Gamal exploit many standard equalities and inequalities in information theory, such as the Fano's inequality, the Jensen's inequality, the Cauchy-Schwarz inequality, the maximum entropy lemma, and the chain rule. There are two important tools in their proof that are crucial but less addressed: the average power constraint on transmitted signals and the introduction of auxiliary random variables to bound the conditional variance for correlated sources. The former is crucial in translating the averaging operation into a single-letter constraint, and the latter is crucial to ensure that the averaging operation formulated

in a "proper" way to make the former translation possible. Here by "proper" we intend to highlight the fact that it is a kind of art to decide when and where to introduce such auxiliary random variables, since only "properly" chosen auxiliary random variables will lead to a nice result.

In our extensions, we start with a simpler network in [DXS11b] where no-cross links exists. We introduce a new random variable to bound the conditional covariance as in (3.16) and (3.17), and we also utilize the concavity of the function $\log(|\cdot|)$ with respect to a positive semi-definite matrix as in (3.24). When there are cross links as in our work [DXS11a], the received signals at destinations have three correlated components which makes the previous approach impossible to apply here. To overcome this difficulty, we introduce in the early stage the *law of total variance* which decomposes the dependence and therefore help us to translate (3.9) into (3.10), after which the previous approach can be applied. When the two source nodes are only partially correlated, as in our work [DXSM], we have to first introduce a genie to remove the ambiguity about the relaying signal at source nodes, and then introduce the *law of total covariance* and develop ourselves two lemma which can be interpreted as the *law of total conditional variance* and the *law of total conditional covariance*. And depends on which auxiliary random variables are introduced, we have developed three different upper bounds, with one complementing another.

Although we have successfully extended this bounding method and managed to have good upper bounds, all the tricks mentioned above may hinder their application to more general and more complex networks in searching for a good upper bound.

3.5 Appendix

3.5.A Proof of the Outer Bound $C_{\mathbf{upp1}}$

We first present two lemmas that will be used in our proof.

Lemma 3.1. *For random variables X, Y, Z on the same probability space, each with finite variance, Y and Z are independent, we have*

$$E[\mathrm{Var}(X|Y)] = E[\mathrm{Var}(X|Y,Z)] + E[\mathrm{Var}(E[X|Z]|Y)].$$

Proof.

$$\begin{aligned} E[\mathrm{Var}(X|Y)] - E[\mathrm{Var}(X|Y,Z)] &= E(E^2[X|Y,Z]) - E(E^2[X|Y]) \\ &\overset{(a)}{=} E(T^2) - E(E^2[T|Y]) \\ &= E[\mathrm{Var}(T|Y)] \\ &= E[\mathrm{Var}(E[X|Y,Z]|Y)] \\ &\overset{(b)}{=} E[\mathrm{Var}(E[X|Z]|Y)], \end{aligned}$$

where (a) comes from variable substitution $T{=}E[X|Y,Z]$ and the fact that

$$E[T|Y]{=}E(E[X|Y,Z]|Y){=}E[X|Y],$$

and (b) is due to the fact that

$$\mathrm{Var}(E[X|Y{=}y,Z]|Y{=}y) = \mathrm{Var}(E[X|Z]|Y{=}y).$$

□

Lemma 3.2. *For random variables X, Y, Z, U on the same probability space, each with finite variance, Z and U are independent, and $X-(Z,U)-Y$ is a Markov chain, then*

$$E[\mathrm{Cov}(X,Y|Z)] = E[\mathrm{Cov}(E[X|U],E[Y|U]|Z)].$$

Proof.

$$\begin{aligned} E[\mathrm{Cov}(X,Y|Z)] &= E(XY) - E(E[X|Z]E[Y|Z]) \\ &= E(E[XY|Z,U]) - E(E(E[X|Z,U]|Z)E(E[Y|Z,U]|Z)) \\ &\overset{(c)}{=} E[\mathrm{Cov}(E[X|Z,U],E[Y|Z,U]|Z)] \\ &\overset{(d)}{=} E[\mathrm{Cov}(E[X|U],E[Y|U]|Z)], \end{aligned}$$

where (c) is due to Markovicity and (d) comes from

$$\mathrm{Cov}(E[X|z,U],E[Y|z,U]|z){=}\mathrm{Cov}(E[X|U],E[Y|U]|z).$$

□

Now we are ready to prove Proposition 3.2. Note that

$$\frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{1,i}|X_{r,i})] \le \frac{1}{n}\sum_{i=1}^{n} E[E(X_{1,i}^2|X_{r,i})] = \frac{1}{n}\sum_{i=1}^{n} E[X_{1,i}^2] \le 1,$$

we introduce an auxiliary variable $\alpha_1 \in [0,1]$ as in [CE79, DXS11b, DXS11a] such that

$$\bar{\alpha}_1 = 1-\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{1,i}|X_{r,i})]. \tag{3.32}$$

It is easy to show that

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{r,i})] = \frac{1}{n}\sum_{i=1}^{n} E[X_{1,i}^2] - \frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{1,i}|X_{r,i})] \le \alpha_1, \tag{3.33}$$

where the inequality comes from the power constraint (3.2).

With the help of the genie, we have $X_1 - (X_r, X_s) - X_2$ with X_r independent of X_s. By Lemma 3.1 and the fact that

$$E[\mathrm{Var}(X_1|X_rX_s)] \leq E[\mathrm{Var}(X_1|X_r)],$$

we define $\rho_1 \in [0,1]$ with $\bar{\rho}_1 = 1-\rho_1$ such that

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})] = \bar{\rho}_1\alpha_1, \\ &\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(E[X_{1,i}|X_{s,i}]|X_{r,i})] \leq \rho_1\alpha_1. \end{aligned} \tag{3.34}$$

Similarly, we can define $\alpha_2, \rho_2 \in [0,1]$ with $\bar{\alpha}_2 = 1-\alpha_2$, $\bar{\rho}_2 = 1-\rho_2$ such that

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} E[E^2(X_{2,i}|X_{r,i})] = \bar{\alpha}_2, \\ &\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{r,i})] \leq \alpha_2, \end{aligned} \tag{3.35}$$

and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{r,i}X_{s,i})] = \bar{\rho}_2\alpha_2, \\ &\frac{1}{n}\sum_{i=1}^{n} E\mathrm{Var}(E[X_{2,i}|X_{s,i}]|X_{r,i}) \leq \rho_2\alpha_2. \end{aligned} \tag{3.36}$$

Since $\mathrm{Cov}(X,Y) \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$, we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}[E(X_{1,i}|X_{r,i}), E(X_{2,i}|X_{r,i})] \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \sqrt{\mathrm{Var}[E(X_{1,i}|X_{r,i})]\mathrm{Var}[E(X_{2,i}|X_{r,i})]} \\ &\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}[E(X_{1,i}|X_{r,i})]\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}[E(X_{2,i}|X_{r,i})]} \\ &\leq \sqrt{\bar{\alpha}_1\bar{\alpha}_2}, \end{aligned} \tag{3.37}$$

where the second inequality is by the Cauchy–Schwarz inequality, and the last inequality is due to $\mathrm{Var}(X) \leq E(X^2)$ with substitution of (3.32) and (3.35). Applying

the same procedure, we can further obtain

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(X_{1,i},X_{r,i})=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(E[X_{1,i}|X_{r,i}],X_{r,i})\le\sqrt{\bar{\alpha}_1}, \tag{3.38}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(X_{2,i},X_{r,i})=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(E[X_{2,i}|X_{r,i}],X_{r,i})\le\sqrt{\bar{\alpha}_2}, \tag{3.39}$$

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(X_{1,i},X_{2,i}|X_{r,i})]\\&=\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(E(X_{1,i}|X_{s,i}),E(X_{2,i}|X_{s,i})|X_{r,i})]\\&\le\frac{1}{n}\sum_{i=1}^{n}\sqrt{E[\mathrm{Var}(E[X_{1,i}|X_{s,i}]|X_{r,i})]E[\mathrm{Var}(E[X_{2,i}|X_{s,i}]|X_{r,i})]}\\&\le\sqrt{\rho_1\rho_2\alpha_1\alpha_2},\end{aligned} \tag{3.40}$$

where the equality in (3.40) is due to Lemma 3.2.

By the *Law of total covariance* [Gus10], we have

$$\mathrm{Cov}(X,Y)=E[\mathrm{Cov}(X,Y|Z)]+\mathrm{Cov}(E[X|Z],E[Y|Z]),$$

which, combined with (3.40) and (3.37), leads to

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(X_{1,i},X_{2,i})\le\sqrt{\rho_1\rho_2\alpha_1\alpha_2}+\sqrt{\bar{\alpha}_1\bar{\alpha}_2}. \tag{3.41}$$

Similar to (3.40), we can obtain by Lemma 3.2 that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(X_{1,i},X_{2,i}|X_{s,i})]\\&=\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(E(X_{1,i}|X_{r,i}),E(X_{2,i}|X_{r,i})|X_{s,i})]\\&\le\frac{1}{n}\sum_{i=1}^{n}\sqrt{E[\mathrm{Var}(E(X_{1,i}|X_{r,i})|X_{s,i})]E[\mathrm{Var}(E(X_{2,i}|X_{r,i})|X_{s,i})]}\\&\le\frac{1}{n}\sum_{i=1}^{n}\sqrt{E(E^2(X_{1,i}|X_{r,i}))E(E^2(X_{2,i}|X_{r,i}))}\\&\le\sqrt{\bar{\alpha}_1\bar{\alpha}_2},\end{aligned} \tag{3.42}$$

where the second inequality comes from

$$E[\mathrm{Var}(X|Z)]=E(X^2)-E(E^2(X|Z))\le E(X^2). \tag{3.43}$$

By Lemma 3.1 and (3.43), we can obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{s,i})] \qquad (3.44)\\
&= \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{r,i}, X_{s,i})] + E[\mathrm{Var}(E[X_{1,i}|X_{r,i}]|X_{s,i})]\\
&\le \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{1,i}|X_{r,i}, X_{s,i})] + \frac{1}{n}\sum_{i=1}^{n} E(E^2[X_{1,i}|X_{r,i}])\\
&\le \bar{\rho}_1\alpha_1 + \bar{\alpha}_1,
\end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{s,i})] \le \bar{\rho}_2\alpha_2 + \bar{\alpha}_2. \qquad (3.45)$$

By symmetry, we only need to bound the following six mutual information constraints in (3.3) (the rest can be bounded using the same method):

$$\begin{aligned}
&I(X_1^{(n)}X_2^{(n)}; Y_1^{(n)}Y_r^{(n)}|X_r^{(n)}), I(X_1^{(n)}X_2^{(n)}X_r^{(n)}; Y_1^{(n)}|X_s^{(n)}),\\
&I(X_1^{(n)}X_r^{(n)}; Y_1^{(n)}|X_2^{(n)}X_s^{(n)}), I(X_1^{(n)}X_2^{(n)}X_r^{(n)}; Y_1^{(n)}),\\
&I(X_1^{(n)}X_2^{(n)}; Y_1^{(n)}Y_r^{(n)}|X_r^{(n)}X_s^{(n)}), I(X_1^{(n)}; Y_1^{(n)}Y_r^{(n)}|X_2^{(n)}X_r^{(n)}X_s^{(n)}).
\end{aligned}$$

Since

$$\begin{aligned}
\mathrm{Var}[Y_{1,i}] = 1 &+ \gamma_{11}\mathrm{Var}[X_{1,i}] + \gamma_{21}\mathrm{Var}[X_{2,i}] + \gamma_{r1}\mathrm{Var}[X_{r,i}] + 2\sqrt{\gamma_{11}\gamma_{21}}\mathrm{Cov}(X_{1,i}, X_{2,i})\\
&+ 2\sqrt{\gamma_{11}\gamma_{r1}}\mathrm{Cov}(X_{1,i}, X_{r,i}) + 2\sqrt{\gamma_{21}\gamma_{r1}}\mathrm{Cov}(X_{2,i}, X_{r,i}),
\end{aligned}$$

we can apply (3.38), (3.39), (3.41) and obtain

$$\begin{aligned}
&\frac{1}{n}I(X_1^{(n)}X_2^{(n)}X_r^{(n)}; Y_1^{(n)}) = \frac{1}{n}h(Y_1^{(n)}) - \frac{1}{2}\log(2\pi e) \qquad (3.46)\\
&\le \frac{1}{2}\log(\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}[Y_{1,i}])\\
&\le \mathcal{C}(\gamma_{11} + \gamma_{21} + \gamma_{r1} + 2\sqrt{\bar{\alpha}_1\gamma_{11}\gamma_{r1}} + 2\sqrt{\bar{\alpha}_2\gamma_{21}\gamma_{r1}}\\
&\qquad + 2\sqrt{\gamma_{11}\gamma_{21}}(\sqrt{\rho_1\rho_2\alpha_1\alpha_2} + \sqrt{\bar{\alpha}_1\bar{\alpha}_2})),
\end{aligned}$$

with the first inequality obtained as follows

$$\frac{1}{n}h(Y_1^{(n)}) \le \frac{1}{n}\sum_{i=1}^{n} h(Y_{1,i}) \le \frac{1}{2n}\sum_{i=1}^{n}\log(2\pi e\mathrm{Var}[Y_{1,i}]) \le \frac{1}{2}\log(\frac{2\pi e}{n}\sum_{i=1}^{n}\mathrm{Var}[Y_{1,i}]),$$

where the second inequality is due to the maximum entropy lemma [CT06] and the last step follows from Jensen's inequality. Similarly, by the fact that

$$\begin{aligned}\mathrm{Cov}(X_1, X_r) &= E[\mathrm{Cov}(X_1, X_r|X_s)] + \mathrm{Cov}(E(X_1|X_s), E(X_r)) \\ &= E[\mathrm{Cov}(X_1, X_r|X_s)],\end{aligned}$$

and

$$\begin{aligned}E[\mathrm{Var}(Y_{1,i}|X_{s,i})] = 1 &+ \gamma_{r1}\mathrm{Var}(X_{r,i}) + 2\sqrt{\gamma_{11}\gamma_{21}}E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{s,i})] \\ &+ \gamma_{11}E[\mathrm{Var}(X_{1,i}|X_{s,i})] + 2\sqrt{\gamma_{11}\gamma_{r1}}E[\mathrm{Cov}(X_{1,i}, X_{r,i}|X_{s,i})] \\ &+ \gamma_{21}E[\mathrm{Var}(X_{2,i}|X_{s,i})] + 2\sqrt{\gamma_{21}\gamma_{r1}}E[\mathrm{Cov}(X_{2,i}, X_{r,i}|X_{s,i})],\end{aligned}$$

we can obtain

$$\begin{aligned}&\frac{1}{n}I(X_1^{(n)}X_2^{(n)}X_r^{(n)}; Y_1^{(n)}|X_s^{(n)}) = \frac{1}{n}h(Y_1^{(n)}|X_s^{(n)}) - \frac{1}{2}\log(2\pi e) \qquad (3.47)\\ &\le \frac{1}{2}\log(\frac{1}{n}\sum_{i=1}^{n} E\mathrm{Var}[Y_{1,i}|X_{s,i}]) \\ &\le \mathcal{C}\left(\gamma_{11}(\bar{\alpha}_1 + \bar{\rho}_1\alpha_1) + \gamma_{21}(\bar{\alpha}_2 + \bar{\rho}_2\alpha_2) + \gamma_{r1}\right. \\ &\qquad \left. +2\sqrt{\gamma_{11}\gamma_{21}\bar{\alpha}_1\bar{\alpha}_2} + 2\sqrt{\gamma_{11}\gamma_{r1}\bar{\alpha}_1} + 2\sqrt{\gamma_{21}\gamma_{r1}\bar{\alpha}_2}\right).\end{aligned}$$

Similarly we can obtain

$$\begin{aligned}&E[\mathrm{Var}(Y_{1,i}|X_{2,i}X_{s,i})] \\ &\quad = 1 + E[\mathrm{Var}(\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{r1}}X_{r,i}|X_{2,i}X_{s,i})] \\ &\quad = 1 + E[\mathrm{Var}(\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{r1}}X_{r,i}|X_{r,i}X_{2,i}X_{s,i})] \\ &\qquad + E[\mathrm{Var}(E(\sqrt{\gamma_{11}}X_{1,i} + \sqrt{\gamma_{r1}}X_{r,i}|X_{r,i})|X_{2,i}X_{s,i})] \\ &\quad \le 1 + \gamma_{11}(E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})] + E[E^2(X_{1,i}|X_{r,i})]) \\ &\qquad + \gamma_{r1}E(X_{r,i}^2) + 2\sqrt{\gamma_{11}\gamma_{r1}}E(X_{r,i}E(X_{1,i}|X_{r,i})),\end{aligned}$$

where the second equality is due to Lemma 3.1 and the inequality is from (3.43). After applying the Cauchy–Schwarz inequality and power constraints, we can obtain

$$\begin{aligned}\frac{1}{n}I(X_1^{(n)}X_r^{(n)}; Y_1^{(n)}|X_2^{(n)}X_s^{(n)}) &\le \frac{1}{2}\log\left(\frac{1}{n}\sum_{i=1}^{n} E\mathrm{Var}[Y_{1,i}|X_{2i}X_{s,i}]\right) \\ &\le \mathcal{C}(\gamma_{11}(\bar{\rho}_1\alpha_1 + \bar{\alpha}_1) + \gamma_{r1} + 2\sqrt{\gamma_{11}\gamma_{r1}\bar{\alpha}_1}).\end{aligned}$$

Let $\boldsymbol{A}_i$ be the conditional covariance matrix of $(Y_{1,i}, Y_{r,i})$ given $X_{r,i}$, then

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i\right| \\
=&1+(\gamma_{11}+\gamma_{1r})\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i})]+\frac{(\gamma_{21}+\gamma_{2r})}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{2,i}|X_{r,i})] \\
&+(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\left(\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i})]\right)\left(\frac{1}{n}\sum_{j=1}^{n}E[\mathrm{Var}(X_{2,j}|X_{r,j})]\right) \\
&+2(\sqrt{\gamma_{11}\gamma_{21}}+\sqrt{\gamma_{1r}\gamma_{2r}})\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(X_{1,i},X_{2,i}|X_{r,i})] \\
&-(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\left(\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(X_{1,i},X_{2,i}|X_{r,i})]\right)^2 \\
\leq&1+(\gamma_{11}+\gamma_{1r})\alpha_1+2(\sqrt{\gamma_{11}\gamma_{21}}+\sqrt{\gamma_{1r}\gamma_{2r}})\lambda_1\sqrt{\rho_1\rho_2\alpha_1\alpha_2} \\
&+(\gamma_{21}+\gamma_{2r})\alpha_2+(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\alpha_1\alpha_2(1-\lambda_1^2\rho_1\rho_2),
\end{aligned}$$

where $\lambda_1 \in [0,1]$ is a maximization parameter defined by

$$\lambda_1 = \min\left\{1,\ \frac{\sqrt{\gamma_{11}\gamma_{21}}+\sqrt{\gamma_{1r}\gamma_{2r}}}{(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\sqrt{\rho_1\rho_2\alpha_1\alpha_2}}\right\},$$

with $\lambda_1 = 1$ if $\alpha_1\alpha_2\rho_1\rho_2 = 0$.

We can therefore bound the following mutual information terms

$$\begin{aligned}
&\frac{1}{n}I(X_1^{(n)}X_2^{(n)};Y_1^{(n)}Y_r^{(n)}|X_r^{(n)}) \leq \frac{1}{2}\log\left(\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i\right|\right) \qquad (3.48)\\
&\leq \mathcal{C}\left((\gamma_{11}+\gamma_{1r})\alpha_1+(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\alpha_1\alpha_2(1-\lambda_1^2\rho_1\rho_2)\right. \\
&\qquad \left.+(\gamma_{21}+\gamma_{2r})\alpha_2+2(\sqrt{\gamma_{11}\gamma_{21}}+\sqrt{\gamma_{1r}\gamma_{2r}})\lambda_1\sqrt{\rho_1\rho_2\alpha_1\alpha_2}\right).
\end{aligned}$$

Similarly, let $\boldsymbol{B}_i$ and $\boldsymbol{C}_i$ be the conditional covariance matrix of $(Y_{1,i}, Y_{r,i})$ given $(X_{r,i}, X_{s,i})$ and given $(X_{2,i}, X_{r,i}, X_{s,i})$, respectively, we have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{B}_i\right| \\
&=1+(\gamma_{11}+\gamma_{1r})\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})]+\frac{(\gamma_{21}+\gamma_{2r})}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{2,i}|X_{r,i}X_{s,i})] \\
&+(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})]\frac{1}{n}\sum_{j=1}^{n}E[\mathrm{Var}(X_{2,j}|X_{r,j}X_{s,j})],
\end{aligned}$$

and

$$\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{C}_i\right| = 1 + (\gamma_{11}+\gamma_{1r})\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})].$$

We can therefore bound the following terms

$$\begin{aligned}\frac{1}{n}I(X_1^{(n)}X_2^{(n)};Y_1^{(n)}Y_r^{(n)}|X_r^{(n)}X_s^{(n)}) &\le \frac{1}{2}\log\left(\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{B}_i\right|\right) \qquad (3.49)\\ &\le \mathcal{C}((\gamma_{11}+\gamma_{1r})\bar{\rho}_1\alpha_1 + (\gamma_{21}+\gamma_{2r})\bar{\rho}_2\alpha_2 + (\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2\bar{\rho}_1\bar{\rho}_2\alpha_1\alpha_2),\\ \frac{1}{n}I(X_1^{(n)};Y_1^{(n)}Y_r^{(n)}|X_2^{(n)}X_r^{(n)}X_s^{(n)}) &\le \frac{1}{2}\log\left(\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{C}_i\right|\right)\\ &\le \mathcal{C}\left((\gamma_{11}+\gamma_{1r})\bar{\rho}_1\alpha_1\right). \qquad (3.50)\end{aligned}$$

Using the similar procedures as we demonstrated above, we may bound the remaining inequalities in (3.3), and combining all the results and let $n\to\infty$, we obtain (3.30).

3.5.B Proof of the Outer Bound $C_{\mathbf{upp2}}$

As in Appendix 3.5.A, we first introduce auxiliary random variables $\alpha_1, \alpha_2, \rho_1 \in [0,1]$ such that

$$\begin{aligned}\bar{\alpha}_1 &= \frac{1}{n}\sum_{i=1}^{n}E[E^2(X_{1,i}|X_{r,i})],\\ \bar{\alpha}_2 &= \frac{1}{n}\sum_{i=1}^{n}E[E^2(X_{2,i}|X_{r,i})],\\ \bar{\rho}_1\alpha_1 &= \frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i}X_{s,i})],\end{aligned}$$

we can thus obtain (3.33)–(3.35), (3.37)–(3.39) and (3.42) as in Appendix 3.5.A. Since

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Cov}(X_{1,i},X_{2,i}|X_{r,i})]\\ &\le \sqrt{\frac{1}{n}\sum_{i=1}^{n}E[\mathrm{Var}(X_{1,i}|X_{r,i})]\frac{1}{n}\sum_{j=1}^{n}E[\mathrm{Var}(X_{2,j}|X_{r,j})]}\\ &\le \sqrt{\alpha_1\alpha_2},\end{aligned}$$

we can introduce an auxiliary variable $\rho \in [0,1]$ such that

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})] = \rho\sqrt{\alpha_1\alpha_2}. \tag{3.51}$$

On the other hand, we can obtain from (3.40) that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Cov}(X_{1,i}, X_{2,i}|X_{r,i})]\\ &\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(E(X_{1,i}|X_{s,i})|X_{r,i})] \times \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(E(X_{2,i}|X_{s,i})|X_{r,i})]},\end{aligned}$$

which leads to the following observations

$$\alpha_1\rho^2 \leq \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(E(X_{1,i}|X_{s,i})|X_{r,i})] \leq \rho_1\alpha_1, \tag{3.52}$$

$$\alpha_2\rho^2/\rho_1 \leq \frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(E(X_{2,i}|X_{s,i})|X_{r,i})] \leq \alpha_2, \tag{3.53}$$

$$\frac{1}{n}\sum_{i=1}^{n} E[\mathrm{Var}(X_{2,i}|X_{r,i}X_{s,i})] \leq (1-\rho^2/\rho_1)\alpha_2, \tag{3.54}$$

where $\rho^2/\rho_1 = 0$ when $\rho = \rho_1 = 0$.

Now, by following the same procedure as in Appendix 3.5.A, replacing ρ_2 by ρ^2/ρ_1, and bounding

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_i\right| \leq\ & 1 + (\sqrt{\gamma_{11}\gamma_{2r}} - \sqrt{\gamma_{21}\gamma_{1r}})^2\alpha_1\alpha_2(1-\rho^2) + (\gamma_{11}+\gamma_{1r})\alpha_1\\ &+ (\gamma_{21}+\gamma_{2r})\alpha_2 + 2(\sqrt{\gamma_{11}\gamma_{21}} + \sqrt{\gamma_{1r}\gamma_{2r}})\rho\sqrt{\alpha_1\alpha_2},\end{aligned}$$

we can obtain (3.31).

Chapter 4

Decoding Based Cooperation Schemes

We investigate cooperative network coding strategies for relay-aided two-source two-destination wireless relay networks with a backhaul connection between the source nodes, whose outer bounds have already been investigated in Chapter 3. We study cooperative strategies based on different network coding schemes, namely, finite field and linear network coding, and lattice coding. We measure the performance in terms of achievable rates over Gaussian channels and observe significant gains over benchmark schemes. We also propose a new coding scheme, partial-decode-and-forward based linear network coding, which is essentially a hybrid scheme utilizing rate-splitting and messages exchange at the source nodes, partial decoding and linear network coding at the relay, and joint decoding at each destination. We derive achievable rate regions for these schemes and show that for high-rate backhaul, the cut-set bound can be achieved by network coding based beamforming when the signal-to-noise ratios lie in the sphere defined by the source-relay and relay-destination channel gains.

4.1 Cooperative Schemes with Full Source Cooperation

Similar to [CE79, HMZ05, KGG05, SE07], the source $\mathcal{S}_i$, $i = 1, 2$, divides its messages W_i into B blocks $W_{i,1}, \ldots, W_{i,B}$ with nR_i bits each. The transmission is completed over $B + 1$ blocks. In the first block the two sources exchange $W_{i,1}$ over the backhaul and also broadcast their own messages over the relay channels. In block t, source $\mathcal{S}_i$ exchanges $W_{i,t}$ through the backhaul and broadcasts its codeword $X_{i,t}^{(n)}$, which is a function of $(W_{i,t}, W_{1,t-1}, W_{2,t-1})$, over the channels; in block $B + 1$ a codeword only depends on $W_{i,B}$ is broadcasted. As each block is transmitted over n channel uses, and assuming the backhaul is used for free, the overall rate is $\frac{BnR_i}{(B+1)n}$ bits per channel use, which converges to R_i when B goes to infinity. Three decoding protocols, namely *forward decoding* [CE79], *backward*

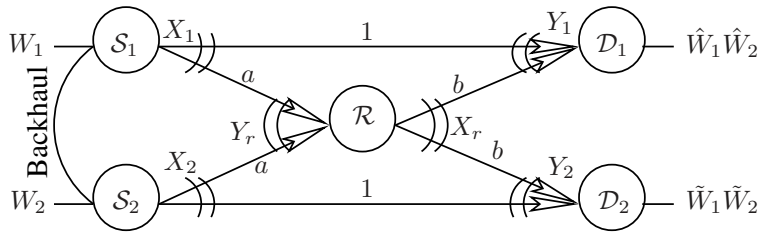


Figure 4.1. Two source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with backhaul, multicast information W_1 and W_2 respectively to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$, with aid from a full-duplex relay node $\mathcal{R}$.

decoding [Car82], and *sliding-window decoding* [Wil82], have been summarized and extended to multiple-source or multiple-relay scenarios in [KGG05]. We implement these protocols at relay/destination nodes depending on the cooperative NC strategy under consideration. Unless otherwise stated, random coding is used for encoding and joint-typicality is used for decoding. Each codeword is generated randomly in a memoryless fashion [CT06]: for transmitting messages in $\{W\}$, each of nR bits, we create a codebook consisting of 2^{nR} randomly and independently generated sequences $\{U^{(n)}\}$, each of length n, according to the distribution $\Pi_{i=1}^{n} p(u_i)$. We assign a codeword $U^{(n)}$ to each message W and associate them via an encoding function $U^{(n)}(W)$, omitting the explicit relation where convenient.

To simplify our analysis, we first consider the symmetric channel gain scenario without cross channels between $\mathcal{S}_1$ and $\mathcal{D}_2$, or $\mathcal{S}_2$ and $\mathcal{D}_1$, as illustrated in Figure 4.1. In this setup, we will provide several cooperative network coding schemes and compare their performance to the cut-set outer bound and a time-sharing benchmark. We then extend the coding schemes to more general channel setups with cross-link in Chapter 4.1.8.

Following the notation defined in (2.1), the system in Figure 4.1 can be modelled as

$$\begin{aligned} Y_1^{(n)} &= X_1^{(n)} + bX_r^{(n)} + Z_1^{(n)}, \\ Y_2^{(n)} &= X_2^{(n)} + bX_r^{(n)} + Z_2^{(n)}, \\ Y_r^{(n)} &= aX_1^{(n)} + aX_2^{(n)} + Z_r^{(n)}, \end{aligned} \tag{4.1}$$

where $X_i^{(n)}$, $Y_i^{(n)}$, $Z_i^{(n)}$, $i{=}1,2,r$ are n-dimensional vectors for the transmitted signals, received signals, and additive noise, respectively. $a \geq 0$ is the normalized channel gain for the source-relay links and $b \geq 0$ for the relay-destination links. The transmitted signals are subject to individual average power constraints as in (2.2). While, in general, the signal from $\mathcal{S}_i$ would be heard also at $\mathcal{D}_j$, $j \neq i$, our assumption can be motivated for example in scenarios where the cross links are too weak, or are technically suppressed. In any case we consider any contribution directly from $\mathcal{S}_i$ at $\mathcal{D}_j$ $(j \neq i)$ not to be useful and therefore part of the noise.

Table 4.1. Illustration of the encoding and decoding process for DF+FNC, with $W_{r,t} = W_{1,t-1} \oplus W_{2,t-1}$, $W_{r,1} = 1$, and $B = 3$.

$t =$	1	2	3	4
Backhaul	$W_{1,1} \Leftrightarrow W_{2,1}$	$W_{1,2} \Leftrightarrow W_{2,2}$	$W_{1,3} \Leftrightarrow W_{2,3}$	/
$\mathcal{S}_1$ transmits	$(W_{1,1}, 1)$	$(W_{1,2}, W_{r,2})$	$(W_{1,3}, W_{r,3})$	$(1, W_{r,4})$
$\mathcal{S}_2$ transmits	$(W_{2,1}, 1)$	$(W_{2,2}, W_{r,2})$	$(W_{2,3}, W_{r,3})$	$(1, W_{r,4})$
$\mathcal{R}$ transmits	1	$W_{r,2}$	$W_{r,3}$	$W_{r,4}$
$\mathcal{R}$ decodes	$W_{1,1}, W_{2,1}$	$W_{1,2}, W_{2,2}$	$W_{1,3}, W_{2,3}$	/
$\mathcal{D}_1$ decodes	$W_{1,1}$	$W_{1,2}, W_{r,2}$	$W_{1,3}, W_{r,3}$	$W_{r,4}$
recovers by $\oplus$	/	$W_{2,1}$	$W_{2,2}$	$W_{2,3}$

4.1.1 Finite-field Network Coding (DF+FNC)

At the end of block $t-1$, the relay decodes $(W_{1,t-1}, W_{2,t-1})$ jointly from its received signal $Y_{r,t-1}^{(n)}$ and then creates a new message $W_{r,t} = W_{1,t-1} \oplus W_{2,t-1}$ (bit-wise GF(2) addition). If the lengths of $W_{1,t-1}$ and $W_{2,t-1}$ are not equal, i.e., $R_1 \neq R_2$, we can append zeros at the end of the shorter message. During block t, $\mathcal{R}$ transmits $W_{r,t}$ using an independent random codebook $\{U^{(n)}\}$ of size 2^{nR} (where $R = \max(R_1, R_2)$),

$$X_{r,t}^{(n)} = \sqrt{P_r} U^{(n)}(W_{r,t}). \tag{4.2}$$

$\mathcal{S}_1$ and $\mathcal{S}_2$, on the other hand, transmit their information via independent random codebooks $\{V_1^{(n)}\}$ of size 2^{nR_1} and $\{V_2^{(n)}\}$ of size 2^{nR_2}, respectively. Since $W_{1,t-1}$ and $W_{2,t-1}$ have been exchanged via the backhaul in block $t-1$, $\mathcal{S}_1$ and $\mathcal{S}_2$ also know $W_{r,t}$. Therefore, to exploit the possibility of coherent combining gain $\mathcal{S}_1$ and $\mathcal{S}_2$ can coordinate their transmission with $\mathcal{R}$ as follows,

$$\begin{aligned} X_{1,t}^{(n)} &= \sqrt{\alpha_1 P_1} V_1^{(n)}(W_{1,t}) + \sqrt{(1-\alpha_1) P_1} U^{(n)}(W_{r,t}), \\ X_{2,t}^{(n)} &= \sqrt{\alpha_2 P_2} V_2^{(n)}(W_{2,t}) + \sqrt{(1-\alpha_2) P_2} U^{(n)}(W_{r,t}), \end{aligned} \tag{4.3}$$

where $0 \leq \alpha_1, \alpha_2 \leq 1$ are power allocation parameters. The received signals are therefore

$$\begin{aligned} Y_{1,t}^{(n)} &= \sqrt{\alpha_1 P_1} V_1^{(n)} + (\sqrt{(1-\alpha_1) P_1} + b\sqrt{P_r}) U^{(n)} + Z_{1,t}^{(n)}, \\ Y_{2,t}^{(n)} &= \sqrt{\alpha_2 P_2} V_2^{(n)} + (\sqrt{(1-\alpha_2) P_2} + b\sqrt{P_r}) U^{(n)} + Z_{2,t}^{(n)}, \\ Y_{r,t}^{(n)} &= a \left[\sqrt{\alpha_1 P_1} V_1^{(n)} + \sqrt{\alpha_2 P_2} V_2^{(n)} + (\sqrt{(1-\alpha_1) P_1} + \sqrt{(1-\alpha_2) P_2}) U^{(n)} \right] + Z_{r,t}^{(n)}. \end{aligned} \tag{4.4}$$

Forward decoding is implemented at both the relay and the two destination nodes: assuming $W_{1,t-1}$ has been successfully decoded by $\mathcal{D}_1$, at the end of block t, $\mathcal{D}_1$ recovers $(W_{1,t}, W_{r,t})$ jointly from $Y_{1,t}^{(n)}$, and then retrieves $W_{2,t-1} = W_{r,t} \oplus W_{1,t-1}$. This approach is also used by $\mathcal{D}_2$. The relay $\mathcal{R}$ decodes jointly $(W_{1,t}, W_{2,t})$ from $Y_{r,t}^{(n)}$ by first cancelling out $U^{(n)}$. The encoding/decoding process is illustrated in Table 4.1.

Proposition 4.1. *The achievable rate region for DF+FNC is the union over all* (R_1, R_2) *satisfying* $R_1 \geq 0$, $R_2 \geq 0$, *and*

$$\begin{aligned} R_1 &< \min\left\{\mathcal{C}(a^2\alpha_1 P_1), \mathcal{C}(\alpha_1 P_1), \mathcal{C}((\sqrt{(1-\alpha_2)P_2}+b\sqrt{P_r})^2)\right\}, \\ R_2 &< \min\left\{\mathcal{C}(a^2\alpha_2 P_2), \mathcal{C}(\alpha_2 P_2), \mathcal{C}((\sqrt{(1-\alpha_1)P_1}+b\sqrt{P_r})^2)\right\}, \\ R_1+R_2 &< \min\left\{\mathcal{C}(a^2\alpha_1 P_1+a^2\alpha_2 P_2), \mathcal{C}\left(P_1+b^2P_r+2b\sqrt{(1-\alpha_1)P_1P_r}\right),\right. \\ &\qquad \left. \mathcal{C}\left(P_2+b^2P_r+2b\sqrt{(1-\alpha_2)P_2P_r}\right)\right\}, \end{aligned} \tag{4.5}$$

where the union is taken over $0 \leq \alpha_1, \alpha_2 \leq 1$.

The constraint on R_1 corresponds to the condition that W_1 needs to be decoded reliably at $\mathcal{R}$ and $\mathcal{D}_1$, and that the NC message W_r be decoded at $\mathcal{D}_2$, and similarly for R_2 and $R_1 + R_2$. Note that our scheme is similar to the strategy in [GSG+10]: $\mathcal{D}_1$ recovers W_1 from the direct link and W_r from the $\mathcal{R}$–$\mathcal{D}_1$ link, and then retrieves W_2 from W_r based on the previous observation of W_1. However there are two main differences: finite-field NC rather than lattice coding is used; both source nodes know W_r thanks to the backhaul and therefore they cooperate with $\mathcal{R}$ to get a coherent combining gain.

Proof. The achievable rate regions derived for MARC in [KvW00, KGG05] involving multiple sources and a full-duplex DF relay can be directly applied here. Observing that $X_r^{(n)}$ is fully determined by $U^{(n)}$ as stated by (4.2), the rate regions defined by [KvW00, (5)] and [KGG05, (24) and (25)] can be translated to the FNC strategy as follows

$$\begin{aligned} nR_1 &< \min\{I(X_1^{(n)}; Y_r^{(n)}|U^{(n)}X_2^{(n)}), I(X_1^{(n)}; Y_1^{(n)}|U^{(n)}), I(X_r^{(n)}; Y_2^{(n)}|V_2^{(n)})\}, \\ nR_2 &< \min\{I(X_2^{(n)}; Y_r^{(n)}|U^{(n)}X_1^{(n)}), I(X_2^{(n)}; Y_2^{(n)}|U^{(n)}), I(X_r^{(n)}; Y_1^{(n)}|V_1^{(n)})\}, \\ n(R_1+R_2) &< \min\{I(X_1^{(n)}X_2^{(n)}; Y_r^{(n)}|U^{(n)}), I(X_1^{(n)}X_r^{(n)}; Y_1^{(n)}), I(X_2^{(n)}X_r^{(n)}; Y_2^{(n)})\}, \end{aligned} \tag{4.6}$$

where U, V_1, V_2 are auxiliary random variables and the joint probability factorizes as follows

$$p(v_1, v_2, u, x_1, x_2, x_r) = p(x_1, v_1|u)p(x_2, v_2|u)p(x_r, u).$$

For the constraints on R_1 in (4.6), the first term corresponds to successful decoding of W_1 at $\mathcal{R}$ given that the relaying signal $X_r^{(n)}$ has been cancelled out and $\mathcal{S}_2$ is not transmitting. The second term refers to the decoding of W_1 at $\mathcal{D}_1$ given correctly decoded W_r. The last term indicates the successful decoding of W_r (hence W_1 after NC decoding) at $\mathcal{D}_2$ given correctly decoded W_2. It is similar for R_2 and $R_1 + R_2$.

Since $X_1 - U - X_2$ form a Markov chain, following similar arguments as in the proof of *Lemma 3* in [BLW08], one can show that there exist joint Gaussian variables $(X_1^{(n)}, X_2^{(n)}, U^{(n)})$ that achieve the largest rate region defined in (4.6). By

choosing U, V_1, V_2 in (4.2) and (4.3) as i.i.d. zero-mean unit-variance Gaussian random variables, we can see that (X_1, X_r, X_2) is a zero-mean jointly Gaussian tuple satisfying the power constraint and $X_1 - X_r(U) - X_2$ form a Markov chain. By substituting (4.4) into (4.6), one can get (4.5) straightforwardly. □

For the symmetric scenario with $P_1 = P_2 = P_r = P$ and $R_1 = R_2 = R$, the symmetric rate R achieved by DF+FNC can be obtained straightforwardly from (4.5) by setting $\alpha_1 = \alpha_2 = \alpha$ as follows

$$R < \max_{0\le\alpha\le 1} \min\left\{\mathcal{C}(\alpha P), \frac{1}{2}\mathcal{C}(2a^2P\alpha), \frac{1}{2}\mathcal{C}((1+b^2+2b\sqrt{1-\alpha})P)\right\}. \tag{4.7}$$

Without the backhaul, $\mathcal{S}_1$ and $\mathcal{S}_2$ cannot know/estimate W_r and therefore cannot cooperate with $\mathcal{R}$, i.e. $\alpha_1 = \alpha_2 = 1$. Hence, no coherent combining gain can be achieved.

Remark 4.1. *FNC induces extra rate penalty in asymmetric channel setups: the rate carried by $U^{(n)}$ should be the the maximum of the two individual rates, but at the same time it should be supported by the weakest relay-destination link, which results in the following constraint,*

$$\max(R_1, R_2) < \min(I(X_r; Y_1|V_1), I(X_r; Y_2|V_2)).$$

4.1.2 Linear Network Coding (DF+LNC)

When LNC is used in the signal domain, $\mathcal{R}$ essentially performs superposition coding. The scheme presented here is a natural extension of the one in Theorem 1 of [SE07] which is designed for transmitting both private and common messages via the interference relay channel (IFRC). In our case, only common messages are transmitted (i.e., multicast). Unlike in [SE07] where each source can only cooperate with node $\mathcal{R}$ regarding its own message in $X_r^{(n)}$, here the two source nodes can cooperate to transmit both messages thanks to the backhaul. We first generate two independent random codebooks $\{U_1^{(n)}\}$ of size 2^{nR_1} and $\{U_2^{(n)}\}$ of size 2^{nR_2}. At the end of block $t-1$, $\mathcal{R}$ decodes $(W_{1,t-1}, W_{2,t-1})$ and then picks up codewords $U_1^{(n)}(W_{1,t-1})$ and $U_2^{(n)}(W_{2,t-1})$ from the two codebooks respectively, and transmits the superposition of these in block t with power allocation parameter $0 \le \alpha_r \le 1$

$$X_{r,t}^{(n)} = \sqrt{\alpha_r P_r}U_1^{(n)}(W_{1,t-1}) + \sqrt{(1-\alpha_r)P_r}U_2^{(n)}(W_{2,t-1}).$$

For each codeword $U_1^{(n)}(W_{1,t-1})$, we generate an independent codebook $\{V_1^{(n)}\}$ of size 2^{nR_1}, and then use this codebook to encode the new message $W_{1,t}$. We denote the selected codeword for $W_{1,t}$ given $W_{1,t-1}$ as $V_1^{(n)}(W_{1,t}, W_{1,t-1})$. Similarly we choose $V_2^{(n)}(W_{2,t}, W_{2,t-1})$ for $W_{2,t}$. With power allocation parameters $0 \le \alpha_i', \alpha_i'' \le$

1, $i = 1, 2$ to cooperate with $\mathcal{R}$, the transmitted signals at $\mathcal{S}_1$ and $\mathcal{S}_2$ are therefore

$$X_{1,t}^{(n)} = \sqrt{\alpha_1' P_1} U_1^{(n)} + \sqrt{\alpha_1'' P_1} U_2^{(n)} + \sqrt{(1-\alpha_1'-\alpha_1'')P_1} V_1^{(n)},$$
$$X_{2,t}^{(n)} = \sqrt{\alpha_2' P_2} U_2^{(n)} + \sqrt{\alpha_2'' P_2} U_1^{(n)} + \sqrt{(1-\alpha_2'-\alpha_2'')P_2} V_2^{(n)},$$

respectively. The received signals at the destinations and the relay are

$$\begin{aligned}
Y_1^{(n)} &= \sqrt{(1-\alpha_1'-\alpha_1'')P_1} V_1^{(n)} + (\sqrt{\alpha_1' P_1} + b\sqrt{\alpha_r P_r}) U_1^{(n)} \\
&\quad + (\sqrt{\alpha_1'' P_1} + b\sqrt{(1-\alpha_r)P_r}) U_2^{(n)} + Z_1^{(n)}, \\
Y_2^{(n)} &= \sqrt{(1-\alpha_2'-\alpha_2'')P_2} V_2^{(n)} + (\sqrt{\alpha_2'' P_2} + b\sqrt{\alpha_r P_r}) U_1^{(n)} \\
&\quad + (\sqrt{\alpha_2' P_2} + b\sqrt{(1-\alpha_r)P_r}) U_2^{(n)} + Z_2^{(n)}, \\
Y_r^{(n)} &= a \Big[\sqrt{(1-\alpha_1'-\alpha_1'')P_1} V_1^{(n)} + \sqrt{(1-\alpha_2'-\alpha_2'')P_2} V_2^{(n)} + (\sqrt{\alpha_1' P_1} + \sqrt{\alpha_2'' P_2}) U_1^{(n)} \\
&\quad + (\sqrt{\alpha_1'' P_1} + \sqrt{\alpha_2' P_2}) U_2^{(n)} \Big] + Z_r^{(n)},
\end{aligned} \tag{4.8}$$

respectively. Decoding follows directly from [SE07]: the relay performs *successive decoding* and the destinations use *backward decoding*. $\mathcal{R}$ decodes $(W_{1,t}, W_{2,t})$ reliably from $Y_{r,t}^{(n)}$ at the end of block t. $\mathcal{D}_1$ and $\mathcal{D}_2$ start decoding when transmission is finished. In block $B+1$ no new message is transmitted and the received signal at $\mathcal{D}_1$ ($\mathcal{D}_2$) only depends on $(W_{1,B}, W_{2,B})$. After decoding $(W_{1,B}, W_{2,B})$ successfully, only $W_{1,B-1}$ ($W_{2,B-1}$) is unknown in $Y_{1,B}^{(n)}$ ($Y_{2,B}^{(n)}$), and we repeat this process backwards until all messages are recovered.

Proposition 4.2. *The achievable rate region for DF+LNC is the union of* (R_1, R_2) *with* $R_1 \geq 0$, $R_2 \geq 0$, *and*

$$\begin{aligned}
R_1 < \min \Big\{ & \mathcal{C}(a^2 P_1 (1-\alpha_1'-\alpha_1'')), \\
& \mathcal{C}\left((1-\alpha_1'')P_1 + b^2 \alpha_r P_r + 2b\sqrt{\alpha_1' \alpha_r P_1 P_r} \right), \\
& \mathcal{C}\left(\alpha_2'' P_2 + b^2 \alpha_r P_r + 2b\sqrt{\alpha_2'' \alpha_r P_2 P_r} \right) \Big\}, \\
R_2 < \min \Big\{ & \mathcal{C}(a^2 P_2 (1-\alpha_2'-\alpha_2'')), \\
& \mathcal{C}\left((1-\alpha_2'')P_2 + b^2 (1-\alpha_r) P_r + 2b\sqrt{\alpha_2' (1-\alpha_r) P_2 P_r} \right), \\
& \mathcal{C}\left(\alpha_1'' P_1 + b^2 (1-\alpha_r) P_r + 2b\sqrt{\alpha_1'' (1-\alpha_r) P_1 P_r} \right) \Big\}, \\
R_1 + R_2 < \min \Big\{ & \mathcal{C}(a^2 (1-\alpha_1'-\alpha_1'') P_1 + a^2 (1-\alpha_2'-\alpha_2'') P_2), \\
& \mathcal{C}\left(P_1 + b^2 P_r + 2b\sqrt{P_1 P_r} \left[\sqrt{\alpha_1' \alpha_r} + \sqrt{\alpha_1''(1-\alpha_r)} \right] \right), \\
& \mathcal{C}\left(P_2 + b^2 P_r + 2b\sqrt{P_2 P_r} \left[\sqrt{\alpha_2'' \alpha_r} + \sqrt{\alpha_2'(1-\alpha_r)} \right] \right) \Big\},
\end{aligned} \tag{4.9}$$

where the union is taken over all $0 \leq \alpha_r, \alpha_1', \alpha_1'', \alpha_2', \alpha_2'' \leq 1$*, with* $\alpha_1' + \alpha_1'' \leq 1$ *and* $\alpha_2' + \alpha_2'' \leq 1$.

The constraint on R_1 refers to the condition that W_1 needs to be decoded successfully at $\mathcal{R}$, $\mathcal{D}_1$, and $\mathcal{D}_2$, respectively, and similarly for R_2 and $R_1 + R_2$.

Proof. By Theorem 1 of [SE07], $\mathcal{R}$ can decode $(W_{1,t}, W_{2,t})$ reliably if n is large, its past detection is correct, and

$$nR_1 < I(X_1^{(n)}; Y_r^{(n)} | U_1^{(n)} U_2^{(n)} X_2^{(n)} X_r^{(n)}), \tag{4.10a}$$

$$nR_2 < I(X_2^{(n)}; Y_r^{(n)} | U_1^{(n)} U_2^{(n)} X_1^{(n)} X_r^{(n)}), \tag{4.10b}$$

$$n(R_1 + R_2) < I(X_1^{(n)} X_2^{(n)}; Y_r^{(n)} | U_1^{(n)} U_2^{(n)} X_r^{(n)}), \tag{4.10c}$$

where U_1, U_2 are auxiliary random variables and the joint probability factorizes as follows

$$p(u_1, u_2, x_1, x_2, x_r) = p(x_1, u_1) p(x_2, u_2) p(x_r | u_1, u_2).$$

For $i = 1, 2$, $\mathcal{D}_i$ can decode $W_{i,t-1}$ reliably if n is large, its previously detection of $W_{i,t}$ is correct, and

$$\begin{aligned} nR_1 &< \min\left\{ I(X_1^{(n)} X_r^{(n)}; Y_1^{(n)} | U_2^{(n)}), I(X_r^{(n)}; Y_2^{(n)} | U_2^{(n)} X_2^{(n)}) \right\}, \\ nR_2 &< \min\left\{ I(X_2^{(n)} X_r^{(n)}; Y_2^{(n)} | U_1^{(n)}), I(X_r^{(n)}; Y_1^{(n)} | U_1^{(n)} X_1^{(n)}) \right\}, \\ n(R_1{+}R_2) &< \min\left\{ I(X_1^{(n)} X_r^{(n)}; Y_1^{(n)}), I(X_2^{(n)} X_r^{(n)}; Y_2^{(n)}) \right\}. \end{aligned} \tag{4.11}$$

By choosing U_1, U_2, V_1, V_2 i.i.d. zero-mean unit-variance Gaussian random variables and applying them in (4.10) and (4.11), we obtain the rate region as defined in (4.9). □

For the symmetric scenario, following from (4.9) directly by setting $\alpha_1' = \alpha_2' = \alpha'$, $\alpha_1'' = \alpha_2'' = \alpha''$, and $\alpha_r = \frac{1}{2}$, the following equal rate constraints apply

$$\begin{aligned} R < \max_{\substack{\alpha' \geq 0,\ \alpha'' \geq 0 \\ \alpha' + \alpha'' \leq 1}} \min\Bigg\{ & \mathcal{C}\left(\left(\alpha'' + \frac{1}{2} b^2 + b\sqrt{2\alpha''} \right) P \right), \mathcal{C}\left(\left(1 - \alpha'' + \frac{1}{2} b^2 + b\sqrt{2\alpha'} \right) P \right), \\ & \frac{1}{2} C\left(2a^2 P(1 - \alpha' - \alpha'') \right), \frac{1}{2} C\left((1 + b^2 + b\sqrt{2\alpha'} + b\sqrt{2\alpha''}) P \right) \Bigg\}. \end{aligned} \tag{4.12}$$

Without backhaul, X_r would only be partially known by the source nodes, i.e., $\alpha_1'' = \alpha_2'' = 0$.

4.1.3 Physical Layer Network Coding by Lattice Coding

In contrast to Sec. 4.1.1 where $\mathcal{R}$ first decodes (W_1, W_2) and then encodes into a joint NC message W_r, the relay can decode the NC message directly from $Y_r^{(n)}$ by using lattice encoding at the sources and lattice decoding at the relay, as in [GSG+10, WNPS10] where only the case of symmetric powers is considered. We propose a protocol based on superposition of a lattice code and a random code to be able to handle the case of non-symmetric powers. Without loss of generality, we assume that $P_1 \leq P_2$ (hence $R_1 \leq R_2$ due to symmetric channel setups). $\mathcal{S}_2$ splits its message $W_{2,t}$ into two parts $[W'_{2,t}, W''_{2,t}]$, where $W'_{2,t}$ has the same length as $W_{1,t}$. $\mathcal{S}_1$ encodes $W_{1,t}$ based on a nested lattice code [EZ04], and we denote the corresponding transmitted codeword by $V_1^{(n)}(W_{1,t})$. $\mathcal{S}_2$ encodes $W'_{2,t}$ using the same nested lattice code as $\mathcal{S}_1$, denoting the corresponding codeword by $V_2^{(n)}(W'_{2,t})$, and encodes $W''_{2,t}$ using a random codebook $\{V_3^{(n)}\}$ of size $2^{n(R_2-R_1)}$. Note that codewords $V_1^{(n)}$ and $V_2^{(n)}$ are independent even though they are generated by the same nested lattice code, since the dither vectors used at $\mathcal{S}_1$ and $\mathcal{S}_2$ are independent [GSG+10, EZ04]. The relay, after decoding $W''_{2,t-1}$ via a single-user joint-typicality decoder and the NC message $W_{1,t-1} \oplus W'_{2,t-1}$ using a lattice decoder, encodes all these new messages by using an independent random codebook $\{U^{(n)}\}$ of size 2^{nR_2},

$$X_{r,t}^{(n)} = \sqrt{P_r} U^{(n)}(W_{1,t-1} \oplus W'_{2,t-1}, W''_{2,t-1}).$$

Since $W_{1,t-1}$ and $W_{2,t-1}$ are known both at $\mathcal{S}_1$ and $\mathcal{S}_2$ thanks to the backhaul, $U^{(n)}(W_{1,t-1} \oplus W'_{2,t-1}, W''_{2,t-1})$ is also known. Therefore $\mathcal{S}_1$ and $\mathcal{S}_2$ cooperate with $\mathcal{R}$ as follows

$$\begin{aligned} X_{1,t}^{(n)} &= \sqrt{\delta} V_1^{(n)}(W_{1,t}) + \sqrt{P_1 - \delta} U^{(n)}, \\ X_{2,t}^{(n)} &= \sqrt{\delta} V_2^{(n)}(W'_{2,t}) + \sqrt{\epsilon} V_3^{(n)}(W''_{2,t}) + \sqrt{P_2 - \delta - \epsilon} U^{(n)}, \end{aligned} \tag{4.13}$$

where $0 \leq \delta \leq P_1$ and $0 \leq \epsilon \leq P_2 - \delta$ are the power[1] allocated to transmit the new messages. The corresponding received signals at the relay and destinations are

$$\begin{aligned} Y_{r,t}^{(n)} &= a\sqrt{\delta}\left(V_1^{(n)} + V_2^{(n)}\right) + a\sqrt{\epsilon} V_3^{(n)} + a\left(\sqrt{P_1 - \delta} + \sqrt{P_2 - \delta - \epsilon}\right) U^{(n)} + Z_{r,t}^{(n)}, \\ Y_{1,t}^{(n)} &= \sqrt{\delta} V_1^{(n)} + \left(\sqrt{P_1 - \delta} + b\sqrt{P_r}\right) U^{(n)} + Z_{1,t}^{(n)}, \\ Y_{2,t}^{(n)} &= \sqrt{\delta} V_2^{(n)} + \sqrt{\epsilon} V_3^{(n)} + \left(\sqrt{P_2 - \delta - \epsilon} + b\sqrt{P_r}\right) U^{(n)} + Z_{2,t}^{(n)}, \end{aligned} \tag{4.14}$$

respectively. $\mathcal{D}_1$ performs *successive forward decoding*: at the end of block t, $\mathcal{D}_1$ decodes $(W_{1,t-1} \oplus W'_{2,t-1}, W''_{2,t-1})$ from $Y_{1,t}^{(n)}$ by joint typicality and recovers $W'_{2,t-1}$ by using $W_{1,t-1}$ which has been recovered successfully from block $t-1$; after cancelling out $U^{(n)}$ the new information $W_{1,t}$ can be decoded. This approach is also used by $\mathcal{D}_2$.

[1] Multiplicative rather than additive power allocation factors are used here to ensure that the lattice codewords arrive at the relay node with the same power.

Proposition 4.3. *Using lattice coding, an achievable rate region is given by the union of* (R_1, R_2)*, where* $R_1 \geq 0$*,* $R_2 \geq 0$*, and*

$$\begin{aligned} R_1 &< \min\left\{\mathcal{C}\left(-\frac{1}{2}+a^2\delta\right),\ C(\delta)\right\},\\ R_2 &< \min\left\{\mathcal{C}\left(-\frac{1}{2}+a^2\delta+\frac{a^2\epsilon}{2}\right),\ \mathcal{C}(\delta+\epsilon)\right\},\\ R_1+R_2 &< \min\Big\{\mathcal{C}\left(P_1+b^2P_r+2b\sqrt{P_r(P_1-\delta)}\right),\\ &\qquad \mathcal{C}\left(P_2+b^2P_r+2b\sqrt{P_r(P_2-\delta-\epsilon)}\right)\Big\}, \end{aligned} \tag{4.15}$$

with the union taken over $0 \leq \delta \leq P_1$ *and* $0 \leq \epsilon \leq P_2 - \delta$*.*

The first term in R_1 (R_2) refers to the decoding constraint at $\mathcal{R}$ for the nested lattice code [EZ04].

Proof. After cancelling out $U^{(n)}$ from $Y_{r,t}^{(n)}$, the relay can reliably decode $W_{2,t}''$ (by using a Gaussian codebook) if

$$R_2 - R_1 < \frac{1}{2}\log\left(1+\frac{a^2\epsilon}{1+2a^2\delta}\right).$$

Then $\mathcal{R}$ can further cancel out $V_3^{(n)}$ and use the remaining signal to decode the NC message by using lattice decoding [EZ04, WNPS10] if

$$R_1 < \frac{1}{2}\log\left(\frac{1}{2}+a^2\delta\right).$$

Therefore decoding at $\mathcal{R}$ will introduce the following constraints:

$$\begin{aligned} R_1 &< \mathcal{C}\left(-\frac{1}{2}+a^2\delta\right), \qquad (4.16)\\ R_2 &< \frac{1}{2}\log\left(\left(\frac{1}{2}+a^2\delta\right)\left(1+\frac{a^2\epsilon}{1+2a^2\delta}\right)\right) = \mathcal{C}\left(-\frac{1}{2}+a^2\delta+\frac{a^2\epsilon}{2}\right). \end{aligned}$$

$\mathcal{D}_1$ and $\mathcal{D}_2$ can successfully decode $W_{1,t}$ and $W_{2,t} = [W_{2,t}', W_{2,t}'']$, respectively, if

$$\begin{aligned} R_1 &< \frac{1}{2}\log(1+\delta),\\ R_2 &< \frac{1}{2}\log(1+\delta+\epsilon). \end{aligned} \tag{4.17}$$

By using *successive decoding* at both $\mathcal{D}_1$ and $\mathcal{D}_2$, the following constraints apply

$$\begin{aligned} R_1+R_2 &< \frac{1}{2}\log(1+\delta)+\frac{1}{2}\log\left(1+\frac{(\sqrt{P_1-\delta}+b\sqrt{P_r})^2}{1+\delta}\right), \qquad (4.18)\\ R_1+R_2 &< \frac{1}{2}\log(1+\delta+\epsilon)+\frac{1}{2}\log\left(1+\frac{(\sqrt{P_2-\delta-\epsilon}+b\sqrt{P_r})^2}{1+\delta+\epsilon}\right). \end{aligned}$$

Combining (4.16), (4.17), and (4.18) together we can obtain (4.15). □

By setting $\epsilon = 0$ and $\delta = P\alpha$ in (4.19), we can obtain the following achievable rate for the symmetric scenario

$$R < \max_{0\le\alpha\le 1} \min\left\{\mathcal{C}(\alpha P),\ \mathcal{C}\left(-\frac{1}{2} + a^2 P\alpha\right), \frac{1}{2}\mathcal{C}\left(\left(1 + b^2 + 2b\sqrt{1-\alpha}\right) P\right)\right\}. \quad (4.19)$$

Without backhaul, the NC message would not be known at the sources, i.e., $\delta = P_1$ and $\epsilon = P_2 - P_1$.

4.1.4 Network Coding Based Beamforming (DF+NBF)

To further exploit the available coherent combining (beam-forming) gain [CE79, HMZ05, KGG05] at the sinks, we propose a new strategy that performs network coding at both $\mathcal{S}_1$ and $\mathcal{S}_2$ but not at the relay (decreasing the complexity at $\mathcal{R}$). We refer to this scheme as network coding based beamforming (NBF) since the signals transmitted at $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{R}$ are formed in a beam-forming fashion. NBF requires B+2 blocks[2] in total: $(W_{1,t-1}, W_{2,t-1})$ are exchanged via the backhaul during block t−1; in block t the network coded message $W_t = f(W_{1,t-1}, W_{2,t-1})$ is transmitted by both $\mathcal{S}_1$ and $\mathcal{S}_2$; W_t is transmitted by $\mathcal{R}$ in block t+1. Note that the network coding operation $W_t = f(W_{1,t-1}, W_{2,t-1})$ at both source nodes is indeed symbolic rather than operational, since the operator $f(\cdot,\cdot)$ incurs no rate loss and does not explicitly specify how the two messages are combined together. There are many options for this network coding operator $f(\cdot,\cdot)$. For example, it can be superposition encoding, or double-index encoding, or simply appending one message after the other. Note that these options will make no difference under the encoding and decoding schemes stated below for the current network setup, but we would like to keep $f(\cdot,\cdot)$ in general format since NBF scheme can be applied to more general channel setups and be equipped with more general encoding/decoding methods, which might result in different achievable rate regions for different network coding operations.

The relay transmits W_{t-1} using a random codebook $\{U^{(n)}\}$ of size $2^{n(R_1+R_2)}$. For each codeword $U^{(n)}(W_{t-1})$, we generate an independent random codebook $\{V^{(n)}\}$ of size $2^{n(R_1+R_2)}$, and then use it to encode the new message W_t. We denote the selected codeword for W_t given W_{t-1} as $V^{(n)}(W_t, W_{t-1})$. At block t, the transmitted signals are

$$\begin{aligned} X_{r,t}^{(n)} &= \sqrt{P_r}U^{(n)}(W_{t-1}), \quad (4.20)\\ X_{1,t}^{(n)} &= \sqrt{\alpha_1 P_1}V^{(n)}(W_t, W_{t-1}) + \sqrt{(1-\alpha_1)P_1}U^{(n)}(W_{t-1}),\\ X_{2,t}^{(n)} &= \sqrt{\alpha_2 P_2}V^{(n)}(W_t, W_{t-1}) + \sqrt{(1-\alpha_2)P_2}U^{(n)}(W_{t-1}), \end{aligned}$$

[2]The first block only includes message exchange via backhaul but no signal transmission in the wireless channel. Therefore the power consumption in NBF scheme is the same as in other schemes as we assume no cost for message exchange via backhaul.

where $0 \leq \alpha_1, \alpha_2 \leq 1$ are power allocation parameters. The corresponding received signals[3] are

$$\begin{aligned} Y_{1,t}^{(n)} &= \sqrt{\alpha_1 P_1} V^{(n)} + (b\sqrt{P_r} + \sqrt{(1-\alpha_1)P_1}) U^{(n)} + Z_{1,t}^{(n)}, \\ Y_{2,t}^{(n)} &= \sqrt{\alpha_2 P_2} V^{(n)} + (b\sqrt{P_r} + \sqrt{(1-\alpha_2)P_2}) U^{(n)} + Z_{2,t}^{(n)}, \\ Y_{r,t}^{(n)} &= a\left(\sqrt{\alpha_1 P_1} + \sqrt{\alpha_2 P_2}\right) V^{(n)} + a\left(\sqrt{(1-\alpha_1)P_1} + \sqrt{(1-\alpha_2)P_2}\right) U^{(n)} + Z_{r,t}^{(n)}. \end{aligned} \tag{4.21}$$

The decoding process is similar as in the other cooperative strategies: the relay performs *forward decoding* and the destinations utilize *backward decoding*.

Proposition 4.4. *The achievable rate region for NBF is defined by the union of* (R_1, R_2) *satisfying* $R_1 \geq 0$, $R_2 \geq 0$, *and*

$$\begin{aligned} R_1 + R_2 < \min \Big\{ & \mathcal{C}\left(P_1 + b^2 P_r + 2b\sqrt{(1-\alpha_1)P_1 P_r}\right), \\ & \mathcal{C}\left(P_2 + b^2 P_r + 2b\sqrt{(1-\alpha_2)P_2 P_r}\right), \\ & \mathcal{C}\left(a^2\left(\alpha_1 P_1 + \alpha_2 P_2 + 2\sqrt{\alpha_1\alpha_2 P_1 P_2}\right)\right)\Big\}, \end{aligned} \tag{4.22}$$

with the union taken over the power allocation parameters $0 \leq \alpha_1, \alpha_2 \leq 1$.

The terms in (4.22) indicate the constraints at $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{R}$, respectively.

Proof. Since $\mathcal{S}_1$ and $\mathcal{S}_2$ transmit the same NC message W_t, the achievable sum rate can be split arbitrarily between them. Therefore in the NBF strategy only the constraints for the sum rate matter. Following similar arguments as in Sec. 4.1.1, the sum rate constraint in (4.6) still holds here, i.e.,

$$n(R_1+R_2) < \min\{I(X_1^{(n)}X_2^{(n)}; Y_r^{(n)}|U^{(n)}), I(X_1^{(n)}X_r^{(n)}; Y_1^{(n)}), I(X_2^{(n)}X_r^{(n)}; Y_2^{(n)})\}.$$

By applying *successive decoding* to $Y_{r,t}^{(n)}$ and *backward decoding* to $Y_{1,t}^{(n)}$ and $Y_{2,t}^{(n)}$, the jointly Gaussian distributed random variables $(V^{(n)}, U^{(n)})$ will translate the above rate constraints into (4.22). □

For the symmetric scenario, the achievable rate can be obtained from (4.22) by setting $\alpha_1 = \alpha_2 = \alpha$ as follows

$$R < \max_{0\leq\alpha\leq 1} \min\left\{\frac{1}{2}\mathcal{C}(4a^2 P\alpha), \frac{1}{2}\mathcal{C}\left(\left(1 + b^2 + 2b\sqrt{1-\alpha}\right)P\right)\right\}. \tag{4.23}$$

Without the backhaul, this strategy is impossible.

Remark 4.2. *Compared to the lattice coding based scheme in Sec. 4.1.3, DF+NBF enjoys coherent combining gain at the relay but at the same time suffers from the decoding constraint at the relay node. Therefore these two schemes will complement each other in different channel setups.*

[3]Beamforming is implicitly performed here to ensure coherent addition of codewords/signals at relay and destinations, as stated in Section 2.1.

4.1.5 Benchmark: Time Sharing Relay

In contrast to the orthogonal scheme described in [SE07] for the case of the IFRC, $\mathcal{S}_1$ and $\mathcal{S}_2$ here cooperate with $\mathcal{R}$ to convey both messages. We first generate two independent random codebooks $\{U_1^{(n)}\}$ of size 2^{nR_1} and $\{U_2^{(n)}\}$ of size 2^{nR_2} that will be used by $\mathcal{R}$ to help $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. For each codeword in $\{U_2^{(n)}\}$, we generate an independent random codebook $\{V_1^{(n)}\}$ of size 2^{nR_1}, and then use it to encode the new message at $\mathcal{S}_1$. Similarly, we generate a random codebook $\{V_2^{(n)}\}$ of size 2^{nR_2} for each codeword in $\{U_1^{(n)}\}$. During block t, $W_{1,t}$ and $W_{2,t}$ are exchanged via the backhaul, and the transmission during block t is divided into two parts. During the first part of block t, the transmitted signals are

$$
\begin{aligned}
X_{r,t_1}^{(n)} &= \sqrt{P_r}U_2^{(n)}(W_{2,t-1}),\\
X_{2,t_1}^{(n)} &= 0,\\
X_{1,t_1}^{(n)} &= \sqrt{\frac{\alpha_1 P_1}{\beta}}V_1^{(n)}(W_{1,t}, W_{2,t-1}) + \sqrt{\frac{P_1(1-\alpha_1)}{\beta}}U_2^{(n)}(W_{2,t-1}),
\end{aligned}
$$

where $0 \le \alpha_1 \le 1$ is the power allocation parameter and $0 \le \beta \le 1$ is the time sharing parameter. Transmission power $\frac{P_1}{\beta}$ is used in $X_{1,t_1}^{(n)}$ to meet the power constraint. The received signals are

$$
\begin{aligned}
Y_{2,t_1}^{(n)} &= bX_{r,t_1}^{(n)} + Z_{2,t_1}^{(n)} = b\sqrt{P_r}U_2^{(n)} + Z_{2,t_1}^{(n)},\\
Y_{r,t_1}^{(n)} &= a\sqrt{\frac{\alpha_1 P_1}{\beta}}V_1^{(n)} + a\sqrt{\frac{P_1(1-\alpha_1)}{\beta}}U_2^{(n)} + Z_{r,t_1}^{(n)},\\
Y_{1,t_1}^{(n)} &= \sqrt{\frac{\alpha_1 P_1}{\beta}}V_1^{(n)} + \left(\sqrt{\frac{P_1(1-\alpha_1)}{\beta}} + b\sqrt{P_r}\right)U_2^{(n)} + Z_{1,t_1}^{(n)}.
\end{aligned}
\tag{4.24}
$$

The relay decodes $W_{1,t}$ given $W_{2,t-1}$ and then encodes it to $U_1^{(n)}(W_{1,t})$. During the remaining part of block t, the transmitted signals are

$$
\begin{aligned}
X_{r,t_2}^{(n)} &= \sqrt{P_r}U_1^{(n)}(W_{1,t}),\\
X_{1,t_2}^{(n)} &= 0,\\
X_{2,t_2}^{(n)} &= \sqrt{\frac{\alpha_2 P_2}{1-\beta}}V_2^{(n)}(W_{2,t}, W_{1,t}) + \sqrt{\frac{P_2(1-\alpha_2)}{1-\beta}}U_1^{(n)}(W_{1,t}).
\end{aligned}
$$

The corresponding received signals are

$$\begin{aligned}
Y_{1,t_2}^{(n)} &= bX_{r,t_2}^{(n)} + Z_{1,t_2}^{(n)} = b\sqrt{P_r}U_1^{(n)} + Z_{1,t_2}^{(n)}, \\
Y_{r,t_2}^{(n)} &= a\sqrt{\frac{\alpha_2 P_2}{1-\beta}}V_2^{(n)} + a\sqrt{\frac{P_2(1-\alpha_2)}{1-\beta}}U_1^{(n)} + Z_{2,t_2}^{(n)}, \\
Y_{2,t_2}^{(n)} &= \sqrt{\frac{\alpha_2 P_2}{1-\beta}}V_2^{(n)} + \left(\sqrt{\frac{(1-\alpha_2)P_2}{1-\beta}} + b\sqrt{P_r}\right)U_1^{(n)} + Z_{2,t_2}^{(n)}.
\end{aligned} \tag{4.25}$$

At the end of block t, $\mathcal{R}$ decodes $W_{2,t}$ given $W_{1,t}$, and $\mathcal{D}_1$ can retrieve $(W_{1,t}, W_{2,t-1})$ reliably using *sliding-window decoding* based on the received signals during block t. Similarly, after the first part of block $t+1$, $\mathcal{D}_2$ can decode $(W_{2,t}, W_{1,t})$ reliably based on signals received from the first part of block $t+1$ and the second part of block t.

Following the same method as in Sec. 4.1.2, by applying Gaussian random variables and noting the dependence stated in (4.24) and (4.25), the achievable rate region for this time sharing strategy can be defined by the union of (R_1, R_2) satisfying $R_1 \geq 0$, $R_2 \geq 0$, and

$$\begin{aligned}
R_1 < \min\Bigg\{ & \beta\mathcal{C}\left(\frac{\alpha_1 a^2 P_1}{\beta}\right),\ \beta\mathcal{C}\left(\frac{\alpha_1 P_1}{\beta}\right) + (1-\beta)\mathcal{C}(b^2 P_r), \\
& (1-\beta)\mathcal{C}\left(b^2 P_r + \frac{P_2}{1-\beta} + 2b\sqrt{\frac{(1-\alpha_2)P_2 P_r}{1-\beta}}\right)\Bigg\}, \\
R_2 < \min\Bigg\{ & (1-\beta)\mathcal{C}\left(\frac{\alpha_2 a^2 P_2}{1-\beta}\right), (1-\beta)\mathcal{C}\left(\frac{\alpha_2 P_2}{1-\beta}\right) + \beta\mathcal{C}(b^2 P_r), \\
& \beta\mathcal{C}\left(b^2 P_r + \frac{P_1}{\beta} + 2b\sqrt{\frac{P_1 P_r(1-\alpha_1)}{\beta}}\right)\Bigg\}, \\
R_1 + R_2 < \min\Bigg\{ & (1-\beta)\mathcal{C}(b^2 P_r) + \beta\mathcal{C}\left(b^2 P_r + \frac{P_1}{\beta} + 2b\sqrt{\frac{P_1 P_r(1-\alpha_1)}{\beta}}\right), \\
& \beta\mathcal{C}(b^2 P_r) + (1-\beta)\mathcal{C}\left(b^2 P_r + \frac{P_2}{1-\beta} + 2b\sqrt{\frac{P_2 P_r(1-\alpha_2)}{1-\beta}}\right)\Bigg\},
\end{aligned} \tag{4.26}$$

with the union taken over all $0 \leq \alpha_1, \alpha_2 \leq 1$ and the time sharing parameter $0 \leq \beta \leq 1$.

Constraints in R_1 (R_2) correspond to the condition of successful decoding of W_1 (W_2) at $\mathcal{R}$, $\mathcal{D}_1$ ($\mathcal{D}_2$), and $\mathcal{D}_2$ ($\mathcal{D}_1$), respectively. Constraints in $R_1 + R_2$ refer to successful decoding at $\mathcal{D}_1$ and $\mathcal{D}_2$.

By setting $\alpha_1 = \alpha_2 = \alpha$ and $\beta = 1/2$, (4.26) can be translated into the sym-

metric rate constraint

$$R < \max_{0\le\alpha\le 1} \min\left\{ \frac{1}{2}\mathcal{C}\left(P\left(2+b^2+2b\sqrt{2-2\alpha}\right)\right), \frac{1}{2}\mathcal{C}\left((2\alpha+b^2+2\alpha b^2 P)P\right), \right. \\ \left. \frac{1}{2}\mathcal{C}(2a^2P\alpha), \frac{1}{4}\left[\mathcal{C}(b^2P)+\mathcal{C}\left(\left(2+b^2+2b\sqrt{2-2\alpha}\right)P\right)\right]\right\}. \quad (4.27)$$

Without backhaul, sources can only encode over their own messages. Therefore we have $\alpha_1 = \alpha_2 = 1$ and the first term in (4.27) reduces to $\frac{1}{2}\mathcal{C}(b^2P)$.

4.1.6 Capacity Achieving Special Case

The cut-set bound defined in (3.4) can be applied to the symmetric channel setup by setting $\gamma_{11} = \gamma_{22} = P$, $\gamma_{1r} = \gamma_{2r} = Pa^2$, $\gamma_{r1} = \gamma_{r2} = Pb^2$, and $\gamma_{12} = \gamma_{21} = 0$ (i.e., no cross links), resulting in the following constraint (after setting $\alpha = \alpha_1 = \alpha_2$ due to channel symmetry)

$$R < \sup_{0\le\alpha,\rho\le 1} \min\left\{ \frac{1}{2}\mathcal{C}\left(P\left(1+b^2+2b\sqrt{1-\alpha}\right)\right), \right. \quad (4.28) \\ \left. \frac{1}{2}\mathcal{C}\left(P[(1+2a^2)\alpha+2a^2\rho\alpha+a^2(1-\rho^2)\alpha^2 P]\right)\right\}.$$

Proposition 4.5. *In the symmetric scenario where $R_1 = R_2 = R$, DF+NBF can achieve the cut-set bound, i.e.* (4.23) *and* (4.28) *are equivalent, if and only if (a^2, b^2, P) satisfy the following two constraints simultaneously,*

$$\begin{cases} 4a^2 > \max\{2,\ 1+b^2\}, \\ 0 < P \le \dfrac{8a^2(2a^2-1)}{2a^2(1+b^2)-b^2+\sqrt{(4a^2-b^2)(4a^2-1)b^2}}. \end{cases} \quad (4.29)$$

Proof. The proof can be found in Appendix 4.4.A. □

Proposition 4.5 states that there exists a large set of different source-relay channel gains and relay-destination channel gains, where the cut-set bound can be achieved by the DF+NBF strategy if the normalized ($\sigma^2 = 1$) transmit power constraint P is no larger than an upper bound defined in (4.29), as shown in Figure 4.2. Therefore we can claim that even for non-degraded Gaussian relay channels, the capacity region for the system defined in Figure 4.1 can be known for the scenarios defined by Proposition 4.5.

An intuitive interpretation of Proposition 4.5 is that (4.29) ensures the successful decoding at the relay node $\mathcal{R}$. In this scenario, the NBF achievable rate (4.23) and the cut-set bound (4.28) have the same active constraint on the MAC at $\mathcal{D}_1$ and $\mathcal{D}_2$, and therefore leads to tight capacity bounds. The upper bound on P in (4.29) is to make sure that, given a^2 and α, the second term (the constraint at $\mathcal{R}$) in cut-set bound (4.28) cannot be increased by reducing ρ (otherwise we can increase (4.28) simply by decreasing α and ρ).

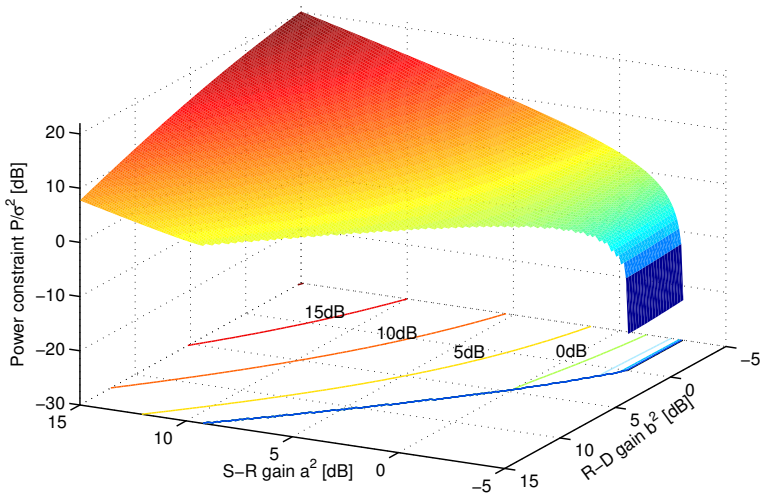


Figure 4.2. Upper bound for the normalized transmit power constraint P/σ^2 given different source-relay and relay-destination channel gains. Contour plots of the upper bound are shown at the bottom.

4.1.7 Numerical Results

In this section we illustrate numerically the achievable rate regions and the achievable rates R in symmetric setups for FNC, LNC, Lattice code, and NBF strategies, and compare them to time-sharing benchmark and the cut-set bound.

Achievable Rate Regions

In Figure 4.3, we plot the achievable rate regions for a scenario where the source $\mathcal{S}_1$ has transmit power P_1/σ^2=10dB, $\mathcal{S}_2$ has a power budget P_2/σ^2=5dB, $\mathcal{R}$ has transmit power constraint P_r/σ^2=5dB, the source-relay channel gain a^2=5dB, and the relay-destination channel gain b^2=5dB. Note that the NBF scheme actually achieves the cut-set bound in this setup, see Sec. 4.1.6 for a formal proof in the symmetric scenarios. The curves for FNC and Lattice code coincide each other as decoding at relay is not a constraint due to strong source-relay channels.

Symmetric Achievable Rates

In Figure 4.4 we investigate the impact of the source-relay link quality a^2 on the achievable rates for different cooperative strategies, with fixed transmit power P/σ^2=5dB and relay-destination channel gain b^2=0dB. Rate gains of network coding are significant in a large range of a^2 values. Note that when the source-relay link quality is comparable to the source-destination link, i.e., when a^2 is around 0dB, the lattice coding strategy is preferred over LNC or FNC due to the relaxed

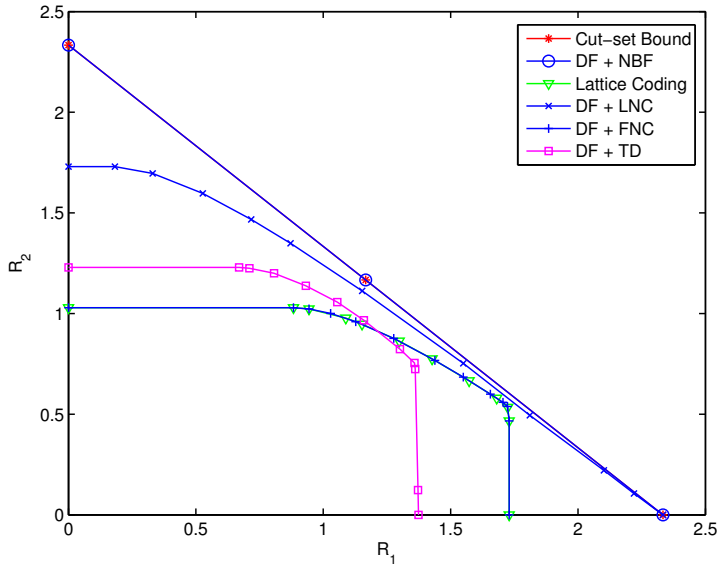


Figure 4.3. Achievable rate regions with transmit power $P_1/\sigma^2 = 10$dB, $P_2/\sigma^2 = 5$dB, $P_r/\sigma^2 = 5$dB, source-relay channel gain $a^2 = 5$dB, and relay-destination channel gain $b^2 = 5$dB. The cut-set outer bound is also plotted for reference. The curves for FNC and Lattice coding coincide with each other.

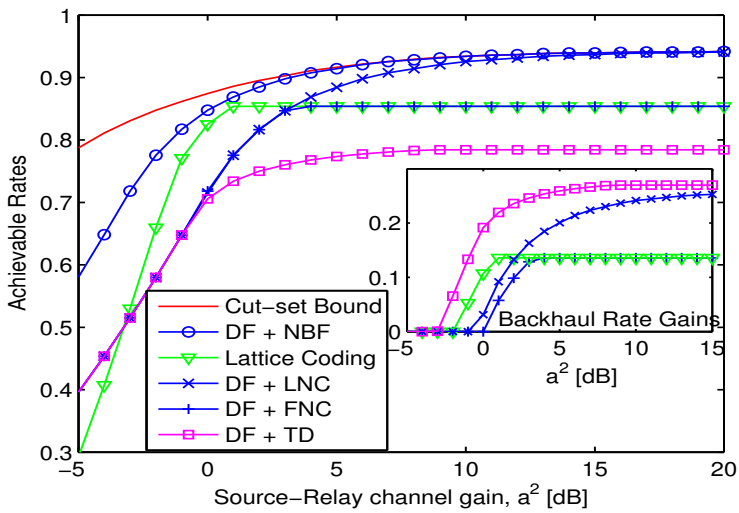


Figure 4.4. Effects of the source-relay channel gain a^2 on the achievable rates with $P/\sigma^2 = 5$dB and the relay-destination channel gain $b^2 = 0$dB.

decoding constraint at the relay node. The gain by using backhaul, as illustrated in the sub-figure, is significant for all schemes for a^2 larger than 0dB.

In Figure 4.5, we investigate the impact of the relay-destination link quality b^2 on the achievable rates for different cooperative strategies, with fixed transmit power P/σ^2=5dB and source-relay channel gain a^2=10dB (upper) and a^2=5dB (lower). With backhaul, substantial rate gains can be achieved by performing LNC or NBF compared to the time sharing relay. Significant gains can be achieved by utilizing the backhaul in the case of a poor relay-destination link (small b^2).

Comparison of NBF with Lattice Coding

As discussed in Remark 4.2, NBF and lattice coding based cooperation schemes complement each other in different channel setups. As illustrated in Figure 4.6, we compared their performance at a fixed transmit power P/σ^2=7dB but with varying source-relay and relay-destination channel qualities. The relative rate gain of NBF compared with lattice coding given different P/σ^2 is shown in the contour plots. NBF outperforms lattice coding uniformly in low SNR regions ($P/\sigma^2 \leq$5dB) and in medium SNR regions (5$<P/\sigma^2<$20dB) with relatively strong source-relay gain a^2. For high SNR regions ($P/\sigma^2>$20dB), lattice coding outperforms NBF for most of channel conditions.

In general, network coding based beamforming (NBF) strategies give the best performance, and the gap between the cut-set bound and the NBF achievable rate is not large. In high SNR regions, however, the lattice code based strategy is preferred. FNC, which only performs modulo-2 addition in the finite field, suffers limited performance loss in most of the cases. Further, and more importantly, we show significant rate gains compared to the scenarios without backhaul in various channel conditions.

4.1.8 Extension to General Channel Setups with Cross Links

With a high-rate backhaul the extension to non-symmetric channel gains is straightforward: labelling the channel gains by g_{ij} as in (2.1) and substituting them into the previous analysis where appropriate we will get the achievable rate regions and the cut-set bound in the general case. However, the results for the symmetric scenario where $R_1 = R_2 = R$ have to be modified since setting $\alpha_1 = \alpha_2$ may no longer be the optimal solution.

The extension to the case with cross links (i.e., $g_{12} > 0$ and $g_{21} > 0$) is also straightforward: replacing the channel model (4.1) with (2.1), formulating the mutual information constraints induced by decoding at destination $d \in \{\mathcal{D}_1, \mathcal{D}_2\}$ as $I(X_1X_2X_r; Y_d|\cdots)$, and translating the rate constraints into $\mathcal{C}(\cdot)$ expressions. However, it is not a good idea to apply such extension to the DF+FNC scheme since the signal from the cross-link will then be simply treated as noise. Therefore we should redesign DF+FNC to make better use of the cross links.

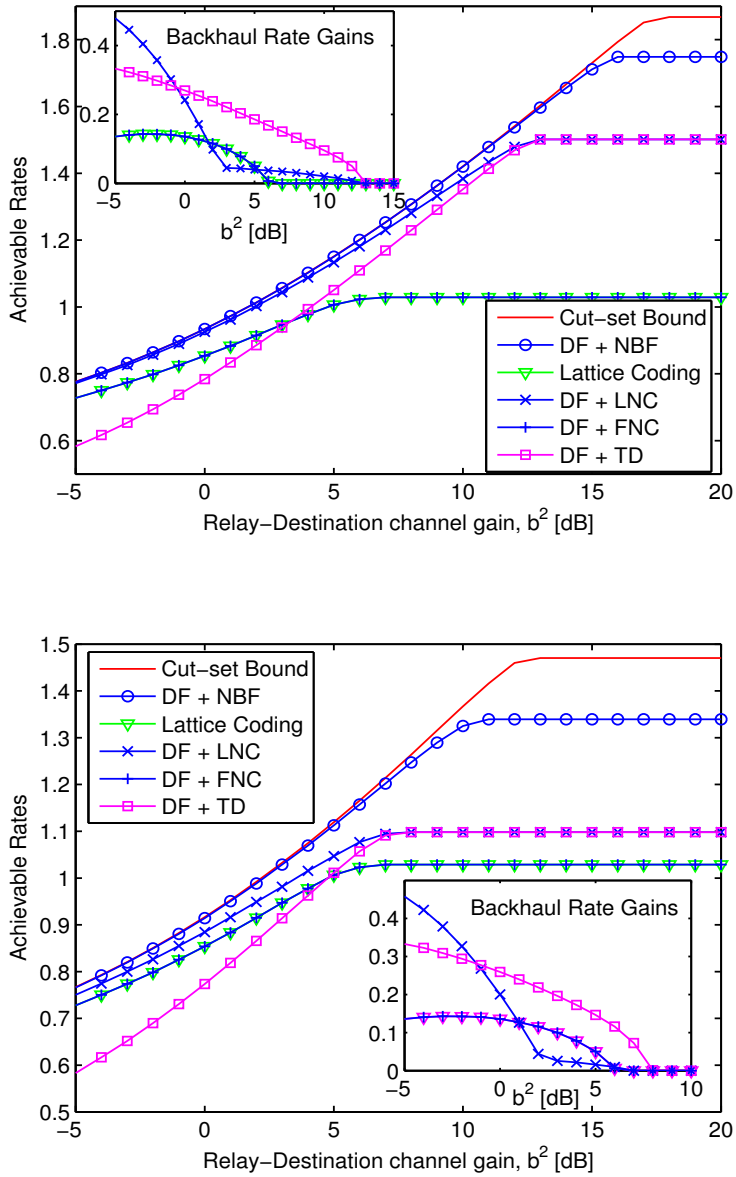


Figure 4.5. Effects of the relay-destination channel gain b^2 on the achievable rates with backhaul, when $P/\sigma^2 = 5$dB and the source-relay channel gain $a^2 = 10$dB (upper) and 5dB (lower). The rate gains (in bits per channel use) compared to the schemes without backhaul are also presented in the sub-figure.

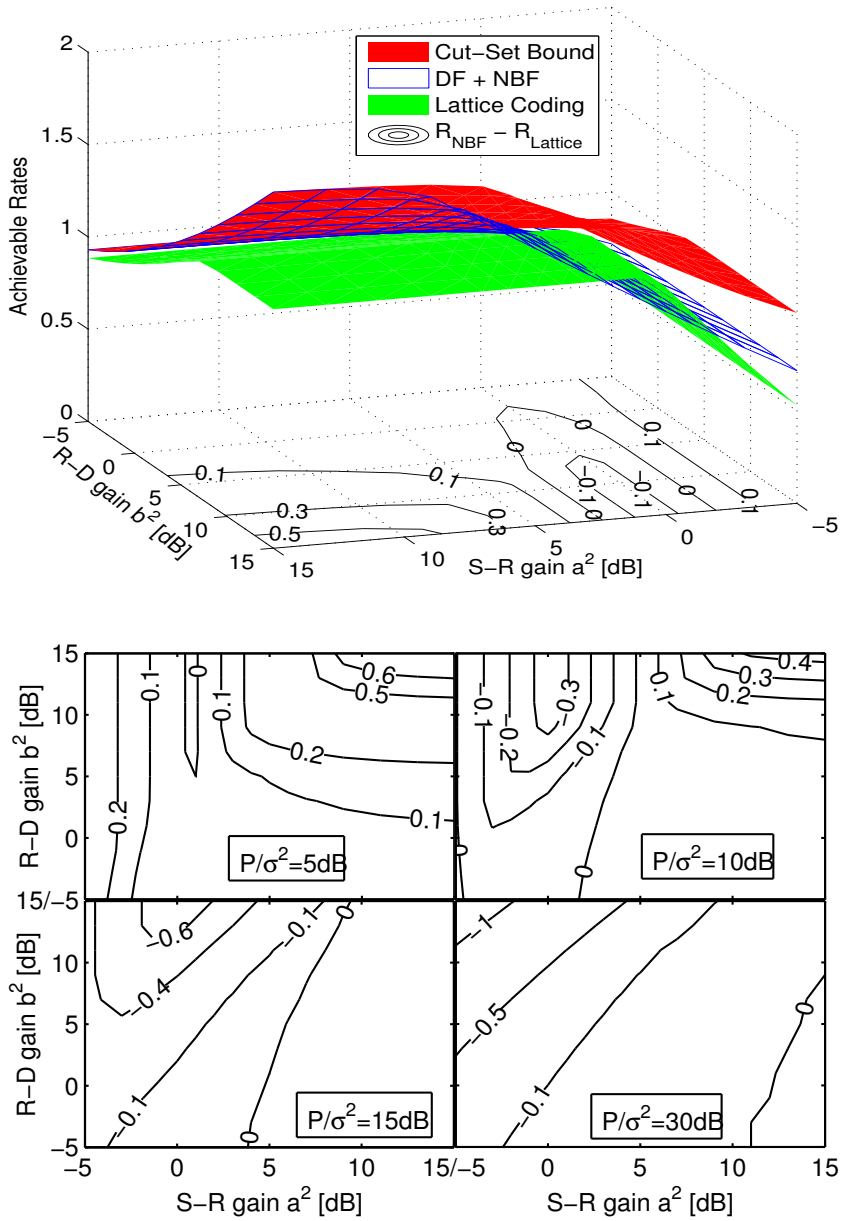


Figure 4.6. Comparison of DF+NBF (4.23) and Lattice Coding (4.19) with P/σ^2=7dB (upper). Contour plots given different transmit power constraints are also shown (lower).

As discussed in Remark 4.1, FNC induces extra rate penalty in asymmetric channel setups: the XOR operation appends zeros to the shorter message which is a waste of spectrum efficiency, and the achievable rate is constrained by the weakest relay-destination link quality. By utilizing a double-index codebook $U(W_1, W_2)$ instead of $U(W_1 \oplus W_2)$, the former penalty can be avoided and the latter can be reduced by backward decoding.

We first generate a random codebook $\{U^{(n)}\}$ of size $2^{n(R_1+R_2)}$ and label the codewords as $U^{(n)}(s,t)$, with $s \in \{1,...,2^{nR_1}\}$ and $t \in \{1,...,2^{nR_2}\}$. At the end of block $t-1$, $\mathcal{R}$ decodes $(W_{1,t-1}, W_{2,t-1})$ and then transmits the codeword $U^{(n)}(W_{1,t-1}, W_{2,t-1})$. For each codeword $U^{(n)}(W_{1,t-1}, W_{2,t-1})$, we generate a random codebook $\{V_1^{(n)}\}$ of size 2^{nR_1} , and then use this codebook to encode the new message $W_{1,t}$. We denote the selected codeword for $W_{1,t}$ as $V_1^{(n)}(W_{1,t}, W_{1,t-1}, W_{2,t-1})$. Similarly we generate $\{V_2^{(n)}\}$ of size 2^{nR_2} and choose $V_2^{(n)}(W_{2,t}, W_{1,t-1}, W_{2,t-1})$ for $W_{2,t}$. With power allocation parameters $0 \leq \alpha_1, \alpha_2 \leq 1$ to cooperate with $\mathcal{R}$, the transmitted signals at block t are therefore

$$\begin{aligned}
X_{1,t}^{(n)} &= \sqrt{\alpha_1 P_1} V_1^{(n)}(W_{1,t}, W_{1,t-1}, W_{2,t-1}) + \sqrt{(1-\alpha_1)P_1} U^{(n)}(W_{1,t-1}, W_{2,t-1}), \\
X_{2,t}^{(n)} &= \sqrt{\alpha_2 P_2} V_2^{(n)}(W_{2,t}, W_{1,t-1}, W_{2,t-1}) + \sqrt{(1-\alpha_2)P_2} U^{(n)}(W_{1,t-1}, W_{2,t-1}), \\
X_{r,t}^{(n)} &= \sqrt{P_r} U^{(n)}(W_{1,t-1}, W_{2,t-1}).
\end{aligned} \tag{4.30}$$

The relay performs *forward decoding* and the destinations use *backward decoding*. At the end of block t, the relay decodes $(W_{1,t}, W_{2,t})$ reliably from $Y_{r,t}^{(n)}$ and then selects the new codeword $U^{(n)}(W_{1,t}, W_{2,t})$ to be transmitted in next block. $\mathcal{D}_1$ and $\mathcal{D}_2$ start decoding after transmission in block $B+1$ has finished. In block $B+1$, no new message is transmitted and the received signal at $\mathcal{D}_1$ ($\mathcal{D}_2$) only depends on $(W_{1,B}, W_{2,B})$. After decoding $(W_{1,B}, W_{2,B})$ successfully, only $W_{1,B-1}$ ($W_{2,B-1}$) is unknown in $Y_{1,B}^{(n)}$ ($Y_{2,B}^{(n)}$), and we repeat this process backwards until all messages are recovered. Following a similar method as in Proposition 4.2, the achievable rate region can be characterized as follows

$$\begin{aligned}
R_1 &< \min\{I(X_1; Y_r|X_2X_r), I(X_1X_2X_r; Y_1|V_2), I(X_1X_2X_r; Y_2|V_2)\}, \\
R_2 &< \min\{I(X_2; Y_r|X_1X_r), I(X_1X_2X_r; Y_1|V_1), I(X_1X_2X_r; Y_2|V_1)\}, \\
R_1 + R_2 &< \min\{I(X_1X_2; Y_r|X_r), I(X_1X_2X_r; Y_1), I(X_1X_2X_r; Y_2)\},
\end{aligned} \tag{4.31}$$

which can be easily translated into $\mathcal{C}(\cdot)$ expressions by substituting (4.30) into it and choosing (U, V_1, V_2) to be Gaussian.

Note that the above scheme can be regarded as an extension of the DF+LNC scheme with the superposition encoding at the relay being replaced by the double-index encoding. A further extension is to combine the double-index encoding with block Markov encoding, which is exactly what we have done in DF+NBF. Therefore we will concentrate on NBF (rather than FNC or LNC) in the rest of this thesis for full source cooperation in general channel setups. The numerical results of

DF+NBF in general channel setups will be postponed to next chapter when CF/AF results are ready for comparison.

4.2 Cooperative Scheme with Partial Source Cooperation

To gain deeper understanding of source cooperation in a more realistic setting and demonstrate the benefit of combining source cooperation with relaying, in this section we focus on the low-rate backhaul scenario ($0 \le C_{12} < R_1, 0 \le C_{21} < R_2$).

As demonstrated in Sec. 4.1.7, the network coding based beamforming (NBF) strategy gives the best performance in general as it can greatly benefit from a coherent combining gain. In low-rate backhaul scenario, we can still enjoy such benefit by rate-splitting [Car78, RU96] and message exchange via backhaul. On the other hand, linear network coding (LNC) based scheme outperforms finite-field network coding (FNC) and time-sharing (TD) based schemes. Therefore in the following we only consider LNC as the combining method at the relay node. The lattice coding based scheme, which can beat NBF when decoding at relay becomes problematic, is non-trivial to be extended to the scenario with partial source cooperation. We will instead relax the decoding constraint at the relay node by allowing partial decoding at relay.

We first partition each source message into two parts $W_1=[W_{1c},\ W_{1p}]$ and $W_2=[W_{2c},\ W_{2p}]$, and then divide all the four messages evenly into B blocks $W_{1c,t}$, $W_{1p,t}$, $W_{2c,t}$, $W_{2p,t}$, each with nR_{1c}, nR_{1p}, nR_{2c}, nR_{2p}, bits, respectively. The transmission is completed in $B+2$ blocks[4], each with n channel uses. During block $t-1$, the sources exchange $(W_{1c,t}, W_{2c,t})$ over the conferencing links at rate $R_{1c} \le C_{12}$ and $R_{2c} \le C_{21}$, respectively, to formulate a common message $W_{c,t}=[W_{1c,t}, W_{2c,t}]$; during block t, $\mathcal{S}_1$ broadcasts $[W_{c,t}, W_{1p,t}]$ and $\mathcal{S}_2$ broadcasts $[W_{c,t}, W_{2p,t}]$ over the channel in cooperation with the relay's transmission.

4.2.1 Partial-Decode-and-Forward Relaying with Linear Network Coding (pDF+LNC)

Unlike the cooperative strategies with DF relaying proposed in [DXS11b], $\mathcal{R}$ here only needs to decode and forward some or all of the messages $(W_{1p,t}, W_{2p,t}, W_{c,t})$ depending on the channel quality, owning to the existence of cross-links. We propose a hybrid coding scheme termed partial-decode-and-forward based linear network coding (pDF+LNC). It essentially performs rate-splitting at the source nodes to exchange messages, partial decoding and LNC at the relay to reduce the rate constraints and superpose the decoded messages, and joint decoding at the destinations to enlarge the rate region. The codebook generation and encoding/decoding process are a natural extension of Theorem 1 of [SE07]. Given i.i.d. random variables $V_{1p}, V_{2p}, V_c \sim \mathcal{N}(0,1)$, we first generate independent codebooks $\{V_{1p}^{(n)}\}$, $\{V_{2p}^{(n)}\}$, and

[4]The first block ($t = 0$) involves only message exchanging via error-free backhaul but no transmission over the relay channel.

Table 4.2. Illustration of the encoding/decoding process for pDF+LNC with $W_{c,t} = [W_{1c,t}, W_{2c,t}]$ for $B = 2$ and full decoding at the relay.

$t=$	0	1	2	3
Backhaul	$W_{1c,1} \Leftrightarrow W_{2c,1}$	$W_{1c,2} \Leftrightarrow W_{2c,2}$	/	/
$\mathcal{S}_1$ TX	/	$W_{1p,1}, W_{c,1}$	$W_{1p,1}, W_{1p,2}, W_{c,1}, W_{c,2}$	$W_{1p,2}, W_{c,2}$
$\mathcal{S}_2$ TX	/	$W_{2p,1}, W_{c,1}$	$W_{2p,1}, W_{2p,2}, W_{c,1}, W_{c,2}$	$W_{2p,2}, W_{c,2}$
$\mathcal{R}$ RX	/	$W_{1,1}, W_{2,1} \rightarrow$	$W_{1,2}, W_{2,2} \rightarrow$	/
$\mathcal{R}$ TX	/	/	$W_{1p,1}, W_{2p,1}, W_{c,1}$	$W_{1p,2}, W_{2p,2}, W_{c,2}$
$\mathcal{D}$ RX	/	/	$W_{1,1}, W_{2,1}$	$\leftarrow W_{1,2}, W_{2,2}$

$\{V_c^{(n)}\}$, each of size $2^{nR_{1p}}$, $2^{nR_{2p}}$, and 2^{nR_c}, respectively. Then, for each index $k \in \{1, ..., 2^{nR_{1p}}\}$, we generate independently $2^{nR_{1p}}$ codewords $X_{1p}^{(n)}$ using distribution $\prod p(x_{1p}|v_{1p}(k))$, and label the codewords as $X_{1p}^{(n)}(n,k)$, where $n \in \{1, ..., 2^{nR_{1p}}\}$. We generate $X_{2p}^{(n)}$ and $X_c^{(n)}$ in a similar way. At block t, $\mathcal{S}_1$ transmits $[W_{c,t}, W_{1p,t}]$ and $\mathcal{S}_2$ transmits $[W_{c,t}, W_{2p,t}]$ in cooperation with the relay's transmission as follows

$$\begin{aligned}
X_{r,t}^{(n)} &= \sqrt{\alpha_r'} V_{1p}^{(n)}(W_{1p,t-1}) + \sqrt{\alpha_r''} V_{2p}^{(n)}(W_{2p,t-1}) + \sqrt{1-\alpha_r'-\alpha_r''} V_c^{(n)}(W_{c,t-1}), \\
X_{1,t}^{(n)} &= \sqrt{\alpha_1} X_{1p}^{(n)}(W_{1p,t}, W_{1p,t-1}) + \sqrt{\alpha_2} X_c^{(n)}(W_{c,t}, W_{c,t-1}) + \sqrt{\alpha_3} V_{1p}^{(n)} + \sqrt{\alpha_4} V_c^{(n)}, \\
X_{2,t}^{(n)} &= \sqrt{\alpha_5} X_{2p}^{(n)}(W_{2p,t}, W_{2p,t-1}) + \sqrt{\alpha_6} X_c^{(n)}(W_{c,t}, W_{c,t-1}) + \sqrt{\alpha_7} V_{2p}^{(n)} + \sqrt{\alpha_8} V_c^{(n)},
\end{aligned} \tag{4.32}$$

where $0 \le \alpha_r', \alpha_r'', \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8 \le 1$ are power allocation parameters with $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$ and $\alpha_5 + \alpha_6 + \alpha_7 + \alpha_8 = 1$.

The encoding/decoding process when $\mathcal{R}$ decodes all messages with $B = 3$ is illustrated in Table 4.2. $\mathcal{R}$ recovers $(W_{1p,t}, W_{2p,t}, W_{c,t})$ jointly at the end of block t by *forward decoding* [CE79] based on $Y_{r,t}^{(n)}$ after cancelling out $(W_{1p,t-1}, W_{2p,t-1}, W_{c,t-1})$. The destinations carry out *backward decoding* [Car82]: the received signal $Y_{1,B+1}^{(n)}$ ($Y_{2,B+1}^{(n)}$) only depends on $(W_{1p,B}, W_{2p,B}, W_{c,B})$, which can be retrieved by a joint typicality decoder; then we can proceed to $Y_{1,B}^{(n)}$ ($Y_{2,B}^{(n)}$) and repeat this process backwards until all messages are recovered.

Proposition 4.6. *Define $T = \{1p, 2p, c\}$, $T_Q \subseteq T$ and $T_Q \neq \emptyset$, the achievable rate region for pDF+LNC is the union of all (R_1, R_2) satisfying $R_1 \ge 0$, $R_2 \ge 0$, and*

$$\begin{cases}
R_1 = R_{1p} + R_{1c}, \\
R_2 = R_{2p} + R_{2c}, \\
R_c = R_{1c} + R_{2c}, \\
0 \le R_{1c} \le C_{12},\ 0 \le R_{2c} \le C_{21},\ R_{1p} \ge 0,\ R_{2p} \ge 0, \\
R(S \subseteq T) < \min\limits_{d \in \{1,2\}} I(X(S)X_r; Y_d | X(S^c)V(S^c)), \\
R(S_Q \subseteq T_Q) < I(X(S_Q); Y_r | X(S_Q^c)V(T_Q)),
\end{cases} \tag{4.33}$$

where $R(S) = \sum_{k\in S} R_k$, S^c (S_Q^c) *is the complementary subset of* S (S_Q) *with* $S \cup S^c = T$ ($S_Q \cup S_Q^c = T_Q$), *and the union is taken over all the power allocation parameters, and over all possible rate constraints* (R_{1p}, R_{2p}, R_c) *that are determined by the corresponding partial-DF cooperation strategies indicated by* T_Q. *Intermediate variables* R_{1p}, R_{2p}, R_c, R_{1c} *and* R_{2c} *can be easily removed by performing Fourier-Motzkin elimination.*

Proof. Proof outline.

There are 7 different partial decoding options at the relay, namely, decoding only $W_{1p,t}, W_{2p,t}, W_{c,t}$, $(W_{1p,t}, W_{2p,t})$, $(W_{1p,t}, W_{c,t})$, $(W_{2p,t}, W_{c,t})$, or $(W_{1p,t}, W_{2p,t}, W_{c,t})$, resulting in 7 different rate constraints (R_{1p}, R_{2p}, R_c). We therefore introduce an auxiliary random variable Q to indicate different partial DF strategies and any combinations of them by arbitrary time-sharing. If the relay decodes $(W_{1p,t}, W_{2p,t}, W_{c,t})$ (i.e., $T_Q = T$), by performing *forward decoding* [CE79] at the relay and *backward decoding* [Car82] at destinations, we can get from Theorem 1 of [SE07] that

$$\begin{aligned} R(S \subseteq T) < \min\{ & I(X(S); Y_r | X(S^c) X_r), \\ & I(X(S) X_r; Y_1 | X(S^c) V(S^c)), \\ & I(X(S) X_r; Y_2 | X(S^c) V(S^c))\}, \end{aligned} \tag{4.34}$$

with variables defined as in (4.32) and (3.1). By enforcing $p(x_{1p}|v_{1p})$, $p(x_{2p}|v_{2p})$, and $p(x_c|v_c)$ to be normal distributions, and applying the fact that $V_{1p}, V_{2p}, V_c \sim \mathcal{N}(0,1)$ into (4.32), all the mutual information constraints in (4.34) can be translated into corresponding $\mathcal{C}(\cdot)$ expressions

$$R_{1p} = \min\left\{ \mathcal{C}(\alpha_1\gamma_{1r}), \min_{d\in\{1,2\}} \mathcal{C}(\alpha_1\gamma_{1d} + (\sqrt{\alpha_3\gamma_{1d}} + \sqrt{\alpha_r'\gamma_{rd}})^2) \right\},$$

$$R_{2p} = \min\left\{ \mathcal{C}(\rho_1\gamma_{2r}), \min_{d\in\{1,2\}} \mathcal{C}(\rho_1\gamma_{2d} + (\sqrt{\rho_3\gamma_{2d}} + \sqrt{\alpha_r''\gamma_{rd}})^2) \right\},$$

$$\begin{aligned} R_c = \min\Big\{ & \mathcal{C}((\sqrt{\alpha_2\gamma_{1r}} + \sqrt{\rho_2\gamma_{2r}})^2), \\ & \min_{d\in\{1,2\}} \mathcal{C}((\sqrt{\alpha_2\gamma_{1d}} + \sqrt{\rho_2\gamma_{2d}})^2 + (\sqrt{\bar{\alpha}_r\gamma_{rd}} + \sqrt{\alpha_4\gamma_{1d}} + \sqrt{\rho_4\gamma_{2d}})^2) \Big\}, \end{aligned}$$

$$\begin{aligned} R_{1p} + R_{2p} = \min\Big\{ & \mathcal{C}(\alpha_1\gamma_{1r} + \rho_1\gamma_{2r}), \\ & \min_{d\in\{1,2\}} \mathcal{C}(\alpha_1\gamma_{1d} + \rho_1\gamma_{2d} + (\sqrt{\alpha_3\gamma_{1d}} + \sqrt{\alpha_r'\gamma_{rd}})^2 + (\sqrt{\rho_3\gamma_{2d}} + \sqrt{\alpha_r''\gamma_{rd}})^2) \Big\}, \end{aligned}$$

$$\begin{aligned} R_{1p} + R_c = \min\Big\{ & \mathcal{C}(\alpha_1\gamma_{1r}) + (\sqrt{\alpha_2\gamma_{1r}} + \sqrt{\rho_2\gamma_{2r}})^2), \\ & \min_{d\in\{1,2\}} \mathcal{C}(\alpha_1\gamma_{1d} + (\sqrt{\alpha_3\gamma_{1d}} + \sqrt{\alpha_r'\gamma_{rd}})^2 + (\sqrt{\alpha_2\gamma_{1d}} + \sqrt{\rho_2\gamma_{2d}})^2 \\ & \qquad + (\sqrt{\bar{\alpha}_r\gamma_{rd}} + \sqrt{\alpha_4\gamma_{1d}} + \sqrt{\rho_4\gamma_{2d}})^2) \Big\}, \end{aligned}$$

and

$$R_{2p}+R_c=\min\Big\{\ \mathcal{C}((\rho_1\gamma_{2r}+\sqrt{\alpha_2\gamma_{1r}}+\sqrt{\rho_2\gamma_{2r}})^2),$$
$$\min_{d\in\{1,2\}}\mathcal{C}(\rho_1\gamma_{2d}+(\sqrt{\rho_3\gamma_{2d}}+\sqrt{\alpha_r''\gamma_{rd}})^2+(\sqrt{\alpha_2\gamma_{1d}}+\sqrt{\rho_2\gamma_{2d}})^2$$
$$+(\sqrt{\bar{\alpha}_r\gamma_{rd}}+\sqrt{\alpha_4\gamma_{1d}}+\sqrt{\rho_4\gamma_{2d}})^2)\Big\},$$

$$R_{1p}+R_{2p}+R_c=\min\Big\{\ \mathcal{C}(\gamma_{1r}(\alpha_1+\alpha_2)+\gamma_{2r}(\rho_1+\rho_2)+2\sqrt{\alpha_2\rho_2\gamma_{1r}\gamma_{2r}}),$$
$$\min_{d\in\{1,2\}}\mathcal{C}(\gamma_{1d}+\gamma_{2d}+\gamma_{rd}+2\sqrt{\gamma_{1d}\gamma_{2d}}(\sqrt{\alpha_2\rho_2}+\sqrt{\alpha_4\rho_4})$$
$$+2\sqrt{\gamma_{1d}\gamma_{rd}}(\sqrt{\alpha_3\alpha_r'}+\sqrt{\alpha_4\bar{\alpha}_r})+2\sqrt{\gamma_{2d}\gamma_{rd}}(\sqrt{\rho_3\alpha_r''}+\sqrt{\rho_4\bar{\alpha}_r}))\Big\}.$$

By performing Fourier-Motzkin elimination we can straightforwardly translate the above expressions into the corresponding rate region of (R_1,R_2) for $T_Q=T$.

If the relay only decodes $W_{c,t}$ ($T_Q=\{c\}$), we have

$$\begin{aligned}R(S\subseteq T)&<\min_{d\in\{1,2\}}I(X(S)X_r;Y_d|X(S^c)V(S^c)),\\ R_c&<I(X_c;Y_r|V_c),\end{aligned}\tag{4.35}$$

with variables defined as in (4.32) and (3.1) but with $\alpha_r'=\alpha_r''=\alpha_3=\alpha_7=0$. The case when the relay only decodes $W_{1p,t}$ (for $T_Q=\{1p\}$) or $W_{2p,t}$ (for $T_Q=\{2p\}$) is handled similarly.

If the relay decodes $W_{1p,t},X_{2p,t}$ but not $W_{c,t}$ (for $T_Q=\{1p,2p\}$), we can obtain

$$\begin{aligned}R(S\subseteq T)&<\min_{d\in\{1,2\}}I(X(S)X_r;Y_d|X(S^c)V(S^c)),\\ R_{1p}&<I(X_{1p};Y_r|X_{2p}V_{1p}V_{2p}),\\ R_{2p}&<I(X_{2p};Y_r|X_{1p}V_{1p}V_{2p}),\\ R_{1p}+R_{2p}&<I(X_{1p}X_{2p};Y_r|V_{1p}V_{2p}),\end{aligned}\tag{4.36}$$

with variables defined as in (4.32) and (3.1) but with $\alpha_r'+\alpha_r''=1$ and $\alpha_4=\alpha_8=0$. It is similar for scenarios when the relay does not decode $W_{1p,t}$ (for $T_Q=\{2p,c\}$) or $W_{2p,t}$ (for $T_Q=\{1p,c\}$).

For other values of Q, different partial DF strategies are used in a time-sharing fashion. The achievable rate region in (4.33) is therefore the union of all the different regions resulting from different partial decoding strategies. □

The pDF+LNC strategy requires a smart relay which can adopt a proper encoding/decoding scheme depending on the effective link SNR γ, in addition to a powerful joint typicality decoder. Based on the design metric (e.g. maximizing the sum rate) and the values of $\gamma_{i,j}$, the same optimization process can be carried out at both the relay and source nodes, resulting in an operation point (R_1,R_2)

on the boundary of the achievable rate region together with a group of operating parameters $(Q, R_{1p}, R_{2p}, R_{1c}, R_{2c}, \alpha_r', \alpha_r'', \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)$. Ideally such optimization process can be carried out on the fly to adaptively update the operating parameters. For practical implementation however, we need to form a lookup table for $\mathcal{R}$ which contains $(Q, R_{1p}, R_{2p}, R_{1c}, R_{2c}, \alpha_r', \alpha_r'')$ and is indexed by quantized link SNR $\tilde{\gamma}_{ij}$, $i, j = 1, 2, r$. The lookup tables for $\mathcal{S}_1$ and $\mathcal{S}_2$ are created in a similar way. Note that the quantization should satisfy $\tilde{\gamma} \leq \gamma$ to avoid link outage and hence results in a loss of spectrum efficiency. If the quantization resolution is properly selected, the complexity of implementing a lookup table can be marginal compared to the joint typicality decoder equipped by the DF relay.

4.3 Summary

In this chapter we have investigated DF relaying based cooperative strategies with different NC schemes, namely, finite field network coding, linear network coding, lattice coding. For high-rate backhaul, we have shown that the cut-set bound can be achieved by network coding based beamforming when the signal-to-noise ratios lie in the sphere defined by the source-relay and relay-destination channel gains. In general, the network coding based beamforming strategy gives the best performance. In high SNR regions, however, the lattice code based strategy is preferred. FNC, which only performs modulo-2 addition in the finite field, suffers limited performance loss in most of the cases. Further, and more importantly, we have shown significant rate gains compared to the scenarios without backhaul in various channel conditions. For low-rate backhaul scenarios, we propose a new coding scheme, partial-decode-and-forward based linear network coding, which is essentially a hybrid scheme utilizing rate-splitting and messages conferencing at the source nodes, partial decoding and linear network coding at the relay, and joint decoding at each destination.

4.4 Appendix

4.4.A Proof of the Capacity Achieving Case

From (4.23) and (4.28) we can capture the effective power gain as follows

$$g_{\text{NBF}} = \max_{0 \leq \alpha \leq 1} \min\left\{4a^2\alpha,\ 1 + b^2 + 2b\sqrt{1-\alpha}\right\}, \tag{4.37}$$

$$g_{\text{cut-set}} = \sup_{0 \leq \alpha, \rho \leq 1} \min\left\{1 + b^2 + 2b\sqrt{1-\alpha},\ \alpha + a^2(2\alpha + 2\alpha\rho + (1-\rho^2)\alpha^2 P)\right\}. \tag{4.38}$$

From (4.37) it is straightforward to shown that

$$g_{\text{NBF}} = \begin{cases} 4a^2, & \text{if } 4a^2 \leq 1 + b^2; \\ 1 + b^2 + 2b\sqrt{1-\alpha^*}, & \text{otherwise,} \end{cases} \tag{4.39}$$

where $a^* \in [0,1]$ satisfies $4a^2\alpha^* = 1+b^2+2b\sqrt{1-\alpha^*}$. For the second part of (4.38), we have

$$\begin{aligned}
&\max_{0\le\alpha,\rho\le 1} \alpha + a^2(2\alpha + 2\alpha\rho + (1-\rho^2)\alpha^2 P) \\
&= \max_{0\le\alpha,\rho\le 1} \alpha + a^2(2\alpha + \alpha^2 P + 1/P - P(1/P - \alpha\rho)^2) \\
&= \begin{cases} 1 + a^2(2 + P + \frac{1}{P}), & \text{if } P>1, \text{ [by setting } \alpha=1, \rho=\frac{1}{P}] \\ 1 + 4a^2, & \text{if } P\le 1, \text{ [by setting } \alpha=\rho=1] \end{cases}
\end{aligned} \tag{4.40}$$

By combining (4.38) and (4.40) one can easily conclude that

$$g_{\text{cut-set}} = \begin{cases} 1 + 4a^2, & \text{if } 4a^2 \le b^2 \text{ and } P \le 1; \\ > 1 + 4a^2, & \text{if } 4a^2 \le b^2 \text{ and } P > 1; \\ 1 + b^2 + 2b\sqrt{1-\alpha^*}, & \text{otherwise;} \end{cases} \tag{4.41}$$

where $0 \le \alpha^* \le 1$ satisfies the equality

$$1 + b^2 + 2b\sqrt{1-\alpha^*} = \alpha^* + a^2(2\alpha^* + 2\alpha^*\rho + (1-\rho^2)(\alpha^*)^2 P).$$

From (4.41) it clearly follows that $g_{\text{cut-set}} > 4a^2$ for the scenarios when $4a^2 \le 1 + b^2$. Therefore $g_{\text{NBF}} = g_{\text{cut-set}}$ is possible only if $4a^2 > 1 + b^2$, i.e., there should exist two variables $0 \le \alpha^*, \rho \le 1$ such that

$$4a^2\alpha^* = 1 + b^2 + 2b\sqrt{1-\alpha^*}, \tag{4.42a}$$

$$4a^2\alpha^* = \alpha^* + a^2(2\alpha^* + 2\alpha^*\rho + (1-\rho^2)(\alpha^*)^2 P). \tag{4.42b}$$

By subtracting $1 + b^2$ from both sides of (4.42a) and then taking square, we have

$$16a^4(\alpha^*)^2 - \alpha^*(8a^2 + 8a^2b^2 - 4b^2) + (1-b^2)^2 = 0,$$

which has only one true root for (4.42a) (must satisfy $4a^2\alpha^* > 1 + b^2$)

$$\alpha^* = \frac{2a^2(1+b^2) - b^2 + \sqrt{(4a^2 - b^2)(4a^2 - a)b^2}}{8a^4}. \tag{4.43}$$

From (4.42b) we get

$$\rho\alpha^* = 1/P + \sqrt{\alpha^*/(a^2P) + (\alpha^* - 1/P)^2},$$

or

$$\rho\alpha^* = 1/P - \sqrt{\alpha^*/(a^2P) + (\alpha^* - 1/P)^2}.$$

Since $0 \le \rho\alpha^* \le \alpha^*$, the first root is obviously a false root and therefore omitted. To make the second root satisfy the constraint, we must have

$$0 \le 1/P - \sqrt{\alpha^*/(a^2P) + (\alpha^* - 1/P)^2} \le \alpha^*.$$

The second inequality is self-evident, and the first inequality requires

$$a^2 > 1/2 \text{ and } \alpha^* \leq \frac{1}{P}\left(2 - \frac{1}{a^2}\right). \tag{4.44}$$

We therefore conclude from (4.43) and (4.44), given $4a^2 > 1 + b^2$ and $P > 0$, that (4.42) holds if and only if $4a^2 > 2$ and

$$\frac{2a^2(1+b^2) - b^2 + \sqrt{(4a^2 - b^2)(4a^2 - a)b^2}}{8a^4} \leq \frac{1}{P}\left(2 - \frac{1}{a^2}\right).$$

Combined with the finding that $g_{\text{NBF}} = g_{\text{cut-set}}$ is impossible for $4a^2 \leq 1 + b^2$, we can conclude that $g_{\text{NBF}} = g_{\text{cut-set}}$, i.e. (4.23) and (4.28) are identical, if and only if (a^2, b^2, P) satisfies (4.29).

Chapter 5

Compression/Amplification Based Cooperation Schemes

In this Chapter we extend noisy network coding to use source cooperation with the help of the theory of network equivalence. We show that when partial cooperation between source nodes is possible, short-message noisy network coding with message exchange can achieve a strictly larger rate region than noisy network coding. A low-complexity alternative scheme, analog network coding based on amplify-and-forward relaying, is also investigated and shown to benefit greatly from the help of the backhaul and it can even outperform noisy network coding when the coherent combining gain is dominant.

5.1 Noisy Network Coding (NNC)

The standard compress-and-forward (CF) relaying strategy [CE79] provides the destination node(s) with a noisy yet structured observation (compression) of its received signal via the use of an independent codebook. When CF relaying is used, the source performs block Markov encoding and transmits an independent short message in each time slot via random coding. The relay performs compression based on Wyner–Ziv binning and the destination performs block-by-block forward successive decoding (first decoding the bin index and then the message). The use of Wyner–Ziv binning at the relay makes it nontrivial to extend CF to multiple relays [KGG05].

The quantize-map-and-forward (QMF) protocol proposed in [ADT11], which utilizes symbol-by-symbol scalar quantization at the relay and joint decoding at the destination, has been proved to be approximately optimal (within a constant gap to the cut-set bound) for unicast layered networks with multiple relays, and for non-layered unicast networks via time extension. The noisy network coding (NNC) protocol [LKEC11] can be regarded as an extension of QMF. NNC, which performs repetition coding at source nodes, compression without using Wyner–Ziv binning

(vector quantization) at relays, and simultaneous joint message and compression index decoding at destinations, can be easily extended to multiple-source and/or multiple-relay scenarios. Contrary to CF where one large message is first partitioned into many blocks and then transmitted, NNC encodes the large message directly and transmits over all the time slots, each with an independent codebook. After all the transmissions are completed, the destination decodes the large message and all the compression indices jointly. NNC has been shown in [LKEC11] to recover the rate region achieved by CF in the classical 3-node relay model, and outperform CF in the two-way relay channel, the interference relay channel, and the multiple-relay channel discussed in [KGG05].

The basic principle of NNC, as described in [LKEC11], is to convey a "super message" N times, each time using an independent codebook and letting $N\to\infty$, before the destination(s) can successfully decode the message. Given the orthogonal (i.e., out-of-band) conferencing bit-pipes between source nodes, cooperation in NNC can be realized in the following two ways: compression forwarding or message exchange. Since compression forwarding in NNC is realized by noisy transmission (relaying) among source/relay nodes, it is not clear how it can be implemented via the noiseless backhaul. On the other hand, since NNC with message exchange will require the common message to be identified in the first block already such that the repetition coding can start with, it requires a $N\to\infty$ times higher backhaul rate to exchange the super message. How to optimally utilize the rate-limited backhaul in NNC is interesting but yet to be determined.

5.1.1 NNC with Message Exchange via Backhaul

To avoid excessive delay incurred by super message exchange via backhaul, NNC with message exchange can be done in a segment-by-segment fashion[1]: partition the total message blocks into K segments, each with B blocks; during the first segment NNC is used without cooperation but the backhaul is used to exchange messages to be transmitted in segment 2; in segment $m>1$ the common message exchanged via backhaul in segment $m-1$ is transmitted together with private messages at each source node using superposition encoding. This however requires long memories and long encoding delay that grows with $n \times B$, where n is codeword length (i.e., the number of channel uses in each transmission).

On the other hand, as stated in [LKEC11], the achievable rate of NNC with finite B and sufficiently long codeword length ($n \to \infty$) for the classical 3-node relay network shown in Figure 2.3 can be characterized by

$$R_B < \frac{B-1}{B} R_\infty - \frac{I(\hat{Y}_2; Y_2|X_2)}{B}, \tag{5.1}$$

[1]The cooperation for NNC will be done over message blocks within the same segment and no cooperation across segments.

where

$$R_{\infty}=\min\{I(X_1;Y_3\hat{Y}_2|X_2), I(X_1X_2;Y_3)-I(Y_2;\hat{Y}_2|X_1X_2Y_3)\} \tag{5.2}$$

is the rate of NNC when $B\to\infty$. Therefore NNC with segment-wise cooperation can achieve rate R_B for all the K segments except for the first one where no cooperation is done (hence a lower rate than R_B). As the cooperation block number B cannot be very large due to memory constraint, NNC with cooperation will still incur a rate loss.

5.1.2 NNC with Compression Forwarding via Backhaul

At transmission block $t=1,...,B$, source node $\mathcal{S}_1$ forwards a compression

$$X_{s1,t}^{(n)}=f_1(W_1,X_{s1}^{t-1},X_{s2}^{t-1})$$

via backhaul to $\mathcal{S}_2$ at rate C_{12} bits per channel use and $\mathcal{S}_2$ forwards

$$X_{s2,t}^{(n)}=f_2(W_2,X_{s2}^{t-1},X_{s1}^{t-1})$$

to $\mathcal{S}_1$, where

$$X_{s1}^{t-1}=[X_{s1,1}^{(n)},...,X_{s1,t-1}^{(n)}]$$

and $f_1(\cdot)$ and $f_2(\cdot)$ are some compression functions. At block $t+1$, $\mathcal{S}_1$ broadcasts $X_{1,t+1}^{(n)}$ which is determined based on W_1 and $X_{s2,t}^{(n)}$. With compression forwarding through the backhaul, the source nodes essentially behave as a relay node to help the delivery of each other's messages, which fits well into the framework of NNC and therefore only some slight modification on the encoding/decoding process is needed for the extension.

Although there are many different ways to design the compression functions, the optimal way is yet to be determined based on the available backhaul capacity, channel settings, and message delivery requirement. One way to design the compression functions is to mimic a noisy observation of each other's signal and such noise observation is conveyed via the backhaul and then used to generate a compression index as in the normal NNC strategy, as presented in the following.

According to the theory of network equivalence [KEM11], the capacity of a network is unchanged if any independent[2], memoryless, point-to-point channel in this network is replaced by a noiseless bit-pipe with throughput equal to the removed channel's capacity. Since the conferencing bit-pipes between two source nodes are independent and orthogonal to all the other transmissions, they can be

[2]The transition probability of the channel can be separated from the transition probability of the rest of the network, i.e., a product of these two functions can represent the transition of the whole network. A more formal definition can be found in Definition 6.1.

replaced [KEM11] by noisy channels with the same capacity as follows:

$$C_{12}: Y_{s2} = \sqrt{P_1} X_{s1} + Z_{s2}, \text{ with } \mathcal{C}(P_1) = \frac{1}{2}\log(1+P_1) = C_{12},$$
$$C_{21}: Y_{s1} = \sqrt{P_2} X_{s2} + Z_{s1}, \text{ with } \mathcal{C}(P_2) = \frac{1}{2}\log(1+P_2) = C_{21}, \tag{5.3}$$

where $X_{s1}, X_{s2}, Z_{s1}, Z_{s2}$ are independent Gaussian[3] random variables with zero-mean and unit-variance, P_1, P_2 are corresponding power constraints, and Y_{s1}, Y_{s2} are the conferencing outputs at source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. Note that signals in (5.3) are orthogonal to all the other transmissions and therefore will not mix with signals (e.g. X_1, X_2) in (3.1). Now we can extend the NNC scheme [LKEC11], originally designed for noisy cooperation (relaying) among source/relay nodes, to our setup with orthogonal conferencing error-free bit-pipes.

Proposition 5.1. *An achievable rate region of NNC with conferencing encoders is obtained as the union of all rate pairs (R_1, R_2) that satisfy $R_1 \geq 0$, $R_2 \geq 0$, and*

$$\begin{aligned}
R_1 < \Delta_{R_1} + \min\Bigg\{ & \mathcal{C}\left(\gamma_{11} + \frac{\gamma_{1r}}{1+\sigma_r^2}\right),\ \mathcal{C}\left(\gamma_{12} + \frac{\gamma_{1r}}{1+\sigma_r^2}\right), \\
& \mathcal{C}(\gamma_{11}+\gamma_{r1}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right),\ \mathcal{C}(\gamma_{12}+\gamma_{r2}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right)\Bigg\}, \\
R_2 < \Delta_{R_2} + \min\Bigg\{ & \mathcal{C}\left(\gamma_{21} + \frac{\gamma_{2r}}{1+\sigma_r^2}\right),\ \mathcal{C}\left(\gamma_{22} + \frac{\gamma_{2r}}{1+\sigma_r^2}\right), \\
& \mathcal{C}(\gamma_{21}+\gamma_{r1}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right),\ \mathcal{C}(\gamma_{22}+\gamma_{r2}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right)\Bigg\}, \\
R_1 + R_2 < \Delta_{R_s} + \min\Bigg\{ & \mathcal{C}(\gamma_{11}+\gamma_{21}+\gamma_{r1}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right),\ \mathcal{C}(\gamma_{12}+\gamma_{22}+\gamma_{r2}) - \mathcal{C}\left(\frac{1}{\sigma_r^2}\right), \\
& \mathcal{C}\left(\gamma_{11}+\gamma_{21} + \frac{\gamma_{1r}+\gamma_{2r} + (\sqrt{\gamma_{11}\gamma_{2r}} - \sqrt{\gamma_{21}\gamma_{1r}})^2}{1+\sigma_r^2}\right), \\
& \mathcal{C}\left(\gamma_{12}+\gamma_{22} + \frac{\gamma_{1r}+\gamma_{2r} + (\sqrt{\gamma_{12}\gamma_{2r}} - \sqrt{\gamma_{22}\gamma_{1r}})^2}{1+\sigma_r^2}\right)\Bigg\},
\end{aligned} \tag{5.4}$$

where

$$\Delta_{R_1} = \mathcal{C}\left(\frac{P_1}{1+\sigma_2^2}\right) - \mathcal{C}\left(\frac{1}{\sigma_1^2}\right),$$
$$\Delta_{R_2} = \mathcal{C}\left(\frac{P_2}{1+\sigma_1^2}\right) - \mathcal{C}\left(\frac{1}{\sigma_2^2}\right),$$
$$\Delta_{R_s} = -\mathcal{C}\left(\frac{1}{\sigma_1^2}\right) - \mathcal{C}\left(\frac{1}{\sigma_2^2}\right),$$

[3]In [KEM11] the noisy channel is only required to have the same capacity as the bit-pipe's throughput, with no restriction on the channel input or output. By restricting ourselves to Gaussian signals, the capacity of the overall network will not be increased, and therefore we still have a valid inner bound to the capacity.

the value of P_1, P_2 is determined by conferencing rate (C_{12}, C_{21}) as defined in (5.3)*, and the union is taken over all $\sigma_1^2, \sigma_2^2, \sigma_r^2 > 0$.*

Proof. Given the set of transmitting nodes $T = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{R}\}$ and the set of sink nodes $D = \{\mathcal{D}_1, \mathcal{D}_2\}$, an achievable rate region of NNC for the multicast relay network in Figure 3.1 can be specialized from Theorem 1 of [LKEC11] as follows

$$\sum_{k\in S} R_k < \min_{d\in D} I(X(S); \hat{Y}(S^c)Y(d)|X(S^c)Q) - I(Y(S); \hat{Y}(S)|X(T)\hat{Y}(S^c)Y(d)Q),$$

where $\hat{Y}$ is the compressed versions of Y, Q is the time-sharing parameter, S, S^c are any pair of complementary subsets of T, i.e., $S \cup S^c = T$ and $S \cap S^c = \emptyset$, with

$$\begin{aligned}
&X(\mathcal{S}_1) = \{X_1, X_{s1}\},\ X(\mathcal{S}_2) = \{X_2, X_{s2}\},\ X(\mathcal{R}) = X_r,\\
&X(T) = \{X_1X_2X_rX_{s1}X_{s2}\},\\
&Y(\mathcal{S}_1) = Y_{s1},\ Y(\mathcal{S}_2) = Y_{s2},\ Y(\mathcal{R}) = Y_r,\ Y(\mathcal{D}_1) = Y_1,\ Y(\mathcal{D}_2) = Y_2.
\end{aligned}$$

The achievable rate region is therefore defined by

$$\begin{aligned}
R_1 < \min_{d\in\{1,2\}} \Big\{ & \\
& I(X_1X_{s1}; \hat{Y}_{s2}\hat{Y}_rY_d|X_2X_{s2}X_rQ) - I(Y_{s1}; \hat{Y}_{s1}|X_1X_2X_rX_{s1}X_{s2}\hat{Y}_{s2}\hat{Y}_rY_dQ),\\
& I(X_1X_{s1}X_r; \hat{Y}_{s2}Y_d|X_2X_{s2}Q) - I(Y_rY_{s1}; \hat{Y}_r\hat{Y}_{s1}|X_1X_2X_rX_{s1}X_{s2}\hat{Y}_{s2}Y_dQ)\Big\},
\end{aligned} \tag{5.5}$$

$$\begin{aligned}
R_2 < \min_{d\in\{1,2\}} \Big\{ & \\
& I(X_2X_{s2}; \hat{Y}_{s1}\hat{Y}_rY_d|X_1X_{s1}X_rQ) - I(Y_{s2}; \hat{Y}_{s2}|X_1X_2X_rX_{s1}X_{s2}\hat{Y}_{s1}\hat{Y}_rY_dQ),\\
& I(X_2X_{s2}X_r; \hat{Y}_{s2}Y_d|X_1X_{s1}Q) - I(Y_rY_{s2}; \hat{Y}_r\hat{Y}_{s2}|X_1X_2X_rX_{s1}X_{s2}\hat{Y}_{s1}Y_dQ)\Big\},
\end{aligned}$$

$$\begin{aligned}
R_1+R_2 < \min_{d\in\{1,2\}} \Big\{ & \\
& I(X_1X_2X_{s1}X_{s2}; \hat{Y}_rY_d|X_rQ) - I(Y_{s1}Y_{s2}; \hat{Y}_{s1}\hat{Y}_{s2}|X_1X_2X_rX_{s1}X_{s2}\hat{Y}_rY_dQ),\\
& I(X_1X_2X_rX_{s1}X_{s2}; Y_d|Q) - I(Y_rY_{s1}Y_{s2}; \hat{Y}_r\hat{Y}_{s1}\hat{Y}_{s2}|X_1X_2X_rX_{s1}X_{s2}Y_dQ)\Big\},
\end{aligned}$$

with the joint probability $p(q, x_1, x_2, x_r, x_{s1}, x_{s2}, y_r, y_{s1}, y_{s2}, \hat{y}_r, \hat{y}_{s1}, \hat{y}_{s2})$ partitioned as

$$\begin{aligned}
&p(q)p(x_1|q)p(x_2|q)p(x_r|q)p(x_{s1}|q)p(x_{s2}|q)\\
&\qquad\qquad \times p(\hat{y}_r|x_r, y_r, q)p(\hat{y}_{s1}|x_1, y_{s1}, q)p(\hat{y}_{s2}|x_2, y_{s2}, q).
\end{aligned}$$

By setting $Q{=}\emptyset$ and

$$\begin{aligned}
\hat{Y}_r^{(n)} &= Y_r^{(n)} + \hat{Z}_r^{(n)},\ \text{with } \hat{Z}_r \sim \mathcal{N}(0, \sigma_r^2),\\
\hat{Y}_{s1}^{(n)} &= Y_{s1}^{(n)} + \hat{Z}_1^{(n)},\ \text{with } \hat{Z}_1 \sim \mathcal{N}(0, \sigma_1^2),\\
\hat{Y}_{s2}^{(n)} &= Y_{s2}^{(n)} + \hat{Z}_2^{(n)},\ \text{with } \hat{Z}_2 \sim \mathcal{N}(0, \sigma_2^2),
\end{aligned}$$

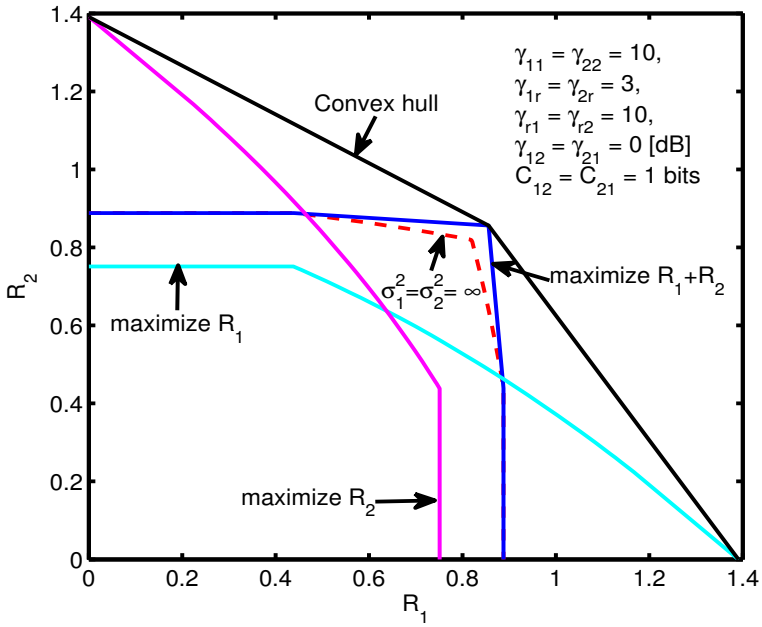


Figure 5.1. Achievable rate region of NNC with compression forwarding via backhaul, achieved by time-sharing among rate optimization of R_1, R_2, and R_1+R_2, respectively. The SNR parameters are heuristically chosen.

and applying (3.1), (3.2) and (5.3) into (5.5), we can find the region (5.4). □

Note that σ_i^2, $i=1,2,r$ refers to the controllable quantization noise power induced by noisy compression at $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{R}$, respectively, which leads to a rate penalty $-\mathcal{C}(\frac{1}{\sigma_i^2})$. Rate contributions $\mathcal{C}(\frac{P_1}{1+\sigma_2^2})$ and $\mathcal{C}(\frac{P_2}{1+\sigma_1^2})$ are due to noisy relaying of the conferencing messages. Since $\Delta_{R_s}\leq 0$ with equality if and only if $\sigma_1^2=\sigma_2^2=\infty$, i.e., no source cooperation via conferencing links, we have to compute the rate region for noisy NC in three steps: first generate the rate region of noisy NC without utilizing conferencing links; then compute rate regions by maximizing R_1, R_2, and R_1+R_2, respectively; finally, apply time-sharing among different optimization schemes to get the rate region, as illustrated in Figure 5.1. Note that maximization of R_1+R_2 is not always necessary. For example, if $0<C_{12}, C_{21}\leq\frac{1}{2}$, we have $P_1\leq 1$ and $P_2\leq 1$ according to (5.3). In such scenario, for any $0<\sigma_1^2, \sigma_2^2<\infty$ we have

$$\Delta_{R_s} < 0,$$
$$\Delta_{R_1}+\Delta_{R_2} = \mathcal{C}\left(\frac{P_1}{1+\sigma_2^2}\right)-\mathcal{C}\left(\frac{1}{\sigma_2^2}\right)+\mathcal{C}\left(\frac{P_2}{1+\sigma_1^2}\right)-\mathcal{C}\left(\frac{1}{\sigma_1^2}\right)<0,$$

which means the sum rate R_1+R_2 cannot be increased. Details on optimization of R_1, R_2, and R_1+R_2 can be found in Appendix 5.6.A.

5.2 Short-message Noisy Network Coding (SNNC)

Wu and Xie have pointed out in [WX10] that the superiority of NNC over CF is due to the postponed decoding process, rather than the large-message repetition coding or the joint decoding of message and compression indexes. By using classical short-message encoding at the source node, compression without binning at relay nodes, block-by-block backward decoding (either successively or jointly) at the destination, one can achieve the same rate as NNC, as long as the relaying signal can be treated as noise at destinations. This new scheme is coined short-message noisy network coding (SNNC) in [KH11b] and the rate-equivalence proof of NNC and SNNC for the single-source multiple-relay network is given in [WX10]. A simpler alternative rate-equivalence proof has been given for the single-source single-relay network [KH11b] and the single-source multiple-relay network [KH11a]. The rate equivalence can also be established for the multiple-source multiple-relay network [Kra12].

Using SNNC instead of regular NNC reduces encoding delay while still allowing for extensions to multiple-source/relay/destination networks. In addition, the short-message transmission facilitates source cooperation, in the presence of backhaul or a conferencing channel between source nodes.

In the rest of this section, we show that when partial cooperation between source nodes is possible, SNNC can achieve a strictly larger rate region than NNC by performing rate-splitting [RU96], message exchanging, and superposition coding with proper power allocation at source nodes.

5.2.1 SNNC with Partial Source Cooperation

SNNC divides the super message W_1 (W_2) evenly into B short messages $W_{1,t}$ ($W_{2,t}$), $t = 1, ..., B$, which facilitates message exchange and therefore can benefit from coherent combining gain when cooperation is done as follows:

1) rate splitting at source nodes:
   $W_{1,t} = [W_{1p,t}, W_{1c,t}], W_{2,t} = [W_{2p,t}, W_{2c,t}]$;

2) $W_{1c,t}$ and $W_{2c,t}$ are exchanged via backhaul and then formulate $W_{c,t} = [W_{1c,t}, W_{2c,t}]$ before transmission t;

3) source node performs superposition coding:
   $X_1^{(n)} = \sqrt{\alpha_1} X_{1p}^{(n)}(W_{1p,t}) + \sqrt{\bar{\alpha}_1} X_c^{(n)}(W_{c,t})$, and
   $X_2^{(n)} = \sqrt{\alpha_2} X_{2p}^{(n)}(W_{2p,t}) + \sqrt{\bar{\alpha}_2} X_c^{(n)}(W_{c,t})$.

In the following we will analyze the performance of SNNC in the multiple-source multicast relay network as in Figure 3.1. Note that the source cooperation scheme for SNNC proposed above applies to any network where a pair of source nodes are connected by conferencing links, although the corresponding achievable

rate expressions have to be adjusted according to the message delivery requirement (unicast, multicast, etc.).

With proper power allocation $\alpha_1, \alpha_2 \in [0,1]$, the received signals can be written as follows

$$\begin{aligned}
Y_1^{(n)} &= \sqrt{\alpha_1\gamma_{11}}X_{1p}^{(n)} + \sqrt{\alpha_2\gamma_{21}}X_{2p}^{(n)} + (\sqrt{\bar{\alpha}_1\gamma_{11}} + \sqrt{\bar{\alpha}_2\gamma_{21}})X_c^{(n)} + \sqrt{\gamma_{r1}}X_r^{(n)} + Z_1^{(n)},\\
Y_2^{(n)} &= \sqrt{\alpha_1\gamma_{12}}X_{1p}^{(n)} + \sqrt{\alpha_2\gamma_{22}}X_{2p}^{(n)} + (\sqrt{\bar{\alpha}_1\gamma_{12}} + \sqrt{\bar{\alpha}_2\gamma_{22}})X_c^{(n)} + \sqrt{\gamma_{r2}}X_r^{(n)} + Z_2^{(n)},\\
Y_r^{(n)} &= \sqrt{\alpha_1\gamma_{1r}}X_{1p}^{(n)} + \sqrt{\alpha_2\gamma_{2r}}X_{2p}^{(n)} + (\sqrt{\bar{\alpha}_1\gamma_{1r}} + \sqrt{\bar{\alpha}_2\gamma_{2r}})X_c^{(n)} + Z_r^{(n)}.
\end{aligned} \tag{5.6}$$

By message exchanging and superposition coding at the source nodes, we have transferred the 2-source relay network described in (3.1) to a network with 3 independent sources $(X_{1p}^{(n)}, X_{2p}^{(n)}, X_c^{(n)})$. The corresponding rate constraints R_{1p}, R_{2p}, R_c can be obtained straightforwardly by applying SNNC/NNC results on independent sources.

Define $T = \{\mathcal{S}_{1p}, \mathcal{S}_{2p}, \mathcal{S}_c\}$ and $D = \{\mathcal{D}_1, \mathcal{D}_2\}$ with

$$\begin{aligned}
X(T) = \{X_{1p}X_{2p}X_c\},\ X(\mathcal{S}_{1p}) = X_{1p},\ X(\mathcal{S}_{2p}) = X_{2p},\\
X(\mathcal{S}_c) = X_c, Y(D) = \{Y_1Y_2\}, Y(\mathcal{D}_1) = Y_1,\ Y(\mathcal{D}_2) = Y_2.
\end{aligned}$$

Let S, S^c be any pair of complementary subsets of T, i.e., $S \cup S^c = T$ and $S \cap S^c = \emptyset$, we can define the following rate

$$R(S) = \min_{d\in D} \max \left\{ \begin{array}{l} I(X(S); Y(d)|Q), \\ \min[I(X(S); \hat{Y}_r Y(d)|X(S^c)X_rQ), \\ \qquad I(X(S)X_r; Y(d)|X(S^c)Q) - I(\hat{Y}_r; Y_r|X(T)X_rY(d)Q)] \end{array} \right\}.$$

By applying the results of SNNC/NNC [LKEC11, Theorem 1] into (5.6), the achievable rate region of SNNC with partial source cooperation can therefore be described as the union of all rate pairs (R_1, R_2) that satisfy

$$\left\{ \begin{array}{l} R_1 = R_{1p} + R_{1c}, \\ R_2 = R_{2p} + R_{2c}, \\ 0 \le R_{1c} \le C_{12}, \\ 0 \le R_{2c} \le C_{21}, \\ 0 \le R_{1p} < R(\{\mathcal{S}_{1p}\}), \\ 0 \le R_{2p} < R(\{\mathcal{S}_{2p}\}), \\ R_{1c} + R_{2c} < R(\{\mathcal{S}_c\}), \\ R_{1p} + R_{2p} < R(\{\mathcal{S}_{1p}, \mathcal{S}_{2p}\}), \\ R_{1p} + R_{1c} + R_{2c} < R(\{\mathcal{S}_{1p}, \mathcal{S}_c\}), \\ R_{2p} + R_{1c} + R_{2c} < R(\{\mathcal{S}_{2p}, \mathcal{S}_c\}), \\ R_{1p} + R_{2p} + R_{1c} + R_{2c} < R(\{\mathcal{S}_{1p}, \mathcal{S}_{2p}, \mathcal{S}_c\}), \end{array} \right. \tag{5.7}$$

with the union taken over all joint distributions that can be factorized as

$$p(q)p(x_{1p}|q)p(x_{2p}|q)p(x_c|q)p(x_r|q)p(\hat{y}_r|x_r, y_r, q). \tag{5.8}$$

By setting $Q{=}\emptyset$ and $\hat{Y}_r^{(n)}{=}Y_r^{(n)}{+}\hat{Z}_r^{(n)}$ with $\hat{Z}_r{\sim}\mathcal{N}(0,\sigma^2)$, all these mutual information constraints in (5.7) can be translated into $\mathcal{C}(\cdot)$ expressions which are functions of $(\alpha_1,\alpha_2,\sigma^2)$. We then introduce notations $I_A,...,I_G$ to indicate these constraints (depending only on $\alpha_1,\alpha_2,\sigma^2$ and link SNR γ_{ij}) as follows

$$\begin{aligned}
I_A &= R(\{\mathcal{S}_{1p}\}),\\
I_B &= R(\{\mathcal{S}_{2p}\}),\\
I_C &= R(\{\mathcal{S}_c\}),\\
I_D &= R(\{\mathcal{S}_{1p},\mathcal{S}_{2p}\}),\\
I_E &= R(\{\mathcal{S}_{1p},\mathcal{S}_c\}),\\
I_F &= R(\{\mathcal{S}_{2p},\mathcal{S}_c\}),\\
I_G &= R(\{\mathcal{S}_{1p},\mathcal{S}_{2p},\mathcal{S}_c\}).
\end{aligned}$$

Remark 5.1. *Given $(\alpha_1,\alpha_2,\sigma^2,\gamma_{ij})$, we can conclude that $I_A \le I_E$ and $I_C \le I_E$ but NOT $I_E \le I_A{+}I_C$ due to the minimization in $R(S)$. Similarly, we have $\max\{I_D,I_E,I_F\} \le I_G$ but NOT $I_G \le I_A+I_F$, $I_G \le I_B+I_E$, or $I_G \le I_C+I_D$.*

After performing Fourier–Motzkin elimination over (5.7), the achievable rate region of SNNC with partial source cooperation for the system (3.1) is the union of all rate pairs (R_1,R_2) that satisfy $R_1 \ge 0$, $R_2 \ge 0$, and

$$\left\{\begin{array}{rl}
R_1 & < \min\{I_A+C_{12},\ I_A+I_C,\ I_E\},\\
R_2 & < \min\{I_B+C_{21},\ I_B+I_C,\ I_F\},\\
R_1+R_2 & < \min\{I_G, I_A+I_B+I_C, I_E+I_B, I_F+I_A,\\
& \qquad I_D+I_C, I_D+C_{12}+C_{21}, \dfrac{I_D+I_E+I_F}{2}\Big\},\\
2R_1+R_2 & < I_D+I_E+C_{12},\\
R_1+2R_2 & < I_D+I_F+C_{21},
\end{array}\right. \tag{5.9}$$

where the union operation is taken over all $\alpha_1,\alpha_2 \in [0,1]$ and $\sigma^2 \in (0,\infty)$.

The standard CF scheme, which uses short messages, can also benefit from source cooperation in the same way as in SNNC. When a single Wyner–Ziv binning process is used at the relay node, by generalizing the results in [GSG+10] for the scenario with a single relay and two independent sources, the achievable rate region of the standard CF with partial source cooperation can be described in the same way as in (5.7) but with

$$R(S) = \min_{d\in D}\ I(X(S);\hat{Y}_rY(d)|X(S^c)X_rQ), \tag{5.10}$$

$$\text{subject to } \max_{d\in D}\ I(\hat{Y}_r;Y_r|X_rY(d)Q) \le \min_{d\in D} I(X_r;Y(d)|Q), \tag{5.11}$$

where the distribution is partitioned as in (5.8).

Remark 5.2. *The achievable rates by CF with cooperation will in general be smaller than the rates obtained by SNNC. For scenarios with multiple sources and*

multiple destinations (e.g. interference relay channel), decoding at different destinations require different level of side information, which is impossible when a single Wyner–Ziv binning process is used. Besides, successful decoding of the same binning index is confined by the weakest channel, as demonstrated in (5.11)*. The gain of NNC (and thus SNNC) over CF in terms of sum rate has been demonstrated in [LKEC11] for the two-way relay channel and the interference relay channel. In addition, SNNC with source cooperation can be easily extended to multiple-source multiple-relay scenarios, which is not the case for standard CF due to to the presence of Wyner–Ziv binning.*

5.2.2 SNNC vs. NNC

For NNC with compression forwarding via backhaul, the transmitted signal $X_{1,t}^{(n)}$ ($X_{2,t}^{(n)}$) is a function of the large message W_1 (W_2), which is not known at the other source. Hence there is no coherent combining gain for NNC with compression forwarding, which is not the case for SNNC and CF where rate-splitting and message exchange is used with superposition coding.

NNC with message exchange under segment-wise cooperation can achieve the rate R_B as defined in (5.1) for all the K segments except for the first one where no cooperation is done (hence a lower rate than R_B). The corresponding achievable rate for SNNC with the same number of message blocks is

$$R_{B\times K} < \frac{B\times K-1}{B\times K+M}R_\infty - \frac{I(\hat{Y}_r;Y_r|X_r)}{B\times K},$$

where R_∞ is defined as in (5.2), and $M \ll B\times K$ is the extra blocks to ensure the start of backward decoding [WX10]. As the cooperation block number B cannot be very large due to memory constraint, NNC with message exchange will still incur a rate loss compared to SNNC.

In absence of memory and encoding delay constraints (i.e., $B\to\infty$), and introducing fading (known only at the receiver side, either relay or destinations), NNC may actually be advantageous, as there will be many blocks conveying the same message, ergodic rates prevail while short messages, even with backward decoding are subject to outages. This issue will be investigated in the future work.

To illustrate the gain of SNNC with partial source cooperation over NNC and CF, we have plotted in Figure 5.2 the achievable rate region for the system in Figure 3.1 in the asymmetric channel setting. The outer bound is obtained based on the genie-aided cut-set bound proposed in Sec. 3.3 and the inner bounds of NNC are obtained with and without compression forwarding as proposed in Sec. 5.1.2. With partial cooperation, SNNC achieves a strictly larger rate region than NNC with compression forwarding, and the gain comes from coherent combining at all the receiving nodes. CF with message exchange, on the other hand, performs better than NNC (even with compression forwarding) when coherent combining gain is large ($C_{12}=0.9$), but worse than NNC (even without cooperation) when the

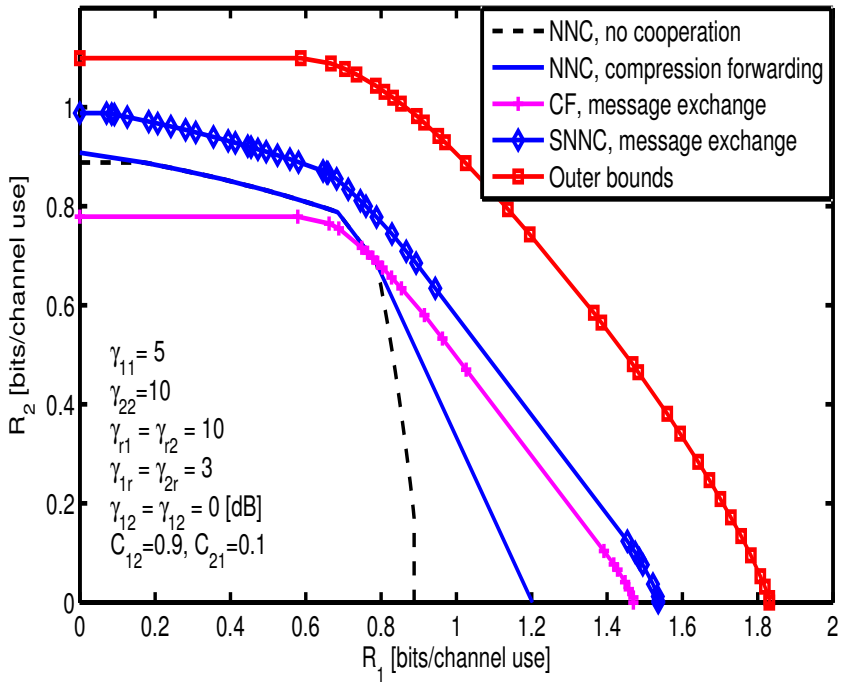


Figure 5.2. Achievable rate regions of SNNC in the asymmetric channel setting with source conferencing rate $C_{12} = 0.9$ and $C_{21} = 0.1$ bits per channel use. NNC with no cooperation is plotted as reference.

asymmetric channel setting effect in (5.11) dominates the coherent combining gain $C_{21} = 0.1$). The degradation of CF compared to SNNC and is due to the asymmetric channel setting which causes different requirement on successive decoding of the Wyner–Ziv binning index, as explained in Remark 5.2, see also [WX10].

5.3 Amplify-and-Forward as Analog Network Coding (AF+ANC)

When the relay cannot support encoding/decoding or interference cancellation, advanced cooperative strategies such as pDF or NNC/SNNC cannot be used. As suggested by [KMG+07, MGM12], AF relaying as analog NC (AF+ANC) is an attractive option in high SNR regimes. In this setup, the relay forwards a scaled version of the signal received during the previous period. Three independent random codebooks $\{V_{1p,t}^{(n)}\}$ of size $2^{nR_{1p}}$, $\{V_{2p,t}^{(n)}\}$ of size $2^{nR_{2p}}$, and $\{V_{c,t}^{(n)}\}$ of size $2^{n(R_{1c}+R_{2c})}$, are generated to encode $W_{1p,t}$, $W_{2p,t}$ and $W_{c,t}$, respectively. At block

Table 5.1. Illustration of the encoding/decoding process for AF+ANC with $W_{c,t} = [W_{1c,t}, W_{2c,t}]$ for $B = 2$.

$t=$	0	1	2	3
Backhaul	$W_{1c,1} \Leftrightarrow W_{2c,1}$	$W_{1c,2} \Leftrightarrow W_{2c,2}$	/	/
$\mathcal{S}_1$ TX	/	$(W_{1p,1}, W_{c,1})$	$(W_{1p,2}, W_{c,2})$	
$\mathcal{S}_2$ TX	/	$(W_{2p,1}, W_{c,1})$	$(W_{2p,2}, W_{c,2})$	
$\mathcal{R}$ TX	/	/	$W_{1p,1}, W_{2p,1}, W_{c,1}$	$W_{1p,2}, W_{2p,2}, W_{c,2}$
$\mathcal{D}$ RX	/	/	$W_{1,1}, W_{2,1} \rightarrow$	$W_{1,2}, W_{2,2}$

t, the transmitted signals are

$$\begin{aligned}
X_{1,t}^{(n)} &= \sqrt{\bar{\alpha}_1} V_{1p,t}^{(n)}(W_{1p,t}) + \sqrt{\alpha_1} V_{c,t}^{(n)}(W_{c,t}), \\
X_{2,t}^{(n)} &= \sqrt{\bar{\alpha}_2} V_{2p,t}^{(n)}(W_{2p,t}) + \sqrt{\alpha_2} V_{c,t}^{(n)}(W_{c,t}), \\
X_{r,t}^{(n)} &= \beta \left(\sqrt{\gamma_{1r}\bar{\alpha}_1} V_{1p,t-1}^{(n)} + \sqrt{\gamma_{2r}\bar{\alpha}_2} V_{2p,t-1}^{(n)} + (\sqrt{\gamma_{1r}\alpha_1} + \sqrt{\gamma_{2r}\alpha_2}) V_{c,t-1}^{(n)} + Z_{r,t-1}^{(n)} \right),
\end{aligned} \tag{5.12}$$

where $0 \leq \alpha_1, \alpha_2 \leq 1$ are power allocation parameters with $\bar{\alpha}_1 = 1-\alpha_1$ and $\bar{\alpha}_2 = 1-\alpha_2$. β is the amplifying factor at the relay to satisfy the power constraint (3.2), i.e.,

$$\beta^2 = \frac{1}{E[\mathrm{Var}(Y_{r,t})]} = \frac{1}{1 + \gamma_{1r} + \gamma_{2r} + 2\sqrt{\gamma_{1r}\gamma_{2r}\alpha_1\alpha_2}}.$$

Note that β is defined in a different way from that in [MGM12] to guarantee the SNR level at the destination nodes after multiple-hop AF relaying. The corresponding received signals at $\mathcal{D}_1$ at block t and $t+1$ are

$$\begin{aligned}
Y_{1,t}^{(n)} &= \beta\sqrt{\gamma_{r1}} \left(\sqrt{\gamma_{1r}\bar{\alpha}_1} V_{1p,t-1}^{(n)} + \sqrt{\gamma_{2r}\bar{\alpha}_2} V_{2p,t-1}^{(n)} + (\sqrt{\gamma_{1r}\alpha_1} + \sqrt{\gamma_{2r}\alpha_2}) V_{c,t-1}^{(n)} \right) \\
&\quad + \sqrt{\gamma_{11}\bar{\alpha}_1} V_{1p,t}^{(n)} + \sqrt{\gamma_{21}\bar{\alpha}_2} V_{2p,t}^{(n)} + (\sqrt{\gamma_{11}\alpha_1} + \sqrt{\gamma_{21}\alpha_2}) V_{c,t}^{(n)} + \tilde{Z}_{1,t}^{(n)}, \qquad (5.13) \\
Y_{1,t+1}^{(n)} &= \beta\sqrt{\gamma_{r1}} \left(\sqrt{\gamma_{1r}\bar{\alpha}_1} V_{1p,t}^{(n)} + \sqrt{\gamma_{2r}\bar{\alpha}_2} V_{2p,t}^{(n)} + (\sqrt{\gamma_{1r}\alpha_1} + \sqrt{\gamma_{2r}\alpha_2}) V_{c,t}^{(n)} \right) \qquad (5.14) \\
&\quad + \sqrt{\gamma_{11}\bar{\alpha}_1} V_{1p,t+1}^{(n)} + \sqrt{\gamma_{21}\bar{\alpha}_2} V_{2p,t+1}^{(n)} + (\sqrt{\gamma_{11}\alpha_1} + \sqrt{\gamma_{21}\alpha_2}) V_{c,t+1}^{(n)} + \tilde{Z}_{1,t+1}^{(n)},
\end{aligned}$$

where $\tilde{Z}_{1,t}$, $t = 1, ..., B+1$ are i.i.d. Gaussian with zero-mean and variance $\sigma_1^2 = 1 + \beta^2\gamma_{r1}$. Similarly, we can write $(Y_{2,t}^{(n)}, Y_{2,t+1}^{(n)})$ by introducing i.i.d. Gaussian $\tilde{Z}_{2,t}$ with zero-mean and variance $\sigma_2^2 = 1 + \beta^2\gamma_{r2}$. The destination nodes $\mathcal{D}_1$ and $\mathcal{D}_2$ perform *sliding window* [Wil82] joint decoding: at the end of block $t+1$, assuming $W_{1,t-1}$ and $W_{2,t-1}$ have been decoded successfully, $\mathcal{D}_1$ can jointly decode $W_{1,t}$ and $W_{2,t}$ from $(Y_{1,t}^{(n)}, Y_{1,t+1}^{(n)})$ and $\mathcal{D}_2$ can decode based on $(Y_{2,t}^{(n)}, Y_{2,t+1}^{(n)})$. The encoding/decoding process is illustrated in Table 5.1.

Proposition 5.2. *Define $T = \{1p, 2p, c\}$ and a pair of its complementary subsets S and S^c, i.e. $S \cup S^c = T$ and $S \cap S^c = \emptyset$, the achievable rate region for AF+ANC*

is the union over all (R_1, R_2) *satisfying* $R_1 \geq 0$, $R_2 \geq 0$, *and*

$$\begin{cases} R_1 = R_{1p} + R_{1c}, \\ R_2 = R_{2p} + R_{2c}, \\ R_c = R_{1c} + R_{2c}, \\ 0 \leq R_{1c} \leq C_{12}, \\ 0 \leq R_{2c} \leq C_{21}, \\ R_{1p} \geq 0,\ R_{2p} \geq 0, \\ \sum_{k \in S} R_k < \min\limits_{d \in \{1,2\}} \{I(V_t(S); Y_{d,t} Y_{d,t+1} | V_{t-1}(T) V_t(S^c))\}, \end{cases} \tag{5.15}$$

where

$$V_{t-1}(T) = \{V_{1p,t-1}, V_{2p,t-1}, V_{c,t-1}\},$$

and the union is taken over all subsets $S \subseteq T$ *and over all power allocation parameters* $0 \leq \alpha_1, \alpha_2 \leq 1$. *The compact rate region described by* (R_1, R_2) *can be straightforwardly obtained by performing Fourier-Motzkin elimination to remove the intermediate variables* R_{1p}, R_{2p}, R_c, R_{1c} *and* R_{2c}.

Proof. The proof can be found in Appendix 5.6.B. □

5.4 Numerical Results

5.4.1 Full Source Cooperation

We compare the lower bounds obtained by the SNNC and the AF+ANC schemes with different link quality to the cut-set bound developed in Section 3.2 and the NBF lower bound developed in Section 4.1.4. Unless stated otherwise, the following heuristic parameters will be used: The $\mathcal{S}_1$-$\mathcal{D}_1$ link SNR γ_{11}=5dB, the $\mathcal{S}_2$-$\mathcal{D}_2$ link SNR γ_{22}=10dB, the source-relay link SNR γ_{1r}=γ_{2r}=10dB, the relay-destination link SNR γ_{r1}=γ_{r2}=10dB, and the cross-link SNR γ_{12}=γ_{21}=0dB.

We investigate the impact of the relay-destination link quality on the capacity region. When the $\mathcal{R}$-$\mathcal{D}_1$ link is not strong, the $\mathcal{R}$-$\mathcal{D}_2$ link SNR γ_{r2} is not a limiting factor and therefore the capacity will monotonically increase with γ_{r1} until it is large enough to reach the bottleneck set by γ_{r2}, as demonstrated in Figure 5.3. Similar observation can be obtained for SNNC with message exchange. The gap between the cut-set bound the NBF lower bound is less than 0.05 bits/channel use. Note that the achievable rate for AF+ANC is not monotonically increasing with γ_{r1} due to the fact that both signal and noise are amplified. Such non-monotonic behavior will diminish as the source-relay channel becomes better, i.e., when AF becomes asymptotically optimal.

In Figure 5.4 we show the impact of the source-relay link with strong cross-link quality. When the source-relay link is poor, the NBF scheme suffers great rate loss due to the decoding constraint at the relay node. SNNC with message exchange and AF+ANC, on the other hand, have no such constraint and therefore can benefit

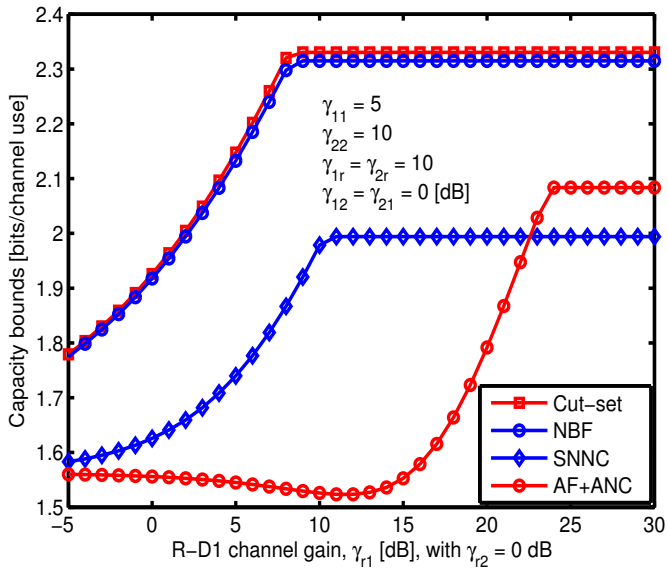


Figure 5.3. Capacity bounds for varying $\mathcal{R}$-$\mathcal{D}_1$ link SNR γ_{r1}, with $\gamma_{r2} = 0$dB.

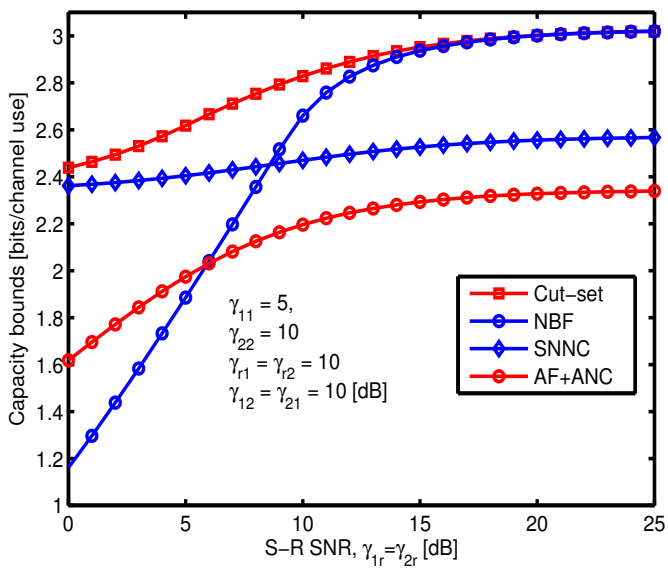


Figure 5.4. Capacity bounds for varying symmetric source-relay SNR $\gamma_{1r} = \gamma_{2r}$, with cross-link SNR γ_{12}=γ_{21} = 10dB.

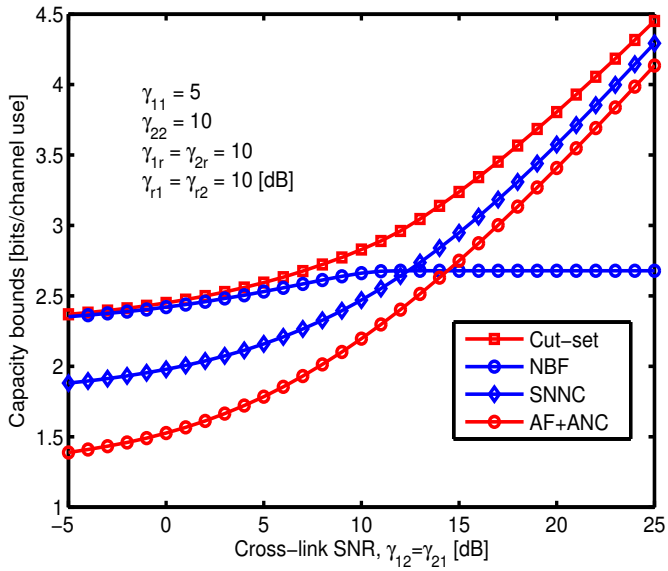


Figure 5.5. Capacity bounds for varying symmetric cross-link SNR γ_{12}=γ_{21}.

from the strong cross-link. As γ_{sr} increases, decoding at the relay node becomes trivial and NBF can benefit greatly from source-relay cooperation and therefore outperforms NNC/SNNC with message exchange and AF+ANC.

When the source-relay link is good, as shown in Figure 5.5, the achievable rate of NBF increases with improved cross-link quality until the decoding at the relay becomes the bottleneck. The cut-set bound and the other two schemes have no restriction on decoding at $\mathcal{R}$ and therefore will increase with the improved cross-link quality.

5.4.2 Partial Source Cooperation

We illustrate the inner and outer bounds to the capacity region based on numerical computation, with channel SNR chosen heuristically. The outer bounds are obtained based on the genie-aided cut-set bounds developed in Section 3.3. The inner bound of NNC with compression forwarding is based on the scheme proposed in Section 5.1.2. As stated in Sec. 4.2.1, computation of the rate region of pDF+LNC requires a union operation over eight independent auxiliary variables and seven partial decoding combinations, making it hard to characterize the exact inner bound numerically. In the following results we simply set α_3=α_4=0 and α_7=α_8=0 in (4.32), i.e. no source-relay cooperation, to lower bound the performance of pDF+LNC when source conferencing is possible.

The benefit of using the backhaul has been illustrated in Figure 5.6. Without source cooperation, achievable rates for AF/DF relaying based schemes are limited by noise propagation and decoding constraints when the source-relay link is poor. When source cooperation is possible, these constraints can be greatly reduced. The difference of rate regions for SNNC with message exchange and NNC with compression forwarding in Figure 5.6 (lower) is the gain of message exchange over compression forwarding. A benchmark scheme based on DF relaying with no source conferencing from [GSG$^+$10, Proposition 4] with R_3=0 has been plotted in Figure 5.6 (upper) for reference. The gain of pDF+LNC (with source-relay cooperation) over the benchmark is due to partial decoding at the relay. The gap between outer and inner bounds is within 0.2 bits for no cooperation and within 0.3 bits for C_{12}=C_{21}=0.5 bits per channel use (bpcu). The difference between different outer bounds is within 0.01 bits in both cases.

In Figure 5.7 we compare the rate regions for asymmetric relay-destination links (γ_{r1} = 20dB, and γ_{r2} = 0dB) with and without source conferencing. All the achievable schemes benefit from source conferencing, especially when message exchange is used, which brings in coherent combining gain. As the relay-destination links are asymmetric (γ_{r1} = 20dB, and γ_{r2} = 0dB), all the compression based schemes (CF, NNC, and SNNC) are constraint by the weakest relay-destination link (i.e. γ_{r2}) due to usage of a single compression index by the relay, despite the fact that NNC and SNNC are more capable than CF in handing the asymmetric relay-destination links and therefore suffer less compared to CF, as explained in Remark 5.2. The simple AF based scheme, on the other hand, has no such problem with the asymmetric channel setups. Therefore, with $C_{12} = C_{21} = 0.5$ or even no backhaul, the simple AF based scheme can beat SNNC (hence CF and NNC) in some region. The gap between inner and outer bounds is within 0.3 bits when the relay-destination links are weak, but decreases to within 0.1 bits for strong relay-destination links. The difference between different outer bounds is negligible.

5.5 Summary

In this chapter we have studied various cooperative strategies when compression or amplification is carried by the relay node. With partial source cooperation, we have investigated the achievable rate regions of NNC with compression forwarding and with message exchange via the backhaul, and demonstrated that encoding delay and memory constraints can affect the achievable rate of regular NNC, and employing instead short-message NNC can provide significant gains and therefore can achieve a strictly larger rate region than NNC. We also investigated the achievable rate region by AF+ANC, and shown that it can even outperform NNC when the coherent combining gain is dominant and outperform CF/SNNC when the asymmetric channel constraint is significant.

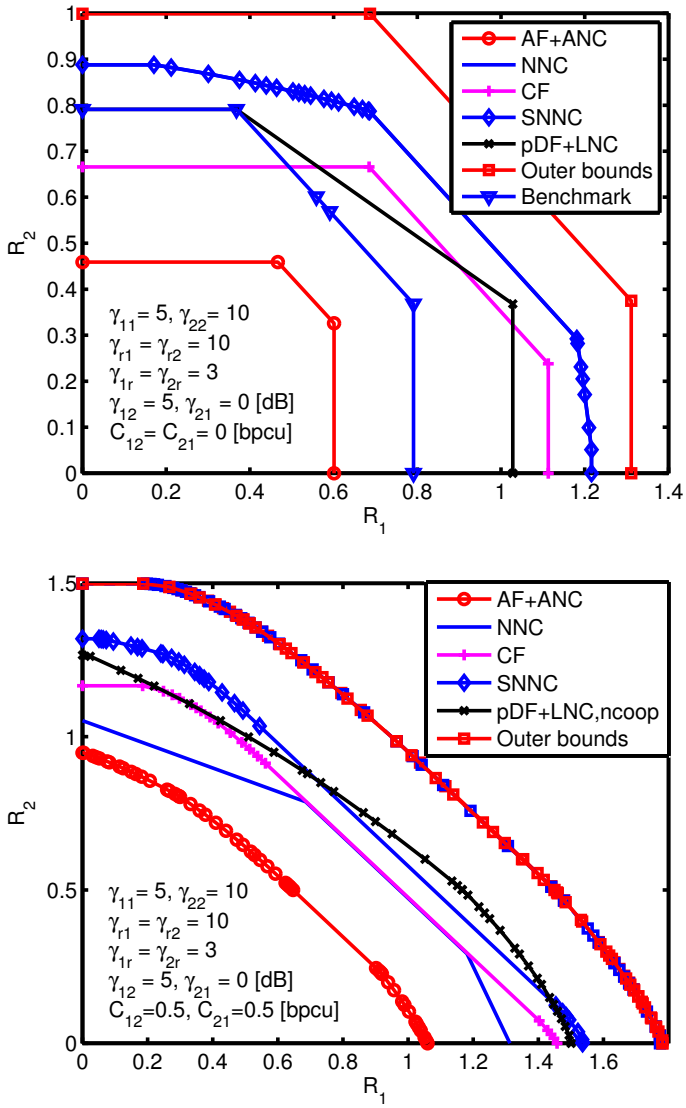


Figure 5.6. Achievable rate region of AF+ANC, CF with message exchange, NNC with compression forwarding, SNNC with message exchange, and pDF+LNC, as well as the capacity outer bounds, for channels setups with direct links $\gamma_{11} = 5$dB and $\gamma_{22} = 10$dB, cross-links $\gamma_{12} = 5$dB and $\gamma_{21} = 0$dB, source-relay links $\gamma_{1r} = \gamma_{2r} = 3$dB and relay-destination links $\gamma_{r1} = \gamma_{r2} = 10$dB, without backhaul (upper) and with symmetric conferencing rates $C_{12} = C_{21} = 0.5$ bits per channel use (lower). The benchmark refers to [GSG+10, Proposition 4] with $R_3 = 0$.

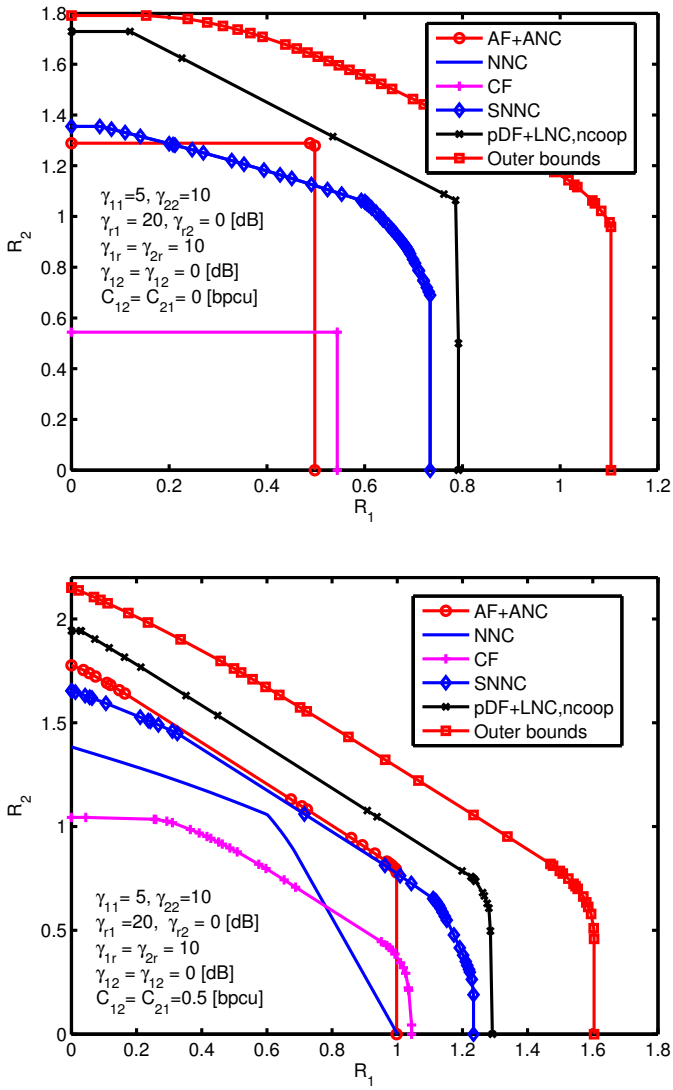


Figure 5.7. Achievable rate region of AF+ANC, NNC with compression forwarding, SNNC with message exchange, and pDF+LNC, as well as the capacity outer bounds, for channels setups with direct links $\gamma_{11} = 5$dB and $\gamma_{22} = 10$dB, cross-links $\gamma_{12} = \gamma_{21} = 0$dB, source-relay links $\gamma_{1r} = \gamma_{2r} = 10$dB and weak/strong relay-destination links $\gamma_{r1} = 20$dB and $\gamma_{r2} = 0$dB, without backhaul (upper) and with symmetric conferencing rates $C_{12} = C_{21} = 0.5$ bits per channel use (lower).

5.6 Appendix

5.6.A Rate Optimization for NNC with Compression Forwarding

We can rewrite (5.4) as

$$\begin{aligned} R_1 &< \Delta_{R_1} + \tilde{R}_1, \\ R_2 &< \Delta_{R_2} + \tilde{R}_2, \\ R_1 + R_2 &< \Delta_{R_s} + \tilde{R}_s, \end{aligned} \tag{5.16}$$

where $(\tilde{R}_1, \tilde{R}_2, \tilde{R}_s)$ denotes the achievable rates when the conferencing links are not used.

To maximize R_1, we should set $\sigma_1^2 = \infty$, i.e., $\mathcal{S}_1$ will not relay the message transmitted by $\mathcal{S}_2$, since this can only leads to decreased value for both Δ_{R_1} and Δ_{R_s}. Therefore (5.16) can be translated to

$$\begin{aligned} R_1 &< \tilde{R}_1 + \mathcal{C}\left(\frac{P_1}{1+\sigma_2^2}\right), \\ R_2 &< \tilde{R}_2 - \mathcal{C}\left(\frac{1}{\sigma_2^2}\right), \\ R_1 + R_2 &< \tilde{R}_s - \mathcal{C}\left(\frac{1}{\sigma_2^2}\right). \end{aligned} \tag{5.17}$$

To maximize $\tilde{R}_1 + \mathcal{C}(\frac{P_1}{1+\sigma_2^2})$ while keeping $\tilde{R}_1 + \mathcal{C}(\frac{P_1}{1+\sigma_2^2}) \le \tilde{R}_s - \mathcal{C}(\frac{1}{\sigma_2^2})$ and $\tilde{R}_2 - \mathcal{C}(\frac{1}{\sigma_2^2}) \ge 0$, we can first find $\sigma_2^* > 0$ such that

$$\tilde{R}_1 + \mathcal{C}\left(\frac{P_1}{1+\sigma_2^{*2}}\right) = \tilde{R}_s - \mathcal{C}\left(\frac{1}{\sigma_2^{*2}}\right),$$

which leads to

$$\sigma_2^* = \sqrt{\frac{1+P_1}{4^{(\tilde{R}_s - \tilde{R}_1)} - 1}}.$$

If $\tilde{R}_2 - \mathcal{C}(\frac{1}{\sigma_2^{*2}}) < 0$, which violates the precondition of a non-negative rate R_2, we should set $\sigma_2^* = \sqrt{\frac{1}{4^{\tilde{R}_2}-1}}$ instead. We can maximize R_2 in the same way.

If $\tilde{R}_1 + \tilde{R}_2 \ge \tilde{R}_s$ or $0 \le P_1, P_2 \le 1$, the sum rate constraint cannot be improved by optimizing over (σ_1, σ_2). In the following we only focus on the scenario when $d_R \triangleq \tilde{R}_s - (\tilde{R}_1 + \tilde{R}_2) > 0$.

When $C_{12} + C_{21} \le d_R$, for all $0 < \sigma_1, \sigma_2 < \infty$ we will have

$$\Delta_{R_1} + \Delta_{R_2} + \tilde{R}_1 + \tilde{R}_2 < \Delta_{R_s} + \tilde{R}_s.$$

Therefore we only need to maximize $\Delta_{R_1}+\Delta_{R_2}+\tilde{R}_1+\tilde{R}_2$ over all possible values of σ_1,σ_2. Since

$$\Delta_{R_1}+\Delta_{R_2}=\frac{1}{2}\log\left(\frac{\sigma_1^2(1+P_2+\sigma_1^2)}{(1+\sigma_1^2)^2}\frac{\sigma_2^2(1+P_1+\sigma_2^2)}{(1+\sigma_2^2)^2}\right),$$

it is straightforward to figure out that

$$\begin{aligned}\sigma_1^{*2}&=\begin{cases}\frac{P_2+1}{P_2-1}, & \text{when } P_2>1;\\ \infty, & \text{otherwise;}\end{cases}\\ \sigma_2^{*2}&=\begin{cases}\frac{P_1+1}{P_1-1}, & \text{when } P_1>1;\\ \infty, & \text{otherwise.}\end{cases}\end{aligned}\tag{5.18}$$

When $C_{12}+C_{21}>d_R$, there exist some $\sigma_1,\sigma_2>0$ such that

$$\Delta_{R_1}+\Delta_{R_2}+\tilde{R}_1+\tilde{R}_2=\Delta_{R_s}+\tilde{R}_s.$$

Therefore we can optimize the sum rate by solving the following optimization problem

$$\begin{aligned}&\max_{\sigma_1>0,\sigma_2>0}\ \tilde{R}_s-\mathcal{C}\left(\frac{1}{\sigma_1^2}\right)-\mathcal{C}\left(\frac{1}{\sigma_2^2}\right)\\ &\text{subject to }\frac{1}{2}\log\left(\left(1+\frac{P_1}{1+\sigma_2^2}\right)\left(1+\frac{P_2}{1+\sigma_1^2}\right)\right)=\tilde{R}_s-(\tilde{R}_1+\tilde{R}_2)=d_R,\end{aligned}$$

which is equivalent to

$$\begin{aligned}&\min_{\sigma_1>0,\sigma_2>0}\ \left(1+\frac{1}{\sigma_1^2}\right)\left(1+\frac{1}{\sigma_2^2}\right)\\ &\text{subject to }\left(1+\frac{P_1}{1+\sigma_2^2}\right)\left(1+\frac{P_2}{1+\sigma_1^2}\right)=4^{d_R}.\end{aligned}$$

When $P_1>1$ and $P_2>1$, by denoting

$$\mu=1+\frac{P_1}{1+\sigma_2^2},\ \omega=1+\frac{P_2}{1+\sigma_1^2},$$

we have

$$\sigma_1^2=\frac{1+P_2-\omega}{\omega-1},\ \sigma_2^2=\frac{1+P_1-\mu}{\mu-1}.$$

Therefore we can rewrite the minimization problem as

$$\begin{aligned}\min\ \left(1+\frac{1}{\sigma_1^2}\right)\left(1+\frac{1}{\sigma_2^2}\right)&=\max\ (1+P_2-\omega)(1+P_1-\mu)\\ &=\min\ \omega(1+P_1)+\mu(1+P_2),\end{aligned}$$

where

$$\omega(1+P_1)+\mu(1+P_2)\geq 2\sqrt{\omega\mu(1+P_1)(1+P_2)},$$

with equality achieved when $\omega(1+P_1)=\mu(1+P_2)$.

On the other hand, as $\omega\mu=4^{d_R}$, we can easily figure out that

$$\omega=\sqrt{\frac{1+P_2}{1+P_1}}2^{d_R},\ \mu=\sqrt{\frac{1+P_1}{1+P_2}}2^{d_R}.$$

When $P_1>1$ but $P_2\leq 1$, forwarding the message from $\mathcal{S}_2$ can only reduce $\Delta_{R_1}+\Delta_{R_2}$. Therefore it is optimal to set $\sigma_1=\infty$. The above optimization problem becomes trivial and we can easily get

$$\sigma_2^2=\frac{P_1}{4^{d_R}-1}-1.$$

Similarly, when $P_2>1$ but $P_1\leq 1$, we have $\sigma_2=\infty$ and

$$\sigma_1^2=\frac{P_2}{4^{d_R}-1}-1.$$

5.6.B Proof of the Achievable Rate Region of AF+ANC

Since $X_{1,t}-V_{c,t}-X_{2,t}$ form a Markov chain, by *sliding window* joint decoding based on $(Y_{1,t}^{(n)},Y_{1,t+1}^{(n)})$ at $\mathcal{D}_1$ and $(Y_{2,t}^{(n)},Y_{2,t+1}^{(n)})$ at $\mathcal{D}_2$, respectively, we can obtain the following results

$$\begin{aligned}\sum_{k\in S}R_k<\min\{&I(V_t(S);Y_{1,t}Y_{1,t+1}|V_{t-1}(T)V_t(S^c)),\\ &I(V_t(S);Y_{2,t}Y_{2,t+1}|V_{t-1}(T)V_t(S^c))\}.\end{aligned}\tag{5.19}$$

Note that

$$\begin{aligned}&R_{1p}+R_{2p}+R_c<I(V_{1p,t}V_{2p,t}V_{c,t};Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1})\\
=&h(Y_{1,t},Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1})-h(\tilde{Z}_t)-h(\sqrt{\gamma_{11}}X_{1,t+1}+\sqrt{\gamma_{21}}X_{2,t+1}+\tilde{Z}_{t+1})\\
\leq&\frac{1}{2}\log(|\boldsymbol{K}_y|)-\frac{1}{2}\log(\sigma_1^2)-\frac{1}{2}\log(\sigma_1^2+\mathrm{Var}(\sqrt{\gamma_{11}}X_{1,t+1}+\sqrt{\gamma_{21}}X_{2,t+1}))\qquad(5.20)\\
=&\mathcal{C}\left(\frac{\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}{1+\beta^2\gamma_{r1}}+\frac{\beta^2\gamma_{r1}(\gamma_{1r}+\gamma_{2r}+2\sqrt{\gamma_{1r}\gamma_{2r}\alpha_1\alpha_2})}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right.\\
&\left.+\frac{\beta^2\gamma_{r1}(1-\alpha_1\alpha_2)(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2}{(1+\beta^2\gamma_{r1})(1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2})}\right),\qquad(5.21)\end{aligned}$$

where $\tilde{Z}_{1,t},\ t=1,...,B+1$ are i.i.d. Gaussian with zero-mean and variance $\sigma_1^2=1+\beta^2\gamma_{r1}$. The inequality in (5.20) comes from the *Maximum Entropy Lemma* and the *Entropy Power Inequality* [CT06], with equality achieved by the joint Gaussian distribution. $\boldsymbol{K}_y$ is the conditional covariance matrix of $(Y_{1,t},Y_{1,t+1})$ given

$(V_{1p,t-1}V_{2p,t-1}V_{c,t-1})$, and the equality in (5.21) is obtained by applying (5.12) into (3.1). Following a similar procedure, it is easy to show that the other mutual information terms in (5.19) are simultaneously maximized by Gaussian distributions. Therefore it is straightforward to translate the remaining constraints to corresponding $\mathcal{C}(\cdot)$ expressions, which are listed as follows.

$$
\begin{aligned}
R_{1p}+R_{2p}+R_c &< I(V_{1p,t}V_{2p,t}V_{c,t};Y_{2,t}Y_{2,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1})\\
&= \mathcal{C}\left(\frac{\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2}}{1+\beta^2\gamma_{r2}}+\frac{\beta^2\gamma_{r2}(\gamma_{1r}+\gamma_{2r}+2\sqrt{\gamma_{1r}\gamma_{2r}\alpha_1\alpha_2})}{1+\beta^2\gamma_{r2}+\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2}}\right.\\
&\qquad \left.+\frac{\beta^2\gamma_{r2}(1-\alpha_1\alpha_2)(\sqrt{\gamma_{12}\gamma_{2r}}-\sqrt{\gamma_{22}\gamma_{1r}})^2}{(1+\beta^2\gamma_{r2})(1+\beta^2\gamma_{r2}+\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2})}\right).\\
R_c &< I(V_{c,t};Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{1p,t}V_{2p,t})\\
&= \mathcal{C}\left(\frac{(\sqrt{\gamma_{11}\alpha_1}+\sqrt{\gamma_{21}\alpha_2})^2}{1+\beta^2\gamma_{r1}}+\frac{\beta^2\gamma_{r1}(\sqrt{\gamma_{1r}\alpha_1}+\sqrt{\gamma_{2r}\alpha_2})^2}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right),\\
R_c &< I(V_{c,t};Y_{2,t}Y_{2,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{1p,t}V_{2p,t})\\
&= \mathcal{C}\left(\frac{(\sqrt{\gamma_{12}\alpha_1}+\sqrt{\gamma_{22}\alpha_2})^2}{1+\beta^2\gamma_{r2}}+\frac{\beta^2\gamma_{r2}(\sqrt{\gamma_{1r}\alpha_1}+\sqrt{\gamma_{2r}\alpha_2})^2}{1+\beta^2\gamma_{r2}+\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2}}\right),\\
R_{1p} &< I(V_{1p,t};Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{2p,t}V_{c,t})\\
&= \mathcal{C}\left(\frac{\gamma_{11}\bar{\alpha}_1}{1+\beta^2\gamma_{r1}}+\frac{\beta^2\gamma_{r1}\gamma_{1r}\bar{\alpha}_1}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right),\\
R_{1p} &< I(V_{1p,t};Y_{2,t}Y_{2,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{2p,t}V_{c,t})\\
&= \mathcal{C}\left(\frac{\gamma_{12}\bar{\alpha}_1}{1+\beta^2\gamma_{r2}}+\frac{\beta^2\gamma_{r2}\gamma_{1r}\bar{\alpha}_1}{1+\beta^2\gamma_{r2}+\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2}}\right),\\
R_{2p} &< I(V_{2p,t};Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{1p,t}V_{c,t})\\
&= \mathcal{C}\left(\frac{\gamma_{21}\bar{\alpha}_2}{1+\beta^2\gamma_{r1}}+\frac{\beta^2\gamma_{r1}\gamma_{2r}\bar{\alpha}_2}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right),\\
R_{2p} &< I(V_{2p,t};Y_{2,t}Y_{2,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{1p,t}V_{c,t})\\
&= \mathcal{C}\left(\frac{\gamma_{22}\bar{\alpha}_2}{1+\beta^2\gamma_{r2}}+\frac{\beta^2\gamma_{r2}\gamma_{2r}\bar{\alpha}_2}{1+\beta^2\gamma_{r2}+\gamma_{12}+\gamma_{22}+2\sqrt{\gamma_{12}\gamma_{22}\alpha_1\alpha_2}}\right).\\
R_{1p}+R_{2p} &< I(V_{1p,t}V_{2p,t};Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{c,t})\\
&= \mathcal{C}\left(\frac{\bar{\alpha}_1\gamma_{11}+\bar{\alpha}_2\gamma_{21}}{1+\beta^2\gamma_{r1}}+\frac{\beta^2\gamma_{r1}(\bar{\alpha}_1\gamma_{1r}+\bar{\alpha}_2\gamma_{2r})}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right.\\
&\qquad \left.+\frac{\beta^2\gamma_{r1}\bar{\alpha}_1\bar{\alpha}_2(\sqrt{\gamma_{11}\gamma_{2r}}-\sqrt{\gamma_{21}\gamma_{1r}})^2}{(1+\beta^2\gamma_{r1})(1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2})}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
R_{1p} + R_c &< I(V_{1p,t}V_{c,t}; Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{2p,t}) \\
&= \mathcal{C}\left(\frac{\gamma_{11} + \alpha_2\gamma_{21} + 2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}{1+\beta^2\gamma_{r1}} + \frac{\beta^2\gamma_{r1}(\gamma_{1r} + \alpha_2\gamma_{2r} + 2\sqrt{\gamma_{1r}\gamma_{2r}\alpha_1\alpha_2})}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right. \\
&\qquad \left. + \frac{\beta^2\gamma_{r1}\bar{\alpha}_1\alpha_2(\sqrt{\gamma_{11}\gamma_{2r}} - \sqrt{\gamma_{21}\gamma_{1r}})^2}{(1+\beta^2\gamma_{r1})(1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2})}\right), \\
R_{2p} + R_c &< I(V_{2p,t}V_{c,t}; Y_{1,t}Y_{1,t+1}|V_{1p,t-1}V_{2p,t-1}V_{c,t-1}V_{1p,t}) = \\
&= \mathcal{C}\left(\frac{\alpha_1\gamma_{11} + \gamma_{21} + 2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}{1+\beta^2\gamma_{r1}} + \frac{\beta^2\gamma_{r1}(\alpha_1\gamma_{1r} + \gamma_{2r} + 2\sqrt{\gamma_{1r}\gamma_{2r}\alpha_1\alpha_2})}{1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2}}\right. \\
&\qquad \left. + \frac{\beta^2\gamma_{r1}\alpha_1\bar{\alpha}_2(\sqrt{\gamma_{11}\gamma_{2r}} - \sqrt{\gamma_{21}\gamma_{1r}})^2}{(1+\beta^2\gamma_{r1})(1+\beta^2\gamma_{r1}+\gamma_{11}+\gamma_{21}+2\sqrt{\gamma_{11}\gamma_{21}\alpha_1\alpha_2})}\right).
\end{aligned}
$$

Chapter 6

General Bounding Models for Networks with Independent Noise

In this chapter we present general capacity bounding models for wireless networks with independent noise, by construction of upper and lower capacity bounding networks consisting of only noiseless bit-pipe channels. The work presented in this chapter is motivated by the elegant framework of network equivalence theory [KEM11, KEM10] and the one-shot bounding method [CME11].

We will first give a brief introduction of the network equivalence theory and the one-shot method, and then present our improvement on the bounding models. For non-coupled networks[1] with independent noise, we apply the bounding models directly and construct the corresponding upper and lower bounds. For coupled networks, we propose a network decoupling method to obtain both upper and lower bounding networks.

6.1 Introduction

In the seminal work [KEM11], an elegant theory of network equivalence has been established and the equivalence of an independent point-to-point noisy channel and a noiseless bit-pipe under any arbitrary networks has been proved as long as the throughput of the latter equals to the capacity of the former. For independent multi-terminal networks, such as multiple-access channels, broadcast channels, and interference channels with two sources and/or two destinations, both upper and lower bounding models have been proposed in [KEM10, KEM09]. A one-shot bounding method proposed in [CME11] gives simple yet tight bounding models for general MAC and for BC with two receivers. Note that in [KRV11] a layering approach with a global information flow routing technique has been proposed to model non-coupled wireless networks with multiple unicast transmission, and this approach is proven to be asymptotically optimal in the sense that the capacity lower

[1] The definition of coupled and non-coupled networks will be introduced in Section 6.1.1.

bounds are within a multiplicative gap from the cut-set bound. We will not follow the approach developed in [KRV11], since we are interested in developing a general method that can construct bounding networks for both coupled and non-coupled noisy networks, rather than developing a scheme to find the capacity of a noiseless network, which itself is a very difficult problem [LL03, CG08] unsolved in general. In this chapter, we will provide efficient and simple methods that can construct both upper and lower bounding networks consisting of only noiseless bit-pipes, which can serve as the basis to bound the capacity of general wireless networks.

6.1.1 Basic Definitions

Before we proceed, we will first present some definitions that are frequently used in this chapter.

Definition 6.1. ***Independent Channel***
A point-to-point channel $\mathfrak{N}_i = (\mathcal{X}_i, p(y_i|x_i), \mathcal{Y}_i)$ *with input alphabet* $\mathcal{X}_i$*, output alphabet* $\mathcal{Y}_i$*, and the conditional transition probability* $p(y_i|x_i)$ *within a network* $\mathfrak{N}_T = (\prod_{n\in T}\mathcal{X}_n, p(\boldsymbol{y}|\boldsymbol{x}), \prod_{n\in T}\mathcal{Y}_n)$ *is said to be independent if the network transition probability* $p(\boldsymbol{y}|\boldsymbol{x})$ *can be partitioned as*

$$p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y}_{/i}|\boldsymbol{x}_{/i})p(y_i|x_i),$$

where $\boldsymbol{x}_{/i}$ *denotes the vector of* $\boldsymbol{x}$ *without element* x_i*, and similarly for* $\boldsymbol{y}_{/i}$*. Therefore we can rewrite the expression of network* $\mathfrak{N}_T$ *as*

$$\mathfrak{N}_T = \mathfrak{N}_i \times (\prod_{n\neq i}\mathcal{X}_n, p(\boldsymbol{y}_{/i}|\boldsymbol{x}_{/i}), \prod_{n\neq i}\mathcal{Y}_n) \triangleq \mathfrak{N}_i \times \mathfrak{N}_{T/i}.$$

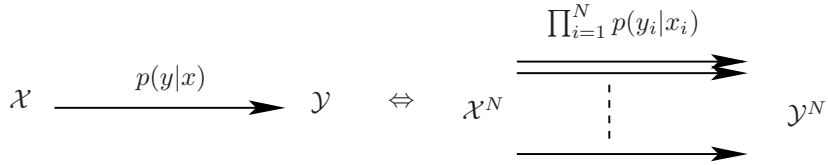
Similarly, a multi-terminal channel/network $\mathfrak{N}_S$ within $\mathfrak{N}_T$ is said to be independent, denoted by $\mathfrak{N}_T = \mathfrak{N}_S \times \mathfrak{N}_{S^c}$, if the network transition probability can be partitioned as $p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y}_S|\boldsymbol{x}_S)p(\boldsymbol{y}_{S^c}|\boldsymbol{x}_{S^c})$, where $S, S^c \subset T$ are a pair of complementary subsets of T.

Definition 6.2. ***Capacity Bounding Models***
Given two channels/networks $\mathfrak{C}$ *and* $\mathfrak{N}$*,* $\mathfrak{C}$ *is said to upper bound* $\mathfrak{N}$*, or equivalently* $\mathfrak{N}$ *lower bounds* $\mathfrak{C}$*, if the capacity (region) of* $\mathfrak{N}\times\mathfrak{W}$ *is a subset of that for* $\mathfrak{C}\times\mathfrak{W}$ *for any network* $\mathfrak{W}$*. We denote their relationship by* $\mathfrak{N}\subseteq\mathfrak{C}$*.* $\mathfrak{C}$ *and* $\mathfrak{N}$ *is said to be equivalent if* $\mathfrak{C}\subseteq\mathfrak{N}\subseteq\mathfrak{C}$*.*

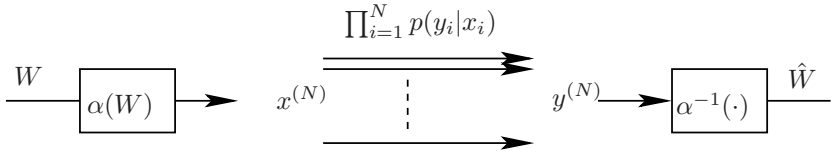
For channels/networks $\mathfrak{C}_u$ *and* $\mathfrak{C}_l$ *consisting of only noiseless bit-pipes, we say that* $\mathfrak{C}_u$ *is the upper bounding model and* $\mathfrak{C}_l$ *is the lower bounding model for* $\mathfrak{N}$ *if*

$$\mathfrak{C}_l \subseteq \mathfrak{N} \subseteq \mathfrak{C}_u.$$

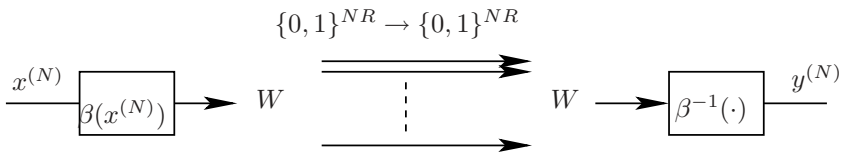
Definition 6.3. ***Coupled/Non-coupled Network***
A network is said to be coupled if any of its point-to-point connections is part of a multiple-access channel and a broadcast channel simultaneously. Otherwise the network is non-coupled.



(a) A noisy channel $\mathfrak{N}$ with capacity C and its stacked network with N replicas



(b) The channel coding argument to prove $\mathfrak{C} \subseteq \mathfrak{N}$ for all $R < C$



(c) The lossy source coding argument to prove $\mathfrak{N} \subseteq \mathfrak{C}$ for all $R > C$

Figure 6.1. A point-to-point noisy channel $\mathfrak{N} = (\mathcal{X}, p(y|x), \mathcal{Y})$ with capacity C and a noiseless bit-pipe $\mathfrak{C}$ of rate R are equivalent if $R = C$. The input/output of their corresponding stack networks are $x^{(N)} \in \mathcal{X}^N$, $y^{(N)} \in \mathcal{Y}^N$, and $W, \hat{W} \in \{1, \ldots, 2^{NR}\}$.

As expected, wireless networks are in general coupled due to the broadcast nature of microwave propagation.

6.1.2 Network Equivalence Theory for Independent Channels

In [KEM11], the equivalence between an independent point-to-point noisy channel $\mathfrak{N}$ with capacity C and a noiseless bit-pipe $\mathfrak{C}$ of the same capacity has been established by showing that any code that runs on a network $\mathfrak{N} \times \mathfrak{W}$ can also be run on $\mathfrak{C} \times \mathfrak{W}$ with asymptotically vanishing error probability. Their argument is based on a stacked network emulation approach where a large stack of N parallel replicas of the network have been put together to run the code, as illustrated in Figure 6.1.

The proof of $\mathfrak{C} \subseteq \mathfrak{N}$ employs a channel coding argument over the stack of N channel replicas as illustrated in Figure 6.1(b): A message W of 2^{NR} bits is mapped by the channel encoder $\alpha(\cdot)$ onto a codeword $x^{(N)}$ of length N, and then transmitted over the N-stack noisy channels, with one symbol on each replica, such that reliable transmission over the noisy stacked network can be realized with arbitrary small error probability as N goes to infinity for all $R < C$.

The proof of $\mathfrak{N} \subseteq \mathfrak{C}$ is based on a lossy source coding argument as illustrated in Figure 6.1(c): The input sequence $x^{(N)}$ to the noisy stacked network is first quan-

tized/compressed by a lossy source encoder $\beta(\cdot)$ into 2^{NR} bits, represented by the message W which is then transmitted through the noiseless stacked network, and the reconstructed sequence $y^{(N)}$ is selected in such a way that it looks jointly typical with the transmitted sequence $x^{(N)}$, as contrast to the usual distortion measure. The existence of a good lossy source coding codebook for any $R > C$ is proved by random coding argument, i.e., by showing that the average error probability over the randomly chosen ensemble of codebooks is small. The equivalence between $\mathfrak{N}$ of capacity C and $\mathfrak{C}$ of throughput R can then be established when $R = C$ based on the continuity of the capacity region. Readers are kindly referred to [KEM11] for a rigorous and thorough treatment of the proof.

The concept of capacity upper and lower bounding models developed in [KEM11] has been extended to independent multi-terminal channels in [KEM10, KEM09] following similar arguments as illustrated in Figure 6.1, and multiplicative and additive gaps between lower and upper bounding models for independent multi-terminal channels have been established. Illustrative upper and lower bounding models for multiple-access, broadcast, and interference channels which involve either two sources and/or two destinations have been demonstrated. Given a noisy network consisting of independent building blocks whose upper and lower bounding models are available, we can replace these building blocks with their corresponding upper (lower) bounding models and then characterize an outer (inner) bound for its capacity region based on the resulting noiseless network models.

The lossy source coding argument for upper bounding models has to be exploited for each network building block and the channel coding argument for lower bounding models has to be specified for each of the established rate-achievable coding schemes. For coupled networks involving many transmitting nodes and receiving nodes, applying bounding models for small size building blocks will result in a very large gap, while finding bounding models for large building blocks might be very complex as we have demonstrated in Chapter 3 for characterizing upper bounds for the small multicast relay networks.

6.1.3 Equivalent One-shot Bounding Models

Instead of exploiting channel/source code emulation to construct bounding models as in [KEM11, KEM10, KEM09], a class of one-shot bounding tools have been proposed in [CME11] for independent multiple-access and broadcast channels. As illustrated in Figure 6.2, auxiliary operation nodes are introduced to specify the rate constraints on the sum rate and on individual rates, and the channel emulation is realized in each time instance and hence called "one-shot" approach, as contrast to [KEM11, KEM10, KEM09] where channel emulation is realized in each block of channel uses. While the lower bounding models can be constructed based on achievable rate regions for multiple-access channels with independent source nodes and for broadcast channels with non-cooperating destination nodes, the upper bounding models require special treatment, which will be outlined below.

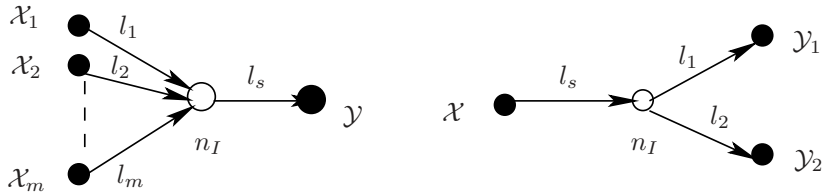


Figure 6.2. The one-shot bounding models for multiple-access channels with m transmitters and for broadcast channels with two receivers. The white nodes indicated by n_I are auxiliary operation nodes to specify the rate constraints on the sum rate and on individual rates. All the channels are noiseless bit-pipes and independent from others. The one-shot bounding models are fully characterized by the rate vector $(R_{l_s}, R_{l_1}, \ldots, R_{l_m})$, where R_{l_i} is the rate of the noiseless bit-pipe l_i.

Upper Bounding Models for Multiple-Access Channels

For multiple-access channels with m transmitters, each with transmit alphabet $\mathcal{X}_i$, $i = 1, \ldots, m$, we can introduce an auxiliary operation node n_I to emulate the MAC channel by a small network with m independent and orthogonal input channels to node n_I, denoted as l_i with $i = 1, \ldots, m$, to emulate the individual rates, and one channel l_s between n_I and the destination node to emulate the sum rate. We then replace all the independent channels with noiseless bit-pipes and define a rate vector $(R_{l_s}, R_{l_1}, \ldots, R_{l_m})$ to describe the rates of channels $(l_s, l_1, \ldots, l_m)$. As the one-shot bounding model is fully characterized by this rate vector, we will utilize it to represent the corresponding bounding model $\mathfrak{C}$.

There are two ways to emulate the MAC channel. One can let each input channel l_i carry exactly what the corresponding source node transmits, hence requiring a noiseless bit-pipe of rate $R_{l_i} \geq \log(|\mathcal{X}_i|)$, where $|\mathcal{X}|$ is the cardinality of the alphabet set $\mathcal{X}$. Note that for continuous-valued random variables we will have $|\mathcal{X}| = \infty$. The auxiliary nodes then combine all the inputs in such a way that the output signal at the destination node is exactly the same as in the original network. By the point-to-point network equivalence theory, successful emulation of the original channel requires

$$R_{l_s} \geq R_{MAC} \triangleq \max_{p(x_1,\ldots,x_m)} I(X_1, \ldots, X_m; Y).$$

Hence we can construct the upper bounding model as

$$\mathfrak{C}_{u,MAC,1} = (R_{MAC}, \log(|\mathcal{X}_1|), \ldots, \log(|\mathcal{X}_m|)).$$

Although $\mathfrak{C}_{u,MAC,1}$ is sharp on sum rate in the sense that there are some kind of networks where the sum rate constraint R_{MAC} is the capacity, the constraints on individual rates are somehow loose.

Denoting z_i as the channel output of l_i, and letting the auxiliary node emulate the output at the destination node exactly via a predefined function $g(\cdot)$, i.e.,

$$y = g(z_1, z_2, \ldots, z_m),$$

the transition probability $p(y|x_1,\ldots,x_m)$ and all its marginal distributions can then be exactly emulated as follows

$$p(y|x_1,\ldots,x_m) = \sum_{\substack{z_1,\ldots,z_m:\\ y=g(z_1,\ldots,z_m)}} \prod_{i=1}^{m} p(z_i|x_i).$$

We can therefore formulate an alternative upper bounding model as

$$\mathfrak{C}_{u,MAC,2} = (\log(|\mathcal{Y}|), R_1, \ldots, R_m),$$

where for $i = 1,\ldots,m$,

$$R_i \triangleq \max_{p(x_i)} I(X_i; Z_i).$$

One way to construct the upper bounding model $\mathfrak{C}_{u,MAC,2}$ for multiple-access channels is based on a noise partitioning approach, i.e., the additive noise at the destination is partitioned into independent parts and allocated to the parallel channels. For Gaussian multiple-access channel with two transmitters, as demonstrated in [CME11], the corresponding upper bounding model is

$$\mathfrak{C}_{u,MAC,2} = \left(\infty,\ \frac{1}{2}\log(1+\frac{\mathrm{SNR}_1}{\alpha}),\ \frac{1}{2}\log(1+\frac{\mathrm{SNR}_2}{1-\alpha})\right),$$

where SNR_i is the corresponding SNR at the receiver when only transmitter i is active, and $\alpha \in (0,1)$ is the noise power partitioning parameter chosen to minimize the total sum rate. The sum rate constraint is ∞ since the alphabet for Gaussian variable is continuous and with infinite cardinality. When calculating capacity upper bounds, this infinite upper constraint can be replaced by

$$R_{MAC} = \frac{1}{2}\log(1+\frac{\mathrm{SNR}_1}{\alpha}) + \frac{1}{2}\log(1+\frac{\mathrm{SNR}_2}{1-\alpha}),$$

i.e., the sum of the total input rates to the auxiliary node. For binary symmetric MAC channels with distortion parameter ϵ, the corresponding distortion ϵ_i for channel l_i should satisfy

$$\epsilon = \epsilon_1(1-\epsilon_2) + \epsilon_2(1-\epsilon_1).$$

Upper Bounding Models for Broadcast Channels with Two Receivers

Upper bounding models for broadcast channels with two receivers have also been constructed in [CME11] for the scenario when the noise at two receivers are independent, i.e., the transition probability can be factorized as

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|x).$$

Similar to the upper bounding models for multiple-access channels, there are also two different bounding models for broadcast channels, i.e.,

$$\mathfrak{C}_{u,BC,1} = (R_{BC}, \log(|\mathcal{Y}_1|), \log(|\mathcal{Y}_2|)),$$

where

$$R_{BC} \triangleq \max_{p(x)} I(X; Y_1, Y_2),$$

and

$$\mathfrak{C}_{u,BC,2} = (\log|\mathcal{X}|, R_1, R_2),$$

where for $i = 1, 2$,

$$R_i \triangleq \max_{p(x)} I(X; Y_i).$$

Gap between the One-shot Bounding Models

The gap between the upper and lower bounding models for Gaussian channels and for binary symmetric channels have been examined in [CME11] where a gap less than $\frac{1}{2}$ bit per channel use has been established for two-user setups, i.e., multiple-access channels with two transmitters, and broadcast channels with two receivers.

6.2 Bounding Models for Non-coupled Networks

For non-coupled networks, we generalize the noise partitioning approach developed in [CME11] for multiple-access channels to setups with more than two transmitters, and extend the upper and lower bounding models for broadcast channels to scenarios with more than two receivers.

6.2.1 Noise Partitioning for Multiple-Access Channels with More than Two Transmitters

The noise partitioning problem for a multiple-access channel with m transmitters, each with normalized receiver-side SNR γ_i, $i = 1, ..., m$, can be formulated as follows

$$\begin{aligned} \min_{\alpha_1,...,\alpha_m} \quad & \sum_{i=1}^{m} \log\left(1 + \frac{\gamma_i}{\alpha_i}\right), \\ \text{subject to} \quad & \sum_{i=1}^{m} \alpha_i = 1, \\ & \alpha_i > 0, \end{aligned} \tag{6.1}$$

which is a convex optimization problem and can be solved efficiently using barrier methods or primal-dual interior point methods for convex optimization [BV04].

Alternatively, we can first solve the optimization problem with two transmitters analytically, as presented in (6.3) stated below, then apply the result to sequentially

Input: received signal power γ_1, γ_2, γ_3
Output: partitioned noise power α_1, α_2, α_3
Initialize α_1 with a constant in $(0, 1)$;
Predefine a small number $\epsilon > 0$ for precision check.
foreach *Iteration* i **do**
  Calculate α_2, α_3 by (6.3) given γ_2, γ_3, with $\alpha = 1 - \alpha_1$;
  Calculate z_1, z_2 by (6.3) given γ_1, γ_2, with $\alpha = 1 - \alpha_3$;
  Calculate n_1, n_3 by (6.3) given γ_1, γ_3, with $\alpha = 1 - \alpha_2$;
  if $|\alpha_1 - z_1| + |\alpha_1 - n_1| < \epsilon$ **then**
    break;
  else
    $\alpha_1 = \dfrac{\alpha_1 + z_1}{2}$;
  end
end

Algorithm 1: Noise partitioning algorithm for MAC with three transmitters

redistribute noise power among transmitters in a pairwise manner, and update the noise partitioning (e.g., take average) after each iteration. An example of the noise partitioning algorithm for a MAC with three transmitters is shown in Algorithm 1. In general, the convergence of the algorithm is fast.

Given total noise power α, the noise partitioning problem can be formulated as

$$\begin{array}{ll} \min\limits_{\alpha_1,\alpha_2} & \left(1+\dfrac{\gamma_1}{\alpha_1}\right)\left(1+\dfrac{\gamma_2}{\alpha_2}\right) \\ \text{subject to} & \alpha_1+\alpha_2=\alpha, \\ & \alpha_1>0,\ \alpha_2>0, \end{array} \tag{6.2}$$

whose solution can be figured out straightforwardly by introducing the Lagrange multiplier and then taking partial derivative,

$$\alpha_1 = \alpha\left[1+\sqrt{\frac{1+\gamma_1^{-1}}{1+\gamma_2^{-1}}}\right]^{-1}, \quad \alpha_2 = \alpha - \alpha_1. \tag{6.3}$$

The sum rate of Gaussian multiple-access channels with m transmitters is upper bounded by

$$R_{u,s} = \frac{1}{2}\log\left(1+\left(\sum_{i=1}^{m}\sqrt{\gamma_i}\right)^2\right),$$

which is the sum rate capacity when full cooperation among all transmitters is possible. On the other hand, the following sum rate is achievable by exploiting independent codebooks at transmitters and successive interference cancellation de-

coding at the receiver,

$$R_{l,s} = \frac{1}{2}\log\left(1 + \sum_{i=1}^{m}\gamma_i\right).$$

The gap between the upper and the lower bounds on sum rate, measured in bits per channel use, is therefore

$$\begin{aligned}\Delta_{MAC} = R_{u,s} - R_{l,s} &= \frac{1}{2}\log\left(\frac{1 + (\sum_{i=1}^{m}\sqrt{\gamma_i})^2}{1 + \sum_{i=1}^{m}\gamma_i}\right) \\ &\leq \frac{1}{2}\log\left(\frac{1 + m\sum_{i=1}^{m}\gamma_i}{1 + \sum_{i=1}^{m}\gamma_i}\right) \\ &< \frac{1}{2}\log(m),\end{aligned} \tag{6.4}$$

where the first inequality comes from Jensen's inequality based on the convexity of the function $f(x) = x^2$. Hence, for Gaussian multiple-access channels with transmitters in isolation, feedback and transmitter cooperation can increase the sum capacity by at most $\frac{1}{2}\log(m)$ bits per channel use. The gap becomes considerably smaller at low SNR or when the SNR for each link diverges.

6.2.2 Broadcast Channels with More than Two Receivers

In [CME11] upper and lower bounding models for broadcast channels have been constructed for the setup with two receivers, in contrast to the multiple-access channels where bounding models for m transmitters have been proposed. The main difference between multiple-access channels and broadcast channels is the encoding process when the transmitters/receivers cannot cooperate: distributed encoding is performed in multiple-access channels while centralized encoding is done in broadcast channels. As a consequence, only one rate constraint for each parallel channel l_i together with a sum rate constraint is sufficient in multiple-access setups. The broadcast setup, however, requires 2^m rate constraints to describe each of the individual channel, and any subset of them, including the rate constraint on the total sum rate.

The upper bounding model for broadcast channels with m receivers can be generalized straightforwardly from the structure developed in [CME11] by omitting the rate constraints on some subsets of parallel channels. Note that utilizing fewer rate constraints may result in a looser upper bounding model for general setups. The upper bounding models with m receivers can therefore be written as follows

$$\begin{aligned}\mathfrak{C}_{u,BC,1} =& (R_{BC}, \log(|\mathcal{Y}_1|), \ldots, \log(|\mathcal{Y}_m|)), \\ \mathfrak{C}_{u,BC,2} =& (\log(|\mathcal{X}|), R_1, \ldots, R_m),\end{aligned}$$

where

$$R_{BC} \triangleq \max_{p(x)} I(X; Y_1, \ldots, Y_m),$$
$$R_i \triangleq \max_{p(x)} I(X; Y_i),\ i = 1, \ldots, m.$$

Generalization of the lower bounding models for broadcast channels with m receivers requires new techniques. We first introduce a rate vector of length 2^m,

$$\boldsymbol{R} = [R_i :\ i = 0, 1, \ldots, 2^m - 1],$$

to describe the necessary rate constraints at the transmitter. R_0 is reserved for the constraint on the sum rate, and R_i, $i = 1, \ldots, 2^m - 1$, is the rate constraint to ensure simultaneously successful decoding of the unicast/multicast message by receivers indicated by the index i expressed in binary format of length m. For example, $R_{2^{n-1}}$, $n = 1, \ldots, m$, is the constraint for unicast rate over channel l_n to receiver n, R_5 is the constraint for multicast rate to receiver 1 and receiver 4, and R_{2^m-1} is the constraint for multicast rate to all receivers. For non-degraded channels, the length of the rate vector can be up to 2^m in the worst case scenario. However, for statistically degraded parallel channels l_i, $i = 1, \ldots, m$, which is the case for Gaussian broadcast channels, the length of the rate vector is at most m. We first illustrate this by a toy example and then present the results for general cases.

A Toy Example: Gaussian Broadcast Channel with Two Receivers

Let γ_1 and γ_2 be the effective link SNR at receiver 1 and receiver 2, respectively.

If $\gamma_1 = \gamma_2 = \gamma$, then we have

$$\boldsymbol{R}_{BC} = [R_0, R_3] = [0.5 \log(1 + \gamma), 0.5 \log(1 + \gamma)].$$

That is, only a common message can be decoded by both receivers.

If $\gamma_1 < \gamma_2$, we introduce the power allocation parameter $\beta \in [0, 1]$ for superposition encoding, which results in the following individual rates

$$R_3 = 0.5 \log\left(1 + \frac{(1-\beta)\gamma_1}{1 + \beta\gamma_1}\right),$$
$$R_2 = 0.5 \log(1 + \beta\gamma_2),$$

where the common message (of rate R_3) can be decoded by both receivers and the private message (of rate R_2) only by the more powerful receiver (receiver 2 in this case). Hence the corresponding lower bounding model is

$$\boldsymbol{R}_{BC} = [R_0, R_3, R_2] = [R_3 + R_2, R_3, R_2].$$

General Case: Gaussian Broadcast Channels with m Receivers

Here we look at bounds for the Gaussian broadcast channels with m receivers. Without loss of generality, assuming $\gamma_1 \leq \gamma_2 \leq ... \leq \gamma_m$, we can encode the codeword W_i by superposition coding with corresponding power allocation parameters $\beta_i \in [0, 1]$, and then utilize successive interference cancellation at receivers. An alternative encoding/decoding approach is to encode codewords W_1 to W_m by dirty paper coding [Cos83] successively and a maximum likelihood decoder at each receiver. Successful decoding of W_1 can be realized at all receivers, and the corresponding rate constraint should be stored as R_{2^m-1}. Successful decoding of W_n, $n = 2, \ldots, m$ can be realized at all receivers except receivers 1 to $n-1$, and therefore the corresponding rate constraint should be stored as $R_{2^m-2^{n-1}}$. The resulting rate vector is therefore

$$\boldsymbol{R} = [R_0, R_{2^m-2^{i-1}} : \ i = 1, \ldots, m],$$

where for $i = 1, \ldots, m$,

$$R_{2^m-2^{i-1}} = \frac{1}{2} \log \left(1 + \frac{\beta_i \gamma_i}{1 + \gamma_i \sum_{j=i+1}^{m} \beta_j} \right), \tag{6.5}$$

is the rate constraint at which message W_i can be transmitted successfully via channel l_i to l_m. The constraint R_0 on the sum rate is

$$\begin{aligned} R_0 &= \sum_{i=1}^{m} R_{2^m-2^{i-1}} \\ &= \sum_{i=1}^{m} \frac{1}{2} \log \left(\frac{1 + \gamma_i \sum_{j=i}^{m} \beta_j}{1 + \gamma_i \sum_{j=i+1}^{m} \beta_j} \right) \\ &= \frac{1}{2} \log(1 + \gamma_1) + \frac{1}{2} \sum_{i=2}^{m} \log \left(\frac{1 + \gamma_i \sum_{j=i}^{m} \beta_j}{1 + \gamma_{i-1} \sum_{j=i}^{m} \beta_j} \right), \end{aligned}$$

where the last equality comes from the fact that

$$\sum_{i=1}^{m} \beta_i = 1.$$

Since $\gamma_{i-1} \leq \gamma_i$, the function $\frac{1+x\gamma_i}{1+x\gamma_{i-1}}$ is monotonically increasing on $x \in [0, 1]$, with its maximum $\frac{1+\gamma_i}{1+\gamma_{i-1}}$ achieved when $x = 1$, it is straightforward to figure out that

$$R_0 \leq \frac{1}{2} \log(1 + \gamma_m),$$

where the equality is achieved when $\beta_m = 1$ (i.e, $\beta_i = 0$ for all $i \neq m$).

The sum rate of Gaussian broadcast channels with m receivers is upper bounded by

$$R_s = \frac{1}{2}\log\left(1+\sum_{i=1}^{m}\gamma_i\right),$$

which can be achieved only when full cooperation among receivers is possible. The gap between the upper and the lower bounds on sum rate, measured in bits per channel use, is therefore

$$\Delta_{BC} = R_s - R_0 = \frac{1}{2}\log\left(\frac{1+\sum_{i=1}^{m}\gamma_i}{1+\gamma_m}\right) \le \frac{1}{2}\log\left(\frac{1+m\gamma_m}{1+\gamma_m}\right) < \frac{1}{2}\log(m), \quad (6.6)$$

where the first inequality comes from the assumption $\gamma_i \le \gamma_m$ for all i. Hence, for Gaussian broadcast channels with receivers in isolation, feedback and receiver cooperation can increase the sum capacity by at most $\frac{1}{2}\log(m)$ bits per channel use. The gap becomes considerably smaller at low SNR or when the SNR for each link diverges.

6.2.3 Illustrative Example

Given a wireless network $\mathfrak{N}$ with independent noise, when constructing upper (lower) bounding networks, we first replace each MAC and BC modules with the corresponding one-shot upper (lower) bounding models. As there are more than one bounding model for each building block, there are many different combinations, each resulting in a valid noiseless upper bounding network $\mathfrak{C}_{u,i}$ (lower bounding network $\mathfrak{C}_{l,i}$) on the original network $\mathfrak{N}$. We should take intersection of all the valid capacity upper bounds $\mathfrak{C}_{u,i}$ to get the final (and tighter) upper bound, and this operation is indicated by (with abuse of notation)

$$\mathfrak{N} \subseteq \bigcap_i \mathfrak{C}_{u,i}.$$

For the lower bound, we should compute the achievable rate region for each of the valid lower bounding network $\mathfrak{C}_{l,i}$ and then take union of them to get the final (and tighter) lower bound, i.e.,

$$\bigcup_i \mathfrak{C}_{l,i} \subseteq \mathfrak{N}.$$

We illustrate the above bounding procedure by a non-coupled noisy network $\mathfrak{N}$ shown in Figure 6.3, where the source node $\mathcal{S}_1$ multicasts information W_1 at rate R_1 to destinations $\mathcal{D}_1$ and $\mathcal{D}_3$, and $\mathcal{S}_2$ multicasts W_2 at rate R_2 to destinations $\mathcal{D}_2$ and $\mathcal{D}_3$, with aid from a full-duplex relay $\mathcal{R}$. The point-to-point connections $\mathcal{S}_1$-$\mathcal{D}_1$ and $\mathcal{S}_2$-$\mathcal{D}_2$ are binary symmetric channels with crossover probability ϵ_1 and ϵ_2, respectively. The multiple-access channel ending at $\mathcal{R}$ and the broadcast channel originating from $\mathcal{R}$ are Gaussian channels with effective link SNR γ_{ir} and γ_{rj}, $i = 1, 2$, $j = 1, 2, 3$, respectively. Without loss of generality, we assume $\gamma_{r1} \le \gamma_{r2} \le \gamma_{r3}$

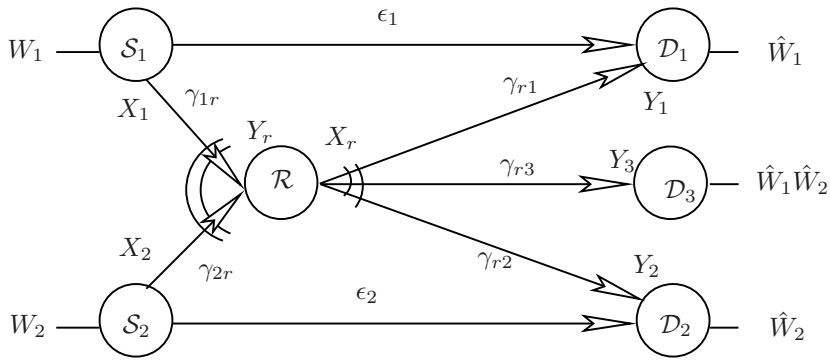


Figure 6.3. A non-coupled noisy network $\mathfrak{N}$ with source node $\mathcal{S}_1$ multicasting information W_1 at rate R_1 to destinations $\mathcal{D}_1$ and $\mathcal{D}_3$, and $\mathcal{S}_2$ multicasting W_2 at rate R_2 to destinations $\mathcal{D}_2$ and $\mathcal{D}_3$, with aid from a full-duplex relay $\mathcal{R}$. The point-to-point connections $\mathcal{S}_1$-$\mathcal{D}_1$ and $\mathcal{S}_2$-$\mathcal{D}_2$ are binary symmetric channels with crossover probability ϵ_1 and ϵ_2, respectively. The multiple-access channel ending at $\mathcal{R}$ and the broadcast channel originating from $\mathcal{R}$ are Gaussian channels.

as their order will determine the structure of the lower bounding model for the broadcast channel.

The point-to-point connections $\mathcal{S}_1$-$\mathcal{D}_1$ and $\mathcal{S}_2$-$\mathcal{D}_2$ are independent channels in the original network, and therefore they can be replaced by bit-pipe with throughput equals to $1 - H(\epsilon_1)$ and $1 - H(\epsilon_1)$, respectively, without affecting the capacity of the original network.

There are two types of upper bounding models for the multiple-access channel, i.e.,

$$\begin{aligned}\mathfrak{C}_{u,MAC,1} &= (R_{MAC}, \infty, \infty),\\ \mathfrak{C}_{u,MAC,2} &= (\infty, R_{d_1}, R_{d_2}),\end{aligned}$$

where

$$\begin{aligned}R_{MAC} &= \mathcal{C}\left((\sqrt{\gamma_{1r}} + \sqrt{\gamma_{2r}})^2\right),\\ R_{d_1} &= \mathcal{C}\left(\frac{\gamma_{1r}}{\alpha}\right),\\ R_{d_2} &= \mathcal{C}\left(\frac{\gamma_{2r}}{1-\alpha}\right),\\ \alpha &= \left(1 + \sqrt{\frac{1+1/\gamma_{1r}}{1+1/\gamma_{2r}}}\right)^{-1}.\end{aligned}$$

Since X_1 and X_2 are independent, the upper bounding model $\mathfrak{C}_{u,MAC,1}$ is loose as there will be no coherent combining gain at $\mathcal{R}$. Therefore only $\mathfrak{C}_{u,MAC,2}$ will be

used as the upper bounding model for the multiple-access channel. There are also two types of upper bounding models for the broadcast channel, i.e.,

$$\mathfrak{C}_{u,BC,1} = (R_{BC}, \infty, \infty, \infty),$$
$$\mathfrak{C}_{u,BC,2} = (\infty, R_{l_1}, R_{l_2}, R_{l_3}),$$

where

$$R_{BC} = \mathcal{C}\left(\gamma_{r1} + \gamma_{r2} + \gamma_{r3}\right),$$
$$R_{l_1} = \mathcal{C}\left(\gamma_{r1}\right),\ R_{l_2} = \mathcal{C}\left(\gamma_{r2}\right),\ R_{l_3} = \mathcal{C}\left(\gamma_{r3}\right).$$

Since there is no destination cooperation among Y_1, Y_2, and Y_3, the upper bounding model $\mathfrak{C}_{u,BC,1}$ is loose. Therefore only $\mathfrak{C}_{u,BC,2}$ will be used to upper bound the broadcast channel. The resulting upper bounding network $\mathfrak{C}_u$ is shown in Figure 6.4. The capacity outer bound can be calculated by applying the max-flow min-cut principle in the noiseless network $\mathfrak{C}_u$, which is defined by the pentagon

$$\boldsymbol{C}_u(R_1, R_2) = \begin{cases} 0 \le R_1 \le 1 - H(\epsilon_1) + \min\{R_{d_1}, R_{l_1}\}, \\ 0 \le R_2 \le 1 - H(\epsilon_2) + \min\{R_{d_2}, R_{l_2}\}, \\ R_1 + R_2 \le \min\{2 - H(\epsilon_1) - H(\epsilon_2) + R_{d_1} + R_{d_2}, \\ \qquad\qquad 2 - H(\epsilon_1) - H(\epsilon_2) + R_{l_1} + R_{l_2},\ R_{l_3}\}. \end{cases} \tag{6.7}$$

The lower bounding model for the multiple-access channel is given by the capacity region of the corresponding MAC channel, i.e.,

$$\mathfrak{C}_{l,MAC} = (R'_{d_s}, R'_{d_1}, R'_{d_2}) = (\mathcal{C}(\gamma_{1r} + \gamma_{2r}), \mathcal{C}(\gamma_{1r}), \mathcal{C}(\gamma_{2r})).$$

The lower bounding model for the broadcast channel can be constructed according to the description in Section 6.2.2 based on the realization of the channel SNR γ_{rj}, $j = 1, 2, 3$. Since we have $\gamma_{r1} \le \gamma_{r2} \le \gamma_{r3}$, the corresponding lower bounding can be written as

$$\mathfrak{C}_{l,BC} = [R'_0, R'_4, R'_6, R'_7], \tag{6.8}$$

where

$$R'_0 = R'_4 + R'_6 + R'_7,$$
$$R'_7 = \mathcal{C}\left(\frac{\gamma_{r1}\beta_1}{1 + \gamma_{r1}(1 - \beta_1)}\right),$$
$$R'_6 = \mathcal{C}\left(\frac{\gamma_{r2}\beta_2}{1 + \gamma_{r2}(1 - \beta_1 - \beta_2)}\right),$$
$$R'_4 = \mathcal{C}\left(\gamma_{r3}(1 - \beta_1 - \beta_2)\right),$$

with $\beta_1, \beta_2 \ge 0$, $\beta_1 + \beta_2 \le 1$ as power allocation parameters for superposition encoding at $\mathcal{R}$. R'_4 represents the rate at which the same message can be decoded

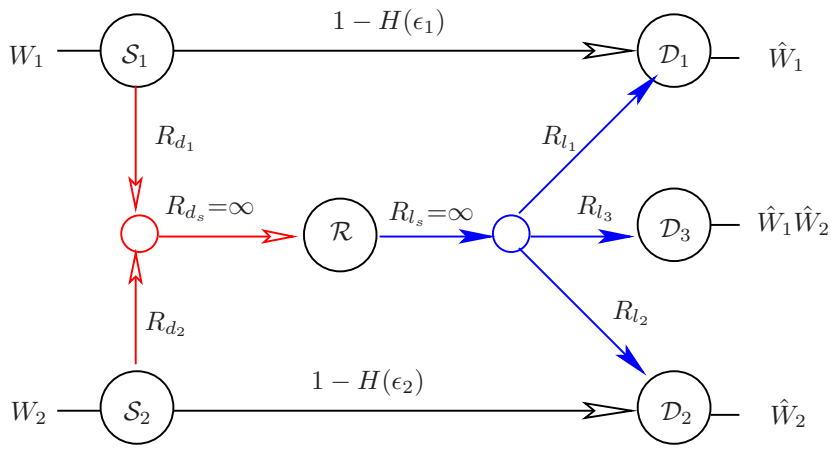


Figure 6.4. The upper bounding network $\mathfrak{C}_u$ where the multiple-access channel is replaced by $\mathfrak{C}_{u,MAC,2}$ and the broadcast channel is replaced by $\mathfrak{C}_{u,BC,2}$.

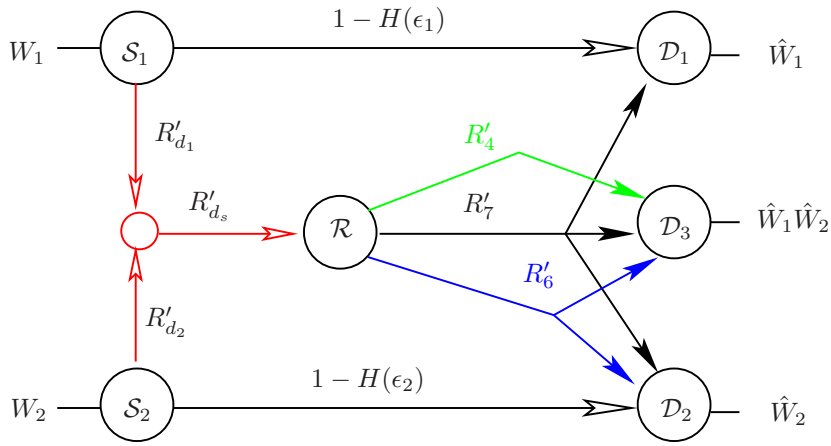


Figure 6.5. The lower bounding network $\mathfrak{C}_l$ where the multiple-access channel is replaced by $\mathfrak{C}_{l,MAC}$ and the broadcast channel is replaced by $\mathfrak{C}_{l,BC}$.

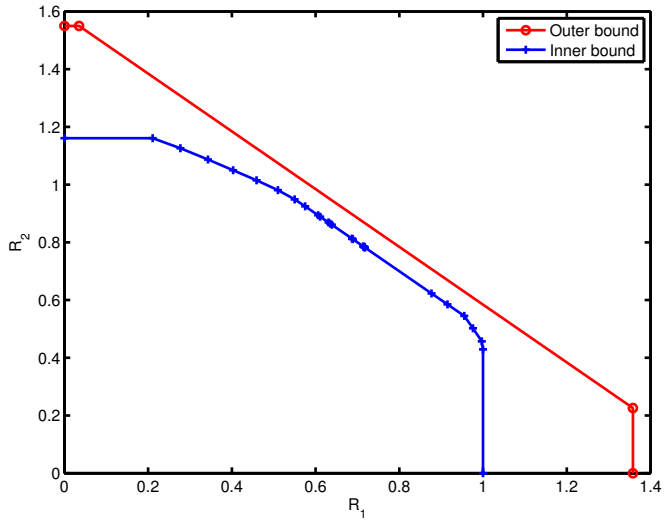


Figure 6.6. The capacity inner and outer bounds for the noisy network $\mathfrak{N}$ obtained from noiseless bounding networks $\mathfrak{C}_l \subseteq \mathfrak{N} \subseteq \mathfrak{C}_u$, with $\epsilon_1 = 0.02$, $\epsilon_2 = 0.04$, $\gamma_{1r} = 3$, $\gamma_{2r} = 4$, $\gamma_{r1} = 1$, $\gamma_{r2} = 2$, and $\gamma_{r3} = 8$.

only by $\mathcal{D}_3$, R'_6 represents the rate for success decoding only by $\mathcal{D}_2$ and $\mathcal{D}_1$, and R'_7 represents the rate for successful decoding by all the destination nodes.

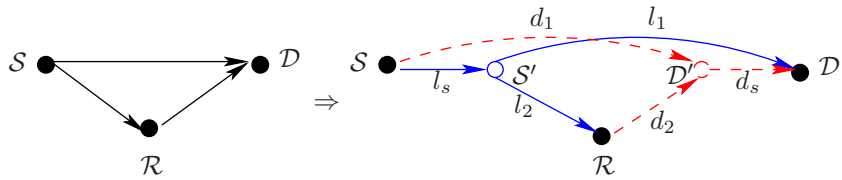
The resulting lower bounding network $\mathfrak{C}_l$ is shown in Figure 6.5, where the redundant bit-pipe constraint R'_0 is removed (therefore the auxiliary node for the broadcast channel is merged into $\mathcal{R}$). Given each valid realization of (β_1, β_2), there is a valid lower bounding network $\mathfrak{C}_l(\beta_1, \beta_2)$, which can be utilized to construct a valid capacity inner bound for the original network $\mathfrak{N}$. Note that for networks with general topology, there are many heuristic ways to construct a valid capacity inner bound, depending on how the routing scheme combines with network coding strategies. For the lower bounding network $\mathfrak{C}_l(\beta_1, \beta_2)$ in Figure 6.5, the capacity inner bound can be described by the pentagon

$$\boldsymbol{C}_l(\beta_1, \beta_2) = \left\{ \begin{array}{l} 0 \le R_1 \le \min\{1 - H(\epsilon_1) + R'_7,\ R'_{d_1}\}, \\ 0 \le R_2 \le \min\{1 - H(\epsilon_2) + R'_6,\ R'_{d_2}\}, \\ R_1 + R_2 \le \min\{R'_4 + R'_6 + R'_7,\ R'_{d_s}\}. \end{array} \right. \tag{6.9}$$

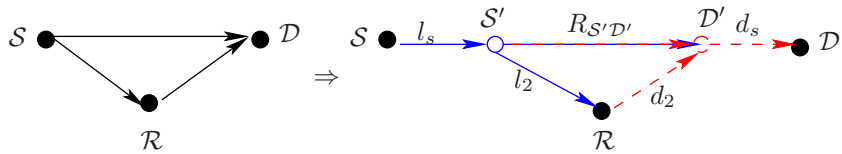
The final capacity inner bound is therefore

$$\boldsymbol{C}_l(R_1, R_2) = \bigcup_{\substack{\beta_1, \beta_2 \ge 0 \\ \beta_1 + \beta_2 \le 1}} \boldsymbol{C}_l(\beta_1, \beta_2).$$

As shown in Figure 6.6, we have compared the capacity inner and outer bounds obtained from noiseless bounding networks $\mathfrak{C}_l$ and $\mathfrak{C}_u$, respectively, for the original



(a) equivalent network after applying bounding models



(b) bounding network by merging constraints on bit-pipe $\mathcal{S}'$–$\mathcal{D}'$

Figure 6.7. Bounding network for the classical 3-node relay network by applying the one-shot bounding models. The bit-pipe l_1 between auxiliary nodes $\mathcal{S}'$ and $\mathcal{D}$ is due to the broadcast channel starting from source node $\mathcal{S}$, and the bit-pipe d_1 is due to the multiple-access channel ending at destination node $\mathcal{D}$.

noisy network $\mathfrak{N}$ with $\epsilon_1 = 0.02$, $\epsilon_2 = 0.04$, $\gamma_{1r} = 3$, $\gamma_{2r} = 4$, $\gamma_{r1} = 1$, $\gamma_{r2} = 2$, and $\gamma_{r3} = 8$. The gap between the inner and outer bounds is within 0.1 bits on the sum rate, and within 0.4 bits on individual rates.

6.3 Bounding Models for Coupled Networks

6.3.1 Channel Decoupling

The one-shot bounding model proposed in [CME11] and its extensions presented in Section 6.2 are designed for non-coupled networks, where multiple-access and broadcast channels are independent. In wireless networks, however, these multiple-access and broadcast channels are usually coupled together due to the broadcast nature of wireless transmission.

For each broadcasting node with m receivers, we introduce an auxiliary node with one input bit-pipe channel and m output bit-pipe channels. For each receiving node with n input sources, we introduce an auxiliary node with n input bit-pipe channels and one output bit-pipe channel. For the full-duplex nodes in the network, we introduce two auxiliary nodes, one for the receiving functionality and the other for the transmitting functionality. When a wireless connection between two points is part of both a broadcast channel and a multiple-access channel, there will be two bounding bib-pipe channels for this connection after applying the one-shot solution, as illustrated in Figure 6.7(a) for the classical 3-node relay channel where the connection between the source node $\mathcal{S}$ and the destination node $\mathcal{D}$ is part of both the broadcast channel starting from $\mathcal{S}$ and the multiple-access channel ending at $\mathcal{D}$. As the bit-pipe channels $\mathcal{S}$–$\mathcal{D}'$ and $\mathcal{S}'$–$\mathcal{D}$ are constraints on the same

connection $\mathcal{S}$–$\mathcal{D}$, we can merge these two bit-pipe constraints on the same point-to-point connection $\mathcal{S}'$–$\mathcal{D}'$, as illustrated in Figure 6.7(b). Note that the constraint l_1 imposed by the broadcast channel and the constraint d_1 imposed by the multiple-access channel should be simultaneously respected in the final bounding networks. We will formally prove this in Theorem 6.1 for the upper bounding models and demonstrate it in Section 6.3.3 for the lower bounding models.

6.3.2 Outer Bounds

Theorem 6.1. *Given a noisy network with independent noise, if a noisy connection between two nodes is part of both a broadcast channel and a multiple-access channel, the corresponding bit-pipe in the upper bounding network should have a throughput that equals to the maximum of the two individual rate requirements imposed by the bonding models for corresponding broadcast and multiple-access channels.*

Proof. Proof outline.
We first look at the classical relay network as illustrated in Figure 6.7, and then extend our proof to general setups. The one-shot upper bounding models for the relay network can be written as

$$\begin{aligned}
\mathfrak{C}_{u,BC,1} &= (R_{BC}, \log(|\mathcal{Y}|), \log(|\mathcal{Y}_r|)),\\
\mathfrak{C}_{u,BC,2} &= (\log|\mathcal{X}|, R_{l_1}, R_{l_2}),\\
\mathfrak{C}_{u,MAC,1} &= (R_{MAC}, \log(|\mathcal{X}|), \log(|\mathcal{X}_r|)),\\
\mathfrak{C}_{u,MAC,2} &= (\log(|\mathcal{Y}|), R_{d_1}, R_{d_2})
\end{aligned}$$

where

$$\begin{aligned}
R_{BC} &\triangleq \max_{p(x)} I(X; Y_r, Y),\\
R_{l_1} &\triangleq \max_{p(x)} I(X; Y),\\
R_{l_2} &\triangleq \max_{p(x)} I(X; Y_r)\\
R_{MAC} &\triangleq \max_{p(x,x_r)} I(X, X_r; Y),\\
R_{d_1} &\triangleq \max_{p(x)} I(X; Z_1),\\
R_{d_2} &\triangleq \max_{p(x_r)} I(X_r; Z_2),
\end{aligned}$$

and Z_1, Z_2 are intermediate auxiliary variables to emulate the channel output Y via a predefined deterministic function $y = g(z_1, z_2)$ such that the channel transition probability can be partitioned as

$$p(y|x, x_r) = \sum_{\substack{z_1,z_2:\\ y=g(z_1,z_2)}} p(z_1|x)p(z_r|x_r),$$

and the sum rate $R_{d_1} + R_{d_2}$ can be minimized.

There are in total four combinations after applying the one-shot bounding models for the broadcast and the multiple-access channels. In the following we will show that, for each combination, merging rate constraints on $\mathcal{S}$–$\mathcal{D}'$ and $\mathcal{S}'$–$\mathcal{D}$ together and replacing these links by a connection $\mathcal{S}'$–$\mathcal{D}'$ is sufficient to construct a valid upper bounding network.

When $\mathfrak{C}_{u,BC,2}$ and $\mathfrak{C}_{u,MAC,2}$ are used for the BC and the MAC, respectively, it is obvious that $\mathcal{S}'$ in $\mathfrak{C}_{u,BC,2}$ behaves exactly the same as $\mathcal{S}$ and therefore replacing the bit-pipe $\mathcal{S}$–$\mathcal{D}'$ by $\mathcal{S}'$–$\mathcal{D}'$ will not affect the model $\mathfrak{C}_{u,MAC,2}$ as

$$R_{l_1} = \max_{p(x)} I(X;Y) \leq \log(|\mathcal{Y}|).$$

For discrete random variable Y, the inequality comes from the fact that $I(X;Y) \leq H(Y)$ and the concavity of the function $f(x) = -x\log(x)$. If Y is a continuous-valued random variable, the inequality holds since $|\mathcal{Y}| = \infty$. Following similar argument we can replace the bit-pipe $\mathcal{S}'$–$\mathcal{D}$ by $\mathcal{S}'$–$\mathcal{D}'$ due to the fact that $R_{d_1} \leq \log(|\mathcal{X}|)$. By choosing the throughput of $\mathcal{S}'$–$\mathcal{D}'$ as

$$R_{\mathcal{S}'\mathcal{D}'} = \max(R_{l_1}, R_{d_1}),$$

the requirement of both upper bounding models $\mathfrak{C}_{u,BC,2}$ and $\mathfrak{C}_{u,MAC,2}$ can be fulfilled over the noiseless network.

When $\mathfrak{C}_{u,BC,2}$ and $\mathfrak{C}_{u,MAC,1}$ are used, the throughput of the bit-pipe $\mathcal{S}'$–$\mathcal{D}'$ can be set as

$$R_{\mathcal{S}'\mathcal{D}'} = \max(R_{l_1}, \log(|\mathcal{X}|)).$$

Then the bit-pipe $\mathcal{S}'$–$\mathcal{D}$ originating from $\mathfrak{C}_{u,BC,2}$ can be replaced by $\mathcal{S}'$–$\mathcal{D}'$ as $R_{l_1} \leq R_{MAC}$, and the inequality comes from the fact that

$$I(X;Y) \leq I(X;Y) + I(X_r;Y|X) = I(X, X_r;Y).$$

The bit-pipe $\mathcal{S}$–$\mathcal{D}'$ originating from $\mathfrak{C}_{u,MAC,1}$ can be replaced by $\mathcal{S}'$–$\mathcal{D}'$ as the bounding channels l_s and d_1 have the same throughput $\log(|\mathcal{X}|)$.

When $\mathfrak{C}_{u,BC,1}$ and $\mathfrak{C}_{u,MAC,1}$ are used, we set the throughput of the bit-pipe $\mathcal{S}'$–$\mathcal{D}'$ as

$$R_{\mathcal{S}'\mathcal{D}'} = \max(\log(|\mathcal{X}|), \log(|\mathcal{Y}|)).$$

Then we need to show that the resulting bounding model is still valid after we replace $\mathcal{S}'$–$\mathcal{D}$ and $\mathcal{S}$–$\mathcal{D}'$ by $\mathcal{S}'$–$\mathcal{D}'$. If X is a continuous-valued random variable, we have $|\mathcal{X}| = \infty$ while R_{BC} is finite[2], which leads to $R_{BC} < \log(|\mathcal{X}|)$. If X is discrete random variable, we have

$$R_{BC} = \max_{p(x)} I(X; Y_r, Y) \leq \max_{p(x)} H(X) \leq \log(|\mathcal{X}|).$$

[2] Y, Y_r are noisy version of X and therefore $I(X; Y_r, Y)$ is finite.

Similarly we have $R_{MAC} \leq \log(|\mathcal{Y}|)$. Therefore we need a new approach to proving the validity of the resulting upper bounding model. It is obvious that R_{BC} and R_{MAC} equal to the corresponding rate constraints obtained by applying the cut-set bound at the source node and the destination node, respectively. To make our bounding model a valid upper bound as suggested by the cut-set bound constraints, we can set

$$R_{\mathcal{S}'\mathcal{D}'} = \max(R_{BC}, R_{MAC}).$$

Since $R_{BC} < \log(|\mathcal{X}|)$ and $R_{MAC} \leq \log(|\mathcal{Y}|)$, setting

$$R_{\mathcal{S}'\mathcal{D}'} = \max(\log(|\mathcal{X}|), \log(|\mathcal{Y}|))$$

will not disqualify our upper bounding model.

When $\mathfrak{C}_{u,BC,1}$ and $\mathfrak{C}_{u,MAC,2}$ are used, we set the throughput of the bit-pipe $\mathcal{S}'$–$\mathcal{D}'$ as $R_{\mathcal{S}'\mathcal{D}'} = \max(\log(|\mathcal{Y}|), R_{d_1})$. The bit-pipe $\mathcal{S}'$–$\mathcal{D}$ originating from $\mathfrak{C}_{u,BC,1}$ can be replaced by $\mathcal{S}'$–$\mathcal{D}'$ as the bounding channels l_1 and d_s have the same throughput $\log(|\mathcal{Y}|)$. Since $R_{d_1} \leq R_{BC}$ might not hold, the replacement of $\mathcal{S}$–$\mathcal{D}'$ by $\mathcal{S}'$–$\mathcal{D}'$ has to be validated following the similar argument as before thanks to the cut-set rate constraint R_{BC}.

We can now conclude that Theorem 6.1 holds for the classical relay network.

For general network setups, we first divide each full-duplex node into two parts, one for transmission and one for reception, interconnected by a directed bit-pipe of infinite rate from the reception part to the transmission part. Assuming the self-interference can be perfectly cancelled, there will be no backward connection from the transmission part to the reception part. Such transmission/reception part will be treated as an independent node when applying the one-shot bounding model, and we will not differentiate them from ordinary nodes (either transmission or reception, but not both).

A noisy connection $\mathfrak{n}_{i,j} = (\mathcal{X}_i, p(y_j|x_i), \mathcal{Y}_j)$ from a node transmitting X_i to another node whose received signal is Y_j in a coupled network can be described by the marginal distribution

$$p(y_j|x_i) = \sum_{\boldsymbol{x}_{/i}, \boldsymbol{y}_{/j}} p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}_{/i}),$$

where

$$p(\boldsymbol{y}|\boldsymbol{x}) \triangleq p(y_1, \ldots, y_j, \ldots, y_q | x_1, \ldots, x_i, \ldots, x_p)$$

is the transition probability of the *smallest* multiple input multiple output network $\mathfrak{N} = (\prod_{n=1}^{p} \mathcal{X}_n, p(\boldsymbol{y}|\boldsymbol{x}), \prod_{k=1}^{q} \mathcal{Y}_k)$ which contains $\mathfrak{n}_{i,j}$, i.e., the broadcast channel with input X_i is necessarily and sufficiently described by the transition probability $p(\boldsymbol{y}|x_i)$ and the multiple-access channel with output Y_j is necessarily and sufficiently described by $p(y_j|\boldsymbol{x})$. The noisy connection $\mathfrak{n}_{i,j}$ and the smallest multiple input multiple output network $\mathfrak{N}$ that contains it are illustrated in the left part of Figure 6.8.

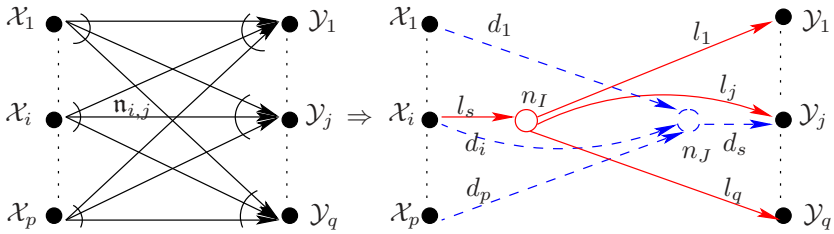


Figure 6.8. The noisy connection $\mathfrak{n}_{i,j}$ from the node with transmitting alphabet $\mathcal{X}_i$ to the node with receiving alphabet $\mathcal{Y}_j$ in a coupled network can be fully described by the smallest multiple input multiple output network $\mathfrak{N}$ that contains it. Applying the one-shot bounding models to the broadcast and the multiple channels that contain $\mathfrak{n}_{i,j}$ will incur two parallel rate constraints. The bit-pipe l_j is due to the broadcast channel and the bit-pipe d_i comes from the multiple-access channel. Both l_j and d_i should be replaced by a single bit-pipe from auxiliary node n_I to auxiliary node n_J when finalizing the noiseless bounding network.

As $\mathfrak{n}_{i,j}$ can be fully described by $\mathfrak{N}$, we can apply the one-shot bounding models to the broadcast and the multiple channels that contain $\mathfrak{n}_{i,j}$ to construct the bounding networks. The upper bounding model for the broadcast channel with input X_i can be written as

$$\mathfrak{C}_{u,BC} = (R_{l_s}, R_{l_1}, \ldots, R_{l_j}, \ldots, R_{l_q}),$$

where R_{l_j} corresponds to the rate constraint on the parallel channel l_j with output Y_j. The upper bounding model for the multiple-access channel with output Y_j can be written as

$$\mathfrak{C}_{u,MAC} = (R_{d_s}, R_{d_1}, \ldots, R_{d_i}, \ldots, R_{d_p}),$$

where R_{d_i} corresponds to the rate constraint on the parallel channel d_i with input X_i, as illustrated in the right part of Figure 6.8. Since the auxiliary nodes n_I and n_J are introduced to emulate the original nodes, we could create a bit-pipe of throughput $R_{i,j}$ from n_I to n_J to replacing both l_j and d_i. Following the same argument as we have done for the relay network, we can show that $R_{i,j} = \max(R_{l_j}, R_{d_i})$ is sufficient to fulfill the requirement of both tasks.

□

Key Steps to Construct Capacity Outer Bounds

Given a coupled network, we can construct the capacity outer bound step by step as follows.

Step I: Channel Decoupling for Network Decomposition

For each full-duplex node $\mathcal{R}$ in the network with input signal Y and output signal X, we introduce a virtual node $\mathcal{R}^r$ representing its functionality of signal reception if Y is the output of a multiple-access channel ending at $\mathcal{R}$. Similarly we introduce a virtual node $\mathcal{R}^t$ representing its functionality of signal transmission if

X is the input of a broadcast channel originating from $\mathcal{R}$. These virtual nodes (if any) will be connected to the original node $\mathcal{R}$ by an infinite-rate bit-pipe. Such transmission/reception virtual nodes will be treated as independent nodes when applying the one-shot bounding model, and we will not differentiate them from ordinary nodes (can be the starting point of a broadcast channel or the ending point of a multiple-access channel, but not both simultaneously).

Step II: Apply One-shot Bounding Models

For each independent point-to-point connection, we replace it with a bit-pipe whose throughput equals to its capacity [KEM11]. Similarly, for any independent multiple-access channel or broadcast channel, replace it with the corresponding one-shot bit-pipe models. Then, for each pair of coupled broadcast channel and multiple-access channel, i.e., there is a noisy connection that is part of both the broadcast and the multiple-access channel, we identify the smallest multiple input multiple output network that contains this noisy connection, and hence also contains the pair of coupled broadcast and multiple-access channels. We apply the one-shot bounding models to this multiple input multiple output network, where the resulting bounding model contains a point-to-point connection that satisfies rate constraints placed by both the broadcast channel and the multiple-access channel, as described in Theorem 6.1.

Step III: Construct a Valid Outer Bound

According to the max-flow min-cut theorem, the maximum throughput from source to sink can be no larger than the value of the minimum cut in between. For each transmission task (unicast or multiple cast), we identify all the cuts in the resulting bounding network (which contains only noiseless point-to-point connections) and calculate the flows across each cut. The resulting capacity region is therefore an outer bound[3] for the upper bounding network, and hence also a valid outer bound for the original coupled noisy network.

Step IV: Take Intersection of all Valid Outer Bounds

For each MAC or BC component, there are two different upper bounding models, which results in 2^M different upper bounding networks after Step II, where M is number of MAC/BC components. Each of these upper bounding networks will result in a valid outer bound as described in Step III. We can then take intersection of all these outer bounds to get a tighter outer bound.

6.3.3 Inner Bounds with Updated Lower Bounding Models

The lower bounding model designed for non-coupled networks assumes isolated source/destination nodes, without taking into account the possibility that the transmitted signal can be designed for multiple destinations (in broadcast channels) and

[3] As shown in [KM03], the max-flow min-cut theorem is tight on some noiseless networks, which include noiseless networks associated with single-source multiple-unicast transmission, single-source (two-level) multicast transmission, and multi-source multicast transmission. Therefore the bound we obtained by the max-flow min-cut theorem might be the capacity region for the corresponding upper bounding network.

the received signal may consist of signals from several source nodes (in multiple-access channels) and thus interfere with each other. When a noisy connection between two nodes is part of both a broadcast and a multiple-access channel, the bounding models for the broadcast channel and for the multiple-access channel have to be updated. We will demonstrate how the update should be done step by step as follows. Note that the proposed method will ensure a valid lower bounding network, without claiming its optimality.

Key Steps to Construct Capacity Lower Bound

Given a coupled network, we can construct the capacity inner bound step by step as follows.

Step I: Channel Decoupling for Network Decomposition

Following the same procedure as in Section 6.3.2 we first decompose the original network such that every node in the resulting network can only be the starting point of a broadcast channel or the ending point of a multiple-access channel, but not both simultaneously.

Step II: Apply Lower Bounding Models for Broadcast Channels and for Non-coupled Multiple-Access Channels

For each broadcast channel in the resulting network, we apply the lower bounding models for broadcast channels as described in Section 6.2.2. For each non-coupled multiple-access channel, i.e., neither of its input signals is part of the output of a broadcast channel, we replace it with a lower bounding model according to existing achievable-rate coding schemes.

Step III: Construct Lower Bounding Models for Coupled Multiple-Access Channels

If (some of) the input signals to a multiple-access channel are the output signals from broadcast channels, part of the input signals to the multiple-access channel can not be decoded by the receiver and therefore behaves as interference. The original lower bounding models for non-coupled multiple-access channels, which assume that all the input signals can be decoded, need to be updated based on the sum power of the interfering signals, which could be calculated by taking into account the the signal structure of each input source nodes. We illustrate the principle of constructing such new lower bounding models by a coupled Gaussian multiple channel as follows.

Example: Gaussian Multiple-Access Channel with m ***Transmitters***

Given a Gaussian multiple-access channel $\mathfrak{N} = (\prod_{i=1}^{m} \mathcal{X}_i, p(y|\boldsymbol{x}), \mathcal{Y})$, where X_i $(i = 1, \ldots, m)$ is the input signal generating the received signal power γ_i (product of transmitted signal power and the corresponding channel gain). If X_i can only be received by the receiver in channel $\mathfrak{N}$, all the components of X_i can be fully decoded by the receiver. If X_i can also be received by other nodes, i.e., X_i is the transmitted signal from a broadcast source node, it may contains components that do not intend to be decoded by the receiver due to rate and power allocation at the broadcast node, as described in Section 6.2.2 for constructing the lower bounding model for

Gaussian broadcast channels. The remaining component of X_i cannot be decoded by the receiver and therefore behaves as interference during the decoding process. We denote the power of the interfering component by Γ_i, and the exact value can be obtained from the lower bounding model with the knowledge of the corresponding power allocation parameters chosen by the broadcast channel transmitting X_i. We will have $\Gamma_i = 0$ if all messages contained in X_i are intended for successful decoding, and $\Gamma_i = \gamma_i$ if nothing is to be decoded.

After careful examination of the structure of all the input signals, we can calculate the total power of interfering components contained in Y as follows

$$P_I = \sum_{i=1}^{m} \Gamma_i, \tag{6.10}$$

out of which

$$P_{I,i} = P_I - \Gamma_i = \sum_{j \neq i} \Gamma_j \tag{6.11}$$

is the amount of interference power introduced by input signals other than X_i. We call $P_{I,i}$ the "extrinsic interference" of X_i.

We can now construct the lower bounding model for the multiple-access channel $\mathfrak{N}$ based on the total power of interfering components P_I obtained from (6.10). The throughput constraints on the bit-pipe d_i (corresponding to the connection X_i-Y) in the lower bounding model is therefore

$$R_{d_i} = \mathcal{C}\left(\frac{\gamma_i - \Gamma_i}{1 + P_I}\right), \tag{6.12}$$

and the constraints on sum rate are

$$\sum_{i \in S} R_{d_i} = \mathcal{C}\left(\frac{\sum_{i \in S}(\gamma_i - \Gamma_i)}{1 + P_I}\right),$$

where $S \subseteq \{1, 2, \ldots, m\}$ is any subset of transmitting nodes.

Step IV: Rate Adjustment for Coupled Broadcast Channels

If a broadcast channel is coupled with a multiple-access channel, the lower bounding model for this broadcast channel should also be adjusted. As described in Section 6.2.2 for Gaussian broadcast channels, the lower bounding model for a non-coupled broadcast channel depends on the effective link SNR of each individual channels. Assume the noisy connection X_n-Y_k is both part of a broadcast channel transmitting X_n and a multiple-access channel with received signal Y_k, we denote the corresponding link SNR by γ_{nk}. After step III, we can obtain by (6.11) the extrinsic interference power $P_{I,nk}$, caused by input signals other than X_n in the coupled multiple-access channel with output Y_k. We can now adjust the lower bounding model for the broadcast channel transmitting X_n with new rate

constraints $R'_{2^m-2^{i-1}}$, corresponding to successful transmission of message W_i via channels l_i to l_m, by taking into account the extrinsic interference power $P_{I,nk}$ for $k = i, \ldots, m$. Therefore (6.5) in Section 6.2.2 which defines $R_{2^m-2^{i-1}}$, should be adjusted to

$$R'_{2^m-2^{i-1}} = \min_{k\in\{i,i+1,\ldots,m\}} \frac{1}{2}\log\left(1+\frac{\gamma_{nk}\beta_i}{1+P_{I,nk}+\gamma_{nk}\sum_{j=i+1}^{m}\beta_j}\right). \tag{6.13}$$

The sum rate constraint should be adjusted by

$$R'_0 = \sum_{i=1}^{m} R'_{2^m-2^{i-1}}.$$

Note that the minimum operation in (6.13) comes from the fact that given $\gamma_{ni} \le \ldots \le \gamma_{nm}$ we cannot guarantee

$$\frac{\gamma_{ni}}{1+P_{I,ni}} \le \ldots \le \frac{\gamma_{nm}}{1+P_{I,nm}}$$

due to the effect of the extrinsic interference caused by coupled multiple-access channels. Here we simply keep the structure of the original lower bounding model unchanged without claiming its optimality.

Step V: Construct Inner Bound for the Resulting Noiseless Network

After applying one-shot bounding models, there will be two rate constraints on the same bit-pipe if the corresponding noisy connection is both part of a broadcast channel and a multiple-access channel. The rate constraint on this bit-pipe in the resulting lower bounding network is therefore the minimum of these two rate constraints.

The resulting lower bounding network consists of only noiseless bit-pipes, but it may contain hyper-arcs that carry the same data from one point to multiple points if the original noisy network has broadcast channels. How to find the optimal scheme to manage the data flows over such noiseless networks is in general an open problem. However, there exist many heuristic (and thus suboptimal in general) methods, see [THMK] for example, to construct a valid inner bound.

Step VI: Take Union of all Valid Inner Bounds

When superposition coding is used in constructing the lower bounding model for broadcast channels, there are infinite[4] number of lower bounding models due to power allocation parameters. We can construct a valid lower bound for each of these bounding models and then take the union of all the valid lower bounds to form the final inner bound.

[4]In practice we have finite number of power allocation parameters, which will result in a slightly looser inner bound.

6.3.4 Illustrative Examples

A Coupled Noisy Network Consisting of Two-User MAC/BC

We first illustrate the channel decoupling method and the upper bounding network construction based on a wireless multicast relay network shown in Figure 6.9, where two source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with backhaul (rate C_{12} and C_{12}), multicast information W_1 at rate R_1 and W_2 at rate R_2, respectively, to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ through Gaussian channels, with aid from a full-duplex relay $\mathcal{R}$. All the broadcast channels consist of two receivers and all the multiple-access channels have two transmitters. The bounding models (k_s, k_1, k_2), (l_s, l_1, l_2), and (r_s, r_1, r_2) come from the broadcast channels originating from $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{R}$, respectively. The bounding models (p_s, p_1, p_2), (q_s, q_1, q_2), and (t_s, t_1, t_2) come from the multiple-access channels ending at $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{R}$, respectively. Note that the bit-pipes k_1 and p_1 refer to the same point-to-point noiseless connection between $\mathcal{S}_1'$ and $\mathcal{D}_1'$, and therefore can be replaced by a single bit-pipe, as described by Theorem 6.1 for constructing upper bounding models and by Section 6.3.3 for constructing lower bounding models. The resulting upper bounding network is shown in the bottom of Figure 6.9, where the maximum operation comes from Theorem 6.1, and the minimum operation $\min\{R_{r_s}, R_{t_s}\}$ comes from the fact that the maximum flow passing through a bit-pipe is constraint by the bottleneck of that bit-pipe.

The upper and lower bounds on the sum rate obtained from the corresponding upper and lower bounding networks have been illustrated for high-rate backhaul scenarios in Figure 6.10 with respect to varying source-relay channel quality, and in Figure 6.11 with respect to varying relay-destination channel quality. The cut-set bound developed in Section 3.2, and lower bounds obtained by the NBF scheme developed in Section 4.1.4 and by the SNNC with message exchange scheme developed in Section 5.2.1, are plotted as references. The upper bound obtained directly from the noiseless upper bounding network is relatively good[5] in all the regions we have demonstrated, normally within a marginal difference from the cut-set bound that is derived based on vigorous and lengthy analysis on the whole network. The lower bound is always within 0.5 bits from the cut-set bound.

In Figure 6.10 the lower bonding models for the multiple-access channels at $\mathcal{D}_1$ and $\mathcal{D}_2$ are constraint by the weak source-relay and relay-destination channel gains, and they become the bottleneck in the resulting noiseless lower bounding network when the source-relay channel gain is strong. The lower bound on the sum rate derived from this lower bounding network is therefore unchanged despite the improved channel quality of the source-relay link. The lower bound in Figure 6.11 is constrained by the multiple-access channels at destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ for weak relay-destination channel quality ($g_{rd}^2 < 15$dB), and constrained by the broadcast channels originating from $\mathcal{S}_1$ and $\mathcal{S}_2$ when the relay-destination channels are strong.

[5] In the sense that the gap from the cut-set bound is small.

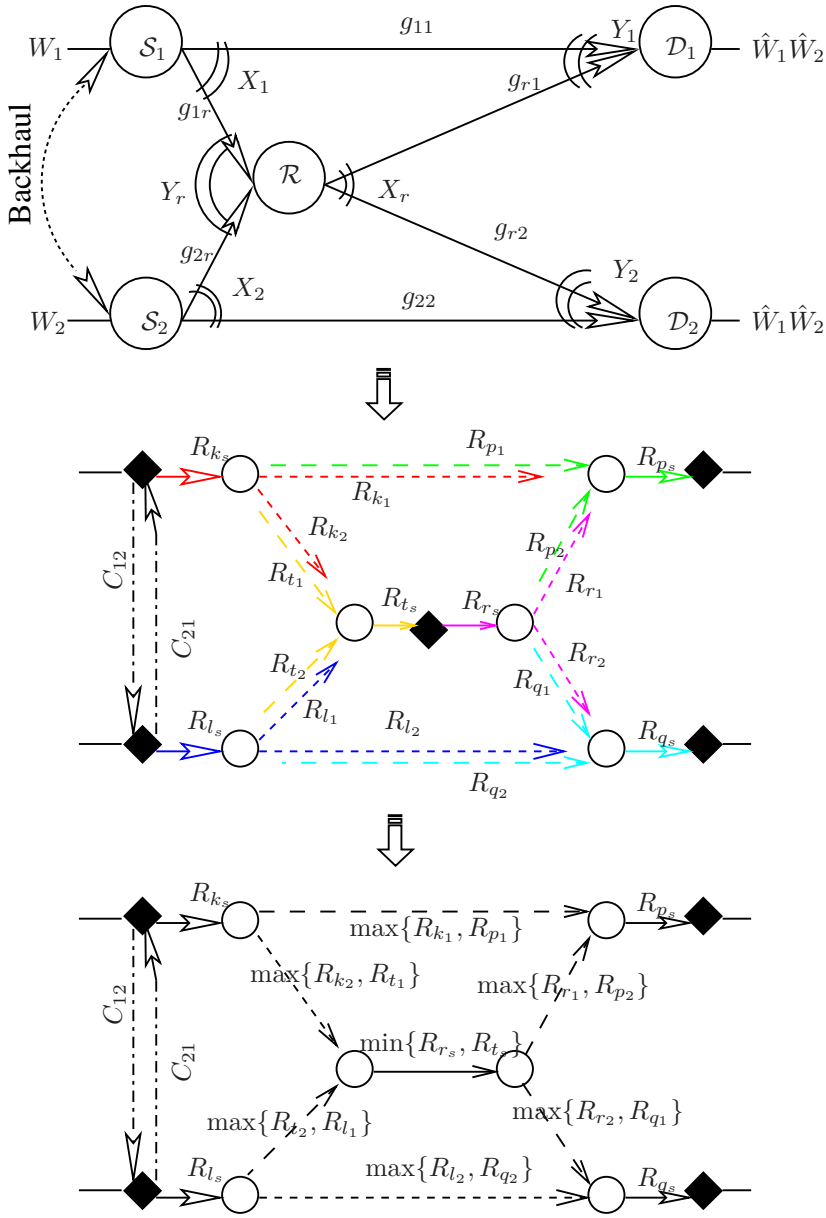


Figure 6.9. Channel decoupling for a wireless multiple multicast relay network, where two source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with backhaul (rate C_{12} and C_{12}), multicast information W_1 at rate R_1 and W_2 at rate R_2 respectively to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ through Gaussian channels, with aid from a full-duplex relay $\mathcal{R}$.

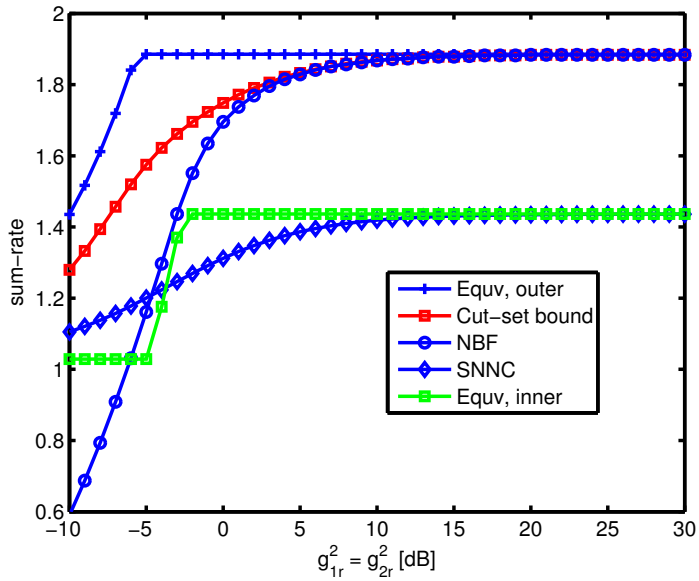


Figure 6.10. Capacity bounds on sum rate for high-rate backhaul scenario with $P_1 = P_2 = P_r = 5$dB and $g_{11} = g_{22} = g_{r1} = g_{r2} = 0$dB.

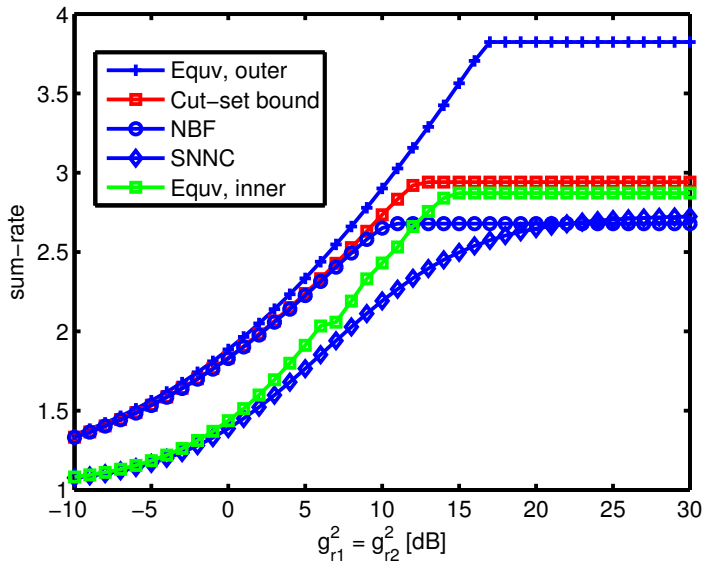


Figure 6.11. Capacity bounds on sum rate for high-rate backhaul scenario with $P_1 = P_2 = P_r = 5$dB, $g_{11} = g_{22} = 0$dB, and $g_{1r}^2 = g_{2r}^2 = 5$dB.

A Coupled Network Contains MAC/BC with Three Users

When the cross-links $\mathcal{S}_1$-$\mathcal{D}_2$ and $\mathcal{S}_2$-$\mathcal{S}_1$ exist, as shown in Figure 6.12, the broadcast channels originating from source nodes contain three receivers and the multiple-access channels ending at destination nodes contain three transmitters. We can first decompose the original network and then apply the bounding models, which will result in a bounding network as shown in Figure 6.12 (II). As both destination nodes need to decode both messages, we may treat them one-by-one when constructing the upper bounding networks, as illustrated in Figure 6.12 (I). We can then derive a valid upper bound based on the bounding network consisting only $\mathcal{D}_1$, and then find another upper bound based on the bounding network consisting only $\mathcal{D}_2$, and finally take intersection of them to get a valid upper bound for the original network. As shown in (6.4) and (6.6), the difference in sum rate between the upper and lower bounding models for broadcast channel with m receivers and for multiple-access channel with m transmitters is at most $0.5\log(m)$ bits per channel use. Therefore the outer bound obtained from bounding network (I) will be tighter than the one obtained from (II) due to smaller gap between upper and lower bound for the broadcast channels originating from source nodes. The lower bound, however, may not be derived from the bounding network (I) due to the change of channel setup. We will therefore construct lower bound only based on (II).

We first focus on the sum rate under various channel setups with high rate backhaul. We compare the resulting two upper bounds and the lower bound to the cut-set bound developed in Section 3.2, and the lower bounds obtained by the NBF scheme developed in Section 4.1.4 and by the SNNC with message exchange scheme developed in Section 5.2.1.

When the source-relay channels are good, as shown in Figure 6.13, the broadcast channels at source nodes are not a bottleneck and therefore there is no difference between the upper bounds obtained from bounding networks (I) and (II). When the source-relay channels are weak, as shown in Figure 6.14, the upper bound obtained from (I) is much better than the bound obtained from (II) as the broadcast channels at source nodes become a bottleneck. In both scenarios, the gap between the upper bound obtained from (I) and the cut-set bound is relatively small, usually less than 0.3 bits per channel use. The gap between the lower bound and the cut-set bound varies, from 0.4 bits per channel use up to 0.7 bits per channel use.

When the cross links are strong, as shown in Figure 6.15, the NBF scheme performs relatively poor since the decoding at the relay node becomes the bottleneck. Our lower bound, on the other hand, achieves surprisingly good performance when the cross links are of very good quality. The gap between our lower bound and the cut-set bound is less than 0.5 bits per channel use as shown in Figure 6.15. The good cross link quality helps to create bit-pipes of high throughput connecting source nodes and destinations, leading to high sum rate with the help of the high-rate backhaul. The upper bounds obtained from bounding networks (I) and (II) overlap each other.

For low-rate backhaul scenarios, we investigate the outer and inner bounds on

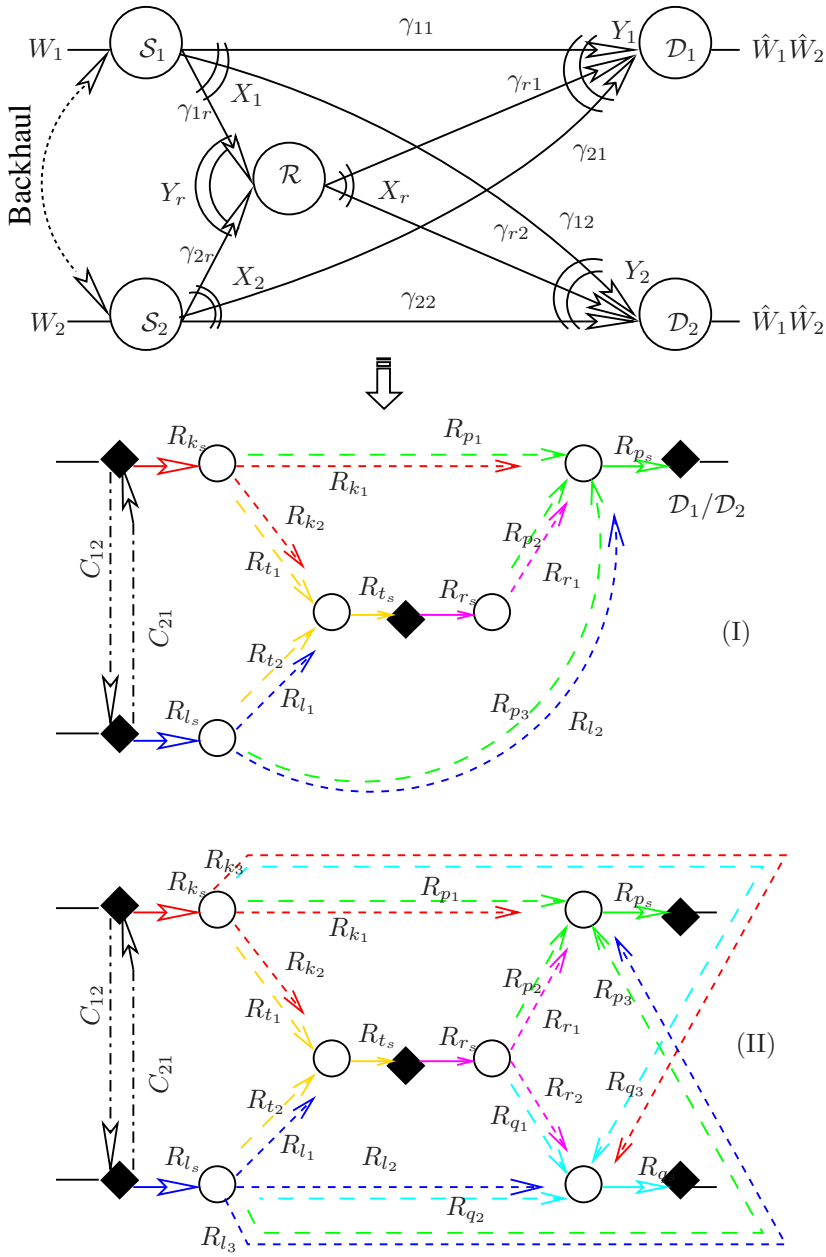


Figure 6.12. Channel decoupling for a wireless multiple multicast relay network, where two source nodes $\mathcal{S}_1$ and $\mathcal{S}_2$, connected with backhaul (rate C_{12} and C_{12}), multicast information W_1 at rate R_1 and W_2 at rate R_2 respectively to both destinations $\mathcal{D}_1$ and $\mathcal{D}_2$ through Gaussian channels, with aid from a full-duplex relay $\mathcal{R}$. The bounding network (I) is only used for constructing upper bounds.

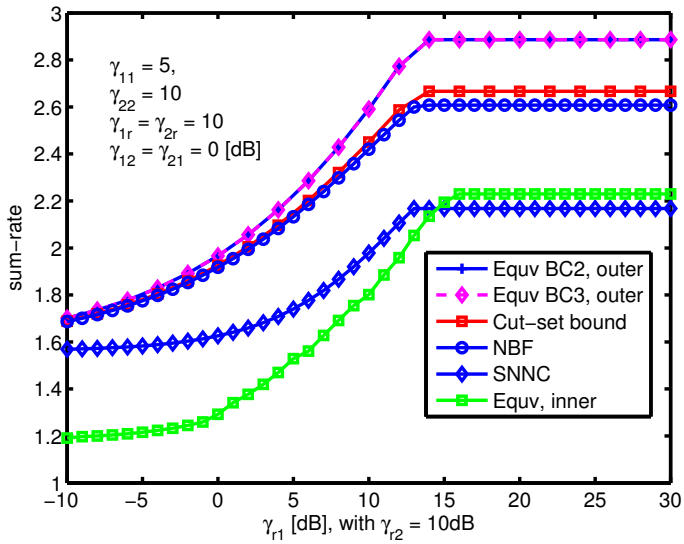


Figure 6.13. Capacity bounds on sum rate for high-rate backhaul scenario with varying $\mathcal{R}$-$\mathcal{D}_1$ link quality. $\gamma_{11} = 5$dB, $\gamma_{22} = 10$dB, $\gamma_{1r} = \gamma_{2r} = 10$dB, $\gamma_{12} = \gamma_{21} = 0$dB, and $\gamma_{r2} = 10$dB.

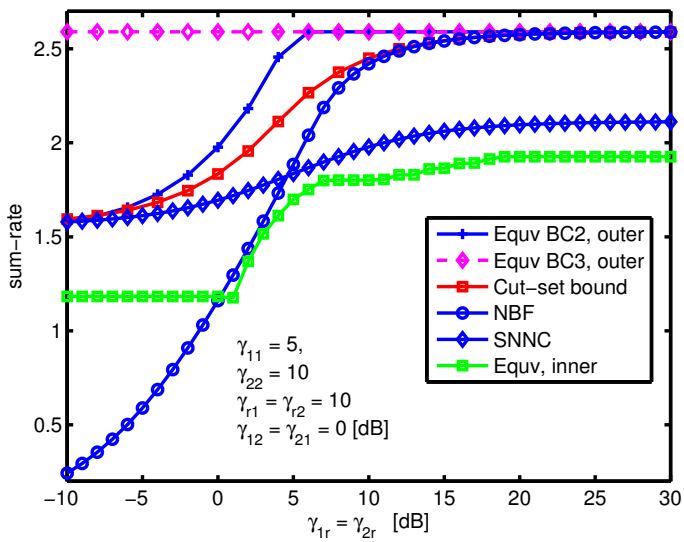


Figure 6.14. Capacity bounds on sum rate for high-rate backhaul scenario with varying source-relay link quality. $\gamma_{11} = 5$dB, $\gamma_{22} = 10$dB, $\gamma_{r1} = \gamma_{r2} = 10$dB, and $\gamma_{12} = \gamma_{21} = 0$dB.

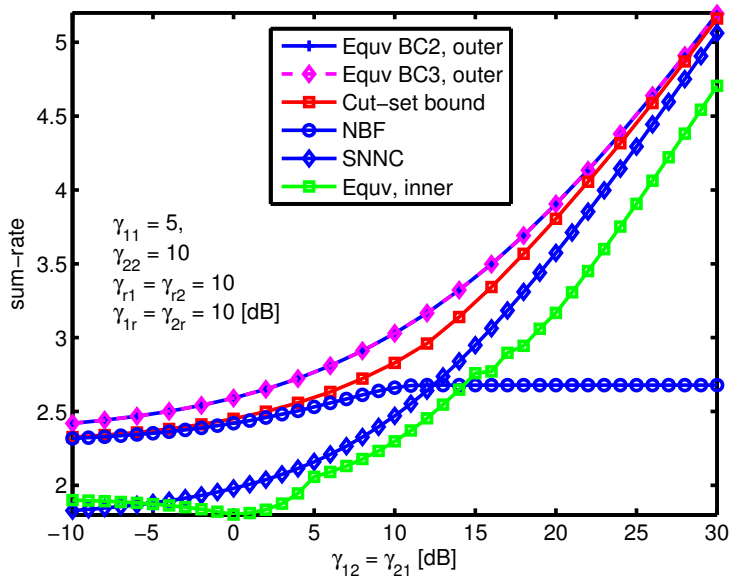


Figure 6.15. Capacity bounds on sum rate for high-rate backhaul scenario with varying cross-link channel quality. $\gamma_{11} = 5$dB, $\gamma_{22} = 10$dB, $\gamma_{1r} = \gamma_{2r} = 10$dB, and $\gamma_{r1} = \gamma_{r2} = 10$dB.

the capacity region obtained from the bounding networks, and compare them to the genie-aided outer bound proposed in Proposition 3.2, the pDF+LNC inner bound developed in Section 4.2.1, and the AF+ANC inner bound developed in Section 5.3. When the relay-destination links and the cross links are weak, both the outer bounds (I) and (II) are within 0.2 bits from the genie-aided outer bound on sum rate, but deviate from it on individual rates, as shown in Figure 6.16. The inner bound is worse than the pDF+LNC bound except for the rate constraint on R_2. The inner bound obtained from the bounding network can beat the AF+ANC region except for the sum rate constraint. The gap between the inner bound and the genie-aided outer bound varies from 0.2 bits to 0.4 bits per channel use.

When the source-relay links and the cross links are weak, as shown in Figure 6.17, the outer bound (I) is very close to the genie-aided outer bound on individual rates, but deviates by 0.5 bits on the sum rate. The outer bound (II) is relatively loose. The inner bound obtained from the bounding network can almost beat the pDF+LNC inner bound except on the sum rate, and it deviates from the genie-aided outer bound by at most 0.3 bits per channel use.

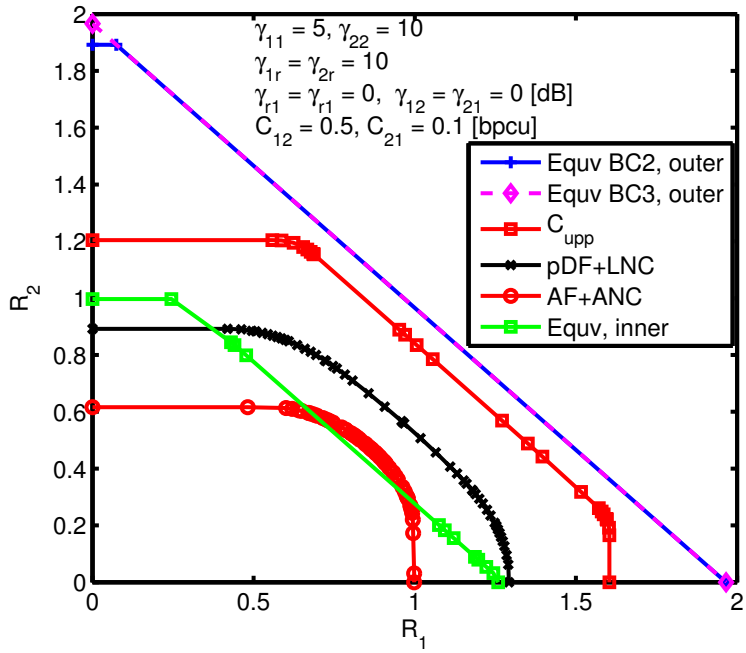


Figure 6.16. Rate regions and capacity outer bounds for channels setups with direct links $\gamma_{11} = 5$dB and $\gamma_{22} = 10$dB, cross-links $\gamma_{12} = \gamma_{21} = 0$dB, source-relay links $\gamma_{1r} = \gamma_{2r} = 10$dB and relay-destination links $\gamma_{r1} = \gamma_{r2} = 0$dB, with asymmetric backhaul rates $C_{12} = 0.5$, $C_{21} = 0.1$ bits per channel use.

6.4 Summary

In this chapter we have presented general capacity bounding models for wireless networks with independent noise, by construction of upper and lower capacity bounding networks consisting of only noiseless bit-pipe channels. We have extended the bounding models for two-user broadcast channels to many-user scenarios and established the gap between upper and lower bounding models. For networks with coupled links, we have proposed a channel decoupling method which can decompose the network into overlapping multiple-access and broadcast channels. We have created an upper bounding network consisting of only bit-pipe channels by applying the one-shot models directly to the decomposed network. When developing the lower bounding network, we have proposed to update the models for coupled broadcast and multiple-access channels. We have demonstrated by some examples that the resulting upper bound is in general very good and the gap between the upper and lower bounds is usually not large.

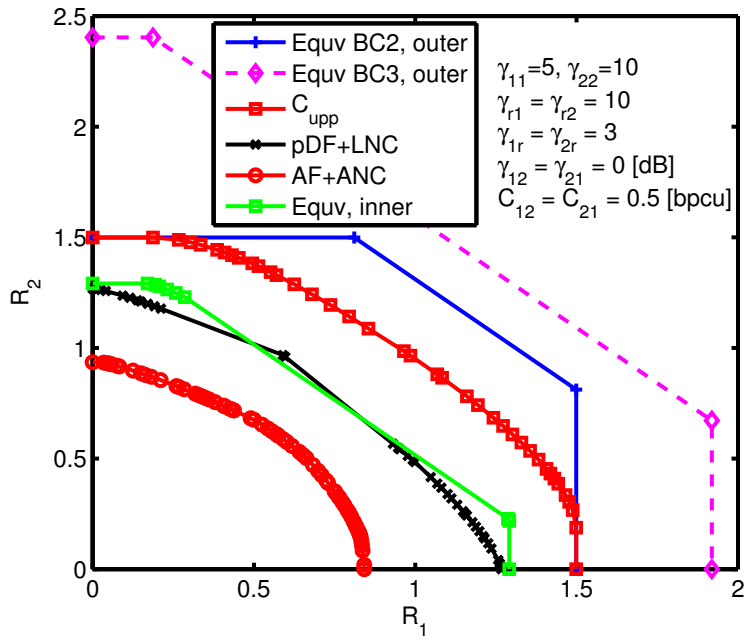


Figure 6.17. Rate regions and capacity outer bounds for channels setups with direct links $\gamma_{11} = 5$dB and $\gamma_{22} = 10$dB, cross-links $\gamma_{12} = \gamma_{21} = 0$dB, relay-destination links $\gamma_{r1} = \gamma_{r2} = 10$dB, and source-relay links $\gamma_{1r} = \gamma_{2r} = 3$dB, with asymmetric backhaul rates $C_{12} = C_{21} = 0.5$ bits per channel use.

Chapter 7

Cooperation by Cancelling Interference at the Transmitter

The capacity bounds established in Chapter 3, 4, 5 are based on arguments involving an infinite number of channel uses, i.e., the transmit signals are of infinite dimensions. To obtain an understanding of what one can achieve in small (or a single) dimensions of signals and at low complexity, we consider a wireless downlink transmission scenario where practical Costa precoding [Cos83] (also known as dirty paper coding) is utilized at the base station to facilitate broadcast transmission to two receivers with the aid from a half-duplex relay node. Specifically, we propose a symbol-by-symbol scheme for the Costa precoding problem. For finite-alphabet signaling and interference, we derive the optimal (in terms of maximum mutual information) modulator under a given power constraint. A sub-optimal modulator is also proposed by formulating an optimization problem that maximizes the minimum distance of the signal constellation, and this non-convex optimization problem is approximately solved by semi-definite relaxation. For the case of binary signaling with binary interference, we obtain a closed-form solution for the sub-optimal modulator, which only suffers little performance degradation compared to the optimal modulator in the region of interest. For more general signal constellations and more general interference distributions, we propose an optimized Tomlinson-Harashima precoder (THP) [Tom71, HM72], which uniformly outperforms conventional THP with heuristic parameters. Bit-level simulation shows that the optimal and sub-optimal modulators can achieve significant gains over the THP benchmark as well as over non-Costa reference schemes, especially when the power of the interference is larger than the power of the noise.

7.1 Cancelling Known Interference in Relay Networks

From information theory, it is known since [Cos83] that the achievable rate of a communication channel remains unchanged if the receiver observes the transmitted

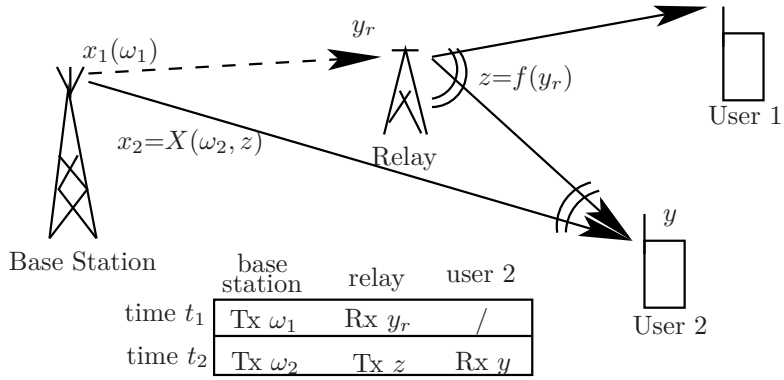


Figure 7.1. The base station transmits ω_1 to user 1 during time slot t_1 and ω_2 to user 2 during time slot t_2. The relaying signal $z{=}f(y_r)$ dedicated for user 1 appears as "interference" for user 2. With non-causal knowledge of z, the base station can design a DPC modulator $x_2 = X(\omega_2, z)$ given the information symbol ω_2 and the interference z.

signal in the presence of additive interference and white Gaussian noise, *provided that the transmitter knows the interference non-causally.* The resulting precoding method is known as "Costa precoding" or "dirty paper coding" (DPC) after the title of [Cos83]. The problem of designing a DPC transmitter is important because the scenario with known interference arises in many contexts, notably, in precoding for inter-symbol interference channels and for the downlink multiuser MIMO wireless channel [TV05, CS03, WSS06, EU10]. In [WSS06] DPC has been shown to be capacity achieving in non-degraded MIMO broadcast channels. DPC can also be applied in a cooperative two-transmitter two-receiver wireless network [NJGM07], in relay-aided broadcast channels [KS09], and in relay interference channels with a cognitive source [ZV09]. Essentially, an information theoretic strategy for achieving capacity is known; it is precisely the achievability proof in [Cos83] and works as follows: First quantize the interference into a number of bins and then, depending on what bin the interference falls into, choose an appropriate code to encode the message at the transmitter. This approach has been used with success in [BBCS06, EtB05], for example, where sophisticated coding schemes were proposed based on superposition coding [BBCS06], lattices and trellis shaping [EtB05]. Trellis and convolutional precoding was used in [YVC05] where the trellis shaping was developed taking into account the knowledge of a non-causal interference sequence.

In this work we study practical DPC schemes in the context of a relay-aided downlink channel. Consider a communication network where the base station transmits information symbols ω_1 and ω_2 to user 1 and user 2, respectively, with the aid of a half-duplex relay (a relay that cannot transmit and receive simultaneously). As illustrated in Figure 7.1, the relay is dedicated to assist user 1 (the weaker/more

distant user) whose direct link with the source fails. The base station transmits x_1 (signal for ω_1) during time slot t_1 and x_2 (signal for ω_2) during t_2. The relay listens to the base station during t_1 and transmits $z = f(y_r)$ during t_2, where y_r is the received signal at the relay during t_1 and $f(\cdot)$ is the relay mapping function. The relaying signal z, which is useful for user 1, appears as interference for user 2. Assuming that the relay function $f(\cdot)$ is known at the base station and that the source-relay link is good, the "interference" z will be known at the base station with high probability before its transmission by the relay node, which effectively results in the Costa problem.

The goal of our work is to obtain an understanding of what one can achieve in small (or a single) dimensions of signals and at low complexity, rather than to achieve the channel capacity. Indeed achieving capacity requires coding over an infinite number of dimensions, as in [Cos83]. More precisely, we consider the design of *one-dimensional*[1] schemes $X(\omega_2, z)$ that map the information symbol $\omega_2 \in \mathbb{Z}$ and an interference symbol $z \in \mathbb{R}$ (known to the base station but *not* to user 2) onto an output symbol $x_2 \in \mathbb{R}$. Thereby, our focus is on symbol-by-symbol modulation rather than on coding. To get a better understanding of how our proposed scheme performs compared to the theoretical limit, we will use the mutual information between the transmitted ω_2 and the received signal at user 2 as the criterion for design.

The well known Tomlinson-Harashima precoder (THP) [Tom71, HM72], originally proposed for channels with inter-symbol interference, is a symbol-by-symbol DPC approach and therefore it serves as a good benchmark. The achievable rate for THP has been investigated in [WC98] and a scaled THP has been invented in [ZSE02]. THP with partial channel knowledge has been studied in [Lia05]. Essentially THP (and its variations) subtracts the interference z from the information-bearing symbol and then performs a modulo operation to avoid a power boost. Another reason for introducing THP is that it already has wide applications. For instance, THP has been proposed as a building block for transmit precoding for the downlink multiuser MIMO channel [AAFS04, WFVH04]. Another symbol-by-symbol DPC scheme proposed in [SL08] minimizes the uncoded symbol error probability by joint design of the modulator and the demodulator. It is omitted here due to the difficulty to evaluate its performance in terms of mutual information.

For binary signaling with binary interference, we present the optimal[2] modulator based on an exhaustive search over all 12 possible mappings, which typically outperforms THP even when the parameters of THP are optimally chosen. To make our modulation design strategy applicable to higher order modulation, we propose

[1]Extension to inphase/quadrature (narrowband) modulation, or to other orthogonal multiplexing formats is immediate by treating each dimension independently. However, we do not claim optimality for such extensions as there exist some non-symmetric constellations that might offer better rates than symmetric ones. One such example can be found in [KAPT09] where a 5-QAM constellation is used during the broadcast phase of a two-way relay channel.

[2]Throughout we use "optimal" in the sense of *maximum mutual information* or *minimum error probability*. When not explicitly stated, we refer jointly to both these criteria.

a mapping set size reduction method that makes the exhaustive search method applicable. We also propose a sub-optimal modulator by formulating an optimization problem targeted at maximizing the minimum constellation distance, which is approximately solved by convex optimization after relaxation. A closed-form solution of the sub-optimal modulator is obtained for the case of binary signaling with binary interference, which suffers a minor performance loss compared to the optimal modulator in most of the interesting scenarios. For arbitrary signal and interference distributions, we propose an optimized THP scheme which demonstrates significant gains over heuristic THP in strong and medium interference scenarios. Our proposed DPC schemes are evaluated in terms of mutual information, coded bit-error-rate (BER), as well as energy efficiency, and compared to two non-DPC approaches, namely orthogonal transmission and receiver centric interference cancellation.

The rest of this chapter is organized as follows. The system model and design criteria are introduced in Section 7.2, where a brief overview of THP is also presented. The optimal modulator for finite alphabet signaling is discussed in Section 7.3, and the sub-optimal modulator by maximizing the minimum distance among constellation points is discussed in Section 7.4. For general signaling, THP with optimized parameters for Gaussian interference is presented in Section 7.5. Two non-DPC schemes are discussed in Section 7.6 as a reference. Simulation results are presented in Section 7.7 and conclusions are drawn in Section 7.8.

7.2 System Model and Tomlinson-Harashima Precoder

From now on, we consider an one-dimensional Gaussian channel, and all quantities are real-valued and scalar. As shown in Figure 7.1, the base station transmits $x_1(\omega_1)$ during time slot t_1 and the relay receives

$$y_r = x_1(\omega_1) + n_r,$$

where n_r is noise with variance σ_r^2, and generates the relaying signal $z = f(y_r)$ dedicated for user 1. During time slot t_2, the base station transmits x_2 to user 2 and the relay transmits z to user 1 through the same channel. Therefore the received signal at user 2 in t_2 can be written as

$$y = x_2 + z + n, \tag{7.1}$$

where n is noise. The design of the optimal relay mapping function $f(\cdot)$ is interesting and challenging, as discussed in [GJ07, KL08]. For example, we can choose the memoryless relaying function proposed in [GJ07] to maximize the generalized signal-to-noise power ratio (GSNR) at user 1, or utilize the constellation rearrangement proposed in [KL08] to maximize the rate for user 1 if its direct link with the source exists. The joint optimization of the relay function and the modulator in the base station is rather complicated. To simplify the analysis and highlight the insights

gained in this paper, hereafter we assume a perfect source-relay link[3] in Figure 7.1 with a deterministic relay mapping $z = \sqrt{\frac{P_r}{P_x}} x_1$. The DPC modulator in the base station that we envision maps an information symbol ω_2 from an M-ary alphabet ($\omega_2 \in \{0, \ldots, M-1\}$), and the interfering relay symbol $z \in \mathbb{R}$, onto a modulated symbol $x_2 \in \mathbb{R}$, through the (nonlinear) modulator mapping as follows

$$x_2 = X(\omega_2, z).$$

User 2 does not know z, but we shall assume that it knows the probability distribution of z, say $p_z(u)$. This assumption is weak if z is drawn from a stationary and ergodic process, because then the base station can provide information about $p_z(u)$ to user 2. We assume that the noise is Gaussian: $n \sim \mathcal{N}(0, \sigma^2)$ where σ^2 is known. Furthermore, we assume that the available average transmit power is fixed to a constant P_x. With the optimization criterion of the mutual information $I(y; \omega_2)$, the problem is then to find the best possible mapping $x_2 = X(\omega_2, z)$ that maximizes $I(y; \omega_2)$, i.e.,

$$X(\omega_2, z) = \arg \max_{X:\ E[x_2^2] \le P_x} I(y; \omega_2), \tag{7.2}$$

where

$$\begin{aligned} I(y;\omega_2) &= H(\omega_2) - H(\omega_2|y) \\ &= \sum_{\omega_2=0}^{M-1} \int_{-\infty}^{\infty} p_y(y, \omega_2) \log(p(\omega_2|y)) dy - \sum_{\omega_2=0}^{M-1} p(\omega_2) \log(p(\omega_2)) \\ &= \sum_{\omega_2=0}^{M-1} \left[\int_{-\infty}^{\infty} p_y(y, \omega_2) \log\left(\frac{p_y(y, \omega_2)}{p_y(y)} \right) dy - p(\omega_2) \log(p(\omega_2)) \right] \\ &= \sum_{\omega_2=0}^{M-1} p(\omega_2) \int_{-\infty}^{\infty} p_y(y|\omega_2) \log\left(\frac{p_y(y|\omega_2)}{\sum_{\omega_2'=0}^{M-1} p_y(y|\omega_2') P(\omega_2')} \right) dy. \end{aligned} \tag{7.3}$$

The last equality comes from the fact that

$$\int_{-\infty}^{\infty} p_y(y|\omega_2) dy = 1, \forall \omega_2.$$

In practice, $I(y; \omega_2)$ can be easily computed by Monte-Carlo integration. Naturally $p_y(y|\omega_2)$ (and $I(y; \omega_2)$) depends on both the specific modulator mapping $X(\omega_2, z)$ and the distribution $p_z(u)$.

7.2.1 Tomlinson-Harashima Precoding (THP)

THP is the best known available baseline for comparison and therefore we outline its principle here. THP first maps ω_2 onto a constellation point by modulating

[3]For $P_x \gg \sigma_r^2$, x_1 can be almost perfectly known/estimated at the relay.

it via $x(\omega_2)$, and then subtracts the interference z from it. A modulo operation $\mathrm{mod}(\cdot, \Lambda)$ is then carried out so that the resulting transmitted signal falls into the region $[-\Lambda/2, \Lambda/2]$. Therefore we have

$$\begin{aligned} x_2 &= X(\omega_2, z) = \mathrm{mod}(x(\omega_2) - z, \Lambda), \\ y &= \mathrm{mod}(x(\omega_2) - z, \Lambda) + z + n \\ &= x(\omega_2) + k\Lambda + n \\ &= x(\omega_2) + e, \end{aligned}$$

where k is an integer which depends both on ω_2 and z. Note that the equivalent noise term

$$e = k\Lambda + n$$

also depends on ω_2. In papers dealing with THP, the following heuristic (and suboptimal) detector is usually used:

$$\hat{\omega_2}_{\mathrm{subopt}} = \underset{\omega_2}{\mathrm{argmin}} \, |\mathrm{mod}(y, \Lambda) - x(\omega_2)|.$$

To find the minimum error-probability receiver for THP, first note that

$$p_y(y|\omega_2) = \sum_{k=-\infty}^{\infty} p(k|\omega_2) p_n(y - x(\omega_2) - k\Lambda),$$

where the integer k is random with the following conditional distribution:

$$\begin{aligned} p(k|\omega_2) &= p\left((x(\omega_2) - z) \in \left[-(k + \frac{1}{2})\Lambda, -(k - \frac{1}{2})\Lambda \right] \, |\omega_2 \right) \\ &= p\left(x(\omega_2) + (k - \frac{1}{2})\Lambda \le z \le x(\omega_2) + (k + \frac{1}{2})\Lambda \, |\omega_2 \right) \\ &= p\left(z \le x(\omega_2) + (k + \frac{1}{2})\Lambda \, |\omega_2 \right) - p\left(z \le x(\omega_2) + (k - \frac{1}{2})\Lambda \, |\omega_2 \right) \\ &= F_z\left(x(\omega_2) + (k + \frac{1}{2})\Lambda \right) - F_z\left(x(\omega_2) + (k - \frac{1}{2})\Lambda \right). \end{aligned} \tag{7.4}$$

In (7.4), $F_z(t) = \int_{-\infty}^{t} p_z(u)du$ is the cumulative distribution function of z. The maximum a posteriori (MAP) receiver finds the most likely ω_2 when y is received:

$$\begin{aligned} \hat{\omega_2}_{\mathrm{MAP}} &= \underset{\omega_2}{\mathrm{argmax}} \, p(\omega_2|y) = \underset{\omega_2}{\mathrm{argmax}} \, p_y(y|\omega_2) \\ &= \underset{\omega_2}{\mathrm{argmax}} \sum_{k=-\infty}^{\infty} p(k|\omega_2) \exp\left(-\frac{(y - x(\omega_2) - k\Lambda)^2}{2\sigma^2} \right), \end{aligned} \tag{7.5}$$

where the second equality comes from the assumption of equally probable ω_2. In practice the sum in (7.5) can be truncated to a few terms since $p(k|\omega_2)$ decreases

rapidly (exponentially if z is Gaussian) as $|k|$ increases. The difference in performance between the two receivers, however, is usually small except for "unlucky" choices of the mapping $x(\omega_2)$ and Λ, i.e., when $p(k \neq k_0|\omega_2)$ is significant where k_0 satisfies

$$\mathrm{mod}(y, \Lambda) = y - k_0\Lambda.$$

7.3 Optimal Modulator Design

In this section we first find the optimal mapping modulator for binary signaling with binary interference and then generalize it to higher order modulations.

7.3.1 Optimal Mapping for Binary Signaling with Binary Interference

For discrete, binary random variables ω_2 and z (over $\mathbb{Z}$ and $\mathbb{R}$, respectively), we assume that

$$\begin{aligned} p(\omega_2 = 0) = p(\omega_2 = 1) &= \frac{1}{2}, \\ p(z = -\beta) = p(z = \beta) &= \frac{1}{2}. \end{aligned} \tag{7.6}$$

That is, the input alphabet is binary ($\omega_2 = 0, 1$) and the interference comes from a scaled BPSK constellation $z = \pm\beta$. Also, ω_2 and z are independent and all combinations of (ω_2, z) are equally likely. Therefore the mapping $X(\omega_2, z)$ can be explicitly written as

$$\begin{aligned} X(\omega_2 = 0, z = -\beta) &\triangleq s_0, \quad X(\omega_2 = 0, z = \beta) \triangleq s_1, \\ X(\omega_2 = 1, z = -\beta) &\triangleq s_2, \quad X(\omega_2 = 1, z = \beta) \triangleq s_3. \end{aligned} \tag{7.7}$$

By symmetry (ω_2 and z have symmetric probability densities), we must have $x \in \{-a, -b, b, a\}$ for some positive constants a, b. The problem is then to find suitable (a, b) and to map $s_0, ..., s_3$ onto the set $\{-a, -b, b, a\}$ such that $I(y; \omega_2)$ stated in (7.3) is maximized. Note that

$$p_y(y|\omega_2) = \sum_{z=\pm\beta} p_{y,z}(y, z|\omega_2) = \sum_{z=\pm\beta} p_y(y|\omega_2, z)p(z), \tag{7.8}$$

where

$$p_y(y|\omega_2, z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - z - X(\omega_2, z))^2}{2\sigma^2}\right).$$

There are $4! = 24$ permutations of the elements in $\{-a, -b, b, a\}$, of which 12 are redundant (a and b are not ordered). The set of all possible mappings (s_0, s_1, s_2, s_3) to be considered are:

$$\begin{array}{lll} \text{(I) } (a, -a, b, -b); & \text{(II) } (a, -b, b, -a); & \text{(III) } (-a, -b, b, a); \\ \text{(IV) } (-a, -b, a, b); & \text{(V) } (-a, b, a, -b); & \text{(VI) } (-a, a, b, -b); \\ \text{(VII) } (-a, a, -b, b); & \text{(VIII) } (-a, b, -b, a); & \text{(IX) } (a, b, -b, -a); \\ \text{(X) } (a, b, -a, -b); & \text{(XI) } (a, -b, -a, b); & \text{(XII) } (a, -a, -b, b). \end{array} \tag{7.9}$$

The mapping $X(\omega_2, z)$ is a deterministic function that assigns one of the values $\{-a, -b, b, a\}$ to x_2 for each possible pair (ω_2, z). Since the variables ω_2 and z are independent and equiprobable (see (7.6)), it follows that all four possibilities for x_2, viz. $x_2 \in \{-a, -b, b, a\}$ are equally likely. Thus the power constraint translates into

$$E[x_2^2] = (a^2 + b^2)/2 \leq P_x.$$

A straightforward approach, as stated in our preliminary work [DLS06], is to perform an exhaustive search over a fine grid which contains all (a, b) that satisfy this constraint. And for each (a, b) we examine all the 12 mappings to identify the optimal modulation which generates the highest mutual information. This optimization process can be carried out off-line and the result can be stored in a look-up table (indexed by P_x/σ^2 and β^2/σ^2) with resolution as required.

The minimum error-probability receiver for the optimal (maximum mutual information) modulator has a rather simple form. To write it out explicitly, note from (7.8) that

$$\hat{\omega_2}_{\text{MAP}} = \underset{\omega_2}{\text{argmax}} \sum_{z=\pm\beta} \exp\left(-\frac{(y - z - X(\omega_2, z))^2}{2\sigma^2}\right).$$

When the assumption of a perfect source-relay link does not hold, i.e., when z is not perfectly known at the relay, the conditional probability $p_y(y|\omega_2)$ must be adjusted to reflect the reliability of z. Given the transmit power P_x and source-relay link noise power σ_r^2, the conditional probability (7.8) should be rewritten as

$$p_y(y|\omega_2) = (1 - P_e) \sum_{z=\pm\beta} p_y(y|\omega_2, z)p(z) + P_e \sum_{z=\pm\beta} p_y(y|\omega_2, -z)p(z),$$

where $P_e = Q(\sqrt{P_x/\sigma_r^2})$ is the error probability of detecting the BPSK modulated ω_1 (hence z).

7.3.2 Extension to Higher Order Modulation

Despite the fact that the optimization can be done off-line, it is not directly feasible to extend the exhaustive search method proposed in Section 7.3.1 to higher-order modulation since the number of possible mappings increases explosively with the order of the modulation. For M-PAM signal with N-PAM interference, in total we have MN combinations for (ω_2, z) and therefore the same number of possible $X(\omega_2, z)$ values. Their amplitudes are symmetric in the real field $\mathbb{R}$ around the origin and hence at most half of them, i.e. $MN/2$, are free to choose under the power constraint. Besides, there are in total $(MN)!$ permutations of the set of MN parameters. Since $MN/2$ of these parameters have no ordering constraint, the number of all possible mappings is $\frac{(MN)!}{(MN/2)!}$. And then for each of these mappings, we still have to do an exhaustive grid search along $MN/2$ dimensions to find the optimal modulator for a particular combination of P_z/σ^2 and P_x/σ^2. For

example, in the case of 4-PAM signaling with BPSK interference, there are in total $8!/4! = 1680$ different mappings and we have to do an exhaustive grid search over 4 dimensions. Therefore for higher-order modulation, the number of candidate mappings can become prohibitively large and makes the off-line exhaustive search computationally impractical. In what follows we will present a method which can greatly reduce the number of mappings.

We start with the special case with binary signaling and binary interference, as stated in (7.6). By comparing all the mappings in (7.9), we come up with the following observations:

1. Two mappings are said to be equivalent if one can be obtained from the other by exchanging $X(0, z)$ and $X(1, z)$ for all z;

2. Mappings satisfying $X(0, z)X(1, z) > 0$ will result in smaller distance between $\omega = 0$ and $\omega = 1$ in the received signal constellations, and therefore should not be considered;

3. Mappings should satisfy $|X(0, z) - X(1, z)| = |X(0, -z) - X(1, -z)|$.

All the equivalent mappings defined by *Observation* 1) are identical in the sense that a and b are commutable, and therefore we will group them together in a pair of parenthesis. For example, we group the following pairs of equivalent mappings together: (III, IX), (IV, X), (V, XI), and (VI, XII). By applying *Observation* 2), mappings I, II, VII, VIII are excluded. By applying *Observation* 3), mappings (IV, X) and (V, XI) are also excluded. Now we only have ***two*** groups left: (III, IX) and (VI, XII). We then search over a fine grid which contains all (a, b) that satisfy the power constraint, and for each (a, b) we only examine the above mentioned ***two*** mappings (one element from each group, say IX and XII) instead of ***twelve*** as in Section 7.3.1.

For the general cases with uniformly distributed information symbols $\omega \in \{0, \ldots, M-1\}$ and uniformly distributed interference $z \in \{z_0, \ldots, z_{N-1}\}$ with N-PAM modulation, by defining the modulation vector associated with ω as

$$\mathcal{X}(\omega) \triangleq [X(\omega, z)|\forall z] = [X(\omega, z_0), \ldots, X(\omega, z_{N-1})], \tag{7.10}$$

the following principles can be applied to reduce the number of mapping candidates:

1. Mappings $X_1(\omega, z)$ and $X_2(\omega, z)$ are equivalent if they have the same vector set[4], i.e., $\{\mathcal{X}_1(\omega)|\forall\omega\} = \{\mathcal{X}_2(\omega)|\forall\omega\}$, where $\mathcal{X}_i(\omega)$, $i = 1, 2$ is defined as in (7.10);

2. For $\omega_i \neq \omega_j$, the elements in received signal constellation subset $\{X(\omega_i, z) + z|\forall z\}$ should be separated as far as possible away from any elements in $\{X(\omega_j, z) + z|\forall z\}$;

[4]The assignment of each modulation vector in $\{\mathcal{X}(0), \ldots, \mathcal{X}(M-1)\}$ to an information symbol $w \in \{0, \ldots, M-1\}$ should not affect the achievable rate or symbol error probability.

3. For each interference pair $(z, -z)$, the subsets $\{X(\omega, z)+z|\forall\omega\}$ and $\{X(\omega, -z)-z|\forall\omega\}$ should be equivalent in the sense that they are symmetric with respect to the origin.

The equivalent mappings defined by the first principle will be grouped together and all the mappings that do not follow the second and the third principles will be deemed "unfavorable" and therefore will be dropped. For example, by applying the above principles, the number of mappings for 4-PAM signaling with BPSK interference can be reduced from 1680 down to 133.

7.4 Sub-optimal Modulator Design via Optimization

A suboptimal modulator by maximizing the minimum distance among constellation points is proposed by formulating an optimization problem.

As stated in Section 7.3.2, the off-line optimization can be greatly simplified by reducing the number of mappings. However, since the optimization in (7.2) is non-convex, the complexity of a grid search will increase exponentially with the number of searching dimensions. Therefore we propose here a low-complexity sub-optimal modulator based on the criterion of maximized minimum distances among the constellation points.

For uniformly distributed information symbols $\omega \in \{0, \ldots, M-1\}$ and uniformly distributed interference z with N-PAM modulation, the distance between received signal constellation points for $\omega_i \neq \omega_j$ (omitting the noise term for simplicity) can be classified into two types

$$\begin{aligned} d_{\mathrm{I}} &= |X(\omega_i, z_m) + z_m - (X(\omega_j, z_m) + z_m)| \\ &= |X(\omega_i, z_m) - X(\omega_j, z_m)|, \\ d_{\mathrm{II}} &= |X(\omega_i, z_m) + z_m - (X(\omega_j, z_n) + z_n)| \\ &= |X(\omega_i, z_m) - X(\omega_j, z_n) + (z_m - z_n)|, \end{aligned} \tag{7.11}$$

where z_m, z_n are interference symbols.

There are in total $N_{\mathrm{I}} = \frac{M(M-1)N}{2}$ type I distances d_{I} and $N_{\mathrm{II}} = \frac{M(M-1)N(N-1)}{2}$ type II distances d_{II}. By reformulating the mapping $X(\omega, z)$ into a vector

$$\boldsymbol{x} = [X(0, z_0), ..., X(0, z_{N-1}), X(1, z_0), ..., X(M-1, z_{N-1})],$$

and denoting $\boldsymbol{x}(n)$ as the nth element of $\boldsymbol{x}$, we can rewrite (7.11) as follows

$$\begin{aligned} &d^2_{\mathrm{I},k} = (\boldsymbol{x}(i_k) - \boldsymbol{x}(j_k))^2 = \boldsymbol{x}\boldsymbol{A}_k\boldsymbol{x}^T, \\ &d^2_{\mathrm{II},l} = (\boldsymbol{x}(i_l) - \boldsymbol{x}(j_l) + \eta_l)^2 = \boldsymbol{x}\boldsymbol{B}_l\boldsymbol{x}^T + 2\boldsymbol{x}\boldsymbol{b}_l^T + \eta_l^2, \\ &i_k, j_k, i_l, j_l \in \{1, ..., MN\},\ k = 1, ..., N_{\mathrm{I}},\ l = 1, ..., N_{\mathrm{II}}, \end{aligned} \tag{7.12}$$

where $\boldsymbol{b}_l$ are $1 \times MN$ sparse vectors each with only two non-zeros elements $\boldsymbol{b}_l(i_l) = \eta_l$ and $\boldsymbol{b}_l(j_l) = -\eta_l$, and $\boldsymbol{A}_k$ ($\boldsymbol{B}_l$) are $MN \times MN$ sparse symmetric matrices each

with only four non-zero elements placed in their diagonal and anti-diagonal positions defined by i_k, j_k (i_l, j_l), i.e.,

$$\boldsymbol{A}_k \text{ or } \boldsymbol{B}_l = \begin{bmatrix} . & . & . & . & . \\ . & 1 & . & -1 & . \\ . & . & \ddots & . & . \\ . & -1 & . & 1 & . \\ . & . & . & . & . \end{bmatrix}, \ \boldsymbol{b}_l^T = \begin{bmatrix} . \\ \eta_l \\ \vdots \\ -\eta_l \\ . \end{bmatrix}. \tag{7.13}$$

The sub-optimal modulator can therefore be formulated based on (7.12) as an inhomogeneous quadratically-constrained quadratic program (QCQP) [BV04] problem,

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{R}^{MN}} \min_{\substack{k=1,\ldots,N_{\mathrm{I}} \\ l=1,\ldots,N_{\mathrm{II}}}} & \{\boldsymbol{x}\boldsymbol{A}_k\boldsymbol{x}^T, \boldsymbol{x}\boldsymbol{B}_l\boldsymbol{x}^T + 2\boldsymbol{x}\boldsymbol{b}_l^T + \eta_l^2\} \\ \text{subject to} \ & \boldsymbol{x}\boldsymbol{x}^T \leq MN \cdot P_x. \end{aligned} \tag{7.14}$$

The solution of (7.14) will yield constellations with large mutual information, since a constellation that offers a large constellation-constraint mutual information also has a large minimum distance. By introducing

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}^T \\ 1 \end{bmatrix} [\boldsymbol{x}, \ 1] = \begin{bmatrix} \boldsymbol{x}^T\boldsymbol{x} & \boldsymbol{x}^T \\ \boldsymbol{x} & 1 \end{bmatrix},$$

and some new matrices

$$\tilde{\boldsymbol{A}}_k = \begin{bmatrix} \boldsymbol{A}_k & 0 \\ 0 & 0 \end{bmatrix}, \ \tilde{\boldsymbol{B}}_l = \begin{bmatrix} \boldsymbol{B}_l & \boldsymbol{b}_l^T \\ \boldsymbol{b}_l & \eta_l^2 \end{bmatrix}, \ \boldsymbol{C} = \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & 0 \end{bmatrix},$$

we can reformulate (7.14) in the homogenous format [BV04] as follows

$$\begin{aligned} \max_{\boldsymbol{X}\in\mathbb{S}^{MN+1}} \ & t \\ \text{subject to} \ & \mathrm{Tr}(\tilde{\boldsymbol{A}}_k\boldsymbol{X}) \geq t, \ k = 1, ..., N_{\mathrm{I}}, \\ & \mathrm{Tr}(\tilde{\boldsymbol{B}}_l\boldsymbol{X}) \geq t, \ l = 1, ..., N_{\mathrm{II}}, \\ & \mathrm{Tr}(\boldsymbol{C}\boldsymbol{X}) \leq MN \cdot P_x, \ \boldsymbol{X} \succeq \underline{0}, \\ & \mathrm{rank}(\boldsymbol{X}) = 1, \\ & \boldsymbol{X}(MN+1, MN+1) = 1, \end{aligned} \tag{7.15}$$

where $\mathbb{S}^n$ denotes the set of $n \times n$ symmetric matrices However, the exact solution of (7.15) is hard to find since the rank-1 constraint makes this problem non-convex. By semi-definite relaxation (SDR) approximation [LMS+10], we can obtain the

following relaxed version of (7.15):

$$\begin{aligned} \max_{\boldsymbol{X}\in\mathbb{S}^{MN+1}} \quad & t \\ \text{subject to} \quad & \mathrm{Tr}(\tilde{\boldsymbol{A}}_k\boldsymbol{X}) \geq t,\ k=1,...,N_{\mathrm{I}}, \\ & \mathrm{Tr}(\tilde{\boldsymbol{B}}_l\boldsymbol{X}) \geq t,\ l=1,...,N_{\mathrm{II}}, \\ & \mathrm{Tr}(\boldsymbol{C}\boldsymbol{X}) \leq MN\cdot P_x,\ \boldsymbol{X}\succeq \underline{0}, \\ & \boldsymbol{X}(MN+1,MN+1)=1. \end{aligned} \tag{7.16}$$

Since (7.16) is an instance of semi-definite programming [BV04], it can be solved in a numerically reliable and efficient fashion by convex optimization software, e.g. CVX [GB]. However, the globally optimal solution $\boldsymbol{X}^*$ to (7.16) in general has rank greater than 1, and therefore is not a feasible solution to the original problem stated in (7.15) and (7.14). We can extract from $\boldsymbol{X}^*$ a feasible (normally sub-optimal) solution $\boldsymbol{x}$ to (7.14) through randomization with provable approximation accuracy, see [LMS+10] and references therein for more details.

Note that (7.14) and (7.16) are actually a realization of the *Principle* 2) stated in Section 7.3.2. Besides, *Principle* 3) can also be utilized to add extra MN linear constraints to (7.14). Then following the same procedure of reformulation and relaxation, we can formulate a new optimization problem similar to (7.16). Detailed discussions on reformulation, relaxation, and approximation are omitted here.

For the special case of $M=N=2$, by confining ourselves to the selected mappings IX and XII in (7.9), we can solve (7.14) analytically (see Appendix 7.9.A for a detailed derivation), resulting in a closed-form solution for the modulation mapping $X(\omega_2, z)$ as follows:

$$\begin{cases} \text{XII}, & a=\sqrt{P_x},\ b=\sqrt{P_x}, & \text{if } P_x\leq\beta^2; \\ \text{XII}, & a=\beta,\ b=\sqrt{2P_x-\beta^2}, & \text{if } \beta^2<P_x<5\beta^2; \\ \text{IX}, & a=\sqrt{P_x-\beta^2}+\beta,\ b=\sqrt{P_x-\beta^2}-\beta, & \text{if } P_x\geq 5\beta^2. \end{cases} \tag{7.17}$$

This sub-optimal modulation can be carried out on-line given the instantaneous channel conditions.

7.5 Optimized THP for Arbitrary Signal and Interference

In Section 7.3 we have discussed the modulator design $x=X(\omega,z)$ given information symbols ω from an M-ary alphabet and an interference signal z modulated with N-PAM. We provided the optimal nonlinear mapping based on an exhaustive grid search, and a sub-optimal mapping based on convex optimization and relaxation. For an interference signal with a more general distribution (say Gaussian), however, it appears impractical (at least without approximations) to design the Costa modulator based the methods proposed in Section 7.3. The THP modulation, however, fits for arbitrary signal and interference constellations and therefore can be regarded as a good candidate for such scenarios. The advantage of staying

within the framework of THP is twofold. First, there are only two parameters to optimize over, as shown later in this section. Second, THP with heuristic parameter choices (which is commonly used in the literature) is known to provide significant gains over no-interference-cancellation.

Let α be half of the minimum distance between the uniformly distributed constellation points, i.e., $x \in \{-\alpha, \alpha\}$ for BPSK and $x \in \{-3\alpha, -\alpha, \alpha, 3\alpha\}$ for 4-PAM modulation, and let Λ be the parameter for modulo operations. Then THP modulation with transmit power constraint P_x has two parameters α, Λ to optimize over. THP with the heuristic parameter choice $\Lambda = 2M\alpha$ for M-PAM modulation appears to be customary and is the choice described in Chapter 10 of [TV05]. This method, referred as heuristic THP hereafter, is rather simple and can be used for general signal constellations and interference distributions. The actual value of Λ (therefore the value of α) is determined by the interference power P_z and the transmit power constraint P_x. For higher-order modulation with equiprobable information symbols, the resulting transmitted signal after modulo operation turns to be approximately uniformly distributed in the region of $[-\frac{\Lambda}{2}, \frac{\Lambda}{2}]$, resulting in a transmit power of $\frac{\Lambda^2}{12}$. Hence we have

$$\Lambda = \sqrt{12P_x}, \qquad \alpha = \Lambda/(2M). \tag{7.18}$$

However, when the modulation order is small, such as for a BPSK modulated signal, this approximation turns out to be biased. For the case of binary signaling with binary interference, the exact value of α (hence also Λ) can be determined as follows (see Appendix 7.9.B for a detailed derivation):

$$\begin{cases} \Lambda = 4\alpha; & \\ \alpha = \sqrt{P_x - \beta^2}, & \text{if } P_x \geq 2\beta^2; \\ \alpha = \frac{2\beta + \sqrt{5P_x - \beta^2}}{5}, & \text{if } \frac{\beta^2}{5} \leq P_x < 2\beta^2; \\ \alpha = \frac{2n\beta + \sqrt{(4n^2+1)P_x - \beta^2}}{4n^2+1}, & \text{if } \frac{\beta^2}{4n^2+1} \leq P_x < \frac{\beta^2}{4(n-1)^2+1}, \\ & \text{with integer } n \geq 2. \end{cases} \tag{7.19}$$

This heuristic parameter choice $\Lambda{=}2M\alpha$, however, appears "unlucky" in some specific situations. Therefore we propose to use optimized parameters α, Λ for THP. This optimization can be accomplished via a similar procedure as described in Section 7.3, i.e., performing a grid search over all Λ, α which satisfy the power constraint P_x. Unlike the optimal modulator where the search dimension increases with the modulation order, the optimization problem here is always two-dimensional. The search over all possible mappings is not necessary either.

7.6 Non-DPC Benchmarks

We present here two non-DPC approaches as a reference to evaluate our DPC schemes.

7.6.1 Relay Uses an Orthogonal Channel

The relay can use an orthogonal channel to help user 1 so that the relaying signal z will not interfere with the reception of x_2, under the same available resource (time, bandwidth, energy) constraints. We use time sharing between the relay and the base station over the same bandwidth to realize the orthogonal transmission. Let $\rho \in [0, 1]$ be the time sharing coefficient for the transmission from the base station to user 2, and assume that the transmitted signal $x_2 = X(\omega_2)$ is uniform M-PAM modulated and subject to the power constraint $E[x_2^2] = \frac{P_x}{\rho}$. Hence the total used energy $\rho \cdot E[x_2^2]$ remains the same as for the other schemes. The mutual information conveyed through this channel is

$$I_\rho(y; \omega_2) = \rho \cdot I(y; \omega_2), \tag{7.20}$$

where $I(y; \omega_2)$ is calculated according to (7.3), with

$$p_y(y|\omega_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - X(\omega_2))^2}{2\sigma^2}\right).$$

Note that ρ affects the throughput of both user 1 and user 2. Therefore, to choose ρ requires total throughput and fairness considerations. We will choose $\rho = \frac{1}{2}$ in our simulation for simplicity.

7.6.2 Interference Cancellation at the Receiver

One can also use no precoding at the base station but perform interference cancellation at user 2. When the interference is much stronger than the signal, user 2 can perform successive interference cancellation (SIC) [TV05]: It first decodes ω_1 treating $x_2 = X(\omega_2)$ as noise, and then subtracts the relaying signal $z(\omega_1)$ from y and uses the remaining signal to decode ω_2. But for moderate and weak interference, SIC will not work. We therefore propose here a new interference cancellation scheme which works for all cases by keeping user 2 receiving signals in both time slot t_1 and t_2. The received signals at user 2 during t_1 and t_2 can be written as

$$\begin{aligned} y_1 &= x_1(\omega_1) + n_1, \\ y &= X(\omega_2) + z(\omega_1) + n, \end{aligned} \tag{7.21}$$

where n_1 is additive white Gaussian noise and $x_1(\omega_1)$ is the signal for user 1 under average power constraint P_x. The mutual information between the transmitted information symbol ω_2 and the received signals (y_1, y) can therefore be written as

$$I(y_1, y; \omega_2) = \sum_{\omega_2} \int_{y_1} \int_{y} p_{y_1,y}(y_1, y|\omega_2) p(\omega_2) \log\left(\frac{p_{y_1,y}(y_1, y|\omega_2)}{\sum_{\omega_2'} p_{y_1,y}(y_1, y|\omega_2') p(\omega_2')}\right) dy dy_1,$$

(7.22)

where $p_{y_1,y}(y_1, y|\omega_2)$ can be obtained from the Bayes rule

$$\begin{aligned} p_{y_1,y}(y_1, y|\omega_2) &= \sum_{\omega_1} p_{y_1,y}(y_1, y|\omega_1, \omega_2)p(\omega_1) \\ &= \sum_{\omega_1} p_{y_1}(y_1|\omega_1)p_y(y|\omega_1, \omega_2)p(\omega_1). \end{aligned} \tag{7.23}$$

The second equality in (7.23) comes from the fact that y_1 and y are independent if ω_1 (and therefore x_1 and z) is known. We further get from (7.21) that

$$\begin{aligned} p_{y_1}(y_1|\omega_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_1 - x_1(\omega_1))^2}{2\sigma^2}\right), \\ p_y(y|\omega_1, \omega_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - X(\omega_2) - z(\omega_1))^2}{2\sigma^2}\right). \end{aligned}$$

The maximum a posteriori receiver given (y_1, y) is therefore

$$\hat{\omega_2}_{\text{MAP}} = \underset{\omega_2}{\text{argmax}}\, p(\omega_2|y_1, y) = \underset{\omega_2}{\text{argmax}}\, p_{y_1,y}(y_1, y|\omega_2), \tag{7.24}$$

where the second equality comes from the assumption of equally probable ω_2.

7.7 Numerical Results

In this section we will evaluate our proposed DPC schemes in terms of mutual information, coded BER and energy efficiency. Bit-level Monte-Carlo simulation is used to obtain the results.

7.7.1 Mutual Information

We first evaluate the performance in terms of mutual information and plot the achievable mutual information values as a function of the signal-to-noise ratio (SNR, P_x/σ^2) at a fixed interference-to-noise ratio (INR, P_z/σ^2) of 6 dB[5]. Strictly speaking, the actual SNR may be less than P_x/σ^2 because the optimal modulator does not necessarily use all available power. Yet we refer to P_x/σ^2 as SNR because this facilitates a well-defined comparison with the no-interference case.

In Figure 7.2 we compare mutual information for binary signaling with binary interference. The optimal modulator and the sub-optimal modulator are slightly worse compared to the no-interference case. THP with optimized parameters (α, Λ), which uniformly outperforms THP with the heuristic parameter choice $\Lambda = 4\alpha$, experiences notable degradation in low to medium SNR regions but converges to the optimal modulator at high SNR. Interference-cancellation performs well in high

[5]Here the INR value is heuristically chosen for illustration. Simulation results at other INR values will generate similar results, though the corresponding gains may vary slightly.

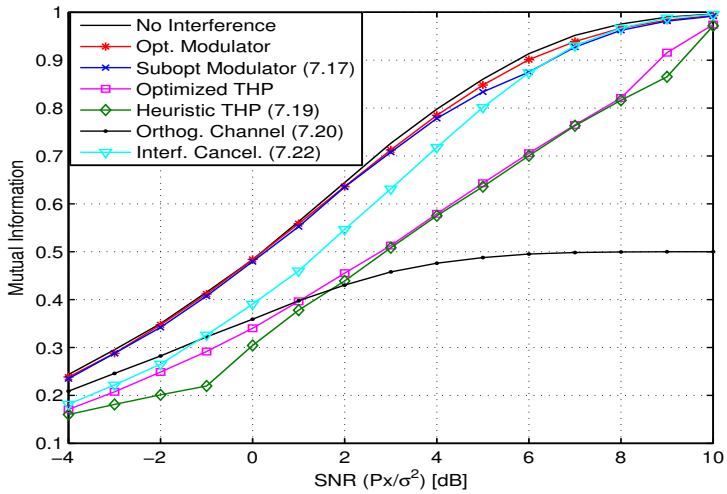


Figure 7.2. Mutual information for binary signaling in binary interference, with $\mathrm{INR}(P_z/\sigma^2) = 6$ dB.

SNR regions but suffers from the corrupted observation of the interference in low to medium SNR regions. By using orthogonal channels with time sharing $\rho = 1/2$, the performance is relatively good in low SNR regions where the benefits of excluding interference dominate (this is the power-limited regime). In medium and high SNR regions (i.e., in the bandwidth-limited regime), however, the penalty of shortening the transmission time (and therefore less channel use) becomes the bottleneck.

Note that although the mapping parameters for the optimal modulator and for the optimized THP might change with the resolution of the searching grid, the actual performance will only differ slightly.

In Figure 7.3 we present the case for quaternary signaling (4-PAM) with binary interference, focusing on the performance of the optimal modulator and the suboptimal modulator (7.16). The optimal modulator, in all SNR regions, suffers only a minor performance degradation compared to the no-interference case and achieves a significant gain (up to 3 dB at low SNR) over heuristic THP. The curve of the suboptimal modulator is not smooth due to the fact that it is an approximate solution of (7.14) based on relaxation and randomization, as discussed in Section 7.4. The performance of optimized THP, not shown here to improve the readability of this figure, lies in between the curves of the optimal modulator and the heuristic THP.

7.7.2 Coded Bit Error Rate

Next, we demonstrate that the gains predicted by calculating the mutual information actually do indicate what one can achieve in practice. Towards this end we

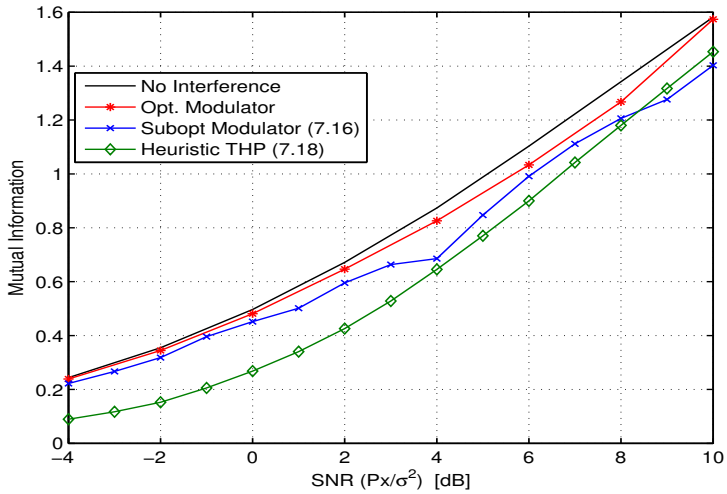


Figure 7.3. Mutual information for quaternary signaling in binary interference, with $\mathrm{INR}(P_z/\sigma^2) = 6$ dB.

think of the system in Figure 7.1 as an inner channel, and concatenate it with a rate-1/3 turbo code [BG96, EPG94] (memory length 2, generators 7_8 and 5_8, block length 1000, interleaver from the WCDMA standard, Max-LOG-MAP decoding, 8 iterations). When time sharing (with $\rho = 1/2$) is used, the encoded bit streams have to be punctured to rate 2/3, with the following puncturing patterns used at the two component encoders

$$\boldsymbol{P}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{P}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The decoding metrics were computed by evaluating $\log P(\omega_2|y)$ (for the interference cancellation scheme, $\log P(\omega_2|y_1, y)$ is used instead). More precisely, the log-likelihood ratios are used as soft input to the turbo decoder as follows:

$$L(y) = \log\left(\frac{p_y(y|\omega_2 = 1)}{p_y(y|\omega_2 = 0)}\right) = \log\left(\frac{p(\omega_2 = 1|y)}{p(\omega_2 = 0|y)}\right),$$

$$L(y_1, y) = \log\left(\frac{p_{y_1,y}(y_1, y|\omega_2 = 1)}{p_{y_1,y}(y_1, y|\omega_2 = 0)}\right) = \log\left(\frac{p(\omega_2 = 1|y_1, y)}{p(\omega_2 = 0|y_1, y)}\right),$$

where $p(\omega_2{=}0){=}p(\omega_2{=}1){=}1/2$ is assumed. As shown in Figure 7.4, for a strong interference scenario with INR=6 dB and for a required BER of 10^{-4}, the optimal modulator (and the sub-optimal modulator) suffers only a 0.1 dB loss compared to the no-interference case, and shows a gain of 1.2 dB to the interference-cancellation scheme and 2.4 dB to the optimized THP. These gains approximately equal the

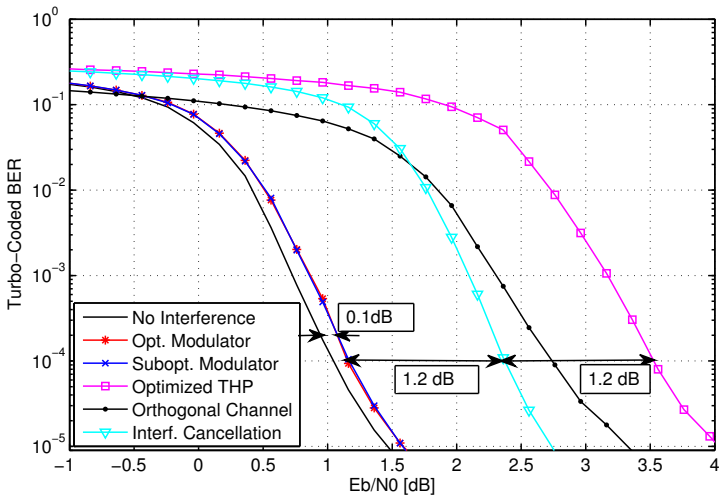


Figure 7.4. BER of binary signaling in binary interference with a rate-1/3 turbo code with INR $(P_z/\sigma^2) = 6$ dB.

difference in required SNR to achieve a mutual information of ~ 0.33 bits/channel use (or slightly larger, since the code is not capacity achieving) as shown in Figure 7.2. The same conclusion holds for heuristic THP (not shown here to simplify the figure).

7.7.3 Energy Efficiency

In order to measure the energy efficiency of the different schemes, we defined "equivalent SNR" γ_X as the required SNR for scheme X to achieve the same mutual information as in the no-interference case despite the presence of interference. That is, given SNR P_x/σ^2 and INR P_z/σ^2, there exists a constant $\gamma_X(P_x/\sigma^2, P_z/\sigma^2)$ such that

$$I_X(\gamma_X) = I_{\text{no interf.}}(P_x/\sigma^2), \tag{7.25}$$

where $I_{\text{no interf.}}(P_x/\sigma^2)$ is the end-to-end mutual information with SNR P_x/σ^2 for the no-interference case, and $I_X(\gamma_X)$ is the mutual information for scheme X with SNR γ_X.

In Figure 7.5, we compare the energy efficiency of the interference-cancellation (left) and of the sub-optimal modulator (right) against the optimal modulator in different channel conditions. Their performance difference in terms of equivalent SNR ($\gamma_{\text{opt. mod.}} - \gamma_X$, in dB) is shown as contour plots. A positive number indicates that the scheme has a better energy efficiency than the optimal modulator. The interference-cancellation scheme achieves slightly better performance than the optimal modulator when the INR is around 0 dB and the SNR is medium or high,

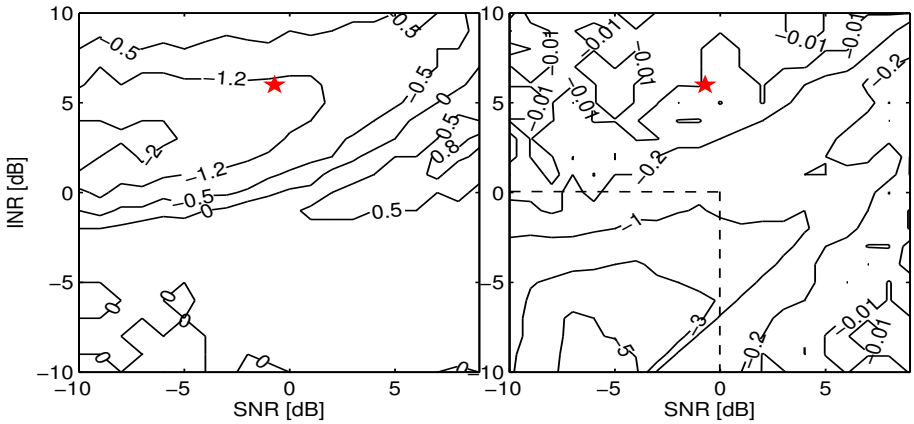


Figure 7.5. Energy efficiency (difference in terms of equivalent SNR, in dB) of interference-cancellation (left) and the sub-optimal modulator (right) compared to the optimal modulator. A positive number indicates better energy efficiency.

with the highest gain being 0.8 dB. When the INR is larger than 0 dB and the SNR is low, the interference-cancellation scheme suffers a performance loss of up to 2 dB. The loss in energy efficiency of the sub-optimal modulator is rather small except in the region indicated by the dashed line where both the SNR and the INR are relatively small[6]. The performance difference demonstrated in Figure 7.4 has been confirmed here, as shown by the $\star$ in Figure 7.5.

The SNR loss for optimized THP compared to the optimal modulator, as shown in Figure 7.6, is up to 3.5 dB when the interference power is comparable to the signal power.

7.7.4 Optimized THP with Gaussian Interference

We next demonstrate partial interference cancellation at the transmitter via the optimized THP proposed in Section 7.5. For this purpose we consider transmission of binary (BPSK) and quaternary (4-PAM) symbols on a channel with Gaussian interference. Figure 7.7 shows the result for both heuristic and optimized THP (computation of the optimal modulator is not directly feasible for this case; cf. Section 7.5). It is clear that we can gain from optimizing the parameters of THP. For quaternary signaling, the gain is significant especially in low signal-to-interference ratio (SIR, P_x/P_z) regions where interference dominates. This indicates that THP is a fairly effective (yet strictly suboptimal) means for combating Gaussian interference known at the transmitter. The gain achieved by optimizing the parameters of THP is much smaller in the binary case, however.

[6]We are more interested in interference dominated channels where DPC is most useful.

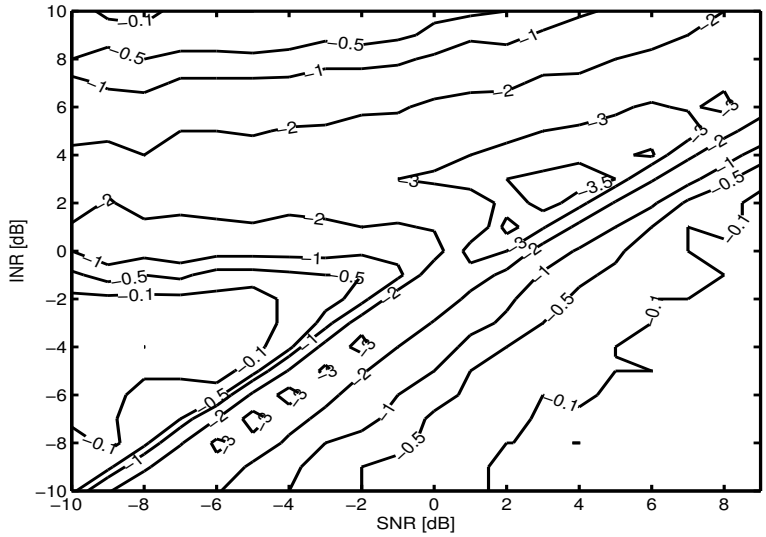


Figure 7.6. Energy efficiency loss (difference in terms of equivalent SNR, in dB) of optimized THP compared to the optimal modulator.

7.8 Summary

In this chapter we have studied DPC solutions for a relay-aided downlink channel that partially solve the Costa precoding problem using symbol-by-symbol processing. We started from the simplest scenario of binary signaling with binary interference, and derived the optimal modulator which maximizes the mutual information between the transmitter and the receiver. By proposing a mapping set size reduction method, we extended this approach to finite-alphabet signaling and interference. We also proposed a sub-optimal modulator based on the maximization of the minimum constellation distance, which was formulated as a QCQP optimization problem and approximately solved by convex optimization after relaxation. A closed-form solution of the sub-optimal modulator was obtained for the case of binary signaling with binary interference, and the performance degradation is very limited compared to the optimal modulator in most interesting scenarios. For arbitrary signal and interference, we proposed an optimized version of THP that outperforms the THP with heuristic parameters. Our proposed DPC solutions were evaluated by simulation of mutual information, coded BER, as well as energy efficiency, and compared to three benchmark schemes, namely THP, orthogonal transmission, and receiver centric interference cancellation. Simulation results showed that both the optimal and sub-optimal modulators typically outperform THP, even when the parameters of the latter are optimally chosen. For example, in high INR scenarios the optimal/sub-optimal modulator can outperform the receiver centric interference cancellation scheme by about 1 dB and outperform

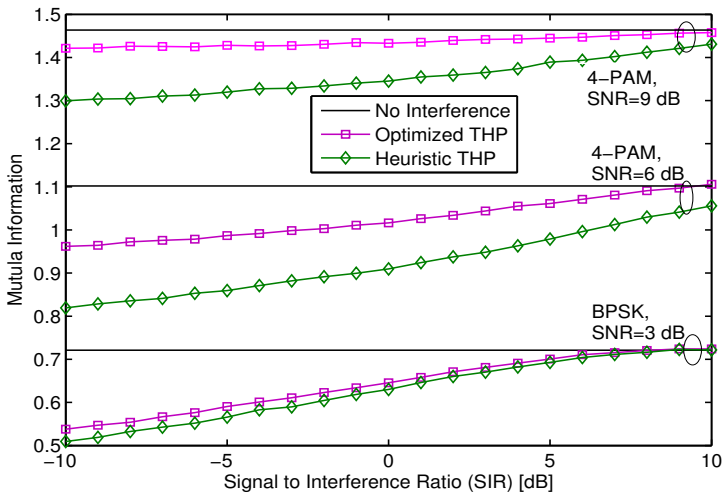


Figure 7.7. Mutual information of the optimized THP versus heuristic THP for binary ($M = 2$) and quaternary ($M = 4$) signaling with Gaussian interference and Gaussian noise.

the optimized THP by 2∼3dB. Furthermore, we have demonstrated that the gains predicted by our analysis translate directly into energy savings in a turbo coded communication link. Mutual information is therefore a relevant performance measure.

Our study indicates that rather impressive transmitter interference cancellation performance can be achieved in a single dimension. This result serves as motivation to further study low-complexity approaches to the Costa problem. Also, an implementation of DPC schemes in practice will likely rely on operations in a space of small dimensions, so the problem studied here appears to be highly relevant.

7.9 Appendix

7.9.A Derivation of the Sub-optimal Modulator

For M=2 and N=2, there are in total four different symbol distances among received signal constellation points, namely,

$$
\begin{aligned}
d_1 &= |X(0,-\beta) - X(1,-\beta)|, \\
d_2 &= |X(0,-\beta) - X(1,\beta) - 2\beta|, \\
d_3 &= |X(0,\beta) - X(1,-\beta) + 2\beta|, \\
d_4 &= |X(0,\beta) - X(1,\beta)|.
\end{aligned}
$$

As stated in Section 7.3.2, only two mappings IX and XII as stated in (7.9) need to be considered. We can find from mappings IX and XII that $d_1 = d_4$, and therefore only $\{d_1, d_2, d_3\}$ are used to identify the modulator $X(\omega_2, z)$.

When $P_x \leq \beta^2$, using mapping XII with $a = b = \sqrt{P_x}$ will result in

$$d_2 = d_3 > d_1 = 2\sqrt{P_x}.$$

The optimal detector which compares $|y|$ with the threshold β will give almost the same performance as the no-interference case.

When $P_x > \beta^2$, we can calculate the maximized minimum distance for each mapping and then identify the larger one. Without loss of generality, we assume $a, b \geq 0$ in the following. For mapping IX we have

$$d_{\min} = \min\{a + b, 2|\beta - a|, 2(\beta + b)\}.$$

If $\beta \geq a$, we have $2(\beta + b) \geq 2(a + b) > a + b$ which means

$$d_{\min} = \min\{a + b, 2(\beta - a)\}.$$

The maximum of $d_{\min}$ is achieved when $a + b = 2(\beta - a)$. Combine this condition with the power constraint $a^2 + b^2 = 2P_x$, we can formulate a new equation of a as

$$5a^2 - 6\beta a + 2\beta^2 - P_x = 0,$$

which has two roots

$$a = \frac{3\beta - \sqrt{5P_x - \beta^2}}{5},\ a = \frac{3\beta + \sqrt{5P_x - \beta^2}}{5}.$$

The former root is valid (greater than 0) only if $P_x \leq 2\beta^2$ and the latter root conflicts with the precondition $a \leq \beta$. Hence for $\beta^2 < P_x \leq 2\beta^2$ we have

$$d^*_{\text{IX}} = \frac{4\beta + 2\sqrt{5P_x - \beta^2}}{5}, \tag{7.26}$$
$$\text{with } a = \frac{3\beta - \sqrt{5P_x - \beta^2}}{5},\ b = \frac{\beta + 3\sqrt{5P_x - \beta^2}}{5}.$$

If $\beta < a$, we have

$$d_{\min} = \min\{a + b,\ 2(a - \beta),\ 2(\beta + b)\},$$

whose maximum is achieved when

$$a + b = 2(a - \beta) = 2(\beta + b),$$

i.e. $b = a - 2\beta$. Combine this with the power constraint, we get

$$a^2 - 2\beta a + 2\beta^2 - P_x = 0,$$

which has a valid solution (only if $P_x \geq 2\beta^2$) as follows

$$a = \beta + \sqrt{P_x - \beta^2},\ b = \sqrt{P_x - \beta^2} - \beta.$$

Hence for $P_x \geq 2\beta^2$ we have

$$\begin{aligned} d^*_{\text{IX}} &= 2\sqrt{P_x - \beta^2}, \\ \text{with } a &= \beta + \sqrt{P_x - \beta^2},\ b = \sqrt{P_x - \beta^2} - \beta. \end{aligned} \tag{7.27}$$

When mapping XII is used,

$$d_{\min} = \min\{a + b,\ |2\beta + b - a|\}.$$

If $2\beta + b - a < 0$, the maximum of $d_{\min}$ is achieved when

$$a + b = a - b - 2\beta,$$

i.e., $b + \beta = 0$ which is impossible. If $2\beta + b - a > 0$, the maximum of $d_{\min}$ is achieved when

$$a + b = 2\beta + b - a,$$

i.e. $a = \beta$. Hence we have

$$\begin{aligned} d^*_{\text{XII}} &= \beta + \sqrt{2P_x - \beta^2}, \\ \text{with } a &= \beta,\ b = \sqrt{2P_x - \beta^2}. \end{aligned} \tag{7.28}$$

Compared with d^*_{IX} in (7.26) and (7.27), d^*_{XII} in (7.28) has greater value for $\beta^2 < P_x < 5\beta^2$ and therefore mapping XII will be selected in this region and mapping IX will be selected when $P_x \geq 5\beta^2$. Together with the finding for $P_x \leq \beta^2$, one can easily conclude the results shown in (7.17).

7.9.B Parameters for Heuristic THP

For binary signaling $w \in \{-\alpha, \alpha\}$ with binary interference $z \in \{-\beta, \beta\}$, there are four different combinations/values ($\beta+\alpha,\ \beta-\alpha,\ -\beta+\alpha,\ -\beta-\alpha$) subject to modulo operation with $\Lambda = 4\alpha$ to ensure $|x| \leq \Lambda/2$. By the assumption of equiprobable signals and interference as stated in (7.6), only 2 out of these 4 values are of interest due to their amplitude symmetry. Without loss of generality, we just select $\beta + \alpha$ and $\beta - \alpha$.

When $\beta \leq \alpha$, all the values satisfy the requirement $|x| \leq \Lambda/2$ and therefore we have

$$P_x = E[x^2] = \alpha^2 + \beta^2 \geq 2\beta^2.$$

Hence for $P_x \geq 2\beta^2$ we have $\alpha = \sqrt{P_x - \beta^2}$. When $\alpha < \beta \leq 3\alpha$, we have

$$\begin{aligned} 2\alpha < \beta + \alpha &\leq 4\alpha, \\ 0 < \beta - \alpha &\leq 2\alpha. \end{aligned}$$

After modulo operation the resulting x has average power

$$P_x = E[x^2] = 5\alpha^2 - 4\beta\alpha + \beta^2 < 2\alpha^2,$$

which has only one feasible solution

$$\alpha = \frac{2\beta + \sqrt{5P_x - \beta^2}}{5}, \text{ for } \beta^2/5 \le P_x < 2\beta^2.$$

When $(2n-1)\alpha < \beta \le (2n+1)\alpha$, for $n = 2, 3, ...$, we have

$$\begin{aligned} 2n\alpha < \beta + \alpha \le (2n+2)\alpha, \\ (2n-2)\alpha < \beta - \alpha \le 2n\alpha. \end{aligned}$$

For n even, the modulo operation will subtract $2n\alpha$ from the above two values and result in $\beta - (2n-1)\alpha$ and $\beta - (2n+1)\alpha$ respectively. For n odd, the modulo operation will subtract $(2n+2)\alpha$ and therefore result in $\beta - (2n+1)\alpha$ and $\beta - (2n-1)\alpha$. Hence we conclude that

$$P_x = E[x^2] = (4n^2+1)\alpha^2 - 4n\beta\alpha + \beta^2 < 2\alpha^2.$$

Similarly, by solving the above equation for α we can get

$$\alpha = \frac{2n\beta + \sqrt{(4n^2+1)P_x - \beta^2}}{4n^2+1}$$

when

$$\frac{\beta^2}{4n^2+1} \le P_x < \frac{\beta^2}{4(n-1)^2+1}.$$

Another solution is infeasible and thus dropped.

Summarize all the above derivation we have proved (7.19).

Chapter 8

Thesis Conclusions

8.1 Conclusions

In this thesis, we have investigated different cooperative communication strategies in wireless networks when relaying is in use. We focus on the fundamental limits of these cooperation schemes and highlight the principles and insights behind the observed gains.

We have considered a relay-aided two-source two-sink wireless multicast network with a backhaul link between the source nodes. We have successfully characterized the capacity outer bounds by extending the proof of the converse developed by Cover and El Gamal [CE79] for the Gaussian relay channel. For the multicast relay network with high-rate backhaul, we find the exact cut-set bound of the capacity region. For low-rate backhaul scenarios, we have provided genie-aided outer bounds by introducing two new lemmas on conditional (co-)variance.

Different cooperative network coding strategies have been investigated when the relay nodes performs either decoding, compression, or amplification. We have characterized the achievable rate regions and compared them with the outer bounds and non-network coding based benchmarks. Significant rate gains have been demonstrated with the help of the conferencing links, especially for amplification or decoding based relaying schemes, or compression based schemes but with message exchange. For high-rate backhaul, we have shown that the cut-set bound can be achieved in certain channel configurations. In general, network coding based beamforming (NBF) strategies give the best performance. In high SNR regions, however, the lattice code based strategy is preferred. FNC, which only performs modulo-2 addition in the finite field, suffers limited performance loss in most of the cases. For low-rate backhaul, i.e., when only partial source cooperation is possible, we have proposed a partial-decode-and-forward based linear network coding scheme which can reduce the constraint of decoding at the relay node. We have demonstrated that encoding delay and memory constraints can affect the achievable rate of regular NNC, and employing instead short-message NNC can provide significant

gains and therefore can achieve a strictly larger rate region than NNC. In absence of encoding delay and memory constraints, however, NNC with message exchange can achieve almost the same rate as SNNC and therefore outperforms NNC with compression forwarding due to the coherent combining gain. We have also shown that AF+ANC can outperform NNC when the coherent combining gain is dominant and outperform CF/SNNC when the asymmetric channel constraint is significant. The gap between inner and outer bounds is small, within 0.3 bits in the scenarios we have considered. By adaptively exploiting these cooperation schemes based on channel quality information, we may achieve a better inner bound and therefore a smaller gap.

For wireless networks with independent noise, we have proposed a simple framework to get outer and inner bounds based on the "one-shot" bounding models. We have extended the bounding models for two-user broadcast channel to many-user scenario and establish the gap between upper and lower bounding models. For networks with coupled links, we have proposed a network decoupling method which can decompose the network into overlapping multiple-access channels and broadcast channels. We then apply the one-shot upper bounding blocks and create an upper bounding network with only bit-pipe channels. When developing the lower bounding network, we have proposed to update these lower bounding models for each coupled broadcast and multiple-access channels. We have demonstrated by some examples that the resulting upper bound is in general very good and the gap between the upper and lower bound is usually not large.

For relay-aided downlink scenarios, we have proposed a cooperation scheme by cancelling interference at the transmitter via symbol-by-symbol processing. We started from the simplest scenario of binary signaling with binary interference, and derived the optimal modulator which maximizes the mutual information between the transmitter and the receiver. By proposing a mapping set size reduction method, we have extended this approach to finite-alphabet signaling and interference. We have also proposed a sub-optimal modulator based on the maximization of the minimum constellation distance, which can be formulated as a QCQP optimization problem and approximately solved by convex optimization after relaxation. For arbitrary signal and interference, we have proposed an optimized version of THP that outperforms the THP with heuristic parameters. We have showed that both the optimal and sub-optimal modulators typically outperform THP, even when the parameters of the latter are optimally chosen. Furthermore, we have demonstrated that the gains predicted by our analysis translate directly into energy savings in a turbo coded communication link. Mutual information is therefore a relevant performance measure.

The results present in this thesis will provide better understanding of various cooperative communication strategies that have been investigated in our research work, which will guide the design and implementation of future cooperative communication systems.

8.2 Future Work

Several open problems remain for future work. As we have demonstrated that the NBF strategy cannot take full advantage of the cross-link in high-rate backhaul scenario, a better cooperating scheme is needed. For NNC with compression forwarding via backhaul, the optimal compression functions is yet to be determined based on the available backhaul capacity, channel settings, and message delivery requirement. In absence of memory and encoding delay constraints, and introducing fading (known only at the receiver side, either relay or destinations), NNC may actually be advantageous, as there will be many blocks conveying the same message, ergodic rates prevail while short messages, even with backward decoding are subject to outages. This issue will be investigated.

Most of the results obtained in this thesis can be extended to more general networks. For example, it is very interesting to figure out a systematic way to extend our outer bounding methods to more networks, either with more relay nodes, or with more source nodes. The various cooperative network coding schemes proposed in this thesis can also be applied to general network setups. For asymmetric channel setups, it is also very interesting to investigate the effect of multi-level compression at the relay nodes for CF/NNC/SNNC based schemes. Our study on symbol-by-symbol DPC has indicated that rather impressive transmitter interference cancellation performance can be achieved in a single dimension. This result serves as motivation to further study low-complexity approaches to the DPC problem.

Bibliography

[AAFS04] M. Airy, R. W. H. J. A. Forenza, and S. Shakkottai, "Practical Costa precoding for the multiple antenna broadcast channel," in *Proceedings IEEE Global Communications Conference*, Dec. 2004.

[ACLY00] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204–1216, Jul. 2000.

[ADSS12] V. Aggarwal, M. Duarte, A. Sabharwal, and N. K. Shankaranarayanan, "Full- or half-duplex? A capacity analysis with bounded radio resources," in *Proceedings IEEE Information Theory Workshop (ITW)*, Sep. 2012.

[ADT11] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless netwrok information flow: a deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, pp. 1872–1905, Apr. 2011.

[AH09] A. Avestimehr and T. Ho, "Approximate capacity of the symmetric half-duplex Gaussian butterfly network," in *Proceedings IEEE Information Theory Workshop (ITW)*, Jun. 2009.

[BBCS06] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 1872–1889, May 2006.

[BG96] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, pp. 1262–1271, Oct. 1996.

[BLW08] S. I. Bross, A. Lapidoth, and M. A. Wigger, "The Gaussian MAC with conferencing encoders," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Jul. 2008.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[Car78] A. B. Carleial, "Interference channels," *IEEE Transactions on Information Theory*, vol. 24, pp. 60–70, Jan. 1978.

[Car82] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 841–850, Nov. 1982.

[CE79] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, pp. 572–584, Sep. 1979.

[CG08] T. Chan and A. Grant, "Dualities between entropy functions and network codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 3129–3139, Oct. 2008.

[CME11] F. P. Calmon, M. Médard, and M. Effros, "Equivalent models for multiterminal channels," in *Proceedings IEEE Information Theory Workshop (ITW)*, Oct. 2011.

[Cos83] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, May 1983.

[CS03] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, Jul. 2003.

[CT98] T. M. Cover and J. A. Thomas, "Determinant inequalities via information theory," *SIAM journal on Matrix Analysis and Applications*, vol. 9, pp. 384–392, Jul. 1998.

[CT06] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[DLS06] J. Du, E. G. Larsson, and M. Skoglund, "Costa precoding in one dimension," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006.

[DLXS11] J. Du, E. G. Larsson, M. Xiao, and M. Skoglund, "Optimal symbol-by-symbol Costa precoding for a relay-aided downlink channel," *IEEE Transactions on Communications*, vol. 59, pp. 2274–2284, Aug. 2011.

[DMT06] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 1813–1827, May 2006.

[DS09] J. Du and S. Signell, "Novel preamble-based channel estimation for OFDM/OQAM systems," in *Proceedings IEEE International Conference on Communications (ICC)*, Jun. 2009.

[DXS10a] J. Du, M. Xiao, and M. Skoglund, "Capacity bounds for relay-aided wireless multiple multicast with backhaul," in *Proceedings International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2010.

[DXS10b] J. Du, M. Xiao, and M. Skoglund, "Cooperative strategies for relay-aided multi-cell wireless networks with backhaul," in *Proceedings IEEE Information Theory Workshop (ITW)*, Aug. 2010.

[DXS11a] J. Du, M. Xiao, and M. Skoglund, "Capacity bounds for backhaul-supported wireless multicast relay networks with cross-links," in *Proceedings IEEE International Conference on Communications (ICC)*, Jun. 2011.

[DXS11b] J. Du, M. Xiao, and M. Skoglund, "Cooperative strategies for relay-aided multi-cell wireless networks with backhaul," *IEEE Transactions on Communications*, vol. 59, pp. 2502–2514, Sep. 2011.

[DXSM] J. Du, M. Xiao, M. Skoglund, and M. Médard, "Wireless multicast relay networks with limited-rate source-conferencing," *IEEE Journal on Selected Areas in Communications*, special issue on Theories and Methods for Advanced Wireless Relays. To appear.

[DXSS12] J. Du, M. Xiao, M. Skoglund, and S. Shamai (Shitz), "Short-message noisy network coding with partial source cooperation," in *Proceedings IEEE Information Theory Workshop (ITW)*, Sep. 2012.

[DXWC12] J. Du, P. Xiao, J. Wu, and Q. Chen, "Design of isotropic orthogonal transform algorithm-based multicarrier systems with blind channel estimation," *IET Communications*, 2012, accepted for publication. DOI: 10.1049/iet-com.2012.0029

[EK10] A. El Gamal and Y.-H. Kim, *Lecture Notes on Network Information Theory*, 2010, arXiv:1001.3404.

[EK11] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[EMZ06] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 1545–1561, Apr. 2006.

[EPG94] J. A. Erfanian, S. Pasupathy, and G. Gulak, "Reduced complexity symbol detectors with parallel structures for ISI channels," *IEEE Transactions on Communications*, vol. 42, pp. 1661–1671, Feb. 1994.

[EtB05] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, pp. 3417–3432, Oct. 2005.

[EU10] E. Ekrem and S. Ulukus, "On Gaussian MIMO broadcast channels with common and private messages," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Jun. 2010.

[EZ04] U. Erez and R. Zamir, "Achieving 1/2 log (1+ SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, pp. 2293–2314, Oct. 2004.

[GB] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming." [Online]. Available: http://www.stanford.edu/ boyd/cvx/

[GJ07] K. S. Gomadam and S. A. Jafar, "Optimal relay functionality for SNR maximization in memoryless relay networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 390–402, Feb. 2007.

[GSG+09] D. Gündüz, O. Simeone, A. J. Goldsmith, H. V. Poor, and S. Shamai, "Relaying simultaneous multicast messages," in *Proceedings IEEE Information Theory Workshop (ITW)*, Jun. 2009.

[GSG+10] D. Gündüz, O. Simeone, A. J. Goldsmith, H. V. Poor, and S. Shamai, "Multiple multicasts with the help of a relay," *IEEE Transactions on Information Theory*, vol. 56, pp. 6142–6158, Dec. 2010.

[Gus10] R. Guseo, "Partial and ecological correlation: a common three-term covariance decomposition," *Statistical Methods and Applications*, vol. 19, pp. 31–46, Mar. 2010.

[HM72] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Transactions on Communications*, vol. COM-20, pp. 774–780, Aug. 1972.

[HM06] A. Høst-Madsen, "Capacity bounds for cooperative diversity," *IEEE Transactions on Information Theory*, vol. 52, pp. 1522–1544, Apr. 2006.

[HMK+06] T. Ho, M. Médard, R. Koetter, M. Effros, D. R. Karger, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, pp. 4413–4430, Oct. 2006.

[HMZ05] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 2020–2040, Jun. 2005.

[KAPT09] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 773–787, Jun. 2009, special Issue on Network Coding for Wireless Communication Networks.

[KEM09] R. Koetter, M. Effros, and M. Médard, "Beyond network equivalence," in *Proceedings 47th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2009.

[KEM10] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence, part I and II," 2010, arXiv:1007.1033.

[KEM11] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence–part I: point-to-point channels," *IEEE Transactions on Information Theory*, vol. 57, pp. 972–995, Feb. 2011.

[KGG05] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, pp. 3037–3063, Sep. 2005.

[KGK07] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proceedings ACM SIGCOMM*, Aug. 2007.

[KH11a] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *Proceedings IEEE Information Theory Workshop (ITW)*, Oct. 2011.

[KH11b] G. Kramer and J. Hou, "Short-message quantize-forward network coding," in *Proceedings 8th International Workshop on Multi-Carrier Systems & Sollutions*, May 2011.

[KL08] M. N. Khormuji and E. G. Larsson, "Rate-optimized constellation rearrangement for the relay channel," *IEEE Communication Letters*, vol. 12, pp. 618–620, Sep. 2008.

[KM03] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 782–795, Oct. 2003.

[KMG+07] S. Katti, I. Marić, A. J. Goldsmith, D. Katabi, and M. Médard, "Joint relaying and network coding in wireless networks," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Jun. 2007.

[Kra07] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, pp. 265–444, 2007.

[Kra12] G. Kramer, "Progress on relaying and noisy network coding," *Keynote lecture at Internation Zürich Seminar*, Feb. 2012.

[KRH+08] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking*, vol. 16, pp. 497–510, Jun. 2008.

[KRV11] S. Kannan, A. Raja, and P. Viswanath, "Local phy + global flow: a layering principle for wireless networks," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Aug. 2011.

[KS09] M. N. Khormuji and M. Skoglund, "Instantaneous forwarding strategies for relay channels with known interference," in *Proceedings IEEE Information Theory Workshop (ITW)*, Jun. 2009.

[KvW00] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Jan. 2000.

[Lia05] A. Liavas, "Tomlinson-Harashima precoding with partial channel knowledge," *IEEE Transactions on Communications*, vol. 53, pp. 5–9, Jan. 2005.

[LK07] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Transactions on Information Theory*, vol. 53, pp. 3517–3535, Oct. 2007.

[LKEC11] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Transactions on Information Theory*, vol. 57, pp. 3132–3152, May 2011.

[LL03] A. R. Lehman and E. Lehman, "Complexity classifications of network information flow problems," in *Proceedings 41st Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2003.

[LMS+10] Z.-Q. Luo, W.-K. Ma, M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, pp. 20–27, May 2010.

[LTW04] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[LYC03] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 371–381, Feb. 2003.

[MGM12] I. Marić, A. J. Goldsmith, and M. Médard, "Multihop analog network coding via amplify-and-forward: the high SNR regime," *IEEE Transactions on Information Theory*, vol. 58, pp. 793–803, Feb. 2012.

[MYK07] I. Marić, R. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Transactions on Information Theory*, vol. 53, pp. 3536–3548, Oct. 2007.

[NG11] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, pp. 6463–6486, Oct. 2011.

[NJGM07] C. T. K. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, "Capacity gain from two-transmitter and two-receiver cooperation," *IEEE Transactions on Information Theory*, vol. 53, pp. 3822–3827, Oct. 2007.

[PV11] V. M. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *IEEE Transactions on Information Theory*, vol. 57, pp. 156–186, Jan. 2011.

[Ray06] S. Ray, "Energy efficient multiple antenna communication," Ph.D. dissertation, MIT, Sep. 2006.

[RU96] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, pp. 364–375, Mar. 1996.

[SE07] O. Sahin and E. Erkip, "Achievable rates for the Gaussian interference relay channel," in *Proceedings IEEE Global Communications Conference*, Nov. 2007.

[SEA03a] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–part I: System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, Nov. 2003.

[SEA03b] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, pp. 1939–1948, Nov. 2003.

[SGP+09] O. Simeone, D. Gündüz, H. V. Poor, A. J. Goldsmith, and S. Shamai, "Compound multiple-access channels with partial cooperation," *IEEE Transactions on Information Theory*, vol. 55, pp. 2425–2441, Jun. 2009.

[SL08] M. Skoglund and E. G. Larsson, "Optimal modulation for known interference," *IEEE Transactions on Communications*, vol. 56, pp. 1892–1899, Nov. 2008.

[SW73] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 47l–480, Jul. 1973.

[THMK] D. Traskov, M. Heindlmaier, M. Médard, and R. Koetter, "Scheduling for network coded multicast," *IEEE/ACM Transactions on Networking*, accepted for publication. DOI: 10.1109/TNET.2011.2180736.

[Tom71] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electronics Letters*, pp. 138–139, Mar. 1971.

[TV05] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[TY11] Y. Tian and A. Yener, "The Gaussian interference relay channel: improved achievable rates and sum rate upperbounds using a potent relay," *IEEE Transactions on Information Theory*, vol. 57, pp. 2865–2879, May 2011.

[vdM71] E. C. van der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, pp. 120–154, 1971.

[WBBJ11] M. Wiese, H. Boche, I. Bjelaković, and V. Jungnickel, "The compound multiple access channel with partially cooperating encoders," *IEEE Transactions on Information Theory*, vol. 57, pp. 3045–3066, May 2011.

[WC98] R. D. Wesel and J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol. 44, pp. 824–831, Mar. 1998.

[WFVH04] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1305–1316, Jul. 2004.

[Wil82] F. M. J. Willems, "Information theoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Katholieke Univ. Leuven, Oct. 1982.

[Wil83] F. M. J. Willems, "The discrete memoryless multiple-access channel with partially cooperating encoders," *IEEE Transactions on Information Theory*, vol. 29, pp. 441–445, May 1983.

[WNPS10] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," *IEEE Transactions on Information Theory*, vol. 56, pp. 5641–5654, Nov. 2010.

[WSS06] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

[WT11] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited transmitter cooperation," *IEEE Transactions on Information Theory*, vol. 57, pp. 2941–2965, May 2011.

[WX10] X. Wu and L.-L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," 2010, arXiv:1009.5959.

[WZ76] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

[YVC05] W. Yu, D. P. Varodayan, and J. M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference pre-substraction," *IEEE Transactions on Communications*, vol. 53, pp. 1220–1230, Jul. 2005.

[ZG11] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Transactions on Information Theory*, vol. 59, pp. 6102–6111, Dec. 2011.

[ZMM08] H. Zhang, N. B. Mehta, and A. F. Molisch, "Asynchronous interference mitigation in cooperative base station systems," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 155–165, Jan. 2008.

[ZSE02] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, pp. 1250–1276, Jun. 2002.

[ZV09] A. Zaidi and L. Vandendorpe, "Achievable rates for the Gaussian relay interferer channel with a cognitive source," in *Proceedings IEEE International Conference on Communications (ICC)*, Jun. 2009.

[ZY11] L. Zhou and W. Yu, "Incremental relaying for the Gaussian interference channel with a degraded broadcasting relay," in *Proceedings 49th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2011.