Approximation for Problems in Multi-user Information Theory

Approximation for Problems in Multi-user Information Theory

Soheil Mohajer

EPFL - Ecole Polytechnique Fédérale de Lausanne

Thesis No. 4816 (September 2010)

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Accepted by the jury:

S. Diggavi
Thesis directors

D. Tse
Expert

C. Fragouli
Expert

C. Tian
Expert

R. Urbanke
President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2010

To my lovely wife, Sara,

and my dear parents, Zari & Hossein,

for all their love and support.

Abstract

The main goal in network information theory is to identify fundamental limits of communication over networks, and design solutions which perform close to such limits. After several decades of effort, many important problems still do not have a characterization of achievable performance in terms of a finite dimensional description. Given this discouraging state of affairs, a natural question to ask is whether there are systematic approaches to make progress on these open questions. Recently, there has been significant progress on several open questions by seeking a (provably) approximate characterization for these open questions. The main goal of approximation in network information theory is to obtain a *universal* approximation gap between the achievable and the optimal performance.

This approach consists of four ingredients: simplify the model, obtain optimal solution for the simplified model, translate this optimal scheme and outer bounds back to the original model, and finally bound the gap between what can be achieved using the obtained technique and the outer bound. Using such an approach, recent progress has been made in several problems such as the Gaussian interference channel, Gaussian relay networks, etc. In this thesis, we demonstrate that this approach is not only successful in problems of transmission over noisy networks, but gives the first approximation for a network data compression problem. We use this methodology to (approximately) resolve problems that have been open for several decades. Not only do we give theoretical characterization, but we also develop new coding schemes that are required to satisfy this approximate optimality property. These ideas could give insights into efficient design of future network communication systems.

This thesis is split into two main parts. The first part deals with the approximation in lossy network data compression. Here, a lossy data compression problem is approximated by a lossless counterpart problem, where all the bits in the binary expansion of the source above the required distortion have to be losslessly delivered to the destination. In particular, we study the multiple description (MD) problem, based on the multi-level diversity (MLD) coding problem. The symmetric version of the MLD problem is well-studied, and we can directly use it to approximate the symmetric MD problem. We formulate the asymmetric multi-level diversity problem, and solve it for three-description case. The optimal solution for this problem, which will be later used to approximate the asymmetric multiple description problem, is based on jointly compressing of independent sources. In both symmetric and asymmetric cases, we derive inner and outer bounds for the achievable rate region, which together with the gap analysis, provide an approximate solution for the problem. In particular, we resolve the symmetric Gaussian MD problem, which has been open for three decades, to within 1 bit.

In the second part, we initiate a study of a Gaussian relay-interference network, in which relay (helper) nodes are to facilitate competing information flows over a wireless network. We focus on a two-stage relay-interference network where there are weak cross-links, causing the networks to behave like a chain of Z Gaussian channels. For these Gaussian ZZ and ZS networks, we establish an approximate characterization of the rate region. The outer bounds to the capacity region are established using genie-aided techniques that yield bounds sharper than the traditional cut-set outer bounds. For the inner bound of the ZZ network, we propose a new interference management scheme, termed interference neutralization, which is implemented using structured lattice codes. This technique allows for over-the-air interference removal, without the transmitters having complete access to the interfering signals. We use insights gained from an exact characterization of the corresponding linear deterministic version of the problem, in order to study the Gaussian network. We resolve the Gaussian relay-interference network to within 2 bits. The new interference management technique (interference neutralization) shows the use of structured lattice codes in the problem.

We also consider communication from a source to a destination over a wireless network with the help of a set of authenticated relays, and presence of an adversarial jammer who wishes to disturb communication. We focus on a special diamond network, and show that use of interference suppression (nulling) is crucial to approach the capacity of the network. The exact capacity characterization for the deterministic network, along with an approximate characterization (to within 4 bits) for the Gaussian network is provided.

The common theme that binds the diverse network communication problems in this thesis is that of approximate characterization, when exact resolutions are difficult. The approach of focusing on the deterministic/lossless problems underlying the noisy/lossy network communication problems has allowed us to develop new techniques to study these questions. These new techniques might be of independent interest in other network information theory problems.

Keywords: network information theory, approximation, degraded message broadcast, multi-level diversity coding, multiple description coding, multiple unicast over deterministic/Gaussian network, relay-interference network, interference neutralization, structured codes, adversarial jamming over network.

Résumé

L'objectif principal de la théorie de l'information des réseaux est d'identifier les limites fondamentales de la communication dans les réseaux et d'explorer des solutions qui atteignent de telles limites au plus près. Après plusieurs décennies d'efforts, de nombreux réseaux n'ont toujours pas trouvé de caractérisation simple en termes de performance. Etant donné cet état de fait, une question naturelle consiste à se damnder s'il existe des approches systématiques pour progresser sur ces questions ouvertes. Récemment, des progrès significatifs ont été accomplis dans cette direction, qui permettent d'obtenir une caractérisation approximative des systèmes étudiés. L'objectif principal de l'approximation dans la théorie de l'information des réseaux est d'obtenir des écarts de nature universelle entre les performances atteignables et le rendement optimal.

Cette approche se compose de quatre ingrédients: simplifier le modèle, obtenir la solution optimale pour le modèle simplifié, ramener ce schéma optimal au modèle original, et fianlement majorer l'écart entre ce qui peut être atteint et la borne supérieure sur le rendement. En utilisant une telle approche, des progrès ont été accomplis récemment concernant plusieurs problèmes, tel que le canal gaussien avec interférence, les réseaux gaussiens avec relais, etc. Dans cette thèse, nous démontrons que cette approche permet non seulement de traiter des problèmes de transmission sur des réseaux bruités, mais permet aussi d'obtenir des approximations pour un problème de compresssion de données dans un réseau. Nous utilisons cette méthodologie pour reśoudre (de manière approximative) des problèmes restés ouverts depuis plusieurs décades. En plus de cette caractérisation théorique, nous développons de nouveaux systèmes de codage qui sont nécesaires pour approcher la performance optimale.

Cette thèse est divisée en deux parties principales. La première partie traite de la compression de données en réseau. Dans ce cas, un problème de compression de donnés avec distorsion est approché par un problème sans distorsion, où tous les bits dans l'expansion binaire de la source au-dessus du niveau de distorsion doivent être livrés sans perte à la destination. En particulier, nous étudions le problème de codage avec description multiple (MD) en nous basant sur le codage avec plusieurs niveaux de diversité (MLD).

La version symétrique du problème MLD est étudiée en détail et nous pouvons directement l'utiliser pour approcher le problème symétrique MD. Nous formulons la version asymétrique du problème de codage avec plusieurs niveaux de diversité (AMLD), et la résolvons pour le cas de trois descriptions. La solution optimale de ce problème, qui sera utilisée plus tard pour approximer le problème de codage avec description multiple asymétrique (AMD), est basée sur la compression conjointe de sources indépendantes. Dans les deux cas, symétrique et asymétrique, nous dérivons des limites inférieures et supérieures pour la région de capacité qui, avec l'analyse de l'écart entre ces limites, fournissent une solution approximative au problème. En particulier, nous résolvons à plus ou moins 1 bit près le problème de description multiple dans le cas gaussien symétrique, resté ouvert depuis 30 ans.

Dans la deuxième partie, nous initions une étude d'un réseau gaussien avec relais et interférences, dans lequel les relais ont pour but de faciliter le flux d'information dans un réseau sans fil. Nous nous concentrons sur un réseau à deux niveaux, avec de faibles liaisons transversales, ce qui correspond à un réseau se comportant comme une chaîne de canaux Z gaussiens. Pour les réseaux gaussiens ZZ et ZS, nous établissons une caractérisation approximative de la région de capacité. Les bornes supérieures de la région de capacité sont établies en utilisant des techniques approfondies qui permettent d'obtenir de meilleures bornes qu'avec les techniques traditionnellement utilisées pour borner la capacité des réseaux. Pour la borne inférieure du réseau ZZ, nous proposons un nouveau système de gestion de l'interférence, appelé la neutralisation des interférences, qui est mis en application en utilisant des codes structurés dits “en treillis”. Cette technique permet d'éliminer l'interférence de manière naturelle, sans que les émetteurs aient nécessairement accès aux signaux d'interférence. Nous utilisons les connaissances acquises à partir de la caractérisation exacte de la version linéaire et déterministe du problème, afin d'étudier les réseaux gaussiens. Nous résolvons le cas du réseau gaussien avec interférence et relais à plus ou moins 2 bits près. La nouvelle technique suggérée pour traiter l'interférence (neutralisation des interférences) montre comment les codes structurés “en treillis” sont utilisés dans le problème.

Nous considérons également la communication d'une source à une destination sur un réseau sans fil, à l'aide d'un ensemble de relais authentifiés, et la présence d'un brouilleur qui veut perturber la communication. Nous nous concentrons sur un réseau spécial dit “en diamant”, et montrons que l'utilisation de la suppression (ou annulation) des interférences est crucial pour approcher la capacité du réseau. Une caractérisation exacte de la capacité est obtenue pur le réseau déterministe, ainsi qu'une caractérisation approximative (à plus ou moins 4 bits près) pour le réseau gaussien.

Dans cette thèse, le thème commun qui relie les différents problèmes de communication dans les réseaux est celui de la caractérisation approximative des performances, quand la résolution exacte du problème s'avère trop difficile. L'approche qui consiste à se focaliser sur les problèmes déterministes/sans distorsion sous-jacents aux problèmes bruités/avec distorsion nous a permis de développer de nouvelles techniques pour étudier ces questions. Ces nouvelles techniques présentent un intérêt potentiel pour d'autres problèmes en théorie de l'information des réseaux.

Mots-clés: théorie de l'information des réseaux, approximation, diffusion de messages dégrades, codage avec plusieurs niveaux de diversité, codage avec description multiple, transmissions multiples sur réseaux déterministes ou gaussiens, réseaux avec relais et interférences, neutralisation des interférences, codes structurés, brouillage dans un réseau.

Acknowledgements[1]

During my several years of study at EPFL, both as a Master's student and a Ph.D. student, I have had the privilege of interacting with so many wonderful colleagues and friends who have been greatly influential in my graduate life. Despite being thousands of miles away from home, thanks to them my graduate studies turned out to be one of the best experiences of my life. These few paragraphs are an attempt to express my deepest gratitude to all those who made such an exciting experience possible.

My foremost gratitude goes to my adviser Suhas Diggavi, for not only his help and advices as a PhD supervisor, but also for being a great friend, whose character in professional life is one of my ultimate goals to achieve. It was a real fortune for me to join Laboratory of Information and Communication Systems (LICOS), where I enjoyed countless hours of discussions I had with Suhas and other colleges in the lab. I am especially thankful to Suhas for being constantly available for discussing different issues even from distant locations and at unusual times. I learned many things from Suhas, both on a technical and on a personal level. I greatly appreciated the kindness, honesty and good humor that were part of every interaction we had. These and many other values will continue to be an inspiration for me in my future professional life. Also his creative approach to teaching and giving engaging presentations was particularly enlightening.

I am indebted to Christina Fragouli, who warmly hosted me in her group, Algorithmic Research in Network Information (ARNI). We had great collaboration within the past two years of my PhD, which led to a few publications. Beside being amazingly kind, her ideas and comments have greatly improved the quality of my work. She also kindly accepted to be a member of my thesis committee.

I would like to thank Amin Shokrollahi, who was the first person introduced EPFL to me. Beside that, he was my first adviser during my Master's study at EPFL, when I did my semester project under his supervision. Moreover, he

[1]This thesis was in part funded by grants from the Swiss National Science Foundation through NCCR-MICS (Grant No. 51NF40-111400), N-CRAVE (Grant No. FP7 ICT-2007-215252), and ERC Starting Grant for NOWIRE (Network Coding for Wireless Networks) Project (Grant No. ERC-2009-StG-240317), that I gratefully acknowledge.

has taught me many things including his wonderful course on modern coding theory.

I am grateful to Chao Tian for all his help and his talent in making wonderful connections between various problems. He was not only a member of my thesis committee for both Master's and PhD, but also an unofficial co-adviser of mine during the few years he was with us in LICOS, when we worked on the first part of this thesis.

I would like to show my gratitude to David Tse, not only for accepting reading my thesis and being in my PhD thesis committee, but also for sharing his knowledge and experience with me during several hours of discussion we had. I learned many things from his brilliant and innovative comments.

My special thanks go to Rüdiger Urbanke for many things I learned from him, plus his unique sense of humor. Moreover, he was the chair of my thesis committee. I'd like to also extend my gratitude to Emre Telatar, who is an infinite dimensional source of information, and always welcomes students for wonderful discussions in various areas. I take this opportunity to thank Bixio Rimoldi, for his help and support, specially his indefatigable effort to teach me how to ski (which he unfortunately failed!). I am also indebted to Olivier Lévêque, for kindly accepting rewriting my awfully written(!) résumé[2] in French, as well as his attractive course on random matrix theory. I would like to thank Nicolas Macris, for his interesting course on quantum information theory. Additionally, I have been always envied his eagerness to read and learn.

I am deeply grateful to Muriel Bardet for all her help inside as well as outside EPFL. It is real pleasure (especially for non French speaking people) to have Muriel in the lab, who takes care of all administrative jobs and allows you to focus only on your work and research. Also, many thanks to Françoise Behn and Yvonne Huskie for all their help and the great social events of our lab. I would also like to thank Damir Laurenzi and Giovanni Cangiani for always being so prompt and friendly when dealing with my (mostly embarrassingly silly) questions. A special thank goes to Damir for being patience for all my grumbles.

I'd like to extend my warmest gratitude to my dear friend, Mahdi Cheraghchi, who was my first office-mate at EPFL, and taught me the basics of theoretical computer sciences in the most understandable way. Beside my adviser, I can hardly name a person at EPFL with whom I have discussed different topics for so many hours as I had with Mahdi. I can never forget his impressive memory, vast knowledge in various areas, and kind patience in answering my questions, particularly regarding LaTeX. Additional thank goes to Amin Karbasi, who was always a real and supportive friend. His exceptional sense of humor as well as leadership skills make him a unique character among my friends. It is a pity to mention his name without recalling our wonderful moments in playing Shelem [3] as partners! I would also like to thank Mahdi Jafari, who involved me in an interesting problem in network coding which contributed

[2] abstract

[3] A trick-taking card game.

to a few joint papers. Among different aspects, I wish to mention his deep and unique thinking style behind his relaxed and peaceful face. A special thank goes to Pedram Pedarsani, on whom and his help I could always count.

I owe my deep gratitude to G. Hossein G. Hamedani, from Marquette University, who has had a significant influence on my entire academic life. I take this opportunity to thank four of my best and most brilliant friends. Payam Pakzad and Ali Kakhbod, with whom I wrote my first paper at EPFL. I had wonderful discussions with Payam during his stay at EPFL, and I have always enjoyed from his approach in thinking about problems, and his confidence in that. Additionally I would like to thank Behrouz Touri, whom I'm privileged to know since high school. I have been lucky to have him as a friend, who has always encouraged me to think about other aspects of life beyond my academic studies. I'm indebted to Mehrdad Khakparaghi, who was probably the first person who made me interested in information theory.

During summer 2009, I spent an internship at Deutsche Telekom Laboratories in Berlin, working with Ruben Merz and BOWL group, where I learned more practical aspects of communication theory. He was always available for discussions and patient to answer to my countless silly questions regarding theoretical versus practical issues.

I'd like to express my great gratitude to my friends who accepted me in their group from the early days of my stay in Switzerland, and shared their experiences with me regarding life and study in a new society. Bijan and Bahar, Arash and Marjan, Masoud and Elaheh, Azad and Azin, Nima and Marjan, Armin and Paris, and Mehdi Molkaraie are definitely in the list. Specially, I would like to thanks Alireza and Faezeh, as well as Amir and Atefeh, for tons of memorials and lovely moments we had together. A warm thank goes to Hamid and Maryam, for all their help and support during my stay in Lausanne. I'm indebted to my great friends, Naser and Sanaz, with whom I rarely felt being away from home.

My former and current colleagues at EPFL made my PhD time memorable: Adel Aziz, Abdel Amraoui, Sibi Bhaskaran, Cyril Measson, Peter Berlin, Etienne Perron, Dinkar Vasudevan, Sanket Dusad, Shrinivas Shrinivas Kudekar, Vishwambhar Rathi , Satish Korada, Ayfer Ozgur, Marius Kleiner, Vojislav Gajic, Christine Neuberg, Dominique Tschopp, Jasper Goseling, Emmanuel Abbe, Eren Sasoglu, Mohammad Karzand, Shirin Saeedi, Ali Hormati, Lorenzo Keller, Javad Ebrahimi, Hamed Hassani, Uday Pulleti, Emre Atsan, Vahid Aref, Marc Vuffray, Aurore Amaudruz, Mine Aslan, Alberto Jimenez, Vinodh Venkatesan, Ghid Matouk, Masoud Alipour, Reza Parhizkar, and Hessam, Jérémie Ezri whom we lost a few month ago, after two years of battle with cancer. I thank them for their friendship and the good times, and for all the help I received from. In addition, I'm grateful to so many amazing friends who made my study in Switzerland an unforgettable stage of my life and full of memorable moments: Ali Mousavi, Ali Ajdari Rad, Hossein Afshari, Hossein Rouhani, Hossein Taghavi, Laleh Golestani, Mani Bastani Parizi, Marjan Hamedani, Maryam Kouchakzadeh, Maryam Zaheri, Mehdi Gholamezaee, Mona Mahmoudi, Noushin Hadadi, Parisa Haghani, Payam Delgosha Pooya

Pakzad, Pouya Dehghani, Ramtin Pedarsani, Vahid Majidzadeh, Vahid Partovinia, and Zahra Sinaei. It is worth mentioning that having prepared the list rather hastily, I'm sure I have missed a lot of nice friends on it, who hopefully forgive my carelessness.

Above all, I express my foremost gratitude to my parents, Hossein and Zahra, my parents in law Forooz and Firouzeh, my brother and his wife Bashir and Narges, my sister Ferdos, and my brother in law Alireza, who filled my life with joy and happiness. Without their love, support, and patience none of my achievements -in particular this thesis- would have been possible. I especially thank my parents for all they went through until I reached this point, and their tremendous patience during the years I was away and could go to home only for a short visits. A special thank goes to my uncle, Amir, who was my classic teacher in my life.

Words can hardly express how deeply thankful I am to my dear wife, Sara, who contributed directly and indirectly to every moments of my PhD. I could make a long list of supports and encouragements I have received from her during these years, specially the last couple of months with all my nagging, but such a list could never be complete. Looking back, I could not complete this thesis without her, and I thank her in the most heartfelt words.

Contents

List of Figures

List of Tables

1 Introduction

Communication networks have revolutionized the way we live and work. However, most of the broadband access and flow of data currently is through wired networks, such as optic fibers, telephone lines, cable, etc. Access through wireless networks is mostly limited to one-hop networks, where nodes communicate to a base-station or hot-spot. The holy grail in network information theory is to understand how multiple users can share a heterogeneous network (consisting of wired and wireless components) efficiently, to characterize fundamental limits of capacity for such situations, and to explore schemes which can approach such limits. A complete theory of network information would have significant impact on design of computer networks, and deployment of next generation of communication networks.

Figure 1.1 depicts the block diagram of a two-terminal communication system. In this system the source (which generates information) wishes to communicate to the destination. *Information* is some knowledge (*e.g.*, data, result of a measurement, etc.) wanted at some destination [1]. The source and destination are separated in space (*e.g.*, in data transmission) or time (*e.g.*, in data compression). The *channel* is in fact the bridge, through which the information can be sent from the source to the destination. Cellular base stations, Internet ISP's, and satellite are examples of communication systems who provide people (consumers) with different type of information such as voice, image, video, computer data, etc., demanded by the users.

A complete understanding of communication theory is only available for a one-hop (two-terminal) communication system. This goes back to the mathematical framework established for understanding the problem of communication in the pioneering work of Claude E. Shannon [2] . Even though this allows coding over potentially large blocks of symbols, the characterization of capacity (maximum reliable transmission rate) for a two-terminal system was

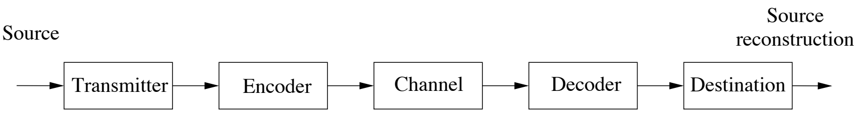


Figure 1.1: A two-terminal communication system.

reduced to a finite dimensional optimization problem. This was a surprising and insightful solution that has set a high standard for the field of information theory for networks.

A communication network (see Figure 1.2), in general, consists of sources, communication nodes (transmitter and/or receivers) who are equipped with transmission devices, and destinations (consumers of information). The destination is interested in reconstructing the source information via what received through the channel. There is a fundamental trade-off between the quality of this reconstructed data and the cost of the communication. This cost is due to different resources have to be utilized to perform communication. For instance, we use channel coding which adds redundant information to the original one. in order to deal with corruption by noise. A direct consequence of adding redundancy is increasing the transmission time, transmission power, or communication bandwidth. Reliable communication is typically desired in spite of the fact that the channel is not perfect, and its output can be random and very different from its input. Hence, it is a system design challenge to overcome the unreliable behavior of the channel. *Encoding* is an effective tool of reducing the cost of transmission and of combating the channel randomness to facilitate communication by transforming the message before entering the channel. This transformation has to be reversed at the output of the channel by the *decoder*. An encoding/decoding system accomplishes reliable communication if the end-to-end transform from the source to the receiver results in reconstructing the transmitted message within certain demanded quality.

Transmission of information is the main purpose of a communication network. It is a quite challenging problem due to the following issues: (i) Disparate demand of users in terms of quality of sources, (ii) Disparate locations of sources and sinks, (iii) Limited resources to be shared between different users, and (iv) Complex signal interaction in a wireless medium.

The general problem is easy to state [3]: *Given a set of sources and their data generation speed (rate), user demands and required quality, and transmission limits imposed by the channel, whether or not the sources can be transmitted to the sinks over the network.* This problem involves centralized and distributed source coding (data compression), and distributed communication (finding the capacity region of the network) .

Many new problems may arise when we deal with complex networks: There may exist more than one path which can be used to transmit data; How can

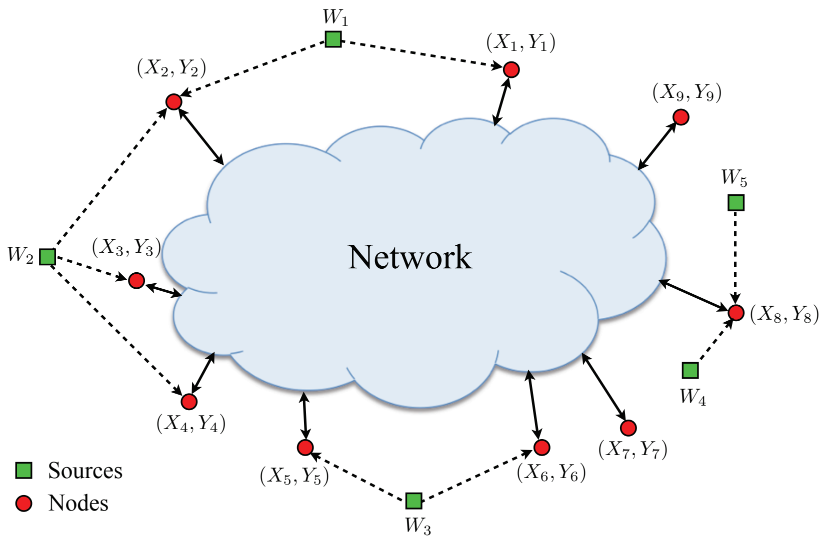


Figure 1.2: A typical communication network, with sources, transmitter nodes, and receivers. The dashed lines denote the direct access of the nodes to the sources data, and solid lines denote connectivity between the nodes in the network.

we get benefit from this diversity in the physical network? A receiver may be interested in different types of data generated and sent by different sources. What is the optimal way to perform different transmissions simultaneously? Various destinations might be interested in receiving a common data send by a unique transmitter through disparate channels to the receivers. What can be a good strategy for the transmitter to provide all of them with the demanded data? Sources can be located separately, and a distributed encoding scheme is needed to perform communication. Different types of information may flow over a wireless network. Interaction between interfering signals is an important problem need to be addressed in network information theory.

Problems in information theory can be expressed as optimization problems, that is, to minimize an *objective function*, subject to the constraints imposed by the network, and user demands. A direct solution for an information theoretic problem allows coding over large blocks of symbols. However, such characterization typically does not give significant insight, since first, such infinite-dimensional mathematical expression cannot be evaluated, and second, they do not give an engineering insight about practical strategies can be used in such situations. However, a single letter characterization[1] is only available for simple networks, which gives limited impact to engineering practice.

[1] versus a multi-letter characterization, which involves an optimization over some vector form of random variables whose length can be arbitrarily large. More formally, it is suggested by Csiszár and Körner [1] to define it through the notion of computability, *i.e.,* a problem has a single-letter solution if there exists an algorithm which can decide if a point belongs to

Several new problems have been also arisen by generalization of Shannon's point-to-point systems to complex communication networks. A network communication is characterized by several objects: (a) what are the sources? (b) what is the demand and structure? (c) what are the probabilistic channels? (d) what quality of source reconstruction is demanded? and (e) what is the cost of communication? On the other hand, a wide range of applications and connections from network information theory to other research fields, such as computer science, control theory, machine learning, biology and genome sequence analysis are discovered. However, in spite of huge effort of various researchers, the general problem is still unsolved.

1.1 Approximation

Unique among many engineering fields, information theory aims for and almost demands exactly optimal solutions to infinite-dimensional design problems. Such a high standard is achieved by Shannon [2] in his analysis for point-to-point systems. After more than 40 years of effort, meeting such a standard has proved to be far more difficult when extending Shannon's theory to networks. *Approximation* is an approach to deal with problems in network information theory, whose main goal is to universally bound the gap between what can be achieved using known techniques and the optimal solutions

We argue that progress can be made by seeking approximate solutions that have a guarantee on the gap from the optimal solution. This can lead to a novel approach for studying a large class of information theory problems that have been intractable up to now.

Focusing on the practically important models of linear Gaussian channels and Gaussian sources, this approach consists of three steps: (i) Simplify the model; (ii) Obtain optimal solution for the simplified model; (iii) Translate the optimal scheme and outer bounds back to the original model. For network channel coding problems, such simplification is performed by replacing the noisy channels by noiseless channels. For network source coding problems, replacing the lossy distortion criterion by lossless one is useful to simplify the problem. This approach also shows a surprising connection between Gaussian problems and network coding problems in wired networks.

Approximation is not only an approach to attack difficult theoretical problems in communication theory, but it is also useful from an engineering perspective. In particular, we gain insights about practical coding schemes and architecture from analysis of the simplified problem. It can be also used to identify the underlying difficulty of the practical problems. Moreover, an approximate solution is good enough in many engineering designs, since in theory we deal with models which can only approximate what happens in reality. From a practical point of view, the goal of the approximation is to design simple architecture for systems who perform close the the performance of the best

an ϵ-neighborhood of the achievable rate region with polynomial complexity in $1/\epsilon$.

schemes. However, further refinement for such proposed scheme is required in a real engineering problem, based on the specific application.

Seeking approximate solutions in network information theory is a counterpart for the approximation algorithms [4] in theoretical computer science. While most of the optimization problems, including those arising in important application areas are NP-hard, and therefore finding their exact solutions are prohibitively time consuming[2], developing polynomial time algorithms which can approximate the optimal solution (within a constant gap) becomes a compelling subject of scientific inquiry in computer science and mathematics. As a similar fact, an approximate perspective to problems in information theory can make them more obtainable.

The most important and fundamental problems in network information theory can be listed [5] as: multiple access channel [6,7] , broadcast channel [8], relay channel [9], interference channel [10], multiple description source coding [11], and distributed lossy compression [12]. These can be taught as basic and fundamental blocks of complex network communication problems. Except the multiple access channel, which is fully solved [6, 7], and the broadcast problem whose capacity is known for degraded channels (including Gaussian broadcast channel with arbitrary number of users) [13, 14], the other problems are still open, despite of many years of effort.

1.2 Gaussian Channel Approximation

Due to the nature of the wireless networks, a signal sent by a node (transmitter) is not only heard by the node of purpose, but also at any other node in proximity of the transmitter. In such situation, flow of information about a message plays a limiting role in communication, for the receiver which is not interested in decoding that message. These limiting and undesired signals (flows) are known as *interference* in the wireless communication literature. Information transmission in a shared medium is one of the fundamental problems in wireless communication. In such situation a wireless channel is shared between several sources/transmitters and receivers, and several information flow are competing for resources. Here, a fundamental question is how to manage interference in a wireless network.

A first question in this context is the study of the Gaussian interference channel, where the network consists of two pairs of source and receiver, each source wishes to communicate its message to its own destination over a shared environment (channel). Etkin *et al.* in [15] provided an approximate capacity characterization for this problem. This characterization includes upper bound for the capacity of the network, as well as encoding/decoding strategies based on Han-Kobayashi scheme [10] , which perform close to be optimal. Moreover, it is shown that the gap between the fundamental information-theoretic bound and what can be achieved using the proposed schemes is provably small. Therefore, the answer to the problem can be approximated by anything within

[2]Unless P=NP.

such narrow gap, although the exact answer is still unknown. This result is further generalized to a wider class of interference channels in [16].

More recently, Avestimehr, Diggavi, and Tse considered the wireless relay networks in [17], wherein a message has to be transmitted from one transmitter to a receiver over a network with an arbitrary number of nodes and network topology. Characterization of the capacity of such network has been an open problem, even for very simple cases since 1979 [9] . They devised a new relaying strategy called *quantizing-and-forward*, which they demonstrated to be within a constant number of bits from any optimal scheme. In particular, they gave an approximate max-flow min-cut characterization for wireless network information flow.

The relay network covers the unicast problem, where the information generated by the source is of interest for a single decoder. It can be generalized, in a straight-forward manner, to the multi-cast scenario, wherein multiple receivers exist in the network, but they are all interested in decoding the same message. In a more general setup, not all the source messages are of interest for all the receivers. In this situation, called *multiple unicast scenario*, the problem is more challenging, since signals carry information about independent messages may get mixed over the wireless network, and become completely useless.

The simplest example in this class of networks is the interference channel problem, which has no relay (helper) node in its setup, as discussed above. However, generalization of this result to a network with possible relay nodes who can facilitate communication lies in a wider research area.

We formulate the relay-interference network in which different information flows are sent through intermediate nodes in the network, *i.e.,* transmission is performed over multiple hops. In particular, we studied the network shown in Figure 5.1 with two sources/transmitters, two relays (intermediate nodes), and two receivers, where each of them is only interested in one of the source messages. While the optimal and exact solution to the simpler problem with no relay node is still unknown, our best hope is to provide approximate solution to this problem. Our work establishes the first additive[3] approximate result in this context.

We apply the deterministic model proposed in [17] to this two-stage interference channel, where the goal is to accommodate multiple unicast flows over the network. The simple layered structure of the networks helps us to focus more on the transmission techniques, rather than synchronization issues, raised in a non-layered network.

Investigation of these networks, suggests a new insight about the transmission techniques, which can be applied in any network. We show that the known techniques to deal with interference, such as interference separation and interference suppression are useful to avoid or remove interference in different regimes. Use of interference alignment is also shown to be essential for some cases, even with two messages transmitted through the network. Moreover,

[3] Interesting results are known for the scaling law of capacity in ad hoc wireless networks [18, 19] which provide multiplicative approximation for the capacity region of large networks.

we discovered a new interference management technique, we term *interference neutralization* [20], in which interference is canceled over air, without the relays necessarily decoding the transmitted signals. This strategy can be used whenever there exist more than one path for the interference. In such situations, the signals carry information about a unique interfering message, can be tuned to act against each other in order to reduce their total limiting effect.

Next, we apply the ideas developed in the analysis of the deterministic relay-interference network to the Gaussian wireless problem to obtain an outer bound for the achievable rates on this network [21]. We propose an encoding scheme to imitate the transmission strategies used in the deterministic model, which gives us an inner bound for the rate region of our interest. We show that the gap between the inner and outer bounds obtained based on analysis of the deterministic problem is bounded above by some universal constant, independent of the problem parameters, and they lead to an approximate solution for the problem. This is the first step for a better understanding the relay-interference network, which is a general framework that can well model the wireless communication networks. Extension of this study to more general cases, would have significant impact on the design of future network generations.

1.3 Gaussian Source Approximation

The approximation for the Gaussian channel is performed by a deterministic model, assuming the bits of the received signal above the noise level are completely clean and bits below are completely useless. In a dual way, a lossy source coding problem with quadratic distortion measure can be approximated by a lossless counterpart problem, where all the bits in the binary expansion of the source above the required distortion have to be losslessly delivered to the destination.

The Multiple Description Source Coding

In this problem a single source is encoded into several descriptions, and sent to the receivers through perfect channels. However, each end user has access to only a subset of the channels, and has to reconstruct the original source with a certain level of quality. The goal is to design the descriptions such that the more description are available at a user, the better reconstruction can be performed. Understanding the relation between the statistical properties of the source, minimum amount of information has to be stored in each description, and the reconstruction quality of each user is the main scope of the multiple description (MD) problem.

The problem is first studied in [11,22], where a complete solution is derived for the case that there are only *two* descriptions, source is generated according to a Gaussian distribution, and a quadratic distortion measure is used to quantify the reconstruction quality. This problem has received significant attention since 1980 [23–29]. However, the 2-description quadratic Gaussian problem is still the only case for which the problem is fully solved.

We discovered connection between this problem and a lossless underlying problem, which is called multi-level diversity coding (MLD) problem [30]. In this new problem, a *family* of independent sources have to be encoded into the descriptions. Each user is assigned with a level which is an integer number, and the user at level k have access to an arbitrary subset of the descriptions of size k, and wishes to *losslessly* decode an *incremental* subset of the sources, *i.e.*, the sources $1, 2, \dots, k$. It is worth mentioning that the MLD problem is lossless, and all reconstructions are perfect, while MD problem is lossy and the main goal is to improve reconstruction quality (see Figure 3.1).

The *symmetric* multiple description (SMD) problem refers to the situation where the reconstruction quality only depends on the number of available descriptions. In contrast and in a general scenario, the quality of reconstruction can be different for each available subset of the descriptions, which is called *asymmetric* multiple description (AMD) problem. We make progress on the SMD problem with *arbitrary number of descriptions*, by providing a novel bound on the descriptions' rates required for a set of quality demands, and proposing a simple coding scheme which can be used by the encoder. This is shown that this scheme can approach the obtained fundamental limit [31]. We show that the desired rate-distortion region is sandwiched between two polytopes, between which the gap can be upper-bounded by constants depending on the number of descriptions, but independent of the descriptions' rates, and the distortion constraints. This approximation is based on the characterization of the symmetric MLD problem, which turns out to be the underlying lossless counterpart of the MD problem.

We also formulated the asymmetric multi-level diversity coding problem (AMLD), and solved it for three descriptions [32]. New techniques, such as source partitioning and jointly encoding of independent sources, have been shown to be necessary and optimal for this problem. Based on this analysis, we managed to approximate the MD rate region for a general set of distortion constraints in [33]. This analysis leads to an outer and lower bound for the rate region of our interest, which sandwich the optimal solution between two polytopes with constant gap. Without the insight we obtain from the solving the deterministic problem, developing such encoding scheme is difficult if not impossible.

The Distributed Lossy Source Coding

In the distributed lossy source coding problem, several correlated sources are observed at different nodes, and encoded separately. The obtained descriptions sent to a joint decoder through perfect channels, whose task is to reconstruct all the sources, each with a pre-determined quality. The main question here is to characterize the rate of the descriptions based on the the set of posed distortion constraints. This problem is only solved for two descriptions, jointly Gaussian sources, and quadratic distortion measure [34]. However, an approximation approach can be used to derive inner and outer bound for the rate region of the problem for various types of correlation [35, 36]. Again, this result is based on the study of the underlying lossless problem, and utilizes the obtained insights for bounding techniques as well as coding schemes. This Analysis of this problem is beyond the scope of this thesis, and only pointed it out here as another approximation result in network information theory.

1.4 Random versus Structured Coding

The work of Avestimehr *et. al* in [17] shows that a random operations at the relay nodes which map the received signals to the transmitting signals are optimal to achieve the capacity of the linear deterministic relay network. This can be translated to (approximate) optimality of use of random codes in the Gaussian relay network.

Also, the Han-Kobayashi scheme which is used as the encoding scheme in the 2-user interference channel [15] allows arbitrary input distributions for messages. Therefore, it can be shown that the random Gaussian input distribution is nearly optimal for this channel.

Both these results are consistent with the folklore in information theory that "Gaussian inputs are good for Gaussian problems" [5]. However, this is not true in general. One important observation through all these works is that random Gaussian coding can be significantly away from being optimal. Structured codes have to be used mostly when different messages/signals have to combined. In a Gaussian interference channel with more than two users, using Gaussian random code at the interfering transmitters, the aggregate interference may cause a significant lost in the performance of the system. But, if instead we use the same lattice [37, 38] code for the interfering users, then interference alignment [39,40] can be utilized. This is because the addition of codewords will remain on the lattice. Now the space in between the codewords is preserved for user 1 to transmit information [5]. Thus, unlike in the two-user case, Gaussian codes are no longer good when there are more users. It is shown in [41] that lattice codes can achieve the capacity of the many-to-one Gaussian to within constant gap independent of the channel gains.

A similar phenomenon happens in the relay-interference network. As mentioned before, in such networks, whenever there are more than one available path for an interfering signal to get to a receiver, these paths can be utilized to (partially) cancel the effect (power) of the interference over air, without post-processing at the receiver. In order to do this, the signals have to be strategically encoded at the transmitters and relay nodes such that the end-to-end interference get neutralized at the receiver. This requires a code for which the summation of two codewords be still a decodable codeword. Clearly, a random code does not satisfy this property, and therefore fails in neutralization, while this can be simply implemented using a structured lattice code [21].

The neutralization technique can be also used in distributed lossy source coding. The encoding scheme proposed in [36] is such that the interfering bits in the binary expansion of the sources get neutralized during the decoding process.

The optimal encoding scheme proposed for the asymmetric multi-level diversity coding problem consists of jointly encoding independent messages (information bits) using linear operations [32]. Roughly speaking, this is required since a single resource (description rate) should be used to carry different independent information for different users with their own side-information received from other available descriptions. Such joint encoding of bit streams will be translated to using a very structured code in the near optimal encoding scheme proposed for the asymmetric multiple description problem [33]. More precisely, the underlying code should have an additive group structure, that is, the addition of any two codewords remains in the codebook. It is clear that a random code does not satisfying this property, and hence, may cause a large gap from the outer bound.

As illustrated above, structured codes play an important role in many of the near optimal coding schemes. This can be understood in a very unified framework. A common phenomenon happens in all of the situations in which a structured codes has to be applied: A single resource (bandwidth in wireless link, description rate, etc.) has be shared between different sources of information which need to be transmitted/stored; If such resource is only used *alone* to reconstruct the encoded data, clearly it has to be physically partitioned between different demands using traditional approaches such as time-sharing, frequency division, etc. However if the role of this resource imposed by problem definition is to *help other resources* to reconstruct different sources, its physical partitioning can be very far from being optimal. In such scenarios, different sources can be encoded jointly using structured codes. Then although this resource might be not useful alone to reconstruct any of the source data, it can help other resources in various manners to decode different encoded data.

1.5 Organization

This thesis is organized in two parts. In Part I we study approximation in lossy data compression, and multiple description problem in particular. Our results for this problem are built on the lossless underlying problem, the multi-level diversity coding. The symmetric version of the multi-level diversity coding problem is a known result from the late 90's [30, 42]. We formulate the asymmetric multi-level diversity coding problem in Chapter 2. A complete solution is provided for this lossless problem for the case of three descriptions. We then build our result for the symmetric multiple description problem based on the symmetric multi-level diversity coding in Chapter 3. We derive inner and outer bounds for the rate region of the SMD problem. Showing a bounded universal gap between the inner and outer bounds, we derive an approximate solution for the problem. In particular, we show that a simple proposed scheme based on successive refinement performs within 1.48 bits/sample of any optimal scheme for any number of descriptions, when the symmetric rate point is considered. This will be generalized to the asymmetric multiple description problem in Chapter 4, where we provide an approximate characterization for the achievable rate region based on the result of Chapter 2.

The main goal in Part II is to understand the relay-interference problem. We formulate the problem and analyze it in the linear deterministic model in Chapter 5. Here, we provide a complete capacity region characterization for the deterministic network. A new technique called *interference neutralization* is introduced to manage interference. This technique is based on jointly encoding of two independent messages into a single signal broadcasting from a node in the network. It is shown that use of this technique is necessary to to achieve the capacity of the network for some regimes of channel parameters. This understanding will be later used in Chapter 6, where we analyze the wireless Gaussian relay-interference network. We derive inner and outer bounds for the capacity region of this network. Bounding the gap between these two regions, we provide an approximate capacity region for the Gaussian network.

The interference neutralization technique is used in Chapters 5 and 6 in order to provide enough degrees of freedom for two legitimate messages to be received at their own destinations. This idea can be also used to neutralize a jamming signal sent through the network by an adversarial node, whose role is to distract data transmission from the source to the receiver in a wireless network. We study a diamond network with an adversarial jammer in Chapter 7. Here the source wishes to send its information to the receiver via two relay nodes. The signal received at the relays is corrupted by the jamming signal. However, we get benefit from existence of two paths for both legitimate and jamming signal to neutralize the jamming part. A complete capacity characterization is provided for the linear deterministic model of the network. The analysis is later generalized to the Gaussian network, wherein we derive lower and upper bounds with constant gap for the capacity of the wireless network.

Finally, we summarize the thesis in Chapter 8. The conclusion is followed by discussion on various possible direction for future works. Most of the technical parts as well as the proof desalt are presented at the end of the thesis in Appendices A-F. We rephrase the short claims before their proofs in the appendix to seek of completeness. However, we avoid repeating longer claims, and they are referred to the main chapter.

Part I

The Gaussian Lossy Source Coding

Overview

In this part we focus on network data compression problems, and explore approximation results for some long standing open questions. In particular, we focus on the multiple description (MD) source coding problem.

Consider a live movie provider over Internet, which streams its program(s) to a set of customers. Due to the packet loss in the network or overload on the server, a connection to a single server can fail, and therefore, parts of the movie never arrive to the customers. A strategy to improve the quality of service, is to use several servers to encode and stream the movie over the network, so that connection to any of them can provide enough packets to watch the movie for the customers. However, most of the time, the packets sent from multiple servers get reliably received to the users. A smart way to utilize such redundant packets is to use them to improve the quality of the movie. In order to that, a common data has to be encoded differently at several servers, so that connection to any of them allow the customer to watch the movie, and simultaneously, any connection to further servers can improve the quality of data reconstruction for the users.

A natural question arises here is to characterize the amount of data (rate) which has to be restored on each of the servers when a set of specific video qualities (distortion) is demanded by users who have reliable access to different subsets of the servers. This problem is terms as multiple description (MD) source coding. Of course, any answer to this question is based on various parameters, such as the statistical properties of stored data (source distribution) and how to quantify the quality of video (distortion measure). The only unique setting of the problem for which a complete rate-distortion characterization is known is when (i) only two servers (descriptions) involved, (ii) the source has a Gaussian distribution, (iii) and reconstruction quality is measured using quadratic distance [11, 22]. In fact, this is difficult and long standing problem in network information theory, which has been open since 1980.

In this part of the thesis, we focus on this problem with more than two descriptions, for Gaussian distributed source[4] and quadratic distortion. Though completely characterizing the rate-distortion region of the Gaussian multiple description problem is difficult, we provide an approximate characterization.

[4]Some of the results hold for general source distribution which are presented in Section 3.8

Underlying this approximation is a lossless problem, called multi-level diversity (MLD) coding.

The MD problem is called *symmetric* if the reconstruction quality demanded by each user only depends on the number of received descriptions, and called *asymmetric* otherwise. Our approximation result for the symmetric MD problem relies on the known results on the symmetric MLD problem [30, 42], which are reviewed in Section 2.1. We later use these results to establish inner and outer bounds for the optimal solution of the symmetric MD problem in Chapter 3. Moreover, we derive lower and upper bounds for the symmetric individual rate, that is when all the rates have to be the same. This resolves an open question to within 1 bit.

We formulated and analyzed the asymmetric version of the MLD problem for three descriptions in Chapter 2. This study reveals fundamental differences between the optimal coding schemes of the symmetric and asymmetric setting. Then, we use these results to attack the asymmetric MD problem in Chapter 4, where we again derive inner and outer bound for the achievable rate region. We further, show that the gap between these two bounds is small, which yields an approximate solution for the problem.

In both symmetric and asymmetric cases, we use simple encoding strategies to derive the lower bound. Our approximate characterization, shows that such simple schemes perform *close* to unknown optimal scheme. This provides a simple underlying architecture for designing such systems which can be of interest from an engineering point of view. However, a practical system design requires further refinements based on the specific application.

2 Multilevel Diversity Coding

With the explosive growth of packet transmission networks, such as the Internet, the transmission of information over unreliable (erasure) channels has recently received considerable attention. Such networks can be efficiently modeled as packet erasure channels [43]. The information source is encoded into a large number of packets, and transmitted over the network, where some of the packets get randomly erased, and the remaining are delivered to the receiver noiselessly. The decoder is expected to obtain a reconstruction of the source based on the packets received.

A *diversity coding* system involves an encoders who observes an information source and encodes it into different descriptions. There are multiple decoders, each with access to a certain subset of the generated descriptions. Each decoder has to reconstruct the source either perfectly or subject to a distortion criterion [44].

This can be understood as coding for erasure channel. The first work on diversity coding was done by Singleton [45] , who required that the decoders reconstruct the source perfectly.

The Read-Solomon codes [46] are a very first class of diversity coding systems, who convert an information sequence of length K from a large enough finite field into N symbols ($N \geq K$), such the original sequence can be perfectly reconstructed at any decoder with access to at least K of the generated symbols. This work is later generalized to the maximum distance separable (MDS) codes by Singleton [45].

The symmetric multilevel diversity coding (symmetric MLD or SMLD) is introduced by Roche [47] and Yeung [44]. In a multilevel diversity coding system, the set of decoders are partitioned into multiple levels. The reconstructions of the source by decoders within the same level are identical and are subject to the same distortion criterion.

In the symmetric (lossless) multi-level diversity coding problem [42], K source sequences are encoded into K descriptions, which are sent to the decoders through noiseless channels. The source sequences have certain levels of importance, indexed by $1, \ldots, K$, where the 1^{st} source sequence is the most important one. Each decoder has access to a non-empty subset of the descriptions. The encoders have to generate the descriptions with certain rates (lengths) such that each decoder with *any* k available descriptions be able to losslessly reconstruct the k most important source sequences. A trivial approach is to encode all the sources in each of the descriptions. However, the main goal here is to minimize the (weighted) sum of the rates (lengths) of the descriptions.

The symmetric MLD problem was motivated by fault-tolerant storage for disk arrays and for incremental priority encoding on packet erasure channels; see [42] for details.

The problem of symmetric multi-level diversity coding for three levels has been considered by Roche *et. al.* and the achievable rate region is completely characterized in [42]. This work has been later extended by Yeung and Zhang in [30], where they generalized the result for an arbitrary number of levels and descriptions. These two papers show that the source separation coding[1] is optimal for the symmetric problem. This means that each source sequence can be compressed separately, and then the descriptions are just produced by concatenating the compressed source sequences appropriately.

We formulate the asymmetric lossless multilevel diversity (AMLD) coding problem. The problem can be understood as a generalization of the symmetric MLD coding problem[2], and it is naturally applicable in distributed disk storage applications with asymmetric (unequal) reliabilities, in contrast to symmetric (equal) reliabilities which motivate the SMLD problem. In this example, users with access to different disks may have different demands. Similarly, for packet erasure applications, the erasure probabilities for the sub-packets may not be equal because the routes over which they are sent may have different reliabilities. As such, in both applications, we may wish to utilize not just the number of the encoders which are accessible, but also their identities, since the descriptions are no longer symmetric. Therefore, the difference between the SMLD and AMLD problem is that the levels of reconstruction are only determined by the number of received descriptions in the former case, while the specific combinations of descriptions available at the decoders determine the reconstruction levels in the latter case.

More precisely, $2^K - 1$ source sequences are encoded into K descriptions at the encoder. The $2^K - 1$ decoders are ordered in a specific way, and the goal of the encoder is to produce the descriptions such that the k-th decoder is able

[1] This coding scheme was originally called superposition coding, but here we adopt the name *source separation coding* as suggested by Raymond Yeung, in order to avoid confusion with the superposition coding in broadcast channels.

[2] A new notion of "ordering level" plays an important role in the characterization of the AMLD problem, which is defined in Definition 2.5. It can be shown that the symmetric MLD problem is subsumed by the AMLD problem for some of its possible ordering levels.

to reconstruct the k most important source sequences, for $k = 1, \ldots, 2^K - 1$, as shown in Figure 2.2.

Here, we only consider the 3-description AMLD case and provide a complete characterization of the achievable rate region. In particular we show that source-separation coding is *not* optimal for this problem, and the source sequences in different levels have to be jointly encoded (like in *network coding*) in an optimal coding strategy. We also show that the scheme using *linear* combination of these compressed sequences is optimal.

We note that various special cases of 3-description problem were studied in[3] [48], where, however, only no more than *three* information sources were considered. The characterization we provide in this work strictly subsumes those considered in [48].

The organization of the chapter is as follows. The first section is dedicated to the SMLD problem wherein we review the problem statement and the main known results from [30, 42]. Then, we formally define the AMLD problem in Section 2.2. A general achievable region characterization for the 3-description AMLD is presented in this section. We further specialize this result for a particular ordering. Next, we prove the converse part of the result for the AMLD problem in Section 2.3, and complete the proof by providing the achievability scheme and its analysis in Section 2.4.

2.1 The Symmetric Multilevel Diversity Coding

The symmetric MLD coding problem considered in [30, 42] can be described as follows. Independent and identically distributed samples of a total of K independent sources $V_1, V_2, \ldots, V_K$ drawn from a (finite) alphabet $\mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_K$ are observed at the encoder, and encoded into K descriptions.

2.1.1 Notation, Definitions, and Problem Formulation

We use V_i^n to denote a length n sequence of V_i, namely, $V_i^n = (V_{i,1}, \ldots, V_{i,n})$. Similarly, $\tilde{V}_i^n$ denotes the set of all n-tuples whose elements come from the alphabet $\tilde{V}_i$. The Shannon entropy of the i-th source is denoted by $H(V_i)$.

The decoders are indexed by binary vectors $\boldsymbol{v} \in \{0,1\}^K$. A decoder $\boldsymbol{v}$ has access to the i-th description if and only if $v_i = 1$. Let

$$\Omega_K^\alpha \triangleq \{\boldsymbol{v} \in \{0,1\}^K : |\boldsymbol{v}| = \alpha\},$$

be the set of decoders at *level* α, where $|\boldsymbol{v}|$ is the Hamming weight of the vector $\boldsymbol{v}$, and

$$\Omega_K \triangleq \bigcup_{\alpha=1}^{K} \Omega_K^\alpha.$$

[3] We would like to thank R. Yeung for bringing this work to our attention.

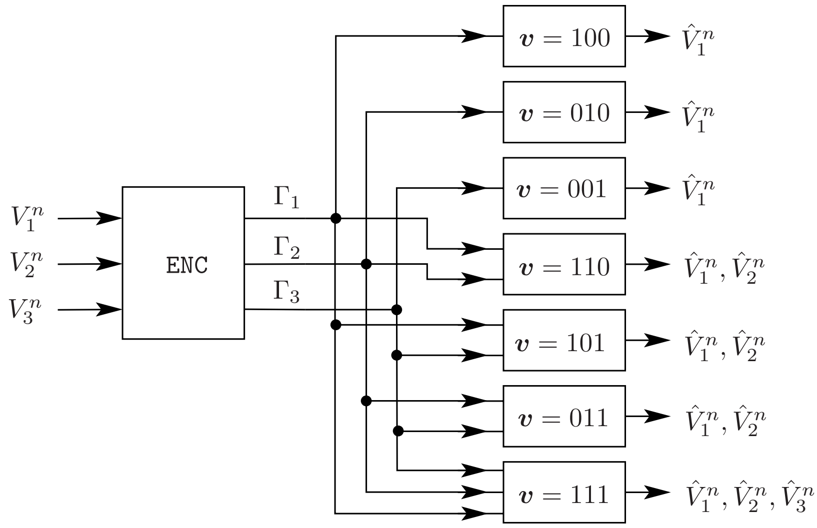


Figure 2.1: The 3-description symmetric multilevel diversity coding problem.

Moreover, for a given $\boldsymbol{v} \in \Omega_K$, and a set of K objects $\mathcal{G} = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_K\}$, we define $\Gamma_{\boldsymbol{v}}$ as the projection of $\mathcal{G}$ onto the the coordinates in $\boldsymbol{v}$, that is $\Gamma_{\boldsymbol{v}} = \{\Gamma_i : v_i = 1\}$. We also use short-hand notation $\mathcal{I}_K$ to denote the set $\{1, 2, \ldots, K\}$.

The descriptions generated at the encoder should be such that any decoder $\alpha \in \Omega_K^\alpha$ be able to perfectly reconstruct the first α source sequences, $V_1, V_2, \ldots, V_\alpha$, in the usual Shannon sense. Figure 2.1 depicts the setup of this problem for $K = 3$ descriptions.

A formal description of the problem is as follows. An $(n, \{M_i; i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\})$ code, where $\mathcal{I}_K = \{1, 2, ..., K\}$, is defined by its encoding functions F_i and decoding functions $G_{\boldsymbol{v}}$

$$F_i : \mathcal{V}_1^n \times \mathcal{V}_2^n \times \cdots \times \mathcal{V}_K^n \longrightarrow \{1, 2, \ldots, M_i\}, \quad i \in \mathcal{I}_K, \tag{2.1}$$

and decoding functions

$$G_{\boldsymbol{v}} : \prod_{j : v_j = 1} \{1, \ldots, M_j\} \longrightarrow \mathcal{V}_1^n \times \mathcal{V}_2^n \times \cdots \times \mathcal{V}_{|\boldsymbol{v}|}^n, \quad \boldsymbol{v} \in \Omega_K, \tag{2.2}$$

where $\prod$ denotes a set product. We define

$$\hat{V}_1^n(\boldsymbol{v}), \ldots, \hat{V}_{\boldsymbol{v}}^n(\boldsymbol{v}) = G_{\boldsymbol{v}}(F_i(V_1^n, \ldots, V_K^n); i : v_i = 1),$$

and

$$\Delta_{\boldsymbol{v}} = \mathbb{E}\left[\sum_{i=1}^{|\boldsymbol{v}|} d_H(V_i^n, \hat{V}^n(\boldsymbol{v}))\right],$$

where $\mathbb{E}$ is the expectation operator, and $d_H(\cdot,\cdot)$ is the Hamming distance, which gives the number of positions at which V_i^n and $\hat{V}^n(\boldsymbol{v})$ are different.

A K-tuple $(R_1, R_2, \ldots, R_K)$ is called SMLD-achievable if for every $\epsilon > 0$, there exists for sufficiently large n an $(n, (M_i, i \in \mathcal{I}_K), (\Delta_{\boldsymbol{v}}, \boldsymbol{v} \in \Omega_K))$ code such that

$$\frac{1}{n} \log M_i \leq R_i + \epsilon, \quad i \in \mathcal{I}_K,$$

and

$$\Delta_{\boldsymbol{v}} \leq \epsilon, \quad \boldsymbol{v} \in \Omega_K.$$

The goal is to characterize $\mathcal{R}_{\mathsf{SMLD}}$, the set of all achievable rates.

2.1.2 Main Result

The following theorem, extracted from [30] gives a full characterization for this rate region.

Theorem 2.1 ([30], Theorem 1)**.** $\mathcal{R}_{\mathsf{SMLD}}$ *is the set of all non-negative rate tuples* $\boldsymbol{R} = (R_1, \ldots, R_K) \geq 0$ *such that*

$$R_i = \sum_{\alpha=1}^{K} r_i^{\alpha}, \quad i = 1, 2, \ldots, K, \tag{2.3}$$

for some $r_i^{\alpha} \geq 0$*,* $\alpha = 1, 2, ..., K$ *such that*

$$\sum_{i:v_i=1} r_i^{|\boldsymbol{v}|} \geq H(V_{|\boldsymbol{v}|}), \quad \boldsymbol{v} \in \Omega_K. \tag{2.4}$$

The main message of this theorem is that source separation coding is in fact optimal for this problem. In this scheme, each source vector V_α^n is encoded independently of the other sources, and the i-th description is allocated rate r_i^{α} for the α-th source source V_α^n. Each description is then the collection of encoded information (codes) produced for all the sources. Clearly the equality in (2.3) can be replaced by $\geq$ without loss of generality.

2.2 The Asymmetric Multilevel Diversity Coding

In this section we formulate the asymmetric multilevel diversity (AMLD) coding problem. The problem can be understood as a refined version of the symmetric MLD coding problem.

The main difference between this problem and the symmetric MLD (SMLD) problem is that the descriptions are no longer symmetric. Therefore, the decoders' ability of reconstruction not only depend on the number of available descriptions, but also on the specific subset of received descriptions.

A K-description AMLD problem consists of $2^K - 1$ independent sources whose independent and identically samples are observed at a encoder. The encoder generates K descriptions based on the observed sources. The sources are ordered according to some predetermined importance.

The $2^K - 1$ decoders are ordered in a specific way, and the goal of the encoder is to produce the descriptions such that the k-th decoder is able to reconstruct the k most important source sequences, for $k = 1, \ldots, 2^K - 1$.

A new notion of *ordering* for the decoders plays an important role in the formulation of the AMLD problem, which is defined in Definition 2.5. It can be shown that the SMLD problem is subsumed by the AMLD problem for some specific ordering levels.

2.2.1 Notation, Definitions, and Problem Formulation

Let $\{(V_{1,t}, V_{2,t}, \ldots, V_{2^K-1,t})\}_{t=1,2,\ldots}$ be an independent and identically distributed process sampled from a finite size alphabet $\mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_{2^K-1}$ with time index t. This can be considered as $2^K - 1$ pieces of independent data streams, namely, $\{V_{1,t}\}, \ldots, \{V_{2^K-1,t}\}$, where each data stream is an independently and identically distributed sequence. The data streams are ordered with decreasing importance, e.g., consecutive refinements of a single source. We use V_i^n to denote a length n sequence of V_i, namely, $V_i^n = (V_{i,1}, \ldots, V_{i,n})$.

We also define the vector random variables U_j as $U_j \triangleq (V_1, \ldots, V_j)$ for $j = 1, \ldots, 2^K - 1$, and $U_0 \triangleq 0$. Similarly, U_j^n is used to denote length n sequences of U_j. We may simply use U^n to denote $U^n_{2^K-1} = (V_1^n, \ldots, V^n_{2^K-1})$ for brevity. Note that U_j^n is a two-dimensional array, whose elements are independent of each other along both directions, $i = 1, \ldots, j$, and $t = 1, \ldots, n$.

The Shannon entropy of the source V_k is denoted by $h_k \triangleq H(V_k)$. We also denote the entropy of U_j by H_j, where the independence of sources V_k's implies

$$H_j = H(U_j) = H(V_1, \ldots, V_j) = \sum_{i=1}^{j} H(V_i) = \sum_{i=1}^{j} h_i. \tag{2.5}$$

The AMLD problem can be described as follows. Consider $2^K - 1$ source sequences which are fed to a single encoder. The encoder produces K descriptions, denoted as $\Gamma_1, \Gamma_2 \ldots, \Gamma_K$ to encode the source sequences. The descriptions are sent over K perfect channels. There are $2^K - 1$ decoders, indicated by non-zero binary K-vector $\boldsymbol{v} \in \Omega_K$, each has access to a non-empty subset of the descriptions, $\Gamma_{\boldsymbol{v}}$, and wishes to losslessly decode the source data streams below a certain *level*, which is a function of the description set $\Gamma_{\boldsymbol{v}}$. Figure 2.2 illustrates the problem setting for $K = 3$, and a specific decoding requirement for the decoders.

Note that in the symmetric MLD problem, the decoders are naturally ordered according to the number available descriptions, whereas here in the asymmetric MLD problem, the decodability requirement for decoders with the same number of descriptions can be different. Therefore, we formally define the no-

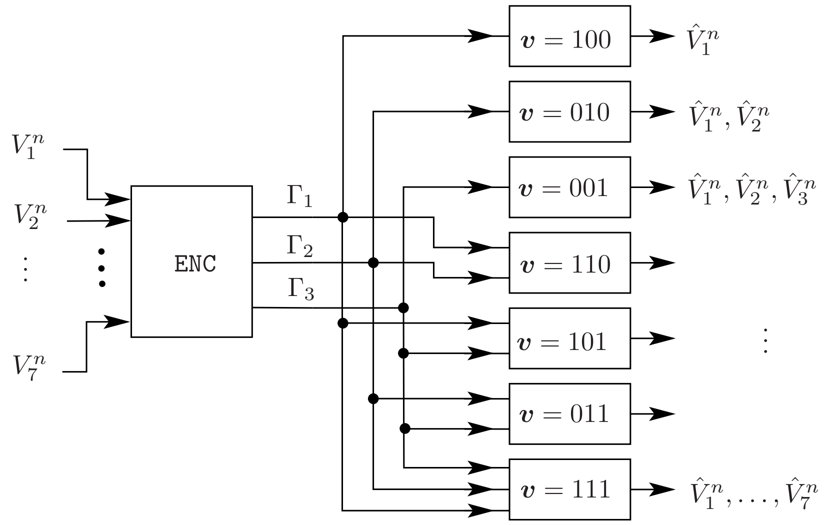


Figure 2.2: The 3-description asymmetric multilevel diversity coding problem for ordering level $\mathscr{L}_1$.

tion of *ordering* to connect the decoding requirement of the decoders to their available description subsets as follows.

Definition 2.2. *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors in $\mathbb{R}^K$. Define $\boldsymbol{u} \leq \boldsymbol{v}$ if and only if $u_i \leq v_i$ for $i = 1, \dots, K$. Similarly, we use $\boldsymbol{u} \geq \boldsymbol{v}$ to denote $u_i \geq v_i$ for $i = 1, \dots, K$. For binary vectors $\boldsymbol{u} \in \{0,1\}^K$ and $\boldsymbol{u} \in \{0,1\}^K$, we write $\boldsymbol{u} \leq \boldsymbol{v}$ if and only if $v_i = 0$ implies $u_i = 0$ for $i = 1, \dots, K$.*

Remark 2.3. *Note that using $\boldsymbol{v}, \boldsymbol{u} \in \{0,1\}^K$ as subset indicators over a set of size K, say $\mathcal{G} = \{\Gamma_1, \Gamma_2, \dots, \Gamma_K\}$, $\boldsymbol{u} \geq \boldsymbol{v}$ is equivalent to $\Gamma_{\boldsymbol{u}} \subseteq \Gamma_{\boldsymbol{v}}$.*

Definition 2.4. *For any $K \in \mathbb{N}$ and $1 \leq j \leq K$, we denote by $\mathbb{1}_j(K)$ the binary vector of length K with all zero elements except at the j-th position,* e.g., $\mathbb{1}_2(5) = 01000$.

Definition 2.5. *A valid* ordering level *(or simply* ordering*) on the decoders of an* AMLD *system, induced by an ordering function on their binary indicators, is a one-to-one mapping $\mathscr{L} : \Omega_K \longrightarrow \{1, \dots, 2^K - 1\}$ satisfying*

(i) $\mathscr{L}(\mathbb{1}_1(K)) < \mathscr{L}(\mathbb{1}_2(K)) < \cdots < \mathscr{L}(\mathbb{1}_K(K))$,

(ii) $\boldsymbol{u} \leq \boldsymbol{v}$ *implies* $\mathscr{L}(\boldsymbol{u}) < \mathscr{L}(\boldsymbol{v})$.

The ordering level will be used to determine the decoding requirements of the decoders, *e.g.,* a decoder with a set of descriptions indicated by $\boldsymbol{v}$ needs to decode the first $\mathscr{L}(\boldsymbol{v})$ source streams. Condition (i) is given to avoid permuted

Table 2.1: The ordering level $\mathscr{L}_1$ for $K = 3$ descriptions.

$\boldsymbol{v}$	100	010	001	110	101	011	111
$\mathscr{L}_1(\boldsymbol{v})$	1	2	3	4	5	6	7

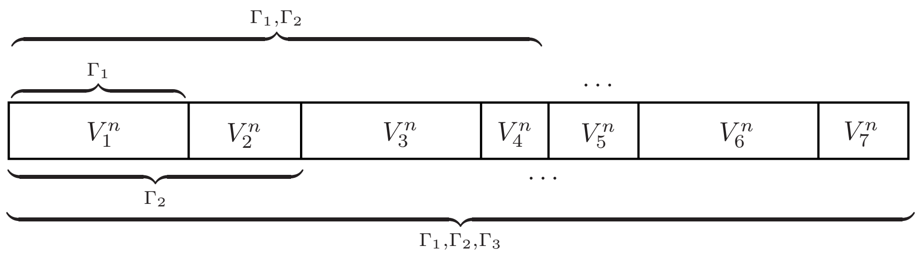


Figure 2.3: Levels assigned to the description subsets determines the recoverable source subsequences. The requirements corresponding to the ordering level $\mathscr{L}_1$ are shown in this figure.

repetition of the levels, where without loss of generality, we assume an initial ordering on the single description decoders. Condition (ii) is a natural fact that if $\Gamma_{\boldsymbol{u}}$ is a subset of $\Gamma_{\boldsymbol{v}}$, then the corresponding decoder can not do better than what decoder $\boldsymbol{v}$ can. The inverse mapping $\mathscr{L}^{-1}(k)$ is well defined, which is the binary sequence $\boldsymbol{v}$ whose ordering level is k.

The following example illustrates an ordering level. We will also use this particular ordering later when specializing the results for a particular ordering is needed for clarification.

Example 2.6. *For 3-description* AMLD*, the ordering level $\mathscr{L}_1$ is given in Table 2.1. The setting of the* AMLD *problem for the ordering level $\mathscr{L}_1$ is shown in Figure 2.2. Figure 2.3 shows the subset of source streams which should be recovered by each subset of descriptions in $\mathscr{L}_1$ setting.*

The following is a formal description of the AMLD problem. An $(n, \mathscr{L}, \{M_i; i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\})$ AMLD-code is defined by a set of encoding functions

$$F_i : \mathcal{V}_1^n \times \mathcal{V}_2^n \times \cdots \times \mathcal{V}_{2^K-1}^n \longrightarrow \{1, 2, \ldots, M_i\}, \quad i \in \{1, 2, \ldots, K\}, \tag{2.6}$$

and decoding functions

$$G_{\boldsymbol{v}} : \prod_{j:v_j=1} \{1, \ldots, M_j\} \longrightarrow \mathcal{V}_1^n \times \mathcal{V}_2^n \times \cdots \times \mathcal{V}_{\mathscr{L}(\boldsymbol{v})}^n, \quad \boldsymbol{v} \in \Omega_K. \tag{2.7}$$

We define

$$\hat{U}_{\mathscr{L}(\boldsymbol{v})}^n(\boldsymbol{v}) \triangleq G_{\boldsymbol{v}}(F_j(U^n); j : v_j = 1), \tag{2.8}$$

and $\hat{V}_i^n(\boldsymbol{v})$ is the corresponding part of $\hat{U}^n_{\mathscr{L}(\boldsymbol{v})}(\boldsymbol{v})$, for $i \leq \mathscr{L}(\boldsymbol{v})$. The difference between the reconstructed source sequence at decoder $\boldsymbol{v}$ and the original one is defined as

$$\Delta_{\boldsymbol{v}} = \mathbb{E} \sum_{i=1}^{\mathscr{L}(\boldsymbol{v})} d_H(V_i^n, \hat{V}^n(\boldsymbol{v})).$$

A rate tuple $\boldsymbol{R}^{\mathscr{L}} = (R_1, R_2, \ldots, R_K)$ is called achievable for a prescribed ordering $\mathscr{L}$, if for any $\epsilon > 0$ and sufficiently large n, there exists an $(n, \mathscr{L}, \{M_i, i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\})$ AMLD-code such that

$$\frac{1}{n} \log M_i \leq R_i + \epsilon, \quad i \in \mathcal{I}_K, \tag{2.9}$$

and

$$\Delta_{\boldsymbol{v}} \leq \epsilon, \quad \boldsymbol{v} \in \Omega_K. \tag{2.10}$$

The main goal in the asymmetric multilevel diversity coding problem is to characterize $\mathcal{R}_{\mathsf{AMLD}}$, the set of all achievable rate tuples $(R_i; i \in \mathcal{I}_K)$ in terms of the entropy rate of the source sequences and the given ordering level. We denote such a rate region by $\mathcal{R}^{\mathscr{L}}_{\mathsf{AMLD}}$ for a specific ordering.

In the rest of this chapter, we study the 3-description AMLD problem. We provide a complete characterization of the achievable rate region. As a consequence we show that unlike the symmetric MLD problem, source-separation coding is *not* optimal here, and the source sequences in different levels have to be jointly encoded (like in *network coding*) in an optimal coding strategy. We also show that the scheme using *linear* combination of these compressed sequences is optimal.

2.2.2 Rate Region Characterization for 3-Description AMLD

In this section we present the main results of the 3-description AMLD problem. We state the theorems in a unified way which hold for all orderings. We then specialize this rate region to the ordering $\mathscr{L}_1$ to facilitate understanding and further discussion.

Note that since the number of descriptions $K = 3$ is fixed here, with slightly abuse of notation, we may drop the variable K and use $\mathbb{1}_i$ instead of $\mathbb{1}_i(K)$.

Theorem 2.7. *Let $\boldsymbol{V} = (V_1, \ldots, V_7)$ be a given sequence of independent sources with entropy sequence $\boldsymbol{H} = (H_1, \ldots, H_7) = (H(V_1), H(V_1, V_2), \ldots, H(V_1, \ldots, V_7))$. For a given ordering level $\mathscr{L}$, the rate region $\mathcal{R}^{\mathscr{L}}_{\mathsf{AMLD}}(\boldsymbol{H})$ is the* *set of all non-negative triples* (R_1, R_2, R_3) *which satisfy*

$$R_i \geq H_{\mathscr{L}(\mathbb{1}_I)}, \qquad i = 1,2,3 \tag{AMLD-1}$$

$$R_i + R_j \geq H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}, \qquad i \neq j \tag{AMLD-2}$$

$$\begin{aligned} 2R_i + R_j + R_k \geq\ & H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}} \\ & + H_{\min\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)\}} + H_{\mathscr{L}(111)}, \qquad i \neq j \neq k \end{aligned} \tag{AMLD-3}$$

$$R_1 + R_2 + R_3 \geq H_{\mathscr{L}(100)} + H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} + H_{\mathscr{L}(111)}, \tag{AMLD-4}$$

$$\begin{aligned} R_1 + R_2 + R_3 \geq\ & H_{\mathscr{L}(100)} + \frac{1}{2}H_{\mathscr{L}(010)} + \frac{1}{2}H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}} \\ & + H_{\mathscr{L}(111)}. \end{aligned} \tag{AMLD-5}$$

In the following corollary, we specialize the bounds for the specific ordering $\mathscr{L}_1$.

Corollary 2.8. *For the ordering level* $\mathscr{L}_1$*, the achievable rate region of the* 3*-description* AMLD *problem is given by the set of all rate triples* (R_1, R_2, R_3) *which satisfy* (Q1)-(Q11)*, on the top of the next page.*

In the following we first prove the converse part of Theorem 2.7 for all orderings, and then the achievability part for ordering $\mathscr{L}_1$. Similar techniques can be used straightforwardly to prove the achievability for all the other orderings, and therefore complete the proof of Theorem 2.7.

2.3 AMLD: The Converse Proof

In this section we show that any achievable rate triple satisfies (AMLD-1)-(AMLD-5). The following important lemma, which simplifies the proof of the theorem, relates the entropy of the original source to the reconstructed one.

Lemma 2.9. *Let* $\boldsymbol{v} \in \Omega_3$ *be the indicator of one of the decoders,* $\Gamma_{\boldsymbol{v}} \subseteq \{\Gamma_1, \Gamma_2, \Gamma_3\}$ *be the corresponding subset of descriptions available at the decoder, and* $i \leq j \leq \mathscr{L}(\boldsymbol{v})$*. Then*

$$H(\Gamma_{\boldsymbol{v}}|U_i^n) \geq H(\boldsymbol{v}|U_j^n) + n(H_j - H_i - \delta_n) \tag{2.11}$$

where $\delta_n \to 0$ *as* n *increases.*

We will present the proof in Appendix A.1. Now, we are ready to prove the converse part of Theorem 2.7.

The converse proof of Theorem 2.7. Let (R_1, R_2, R_3) be any achievable rate triple. and $\mathbb{1}_i$ be a single description with ordering level $\mathscr{L}(\mathbb{1}_i)$. Recall that

$$\begin{aligned}
R_1 &\geq H(V_1), && (Q1)\\
R_2 &\geq H(V_1) + H(V_2), && (Q2)\\
R_3 &\geq H(V_1) + H(V_2) + H(V_3), && (Q3)\\
R_1 + R_2 &\geq 2H(V_1) + H(V_2) + H(V_3) + H(V_4), && (Q4)\\
R_1 + R_3 &\geq 2H(V_1) + H(V_2) + H(V_3) + H(V_4) + H(V_5), && (Q5)\\
R_2 + R_3 &\geq 2H(V_1) + 2H(V_2) + H(V_3) + H(V_4) + H(V_5)\\
&\quad + H(V_6), && (Q6)\\
2R_1 + R_2 + R_3 &\geq 4H(V_1) + 2H(V_2) + 2H(V_3) + 2H(V_4) + H(V_5)\\
&\quad + H(V_6) + H(V_7), && (Q7)\\
R_1 + 2R_2 + R_3 &\geq 4H(V_1) + 3H(V_2) + 2H(V_3) + 2H(V_4) + H(V_5)\\
&\quad + H(V_6) + H(V_7), && (Q8)\\
R_1 + R_2 + 2R_3 &\geq 4H(V_1) + 3H(V_2) + 2H(V_3) + 2H(V_4) + 2H(V_5)\\
&\quad + H(V_6) + H(V_7), && (Q9)\\
R_1 + R_2 + R_3 &\geq 3H(V_1) + 2H(V_2) + 2H(V_3) + H(V_4) + H(V_5)\\
&\quad + H(V_6) + H(V_7), && (Q10)\\
R_1 + R_2 + R_3 &\geq 3H(V_1) + 2H(V_2) + \frac{3}{2}H(V_3) + \frac{3}{2}H(V_4) + H(V_5)\\
&\quad + H(V_6) + H(V_7). && (Q11)
\end{aligned}$$

$U_0^n = 0$, and note that $\hat{U}^n_{\mathscr{L}(\mathbb{1}_i)}(\Gamma_i)$ is a function of Γ_i. Thus

$$\begin{aligned}
nR_i &\geq H(\Gamma_i) = H(\Gamma_i|U_0^n)\\
&\overset{(\star)}{\geq} H(\Gamma_i|U^n_{\mathscr{L}(\mathbb{1}_i)}) + n(H_{\mathscr{L}(\mathbb{1}_i)} - \delta_n)\\
&\geq n(H_{\mathscr{L}(\mathbb{1}_i)} - \delta_n). \qquad (2.12)
\end{aligned}$$

This proves (AMLD-1). Note that here and in the rest of this proof all the inequalities labeled by $(\star)$ are due to Lemma 2.9.

Toward proving (AMLD-2), we can write

$$\begin{aligned}
n(R_i + R_j) &\geq H(\Gamma_i) + H(\Gamma_j)\\
&\geq H(\Gamma_i|U_0^n) + H(\Gamma_j|U_0^n)\\
&\overset{(\star)}{\geq} H(\Gamma_i|U^n_{\mathscr{L}(\mathbb{1}_i)}) + H(\Gamma_j|U^n_{\mathscr{L}(\mathbb{1}_j)}) + n\left[H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} - 2\delta_n\right]\\
&\geq n\left[H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} - 2\delta_n\right] + H(\Gamma_i|U^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})\\
&\quad + H(\Gamma_j|U^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})\\
&\geq n\left[H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} - 2\delta_n\right] + H(\Gamma_i, \Gamma_j|U^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} - 2\delta_n\right] \\
&\quad + n\left[H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)} - H_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} - \delta_n\right] + H(\Gamma_i,\Gamma_j|U^n_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}) \\
&\geq n\left[H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)} - 3\delta_n\right]. \qquad (2.13)
\end{aligned}
$$

For proving (AMLD-3) we can start with

$$
\begin{aligned}
n(2R_i + R_j + R_k) &\geq 2H(\Gamma_i) + H(\Gamma_j) + H(\Gamma_k) \\
&\overset{(\star)}{\geq} [H(\Gamma_i|U^n_{\mathscr{L}(\mathbb{1}_i)}) + H(\Gamma_j|U^n_{\mathscr{L}(\mathbb{1}_j)}) + n(H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} - 2\delta_n)] \\
&\quad + [H(\Gamma_i|U^n_{\mathscr{L}(\mathbb{1}_i)}) + H(\Gamma_k|U^n_{\mathscr{L}(\mathbb{1}_k)}) + n(H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_k)} - 2\delta_n)] \\
&\geq n\left[2H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} + H_{\mathscr{L}(\mathbb{1}_k)} - 4\delta_n\right] \\
&\quad + H(\Gamma_i,\Gamma_j|U^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}}) + H(\Gamma_i,\Gamma_k|U^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}}) \\
&\overset{(\star)}{\geq} n\left[2H_{\mathscr{L}(\mathbb{1}_i)} + H_{\mathscr{L}(\mathbb{1}_j)} + H_{\mathscr{L}(\mathbb{1}_k)} - 4\delta_n\right] \\
&\quad + n\left[H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)} - H_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} - \delta_n\right] + H(\Gamma_i,\Gamma_j|U^n_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}) \\
&\quad + n\left[H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)} - H_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}} - \delta_n\right] + H(\Gamma_i,\Gamma_k|U^n_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)}) \\
&\geq n\Big[H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)} \\
&\qquad -6\delta_n\Big] + H(\Gamma_i,\Gamma_j,\Gamma_k|U^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}) \\
&\overset{(\star)}{\geq} n\Big[H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)} + H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)} \\
&\qquad -6\delta_n\Big] + n\left[H_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j+\mathbb{1}_k)} - H_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}} - \delta_n\right] \\
&= n\Big[H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}} + H_{\min\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}} \\
&\qquad + H_{\min\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}} + H_{\mathscr{L}(111)} - 7\delta_n\Big]. \qquad (2.14)
\end{aligned}
$$

Toward proving (AMLD-4) we can write

$$
\begin{aligned}
n(R_1+R_2+R_3) &\geq H(\Gamma_1) + H(\Gamma_2) + H(\Gamma_3) \\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)} + H_{\mathscr{L}(010)} + H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} - 3\delta_n\right] \\
&\quad + H(\Gamma_1|U^n_{\mathscr{L}(100)}) + H(\Gamma_2|U^n_{\mathscr{L}(010)}) + H(\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} \\
&\geq n\left[H_{\mathscr{L}(100)} + H_{\mathscr{L}(010)} + H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} - 3\delta_n\right] \\
&\quad + H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(010)}) + H(\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}) \\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)} + H_{\mathscr{L}(010)} + H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} - 3\delta_n\right] \\
&\quad + H(\Gamma_1,\Gamma_2|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}) \\
&\quad + n\left[H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}} - H_{\mathscr{L}(010)} - \delta_n\right] \\
&\quad + H(\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})
\end{aligned}
$$

$$
\begin{aligned}
&\geq n\left[H_{\mathscr{L}(100)}+2H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}-4\delta_n\right]\\
&\quad+H(\Gamma_1,\Gamma_2,\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})\\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)}+2H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}-4\delta_n\right]\\
&\quad+n\left[H_{\mathscr{L}(111)}-H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}-\delta_n\right]\\
&\geq n\left[H_{\mathscr{L}(100)}+H_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}+H_{\mathscr{L}(111)}-5\delta_n\right]. \qquad (2.15)
\end{aligned}
$$

We need to consider two different cases in order to obtain the other sum-rate bound in (AMLD-5). First consider the case $\mathscr{L}(001)>\mathscr{L}(110)$. Note that this implies $\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}=\mathscr{L}(110)$. We have

$$
\begin{aligned}
n(R_1+R_2+R_3) &\geq H(\Gamma_1)+H(\Gamma_2)+H(\Gamma_3)\\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)}+H_{\mathscr{L}(010)}+H_{\mathscr{L}(001)}-3\delta_n\right]\\
&\quad+H(\Gamma_1|U^n_{\mathscr{L}(100)})+H(\Gamma_2|U^n_{\mathscr{L}(010)})+H(\Gamma_3|U^n_{\mathscr{L}(001)})\\
&= n\left[H_{\mathscr{L}(100)}+H_{\mathscr{L}(010)}+H_{\mathscr{L}(001)}-3\delta_n\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_1|U^n_{\mathscr{L}(100)})+H(\Gamma_2|U^n_{\mathscr{L}(010)})\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_1|U^n_{\mathscr{L}(100)})+H(\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_2|U^n_{\mathscr{L}(010)})+H(\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\geq n\left[H_{\mathscr{L}(100)}+H_{\mathscr{L}(010)}+H_{\mathscr{L}(001)}-3\delta_n\right] \qquad (2.16)\\
&\quad+\frac{1}{2}\left[H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(010)})+H(\Gamma_1,\Gamma_3|U^n_{\mathscr{L}(001)})+H(\Gamma_2,\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)}+H_{\mathscr{L}(010)}+H_{\mathscr{L}(001)}-3\delta_n\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(110)})+n\left[H_{\mathscr{L}(110)}-H_{\mathscr{L}(010)}-\delta_n\right]\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_1,\Gamma_3|U^n_{\mathscr{L}(001)})+H(\Gamma_2,\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\overset{(a)}{\geq} n\left[H_{\mathscr{L}(100)}+\frac{1}{2}H_{\mathscr{L}(010)}+\frac{1}{2}H_{\mathscr{L}(110)}+H_{\mathscr{L}(001)}-\frac{7}{2}\delta_n\right]\\
&\quad+\frac{1}{2}\left[H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(001)})+H(\Gamma_1,\Gamma_3|U^n_{\mathscr{L}(001)})+H(\Gamma_2,\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\overset{(b)}{\geq} n\left[H_{\mathscr{L}(100)}+\frac{1}{2}H_{\mathscr{L}(010)}+\frac{1}{2}H_{\mathscr{L}(110)}+H_{\mathscr{L}(001)}-\frac{7}{2}\delta_n\right]\\
&\quad+H(\Gamma_1,\Gamma_2,\Gamma_3|U^n_{\mathscr{L}(001)})\\
&\overset{(\star)}{\geq} n\left[H_{\mathscr{L}(100)}+\frac{1}{2}H_{\mathscr{L}(010)}+\frac{1}{2}H_{\mathscr{L}(110)}+H_{\mathscr{L}(001)}-\frac{7}{2}\delta_n\right]\\
&\quad+n\left[H_{\mathscr{L}(111)}-H_{\mathscr{L}(001)}-\delta_n\right]
\end{aligned}
$$

$$= n\left[H_{\mathscr{L}(100)} + \frac{1}{2}H_{\mathscr{L}(010)} + \frac{1}{2}H_{\mathscr{L}(110)} + H_{\mathscr{L}(111)} - \frac{9}{2}\delta_n\right], \qquad (2.17)$$

where in (a) we have used $H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(110)}) \geq H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(001)})$, implied by the assumption $\mathscr{L}(110) > \mathscr{L}(110)$, and (b) is due to the conditional version of Han's inequality [3, page 491]. For the second case, *i.e.,* $\mathscr{L}(001) < \mathscr{L}(110)$, we have from (2.16),

$$\begin{aligned}
n(R_1+R_2+R_3) &\geq n\Big[H_{\mathscr{L}(100)} + H_{\mathscr{L}(010)} + H_{\mathscr{L}(001)} - 3\delta_n\Big]\\
&\quad + \frac{1}{2}\left[H(\Gamma_1,\Gamma_2|U^n_{\mathscr{L}(010)}) + H(\Gamma_1,\Gamma_3|U^n_{\mathscr{L}(\Gamma_3)}) + H(\Gamma_2,\Gamma_3|U^n_{\mathscr{L}(001)})\right]\\
&\overset{(\star)}{\geq} n\Big[H_{\mathscr{L}(100)} + H_{\mathscr{L}(010)} + H_{\mathscr{L}(001)} - 3\delta_n\Big]\\
&\quad + \frac{1}{2}\Big[H(\Gamma_1,\Gamma_2|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}})\\
&\qquad + H(\Gamma_1,\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}})\\
&\qquad + H(\Gamma_2,\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}})\\
&\qquad + n(3H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}} - H_{\mathscr{L}(\Gamma_2)} - 2H_{\mathscr{L}(001)} - 3\delta_n)\Big]\\
&\overset{(c)}{\geq} n\Big[H_{\mathscr{L}(100)} + \frac{1}{2}H_{\mathscr{L}(010)} + \frac{3}{2}H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}}) - \frac{9}{2}\delta_n\Big]\\
&\quad + H(\Gamma_1,\Gamma_2,\Gamma_3|U^n_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}})\\
&\geq n\left[H_{\mathscr{L}(100)} + \frac{1}{2}H_{\mathscr{L}(010)} + \frac{3}{2}H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}} - \frac{9}{2}\delta_n\right]\\
&\quad + n\left[H_{\mathscr{L}(111)} - H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}} - \delta_n\right]\\
&= n\Big[H_{\mathscr{L}(100)} + \frac{1}{2}H_{\mathscr{L}(010)} + \frac{1}{2}H_{\min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\}}\\
&\qquad + H_{\mathscr{L}(111)} - \frac{11}{2}\delta_n\Big]. \qquad (2.18)
\end{aligned}$$

Again we have used the conditional Han's inequality in (c). Putting (2.17) and (2.18) together, we obtain the bound (AMLD-5). □

2.4 AMLD: The Achievability Scheme

In the following we will show that the inequalities (AMLD-1)–(AMLD-5) provide a complete characterization of the achievable rate region of the AMLD problem. However, each individual case given in Table 2.2 needs to be considered separately, due to the specific strategy used in the coding scheme. We may further divide each ordering into sub-regimes corresponding to the relative order of the entropy of different sources in order to simplify the problem for each case. For conciseness, we only present the analysis for the ordering level

Table 2.2: The eight possible level orderings and corresponding sub-regimes.

Ordering	Regime
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(110) < \mathscr{L}(101) < \mathscr{L}(011) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_4 \leq h_3 \leq h_4 + h_5$
	$h_3 \geq h_4 + h_5$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(110) < \mathscr{L}(011) < \mathscr{L}(101) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_4 \leq h_3 \leq h_4 + h_5$
	$h_3 \geq h_4 + h_5$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(101) < \mathscr{L}(110) < \mathscr{L}(011) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_3 \geq h_4$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(101) < \mathscr{L}(011) < \mathscr{L}(110) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_3 \geq h_4$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(011) < \mathscr{L}(110) < \mathscr{L}(101) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_3 \geq h_4$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(011) < \mathscr{L}(101) < \mathscr{L}(110) < \mathscr{L}(111)$	$h_3 \leq h_4$
	$h_3 \geq h_4$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(110) < \mathscr{L}(001) < \mathscr{L}(101) < \mathscr{L}(011) < \mathscr{L}(111)$	$h_3 \leq h_5$
	$h_3 \geq h_5$
$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(110) < \mathscr{L}(001) < \mathscr{L}(011) < \mathscr{L}(101) < \mathscr{L}(111)$	$h_3 \leq h_5$
	$h_3 \geq h_5$

$\mathscr{L}_1$, and provide the details of the achievability scheme for all regimes of this specific ordering. More precisely, we show that any rate triple (R_1, R_2, R_3) satisfying $(Q1)$–$(Q11)$ is achievable, *i.e.,* there exist encoding and decoding functions with the desired rates which are able to reconstruct the required subset of the sources from the corresponding descriptions. This implies $\mathcal{R}^{\mathscr{L}_1}_{\mathsf{AMLD}}(\mathbf{H})$ is achievable, and completes the proof of the theorem for the ordering $\mathscr{L}_1$. Similar proof for other orderings can be straightforwardly completed by applying almost identical techniques. Different cases that needed to be considered are listed in Table 2.2.

Note that the $\mathcal{R}^{\mathscr{L}_1}_{\mathsf{AMLD}}$ is a polytopes specified by several hyper-planes in a three-dimensional space. Therefore, the region $\mathcal{R}^{\mathscr{L}_1}_{\mathsf{AMLD}}$ is a *convex* polytopes, and it suffices to show the achievability only for the *corner points* [49]; that is because a simple *time-sharing* argument can be used to extend the achievability to any arbitrary point in the region $\mathcal{R}^{\mathscr{L}_1}_{\mathsf{AMLD}}$. A complete list of corner points for the other orderings can be found in Appendix A.2.

Depending on the relationship of h_3, h_4, and h_5, some of the inequalities in $(Q1)$–$(Q11)$ may be dominated by the others. Note that $(Q10)$ and $(Q11)$ are of the form

$$\begin{aligned}R_1+R_2+R_3 \geq &3H(V_1)+2H(V_2)+2H(V_3)+H(V_4)+H(V_5)\\&+H(V_6)+H(V_7),\\R_1+R_2+R_3 \geq &3H(V_1)+2H(V_2)+\frac{3}{2}H(V_3)+\frac{3}{2}H(V_4)+H(V_5)\\&+H(V_6)+H(V_7).\end{aligned}$$

It is clear either one of them would be redundant and implied by the other, depending on whether $h_3 \lesseqgtr h_4$. Also if $h_3 \geq h_4+h_5$, inequalities $(Q3)$ and $(Q10)$

$$\begin{aligned}R_3 &\geq H(V_1)+H(V_2)+H(V_3),\\R_1+R_2+R_3 &\geq 3H(V_1)+2H(V_2)+2H(V_3)+H(V_4)+H(V_5)\\&\quad+H(V_6)+H(V_7)\end{aligned}$$

imply

$$\begin{aligned}R_1+R_2+2R_3 &\geq H_1+H_3+H_7+H_3\\&\geq H_1+H_3+H_7+H_2+h_4+h_5\\&= H_1+H_2+H_5+H_7,\end{aligned}$$

which is exactly the inequality given in $(Q9)$, *i.e.,* this inequality is redundant in this regime. Thus, we split the achievability proof into three regimes corresponding to the aforementioned conditions, since the proposed encoding schemes are slightly different for these regimes. We show the achievability of the corner points in each case.

To simplify matters, we perform a lossless *pre-coding*, acting on all the seven source sequences V_i^n's as

$$E_i : \mathcal{V}_i^n \longrightarrow \{0,1\}^{\ell_i}$$

for $i=1,\ldots,7$. This function maps the source sequence V_i^n to $\tilde{V}_i \triangleq E_i(V_i^n)$, which can be used as a new binary source sequence of length ℓ_i. This can be done by using any lossless scheme, and achieves ℓ_i arbitrary close to nh_i for large enough n. With the new source sequences $\tilde{V}_i$, we next perform further coding.

Regime I: $h_3 \geq h_4+h_5$
As mentioned above, the inequalities $(Q9)$ and $(Q11)$ are dominated by the others in this regime. Therefore we only need to consider the remaining nine hyper-planes. In the following we list the corner points of $\mathcal{R}^{\mathsf{AMLD}}_{\mathscr{L}_1}$ in this regime. Each corner point with coordinates (R_1,R_2,R_3) is the intersection of (at least) three hyper-planes, say (Qi), (Qj), and (Qk). Such point is denoted by $\langle Qi,Qj,Qk\rangle : (R_1,R_2,R_3)$. In order to list all the corner points, we first find the intersection of any three hyper-planes, and then check whether

the intersection point satisfies all the other inequalities. We next provide an encoding strategy to achieve the rates prescribed by the corner points of the polytopes.

- $X_1 = \langle Q1,Q4,Q7\rangle : (H_1, H_4, H_7)$

  This corner point is the intersection of the planes Q_1, Q_4, and Q_7, and determines the individual rates of the descriptions as

  $$(R_1, R_2, R_3) = (H_1, H_4, H_7).$$

  The scheme for achieving this rate tuple is as follows. Γ_1 is exactly the pre-coded sequence of V_1^n, *i.e.,* $\tilde{V}_1$. In order to construct Γ_2 it suffices to concatenate the codewords $\tilde{V}_1$, $\tilde{V}_2$, $\tilde{V}_3$, and $\tilde{V}_4$. Similarly, Γ_3 is the concatenation of all the seven codewords. That is,

  $$\Gamma_1 : \tilde{V}_1, \qquad \Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \qquad \Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7.$$

  It is easy to check that the description rates are the same as the rate triple of the corner point, and all the decoding requirements at the seven decoders are satisfied.

  We will only determine the rate triples and illustrate the descriptions construction for the remaining corner points.

- $X_2 = \langle Q1,Q5,Q7\rangle : (H_1, H_7 - h_5, H_5)$

  $$\Gamma_1 : \tilde{V}_1, \qquad \Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \tilde{V}_6, \tilde{V}_7, \qquad \Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \tilde{V}_5.$$

- $X_3 = \langle Q2,Q4,Q8\rangle : (H_1 + h_3 + h_4, H_2, H_7)$

  $$\Gamma_1 : \tilde{V}_1, \tilde{V}_3, \tilde{V}_4, \qquad \Gamma_2 : \tilde{V}_1, \tilde{V}_2, \qquad \Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7.$$

- $X_4 = \langle Q2,Q6,Q8\rangle : (H_1 + h_3 + h_4 + h_7, H_2, H_6)$

  $$\Gamma_1 : \tilde{V}_1, \tilde{V}_3, \tilde{V}_4, \tilde{V}_7, \qquad \Gamma_2 : \tilde{V}_1, \tilde{V}_2, \qquad \Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4, \tilde{V}_5, \tilde{V}_6.$$

- $X_5 = \langle Q3,Q5,Q10\rangle : (H_1 + h_4 + h_5, H_3 + h_6 + h_7, H_3)$

  The encoding schemes for the previous corner points only involve concatenation of different codewords. However, concatenation is not optimal to achieve the rate triple induced by the point X_5, and we need to jointly encode the sources to construct the descriptions. This can be done using a modulo-2 summation of (parts of) the codewords of the same size.

  The description Γ_1 is simply constructed by concatenating $\tilde{V}_1$, $\tilde{V}_4$, and $\tilde{V}_5$. Similarly, Γ_3 is obtained by putting $\tilde{V}_1$, $\tilde{V}_2$, and $\tilde{V}_3$ together. The second description, Γ_2, should be able to help Γ_1 to reconstruct $\tilde{V}_3$ at the decoder with access to $\{\Gamma_1, \Gamma_2\}$, and help Γ_3 to reconstruct $(\tilde{V}_4, \tilde{V}_5)$ at decoder $\{\Gamma_2, \Gamma_3\}$, where $\tilde{V}_3$ is already provided as a part of Γ_3. We

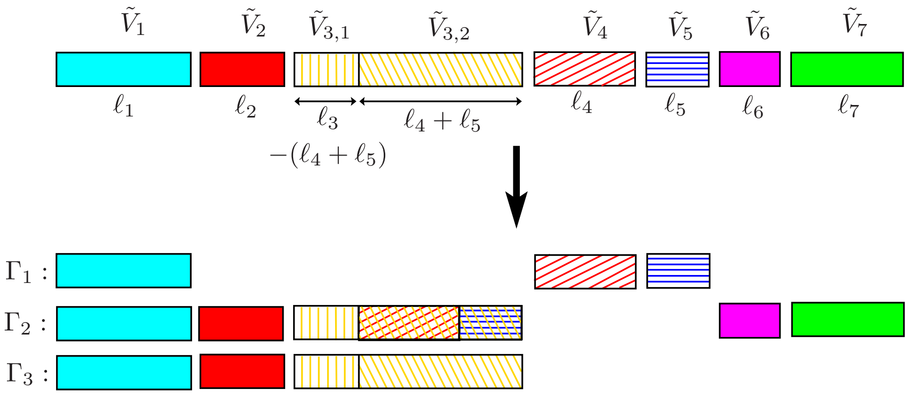


Figure 2.4: Linear encoding for the corner-point X_5

can use this fact to construct Γ_2 as follows. Partition[4] the bit stream $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3-(\ell_4+\ell_5)$ and $\ell_4+\ell_5$, respectively. Compute the modulo-2 summation (binary `xor`) of the bit-streams $\tilde{V}_{3,2}$ and $(\tilde{V}_4,\tilde{V}_5)$. The description Γ_2 is constructed by concatenating this new bit stream with $\tilde{V}_1$, $\tilde{V}_2$, $\tilde{V}_{3,1}$, $\tilde{V}_6$, and $\tilde{V}_7$.

The partitioning and encoding[5] are illustrated in Figure 2.4.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_5,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus (\tilde{V}_4, \tilde{V}_5), \tilde{V}_6, \tilde{V}_7,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

- $X_6 = \langle Q3{,}Q6{,}Q10\rangle : (H_1+h_3+h_7, H_2+h_4+h_5+h_6, H_3)$

  Partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3-(\ell_4+\ell_5)$ and $\ell_4+\ell_5$, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus (\tilde{V}_4, \tilde{V}_5), \tilde{V}_7,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_4, \tilde{V}_5, \tilde{V}_6,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

- $X_7 = \langle Q4{,}Q7{,}Q10\rangle : (H_1+h_4, H_3, H_3+h_5+h_6+h_7)$

[4] Since we are in regime I, we have $h_3 \geq h_4+h_5$ and hence , $\ell_3 \geq \ell_4+\ell_5$.

[5] Note that for this corner point, a part of the description Γ_2 is given by $\tilde{V}_{3,2}\oplus(\tilde{V}_4,\tilde{V}_5)$, which linearly combines independent (compressed) source sequences, just as the network coding idea in the familiar butterfly network [50].

Partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus \tilde{V}_4,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7.$$

- $X_8 = \langle Q4,\, Q8,\, Q10\rangle : (H_1 + h_3, H_2 + h_4, H_3 + h_5 + h_6 + h_7)$

  Partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus \tilde{V}_4,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_4,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7.$$

- $X_9 = \langle Q5,\, Q7,\, Q10\rangle : (H_1 + h_4, H_3 + h_6 + h_7, H_3 + h_5)$

  Partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_6, \tilde{V}_7,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_5.$$

- $X_{10} = \langle Q6,\, Q8,\, Q10\rangle : (H_1 + h_3 + h_7, H_2 + h_4, H_3 + h_5 + h_6)$

  Partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$ of lengths $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_{3,1}, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_7,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_4,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_5, \tilde{V}_6.$$

The associated rate region is shown in Figure 2.5.

Regime II: $h_4 \le h_3 \le h_4 + h_5$

In this regime, $(Q11)$ is dominated by $(Q10)$. Therefore, we only have to consider ten hyperplanes. The rates and encoding scheme for the corner points $Y_1 = \langle Q1,Q4,Q7\rangle$, $Y_2 = \langle Q1,Q5,Q7\rangle$, $Y_3 = \langle Q2,Q4,Q8\rangle$, $Y_4 = \langle Q2,Q6,Q8\rangle$, $Y_7 = \langle Q4,Q7,Q10\rangle$, $Y_8 = \langle Q4,Q8,Q10\rangle$, $Y_9 = \langle Q5,Q7,Q10\rangle$ and $Y_{10} = \langle Q6,Q8,Q10\rangle$ are exactly the same as that of X_1, X_2, X_3, X_4, X_7, X_8, X_9, and X_{10}, respectively. For the remaining corner points, we next provide the encoding schemes.

- $Y_5 = \langle Q3,\, Q5,\, Q9\rangle : (H_1 + h_4 + h_5, H_2 + h_4 + h_5 + h_6 + h_7, H_3)$

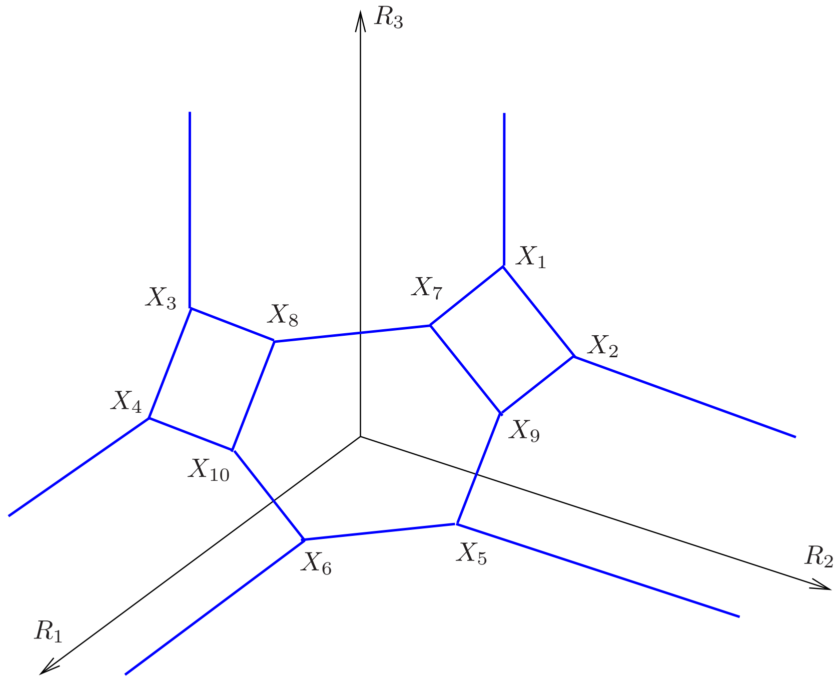


Figure 2.5: Rate region for Regime I: $h_3 \geq h_4 + h_5$

Partition $\tilde{V}_5$ into $\tilde{V}_{5,1}$ and $\tilde{V}_{5,2}$ of lengths $\ell_3 - \ell_4$ and $\ell_4 + \ell_5 - \ell_3$, respectively. Also partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$, of sizes $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_5,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_{3,1} \oplus \tilde{V}_{5,1}, \tilde{V}_{5,2}, \tilde{V}_6, \tilde{V}_7,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

- $Y_6 = \langle Q3,\, Q6,\, Q9 \rangle : (H_1 + h_4 + h_5 + h_7, H_2 + h_4 + h_5 + h_6, H_3)$

  Partition $\tilde{V}_5$ into $\tilde{V}_{5,1}$ and $\tilde{V}_{5,2}$ of lengths $\ell_3 - \ell_4$ and $\ell_4 + \ell_5 - \ell_3$, respectively. Also partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$, of sizes $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_5, \tilde{V}_7,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_{3,1} \oplus \tilde{V}_{5,1}, \tilde{V}_{5,2}, \tilde{V}_6,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

- $Y_{11} = \langle Q5,\, Q9,\, Q10 \rangle : (H_1 + h_3, H_3 + h_6 + h_7, H_2 + h_4 + h_5)$

Partition $\tilde{V}_5$ into $\tilde{V}_{5,1}$ and $\tilde{V}_{5,2}$ of lengths $\ell_3 - \ell_4$ and $\ell_4 + \ell_5 - \ell_3$, respectively. Also partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$, of sizes $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_{5,1},$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_{3,1} \oplus \tilde{V}_{5,1}, \tilde{V}_6, \tilde{V}_7,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{5,2}.$$

- $Y_{12} = \langle Q6,\, Q9,\, Q10\rangle : (H_1 + h_3 + h_7, H_3 + h_6, H_2 + h_4 + h_5)$

  Partition $\tilde{V}_5$ into $\tilde{V}_{5,1}$ and $\tilde{V}_{5,2}$ of lengths $\ell_3 - \ell_4$ and $\ell_4 + \ell_5 - \ell_3$, respectively. Also partition $\tilde{V}_3$ into $\tilde{V}_{3,1}$ and $\tilde{V}_{3,2}$, of sizes $\ell_3 - \ell_4$ and ℓ_4, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_{5,1}, \tilde{V}_7,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_{3,2} \oplus \tilde{V}_4, \tilde{V}_{3,1} \oplus \tilde{V}_{5,1}, \tilde{V}_6,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{5,2}.$$

Figure 2.6 shows the rate region for this regime.

Regime III: $h_3 \le h_4$

It is clear that in this regime $(Q10)$ is dominated by $(Q11)$, and thus $(Q10)$ does not affect the rate region. The remaining ten inequalities characterize the region. The rates and coding schemes for the points $Z_1 = \langle Q1,Q4,Q7\rangle$, $Z_2 = \langle Q1,Q5,Q7\rangle$, $Z_3 = \langle Q2,Q4,Q8\rangle$, and $Z_4 = \langle Q2,Q6,Q8\rangle$ are exactly the same as that of X_1, X_2, X_3, and X_4, respectively. The rate tuples and the corresponding descriptions for the other corner points are as follows.

- $Z_5 = \langle Q3,\, Q5,\, Q9\rangle : (H_1 + h_4 + h_5, H_2 + h_4 + h_5 + h_6 + h_7, H_3)$

  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$ and $\tilde{V}_{4,2}$ of lengths ℓ_3 and $\ell_4 - \ell_3$, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_5,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2}, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

- $Z_6 = \langle Q3,\, Q6,\, Q9\rangle : (H_1 + h_4 + h_5 + h_7, H_2 + h_4 + h_5 + h_6, H_3)$

  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$ and $\tilde{V}_{4,2}$ of lengths ℓ_3 and $\ell_4 - \ell_3$, respectively.

$$\Gamma_1 : \tilde{V}_1, \tilde{V}_4, \tilde{V}_5, \tilde{V}_7,$$
$$\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2}, \tilde{V}_5, \tilde{V}_6,$$
$$\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3.$$

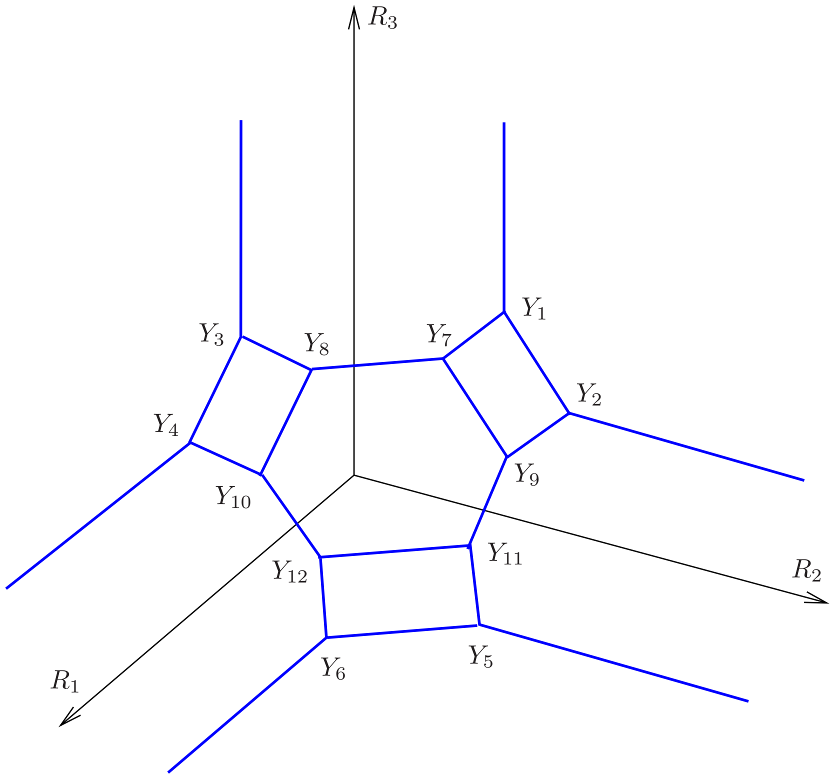


Figure 2.6: Rate region for Regime II of ordering level $\mathscr{L}_1$: $h_4 \le h_3 \le h_4 + h_5$

- $Z_7 = \langle Q4,\, Q7,\, Q8,\, Q11\rangle$:
  $\left(H_1 + \frac{h_3+h_4}{2}, H_2 + \frac{h_3+h_4}{2}, H_2 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7\right)$

  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$, $\tilde{V}_{4,2}$, and $\tilde{V}_{4,3}$ of lengths ℓ_3, $\frac{1}{2}(\ell_4 - \ell_3)$ and $\frac{1}{2}(\ell_4 - \ell_3)$, respectively.

  $$\begin{aligned}&\Gamma_1 : \tilde{V}_1, \tilde{V}_{4,1}, \tilde{V}_{4,2}\\ &\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2} \oplus \tilde{V}_{4,3},\\ &\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{4,3}, \tilde{V}_5, \tilde{V}_6, \tilde{V}_7.\end{aligned}$$

- $Z_8 = \langle Q5,\, Q7,\, Q9,\, Q11\rangle$:
  $\left(H_1 + \frac{h_3+h_4}{2}, H_2 + \frac{h_3+h_4}{2} + h_6 + h_7, H_2 + \frac{h_3+h_4}{2} + h_5\right)$

  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$, $\tilde{V}_{4,2}$, and $\tilde{V}_{4,3}$ of lengths ℓ_3, $\frac{1}{2}(\ell_4 - \ell_3)$ and $\frac{1}{2}(\ell_4 - \ell_3)$, respectively.

$$\begin{aligned}
&\Gamma_1 : \tilde{V}_1, \tilde{V}_{4,1}, \tilde{V}_{4,2}\\
&\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2} \oplus \tilde{V}_{4,3}, \tilde{V}_6, \tilde{V}_7,\\
&\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{4,3}, \tilde{V}_5.
\end{aligned}$$

- $Z_9 = \langle Q6,\ Q8,\ Q11\rangle$:
  $\big(H_1 + \frac{h_3+h_4}{2} + h_7, H_2 + \frac{h_3+h_4}{2}, H_2 + \frac{h_3+h_4}{2} + h_5 + h_6\big)$
  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$, $\tilde{V}_{4,2}$, and $\tilde{V}_{4,3}$ of lengths ℓ_3, $\frac{1}{2}(\ell_4-\ell_3)$ and $\frac{1}{2}(\ell_4-\ell_3)$, respectively.

$$\begin{aligned}
&\Gamma_1 : \tilde{V}_1, \tilde{V}_{4,1}, \tilde{V}_{4,2}, \tilde{V}_7\\
&\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2} \oplus \tilde{V}_{4,3},\\
&\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{4,3}, \tilde{V}_5, \tilde{V}_6.
\end{aligned}$$

- $Z_{10} = \langle Q6,\ Q8,\ Q11\rangle$:
  $\big(H_1 + \frac{h_3+h_4}{2} + h_7, H_2 + \frac{h_3+h_4}{2} + h_6, H_2 + \frac{h_3+h_4}{2} + h_5\big)$
  Partition $\tilde{V}_4$ into $\tilde{V}_{4,1}$, $\tilde{V}_{4,2}$, and $\tilde{V}_{4,3}$ of lengths ℓ_3, $\frac{1}{2}(\ell_4-\ell_3)$ and $\frac{1}{2}(\ell_4-\ell_3)$, respectively.

$$\begin{aligned}
&\Gamma_1 : \tilde{V}_1, \tilde{V}_{4,1}, \tilde{V}_{4,2}, \tilde{V}_7\\
&\Gamma_2 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3 \oplus \tilde{V}_{4,1}, \tilde{V}_{4,2} \oplus \tilde{V}_{4,3}, \tilde{V}_6,\\
&\Gamma_3 : \tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_{4,3}, \tilde{V}_5.
\end{aligned}$$

This region and its corner points are shown in Figure 2.7.

The coding schemes proposed for these three cases give us the achievability proof of the theorem for the specific ordering $\mathscr{L}_1$. As stated before, the coding scheme for other possible orderings listed in Table 2.2 are similar to that of the ordering $\mathscr{L}_1$. There are three main ingredients used in all of them; (i) converting the source sequences into bit-streams, (ii) partitioning the bit streams into sequences of proper length, and (iii) (if required) applying linear coding (binary `xor`) on them. This completes the proof of Theorem 2.7.

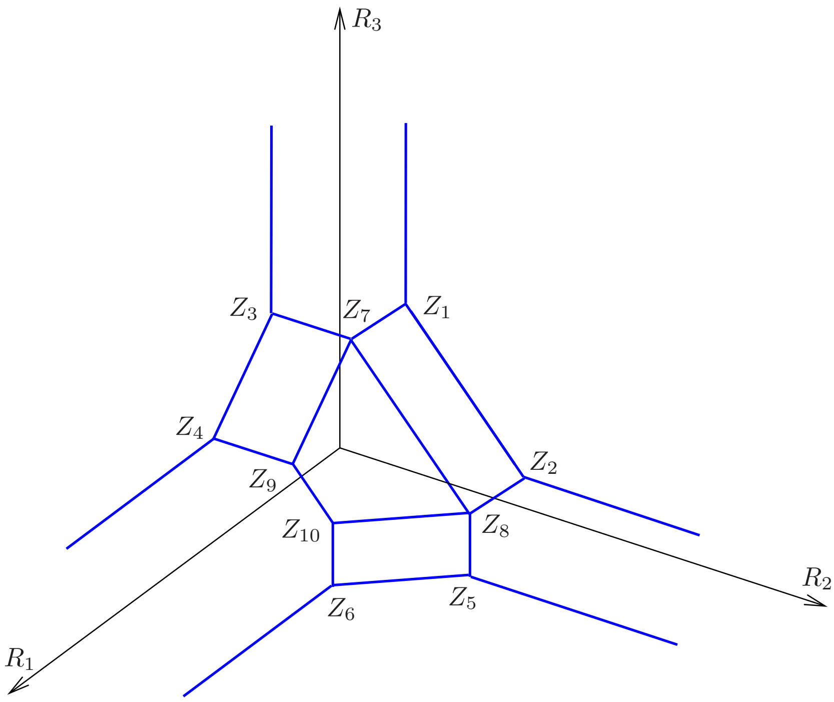


Figure 2.7: Rate region for Regime III of the ordering level $\mathscr{L}_1$: $h_3 \leq h_4$

3 Symmetric Multiple Description Coding

In this chapter we study the symmetric multiple description source coding problem. This problem can be understood as a lossy version of the symmetric multi-level diversity (SMLD) problem discussed in Section 2.1.

In the multiple description (MD) problem, a source is encoded into several descriptions such that any one of them can be used to reconstruct the source with certain quality, and more descriptions can improve the reconstruction. The problem is well motivated by source transmission over unreliable network and distributed storage systems, since there exists uncertainty as to which transmissions are received successfully (or which servers are accessible) by the end user.

In the early works on this problem, for example [11, 22], only two descriptions are considered. Even in this setting, the quadratic Gaussian problem is the only completely solved case [22], for which the achievable rate-distortion region in [11] is tight. Through a counter-example, Zhang and Berger showed that this achievable region is however not tight in general [23], and a complete characterization of the rate-distortion region has not been found to this date. See [24] (and the references therein) for a review of works related to this problem in the information theory literature.

Recent research attention has shifted to the general K-description problem, partly motivated by the availability of multiple transmission paths in modern communication networks. In [25, 26], an achievable scheme was provided for symmetric multiple descriptions, where each description has the same rate, and the distortion constraint depends only on the number of descriptions available. This scheme is based on joint binning of the codebooks, which has a similar flavor as the method often used in distributed source coding problems. Another achievable region was given in [27] using more conventional conditional codebooks. Wang and Viswanath [28, 29] generalized the Gaussian MD problem to

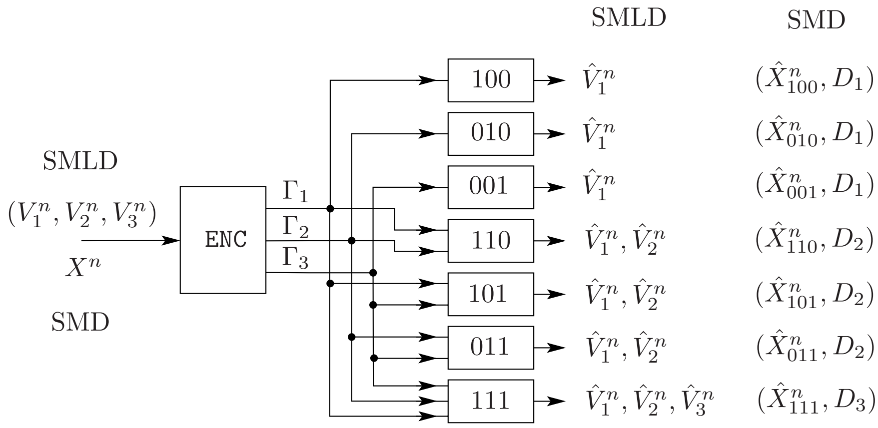


Figure 3.1: SMLD and SMD coding system diagrams for $K = 3$.

vector Gaussian source with many descriptions, and tight sum rate lower bound was established for certain cases with only *two levels* of distortion constraints (see also the outer bound result in [27]).

Here, we consider general multiple description coding with K descriptions under *symmetric distortion constraints*. The distortion constraints are symmetric in the sense that the distortion constraint at each decoder only depends on the number of available descriptions. More precisely, with any $k \leq K$ descriptions, the reconstruction has to satisfy the distortion D_k, regardless of which specific combination of k descriptions is used. Though the distortion constraints are symmetric, the rates of the descriptions are not necessarily the same in this setting, thus generalizing the case treated in [25,26]. Nevertheless the completely symmetric case is indeed an important special case, which we will treat with particular care. Our main focus is on the Gaussian source under the mean squared error (MSE) distortion measure, however the results are also extended to more general sources under the same distortion measure.

Though completely characterizing the rate-distortion region of the Gaussian multiple description problem is difficult, we provide an approximate characterization. Underlying this approximation is the lossless symmetric multi-level diversity (SMLD) coding problem previously studied in [30,42] and discussed in in Section 2.1; see Figure 3.1. The SMLD coding problem can be interpreted as a lossless version of the SMLD problem, and thus one of our main insights is to use the SMLD rate region as a polytopic template for the inner and outer bounds of the SMD rate-distortion region. We show that the SMD rate-distortion region can be sandwiched between two polytopes, between which the gap can be upper bounded by constants dependent on the number of descriptions, but independent of the exact distortion constraints.

One of our main contributions [51] is a novel lower bound to the sum rate for the Gaussian source, under K levels of symmetric distortion constraints.

This generalizes previous results in [22, 28, 29], where only two levels of distortion constraints are enforced in the system. Though the lower bound given here may not be tight as far as we know, it is the first general outer bound when more than two levels of distortion constraints are enforced. We derive this lower bound by generalizing Ozarow's technique in treating the Gaussian two-description problem. More specifically, we expand the probability space of the original problem by more than one auxiliary random variable, and impose certain Markov structure on these random variables. Ozarow's technique has been applied to various problems besides the MD problem [22, 27–29], for example, the multi-terminal source coding problem by Wagner and Anantharam [52], and the joint source channel coding problem by Reznic *et al.* [53]. However, in all these previous works the probability space was expanded by only one additional auxiliary random variable (or one additional auxiliary random vector in [28, 29] since vector source was considered), in contrast with the more general expansion we use here.

Though the SMD sum rate lower bound we derive can be optimized over $K-1$ variables to provide the tightest bound, such an explicit solution for this optimization problem appears difficult, thus we instead choose a specific set of values to provide a sub-optimal lower bound, which nevertheless still offers insight on the problem and allows us to give an approximate characterization of the SMD rate-distortion region.

For the inner bounds, we analyze two achievability schemes: the first is a simple scheme based on successive refinement coding [54–57] coupled with multi-level diversity coding [30, 42, 44, 58], which we call SR − MLD scheme; the second is a generalization of the multilayer coding scheme proposed by Puri, Pradhan and Ramchandran [25, 26], which we will refer to as the PPR multilayer scheme. In the special case of symmetric rate, the first scheme reduces to the unequal loss protection method [43, 59], and we thus also refer to it as the SR − ULP scheme. However, our scheme is more general, in the sense that it works for arbitrary rates, while the distortion constraints are still symmetric. The SR − MLD (or SR − ULP) scheme is a separation-based scheme where the quantization step and lossless source coding step are performed separately, as illustrated in Figure 3.2.

With the inner and outer bounds, we quantify the difference between them. For the symmetric rate problem, the individual-description rate-distortion function can be bounded within a constant. Moreover, regardless of the number of descriptions, the gap between the lower bound and the upper bound using the SR − ULP coding scheme is less than 1.48 bits, and for the PPR multilayer scheme, the gap is less than 0.92 bit.

It is surprising that the simple separation-based scheme is able to achieve performance which is within an additional constant from the optimal scheme. This constant gap is universal, and independent from the expecting rates and distortion constraints. The result implies that this simple scheme may be sufficient in certain practical high rate applications, where additional gain requires more complicated system design. Moreover, when distortion constraints are placed only on the last k levels for the decoders with $K-k+1, K-k+2, ..., K$

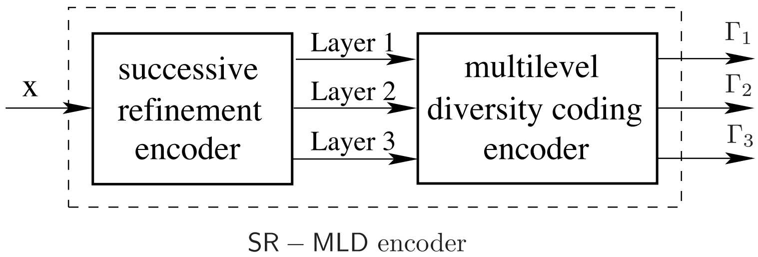


Figure 3.2: The separation approach based on successive refinement and lossless multi-level diversity coding for $K = 3$.

descriptions, we show that even the gap between the lower bound and the upper bound on the sum rate is asymptotically diminishing when the total number of descriptions K becomes large with k fixed. Thus virtually no gain is possible even in terms of sum rate for this case.

The results are then generalized to the K-description problem using the α-resolution approach [30], and the supporting hyper-planes of the rate region are bounded both from above and below between which the gap is bounded, yielding an approximate characterization of the rate-distortion region.

In the following, we first provide a formal definition of the SMD problem in Section 3.1. Since, our results are based on some previously known works, we briefly review them in Section 3.2. Section 3.3 summarizes the main results of the SMD problem for Gaussian sources. Sections 3.4-3.7 are dedicated to prove these results, and an extension of some of the results to arbitrarily distributed sources is presented in Section 3.8. Detailed and technical proofs of this chapter are given in the Appendix B.

3.1 SMD Problem: Notation and Formal Definition

Let $\{X(i)\}_{i=1,2,...}$ be a memoryless stationary source. At each time index i, the random variable $X(i)$ in an alphabet $\mathcal{X}$ is governed by the same distribution law μ_X, and $X(i)$ and $X(j)$ are independent for $i \neq j$. We assume that $\mathcal{X} = \mathbb{R}$, *i.e.*, the real alphabet; moreover the reconstruction alphabet is also usually assumed to be $\hat{\mathcal{X}} = \mathbb{R}$. The vector $X(1), X(2), ..., X(n)$ will be denoted as X^n. Capital letters are used for random variables, and lower-case letters are used for their realizations. Let $d : \mathcal{X} \times \mathcal{X} \to [0, \infty]$ be a single-letter distortion measure, and the multi-letter extension is defined as

$$d(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} d(x(i), y(i)).$$

We are particularly interested in the squared error distortion measure $d(x, y) = (x - y)^2$. As such, it will be assumed without loss of generality that the source has a normalized unit variance. In this context, the most important case is the zero-mean unit-variance Gaussian source $X \sim \mathcal{N}(0,1)$. In fact for the majority of this chapter we shall only consider this Gaussian source, except stated otherwise explicitly.

For the general K-description problem being considered, a length-n block of the source samples is encoded into K descriptions. Let $\vec{v}$ be a vector in $\{0,1\}^K$, and denote the i-th component of $\vec{v}$ by v_i. The sets Ω_K^α for $\alpha = 1, \ldots, K$ and Ω_K are defined as in Section 2.1. Essentially, the set Ω_K provides a compact way to enumerate the possible combinations of the descriptions, or equivalently a compact way to enumerate the possible decoders. Decoder $\vec{v}$, $\vec{v} \in \Omega_K$ has access to the $|\vec{v}|$ descriptions in the set $\Gamma_{\vec{v}} = \{\Gamma_i : v_i = 1\}$.

The *symmetric* distortion constraints are given such that any decoder $\vec{v}$ can reconstruct the source to satisfy a certain distortion $D_{|\vec{v}|}$, i.e., the distortion constraint depends only on the number of descriptions the decoder has access to, but not the particular combination of descriptions.

Formally, an $(n, (M_i; i \in \mathcal{I}_K), (\Delta_{\vec{v}}; \vec{v} \in \Omega_K))$ code, where $\mathcal{I}_K = \{1, 2, ..., K\}$, is defined by its encoding functions F_i and decoding functions $G_{\vec{v}}$

$$F_i : \mathcal{X}^n \to \{1, 2, \ldots, M_i\}, \quad i \in \mathcal{I}_K$$
$$G_{\boldsymbol{v}} : \prod_{i:v_i=1} \{1, 2, \ldots, M_i\} \to \mathcal{X}^n, \quad \vec{v} \in \Omega_K,$$

and

$$\Delta_{\boldsymbol{v}} = \mathbb{E} d(X^n, \hat{X}^n_{\boldsymbol{v}}), \quad \vec{v} \in \Omega_K,$$

where

$$\hat{X}^n_{\boldsymbol{v}} = G_{\boldsymbol{v}}(F_i(X^n); i : v_i = 1),$$

and $\mathbb{E}$ is the expectation operator.

A K-tuple $(R_1, R_2, ..., R_K)$ is $(D_1, D_2, ..., D_K)$-admissible if for every $\epsilon > 0$, there exists for sufficiently large n an $(n, \{M_i; i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\})$ code such that

$$\frac{1}{n} \log M_i \leq R_i + \epsilon, \quad i \in \mathcal{I}_K,$$

and

$$\Delta_{\boldsymbol{v}} \leq D_{|\boldsymbol{v}|} + \epsilon, \quad \boldsymbol{v} \in \Omega_K,$$

where the logarithms are of base 2, such that the rate is measured by bits. We denote by Γ_i the i-th description, which is in fact $F_i(X^n)$.

Let $\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$ be the collection of all $\boldsymbol{D}$-admissible rate vectors, that is

$$\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}) = \{(R_1, R_2, \ldots, R_K) : \\ (R_1, R_2, \ldots, R_K) \text{ is } (D_1, D_2, ..., D_K) - \text{admissible}\}, \tag{3.1}$$

where we shall assume $1 \geq D_1 \geq D_2 \geq ... \geq D_K > 0$ without loss of generality.

One important case of this problem is when the rates of all descriptions are the same, *i.e.*, $M_i = M$ for all $i \in \mathcal{I}_K$, for which the optimal symmetric individual-description rate-distortion ($\mathsf{SID} - \mathsf{RD}$) function is defined as

$$R_{\mathsf{SMD}}(\boldsymbol{D}) = \inf_{\substack{R: R \geq R_i \\ (R_1, R_2, ..., R_K) \in \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})}} R. \tag{3.2}$$

Since $\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$ is a closed set, the infimum can in fact be replaced by a minimum.

Approximation characterization of $\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$ and $R_{\mathsf{SMD}}(\boldsymbol{D})$ are our main goals of this chapter.

As a convention, when a rate R_{SMD} is of interest, we use $\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}$ or $\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}$ to denote its inner (upper) bounds, and use $\underline{R}$ to denote its outer (lower) bound; when rate region $\mathcal{R}_{\mathsf{SMD}}$ is of interest, the same convention is taken.

3.2 Preliminaries and Review of Known Results

In this section we briefly review the bounding techniques and coding schemes which we will use throughout this chapter.

3.2.1 Review of the PPR Multilayer Scheme

In the two-part paper [25] and [26], an achievable symmetric individual rate is given to satisfy symmetric distortion constraints. In the following the main theorem is quoted below together with a necessary definition.

Definition 3.1. *A joint distribution $p(\{y_{\alpha,j}, \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x)$ is called* symmetric *if for all $1 \leq n_i \leq K$, the following holds: conditioned on X, the joint distribution of Y_K and all $(n_1 + n_2 + ... + n_{K-1})$ random variables where any n_α are chosen from the set $\{Y_{\alpha,1}, Y_{\alpha,2}, ..., Y_{\alpha,K}\}$, is the same.*

Theorem 3.2 ([26] Theorem 2)**.** *For any probability distribution*

$$p(x, \{y_{\alpha,j}, \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K) = p(x)p(\{y_{\alpha,j}, \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x),$$

where $p(\{y_{\alpha,j}, \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x)$ is symmetric over $\mathcal{X} \times \mathcal{Y}^{K(K-1)+1}$ and the decoding functions

$$\begin{aligned} G_{\boldsymbol{v}} &: \mathcal{Y}^{|\boldsymbol{v}||\boldsymbol{v}|} \to \mathcal{X}, & \forall \boldsymbol{v} \in \Omega_K \text{ with } |\boldsymbol{v}| < K, \\ G_{\boldsymbol{v}} &: \mathcal{Y}^{K(K-1)+1} \to \mathcal{X}, & |\boldsymbol{v}| = K, \end{aligned}$$

such that

$$\begin{aligned} &\mathbb{E}[d(X, G_{\boldsymbol{v}}(Y_{\alpha,j}; \alpha \in \mathcal{I}_{|\boldsymbol{v}|}, j : v_j = 1))] \leq D_{|\boldsymbol{v}|}, \quad \forall \boldsymbol{v} \in \Omega_K \text{ with } |\boldsymbol{v}| < K, \\ &\mathbb{E}[d(X, G_{\boldsymbol{v}}(\{Y_{\alpha,j}, \alpha \in \mathcal{I}_{|\boldsymbol{v}|}, j \in \mathcal{I}_K\}, Y_K))] \leq D_K, |\boldsymbol{v}| = K, \end{aligned}$$

the following symmetric individual description rate is achievable

$$R = \sum_{\alpha=1}^{K-1} \frac{1}{\alpha} H(\{Y_{\alpha,j}; j \in \mathcal{I}_\alpha\}|\{Y_{i,j}; i \in \mathcal{I}_{\alpha-1}, j \in \mathcal{I}_\alpha\})$$
$$+ \frac{1}{K}\left[H(Y_K|\{Y_{i,j}; i \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}) - H(\{Y_{i,j}; i \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, Y_K|X)\right].$$

We notice that this definition of symmetry is however unnecessarily restrictive and can be straightforwardly relaxed. The following generalized version is thus introduced, which will be useful since our choice of distribution to simplify the inner bound is in this relaxed set, but not in the original more restrictive set.

Definition 3.3. *A joint distribution $p(\{y_{\alpha,j}; \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x)$ is called* generalized symmetric *if for any permutation $\pi(\cdot) : \mathcal{I}_K \to \mathcal{I}_K$, the joint distribution $p(\{y_{\alpha,\pi(j)}; \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x)$ is the same as $p(\{y_{\alpha,j}; \alpha \in \mathcal{I}_{K-1}, j \in \mathcal{I}_K\}, y_K|x)$.*

It is not difficult to check that Theorem 3.2 holds with the generalized version of symmetry. The original version essentially requires the distribution to be invariant under $K-1$ different permutations $\pi_\alpha(\cdot)$, one for each layer $\alpha = 1, 2, ..., K-1$; *i.e.*, if we permute $\{Y_{1,1}, Y_{1,2}, ..., Y_{1,K}\}$, and then permute $\{Y_{2,1}, Y_{2,2}, ..., Y_{3,K}\}$ differently, and so on for each $\alpha = 1, 2, ..., K-1$, the resulting distribution should remain the same as the one before such permutations. This requirement was however not completely utilized in the coding scheme. Instead it in fact only requires invariance under a single permutation $\pi(\cdot)$ which is applied to all the levels simultaneously, *i.e.*, $\pi_\alpha(\cdot) = \pi(\cdot)$, for $\alpha = 1, 2, ..., K-1$. More formally, we state our observation of the generalized result as a theorem.

Theorem 3.4. *The statement of Theorem 3.2 holds when the symmetric distribution requirement is replaced with the generalized symmetric distributions.*

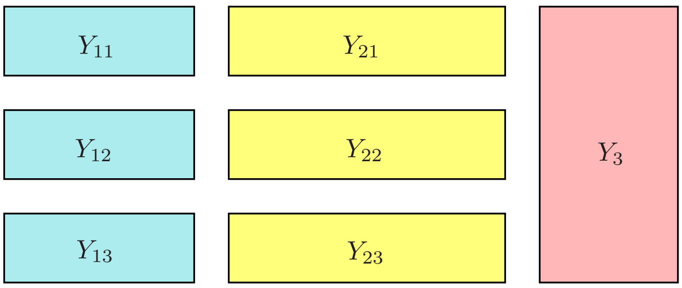


Figure 3.3: Layers of codes in the PPR scheme for $K = 3$.

For the purpose of this work, it is not necessary to change the random coding scheme in [25,26], however, the following properties regarding the overall structure of the codes are important:

- There are a total of K layers of codes, and the α-th layer information ($\alpha \neq K$) in the j-th description is represented by the random variables $Y_{\alpha,j}$; see Figure 3.3. Thus for a set of descriptions indicated by $\boldsymbol{v}$, the (single letter) decoding function $G_{\boldsymbol{v}}$ in Theorem 3.2 can utilize the random variables in the set $\{Y_{\alpha,j}, \alpha \leq |\boldsymbol{v}|, j : v_j = 1\}$. The lossless underlying problem for this part of the scheme is illustrated in Figure 3.4.

- For the j-th description, the overall rate is the rate summation of K layers of codes, the rate of the α-th layer $R_{\alpha,j} = \frac{1}{\alpha}\tilde{H}_\alpha(\boldsymbol{Y})$, where $\tilde{H}_\alpha(\boldsymbol{Y})$ is defined as

$$\begin{aligned}\tilde{H}_\alpha(\boldsymbol{Y}) \triangleq & h(Y_{\alpha,i}, i \in \mathcal{I}_\alpha | \{Y_{k,j}; k \in \mathcal{I}_{\alpha-1}, j \in \mathcal{I}_\alpha\}) \\ & - \frac{\alpha}{K} h(Y_{\alpha,k}; k \in \mathcal{I}_K | X, \{Y_{j,k}, j \in \mathcal{I}_{\alpha-1}, k \in I_K\})\end{aligned} \tag{3.3}$$

for $\alpha = 1, 2, \ldots, K-1$, and for $\alpha = K$, we define

$$\tilde{H}_K(\boldsymbol{Y}) \triangleq I(X; Y_K | \{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}). \tag{3.4}$$

Most importantly, the decoding is done sequentially, and decoding the first α layers of codes does not rely on the $(\alpha+1)$-th to the K-th layers of codes.

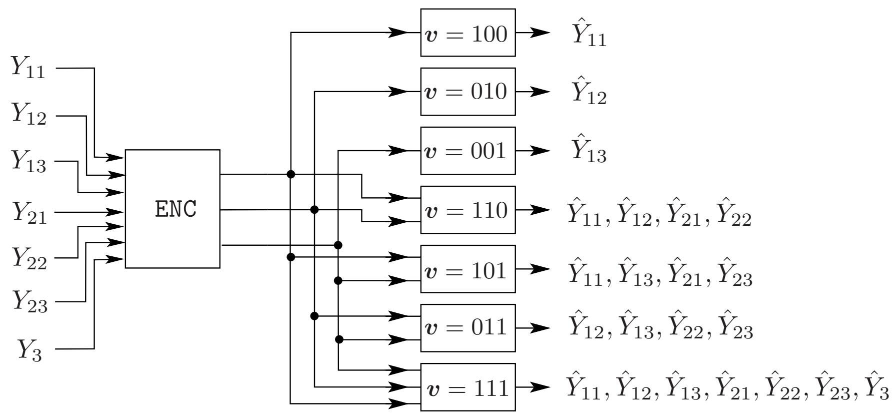


Figure 3.4: The lossless underlying problem for the PPR scheme for $K = 3$.

Note that the set of auxiliary random variables $\{Y_{\alpha,j}\}$ has more than K components, however we still write it as $\boldsymbol{Y}$ for conciseness.

3.2.2 Review of the α-Resolution Approach

As mentioned before, the result of this section is based on the SMLD coding problem, and its rate characterization in [30, 42] (see Section 2.1). The rate region characterization in (2.3) and (2.4) for the MLD coding problem is given in a parametrized form, i.e., involving variables more than the rate tuple of interest $(R_1, R_2, .., .R_K)$. Though for smaller K, it is possible to explicitly investigate its faces and vertex points, for larger value of K this becomes intractable. The α-resolution method was invented in [30] to reveal the inherent structure of this region. We now quote a few definitions and results from [30] and then give several necessary results presented in [31]; some further results will be given after related notation is introduced.

We keep using the partial order for real and binary vectors defined in Definition 2.2. For any $\boldsymbol{A} = (A_1, A_2, ..., A_K) \geq 0$, a mapping $c_\alpha : \Omega_K^\alpha \to \mathbb{R}_+$, where $\mathbb{R}_+$ is the set of non-negative real numbers, satisfying the following properties

$$c_\alpha(\boldsymbol{v}) \geq 0, \quad \text{for all} \quad \boldsymbol{v} \in \Omega_K^\alpha,$$

and

$$\sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})\boldsymbol{v} \leq \boldsymbol{A}$$

is called an α-resolution for $\boldsymbol{A}$; it will be denoted as $\{c_\alpha(\boldsymbol{v})\}$ or simply as c_α. Define a function $f_\alpha : \mathbb{R}_+^K \to \mathbb{R}_+$ for $\alpha \in I_K$ by

$$f_\alpha(\boldsymbol{A}) = \max \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v}), \tag{3.5}$$

where the maximum is taken over all the α-resolution of $\boldsymbol{A}$. If $\{c_\alpha(\boldsymbol{v})\}$ achieves $f_\alpha(\boldsymbol{A})$, then it is called an optimal α-resolution for $\boldsymbol{A}$, or simply α-optimal. Without loss of generality up to a permutation of the rate vector components, we may assume

$$A_1 \geq A_2 \geq ... \geq A_K.$$

Definition 3.5. *Let $\{c_\alpha(\boldsymbol{v})\}$ be an α-resolution of $\boldsymbol{A}$, then $\sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})\boldsymbol{v}$ is called the profile of $\{c_\alpha(\boldsymbol{v})\}$.*

The following lemmas and theorem are extracted form [30] and present some properties of the α-resolution. The corresponding proofs can be found in the original work.

Lemma 3.6 ([30], Lemma 1)**.** *Let $\{c_\alpha(\boldsymbol{v})\}$ be α-optimal for $\boldsymbol{A}$, and assume $(\breve{A}_1, \breve{A}_2, ..., \breve{A}_K)$ is its profile. If there exist $1 \leq i \leq K$ such that $A_i - \breve{A}_i > 0$, then $c_\alpha(\boldsymbol{v}) > 0$ implies $v_i = 1$.*

Lemma 3.7 ([30], Lemma 2)**.** *Let $\{c_\alpha(\boldsymbol{v})\}$ be α-optimal for $\boldsymbol{A}$, and assume $(\breve{A}_1, \breve{A}_2, ..., \breve{A}_K)$ is its profile, then there exists $0 \leq l_\alpha \leq \alpha - 1$ such that $A_i - \breve{A}_i > 0$ if and only if $1 \leq i \leq l_\alpha$.*

Definition 3.8. *For $2 \leq \alpha \leq K$, let c_α and $c_{\alpha-1}$ be α-optimal and $(\alpha-1)$-optimal for $\boldsymbol{A}$, respectively. Then $c_{\alpha-1}$ covers c_α, denoted by $c_{\alpha-1} \succ c_\alpha$, if*

$$\sum_{\boldsymbol{u} \in \Omega_K^{\alpha-1}} c_{\alpha-1}(\boldsymbol{u})H(\Gamma_{\boldsymbol{v}}) \geq \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})H(\Gamma_{\boldsymbol{u}}), \tag{3.6}$$

for any K jointly distributed random variable $\Gamma_1, \Gamma_2, ..., \Gamma_K$.

The following theorem is instrumental for the result presented in [30], and it is also important for us to establish the result on the MD rate-distortion region for $K > 3$.

Theorem 3.9 ([30], Theorem 3)**.** *For any $\boldsymbol{A} \geq 0$, there exist c_α, $1 \leq \alpha \leq K$, where c_α is α-optimal for $\boldsymbol{A}$, such that*

$$c_1 \succ c_2 \succ \cdots \succ c_K. \tag{3.7}$$

The following two lemmas were not in [30], but presented in [31] and play an important role in our analysis.

Lemma 3.10. *Let $c_{\alpha-1}$ and c_α be $(\alpha-1)$-optimal and α-optimal, respectively. If $c_{\alpha-1} \succ c_\alpha$, then $(\alpha-1)f_{\alpha-1}(\boldsymbol{A}) \geq \alpha f_\alpha(\boldsymbol{A})$.*

Proof of Lemma 3.10. Let $S_1, S_2, ..., S_K$ be independently and identically distributed random variables with entropy $H(S_i) = H(S) > 0$ for any $i \in I_K$, then it follows

$$\begin{aligned}
(\alpha-1)f_{\alpha-1}(\boldsymbol{A})H(S) &= \sum_{\boldsymbol{u}\in\Omega_K^{\alpha-1}} c_{\alpha-1}(\boldsymbol{u})(\alpha-1)H(S) \\
&= \sum_{\boldsymbol{u}\in\Omega_K^{\alpha-1}} c_{\alpha-1}(\boldsymbol{u})H(S_{\boldsymbol{u}}) \\
&\geq \sum_{\boldsymbol{v}\in\Omega_K^{\alpha}} c_\alpha(\boldsymbol{v})H(S_{\boldsymbol{v}}) \\
&= \sum_{\boldsymbol{v}\in\Omega_K^{\alpha}} c_\alpha(\boldsymbol{v})\alpha H(S) \\
&= \alpha f_\alpha(\boldsymbol{A})H(S).
\end{aligned}$$

Dividing both ends by $H(S)$ completes the proof. □

The next lemma is straightforward by applying Lemma 3.10 and the definition of $f_\alpha(\boldsymbol{A})$, and we thus omit the proof.

Lemma 3.11. *The following are true.*

- *The optimal 1-resolution is unique, $c_1(\boldsymbol{v}) = A_i$ for $\boldsymbol{v} = \mathbb{1}_i(K)$. Moreover $f_1(\boldsymbol{A}) \triangleq A_{sum} = \sum_{k=1}^{K} A_i$.*
- *The optimal K-resolution is unique $c_K(\boldsymbol{v}) = f_K(\boldsymbol{A}) \triangleq A_{\min} = \min_{i\in I_K} A_i$, where $\boldsymbol{v}$ is the all-one vector of length K.*
- *For any α such that $1 \leq \alpha \leq K$, $f_\alpha(\boldsymbol{A}) \leq \frac{A_{sum}}{\alpha}$.*

3.2.3 Enhanced Distortion Vector

Define the following functions

$$\Phi_\alpha(D) = \frac{\alpha D}{1-D}, \quad \alpha = 1, 2, ..., K. \tag{3.8}$$

For a given distortion vector $\boldsymbol{D} = (D_1, D_2, ..., D_K)$, we shall associate it with an *enhanced distortion vector* $\boldsymbol{D^*} = (D_1^*, D_2^*, ..., D_K^*)$ in a recursive manner. For $\alpha = 1$, we set

$$D_1^* = D_1,$$

and define

$$D_\alpha^* = \begin{cases} \frac{(\alpha-1)D_{\alpha-1}^*}{\alpha - D_{\alpha-1}^*} & \text{if } \Phi_\alpha(D_\alpha) > \Phi_{\alpha-1}(D_{\alpha-1}^*) \\ D_\alpha & \text{otherwise} \end{cases} \tag{3.9}$$

for $\alpha = 2, 3, \ldots, K$. The enhanced distortion vector imposes more stringent distortion constraints than the original one, and it is introduced in order to remove certain cases where the given distortion vectors can not be satisfied with equality using the coding schemes we consider. The following lemma shows such property for the enhanced distortion vector. We will present the proof of the lemma in Appendix B.1.

Lemma 3.12. *For any given non-increasing distortion vector $\boldsymbol{D}$, its enhanced distortion vector $\boldsymbol{D^*}$ defined in* (3.9) *satisfies the following properties.*

- *It imposes more stringent distortion constraints,* i.e.*, $D_\alpha^* \le D_\alpha$, for $\alpha = 1, 2, \ldots, K$.*
- *It is a non-increasing vector,* i.e.*, $D_1^* \ge D_2^* \ge \cdots \ge D_K^*$.*
- *It induces a non-increasing Φ sequence,* i.e.*, $\Phi_1(D_1^*) \ge \Phi_2(D_2^*) \ge \cdots \ge \Phi_K(D_K^*)$.*

Moreover, the enhanced distortion vector has the property that it does not significantly effect the lower bound. We shall also assume $D_1 < 1$ for simplicity at this point, but will discuss the cases when $D_1 = 1$ shortly.

3.3 Main Result

We are now ready to present the main theorem of this chapter, which provides an appropriate rate region characterization for the SMD problem. In this section, we present several theorems which summarize the main results for the Gaussian MD problem. The result on approximating the SID − RD function is first given, followed by the rate-distortion region approximation. More details are given in the following sections and the appendixes. Since the results for general sources are notationally more involved, they are thus delayed to Section 3.8 those sections.

3.3.1 Approximating the Symmetric Individual Rate $R_{\mathsf{SMD}}(\boldsymbol{D})$

In this part, we derive an approximate characterization for the symmetric individual description rate-distortion function, $R_{\mathsf{SMD}}(\boldsymbol{D})$. First, we define a set of rate functions which will be used in the theorem.

Definition 3.13. *For a given distortion vector $\boldsymbol{D} = (D_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K)$, define the following functions.*

$$\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}) \triangleq \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} \tag{3.10}$$

$$\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}) \triangleq \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\left[\log\frac{\alpha-D_{\alpha-1}}{\alpha-1}\right] \tag{3.11}$$

$$\underline{R}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d}) \triangleq \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{(1+d_\alpha)(D_\alpha+d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha+d_\alpha)}, \tag{3.12}$$

where $d_1 \geq d_2 \geq \dots \geq d_{K-1} > 0$, $d_0 \triangleq \infty$, $d_K \triangleq 0$ and $D_0 \triangleq 1$. For convenience we define

$$\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) \triangleq \sup_{d_1\geq d_2\geq\dots\geq d_{K-1}>0}\underline{R}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d}). \tag{3.13}$$

Note that $d_1, d_2, \dots, d_K$ are free variables, and therefore $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d})$ is a parametrized function. Later, d_i's will be set to the variances of some auxiliary random variables used throughout the proofs.

Theorem 3.14 (Bounds on the $\mathsf{SID}-\mathsf{RD}$ Function)**.** *Let $\boldsymbol{D}^*$ be the enhanced version of a given distortion vector $\boldsymbol{D}$, then the Gaussian $\mathsf{SID}-\mathsf{RD}$ function under symmetric distortion constraints satisfies*

$$\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \geq \overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \geq R_{\mathsf{SMD}}(\boldsymbol{D}^*) \geq R_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{R}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d}). \tag{3.14}$$

for any $d_1 \geq d_2 \geq \dots \geq d_{K-1} > d_K = 0$ and $d_0 \triangleq \infty$. Moreover,

$$\begin{aligned}\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) &\leq \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log\alpha - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1)\\ &\triangleq L^{\mathsf{SR}}(K) \leq 1.48,\end{aligned} \tag{3.15}$$

$$\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) \leq \frac{1}{2}\sum_{\alpha=2}^{K}\left[\frac{1}{\alpha-1}-\frac{1}{\alpha}\right]\log\alpha \triangleq L^{\mathsf{PPR}}(K) \leq 0.92. \tag{3.16}$$

Remark 3.15. *In Theorem 3.14, we bound the gaps between the upper and lower bounds by universal constants. This is not necessary, and we will show in*

Table 3.1: Values $L^{\mathsf{SR}}(K)$ and $L^{\mathsf{PPR}}(K)$ for $K = 1, 2, 3, ..., 8$.

K	2	3	4	5	6	7	8
$L^{\mathsf{SR}}(K)$	0.5000	0.7296	0.8648	0.9550	1.0200	1.0693	1.1082
$L^{\mathsf{PPR}}(K)$	0.2500	0.3821	0.4654	0.5235	0.5665	0.6000	0.6268

Section 3.4 that the bounds can in fact be distortion dependent. The numerical values are derived using integral approximation for series which does not yield the tightest bounds possible. Table 3.1 includes a few precise values of these bounds.

An important special case is when only the last several levels have distortion constraints, since usually the packet loss probability is not exceedingly high, and for the majority of the time only a small number of packets can be lost. Though the universal bound in Theorem 3.14 also holds for degenerate cases where only certain levels of distortion constraints exist, applying the general bound $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$ can improve the universal constants significantly. In order to do so, $(d_1, d_2, ..., d_{K-1})$ need to be chosen carefully.

Corollary 3.16. *For the Gaussian source, when only distortion constraints for decoders with access to at least k descriptions,* i.e, $D_{K-k+1}, D_{K-k+2}, ..., D_K$ *(or equivalently* $D_1 = D_2 = ... = D_{K-k} = 1$*), we have*

$$\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}(\boldsymbol{D}) \leq \frac{1}{2} \sum_{\alpha=K-k+2}^{K} \frac{1}{\alpha - 1} \log \alpha - \frac{1}{2} \sum_{\alpha=K-k+2}^{K} \frac{1}{\alpha} \log(\alpha - 1),$$

$$\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}(\boldsymbol{D}) \leq \frac{1}{2} \sum_{\alpha=K-k+2}^{K} \left[\frac{1}{\alpha - 1} - \frac{1}{\alpha} \right] \log \alpha.$$

The proof of this corollary can be found in Subsection 3.4.5.

Remark 3.17. *These bounds are usually significantly tighter than the constants given in Theorem 3.14. It is easily seen that when k is kept fixed and* $K \to \infty$*, the gap approaches zero; in fact, in this case even the sum rate bounds become asymptotically tight. Corollary 3.16 implies that when we are guaranteed to receive all but a constant number of descriptions, the* $\mathsf{SR} - \mathsf{ULP}$ *scheme is even more closer to optimum, and the benefit of using more complicated schemes is diminishing as the number of description increases.*

3.3.2 Approximating the Rate-Distortion Region $\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$

We first define two regions, which are in fact two inner bounds to the Gaussian MD rate region.

Definition 3.18. *Let* $\overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$ *be the set of all non-negative rate vectors* $(R_1, R_2, ..., R_K)$*, such that*

$$R_i \geq \sum_{\alpha=1}^{K} r_i^\alpha, \quad 1 \leq i \leq K, \tag{3.17}$$

for some $r_i^\alpha \geq 0$, $1 \leq \alpha \leq K$*, satisfying*

$$\sum_{i:v_i=1} r_i^{|\boldsymbol{v}|} \geq \hat{H}_{|\boldsymbol{v}|}(\boldsymbol{D}), \quad \forall \boldsymbol{v} \in \Omega_K, \tag{3.18}$$

where

$$\hat{H}_\alpha(\boldsymbol{D}) = \frac{1}{2}\log\frac{D_{\alpha-1}}{D_\alpha}, \quad \alpha = 1, 2, \ldots, K, \tag{3.19}$$

and $D_0 \triangleq 1$.

Since the Gaussian source is successively refinable, the right hand side of (3.19) is the rate for each layer in the successive refinement code; (3.17) and (3.18) are the counterpart of (2.3) and (2.4).

The second region is based on a generalization of the PPR multilayer scheme, discussed in details in Section 3.5.

Definition 3.19. *First let* $\boldsymbol{D^*}$ *be the enhanced distortion vector of* $\boldsymbol{D}$ *and define the following quantities*

$$\tilde{H}_1(\boldsymbol{D}^*) = \frac{1}{2}\log\frac{1}{D_1^*},$$
$$\tilde{H}_\alpha(\boldsymbol{D}^*) = \frac{1}{2}\log\frac{(\alpha-1)D_{\alpha-1}^*}{(\alpha - D_{\alpha-1}^*)D_\alpha^*}, \quad \alpha = 2, 3, \ldots, K.$$

Let $\overline{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D})$ *be the set of non-negative rate vectors* $(R_1, R_2, ..., R_K)$*, such that*

$$R_i \geq \sum_{\alpha=1}^{K} r_i^\alpha, \quad 1 \leq i \leq K,$$

for some $r_i^\alpha \geq 0$, $1 \leq \alpha \leq K$*, satisfying*

$$\sum_{i:v_i=1} r_i^{|\boldsymbol{v}|} \geq \tilde{H}_{|\boldsymbol{v}|}(\boldsymbol{D}^*), \quad \boldsymbol{v} \in \Omega_K.$$

The following theorem establishes that both $\overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$ and $\overline{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D})$ are inner bounds to the Gaussian MD rate-distortion region.

Theorem 3.20 (Inner Bounds on $\mathcal{R}_{\mathsf{SMD}}$)**.** *Let* $\boldsymbol{D^*}$ *be the enhanced distortion vector of a given distortion constraint vector* $\boldsymbol{D}$*. Then*

$$\overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D^*}) \subseteq \overline{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D^*}) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D^*}) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}). \tag{3.20}$$

Next, we define a new rate region which together with the following theorem establishes an outer bound for the achievable rate region of the SMD problem.

Definition 3.21. *For a given distortion vector $\boldsymbol{D}$, vector $\boldsymbol{A} \in \mathbb{R}_+^K$ with $\boldsymbol{A} \neq 0$, and a set of arbitrary variables $d_1 \geq d_2 \geq ... \geq d_{K-1} > d_K = 0$, and $d_0 \triangleq \infty$ define*

$$\underline{R}_{\boldsymbol{A}}(\boldsymbol{D}, \boldsymbol{d}) \triangleq \frac{1}{2} \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{(1+d_\alpha)(D_\alpha + d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha + d_\alpha)}, \tag{3.21}$$

where the function $f_\alpha(\boldsymbol{A})$ is defined in (3.5). Define further the following function

$$\underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) \triangleq \sup_{d_1 \geq d_2 \geq ... \geq d_{K-1} > 0} \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}, \boldsymbol{d}). \tag{3.22}$$

Definition 3.22. *For a given distortion vector $\boldsymbol{D}$ and an arbitrary set of variables satisfying $d_1 \geq d_2 \geq ... \geq d_{K-1} > 0$, $d_0 \triangleq \infty$ and $d_K \triangleq 0$, let $\underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$ be the collection of all rate tuples $(R_1, R_2, \ldots, R_K)$ which satisfy*

$$\boldsymbol{A} \cdot \boldsymbol{R} \geq \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}), \tag{3.23}$$

for any $\boldsymbol{A} \in \mathbb{R}_+^K$ and $\boldsymbol{A} \neq 0$, where "$\cdot$" denotes the usual inner product.

The following theorem introduces an outer bound for the SMD achievable rate region. We will present the proof of the theorem in Section 3.6.

Theorem 3.23 (Outer Bound on $\mathcal{R}_{\mathsf{SMD}}$)**.** *For any given distortion vector $\boldsymbol{D}$ and arbitrary variables satisfying $d_1 \geq d_2 \geq ... \geq d_{K-1} > d_K = 0$ and $d_0 = \infty$, we have*

$$\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d}). \tag{3.24}$$

For large number of descriptions, it is difficult to enumerate the faces of the inner and outer bounds of the achievable rate region. So, we seek to approximately characterize the supporting hyper-planes of the rate-distortion region. An alternate way of characterizing the region is to determine its boundaries. In particular, take an arbitrary radius starting from the origin in the non-negative part of the K-dimensional space, and consider a hyper-plane orthogonal to this radius. This hyper-plane starts from the origin and moves towards the infinity along with the radius until it hits the rate region. In the following we derive an approximate characterization for this hitting point. More precisely, for an $\boldsymbol{A} \in \mathbb{R}_+^K$ and $\boldsymbol{A} \neq 0$,

$$R_{\mathsf{SMD},\boldsymbol{A}}(\boldsymbol{D}) \triangleq \min_{\boldsymbol{R} \in \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})} \boldsymbol{A} \cdot \boldsymbol{R}. \tag{3.25}$$

The minimizer rate tuple $\boldsymbol{R}$ is the minimization above is in fact the first point of the region which meets the hyper-plane orthogonal to $\boldsymbol{A}$. In the following we present bound on $R_{\mathsf{SMD},\boldsymbol{A}}(\boldsymbol{D})$.

The next theorem establishes the upper and lower bounds for the supporting hyper-planes of the rate-distortion region. Since the rate-distortion region is convex, if the upper and lower bounds for the supporting hyper-planes coincide, a complete characterization is then available. The upper and lower bounds given in the following theorem do not coincide in general, however the gap between them is bounded, yielding an approximate characterization of the rate region.

Theorem 3.24. *For the Gaussian source and any $\boldsymbol{A} \geq 0$,*

$$\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\hat{H}_\alpha(\boldsymbol{D}^*) \geq \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\tilde{H}_\alpha(\boldsymbol{D}^*) \geq R_{\mathsf{SMD},\boldsymbol{A}}(\boldsymbol{D}) \geq \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) \geq \underline{R}_{\boldsymbol{A}}(\boldsymbol{D},\boldsymbol{d}) \tag{3.26}$$

for any $d_1 \geq d_2 \geq \ldots \geq d_{K-1} > 0$, $d_0 = \infty$ and $d_K = 0$. Moreover, for any $\boldsymbol{A} \in \mathbb{R}_+^K$ and $\boldsymbol{A} \neq 0$, the gap between the lower and upper bounds can be bounded by

$$\begin{aligned}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\hat{H}_\alpha(\boldsymbol{D}^*) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) &\leq \frac{1}{2}\sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A})\log\alpha - \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A})\log(\alpha-1)\\ &\leq \frac{A_{\text{sum}}}{2}\left[\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log\alpha - \sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\log(\alpha-1)\right] - \frac{A_{\min}}{2}\log(K-1),\end{aligned} \tag{3.27}$$

and

$$\begin{aligned}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\tilde{H}_\alpha(\boldsymbol{D}^*) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) &\leq \frac{1}{2}\sum_{\alpha=2}^{K}[f_{\alpha-1}(\boldsymbol{A}) - f_\alpha(\boldsymbol{A})]\log\alpha\\ &\leq \frac{A_{\text{sum}}}{2}\sum_{\alpha=2}^{K-1}[\frac{1}{\alpha-1} - \frac{1}{\alpha}]\log\alpha + \frac{1}{2}(\frac{A_{\text{sum}}}{K-1} - A_{\min})\log(K).\end{aligned} \tag{3.28}$$

This theorem is proved in Section 3.7.

Remark 3.25. *It is not immediately clear that the outer bound, which is specified in terms of an uncountable number of supporting hyper-planes indexed by $\boldsymbol{A}$, is still a polytopes as for the case $K = 3$. Nevertheless it can indeed be shown that when we specialize these bounds for appropriate choice of $\mathbf{d}$, it is an equivalent characterization of a polytopes. Moreover, the bounds given in (3.27) and (3.28) are established using the bound induced by this specific choice of $\mathbf{d}$. We shall return to this point with more details in Section 3.6.*

Remark 3.26. *Theorem 3.24, which provides approximate characterizations of the rate-distortion region, is given in a similar manner as Theorem 3.14, which provides approximation characterizations of the SID-RD function. The second*

Table 3.2: The values of $f_\alpha(\boldsymbol{A})$ and bounds for $K = 3$.

$\boldsymbol{A}$	$(1,0,0)$	$(1,1,0)$	$(2,1,1)$	$(1,1,1)$
$f_1(\boldsymbol{A})$	1	2	4	3
$f_2(\boldsymbol{A})$	0	1	2	1.5
$f_3(\boldsymbol{A})$	0	0	1	1
$\\|\cdot\\|$ by (3.27)	0	$\frac{1}{\sqrt{2}}$	$\frac{3+2\log 3}{2\sqrt{6}}$	$\frac{4+3\log 3}{4\sqrt{3}}$
$\\|\cdot\\|$ by (3.28)	0	$\frac{1}{2\sqrt{2}}$	$\frac{2+\log 3}{2\sqrt{6}}$	$\frac{3+\log 3}{4\sqrt{3}}$

bound in (3.27) and the second bound in (3.28) are more explicit, whereas the first bounds involve the function $f_\alpha(\boldsymbol{A})$ which requires solving an optimization problem. These bounds imply that the gaps between the supporting hyper-planes of inner and outer bounds is upper-bounded by constants independent of the distortion constraints.

Remark 3.27. *Whether the polytopic inner bound is a good approximate characterization of the rate region does not depend on whether the outer bound is a polytopes, but only on how large the gap is between the inner and outer bounds. Though for the Gaussian source, the outer bound can be specialized to be a polytopes, for general sources this specialization may be difficult. Nevertheless, even for general sources, the inner bound, which is an approximate characterization of the rate region, is still a polytopes.*

Example 3.28 ($K = 3$)**.** *Now we apply the result in Theorem 3.24 to the case of $K = 3$. It can be shown that for this case it suffices to consider the choices of vector $\boldsymbol{A}$ in the following set*

$$\{(1,0,0),(1,1,0),(2,1,1),(1,1,1)\}.$$

In Table 3.2, we list the values of $f_\alpha(\boldsymbol{A})$, which can be easily verified since the α-resolution formulation is a linear optimization problem. Using (3.27) and (3.28), it is straightforward to compute the bounds between the inner and upper bounds, as shown in the last two rows of Table 3.2, where the gap is normalized in terms of Euclidean distance.

3.4 Sum Rate and the Symmetric Individual Rate Approximation

In this section, we provide proof details for results on the $\mathsf{SID-RD}$ function, and therefore prove Theorem 3.14. We first introduce more formally two achievable individual description rates, which are given in a general form that can also be applied to other sources, then the derivation of the lower bounds is discussed. With both the upper and lower bounds, we analyze and bound the gap between them. Finally, we extend the results to general sources under the MSE distortion measure.

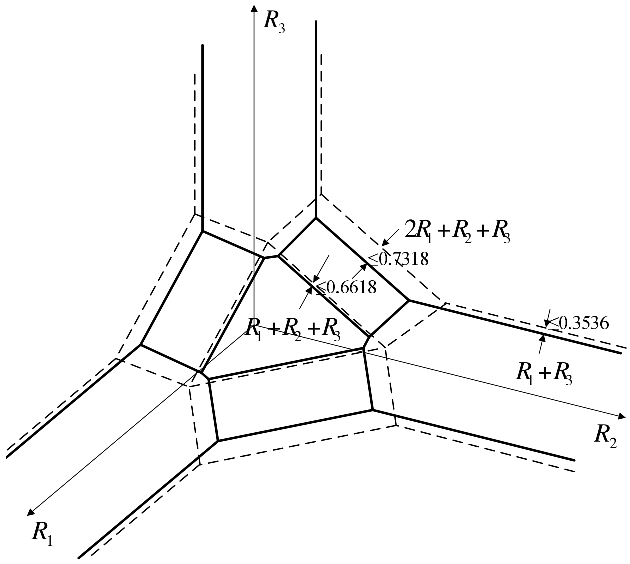


Figure 3.5: Bounding the rate-distortion region for the three-description AMD problem, where the distances between corresponding planes of the inner and outer bounds are measured by Euclidean distance. The inner bound is drawn in with dashed lines, and the outer bound with solid lines.

3.4.1 Achievable Rate Using the SR − MLD Scheme

In this subsection, we derive the first lower bound for $R_{\mathsf{SMD}}(\boldsymbol{D})$ based on the combination of the successive refinement and the multilevel diversity coding. The SR − MLD coding scheme reduces to the SR − ULP scheme when the rate is also symmetric, i.e., $R_1 = R_2 = ... = R_K$. For a general source, we have the following theorem. This theorem is given formally in order to facilitate the analysis for general sources.

Theorem 3.29. *For any given set of random variables $(Y_1, Y_2, ..., Y_K)$ jointly distributed with the source X, such that there exist deterministic functions $G_\alpha : \mathcal{Y}^\alpha \to \mathcal{X}$ to satisfy*

$$\mathbb{E}d(X, G_\alpha(Y_1, Y_2, ..., Y_\alpha)) \leq D_\alpha, \quad \alpha = 1, 2, ..., K, \tag{3.29}$$

we have

$$R_{\mathsf{SMD}}(D_1, D_2, ..., D_K) \leq \sum_{\alpha=1}^{K} \frac{1}{\alpha} I(X; Y_\alpha | Y_1, Y_2, \ldots, Y_{\alpha-1}), \tag{3.30}$$

where $Y_0 \triangleq 0$.

Proof of Theorem 3.29. This theorem is a natural consequence of combining the result on successive refinement [54–57] and the property of the maximum distance separable (MDS) codes.

The random variable Y_α is able to refine the reconstruction based on

$$Y_1, Y_2, \dots, Y_{\alpha-1},$$

such that the expected distortion does not exceed D_α. The excess rate to convey such information required for refinement is $I(X; Y_\alpha | Y_1, Y_2, \dots, Y_{\alpha-1})$. A (K, α) MDS code is then applied to the sequence of Y_α, and encodes it into the K available descriptions, such that any α of them be able to reconstruct Y_α. Therefore each description has to dedicate rate $\frac{1}{\alpha} I(X; Y_\alpha | Y_1, Y_2, \dots, Y_{\alpha-1})$ to Y_α. Repeating this strategy and taking sum over all α's, we get the desired bound. □

In the following, we consider the following natural distribution often seen in the successive refinement problem

$$Y_\alpha = X + \sum_{i=\alpha}^{K} N_i, \qquad \alpha = 1, 2, \dots, K \tag{3.31}$$

where $N_i \sim \mathcal{N}(0, \sigma_i^2)$ are mutually independent and also independent of X. For convenience, we denote $\sum_{i=\alpha}^{K} N_i$ as Z_α. The values of variances σ_i^2, $i = 1, 2, \dots, K$ are chosen such that the distortion constraint at each level is satisfied with equality when the reconstruction is the linear minimum mean squared error estimator (LMMSE), *i.e.*, they are determined by the set of equations

$$D_\alpha = \frac{\sum_{i=\alpha}^{K} \sigma_i^2}{1 + \sum_{i=\alpha}^{K} \sigma_i^2}, \quad \alpha = 1, 2, ..., K. \tag{3.32}$$

It is clear that there always exists a unique and valid solution for these variances when the distortion constraints are given in the natural monotonic order, $D_1 \geq D_2 \geq \cdots \geq D_K$. It is easy to show that this choice of Y_α's gives us

$$I(X; Y_\alpha | Y_1, Y_2, \dots, Y_{\alpha-1}) = \frac{1}{2} \frac{D_{\alpha-1}}{D_\alpha}.$$

Summing up the terms for $\alpha = 1, \dots, K$, we arrive at the function $\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$ defined in (3.10).

3.4.2 Achievable Rate Using the PPR Multilayer Scheme

From Theorem 3.4, an achievable individual description rate can be derived by choosing a specific set of encoding auxiliary random variables, and more specifically, we shall choose the following set of random variables. Let

$$Y_{\alpha,k} = X + \sum_{i=\alpha}^{K-1} N_{i,k}, \quad \alpha = 1, 2, ..., K-1, \quad k = 1, 2, \dots, K \tag{3.33}$$

where $N_{i,k}$ are mutually independent zero-mean Gaussian random variables, which are also independent of X. Their variances are denoted as $\sigma^2_{i,k}$, and they satisfy $\sigma^2_{i,k} = \sigma^2_{i,k'}$ for any $k, k' \in \mathcal{I}_K$; we thus denote $\sigma^2_{i,k}$ as $\tilde{\sigma}^2_i$. The construction of these auxiliary variables is illustrated in Figure 3.6. For convenience, we shall denote $\sum_{i=\alpha}^{K-1} N_{i,k}$ as $Z_{\alpha,k}$. For the last layer, i.e., $\alpha = K$, we use

$$Y_K = X - \mathbb{E}(X|Y_{\alpha,k}, \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K) + N_K, \tag{3.34}$$

where N_K is a zero-mean Gaussian random variable independent of everything else, with variance $\tilde{\sigma}^2_K$. For all the layers except that K-th layer, the lower layers are useless when higher layer random variables are available, thus using only $\{Y_{\alpha-1,i}; i : v_i = 1\}$ in the function $G_{\boldsymbol{v}}$ does not lose optimality.

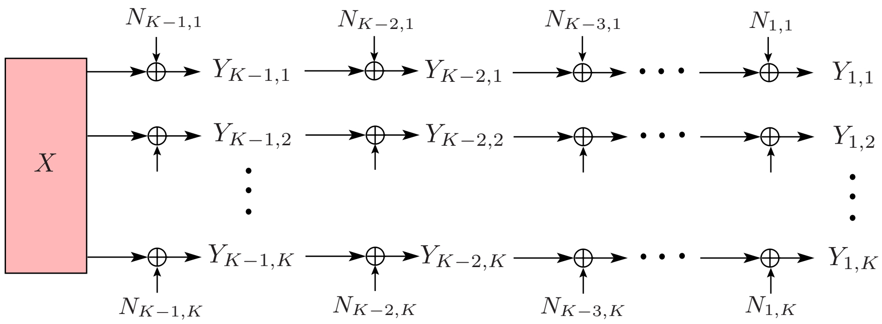


Figure 3.6: structure of the auxiliary random variables used in the PPR multilayer scheme.

It remains to specify the variances of $\{\{N_{\alpha,k}, \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_\alpha\}, N_K\}$, for which the following observation is important. Consider the decoding function $G_{\boldsymbol{v}}$ and the decoding $G_{\boldsymbol{u}}$, where $\boldsymbol{u} = \boldsymbol{v} + \mathbb{1}_j(K)$, *i.e.*, the set of available descriptions at decoder $\boldsymbol{u}$ strictly includes that of decoder $\boldsymbol{v}$, and moreover, it has access to the j-th description. With the auxiliary random variables defined in (3.33) and (3.34), it is clear that we should choose them to have the following forms

$$G_{\boldsymbol{v}}(Y_{|\boldsymbol{v}|,j}; j : v_j = 1) = \mathbb{E}(X|Y_{|\boldsymbol{v}|,j}; j : v_j = 1),$$
$$G_{\boldsymbol{u}}(Y_{|\boldsymbol{u}|,j}; j : u_j = 1) = \mathbb{E}(X|Y_{|\boldsymbol{u}|,j}; j : u_j = 1).$$

However, even when $Y_{|\boldsymbol{v}|,j} = Y_{|\boldsymbol{u}|,j}$, *i.e.*, there is no improvement from the $|\boldsymbol{v}|$-th layer information to the $|\boldsymbol{u}|$-th layer information, the function $G_{\boldsymbol{u}}$ still induces less distortion because it can use $\{Y_{|\boldsymbol{v}|,j} : j : u_j = 1\}$, which strictly includes $\{Y_{|\boldsymbol{v}|,j}; j : v_j = 1\}$. This implies that for certain distortion vector $\boldsymbol{D}$, it is not possible to satisfy all its components with equalities in this scheme because some of them are too loose. The enhanced distortion vector defined given in Subsection 3.2.3 is thus introduced to eliminate this effect.

The enhancement technique has been previously used in [60] and [28] to derive tight outer bound for the MIMO Gaussian broadcast channel capacity and two-level vector Gaussian sum rate, respectively. The enhancement we use here is simpler than theirs without the Lagrangian multiplier, but the essence is the same: the channel or the distortion constraint is enhanced such that the coding problem is less or more stringent, and at the same time the chosen enhancement not only does not change the outer bound (or one specific choice within a set of outer bounds), but also leads to a setting where the coding scheme can satisfy the constraints with equality.

We shall now choose the variances of $\{\{N_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_\alpha\}, N_K\}$ by the following equations

$$\sum_{i=\alpha}^{K-1} \tilde{\sigma}_i^2 = \Phi_\alpha(D_\alpha^*), \quad \alpha = 1, 2, ..., K-1, \tag{3.35}$$

$$\tilde{\sigma}_K^2 = \frac{D_K^* \Phi_{K-1}(D_{K-1}^*)}{(1 - D_K^*)[\Phi_{K-1}(D_{K-1}^*) - \Phi_K(D_K^*)]}. \tag{3.36}$$

The following lemma follows.

Lemma 3.30. *Let $\boldsymbol{D}^*$ be the enhanced distortion vector associated with any valid distortion vector $\boldsymbol{D}$, then the random variables defined by (3.33), (3.34), (3.35) and (3.36) exist, and they satisfy*

$$\mathbb{E}\left[X - \mathbb{E}(X|Y_{|\boldsymbol{v}|,j}; j : v_j = 1)\right]^2 = D_{|\boldsymbol{v}|}^*, \qquad |\boldsymbol{v}| = 1, 2, \ldots, K-1, \tag{3.37}$$

$$\mathbb{E}\left[X - \mathbb{E}(X|\{Y_{K-1,j}; j \in \mathcal{I}_K\}, Y_K)\right]^2 = D_K^*. \tag{3.38}$$

The proof of this lemma can be found in Appendix B.2. Having this theorem, we are now ready to prove the second inner bound.

Proof of $R_{\mathsf{SMD}}(\boldsymbol{D}^*) \leq \overline{R}_{\mathsf{SMD}}^{\mathsf{PPR}}(\boldsymbol{D}^*)$. We first rewrite the rate expressions given in Theorem 3.2. For a fixed set of (generalized symmetric) auxiliary random variables $\boldsymbol{Y} \triangleq \{\{Y_{\alpha,k}; \alpha \in I_{K-1}, k \in I_K\}, Y_K\}$, define

$$\begin{aligned}\check{H}_\alpha(\boldsymbol{Y}) =& h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\ &- \frac{\alpha}{K} h(\{Y_{\alpha,i}; i \in \mathcal{I}_K\}|X, \{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_K\}),\end{aligned}$$

and

$$\check{H}_K(\boldsymbol{Y}) = I(X; Y_K|\{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}).$$

It follows that

$$\begin{aligned}\sum_{\alpha=1}^{K} \frac{1}{\alpha} \tilde{H}_\alpha(\boldsymbol{Y}) = & \sum_{\alpha=1}^{K-1} \frac{1}{\alpha} h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\ &+ \frac{1}{K} h(Y_K|\{Y_{j,k}; j \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}) \\ &- \frac{1}{K} h(\{Y_{j,k}; j \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K|X),\end{aligned}$$

where the right hand side is the rate expression given in Theorem 3.2.

Now for the specific set of random variables defined by (3.33) and (3.35), we have for $\alpha = 2, 3, ..., K-1$

$$\begin{aligned}
\check{H}_\alpha(\boldsymbol{Y}) =& h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
& - \frac{\alpha}{K} h(\{X + Z_{\alpha,i}; i \in \mathcal{I}_K\}|X, \{X + Z_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_K\}) \\
\stackrel{(a)}{=}& h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
& - \frac{\alpha}{K} h(\{Z_{\alpha,i}; i \in \mathcal{I}_K\}|\{Z_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_K\}) \\
\stackrel{(b)}{=}& h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
& - h(\{Z_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Z_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
\stackrel{(c)}{=}& h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
& - h(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}|X, \{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
=& I(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\}; X|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}),
\end{aligned}$$

where (a) and (c) are because X is independent of $Z_{\alpha,i}$; (b) is because of the chain rule and the fact that $Z_{\alpha,i}$ is independent of $\{Z_{\alpha,k}; \alpha \in \mathcal{I}_K, k \neq i\}$. The Markov chain $\{Y_{1,k}; k \in \mathcal{I}_K\} \leftrightarrow \{Y_{2,k}; k \in \mathcal{I}_K\} \leftrightarrow ... \leftrightarrow \{Y_{K-1,k}; k \in \mathcal{I}_K\} \leftrightarrow X$, implies that

$$\begin{aligned}
I(\{Y_{\alpha,i}; i \in \mathcal{I}_\alpha\};& X|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) \\
&= h(X|\{Y_{j,k}; j \in \mathcal{I}_{\alpha-1}, k \in \mathcal{I}_\alpha\}) - h(X|\{Y_{j,k}; j \in \mathcal{I}_\alpha, k \in \mathcal{I}_\alpha\}) \\
&= h(X|\{Y_{\alpha-1,k}; k \in \mathcal{I}_\alpha\}) - h(X|\{Y_{\alpha,k}; k \in \mathcal{I}_\alpha\}) \\
&\stackrel{(d)}{=} \frac{1}{2} \log \frac{(\alpha-1)D^*_{\alpha-1}}{(\alpha - D^*_{\alpha-1})} - \frac{1}{2} \log D^*_\alpha \\
&= \frac{1}{2} \log \frac{(\alpha-1)D^*_{\alpha-1}}{(\alpha - D^*_{\alpha-1})D^*_\alpha},
\end{aligned} \tag{3.39}$$

where (d) follows from the choices of the variances of the Gaussian random variables $N_{\alpha,k}$. For $\alpha = 1$ and $\alpha = K$, it is straightforward to verify that

$$\begin{aligned}
\check{H}_1(\boldsymbol{Y}) &= \frac{1}{2} \log \frac{1}{D^*_1}, \\
\check{H}_K(\boldsymbol{Y}) &= \frac{1}{2} \log \frac{(K-1)D^*_{K-1}}{(K - D^*_{K-1})D^*_K},
\end{aligned} \tag{3.40}$$

which means $\check{H}_\alpha(\boldsymbol{Y}) = \tilde{H}_\alpha(D^*_\alpha)$ for $\alpha = 1, 2, \ldots, K$. Combining (3.39)-(3.40) we have,

$$\sum_{\alpha=1}^{K} \frac{1}{\alpha} \tilde{H}_\alpha(\boldsymbol{D}^*) = \frac{1}{2} \log \frac{1}{D_1^*} + \frac{1}{2} \sum_{\alpha=2}^{K} \frac{1}{\alpha} \log \frac{(\alpha-1)D_{\alpha-1}^*}{(\alpha - D_{\alpha-1}^*)D_\alpha^*}$$
$$= \frac{1}{2} \sum_{\alpha=1}^{K} \log \frac{D_{\alpha-1}^*}{D_\alpha^*} - \frac{1}{2} \sum_{\alpha=2}^{K} \frac{1}{\alpha} \log \frac{\alpha - D_{\alpha-1}^*}{\alpha - 1},$$

where we define $D_0^* \triangleq 1$. Note that the RHS of the above expression is the same as $\overline{R}_{\mathsf{SMD}}^{\mathsf{PPR}}$ defined in (3.11), and it is clearly achievable due to Theorem 3.2. □

3.4.3 Lower Bounding the Sum Rate and $R_{\mathsf{SMD}}(\boldsymbol{D})$ Function

In this subsection, lower bounds toward those used in proving Theorem 3.14 are derived. The following theorem is the key tool to derive these lower bounds.

Theorem 3.31. *For the Gaussian source, the sum-rate under the K-description symmetric distortion satisfies*

$$\sum_{i=1}^{K} R_i \geq \frac{K}{2} \sum_{\alpha=1}^{K} \frac{1}{\alpha} \log \frac{(1+d_\alpha)(D_\alpha + d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha + d_\alpha)}, \tag{3.41}$$

where $d_1 \geq d_2 \geq ... \geq d_{K-1} > d_K = 0$ are arbitrary non-negative values and $d_0 \triangleq \infty$.

Before proving this theorem, we state the following lemma, which plays an important role in the proof. The proof of this lemma can be found in Appendix B.3.

Lemma 3.32. *Let $\{\Gamma_i; i \in \mathcal{I}_K\}$ be a set of descriptions such that there exist decoding functions to satisfy the distortion constraints $\boldsymbol{D} = (D_1, D_2, ..., D_K)$. Let $Y_b = X + N_b$ and $Y_a = X + N_a + N_b$, where N_a and N_b are mutually independent Gaussian random variables, independent of the Gaussian source X, with variance σ_a^2 and σ_b^2, respectively. Then by defining $\sigma_b^2 = d_b$ and $\sigma_a^2 + \sigma_b^2 = d_a$ and $\Gamma_{\boldsymbol{v}} = \{\Gamma_i; i : v_i = 1\}$, we have*

1. ***Mutual information bound between encoding functions and a noisy source***

$$I(\Gamma_{\boldsymbol{v}}; Y_a^n) \geq \frac{n}{2} \log \frac{1 + d_a}{D_{|\boldsymbol{v}|} + d_a}, \tag{3.42}$$

2. ***Bound on mutual information difference between encoding functions and different noisy sources***

$$I(\Gamma_{\boldsymbol{v}}; Y_b^n) - I(\Gamma_{\boldsymbol{v}}; Y_a^n) \geq \frac{n}{2} \log \frac{(1+d_b)(D_{|\boldsymbol{v}|} + d_a)}{(1+d_a)(D_{|\boldsymbol{v}|} + d_b)}. \tag{3.43}$$

Now, we are ready to present the proof of Theorem 3.31.

Proof of Theorem 3.31. The bounding technique extends the method used in [22,28,29], with the additional ingredient that we expand the probability space with more than one additional random variable, and then utilize the special structure in the expanded probability space to bound the sum rate. Consider the variables $\{Y_\alpha; \alpha \in \mathcal{I}_K\}$ defined in (3.31). We have the following chain of inequalities

$$
\begin{aligned}
n\sum_{i=1}^{K}(R_i+\epsilon) &\geq \sum_{i=1}^{K} H(\Gamma_i) - H(\Gamma_i; i \in \mathcal{I}_K | X^n) \\
&\overset{(a)}{=} \sum_{\alpha=1}^{K-1}\left[\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} H(\Gamma_{\boldsymbol{v}}) - \frac{K}{(\alpha+1)\binom{K}{\alpha+1}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} H(\Gamma_{\boldsymbol{v}})\right] \\
&\quad + H(\{\Gamma_i; i \in \mathcal{I}_K\}) - H(\{\Gamma_i; i \in \mathcal{I}_K\}|X^n) \\
&\overset{(b)}{\geq} \sum_{\alpha=1}^{K-1}\left[\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} H(\Gamma_{\boldsymbol{v}}) - \frac{K}{(\alpha+1)\binom{K}{\alpha+1}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} H(\Gamma_{\boldsymbol{v}})\right] \\
&\quad - \sum_{\alpha=1}^{K-1}\left[\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} H(\Gamma_{\boldsymbol{v}}|Y_\alpha^n) - \frac{K}{(\alpha+1)\binom{K}{\alpha+1}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} H(\Gamma_{\boldsymbol{v}}|Y_\alpha^n)\right] \\
&\quad + I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) \\
&= \sum_{\alpha=1}^{K-1}\left[\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - \frac{K}{(\alpha+1)\binom{K}{\alpha+1}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n)\right] \\
&\quad + I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) \\
&= \sum_{i=1}^{K} I(\Gamma_i; Y_1^n) + \sum_{\alpha=2}^{K-1}\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha}\Big[I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - I(\Gamma_{\boldsymbol{v}}; Y_{\alpha-1}^n)\Big] \\
&\quad + \Big[I(\{\Gamma_i, i \in \mathcal{I}_K\}; X^n) - I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n)\Big]
\end{aligned} \tag{3.44}
$$

where (a) is by adding and subtracting the same terms where the positive term in the bracket chases the negative one; (b) is true because the subtracted bracket is non-negative due to conditional version of Han's inequality [3] (p. 668).

For convenience we denote the variance of additive noise in $Y_\alpha = X + \sum_{j=\alpha}^{K} N_j$ by $d_\alpha = \sum_{j=\alpha}^{K} \mathbb{E}[N_j^2] = \sum_{j=\alpha}^{K}\sigma_j^2$. Now we can apply Lemma 3.32 on

(3.44) to get

$$\begin{aligned}\sum_{i=1}^{K}(R_i+\epsilon) \geq \sum_{i=1}^{K}\frac{1}{2}\log\frac{1+d_1}{D_1+d_1} + \sum_{\alpha=2}^{K-1}\frac{K}{\alpha\binom{K}{\alpha}}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha}\frac{1}{2}\log\frac{(1+d_\alpha)(D_\alpha+d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha+d_\alpha)} \\ +\frac{1}{2}\log\frac{(1+d_K)(D_K+d_{K-1})}{(1+d_{K-1})(D_K+d_K)} \\ = \frac{K}{2}\log\frac{1+d_1}{D_1+d_1} + \frac{K}{2}\sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\frac{K}{2}\frac{1}{K}\log\frac{(1+d_\alpha)(D_\alpha+d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha+d_\alpha)},\end{aligned} \tag{3.45}$$

which together with

$$\log\frac{D_1+d_0}{1+d_0} = \log\frac{D_1+\infty}{1+\infty} = 0.$$

give us the desired result. □

The lower bounds in Theorem 3.14 are direct consequences of the above theorem. First by dividing both sides of (3.41) by K, we obtain

$$R_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{R}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d})$$

for any valid choice of $\boldsymbol{d}$. It is clear that by optimizing over $\boldsymbol{d}$ we get tightest bound among this class, that is $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D})$.

Note that the lower bound in Theorem 3.31 is in fact a set of lower bounds, parametrized by $d_1 \geq d_2 \geq \ldots \geq d_{K-1} > 0$. We may optimize it to find the tightest lower bound, however, an explicit optimization is not only difficult, but also fails to offer much insight due to the lack of matching achievability result. Therefore, we do not evaluate $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D})$. Instead we choose a specific set of values to get a (sub-optimal) bound, resulting in the following corollary. We will present the proof of this corollary in Appendix B.4.

Corollary 3.33. *For a given distortion constraint vector $\boldsymbol{D}$ and its enhanced version $\boldsymbol{D}^*$, define*

$$\begin{aligned}\underline{\check{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*) = &\frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D^*_{\alpha-1}}{D^*_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log(\alpha-D^*_{\alpha-1}) \\ &+\frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1).\end{aligned} \tag{3.46}$$

Then, for the Gaussian source, $\underline{\check{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*)$ is a lower bound for $R_{\mathsf{SMD}}(\boldsymbol{D})$ from the class of bounds of the form $\underline{\check{R}}_{\mathsf{SMD}}(\boldsymbol{D},\boldsymbol{d})$ for a particular choice of $\boldsymbol{d}$, i.e.,

$$R_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{\check{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*). \tag{3.47}$$

Remark 3.34. *It is worth noting that the right hand side of (3.46) is only related to the enhanced distortion vector, while it is bounding the* SID − RD, *a function of distortion vector* $\boldsymbol{D}$. *Indeed the enhanced distortion vector is given in such a way that it does not change the lower bound under the chosen value of* $(d_1, d_2, ..., d_{K-1})$. *It is discussed in more detail in the proof of this corollary.*

3.4.4 Bounding the Gap Between Lower and Upper Bounds

In this part we prove the last of Theorem 3.14, which is bounding the gap between the lower and upper bounds on SID − RD function.

For bounding $\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}_{\mathsf{SMD}}(\boldsymbol{D})$, we can simply write

$$\begin{aligned}\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) &\leq \overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \check{\underline{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*)\\ &= \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log(\alpha - D^*_{\alpha-1}) - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1)\\ &\leq \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log\alpha - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1).\end{aligned}$$

Similarly in order to bound the gap between $\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*)$ and $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D})$, we have

$$\begin{aligned}\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) &\leq \overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) - \check{\underline{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*)\\ &= \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log(\alpha - D^*_{\alpha-1}) - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1)\\ &\quad - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\left[\log\frac{\alpha - D^*_{\alpha-1}}{\alpha-1}\right]\\ &= \frac{1}{2}\sum_{\alpha=2}^{K}\left[\frac{1}{\alpha-1} - \frac{1}{\alpha}\right]\log(\alpha - D^*_{\alpha-1})\\ &\leq \frac{1}{2}\sum_{\alpha=2}^{K}\left[\frac{1}{\alpha-1} - \frac{1}{\alpha}\right]\log\alpha\end{aligned}$$

This completes the proof of Theorem 3.14.

Theorem 3.14 provides one possible approximation for the SID − RD function with universal constants. Various improvements can be made, for example, with better choice of $(d_1, d_2, ..., d_{K-1})$, particularly for low-resolution regimes. For the case with only two level distortion constraints, the outer bound in Theorem 3.31 reduces correctly to the one given in [28] and [29]. It was shown in [22], [28] and [29] that for certain cases this bound is indeed tight, which however requires optimization to find the optimal lower bound. Better choice of random variables in the PPR multilayer coding scheme can also lead to improved bounds. The current upper bound is loose even in the high-resolution regime, and through more sophisticated coding schemes, it may be possible to find such asymptotically tight inner bounds. We will not pursue such refinements here, but it can be an interesting question to be addressed in future.

3.4.5 Bounding the Gap when $D_1 = D_2 = \cdots = D_{K-k} = 1$

In order to prove Corollary 3.16, notice that this case implies we can set $\check{d}_1 = \check{d}_2 = ... = \check{d}_{K-k} = \infty$, which makes the term in the summation in (3.12) zero for $\alpha = 1, 2, \ldots, K-1$. Thus the lower bound $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \check{\boldsymbol{d}})$ implies that

$$\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \check{\boldsymbol{d}}) = \frac{1}{2} \sum_{\alpha=K-k+1}^{K} \frac{1}{\alpha} \log \frac{(1+d_\alpha)(D_\alpha + d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha + d_\alpha)}.$$

Next, apply the procedure of computing the enhanced distortion vector on

$$(D_{K-k+1}, D_{K-k+2}, .., D_K)$$

only, and denote the output as $(D^*_{K-k+1}, D^*_{K-k+2}, ..., D^*_K)$. The rest of the proof is exactly the same as the proof of Corollary 3.33, and we skip it.

3.5 Rate-Distortion Region: Achievable Schemes

In this section, we derive two inner bounds for the SMD rate-distortion region. This essentially proves Theorem 3.20. The development is largely parallel to those in the previous section, with the new ingredient of α-resolution to help establish an achievable region by "inflating" the PPR multilayer rate.

3.5.1 An Achievable Region by the SR − MLD Scheme

We first need to define a new rate region as follows.

Definition 3.35. *Let $\hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ be the set of all non-negative rate vectors $(R_1, R_2, ..., R_K)$, such that*

$$R_i \geq \sum_{\alpha=1}^{K} r_i^\alpha, \qquad 1 \leq i \leq K,$$

for some $r_i^\alpha \geq 0$, $1 \leq \alpha \leq K$, satisfying

$$\sum_{i:v_i=1} r_i^{|\boldsymbol{v}|} \geq I(X; Y_{|\boldsymbol{v}|}|Y_1, Y_2, \ldots, Y_{|\boldsymbol{v}|-1}), \qquad \forall \boldsymbol{v} \in \Omega_K.$$

The following theorem is rather immediate by the result of SMLD problem in Section 2.1 and and Theorem 3.29.

Theorem 3.36. *We have*

$$\mathbf{conv}\left(\bigcup \hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})\right) \subseteq \mathcal{R}(\boldsymbol{D}),$$

where $\mathbf{conv}(\cdot)$ *is the convex hull operator, and the union is taken over the set of auxiliary random variables* $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_K)$ *in some alphabets* $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_K$*, which are jointly distributed with* X*, such that there exist deterministic functions* $G_\alpha : \mathcal{Y}_1 \times \mathcal{Y}_2 \times \ldots \times \mathcal{Y}_\alpha \to \mathcal{X}$ *to satisfy*

$$\mathbb{E}d(X, G_\alpha(Y_1, Y_2, \ldots, Y_\alpha)) \leq D_\alpha, \quad \alpha = 1, 2, \ldots, K.$$

The following corollary is a direct consequence of the above theorem.

Corollary 3.37.

$$\overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}).$$

Proof of Corollary 3.37. One particular choice of $\boldsymbol{Y}$ is the one specified by (3.31) and (3.32), for which we have $\hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y}) = \overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$. Therefore, it is clear that $\underline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$ is a (proper) subset of $\mathbf{conv}(\bigcup \hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})$, and thus an achievable region. $\square$

Remark 3.38. *Note that the region* $\mathbf{conv}(\bigcup \hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ *may be a general convex region with curvy boundary, thus not a polytopes. However* $\underline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$ *is a subset of this set by specializing it to a particular distribution, resulting in a polytopes, since it is the projection of the polytopes* $(R_1, R_2, \ldots, R_K; \{r_i^\alpha; \alpha \in \mathcal{I}_K, i \in \mathcal{I}_K\})$ *on the first* K *components.*

Remark 3.39. *It is clear that Corollary 3.37 in particular implies* $\overline{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}^*)$*, by replacing* $\boldsymbol{D}$ *by* $\boldsymbol{D}^*$*, which is claimed in Theorem 3.20.*

3.5.2 An Achievable Region by the PPR Multilayer Scheme

Though the original PPR multilayer scheme is designed for a single symmetric rate point, it can indeed be used to generate an achievable region. However, this directly generalized region is not suitable for the purpose of approximate characterization, and thus we do not discuss it in details. Instead, we shall consider the following region.

Definition 3.40. *Let* $\hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ *be the set of all non-negative rate vectors* $(R_1, R_2, \ldots, R_K)$*, such that*

$$R_i \geq \sum_{\alpha=1}^{K} r_i^\alpha, \qquad 1 \leq i \leq K,$$

for some $r_i^\alpha \geq 0$, $1 \leq \alpha \leq K$*, satisfying*

$$\sum_{i: v_i = 1} r_i^{|\boldsymbol{v}|} \geq \tilde{H}_{|\boldsymbol{v}|}(\boldsymbol{Y}), \qquad \boldsymbol{v} \in \Omega_K,$$

where $\tilde{H}_{|\boldsymbol{v}|}(\boldsymbol{Y})$ *is defined in (3.3) and (3.4).*

The following theorem is the counterpart of Theorem 3.36. However, its proof is more involved.

Theorem 3.41. *We have*

$$\mathbf{conv}\left(\bigcup \hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})\right) \subseteq \mathcal{R}(\boldsymbol{D}),$$

where the union is taken over the set of generalized symmetric auxiliary random variables $\{\{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K\}$ *in the alphabets* $\mathcal{Y}_1^K \times \mathcal{Y}_2^K \times \cdots \times \mathcal{Y}_{K-1}^K \times \mathcal{Y}_K$*, which are jointly distributed with* X*, such that there exist deterministic functions* $G_\alpha : \mathcal{Y}_1^\alpha \times \mathcal{Y}_2^\alpha \times \cdots \times \mathcal{Y}_\alpha^\alpha \to \mathcal{X}$*,* $\alpha \in \mathcal{I}_{K-1}$ *such that*

$$\mathbb{E}\left[d(X, G_\alpha(\{Y_{i,k}; i \in \mathcal{I}_\alpha, k \in \mathcal{I}_\alpha\}))\right] \le D_\alpha, \qquad \alpha = 1, 2, \ldots, K-1,$$

and $G_K : \mathcal{Y}_1^K \times \mathcal{Y}_2^K \times \cdots \times \mathcal{Y}_{K-1}^K \times \mathcal{Y}_K \to \mathcal{X}$*, such that*

$$\mathbb{E}\left[d(X, G_K(\{Y_{i,k}; i \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K))\right] \le D_K.$$

The proof of Theorem 3.41 relies on a result in [30]. We quote this result and the proof of the theorem in Appendix B.5. The following corollary shows the second part of Theorem 3.20.

Corollary 3.42. *Let* $\boldsymbol{D}^*$ *be the enhanced distortion vector of* $\boldsymbol{D}$*, then for the Gaussian source*

$$\overline{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}^*).$$

Proof of Corollary 3.42. Notice that the region $\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D})$ is just $\hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ with the choice of $\{\{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K\}$ defined by (3.33) and (3.34), the variances of which are given by (3.35) and (3.36). Therefore, we have

$$\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \mathbf{conv}\left(\bigcup \hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})\right) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}),$$

and similarly, $\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}^*)$. □

3.6 Rate-Distortion Region: Outer Bound

In this section, we provide a lower bound to the supporting hyper-plane of the Gaussian MD rate-distortion region, which yields in establishing the outer bound in Theorem 3.23.

Proof of Theorem 3.23. Consider a arbitrary vector $\boldsymbol{A} \ge 0$ and let $c_1, c_2, ..., c_K$ be a set of α-resolution for $\boldsymbol{A}$ satisfying

$$c_1 \succ c_2 \succ \ldots \succ c_K.$$

as stated in Theorem 3.9. Then we can start with

$$
\begin{aligned}
n\sum_{i=1}^{K} A_i(R_i+\epsilon) &\geq \sum_{i=1}^{K} A_i H(\Gamma_i)\\
&\overset{(a)}{=} \sum_{\boldsymbol{v}:|\boldsymbol{v}|=1} c_1(\boldsymbol{v}) H(\Gamma_{\boldsymbol{v}})\\
&\overset{(b)}{=} \sum_{\alpha=1}^{K-1}\left[\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}}) - \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}})\right]\\
&\quad + A_{\min}H(\Gamma_i; i\in\mathcal{I}_K) - A_{\min}H(\Gamma_i; i\in\mathcal{I}_K|X^n)\\
&\overset{(c)}{\geq} \sum_{\alpha=1}^{K-1}\left[\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}}) - \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}})\right]\\
&\quad - \sum_{\alpha=1}^{K-1}\left[\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}}|Y_\alpha^n) - \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})H(\Gamma_{\boldsymbol{v}}|Y_\alpha^n)\right]\\
&\quad + A_{\min}I(\{\Gamma_i; i\in\mathcal{I}_K\}; X^n)\\
&= \sum_{\alpha=1}^{K-1}\left[\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha+1} c_{\alpha+1}(\boldsymbol{v})I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n)\right]\\
&\quad + A_{\min}I(\{\Gamma_i; i\in\mathcal{I}_K\}; X^n)\\
&\overset{(d)}{=} \sum_{i=1}^{K} A_i I(\Gamma_i; Y_1^n) + \sum_{\alpha=2}^{K-1}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})\Big[I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - I(\Gamma_{\boldsymbol{v}}; Y_{\alpha-1}^n)\Big]\\
&\quad + A_{\min}\Big[I(\{\Gamma_i; i\in\mathcal{I}_K\}; X^n) - I(\{\Gamma_i; i\in\mathcal{I}_K\}; Y_{K-1}^n)\Big],
\end{aligned}
\tag{3.48}
$$

where in (a) and (d) we used the result of Lemma 3.11 that $c_1(\boldsymbol{v}) = A_i$ for $\boldsymbol{v} = \mathbb{1}_i(K)$; (b) is by adding and subtracting the same terms, and due to the fact that $\{\Gamma_i; i\in\mathcal{I}_K\}$ are deterministic functions of X^n; finally, (c) is by a conditional version of the covering property of the given sequence of the optimal α-resolutions as defined in (3.6). At this point, we can apply Lemma 3.32 to bound each term in (3.48). This results in

$$
\begin{aligned}
\sum_{i=1}^{K} A_i(R_i+\epsilon) \geq &\frac{1}{2}\sum_{i=1}^{K} A_i \log\frac{1+d_1}{D_1+d_1}\\
&+ \sum_{\alpha=2}^{K-1}\sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} c_\alpha(\boldsymbol{v})\left[\frac{1}{2}\log\frac{(1+d_\alpha)(D_\alpha+d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha+d_\alpha)}\right]\\
&+ A_{\min}\left[\frac{1}{2}\log\frac{(1+d_\alpha)(D_\alpha+d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha+d_\alpha)}\right]
\end{aligned}
$$

$$\stackrel{(e)}{=} \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{(1+d_\alpha)(D_\alpha + d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha + d_\alpha)} \tag{3.49}$$

where in (e) we used the definition of $f_\alpha(\boldsymbol{A})$ in (3.5) as well as the fact that

$$\log \frac{D_1 + d_0}{1 + d_0} = \log \frac{D_1 + \infty}{1 + \infty} = 0.$$

This completes the proof of the theorem. □

As mentioned before, one can optimize $\underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$ for $\boldsymbol{d}$ and obtain the tightest bound in this class, $\underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D})$, which is a difficult task to do. Instead, we choose a set of values for $d_1, d_2, \ldots, d_{K-1}$ and derive a non-parametric bound as in the following corollary. This would be useful later for bounding the gap between the inner and outer bounds. We first define the following rate region.

Definition 3.43. *For a given distortion constraint vector $\boldsymbol{D}$ and its enhanced version $\boldsymbol{D^*}$, let $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ be the set all rate tuples $(R_1, R_2, \ldots, R_K)$ that for any $\boldsymbol{A} \in \mathbb{R}_+^K$ and $\boldsymbol{A} \neq 0$*

$$\begin{aligned}\boldsymbol{A}\cdot\boldsymbol{R} \geq & \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D^*_{\alpha-1}) \\ & + \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - 1).\end{aligned} \tag{3.50}$$

Corollary 3.44. *For the Gaussian source, $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ establishes an outer bound for the* SMD *rate-distortion region,* i.e.,

$$\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*}). \tag{3.51}$$

The proof of this corollary is along a similar line as the proof of Corollary 3.33, with the additional application of Lemma 3.10 in one step. We present this proof in Appendix B.6 for completeness.

Next we proceed to establish that the outer bound $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ is indeed a polytopes.

Corollary 3.45. *The rate region $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ is a polytope.*

The proof of this corollary is given in Appendix B.7.

3.7 Rate-Distortion Region: Gap Analysis

We derived inner and outer bounds for the achievable rate region of the SMD problem in Section 3.5 and Section 3.6, respectively. These bounds are defined based on characterization of their supporting hyper-planes. Our goal in this section is to show that these bounds together, provide an approximate characterization for the SMD rate region. In order to do this, we show that the gap between each two parallel hyper-planes is bounded by a constant, independent of the distortion constraints. Therefore, the boundary of the actual rate region would be sandwiched between two boundaries with a gap bounded above by a constant. This is exactly the result of Theorem 3.24, which will prove here.

First of all, note that the first and second inequalities in (3.26) are direct consequences of Theorem 3.36 and Theorem 3.41. Also, Theorem 3.23 together with Definition 3.21 imply the third and the last inequalities in (3.26). So, it remains to prove the difference bounds in (3.27) and (3.28).

The first inequality in (3.27) can be proved as follows.

$$
\begin{aligned}
\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\hat{H}_\alpha(\boldsymbol{D}^*) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) &\overset{(a)}{\leq} \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\hat{H}_\alpha(\boldsymbol{D}^*) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}^*, \check{\boldsymbol{d}}) \\
&\overset{(b)}{\leq} \frac{1}{2}\left[\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha}\right] - \frac{1}{2}\left[\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha}\right. \\
&\quad \left. - \sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D^*_{\alpha-1}) + \sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha-1)\right] \\
&\leq \frac{1}{2}\sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log \alpha - \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha-1) \\
&= \frac{f_1(\boldsymbol{A})}{2}\log 2 + \frac{1}{2}\sum_{\alpha=2}^{K-1} f_\alpha(\boldsymbol{A})[\log(\alpha+1) - \log(\alpha-1)] - \frac{f_K(\boldsymbol{A})}{2}\log(K-1) \\
&\overset{(c)}{\leq} \frac{A_{\text{sum}}}{2}\log 2 + \frac{1}{2}\sum_{\alpha=2}^{K-1} \frac{A_{\text{sum}}}{\alpha}[\log(\alpha+1) - \log(\alpha-1)] - \frac{A_{\min}}{2}\log(K-1) \\
&\leq \frac{A_{\text{sum}}}{2}\sum_{\alpha=2}^{K} \frac{1}{\alpha-1}\log\alpha - \frac{A_{\text{sum}}}{2}\sum_{\alpha=2}^{K-1} \frac{1}{\alpha}\log(\alpha-1) - \frac{A_{\min}}{2}\log(K-1),
\end{aligned}
\tag{3.52}
$$

where both (a) and (b) are due to Corollary 3.44, and in (c) we have used Lemma 3.11.

To prove the bound in (3.28) of Theorem 3.24, we again combine Theorem 3.20, Theorem B.2, Theorem 3.23, Corollary 3.44, and notice the fact that

$$\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\tilde{H}_\alpha(\boldsymbol{D^*}) = \frac{1}{2} f_1(\boldsymbol{A}) \log \frac{1}{D_1^*} + \frac{1}{2}\sum_{\alpha=2}^{K-1} f_\alpha(\boldsymbol{A}) \log \frac{(\alpha-1)D_{\alpha-1}^*}{D_\alpha^*(\alpha - D_{\alpha-1}^*)}$$
$$+ \frac{1}{2} f_K(\boldsymbol{A}) \log \frac{(K-1)D_{K-1}^*}{D_K^*(K - D_{K-1}^*)}$$
$$= \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D_{\alpha-1}^*}{D_\alpha^*} - \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - D_{\alpha-1}^*)$$
$$+ \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - 1).$$

Therefore, similar to (3.52) we can write

$$\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\tilde{H}_\alpha(\boldsymbol{D^*}) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D}) \le \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A})\tilde{H}_\alpha(\boldsymbol{D^*}) - \underline{R}_{\boldsymbol{A}}(\boldsymbol{D^*}, \check{\boldsymbol{d}})$$
$$\le \frac{1}{2}\left[\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D_{\alpha-1}^*}{D_\alpha^*} - \sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - D_{\alpha-1}^*)\right.$$
$$\left. + \sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - 1)\right]$$
$$- \frac{1}{2}\left[\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D_{\alpha-1}^*}{D_\alpha^*} - \sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D_{\alpha-1}^*)\right.$$
$$\left. + \sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - 1)\right]$$
$$\le \frac{1}{2}\sum_{\alpha=2}^{K} [f_{\alpha-1}(\boldsymbol{A}) - f_\alpha(\boldsymbol{A})] \log \alpha$$

which shows that the first inequality in (3.28) holds. The proof of the second inequality is similar to that of (3.27), and by using Lemma 3.11.

3.8 Extension to General Sources

In this section we generalize the result for the Gaussian source to other sources under the MSE distortion measure , and show similar but looser bounds hold for the symmetric individual description rate-distortion (SID – RD) function as well as the rate-distortion region under the quadratic distortion measure. We derive the result using the SR – MLD scheme, but not the PPR multilayer scheme, which appears difficult to analyze for general sources. For $K = 2$ and the symmetric distortion constraints, the sum rate gap between the upper bound derived using the SR – MLD scheme and the rate-distortion function is upper-bounded by 1.5 bits, which is the same value as that derived in [61] for the two description case; nevertheless our result is stronger, since in [61] the achievable scheme is more involved than the $\mathsf{SR-MLD}$ scheme yet the bounding constant is the same.

Some additional definitions are necessary. For a general source X with finite differential entropy, zero mean and unit variance, define the following quantity,

$$\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) = \sum_{\alpha=1}^{K} \frac{1}{\alpha} I(X; Y_\alpha | Y_1, Y_2, \ldots, Y_{\alpha-1})$$

where random variable Y_α, $\alpha = 1, 2, ..., K$ are defined as in (3.31) and (3.32).

The following theorem is the main result of this section.

Theorem 3.46. *For any general source with unit variance under the* MSE *distortion measure, we have*

$$\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) \geq R_{\mathsf{SMD}}(\boldsymbol{D}),$$

moreover,

$$\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) - R_{\mathsf{SMD}}(\boldsymbol{D}) \leq \sum_{\alpha=1}^{K} \frac{1}{2\alpha}.$$

This theorem essentially states the the $\mathsf{SR-MLD}$ ($\mathsf{SR-ULP}$) scheme with the additive Gaussian codebook operates within $\sum_{\alpha=1}^{K}(2\alpha)^{-1}$ of the optimal coding scheme, in terms of individual description rate, for any source with unit variance. The first statement in the theorem is trivial by applying Theorem 3.29, and the second statement is proved in Appendix B.8.

Unlike Theorem 3.14, there is no explicit lower bound on the $\mathsf{SID-RD}$ function. Indeed, in the proof of Theorem 3.46, the outer bound is never explicitly written to have a single letter form or an analytical form that can be computed directly. The key proof idea is to construct the lower and upper bounds in appropriate forms such that certain terms are the same, and then cancel out these terms to bound the remaining terms.

Similar to the $\mathsf{SID-RD}$ approximation, we can extend the rate-distortion region approximation technique to general sources under the MSE distortion measure. It is clear that the definition of $\hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ is not limited to the Gaussian source. Hence, we can define the following rate region for arbitrary distributed source.

Definition 3.47. *Let* $\boldsymbol{Y}$ *be defined as in (3.31) and (3.32) for a general source* X *with zero mean and unit variance, and denote the corresponding rate region* $\hat{\mathcal{R}}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ *by* $\hat{\mathcal{R}}'_{\mathsf{SMD}}(\boldsymbol{D})$ *as define in Theorem 3.36. Moreover, for a given* $\boldsymbol{A} \in \mathbb{R}^K_+$ *and* $\boldsymbol{A} \neq 0$*, defined*

$$\hat{R}'_{\boldsymbol{A}}(\boldsymbol{D}) \triangleq \min_{\boldsymbol{R} \in \hat{\mathcal{R}}'_{\mathsf{SMD}}(\boldsymbol{D})} \boldsymbol{A} \cdot \boldsymbol{R}.$$

Now, we have the following theorem.

Theorem 3.48. *For any general source X with unit variance under the* MSE *distortion measure, we have*

$$\hat{\mathcal{R}}'_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}).$$

Moreover, for any $\boldsymbol{A} \in \mathbb{R}_+^K$ and $\boldsymbol{A} \neq 0$

$$\hat{R}'_{\boldsymbol{A}}(\boldsymbol{D}) - R_{\boldsymbol{A}}(\boldsymbol{D}) \leq \sum_{\alpha=1}^{K} \frac{f_\alpha(\boldsymbol{A})}{2}.$$

The proof of this theorem follows closely the proof for the general source sum-rate in this section, and it is thus omitted.

4 Asymmetric Multiple Description Coding

The multiple description problem analysis in Chapter 3, we derived an approximate characterization for the achievable rate region of the MD problem proved that the distortion constraints only depend on the number of available descriptions at the decoder, which is known as symmetric distortion constraint in the literature. However, asymmetric (unequal) link reliabilities as well as different available bandwidth for the descriptions, cause an asymmetric situation, wherein the quality of source reconstruction depends not only on the number of received descriptions, but also the specific subset of them. This problem is called *asymmetric multiple description* (AMD) coding, which is a generalization of the symmetric version of the problem.

In this problem a source is mapped into K descriptions and sent to $2^K - 1$ decoders. The decoders are required to reconstruct the source sequence within certain distortions using the available descriptions.

Since the AMD problem is an extension of the SMD problem, not surprisingly, the exact rate region characterization for this problem is still open for the quadratic Gaussian unless for the two description case [11, 22].

Compared to the symmetric MD problem, the asymmetric MD problem has received much less attention. Though some vanguard effort can be found in [23, 27, 28, 62], usually with only a subset of all possible distortion constraints, the asymmetric problem has not been systematically investigated for more than two descriptions, to the best of our knowledge.

We developed an intimate connection between the symmetric multilevel diversity (SMLD) coding problem and the symmetric multiple description (SMD) problem in Chapter 3. There we showed that for the SMD problem, achievable rate region based on successive refinement (SR) coding coupled with SMLD coding provides good approximation to the MD rate region under symmetric distortion constraints; perhaps more interestingly, the achievable rate region has the same geometric structure as that of the SMLD coding rate region. In fact, the SMLD coding result is essential for establishing the SMD result in [63].

A similar approach is used here in order to develop an approximate characterization for the achievable rate region of the asymmetric 3-description problem based on its lossless counterpart (see Figure 2.2 and Figure 4.1), the AMLD problem discussed in Chapter 2.

Using the intuitions gained in treating the AMLD problem as well as the sum-rate lower bound for symmetric Gaussian MD problem in Chapter 3, we develop inner and outer bounds for the AMD rate region, both of which bear similar geometric structure to the AMLD coding rate region. Moreover, the gap between the bounds is constant (less than 1.3 bits in terms of the Euclidean distance between the bounding planes), yielding an approximate characterization.

One surprising consequence of the result is that the proposed simple architecture based on successive refinement (SR) [54, 56, 57] and AMLD coding is in fact close to optimality. From an engineering viewpoint, this suggests that one can design simple and flexible MD codes that are approximately optimal.

This result suggests a general approach in treating lossy source coding problems: first solve a corresponding a lossless version of the problem, then extend the results and intuitions to its lossy counterpart to yield an approximate characterization. The result given in this chapter further illustrates the effectiveness of this approach.

4.1 Gaussian AMD: Problem Formulation

Let $\{X(t)\}_{t=1,2,\ldots}$ be a sequence of independent and identically distributed zero mean and unit variance real-valued Gaussian source, *i.e.,* $\mathcal{X} = \mathbb{R}$, with time index t. Moreover, assume $\mathcal{X} = \mathbb{R}$ be the reconstruction alphabet. The vector

$$(X(1), X(2), \ldots, X(n))$$

is denoted by X^n. We use capital letters for random variables, and the corresponding lower-case letters for their realization. The quality of the reconstruction is measured by the quadratic distance between the original sequence x^n and the reconstructed one $\hat{x}^n$. Formally, we define the distortion as

$$d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{k=1}^{n} |x(k) - \hat{x}(k)|^2, \tag{4.1}$$

In a general multiple description setting, the encoders produces K descriptions, namely $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$ based on the source sequence and sends them to the decoders through noiseless channels. Each decoder receives a non-empty subset of the descriptions, and has to reconstruct the source sequence $\hat{x}^n$ which satisfies a certain level of fidelity.

In a manner similar to the last chapters, we denote each decoder by a binary vector $\boldsymbol{v}$, the indicator of the corresponding set of available descriptions. Each decoder $\boldsymbol{v}$ has a distortion constraint $D_{\boldsymbol{v}}$, and needs to reconstruct the source such that the corresponding expected distortion does not exceed this constraint. The main goal in this problem is to characterize the set of achievable rates of the descriptions in a way that such reconstructions are possible. We present a formal definition of the problem next.

An $(n, \{M_i, i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}, \boldsymbol{v} \in \Omega_K\})$ MD-code is defined as a set of encoding functions

$$F_i : \mathcal{X}^n \longrightarrow \{1, 2, \ldots, M_i\}, \quad i \in \mathcal{I}_K, \tag{4.2}$$

and $2^K - 1$ decoding functions

$$G_{\boldsymbol{v}} : \prod_{j:v_j=1} \{1, \ldots, M_j\} \longrightarrow \mathcal{X}^n, \quad \boldsymbol{v} \in \Omega_K, \tag{4.3}$$

with

$$\Delta_{\boldsymbol{v}} = \mathbb{E}\left[d(X^n, \hat{X}^n_{\boldsymbol{v}})\right], \quad \boldsymbol{v} \in \Omega_K, \tag{4.4}$$

where

$$\hat{X}^n_{\boldsymbol{v}} = G_{\boldsymbol{v}}(F_j(X^n), j : v_j = 1). \tag{4.5}$$

where $\prod$ denotes set product, and $\mathbb{E}$ is the expectation operator.

A rate tuple $\boldsymbol{R} = (R_1, R_2, \ldots, R_K)$ is called $\boldsymbol{D} = (D_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K)$-achievable if for every $\epsilon > 0$ and sufficiently large n, there exists an $(n, \{M_i; i \in \mathcal{I}_K\}, \{\Delta_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\})$ MD-code such that

$$\frac{1}{n} \log M_i \leq R_i + \epsilon, \quad i = 1, \ldots, K, \tag{4.6}$$

and

$$\Delta_{\boldsymbol{v}} \leq D_{\boldsymbol{v}} + \epsilon, \quad \boldsymbol{v} \in \Omega_K. \tag{4.7}$$

We denote by $\mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$ the set of all $\boldsymbol{D}$-achievable rate tuples, which we seek to characterize.

We use the notation $\boldsymbol{u} \leq \boldsymbol{v}$ introduced in Definition 2.2. It is clear that if $\boldsymbol{u} \leq \boldsymbol{v}$, then the decoder $\boldsymbol{v}$ can reconstruct the source sequence as *well* as the decoder $\boldsymbol{u}$ does, even if $D_{\boldsymbol{u}} \leq D_{\boldsymbol{v}}$. The following lemma shows that a slight modification of the distortion vector in order to satisfy such property does not change the achievable rate region. The proof of the lemma is presented in Appendix C.1.

Lemma 4.1. *For a given distortion vector $\boldsymbol{D}$, define $\tilde{\boldsymbol{D}}$ as $\tilde{\boldsymbol{D}} = (\tilde{D}_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K)$, where*

$$\tilde{D}_{\boldsymbol{v}} = \min_{\boldsymbol{u}:\boldsymbol{u} \leq \boldsymbol{v}} D_{\boldsymbol{u}}.$$

Then $\mathcal{R}_{\mathsf{AMD}}(\tilde{\boldsymbol{D}}) = \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$.

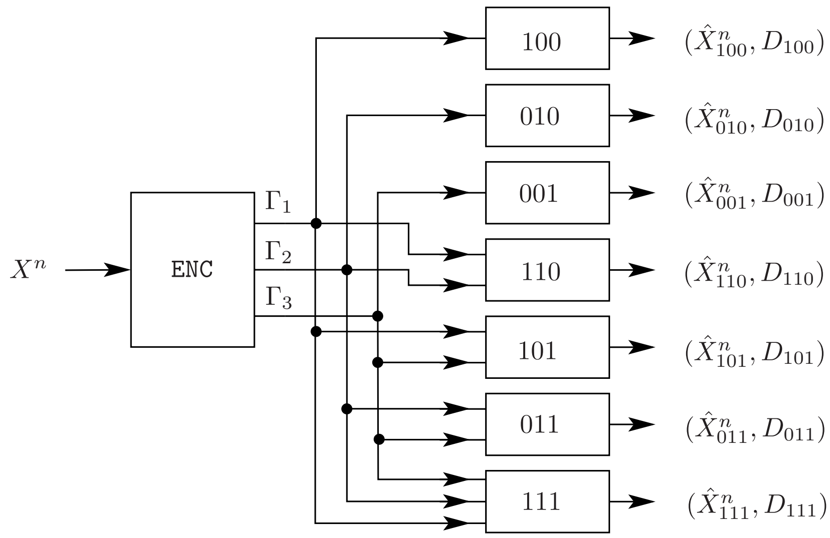


Figure 4.1: The three description source coding problem with ordering $\mathscr{L}_1$.

Given this lemma, we can assume, without loss of generality, that $D_{\boldsymbol{v}} \leq D_{\boldsymbol{u}}$ for all $\boldsymbol{u} \leq \boldsymbol{v}$. These distortion constraints then induce an ordering on the decoders, or equivalently on their associated subset of descriptions, which is consistent with the *ordering* notion defined in Definition 2.5.

In the rest of this chapter, we focus on the three description ($K = 3$) problem, and present the results in general form, *i.e.,* regardless of the exact underlying ordering. Occasionally we shall provide the proof details only for the specific sorted distortion constraints

$$D_{100} \geq D_{010} \geq D_{001} \geq D_{110} \geq D_{101} \geq D_{011} \geq D_{111},$$

which induces the ordering

$$\mathscr{L}(100) < \mathscr{L}(010) < \mathscr{L}(001) < \mathscr{L}(110) < \mathscr{L}(101) < \mathscr{L}(011) < \mathscr{L}(111)$$

on the subsets of descriptions, which is exactly the aforementioned ordering $\mathscr{L}_1$ in Table 2.1. Figure 4.1 shows the setting of this problem for the ordering $\mathscr{L}_1$. It is worth mentioning that the distortion constraints may also induce different ordering of subsets of the descriptions. All possible ordering functions are listed in Table 2.2.

4.2 Main Result

In this section we first establish an outer bound for the achievable rate region of the Gaussian asymmetric 3-description problem. We also present an inner bound for the achievable rate region. An approximate characterization for the AMD problem will be obtained by bounding that the gap between the inner and outer bounds.

We present the theorems in a unified way which hold for all orderings, and also specialize it to the ordering $\mathscr{L}_1$ to facilitate understanding and further discussion in next section.

Definition 4.2. *For a given distortion vector $\boldsymbol{D} = (D_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_3)$, denote by $\underline{\mathcal{R}}_{\mathsf{AMD}(\boldsymbol{D})}$ the set of all rate triples (R_1, R_2, R_3) satisfying*

$$R_i \geq \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \qquad i = 1,2,3 \tag{AMD – O1}$$

$$R_i + R_j \geq \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_j}}\right) + \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_j}} - 1, \qquad i \neq j \tag{AMD – O2}$$

$$\begin{aligned} 2R_i + R_j + R_k \geq{} & \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_j}}\right) + \min\left(\frac{1}{2}\log\frac{1}{D_{1i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_k}}\right) \\ & + \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_j}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_k}}\right) + \frac{1}{2}\log\frac{1}{D_{111}} - 3, \\ & i \neq j \neq k \end{aligned} \tag{AMD – O3}$$

$$\begin{aligned} R_1 + R_2 + R_3 \geq{} & \frac{1}{2}\log\frac{1}{D_{100}} + \min\left(\frac{1}{2}\log\frac{1}{D_{110}}, \frac{1}{2}\log\frac{1}{D_{001}}\right) \\ & + \frac{1}{2}\log\frac{1}{D_{111}} - 2, \end{aligned} \tag{AMD – O4}$$

$$\begin{aligned} R_1 + R_2 + R_3 \geq{} & \frac{1}{4}\log\frac{1}{D_{100}^2 D_{\Gamma_2}} + \min\left(\frac{1}{4}\log\frac{1}{D_{110}}, \frac{1}{4}\log\frac{1}{D_{101}}, \frac{1}{4}\log\frac{1}{D_{011}}\right) \\ & + \frac{1}{2}\log\frac{1}{D_{111}} - \frac{9}{4}. \end{aligned} \tag{AMD – O5}$$

Theorem 4.3 (An Outer Bound for the AMD Rate Region)**.** *Any achievable rate triple (R_1, R_2, R_3) belongs to $\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$,* i.e., $\mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$.

The bound stated in this theorem is a consequence of a more general parametric outer bound $\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}, \boldsymbol{d})$, defined in Theorem 4.11. However, the current form is more convenient for comparison between the inner and outer bounds. We will prove this theorem in Section 4.4.

In order to establish an inner bound for the AMD we need to define the following rate region. This together with Theorem 4.5 give an inner bound for the achievable rate region of the 3-description AMD problem.

Definition 4.4. *For a given distortion vector $\boldsymbol{D} = (D_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_3)$, let $\overline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$ be the set of all rate triples (R_1, R_2, R_3) satisfying*

$$R_i \geq \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \qquad i = 1,2,3 \tag{AMD − I1}$$

$$R_i + R_j \geq \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_j}}\right) + \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_j}}, \qquad i \neq j \tag{AMD − I2}$$

$$\begin{aligned} 2R_i + R_j + R_k \geq{} & \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_j}}\right) + \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_k}}\right) \\ & + \min\left(\frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_j}}, \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i+\mathbb{1}_k}}\right) + \frac{1}{2}\log\frac{1}{D_{111}}, \\ & i \neq j \neq k \end{aligned} \tag{AMD − I3}$$

$$R_1 + R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}} + \min\left(\frac{1}{2}\log\frac{1}{D_{110}}, \frac{1}{2}\log\frac{1}{D_{001}}\right) + \frac{1}{2}\log\frac{1}{D_{111}}, \tag{AMD − I4}$$

$$\begin{aligned} R_1 + R_2 + R_3 \geq{} & \frac{1}{4}\log\frac{1}{D_{100}^2 D_{010}} + \min\left(\frac{1}{2}\log\frac{1}{D_{110}}, \frac{1}{2}\log\frac{1}{D_{101}}, \frac{1}{2}\log\frac{1}{D_{011}}\right) \\ & + \frac{1}{2}\log\frac{1}{D_{111}}. \end{aligned} \tag{AMD − I5}$$

Theorem 4.5 (An Inner Bound for the AMD Rate Region)**.** *Any rate triple* $\boldsymbol{R} \in \overline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$ *is achievable,* i.e, $\overline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$.

This region can be achieved using AMLD coding along with SR, and it can be considered as the asymmetric counterpart of $\mathcal{R}^{\mathsf{SR}}_{\mathsf{SMD}}$ defined in Section 3.3. The proof of this theorem can be found in Section 4.5.

Summarizing the results of Theorem 4.3 and Theorem 4.5 gives the following corollary.

Corollary 4.6.

$$\overline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}). \tag{4.8}$$

The result of this corollary is that the multiple description achievable rate region is bounded between two sets of hyper-planes, which are pair-wise parallel. For each pair of parallel planes, we can compute the distance between them. Denote by $\delta_{(x,y,z)}$ the Euclidean distance between two parallel planes which are orthogonal to the vector (x,y,z). Then for the distortion constraints corresponding to ordering $\mathscr{L}_1$, we have

$$\delta_{(1,0,0)} = 0, \tag{4.9}$$

$$\delta_{(1,1,0)} \leq \frac{1}{\sqrt{2}} = 0.7071, \tag{4.10}$$

$$\delta_{(2,1,1)} \leq \frac{3}{\sqrt{6}} = 1.2247, \tag{4.11}$$

$$\delta_{(1,1,1)} \leq \frac{9}{4\sqrt{3}} = 1.2990, \tag{4.12}$$

where the denominators are the normalizing factors, corresponding to the length of the vector (x, y, z). This shows that the inner and outer bounds provide an approximate characterization for the achievable rate region, for which the Euclidean distance between the bounds in less than 1.3 in the worst case. Figure 4.2 shows a typical pair of inner and outer bounds for $\mathscr{L}_1$ ordering and the case $D_{010}D_{101} \leq D_{001}^2 \leq D_{010}D_{110}$, which is the lossy counterpart of the lossless AMLD problem with $h_4 \leq h_3 \leq h_4 + h_5$, discussed in Section 2.4, under regime II (see also Figure 2.6).

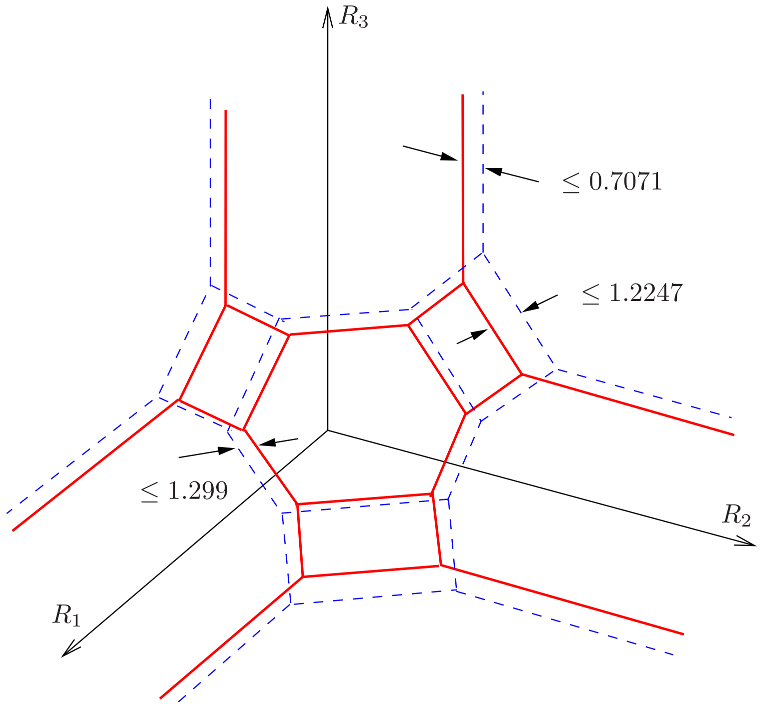


Figure 4.2: The inner and outer bound for the achievable rate region of the Gaussian multiple descriptions problem for distortion constraints corresponding to the ordering $\mathscr{L}_1$, and the case $D_{010}D_{101} \leq D_{001}^2 \leq D_{010}D_{110}$.

4.3 Illustration of the Result for Ordering $\mathscr{L}_1$

In this section we present the inner and outer bounds for the achievable rate region of the Gaussian AMD problem for the specific order $\mathscr{L}_1$. This would be helpful to illustrate the geometric structure of the region.

The following corollary is a consequence of Theorem 4.3 for the specific ordering $\mathscr{L}_1$.

Corollary 4.7. *Any achievable rate triple for a three-description* AMD *with* $\mathscr{L}_1$ *ordering satisfies*

$$R_1 \geq \frac{1}{2}\log\frac{1}{D_{100}}, \tag{AMD $-\mathscr{L}_1-$ O1}$$

$$R_2 \geq \frac{1}{2}\log\frac{1}{D_{010}}, \tag{AMD $-\mathscr{L}_1-$ O2}$$

$$R_3 \geq \frac{1}{2}\log\frac{1}{D_{001}}, \tag{AMD $-\mathscr{L}_1-$ O3}$$

$$R_1 + R_2 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{110}} - 1, \tag{AMD $-\mathscr{L}_1-$ O4}$$

$$R_1 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{101}} - 1, \tag{AMD $-\mathscr{L}_1-$ O5}$$

$$R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{010}D_{011}} - 1, \tag{AMD $-\mathscr{L}_1-$ O6}$$

$$2R_1 + R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}^2D_{110}D_{111}} - 3, \tag{AMD $-\mathscr{L}_1-$ O7}$$

$$R_1 + 2R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{010}D_{110}D_{111}} - 3, \tag{AMD $-\mathscr{L}_1-$ O8}$$

$$R_1 + R_2 + 2R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{010}D_{101}D_{111}} - 3, \tag{AMD $-\mathscr{L}_1-$ O9}$$

$$R_1 + R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{001}D_{111}} - 2, \tag{AMD $-\mathscr{L}_1-$ O10}$$

$$R_1 + R_2 + R_3 \geq \frac{1}{4}\log\frac{1}{D_{100}^2D_{010}D_{110}D_{111}^2} - \frac{9}{4}. \tag{AMD $-\mathscr{L}_1-$ O11}$$

Similarly, in the following corollary we specify the result of Theorem 4.5 for the ordering $\mathscr{L}_1$.

Corollary 4.8. *If the distortion constraints satisfy the ordering* $\mathscr{L}_1$*,* i.e.,

$$D_{100} \geq D_{101} \geq D_{001} \geq D_{110} \geq D_{101} \geq D_{011} \geq D_{111},$$

then any rate triple (R_1, R_2, R_3) *satisfying*

$$R_1 \geq \frac{1}{2}\log\frac{1}{D_{100}}, \tag{AMD $-\mathscr{L}_1-$ I1}$$

$$R_2 \geq \frac{1}{2}\log\frac{1}{D_{010}}, \tag{AMD $-\mathscr{L}_1-$ I2}$$

$$R_3 \geq \frac{1}{2}\log\frac{1}{D_{001}}, \tag{AMD $-\mathscr{L}_1-$ I3}$$

$$R_1 + R_2 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{110}}, \tag{AMD $-\mathscr{L}_1-$ I4}$$

$$R_1 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{101}}, \tag{AMD $-\mathscr{L}_1-$ I5}$$

$$R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{010}D_{011}}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I6})$$

$$2R_1 + R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}^2 D_{110}D_{111}}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I7})$$

$$R_1 + 2R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{010}D_{110}D_{111}}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I8})$$

$$R_1 + R_2 + 2R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{001}^2 D_{111}}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I9})$$

$$R_1 + R_2 + R_3 \geq \frac{1}{2}\log\frac{1}{D_{100}D_{001}D_{111}}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I10})$$

$$R_1 + R_2 + R_3 \geq \frac{1}{4}\log\frac{1}{D_{100}^2 D_{010}D_{110}D_{111}^2}, \qquad (\mathsf{AMD} - \mathscr{L}_1 - \mathsf{I11})$$

is achievable.

Note that both bounds are characterized by a set of hyper-planes, which are pair-wise parallel. Moreover, the distance between the parallel planes are constants, independent of the distortion constraints.

4.4 AMD: The Outer Bound

In this section we prove Theorem 4.3. In order to prove this theorem, we first show a parametric outer-bound for the AMD rate region. Then we specialize the parameters to obtain the bound claimed in the theorem.

We first need to define a set of auxiliary random variables in order to state and prove the parametric bound, which are some noisy versions of the source. The strategy of expanding the probability space by a single auxiliary variable was used to characterize the two descriptions Gaussian MD region [22], and later in [31] extended to include multiple auxiliary random variables with certain built-in Markov structure. These variables are similar to the noisy versions of the source defined in (3.31). We shall continue to use this extended strategy as used in [31].

Definition 4.9. *Let $N_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, \ldots, 6$, be mutually independent zero-mean Gaussian random variables with variance σ_i^2. They are also assumed to be independent of X. A noisy version of the source, Y_i, is defined as*

$$Y_i = X + \sum_{j=i}^{6} N_j, \qquad i = 1, \ldots, 6. \tag{4.13}$$

For convenience, we denote the effective noise in Y_i by $Z_i = \sum_{j=i}^{6} N_j$ for $i = 1, \ldots, 6$. Thus $d_i \triangleq \sum_{j=i}^{6} \sigma_j^2$ would be the variance of the noise Z_i, for $i = 1, \ldots, 6$. We also define $Y_7 = X$ and $d_7 = 0$. Note that incremental noises are added to X to build Y_i's, and therefore they form a Markov chain as

$$(\Gamma_1, \Gamma_2, \Gamma_3) \leftrightarrow X^n \leftrightarrow Y_6^n \leftrightarrow Y_5^n \leftrightarrow \cdots \leftrightarrow Y_1^n. \tag{4.14}$$

The following theorem provides a parametric outer-bound for the rate region of the AMLD problem, depending on d_i variables, which are the noise variances defined above. Such bound holds for any choice of $d_1 \geq d_2 \geq \cdots \geq d_6 > 0$, and can be further optimized to obtain a good non-parametric outer-bound for the rate region. However, we simply derive the bound in Theorem 4.3 by setting the values of d_i's.

Definition 4.10. *For a given distortion vector $\boldsymbol{D} = (D_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_3)$ and a set of variables $d_1 \geq d_2 \geq \cdots \geq d_6 > d_7 = 0$, denote by $\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}, \boldsymbol{d})$ the set of all rate triples (R_1, R_2, R_3) satisfying*

$$R_i \geq \frac{1}{2}\log\frac{1}{D_{\mathbb{1}_i}} \qquad j = 1,2,3 \tag{AMD -- OP1}$$

$$\begin{aligned} R_i + R_j \geq{}& \frac{1}{2}\log\frac{1+d_{\mathscr{L}(\mathbb{1}_i)}}{D_{\mathbb{1}_j}+d_{\mathscr{L}(\mathbb{1}_i)}}\frac{1+d_{\mathscr{L}(\mathbb{1}_j)}}{D_{\mathbb{1}_j}+d_{\mathscr{L}(\mathbb{1}_j)}} \\ &+\frac{1}{2}\log\frac{\left(D_{\mathbb{1}_i+\mathbb{1}_j}+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}}\right)}{\left(1+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}}\right)D_{\mathbb{1}_i+\mathbb{1}_j}} \qquad i \neq j \end{aligned} \tag{AMD -- OP2}$$

$$\begin{aligned} 2R_i + R_j + R_k \geq{}& \frac{1}{2}\log\left(\frac{1+d_{\mathscr{L}(\mathbb{1}_i)}}{D_{\mathbb{1}_i}+d_{\mathscr{L}(\mathbb{1}_i)}}\right)^2\frac{1+d_{\mathscr{L}(\mathbb{1}_j)}}{D_{\mathbb{1}_j}+d_{\mathscr{L}(\mathbb{1}_j)}}\frac{1+d_{\mathscr{L}(\mathbb{1}_k)}}{D_{\mathbb{1}_k}+d_{\mathscr{L}(\mathbb{1}_k)}} \\ &+\frac{1}{2}\log\frac{\left(1+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}\right)\left(D_{\mathbb{1}_i+\mathbb{1}_j}+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}}\right)}{\left(1+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}}\right)\left(D_{\mathbb{1}_i+\mathbb{1}_j}+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}\right)} \\ &+\frac{1}{2}\log\frac{\left(1+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)}\right)\left(D_{\mathbb{1}_i+\mathbb{1}_k}+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}}\right)}{\left(1+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}}\right)\left(D_{\mathbb{1}_i+\mathbb{1}_k}+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)}\right)} \\ &+\frac{1}{2}\log\frac{\left(D_{111}+d_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}\right)}{\left(1+d_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}\right)D_{111}} \qquad i \neq j \neq k \end{aligned} \tag{AMD -- OP3}$$

$$\begin{aligned} R_1 + R_2 + R_3 \geq{}& \frac{1}{2}\log\frac{1+d_{\mathscr{L}(100)}}{D_{100}+d_{\mathscr{L}(100)}} \\ &+\frac{1}{2}\frac{1+d_{\mathscr{L}(010)}}{D_{010}+d_{\mathscr{L}(010)}}\frac{1+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}}{D_{001}+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}} \\ &+\frac{1}{2}\log\frac{\left(1+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}\right)\left(D_{110}+d_{\mathscr{L}(010)}\right)}{\left(1+d_{\mathscr{L}(010)}\right)\left(D_{110}+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}\right)} \\ &+\frac{1}{2}\log\frac{D_{111}+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}}{\left(1+d_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}}\right)D_{111}} \end{aligned} \tag{AMD -- OP4}$$

$$\begin{aligned}R_1+R_2+R_3 \geq \frac{1}{2}\log \frac{1+d_{\mathscr{L}(100)}}{D_{100}+d_{\mathscr{L}(100)}}\frac{1+d_{\mathscr{L}(010)}}{D_{010}+d_{\mathscr{L}(010)}}\frac{1+d_{\mathscr{L}(001)}}{D_{001}+d_{\mathscr{L}(001)}}\\
+\frac{1}{4}\log \frac{(1+d_\alpha)(D_{110}+d_{\mathscr{L}(010)})}{(1+d_{\mathscr{L}(010)})(D_{110}+d_\alpha)}\\
+\frac{1}{4}\log \frac{(1+d_\alpha)(D_{101}+d_{\mathscr{L}(001)})}{(1+d_{\mathscr{L}(001)})(D_{101}+d_\alpha)}\\
+\frac{1}{4}\log \frac{(1+d_\alpha)(D_{011}+d_{\mathscr{L}(001)})}{(1+d_{\mathscr{L}(001)})(D_{011}+d_\alpha)}+\frac{1}{2}\log\frac{D_{111}+d_\alpha}{(1+d_\alpha)D_{111}},\end{aligned} \tag{AMD − OP5}$$

where

$$\alpha = \begin{cases} \mathscr{L}(001) & \textit{if } \mathscr{L}(001) > \mathscr{L}(110),\\ \min\{\mathscr{L}(110),\mathscr{L}(101),\mathscr{L}(011)\} & \textit{if } \mathscr{L}(001) < \mathscr{L}(110).\end{cases}$$

Theorem 4.11 (A Parametric Outer Bound for the AMD Rate Region)**.** *Any achievable rate triple $\boldsymbol{R} \in \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$ belongs to $\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D},\boldsymbol{d})$ for any choice of $d_1 \geq d_2 \geq \cdots \geq d_6 \geq d_7 = 0$,* i.e., $\mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D},\boldsymbol{d})$.

We need some tools in order to prove the theorem presented above. The following lemma extracted from [51] is very similar to Lemma 3.32. However, we rephrase the lemma to adapt it with the notation used here. We skip the proof of the lemma, since it closely follows the proof of Lemma 3.32 given in Appendix B.3

Lemma 4.12. *For any set of descriptions $\Gamma_{\boldsymbol{v}}$ indicated by $\boldsymbol{v} \in \Omega_3$, and noisy version of the source Y_i, $i = 1,2,\ldots,7$, we have*

$$I(\Gamma_{\boldsymbol{v}};Y_i^n) \geq \frac{n}{2}\log\frac{1+d_i}{D_{\boldsymbol{v}}+d_i}. \tag{4.15}$$

Moreover, for any subset of the descriptions $\Gamma_{\boldsymbol{v}}$, and two noisy versions of the source Y_i and Y_j with $i<j$, we have

$$I(\Gamma_{\boldsymbol{v}};Y_j^n) - I(\Gamma_{\boldsymbol{v}};Y_i^n) \geq \frac{n}{2}\log\frac{(1+d_j)(D_{\boldsymbol{v}}+d_i)}{(1+d_i)(D_{\boldsymbol{v}}+d_j)}. \tag{4.16}$$

This lemma is useful to bound the mutual information between the noisy versions of the source and the descriptions. We will use these results in several points in the proof of Theorem 4.11, which are indicated by (†). Now, we are ready to prove the parametric outer-bound.

Proof of Theorem 4.11. The single description inequalities are just straight forward result of Lemma 4.12. For $\boldsymbol{v} = \mathbb{1}_i$ we have

$$nR_i \geq H(\Gamma_i) = H(\Gamma_i) - H(\Gamma_i|X^n) = I(\Gamma_i;X^n) \overset{(\dagger)}{\geq} \frac{n}{2}\log\frac{1}{D_{\mathbb{1}_i}} \tag{4.17}$$

where we used Lemma 4.12 for $Y_7 = X$ and the fact $d_7 = 0$ in the last inequality. This proves $(\mathsf{AMD-OP1})$.

The bound for the two description rates in $(\mathsf{AMD-OP2})$ follows from

$$
\begin{aligned}
n(R_i + R_j) &\geq H(\Gamma_i) + H(\Gamma_j) \\
&\overset{(a)}{\geq} H(\Gamma_i) + H(\Gamma_j) - H(\Gamma_i, \Gamma_j | X^n) \\
&\quad - \Big[H(\Gamma_i | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) + H(\Gamma_j | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) \\
&\qquad - H(\Gamma_i, \Gamma_j | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) \Big] \\
&= I(\Gamma_i; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) + I(\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) \\
&\quad + [I(\Gamma_i \Gamma_j; X^n) - I(\Gamma_i \Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}})] \\
&\overset{(b)}{\geq} I(\Gamma_i; Y^n_{\mathscr{L}(\mathbb{1}_i)}) + I(\Gamma_j; Y^n_{\mathscr{L}(\mathbb{1}_j)}) \\
&\quad + [I(\Gamma_i \Gamma_j; X^n) - I(\Gamma_i \Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}})] \\
&\overset{(\dagger)}{\geq} \frac{n}{2} \log \frac{1 + d_{\mathscr{L}(\mathbb{1}_i)}}{D_{\mathbb{1}_i} + d_{\mathscr{L}(\mathbb{1}_i)}} \frac{1 + d_{\mathscr{L}(\mathbb{1}_j)}}{D_{\mathbb{1}_j} + d_{\mathscr{L}(\mathbb{1}_j)}} \frac{(1+0)(D_{\mathbb{1}_i + \mathbb{1}_j} + d_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}})}{(1 + d_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}})(D_{\mathbb{1}_i + \mathbb{1}_j} + 0)}
\end{aligned}
\tag{4.18}
$$

where the subtracted terms in (a) are positive due to the fact that Γ_i and Γ_j are functions of X^n and non-negativity of mutual information, (b) is by the data processing inequality and the Markov chain in (4.14). Finally, we have used Lemma 4.12 in $(\dagger)$.

The inequality $(\mathsf{AMD-OP3})$ can be proved through the following chain of inequalities.

$$
\begin{aligned}
n(2R_i + R_j + R_k) &\geq 2H(\Gamma_i) + H(\Gamma_j) + H(\Gamma_k) \\
&\overset{(a)}{\geq} 2H(\Gamma_i) + H(\Gamma_j) + H(\Gamma_k) - H(\Gamma_i \Gamma_j \Gamma_k | X^n) \\
&\quad - \Big[H(\Gamma_i | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) + H(\Gamma_j | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_i)\}}) \\
&\qquad - H(\Gamma_i \Gamma_j | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) \Big] \\
&\quad - \Big[H(\Gamma_i | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_k)\}}) + H(\Gamma_k | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_k)\}}) \\
&\qquad - H(\Gamma_i \Gamma_k | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_k)\}}) \Big] \\
&\quad - \Big[H(\Gamma_i \Gamma_j | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i + \mathbb{1}_j), \mathscr{L}(\mathbb{1}_i + \mathbb{1}_k)\}}) \\
&\qquad + H(\Gamma_i \Gamma_k | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i + \mathbb{1}_j), \mathscr{L}(\mathbb{1}_i + \mathbb{1}_k)\}}) \\
&\qquad - H(\Gamma_i \Gamma_j \Gamma_k | Y^n_{\max\{\mathscr{L}(\mathbb{1}_i + \mathbb{1}_j), \mathscr{L}(\mathbb{1}_i + \mathbb{1}_k)\}}) \Big] \\
&= I(\Gamma_i; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) + I(\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_j)\}}) \\
&\quad + I(\Gamma_i; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_k)\}}) + I(\Gamma_k; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i), \mathscr{L}(\mathbb{1}_k)\}})
\end{aligned}
$$

$$+ [I(\Gamma_i\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}) - I(\Gamma_i\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})]$$
$$+ [I(\Gamma_i\Gamma_k; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}) - I(\Gamma_i\Gamma_k; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}})]$$
$$+ [I(\Gamma_i\Gamma_j\Gamma_k; X^n) - I(\Gamma_i\Gamma_j\Gamma_k; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}})] \tag{4.19}$$

where in (a) we have used the fact that all the brackets are non-negative. Now, we will bound each term in (4.19) individually. The single description terms can be bounded as

$$I(\Gamma_i; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_k)\}}) \geq I(\Gamma_i; Y^n_{\mathscr{L}(\mathbb{1}_i)}) \overset{(\dagger)}{\geq} \frac{n}{2}\log\frac{1+d_{\mathscr{L}(\mathbb{1}_i)}}{D_{\mathbb{1}_i}+d_{\mathscr{L}(\mathbb{1}_i)}}, \tag{4.20}$$

and similarly for j and k. Also we can bound the differential terms as

$$\begin{aligned}
I(\Gamma_i\Gamma_j;&Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}}) - I(\Gamma_i\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})\\
&\overset{(b)}{\geq} I(\Gamma_i\Gamma_j; Y^n_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}) - I(\Gamma_i\Gamma_j; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})\\
&\overset{(\dagger)}{\geq} \frac{n}{2}\log\frac{(1+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_i)})(D_{\Gamma_i\Gamma_j}+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})}{(1+d_{\max\{\mathscr{L}(\mathbb{1}_i),\mathscr{L}(\mathbb{1}_j)\}})(D_{\mathbb{1}_i+\mathbb{1}_j}+d_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)})}
\end{aligned} \tag{4.21}$$

where (b) is due to the data processing inequality implied by the Markov chain

$$(\Gamma_i\Gamma_j) \leftrightarrow Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}} \leftrightarrow Y^n_{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j)}$$

implied by (4.14). We also have

$$\begin{aligned}
I(\Gamma_i\Gamma_j\Gamma_k; X^n) &- I(\Gamma_i\Gamma_j\Gamma_k; Y^n_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}})\\
&\overset{(\dagger)}{\geq} \frac{n}{2}\log\frac{(1+0)(D_{111}+d_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}})}{(1+d_{\max\{\mathscr{L}(\mathbb{1}_i+\mathbb{1}_j),\mathscr{L}(\mathbb{1}_i+\mathbb{1}_k)\}})(D_{111}+0)}.
\end{aligned} \tag{4.22}$$

By replacing (4.20)–(4.22) in (4.19) we get the desired inequality.

In order to derive the sum-rate bound in $(\mathsf{AMD}-\mathsf{OP}4)$, we can write

$$\begin{aligned}
n(R_1+R_2+R_3) &\geq H(\Gamma_1)+H(\Gamma_2)+H(\Gamma_3)\\
&\geq H(\Gamma_1)+H(\Gamma_2)+H(\Gamma_3)-H(\Gamma_1\Gamma_2\Gamma_3|X^n)\\
&\quad -\Big[H(\Gamma_1|Y^n_{\mathscr{L}(010)})+H(\Gamma_2|Y^n_{\mathscr{L}(010)})-H(\Gamma_1\Gamma_2|Y^n_{\mathscr{L}(010)})\Big]\\
&\quad -\Big[H(\Gamma_1\Gamma_2|Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})+H(001|Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})\\
&\qquad -H(\Gamma_1\Gamma_2\Gamma_3|Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})\Big]\\
&\geq I(\Gamma_1;Y^n_{\mathscr{L}(100)})+I(\Gamma_2;Y^n_{\mathscr{L}(010)})+I(\Gamma_3;Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})\\
&\quad +\Big[I(\Gamma_1\Gamma_2;Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})-I(\Gamma_1\Gamma_2;Y^n_{\mathscr{L}(010)})\Big]\\
&\quad +\Big[I(\Gamma_1\Gamma_2\Gamma_3;X^n)-I(\Gamma_1\Gamma_2\Gamma_3;Y^n_{\min\{\mathscr{L}(110),\mathscr{L}(001)\}})\Big].
\end{aligned} \tag{4.23}$$

Again, applying Lemma 4.12 we can bound each term in (4.23), and obtain $(\mathsf{AMD}-\mathsf{OP4})$.

It remains to show the inequality in $(\mathsf{AMD}-\mathsf{OP5})$. Recall the proof of (AMLD-5), and consider two cases. If $\mathscr{L}(001) > \mathscr{L}(110)$, then Using a similar argument as in the proof of $(\mathsf{AMD}-\mathsf{OP3})$, we obtain

$$\begin{aligned}
n(R_1+R_2+R_3) &\geq H(\Gamma_1)+H(\Gamma_2)+H(\Gamma_3)-H(\Gamma_1\Gamma_2\Gamma_3|X^n)\\
&\quad-\frac{1}{2}\left[H(\Gamma_1|Y^n_{\mathscr{L}(010)})+H(\Gamma_2|Y^n_{\mathscr{L}(010)})-H(\Gamma_1\Gamma_2|Y^n_{\mathscr{L}(010)})\right]\\
&\quad-\frac{1}{2}\left[H(\Gamma_1|Y^n_{\mathscr{L}(001)})+H(\Gamma_3|Y^n_{\mathscr{L}(001)})-H(\Gamma_1\Gamma_3|Y^n_{\mathscr{L}(001)})\right]\\
&\quad-\frac{1}{2}\left[H(\Gamma_2|Y^n_{\mathscr{L}(001)})+H(\Gamma_3|Y^n_{\mathscr{L}(001)})-H(\Gamma_2\Gamma_3|Y^n_{\mathscr{L}(001)})\right]\\
&\quad-\frac{1}{2}\Big[H(\Gamma_1\Gamma_2|Y^n_{\mathscr{L}(001)})+H(\Gamma_1\Gamma_3|Y^n_{\mathscr{L}(001)})\\
&\qquad\quad+H(\Gamma_2\Gamma_3|Y^n_{\mathscr{L}(001)})-2H(\Gamma_1\Gamma_2\Gamma_3|Y^n_{\mathscr{L}(001)})\Big]\\
&\geq I(\Gamma_1;Y^n_{\mathscr{L}(100)})+I(\Gamma_2;Y^n_{\mathscr{L}(010)})+I(\Gamma_3;Y^n_{\mathscr{L}(001)})\\
&\quad+\frac{1}{2}[I(\Gamma_1\Gamma_2;Y^n_{\mathscr{L}(001)})-I(\Gamma_1\Gamma_2;Y^n_{\mathscr{L}(010)})]\\
&\quad+[I(\Gamma_1\Gamma_2\Gamma_3;X^n)-I(\Gamma_1\Gamma_2\Gamma_3;Y^n_{\mathscr{L}(001)})],
\end{aligned}\tag{4.24}$$

which gives us the desired inequality by using Lemma 4.12 to bound each individual term. Similarly, for the case where $\mathscr{L}(001) < \mathscr{L}(110)$ we can write

$$\begin{aligned}
n(R_1+R_2+R_3) &\geq H(\Gamma_1)+H(\Gamma_2)+H(\Gamma_3)-H(\Gamma_1\Gamma_2\Gamma_3|X^n)\\
&\quad-\frac{1}{2}\left[H(\Gamma_1|Y^n_{\mathscr{L}(010)})+H(\Gamma_2|Y^n_{\mathscr{L}(010)})-H(\Gamma_1\Gamma_2|Y^n_{\mathscr{L}(010)})\right]\\
&\quad-\frac{1}{2}\left[H(\Gamma_1|Y^n_{\mathscr{L}(001)})+H(\Gamma_3|Y^n_{\mathscr{L}(001)})-H(\Gamma_1\Gamma_3|Y^n_{\mathscr{L}(001)})\right]\\
&\quad-\frac{1}{2}\left[H(\Gamma_2|Y^n_{\mathscr{L}(\Gamma_3)})+H(\Gamma_3|Y^n_{\mathscr{L}(001)})-H(\Gamma_2\Gamma_3|Y^n_{\mathscr{L}(001)})\right]\\
&\quad-\frac{1}{2}\Big[H(\Gamma_1\Gamma_2|Y^n_\beta)+H(\Gamma_1\Gamma_3|Y^n_\beta)+H(\Gamma_2\Gamma_3|Y^n_\beta)\\
&\qquad\quad-2H(\Gamma_1\Gamma_2\Gamma_3|Y^n_\beta)\Big]\\
&\geq I(\Gamma_1;Y^n_{\mathscr{L}(100)})+I(\Gamma_2;Y^n_{\mathscr{L}(010)})+I(\Gamma_3;Y^n_{\mathscr{L}(001)})\\
&\quad+\frac{1}{2}[I(\Gamma_1\Gamma_2;Y^n_\beta)-I(\Gamma_1\Gamma_2;Y^n_{\mathscr{L}(010)})]\\
&\quad+\frac{1}{2}[I(\Gamma_1\Gamma_3;Y^n_\beta)-I(\Gamma_1\Gamma_3;Y^n_{\mathscr{L}(001)})]\\
&\quad+\frac{1}{2}[I(\Gamma_2\Gamma_3;Y^n_\beta)-I(\Gamma_2\Gamma_3;Y^n_{\mathscr{L}(001)})]\\
&\quad+[I(\Gamma_1\Gamma_2\Gamma_3;X^n)-I(\Gamma_1\Gamma_2\Gamma_3;Y^n_\beta)].
\end{aligned}\tag{4.25}$$

where $\beta = \min\{\mathscr{L}(110), \mathscr{L}(101), \mathscr{L}(011)\}$. Now, we can use the above-mentioned lemmas again to bound each individual term. It is clear that (4.24) and (4.25) give $(\mathsf{AMD} - \mathsf{OP}5)$. □

Remark 4.13. *Note that there is an one-to-one correspondence between the converse proof of Theorem 2.7 and that of Theorem 4.11. In fact, here we use the description subsets and their capability of lossy recovering the noisy source layers, where they have been used to losslessly reconstruct the source levels in the* AMLD.

Now we are ready to prove Theorem 4.3, which is a direct consequence of Theorem 4.11.

Proof of Theorem 4.3. We can choose arbitrary values of d_i's, the variance of the additive noise in Theorem 4.11, such that $d_1 \geq d_2 \geq \cdots \geq d_6 > 0$. One can optimize the bound in Theorem 4.11 with respect to the values of d_i's, and obtain a bound isolated from d_i's, by replacing them with the optimal choices. Such bound would be the best that can be found using this method. However instead of solving such a difficult optimization problem, we choose $d_i = D_{\mathscr{L}^{-1}(i)}$, for $i = 1, \ldots, 6$. It is clear the d_i's satisfy the desired non-increasing order due to the definition of the ordering. We will later show that this choice gives a bound which is within constant bit gap from the inner bound in Theorem 4.5.

The single description rate inequalities are exactly the same. The proof of the other inequalities is by straightforward evaluation of their counterparts in Theorem 4.11, for $d_i = D_{\mathscr{L}^{-1}(i)}$, and applying simple bounds. We do not repeat the same arguments here, and only illustrate such derivation for one simple case. For the sum of two description rates, we can start with $(\mathsf{AMD} - \mathsf{OP}2)$ and use $d_i = D_{\mathscr{L}^{-1}(i)}$ to get

$$\begin{aligned}
R_i + R_j &\overset{(a)}{\geq} \frac{1}{2}\log \frac{1 + D_{\mathbb{1}_i}}{D_{\mathbb{1}_i} + D_{\mathbb{1}_i}} \frac{1 + D_{\mathbb{1}_j}}{D_{\mathbb{1}_j} + D_{\mathbb{1}_j}} \frac{D_{\mathbb{1}_i+\mathbb{1}_j} + \min(D_{\mathbb{1}_i}, D_{\mathbb{1}_j})}{(1 + \min(D_{\mathbb{1}_i}, D_{\mathbb{1}_j}))D_{\mathbb{1}_i+\mathbb{1}_j}} \\
&\overset{(b)}{=} \frac{1}{2}\log \frac{1 + \max(D_{\mathbb{1}_i}, D_{\mathbb{1}_j})}{4D_{\mathbb{1}_i}D_{\mathbb{1}_j}} \frac{D_{\mathbb{1}_i+\mathbb{1}_j} + \min(D_{\mathbb{1}_i}, D_{\mathbb{1}_j})}{D_{\mathbb{1}_i+\mathbb{1}_j}} \\
&\overset{(c)}{\geq} \frac{1}{2}\log \frac{\min(D_{\mathbb{1}_i}, D_{\mathbb{1}_j})}{4D_{\mathbb{1}_i}D_{\mathbb{1}_j}D_{\mathbb{1}_i+\mathbb{1}_j}} \\
&= -1 + \frac{1}{2}\log \frac{1}{\max(D_{\mathbb{1}_i}, D_{\mathbb{1}_j})} + \frac{1}{2}\log \frac{1}{D_{\mathbb{1}_i+\mathbb{1}_j}},
\end{aligned} \tag{4.26}$$

where we have also used the fact $d_{\max(a,b)} = \min(d_a, d_b)$ in (a) which is implied by decreasing ordering of d_i's, (b) is due to the fact that $(1 + x)(1 + y) = (1+\min(x, y))(1+\max(x, y))$, and (c) holds since $D_{\boldsymbol{v}}$'s are non-negative. Similar simple manipulations give the other bounds in Theorem 4.3.

□

4.5 AMD: The Achievability Scheme and Inner Bound

Our approach to prove Theorem 4.5 is to present a simple scheme with description rates satisfying $(\mathsf{AMD}-\mathsf{I}1)$–$(\mathsf{AMD}-\mathsf{I}5)$ which guarantees the distortion constraints. This scheme is based on the successive refinability of Gaussian sources [54–57], as well as the asymmetric multilevel diversity coding result presented in the previous section. In the encoding scheme, we first produce seven successive refinement layers of the source, and then encode them losslessly.

4.5.1 Successive Refinement Coding

Consider the non-increasing sequence of distortion constraints

$$\boldsymbol{D}' = (D_{\mathscr{L}^{-1}(1)}, D_{\mathscr{L}^{-1}(2)}, \dots, D_{\mathscr{L}^{-1}(7)}).$$

Produce seven layers of successive refinement (SR), Ψ_k for $k = 1, 2, \dots, 7$, such that one can reconstruct the source sequence within distortion constraint $D_{\mathscr{L}^{-1}(k)}$ using $\Psi_1, \dots, \Psi_k$. Since the Gaussian source is successively refinable [56], it is clear that Ψ_k can be encoded to a binary block of length arbitrary close to

$$nh'_k \triangleq nR(D_{\mathscr{L}^{-1}(k)}) - nR(D_{\mathscr{L}^{-1}(k-1)}) \tag{4.27}$$

where $R(D) = -\frac{1}{2}\log D$ is the unit variance Gaussian R-D function, and $D_{\mathscr{L}^{-1}(0)} \triangleq 1$. Note that by using fixed length code in SR coding, these blocks are block-wise independently and identically distributed.

4.5.2 Multilevel Diversity Coding

Now, it only remains to produce the descriptions such that the decoder at level $\mathscr{L}(\boldsymbol{v})$ can losslessly recover the pre-coded bit-stream SR layers $\Psi_1, \dots, \Psi_{\mathscr{L}(\boldsymbol{v})}$, and then reconstruct the Gaussian source sequence within distortion $D_{\boldsymbol{v}}$. Encoding and decoding of the pre-coded SR layers are exactly the AMLD problem. We can simply use the rate region characterization of the AMLD problem in Theorem 2.7 to find the achievable rate region of the proposed scheme for AMD, where only substitution of $V_k = \Psi_k$ and $U_k = (\Psi_1, \dots, \Psi_k)$ is needed. Therefore we have

$$h_k = \frac{1}{2}\log\frac{D_{\mathscr{L}^{-1}(k-1)}}{D_{\mathscr{L}^{-1}(k)}} \tag{4.28}$$

and

$$H_k = \sum_{j=1}^{k} h_j = \frac{1}{2}\log\frac{1}{D_{\mathscr{L}^{-1}(k)}}. \tag{4.29}$$

Replacing the values of H_k's in Theorem 2.7, we obtain Theorem 4.5.

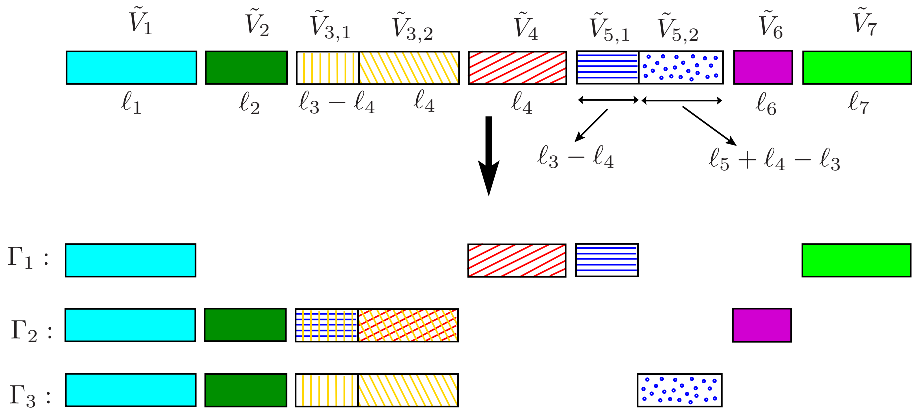


Figure 4.3: Lossless description encoding (AMLD) for a system with ordering $\mathscr{L}_1$ and $D_{010}D_{101} \leq D_{001}^2 \leq D_{010}D_{110}$

It is worth mentioning that although the successive refinement part of the scheme is well-known, producing the descriptions and their rate characterization is not an easy task without the AMLD result. As an example, consider a system with ordering $\mathscr{L}_1$ and assume $D_{010}D_{101} \leq D_{001}^2 \leq D_{010}D_{110}$. An achievable rate triple is

$$(R_1, R_2, R_3) = (\frac{1}{2}\log\frac{D_{010}D_{011}}{D_{100}D_{001}D_{111}}, \frac{1}{2}\log\frac{D_{101}}{D_{001}D_{011}}, \frac{1}{2}\log\frac{D_{001}}{D_{010}D_{101}}), \tag{4.30}$$

which corresponds to the corner point Y_{12} in regime II of the AMLD coding problem. The description encoding for this corner point is illustrated in Figure 4.3. Clearly, the coding scheme for this point matches that for Y_{12} closely, and the SR encoded information in the 3-rd, 4-th and 5-th layers needs to be strategically re-processed using linear codes. Without the underlining AMLD coding scheme, it appears difficult to devise this coding operation directly.

Part II

Multiple Unicast over Gaussian Network

Overview

A wireless network consists of several users in general, each of which has a demand to receive data at a certain rate from a source in the network, and the network has to provide service to all of them simultaneously.

Data transmission over a wireless network is quite challenging problem. This is due to two natural features of wireless communication: (i) broadcast: wireless users communicate over the air and signals from any of the transmitters are heard by multiple nodes in the proximity of the transmitter with possibly different signal strengths; (ii) superposition: a wireless node receives signals from multiple simultaneously transmitting nodes in its proximity, with the received signals all superimposed on top of each other.

Because of these effects, unlike the wired networks, links in a wireless network are never isolated, and complex signal interactions exist between the competing flows. More precisely, undesired signals caused by broadcasting from different nodes create signal interference for a node in their proximity.

Characterization of the capacity region of the network, the set of rates (speed of data transmission) at which data can be transmitted to the users, is an important question in order to design (approximately) optimal systems. This question has been unsolved even for one source and one receiver, when there exist more nodes (helper) in the network to facilitate communication. A recent study by Avestimehr *et. al* [17] provides an approximate solution for this problem when there is only flow of information in the network. That is when one source wishes to communicate to a single[1] receiver over a wireless network. One of the main contributions of this work is introducing the *deterministic model* for wireless network. This model gets rid of the stochastic behavior of the noise and focuses on the signal intercalation.

However, less progress has been made when different users in the network have different data demands. In this situation, called *multiple unicast scenario*, the problem is more challenging, since signals carry information about independent messages may get mixed over the wireless network, and become completely useless. The main question here is how to manage such interference in order to provide the data for the users.

[1] This can be extended to the case where the *same* information has to be sent to multiple receives [17].

The simplest case in this class of networks is the interference channel problem, which consists of two sources each wishes to communicate to its own receiver, and there is no relay (helper) node in the network. Even for this simple one-hop network, the information-theoretic characterization has been open for several decades. Etkin *et al.* in [15] provided an approximate capacity characterization for this problem. However, generalization of this result to a network with possible relay nodes who can facilitate communication lies in a wider research area.

We formulate the relay-interference network in which different information flows are sent through intermediate nodes in the network, *i.e.,* transmission is performed over multiple hops. In particular, we study two specific networks, called ZS and ZZ networks, with two sources/transmitters, two relays (intermediate nodes), and two receivers, where each of them is only interested in one of the source messages. While the optimal and exact solution to the simpler problem with no relay node is still unknown, our best hope is to provide approximate solution to this problem.

We start with analysis of the problem under the deterministic model in Chapter 5, where we derive an *exact* characterization for the set of achievable rates of both networks. Moreover, we introduce a new interference management technique, called *interference neutralization*, which is shown to be required in order to achieve the capacity of the network. These characterizations will be later translated for the noisy Gaussian wireless network in Chapter 6, where we present a 2 bit gap approximate characterization for the original problem. Our work establishes the first additive approximate result in this context. More importantly, the transmission strategies used for the deterministic networks have to be translated and adapted to the Gaussian networks. This provides guideline for system design and coding architecture, which is very important from an engineering point of view.

Finally, we explore the role of an adversarial jammer in a wireless network in Chapter 7. In this situation, a jammer wishes to avoid communication from transmitter to receiver. The jammer broadcasts interfering signal to the network in order to corrupt the signals carry information from the source. The question here is to identify the effect of such an adversarial node on the performance of the system, as well as exploring coding schemes which can be utilized by the nodes in the network to minimize such effect.

5 The Deterministic Relay-Interference Network

The multi-commodity flow problem, where multiple independent unicast sessions need to share network resources, can be solved efficiently over graphs using linear programming [1] techniques [64]. This is not the case for wireless networks, where the broadcast and superposition nature of the wireless medium introduces complex signal interactions between the competing flows. The simplest example is the one-hop interference channel [10], where two transmitters with independent messages are attempting to communicate with their respective receivers over the wireless transmission medium. Even for this simple one-hop network, the information-theoretic characterization has been open for several decades. To study more general networks, there is a clear need to understand and develop sophisticated interference management techniques.

The capacity of the wireless Gaussian interference channel has been (approximately) characterized, within one bit (see [15] and the references therein). Building on this progress, a natural next step is to study the approximate capacity region of small-scale *interference-relay networks*, where there are potentially multiple hops from the sources to destinations through cooperating relays. Studying even simple two-hop topologies could help develop techniques and build insight that would enable a (perhaps approximate) characterization of capacity for more general networks. We are interested in our work in a *universal* type of approximation, in that it should characterize the capacity to within a constant number of bits, independently of the signal-to-noise ratio and the channel parameter values.

The focus of this chapter and the next one is to study the two-stage relay-interference network illustrated in Figure 5.1. In particular, we give an approximate characterization of the capacity region for special cases of these networks

[1] However, if we require that the flows take integer values, it becomes NP-complete.

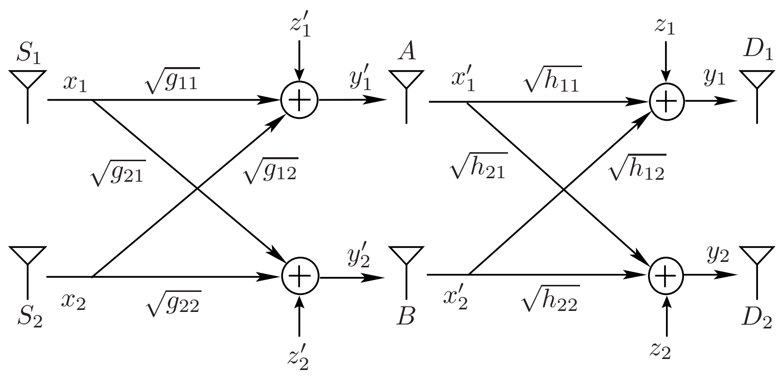


Figure 5.1: Two-stage relay-interference network.

when some of the cross-links are weak. These are illustrated in Figure 6.2 and Figure 6.3, which we refer to as the ZS and ZZ Gaussian models.

The *deterministic* approach, studied by Avestimehr, Diggavi, and Tse [17], simplifies the wireless network interaction model by eliminating the noise. This approach was successfully applied to the relay network in [65], and resulted in insight in terms of transmission techniques. These insights also led to an approximate characterization of the noisy wireless relay network problem [17].

Our approach to analyze these networks is to first apply the deterministic model to these Gaussian networks, and study these problems by using the linear deterministic model. We provide an exact capacity region characterization for this deterministic network in this chapter. This will be translated into a universally approximate characterization for the (noisy) Gaussian network.

In studying these special networks, we discover that many sophisticated techniques are required to (approximately) characterize the network capacity region. The main new ingredients that enable this characterization are as follows: (i) a new interference management technique we term *interference neutralization*, in which interference is canceled over the air, without the relays necessarily decoding the transmitted messages[2]; (ii) a *structured lattice code* that enables interference neutralization over Gaussian networks; (iii) A *network decomposition* technique which enables appropriate rate-splitting of the message and power allocation for the different message components; (iv) genie-aided outer bounding techniques that enable bounds that are tighter than the information-theoretic cut-set outer bounds.

[2] A noise nulling technique is proposed in [66] to mitigate correlated noise in an amplify-forward relaying strategy for a single unicast "diamond" parallel relay network. However, the difference in our technique is that we use the structure of the codebooks (without necessarily decoding information) to neutralize interference, and not noise statistics. Moreover the multiple-unicast nature of the problem necessitates strategic partitioning and rate-splitting of different components of the messages.

A way to interpret the achievability results for the ZS networks is that the relays perform a partial-decoding of strategically split messages from the sources, and then cooperate to deliver the required messages to the destination, again through strategically splitting the messages. The power allocated to each of the sub-messages is determined using the insight derived from the deterministic model, that messages that are not intended be decoded arrive at the noise-level. The achievability for the ZZ network is slightly more sophisticated in that one of the relays is required to only decode a function of the sub-messages. The function is chosen such that its signal in combination with the transmission of the other relay causes the unwanted interference to be canceled (neutralized) at the destination. This interference neutralization is enabled in the Gaussian channel using the group property of a structured lattice code.

Work in the literature over the past decade has examined scaling laws for multiple independent flows over wireless networks, see for example [18, 19, 67]. The goal there is to characterize the order of the wireless network capacity as the network size grows. In contrast, in our work, instead of seeking order arguments and scaling laws, we try to characterize the capacity (perhaps within a universal constant of a few bits) for specific topologies. The interference channel is a special case of such networks, where there is only one-hop communication between the sources and destinations. There has been a surge of recent work on this topic including cooperating destinations [68] and use of feedback in inducing cooperation at the transmitters [69]. The deterministic approach developed in [17] has been successfully applied to the interference channel in [70]. The fundamental role of interference alignment in K-user interference channel (still a one-hop network) has been demonstrated in [39, 40].

The chapter is organized as follows. Section 5.1 gives a broad overview of the deterministic modeling for wireless networks. Our notation and the basic network models we study in this chapter are introduced in Section 5.2. Section 5.3 illustrates the transmission techniques used for interference management over a wireless network through simple deterministic examples. The main results are given in Section 5.4 for the so called ZS and ZZ networks. The achievability and converse for the deterministic ZS network is given in Section 5.5. Similarly, Section 5.6 follows a similar program for the ZZ network, first identifying the capacity region for the deterministic version. This allows illustration of ideas such as interference neutralization, as well as genie-aided outer bounding techniques.

5.1 Review: A Deterministic Approach to Wireless Networks

In this section we first present the deterministic model introduced by Avestimehr, Diggavi, and Tse, and then briefly review the results obtained using this approach in [17].

5.1.1 The Deterministic Model

In a standard and well accepted model for the wireless signal interaction, transmitted signals get attenuated by (complex) gains to which independent (Gaussian) receiver noise is added [71]. More formally, the received signal y_i at node $i \in \mathcal{V}$ at time t is given by,

$$y_i(t) = \sum_{j \in \mathcal{N}_i} h_{ij} x_j(t) + z_i(t), \tag{5.1}$$

where h_{ij} is the complex channel gain between node j and i, x_j is the signal transmitted by node j, and $\mathcal{N}_i$ are the set of nodes that have non-zero channel gains to i. We assume that the average transmit power constraints for all nodes is 1 and the additive receiver Gaussian noise is of unit variance. We use the terminology *Gaussian wireless network* when the signal interaction model is governed by (5.1).

In [65], a simpler deterministic model which captures the essence of wireless interaction was developed. The advantage of this model is its simplicity, which gives insight to strategies for the noisy wireless network model in (5.1). We will utilize this model to develop techniques for the relay-interference network. Our main results are developed for this deterministic model. The deterministic model of [65] simplifies the wireless interaction model in (5.1) by eliminating the noise and discretizing the channel gains through a binary expansion of p bits. Therefore, the received signal Y_i which is a binary vector of size p is modeled as

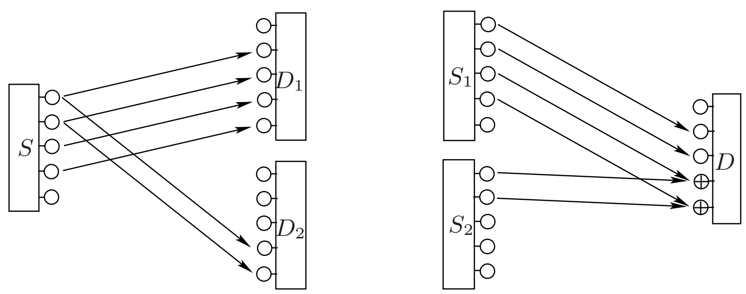
$$Y_i(t) = \sum_{j \in \mathcal{N}_i} N_{ij} X_j(t), \tag{5.2}$$

where N_{ij} is a $p \times p$ binary matrix representing the (discretized) channel transformation between nodes j and i and X_j is the (discretized) transmitted signal. All operations in (5.2) are done over the binary field, $\mathbb{F}_2$. We use the terminology *deterministic wireless network* when the signal interaction model is governed by (5.2). Shift matrix is a special matrix representation for a Gaussian fading channel. This matrix captures the attenuation effect of the signal caused by the channel gain by performing a shift on the binary representation of the input, x_j, and ignoring the bits below the average noise level. More precisely, this model assigns a matrix $\mathbf{J}^{p-n_{ij}}$ to the Gaussian gain h_{ij}, where

$$\mathbf{J} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}_{p \times p}, \tag{5.3}$$

is the shift matrix, and $n_{ij} = \lceil \frac{1}{2} \log |h_{ij}|^2 \rceil$, for real channel gains. Note that multiplying a vector X_j by the matrix $\mathbf{J}^{p-n_{ij}}$ is equivalent to shift down its elements $p - n_{ij}$ times, which results in a vector with $p - n_{ij}$ zeros on the top followed the top n_{ij} elements of X_j.

An illustration of this deterministic model is given in Figure 5.2 for the broadcast and multiple access networks. The transmitter and receiver of each node in the network is equipped with p so called *sub-nodes*, each represents a bit from the binary expansion of the corresponding signal. Figure 5.2(a) shows a deterministic model of the broadcast channel, where the channel from the transmitter to Receiver 1 is stronger than that to Receiver 2. This is represented by the deterministic model developed in [65] with 4 most significant bits (MSB) of the transmitted signal captured by D_1 and only 2 MSBs of the transmitted signal captured by D_2. The deterministic model of the multiple access channel shown in Figure 5.2(b) adds one more ingredient, which is how the bits from two transmitting nodes interact at a receiver. In Figure 5.2(b) the channel from S_1 to D is stronger than that of S_2. Therefore, the interaction is between the 2 MSBs of the message sent by S_2 with the lower 2 significant bits of the message sent by S_1, and the interaction is modeled with an addition over the binary field (*i.e.,* `xor`). This interaction captures the dynamic range of the signal interactions. It was shown in [17], that this model *approximately*[3] captures the wireless interaction model of (5.1) for the broadcast and multiple access channels. For general networks the deterministic model yields insights which, when translated to the noisy wireless network, lead one to develop cooperative strategies for the model in (5.1), which are (provably) approximately[4] optimal [17].



(a) Broadcast channel (b) Multiple access channel

Figure 5.2: The linear deterministic model for a Gaussian broadcast channel (BC) is shown in (a) and for a Gaussian multiple access channel (MAC) is shown in (b).

5.1.2 Single Unicast over the Relay Network

Capacity characterization of the relay network [9] is one of the longest outstanding open problems in network information theory. The capacity of such

[3]The approximation is in the sense that the capacity region of the deterministic model is within 1 bit of the capacity region of the Gaussian counterparts.

[4]It has been shown for single unicast there is an *approximate* max-flow, min-cut result where the difference is within a constant number of bits, which depends on the topology of the network, but not the values of the channel gains [17].

network is unknown even for the simplest network with Gaussian noise and only one relay node.

A natural information-theoretic cut-set bound [3] is known for any relay network which upper bounds the reliable transmission rate R. We need the following definition before presenting this bound.

Definition 5.1. *Consider a network over set of nodes $\mathcal{V}$. For $\mathcal{S} \subseteq \mathcal{V}$ and $\mathcal{D} \subseteq \mathcal{V}$ with $\mathcal{S} \cap \mathcal{D} = \varnothing$, define*[5] *$\Lambda(\mathcal{S};\mathcal{D})$ to be the set of all cuts (Ω, Ω^c) which separate $\mathcal{S}$ from $\mathcal{D}$, that is,*

$$\Lambda(\mathcal{S};\mathcal{D}) \triangleq \{\Omega \subseteq \mathcal{V} : \mathcal{S} \subseteq \Omega, \mathcal{D} \subseteq \Omega^c\}. \tag{5.4}$$

The main idea in the cut-set bound is to consider the amount of information can be conveyed through a cut in the network. This can be obtained by maximizing the mutual information between the received signal at the receiver side of the network, and the transmitting signals at the sender side. More precisely, it is shown [3] that the achievable rate of any relay network is upper bounded as

$$C \leq \max_{p(\{X_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda(S,D)} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}), \tag{5.5}$$

where $X_{\mathcal{U}}$ and $Y_{\mathcal{U}}$ respectively denote the collection of transmitted and received signals at nodes in the subset $\mathcal{U} \subseteq \mathcal{V}$.

This bound, however, is not tight in general. More importantly, it is difficult to evaluate the bound for arbitrary network due to the maximization over the *joint* probability of the transmuting signals. It has been shown in [17] that for the linear finite-field deterministic relay network, all the mutual information expressions appear in (5.5) are simultaneously optimized by independent and uniform distribution of $\{x_i\}_{i\in\mathcal{V}}$. Moreover, it is shown that this bound is achievable, which gives a complete characterization for the capacity of the linear finite-field deterministic relay network as in the following theorem.

Theorem 5.2 ([17] Theorem 4.3)**.** *The capacity of a linear finite-field deterministic relay network is given by*

$$\begin{aligned} C^{\mathsf{det}} &= \max_{\prod_{j\in\mathcal{V}} p(X_j)} \min_{\Omega\in\Lambda(S,D)} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \\ &= \max_{\prod_{j\in\mathcal{V}} p(X_j)} \min_{\Omega\in\Lambda(S,D)} H(Y_{\Omega^c} | X_{\Omega^c}) \\ &= \min_{\Omega\in\Lambda_D} \operatorname{rank}(G_{\Omega,\Omega^c}), \end{aligned} \tag{5.6}$$

where G_{Ω,Ω^c} denotes the transfer matrix associated with cut Ω.

A polynomial-time algorithm is proposed in [72] to perform the encoding strategy at the relays to achieve the min-cut value in binary linear deterministic networks, for the case of a unicast connection. However, one important

[5] *We may use $\Lambda(S,D)$ instead of $\Lambda(\{S\},\{D\})$ when $\mathcal{S}$ and/or $\mathcal{D}$ has only one element.*

message of this result is due to the its achievability argument, that is, a random mapping at each relay node can achieve the capacity of the network. The intuitions gained from the deterministic model were then applied to Gaussian signal interaction in a wireless network, which provides an approximate characterization of the capacity for Gaussian relay network.

Theorem 5.3 ([17] Theorem 4.5)**.** *Consider a Gaussian relay network, with cut-set upper bound*

$$\overline{C} = \max_{p(\{X_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda(S,D)} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}).$$

Any rate R not exceeding $\overline{C} - \kappa$ is achievable over this network. More precisely, the capacity of this network is bounded by

$$\overline{C} - \kappa \leq C^{\mathrm{g}} \leq \overline{C}, \tag{5.7}$$

where $\kappa = \mathcal{O}(|\mathcal{V}|)$ is a constant independent of the channel gains, and $|\mathcal{V}|$ is the number of nodes in the network.

5.1.3 Multiple Unicast over the Relay Network

The result mentioned above gives an approximate characterization of a wireless network where a single source transmits the message to a single receiver. It can be also generalized to multi-cast network, where there are multiple receivers in the network, and all of them are interested in decoding the same message sent by the transmitter. A random relaying strategy is shown to be (approximately) optimal for this scenario.

However, the problem is much more difficult when multiple flows of information exist in the network. That is, multiple messages are encoded at the transmitter(s) and sent over the network. various receivers in the network are interested in decoding different messages. In this situation, the relays receive a combination of the signals describing messages, and a random mapping from the received signal to transmit signal is not optimal anymore.

The simplest setting for a multiple unicast scenario is the Gaussian interference channel [10], where two transmitters wish to communicate their messages to their own receivers, over a shared channel. This problem has been studied using a deterministic approach by Bresler and Tse [73], where they show that the capacity of the real-valued 2-user Gaussian interference channel with arbitrary signal and interference to noise ratios is within 18.6 bits per user of the capacity of a deterministic interference channel with proper gains.

Although this work gives a similar result as in [15] (albeit with a larger gap), it provides a lot of insight to the structure of the various near-optimal schemes for the Gaussian interference channel in the different parameter ranges. The simplicity of the deterministic channel model gives us a better understanding of the near-optimality of the simple private/common message splitting (Han-Kobayashi scheme [10]) for the Gaussian interference channel.

The general form of the cut-set bound for a multi-terminal network with arbitrary number of transmitters and receivers is the following[6].

Theorem 5.4 ([3] Theorem 15.10.1)**.** *Let $\mathcal{N}$ be an arbitrary network over a set of nodes $\mathcal{V}$ with K pairs of source/destination where transmitter S_i wishes to communicate to receiver D_i at rate R_i, for $i = 1, \ldots, K$. Then, any rate tuple $(R_1, R_2, \ldots, R_K)$ satisfies*

$$\sum_{i \in \mathcal{A}} R_i \leq \max_{p(\{X_j\}_{j \in \mathcal{V}})} \min_{\Omega \in \Lambda(\mathcal{S}_\mathcal{A}; \mathcal{D}_\mathcal{A})} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}), \tag{5.8}$$

for all $\mathcal{A} \subseteq \mathcal{I}_K$, where $\mathcal{S}_\mathcal{A} = \{S_i : i \in \mathcal{A}\}$ and $\mathcal{D}_\mathcal{A} = \{D_i : i \in \mathcal{A}\}$.

It can be shown that this upper bound for the achievable rate tuples is not tight in general when there are more than one source/destination pair, even for linear finite-field deterministic network. We will use this bound in the rest of this chapter to upper bound the achievable rates. However, we will see that tighter bounds are needed in order to provide an exact characterization of the capacity of the DZS and DZZ networks.

5.2 Problem Formulation

Our goal in this chapter is to study the capacity region for a class of deterministic 2-user relay-interference networks shown in Figure 5.1, which we call the XX network. This understanding will be used in Chapter 6 to derive an approximate capacity region for the Gaussian relay-interference networks. We start by describing our notation we use for the nodes and the signals within the network.

Two transmitters, S_1 and S_2, encode their messages W_1 and W_2 of rates R_1 and R_2, respectively, and broadcast the obtained signals to the relay nodes, A and B. Denote the transmitted signals by X_1 and X_2, and the received signals at the relays by Y_1' and Y_2'. Then

$$\begin{aligned} Y_1'[t] &= M_{11} X_1[t] + M_{12} X_2[t] \\ Y_2'[t] &= M_{21} X_1[t] + M_{22} X_2[t]. \end{aligned} \tag{5.9}$$

where matrices $\{M_{ij}\}$ are linear shift deterministic matrices, and defined as $M_{ij} = \mathbf{J}^{p-m_{ij}}$. The matrix $\mathbf{J}$ is defined as in (5.3), while variables m_{ij} are non-negative integer number, describe the channel gains.

The relay nodes perform any (causal) processing on their received signal sequences $\{Y_1'[t]\}$ and $\{Y_2'[t]\}$ respectively, to obtain their transmitting signal sequences, $\{X_1'[t]\}$ and $\{X_2'[t]\}$. The received signals at the destination nodes can be written as

$$\begin{aligned} Y_1[t] &= N_{11} X_1'[t] + N_{12} X_2'[t] \\ Y_2[t] &= N_{21} X_1'[t] + N_{22} X_2'[t], \end{aligned} \tag{5.10}$$

[6]We have slightly modified the original expression of the theorem to adapt it with the notations used in this chapter.

where matrices $\{N_{ij}\}$ are defined similarly.

Each destination node D_i, $i = 1, 2$, is interested in decoding its message W_i, using its received signals $\{Y_i[t]\}$. We define a rate pair (R_1, R_2) to be admissible if there exist a transmission scheme under which D_1 and D_2 can decode W_1 and W_2, respectively, with arbitrary small (average) error probability in the standard manner [3]. This would allow two end-to-end reliable unicast sessions at rates (R_1, R_2) for the source/destination pairs (S_1, D_1) and (S_2, D_2).

It is worth mentioning that though this network looks like cascaded interference channels, there is an important difference between this network, and the standard interference channel. Unlike the interference channel, the messages sent by the relays at the second layer of transmission need not to be independent, *i.e.,* we can try to induce cooperation at the relays to transmit information to the final destinations. This distinction makes this network more interesting than a simple cascade of interference channels.

In this work, we focus on two specific realizations of the network, where at least two of the cross links have negligible gains which simplify the connectivity models of (5.9)-(5.10), and therefore can be eliminated from the network. If such weak cross links appear in the same layer, the resulting network would be an standard deterministic interference channel, cascaded by a parallel channel, and its capacity characterization is addressed in [73]. On the other hand, if there is one weak cross link in each layer, the resulting network would be either a ZS or a ZZ networks. We describe these two networks in the following, and give the capacity characterization of their admissible rate region in Section 5.4.

5.2.1 The ZS Network

The deterministic ZS (DZS) network is a special case of the interference-relay network defined in (5.9)-(5.10). In the DZS network one cross link in each layer has a negligible gain, and therefore does not cause interference. In particular, we assume $m_{21} = n_{12} = 0$. The resulting deterministic model for this network is given in Figure 5.3.

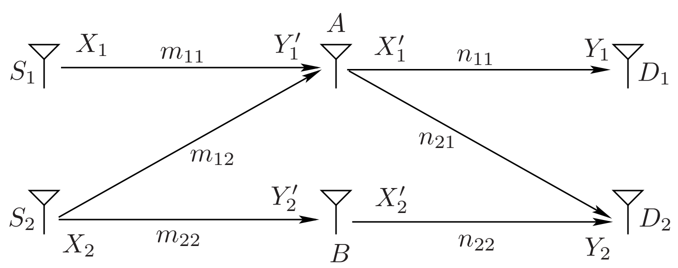


Figure 5.3: The deterministic ZS network.

5.2.2 The ZZ Network

The ZZ network is another special configuration interference-relay network, wherein one cross link in each layer has zero gain. However, the difference is that, here the missing links are in parallel. In particular, we assume $m_{21} = n_{21} = 0$. The deterministic ZZ networks is shown in Figure 5.2.2.

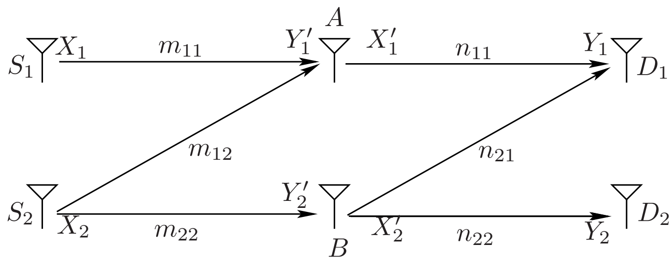


Figure 5.4: The deterministic ZZ network.

5.3 Examples Illustrating Transmission Techniques

In this section, we illustrate through examples some of the main interference management techniques we will use to achieve the capacity of our deterministic relay-interference networks. Although we demonstrate the ideas for deterministic networks, similar techniques will be used later for Gaussian channels as well.

Example 5.5 (Interference Separation)**.** *Consider the network shown in Figure 5.5. It is easy to see that the sum-rate of this network is upper bounded by*

$$R_1 + R_2 \leq 3, \tag{5.11}$$

since the cut which separates the destination nodes from the rest of the network has value equal to 3.

Assume we wish to transmit at rate pair $(R_1, R_2) = (1, 2)$ *from the source nodes to the destination nodes. Namely, source* S_1 *would like to send a message* W_1 *that consists of a single bit* $X_1(1)$ *(per unit time) to* D_1*, while source* S_2 *would like to send message* W_2 *which consists of two bits* $X_2(1)$ *and* $X_2(2)$ *to* D_2*.*

It is easy to see that this rate pair can only be achieved by using a transmission strategy which avoids interference. Since D_1 *receives only one bit from* A*, this bit should be the clear data about* W_1*. Hence,* A *should have received the message from* S_1*, without interference. Therefore, the message* W_2 *should be encoded such that it does not cause interference at* A*. More precisely, in order*

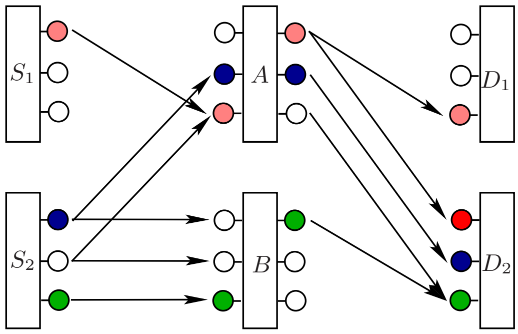


Figure 5.5: Interference separation: $(R_1, R_2) = (1, 2)$ is achievable.

to to communicate at this rate, the transmitters should encode their messages as

$$X_1 = \begin{bmatrix} X_1(1) \\ 0 \\ 0 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} X_2(1) \\ 0 \\ X_2(2) \end{bmatrix}. \tag{5.12}$$

The relay nodes also have to appropriate map their received bits to the signals they broadcast, to avoid interference, as

$$X_1' = \begin{bmatrix} X_1(1) \\ X_2(1) \\ 0 \end{bmatrix}, \qquad X_2' = \begin{bmatrix} X_2(2) \\ 0 \\ 0 \end{bmatrix}. \tag{5.13}$$

It is clear this scheme makes the interference separable from the signal at the nodes A and D_2. The decoding strategy at destination node D_2 is to first decode the interference $X_1(1)$, cancel it, and then retrieve its own message $X_2(1)$ and $X_2(2)$.

Example 5.6 (Interference Suppression)**.** *Depending on the parameters of the network, there are cases in which interference separation is not enough to achieve a transmission rate pair. That is, there does not exist a strategy that avoids or separates interference at the receivers. However, it might be possible to receive a clean copy of the interference, additionally to the copy that interferes with the desired signal. Therefore, one could use the clean copy to remove the interference. This is exactly the situation that occurs in the* ZZ *network shown in Figure 5.6, for the transmission rate pair $(R_1, R_2) = (3, 2)$.*

In this network, the signal observed at node A is interfered by the transmissions of source S_2. It is thus not possible for node A to completely decode W_1 using the 3 *bits it receives. However, the decoder D_1 can recover W_1, by first (partially) decoding W_2, and using this information to remove the interference from W_1. Note that here there are two interfering paths from S_2 to D_1, one through node A, and one through node B. The second path (through B), in fact*

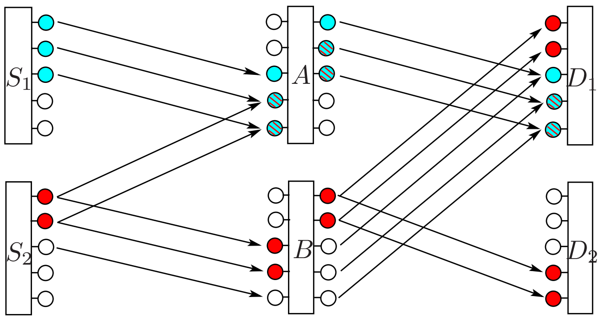


Figure 5.6: Interference suppression: $(R_1, R_2) = (3, 2)$ is achievable.

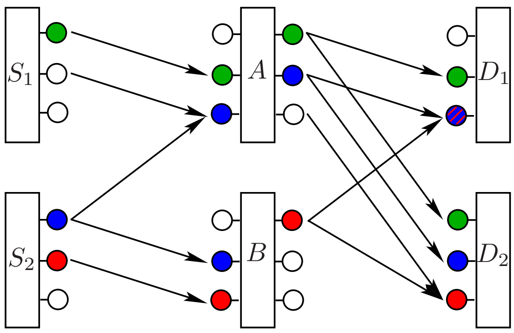


Figure 5.7: Interference alignment; $(R_1, R_2) = (1, 2)$ is achievable.

helps the decoder to remove the interference caused by the first path (through A). This is the only strategy that allows to achieve the rate pair $(R_1, R_2) = (3, 2)$.

Example 5.7 (Interference Alignment)**.** *Interference alignment is a transmission technique that essentially concentrates all the interference into a small dimension. The intuition behind this technique is the fact that whatever scheme one uses to deal with interference, it may occupy a certain number of degrees of freedom of the network. Therefore, one would expect that aligning the footprints of all interference, such that the occupied degrees of freedom coincide, will lead to minimizing the loss.*

Consider the network shown in Figure 5.7, and assume we wish to communicate at the rate pair $(R_1, R_2) = (1, 2)$. Since there is only one link from B to D_2, it is clear that the relay node A should help the S_2-D_2 communication by sending information bits about W_2. Therefore, the destination node D_1 receives two interfering signals (from A and B) which describe X_2. Hence, it would be able to resolve X_1 if and only if the occupied sub-node by these two interference *coincide. More precisely, by encoding the messages at the transmitters as*

$$X_1 = \begin{bmatrix} X_1(1) \\ 0 \\ 0 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} X_2(1) \\ X_2(2) \\ 0 \end{bmatrix}, \tag{5.14}$$

and using the transmission strategy shown in Figure 5.7, the received signal at the destination nodes would be

$$Y_1 = \begin{bmatrix} 0 \\ X_1(1) \\ X_2(1) + X_2(2) \end{bmatrix}, \qquad Y_2 = \begin{bmatrix} X_1(1) \\ X_2(1) \\ X_2(2) \end{bmatrix}. \tag{5.15}$$

Thus, the interfering bits $X_2(1)$ *and* $X_2(2)$ *are aligned at the destination node* D_1*, and occupy only one degree of freedom.*

Interference alignment has been introduced in a X *channel [39] by Maddah-Ali* et. al.*, wherein two transmitters attempt to communicate to two receivers, over an interference channel. Each transmitter has two messages, and each message has to be decoded at one receiver. It has been shown that in order to achieve the capacity of this network, it is necessary to align the two signals carrying information about the irrelevant interfering messages at the receivers. In [39], this technique has been used for the case wherein there are more than two messages have to be transmitted in the network. However, this strategy is also applied to the* K*-user interference channel where each receiver is only interested in decoding its own message, and all the other messages play as interference for it. Cadambe and Jafar [40] showed that a total degrees of freedom of* $K/2$ *can be achieved on this network using interference alignment. In the previous example, we showed through an example that interference alignment might be an essential strategy in a network even with two messages.*

Example 5.8 (Interference Neutralization)**.** *This technique can be used in networks which contain more than one disjoint path from* S_i *to* D_j *for* $i \neq j$*, where* D_j *is not interested in decoding the message sent by the source node* S_i*, and therefore it receives the interference through more than one link. The proposed technique is to tune these interfering signals such that they neutralize each other at the destination node. In words, the interfering signal should be received at the same power level and with different sign such that the effective interference, obtained by adding them, occupies a smaller number of degrees of freedom. To best of our knowledge, this technique is new and was not appeared in the literature before [20].*

Figure 5.8 shows a network in which interference neutralization is essential to achieve the desired rate pair $(R_1, R_2) = (2, 3)$*. Here* D_1 *has only two degrees of freedom, and receives information bits from both* A *and* B *over these sub-nodes. However, notice that there are two disjoint paths* (S_2, A, D_1) *and* (S_2, B, D_1)*, which connect* S_2 *to* D_1*. As it is shown in Figure 5.8, using a proper mapping (permutation) at the relay nodes, one can make the interference neutralized at the destination node* D_1*, and provide two non-interfered*

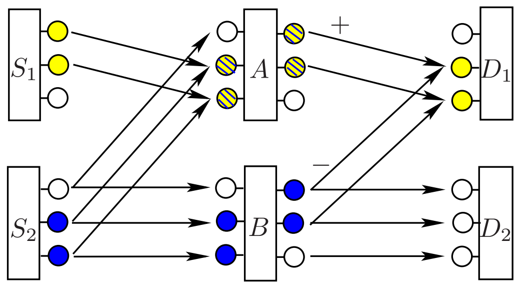


Figure 5.8: Interference neutralization; $(R_1, R_2) = (2, 2)$ is achievable.

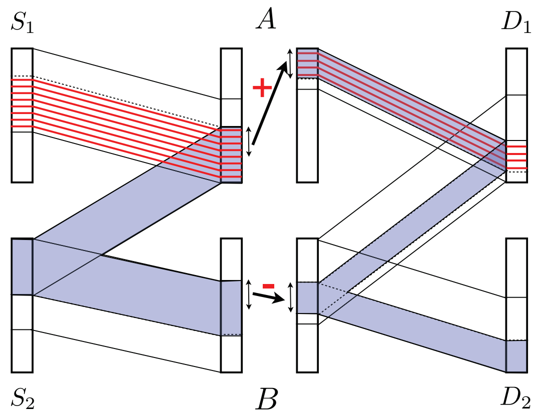


Figure 5.9: Interference neutralization.

links from S_1 to D_2. Note that this permutation does not effect the admissible rate of the other unicast from S_2 to D_2, the cost we pay, is to permute the received bits at D_2. A more general illustration of this phenomenon is given in Figure 5.9.

Example 5.9 (Network Decomposition for the ZS Network)**.** *A deterministic* ZS *network can be always decomposed into two sub node-disjoint networks, where the first partition consists of a set of sub-nodes of S_1, A and D_1, and looks like a line network. The second partition is, however, a diamond network, with a broadcast channel from S_2 to A and B in the first layer, and a multiple access channel from A and B to D_2 in the second layer. This diamond network can be used to send information from S_2 to D_2. Since these two networks are sub-node disjoint, there would be no interfering signal, and each of them can be analyzed separately. This is more illustrated in Figure 5.10.*

In a Gaussian ZS *network the network decomposition can be done using message splitting, superposition coding and proper power allocation. We will*

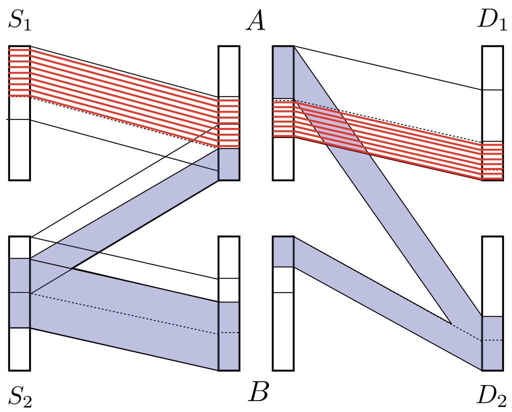


Figure 5.10: Network partitioning for a deterministic ZS network.

use this technique to achieve an approximate capacity for the Gaussian ZS *network.*

5.4 Main Results

In this section we present the main results of this paper, which is the exact capacity characterization of the deterministic ZS and ZZ interference-relay networks. We prove these results in Sections 5.5 and 5.6, respectively. The optimality of these rate regions are shown in the later sections. Both rate regions are polytopes, and therefore it suffices to show that the corner point rate pairs are achievable.

5.4.1 The ZS Network

The ZS network illustrated in Figure 5.3. Theorems 5.10 gives the exact characterizations of the capacity region of this network.

Theorem 5.10 (The capacity region of deterministic ZS network)**.** *The capacity region of the deterministic* ZS *network is specified by* $\mathcal{R}_{\mathsf{DZS}}$*, where* $\mathcal{R}_{\mathsf{DZS}}$ *is the set of all rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \le m_{11}, \tag{DZS-1}$$
$$R_2 \le \max(m_{12}, m_{22}), \tag{DZS-2}$$
$$R_1 + R_2 \le \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+, \tag{DZS-3}$$
$$R_2 \le m_{12} + n_{22}, \tag{DZS-4}$$
$$R_1 + R_2 \le m_{22} + \max(n_{11}, n_{21}), \tag{DZS-5}$$
$$R_1 + R_2 \le \max(m_{11}, m_{12}) + n_{22}, \tag{DZS-6}$$
$$R_1 \le n_{11}, \tag{DZS-7}$$
$$R_2 \le \max(n_{21}, n_{22}), \tag{DZS-8}$$
$$R_2 \le m_{22} + n_{21}, \tag{DZS-9}$$
$$R_1 + R_2 \le \max(n_{21}, n_{22}) + (n_{11} - n_{21})^+. \tag{DZS-10}$$

Here and elsewhere $(x)^+$ *denotes the positive part of* x*, which is formally defined by*

$$(x)^+ = \begin{cases} x & \text{if } x \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

The outer bound for the result above is fairly standard arguments based on reducing a multi-letter mutual information into single-letter forms by appropriately using decodability requirements at the different destinations. The details of these are given in Section 5.5.1 and Appendix D.1 respectively.

The coding strategy achieving this region is based on the idea of *network decomposition* illustrated in Section 5.3, Example 5.9 for the deterministic network. This will be discussed in more detail in Section 5.5.

5.4.2 The ZZ Network

The ZZ network illustrated in Figure 5.2.2. Although superficially the ZS and ZZ networks may look similar, the subtle difference in the network connectivity, makes the two problems completely different, both in terms of capacity characterization, as well as transmission schemes. It will be shown that a new interference management scheme, which we term as interference neutralization, is needed to achieve the capacity of this network. The most intuitive description for interference neutralization is to cancel interference over air without processing at the destinations. This scheme can be used whenever there are more than one path for interference to get received at a destination. We will explain it in more detail in Section 5.6.

Theorem 5.11 gives the exact characterizations for the capacity region of the deterministic ZZ network. Another new ingredient used here is needed a genie-aided outer bound that gives signal sent over the cross link of the first (or correspondingly second) layer to the destination (or correspondingly to the relay). This genie-aided bound presented in Section 5.6.1 allows us to develop outer bounds that are apparently tighter than the information-theoretic cut-set bounds by utilizing the decoding structure needed.

Theorem 5.11 (The capacity region of deterministic ZZ network)**.** *The capacity region of the deterministic* ZZ *network is given by $\mathcal{R}_{\mathsf{DZZ}}$, where $\mathcal{R}_{\mathsf{DZZ}}$ is the set of all rate pairs (R_1, R_2) which satisfy*

$$
\begin{aligned}
R_1 &\leq m_{11}, && \text{(DZZ-1)}\\
R_2 &\leq m_{22}, && \text{(DZZ-2)}\\
R_1 &\leq n_{11}, && \text{(DZZ-3)}\\
R_2 &\leq n_{22}, && \text{(DZZ-4)}\\
R_1 + r_2 &\leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ + n_{12}, && \text{(DZZ-5)}\\
R_1 + R_2 &\leq \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ + m_{12}. && \text{(DZZ-6)}
\end{aligned}
$$

5.5 The Deterministic ZS Network

In this section we prove Theorem 5.10. We study this problem in two parts. First we present the converse proof, which shows any achievable rate pair belongs to $\mathcal{R}_{\mathsf{DZS}}$. Then for any rate pair in this region, we propose an encoding scheme which is able to transmit messages up to the desired rates.

5.5.1 The Outer Bound

In this section we show that any achievable rate pair (R_1, R_2) for the deterministic ZS network belongs to $\mathcal{R}_{\mathsf{DZS}}$. Assume there exists a coding scheme with block length ℓ which can be used to communicate at rates R_1 and R_2 over the network. We use bold-face matrices to denote ℓ copy of them, as the transfer matrix applied over a codeword of length ℓ, *e.g.,* $\mathbf{M}_{11} = I_\ell \otimes M_{11}$.

All of the bounds in the theorem except (DZS-3) and (DZS-10) can be obtained in a straight-forward manner using the generalized cut-set bound in [74], which shows that in a linear finite-field network, the maximum reliable rate can be transmitted through a cut is upper bounded by the rank of the transition matrix of the cut. Here, we only present the proof of (DZS-3) and (DZS-10), which are more involved, and tighter than the cut-set bound. However, we present cut-set bound argument and the proof of the remaining bounds in Appendix D.1 for completeness.

$\triangleright$ (DZS-3) $R_1 + R_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+$
Recall the notation used in Figure 5.3 for transmitting and receiving signals.

In order to prove this bound, we can start with

$$\begin{aligned}
\ell(R_1+R_2) &\leq I(X_1^\ell, X_2^\ell; Y_1^\ell, Y_2^\ell) \\
&\leq I(X_1^\ell, X_2^\ell; Y_1^{'\ell}, Y_2^{'\ell}) \qquad (5.16)\\
&= I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) + I(X_1^\ell, X_2^\ell; Y_2^{'\ell}|Y_1^{'\ell}) \\
&= I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) + H(Y_2^{'\ell}|Y_1^{'\ell}) - H(Y_2^{'\ell}|X_1^\ell, X_2^\ell, Y_1^{'\ell}) \\
&= I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) + H(Y_2^{'\ell}|Y_1^{'\ell}), \qquad (5.17)
\end{aligned}$$

where in (5.16) we used the data-processing inequality for the Markov chain

$$(X_1^\ell, X_2^\ell) \leftrightarrow (Y_1^{'\ell}, Y_2^{'\ell}) \leftrightarrow (X_1^{'\ell}, X_2^{'\ell}) \leftrightarrow (Y_1^\ell, Y_2^\ell), \qquad (5.18)$$

and (5.17) holds since $Y_2^{'\ell}$ is a deterministic function of X_2^ℓ in the Z part of the network. Now, it is clear that

$$I(X_1^\ell, X_2^\ell; Y_1^{'\ell}) \leq \text{rank}\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{21} \end{bmatrix} = \ell \max(m_{11}, m_{12}). \qquad (5.19)$$

In order to bound the second term, we can write

$$\begin{aligned}
H(Y_2^{'\ell}|Y_1^{'\ell}) &= H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell) \qquad (5.20)\\
&\leq H(Y_2^{'\ell}, W_1|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell) \\
&= H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell, W_1) + H(W_1|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell) \\
&\leq H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell, W_1) + \ell\epsilon_\ell \qquad (5.21)\\
&= H(Y_2^{'\ell}|Y_1^{'\ell}, X_1^{'\ell}, Y_1^\ell, W_1, X_1^\ell) + \ell\epsilon_\ell \qquad (5.22)\\
&\leq H(Y_2^{'\ell}|Y_1^{'\ell} - \mathbf{M}_{11}X_1^\ell) + \ell\epsilon_\ell \\
&= H(\mathbf{M}_{22}X_2^\ell|\mathbf{M}_{12}X_2^\ell) + \ell\epsilon_\ell \\
&\leq \ell\text{rank}\begin{bmatrix} M_{12} \\ M_{22} \end{bmatrix} - \ell\text{rank}\,(M_{12}) + \ell\epsilon_\ell \\
&= \ell(m_{22} - m_{12})^+ + \ell\epsilon_\ell, \qquad (5.23)
\end{aligned}$$

where (5.20) holds since $X_1^{'\ell}$ is completely determined by $Y_1^{'\ell}$, and Y_1^ℓ is also a deterministic function of $X_1^{'\ell}$ in the S part of the network. We used Fano's inequality in (5.21), where W_1 should be decodable based on Y_1^ℓ. Finally, (5.22) holds, since X_1^ℓ is determined by W_1. Summing up (5.19) and (5.23), we get the desired bound.

Remark 5.12. *Note that the cut-set bound for the cut-set value for the cut $\Omega = \{S_1, S_2\}$ and $\Omega^c = \{A, B, D_1, D_2\}$ gives us*

$$\ell(R_1+R_2) \leq \text{rank}\,(\mathbf{G}_{\Omega,\Omega^c})\,,$$

where $\mathbf{G}_{\Omega,\Omega^c}$ *is the transfer matrix of the cut from the inputs* $X_\Omega = (X_1, X_2)$ *to outputs* $Y_{\Omega^c} = (Y'_1, Y'_2, Y_1, Y_2)$*, determined as*

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{0} & \mathbf{M}_{22} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

This gives

$$R_1 + R_2 \leq \max(m_{11} + m_{22}, m_{12}). \tag{5.24}$$

in which the RHS can be arbitrarily larger than the RHS of the presented bound in (DZS-3)*. The reason for this difference is the following. It is inherently assumed in deriving the cut-set bound that the receivers can cooperate to decode the messages of rates* R_1 *and* R_2*, and no decodability requirement is posed for individual receivers. However, the setup of this problem impose an extra constraint, that is* B *alone should be able to decode* W_2*. Incorporating this decodability requirement shrinks the set of admissible rates, and gives us a tighter bound that that predicted by the cut-set bound.*

▷ (DZS-10) $R_1 + R_2 \leq \max(n_{21}, n_{11}) + (n_{11} - n_{21})^+$

The last inequality captures the maximum flow of information from the relays to the destinations, such that D_1 and D_2 be able to decode W_1 and W_2, respectively. We again start with

$$\ell(R_1 + R_2) \leq I(X_1^\ell, X_2^\ell; Y_1^\ell, Y_2^\ell) = H(Y_1^\ell, Y_2^\ell) = H(Y_2^\ell) + H(Y_1^\ell|Y_2^\ell). \tag{5.25}$$

The first term can be easily bounded by

$$H(Y_2^\ell) \leq \text{rank}\begin{bmatrix} \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix} = \ell \max(n_{21}, n_{22}). \tag{5.26}$$

In order to bound the second term, we use the fact that W_2 can be decoded from Y_2^ℓ. Therefore,

$$\begin{aligned} H(Y_1^\ell|Y_2^\ell) &\leq H(Y_1^\ell, W_2|Y_2^\ell) \\ &= H(Y_1^\ell|Y_2^\ell, W_2) + H(W_2|Y_2^\ell) \\ &\leq H(Y_1^\ell|Y_2^\ell, W_2) + \ell\epsilon_\ell && (5.27) \\ &= H(Y_1^\ell|Y_2^\ell, W_2, X_2^\ell, {Y'_2}^\ell, {X'_2}^\ell) + \ell\epsilon_\ell \\ &\leq H(Y_1^\ell|Y_2^\ell - \mathbf{N}_{22}{X'_2}^\ell) + \ell\epsilon_\ell \\ &= H(\mathbf{N}_{11}{X'_1}^\ell|\mathbf{N}_{21}{X'_1}^\ell) + \ell\epsilon_\ell \\ &\leq \ell\text{rank}\begin{bmatrix} N_{11} \\ N_{21} \end{bmatrix} - \ell\text{rank}\,(N_{21}) + \ell\epsilon_\ell \\ &= \ell(n_{11} - n_{21})^+ + \ell\epsilon_\ell. && (5.28) \end{aligned}$$

In (5.27) we used the Fano's inequality, as well as the fact that X_2^ℓ, $Y_2'^\ell$, and $X_2'^\ell$ are known having W_2. The bound is obtained by replacing (5.26) and (5.28) in (5.25).

Remark 5.13. *It is worth mentioning that this bound is tighter than the cut-set bound for the cut* $\Omega = \{S_1, S_2, A, B\}$ *and* $\Omega^c = \{D_1, D_2\}$*, which is*

$$R_1 + R_2 \leq \max(n_{11} + n_{22}, n_{12}). \tag{5.29}$$

5.5.2 The Achievability Part

Network Decomposition: The achievability scheme presented here is based on decomposition of the deterministic ZS network into two node-disjoint networks. In fact, such partitioning depends on the demanded rate pair $(R_1, R_2) \in \mathcal{R}_{\mathsf{DZS}}$. The resulting family of separations immediately suggests a simple coding scheme. We will show that this separation is optimal, and does not cause any loss in the admissible rate region of the network.

Before introducing the network decomposition, we define an equivalence class for the sub-nodes (levels) in a network.

Definition 5.14. *In a* Z *(or* S*) deterministic network, two sub-nodes a and b are called* related *sub-nodes, and denoted by $a \sim b$ if any of the following conditions hold:*

- *$a = b$;*
- *a is connected to b;*
- *b is connected to a;*
- *there exists a sub-node c such that c broadcasts to both a and b;*
- *there exists a sub-node d where both a and b are connected to.*

Note that this relation is reflective, symmetric, and transitive. Therefore, it forms equivalence classes for the sub-nodes.

We denote by $\mathcal{N}_1$ and $\mathcal{N}_2$ the partitions of the network. Assume we want to transmit at rate $R_1 \triangleq r \leq \min(m_{11}, n_{11})$ from S_1 to D_1. The first part of the network $\mathcal{N}_1$, includes the top $(m_{11} - m_{12})^+$ levels as well as the lowest $(r - (m_{11} - m_{12})^+)^+$ levels of S_1. It also includes all the related sub-nodes of S_2, and the receiver levels of A and B. Similarly, in the second layer of the network, $\mathcal{N}_1$ includes the lowest $(n_{11} - n_{21})^+$ levels as well as the top $(r - (n_{11} - n_{21})^+)^+$ nodes of the transmitter part of A. All related sub-nodes of the transmitter part of B, as well as D_1 and D_2 also belong to $\mathcal{N}_1$. The second part of the network $\mathcal{N}_2$, is formed by all the remaining nodes. An example of this network decomposition is depicted in Figure 5.11.

We will use $\mathcal{N}_1$ for transmitting data from S_1 to D_1. Similarly $\mathcal{N}_2$ is only used to communicate from S_2 to D_2. Therefore, we have two unicast networks,

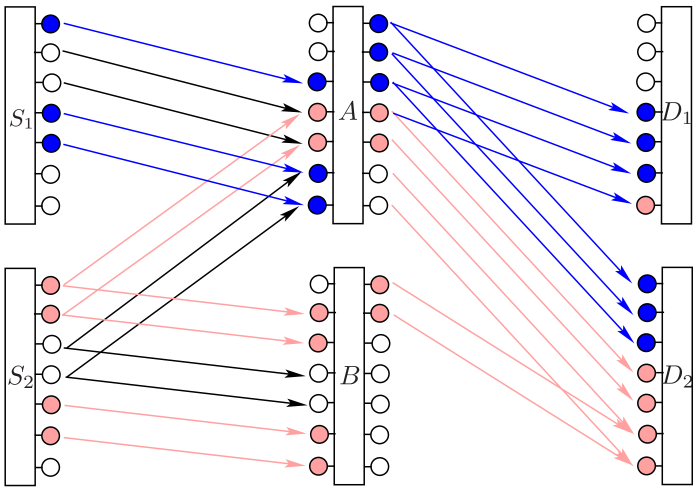


Figure 5.11: Network decomposition for a DZS network for $(R_1, R_2) = (3, 4)$: The blue part of the network is used by S_1/D_1 to communicate W_1 and the pink part is remained for transmitting W_2.

and each pair of transmitter-receiver can communicate up to the capacity of their own partition, which is the min-cut of the partition [65].

It is worth mentioning that any two "related" sub-nodes belong to the same partition. Therefore, these two networks are node-disjoint, and do not cause interference for each other. This allows us to derive the capacity of each network separately, and argue that (R_1, R_2) can be achieved simultaneously for the original network, if R_1 and R_2 are achievable for partitions $\mathcal{N}_1$ and $\mathcal{N}_2$.

Encoding Scheme: A transmission from S_1 and S_2 to D_1 and D_2 is performed as follows. S_1 transmits only on its sub-nodes which belong to $\mathcal{N}_1$, and keeps its other sub-nodes silent. Similarly, S_2 encodes its message on the sub-nodes included in $\mathcal{N}_2$, and sends zero on the other levels. Therefore, the *effective* communication over each partition is a simple uni-cast.

Figure 5.12 shows the effective parts of the network. It is easy to see that the diamond network in Figure 5.12(b) is also a linear shift deterministic networks, with channel gains

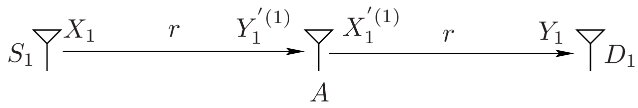
$$m'_{12}(r) = \min(\max(m_{11}, m_{12}) - r, m_{12}), \tag{5.30}$$

$$m'_{22}(r) = \min(\max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ - r, m_{22}), \tag{5.31}$$

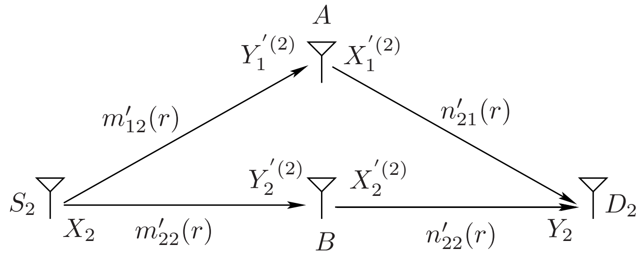
$$n'_{21}(r) = \min(\max(n_{11}, n_{21}) - r, n_{21}), \tag{5.32}$$

$$n'_{22}(r) = \min(\max(n_{11}, n_{21}) + (n_{22} - n_{21})^+ - r, n_{22}), \tag{5.33}$$

where r is the desired rate of communication for S_1/D_1.



(a) Effective channel for (S_1, D_1).



(b) Effective channel for (S_2, D_2).

Figure 5.12: The effective separated ZS network.

Achievable Rate Region: The cut values of the network $\mathcal{N}_1$ can be easily computed as

$$(m_{11} - m_{12})^+ + (r - (m_{11} - m_{12})^+)^+ = \max\{(m_{11} - m_{12})^+, r\} \geq r$$

for $\Omega = \{S_1\}$, and

$$(r - (n_{11} - n_{12})^+)^+ + (n_{11} - n_{12})^+ = \max\{(n_{11} - n_{12})^+, r\} \geq r.$$

for $\Omega = \{S_1, A\}$. Therefore any rate in $\mathcal{R}_{\mathsf{DZS},1}(r) = \{R_1 : R_1 \leq r\}$ can be conveyed from S_1 to D_1 through $\mathcal{N}_1$.

The capacity of $\mathcal{N}_2$ can be found using the generalized max-flow min-cut theorem [65]. Hence, the rate region of the second partition $\mathcal{N}_2$ would be

$$\begin{aligned}\mathcal{R}_{\mathsf{DZS},2}(r) = \{R_2 :& R_2 \leq \max(m'_{12}(r), m'_{22}(r)),\\ & R_2 \leq m'_{22}(r) + n'_{21}(r),\\ & R_2 \leq m'_{12}(r) + n'_{22}(r),\\ & R_2 \leq \max(n'_{21}(r), n'_{22}(r))\}.\end{aligned}$$

Therefore, by using this decomposition, any rate pair in the set $\overline{\mathcal{R}}_{\mathsf{DZS}}(r) \triangleq \mathcal{R}_{\mathsf{DZS},1}(r) \times \mathcal{R}_{\mathsf{DZS},2}(r) = \{(R_1, R_2) : R_1 \in \mathcal{R}_{\mathsf{DZS1}}(r), R_2 \in \mathcal{R}_{\mathsf{DZS},2}(r)\}$ can be achieved. It remains to prove the following lemma.

Lemma 5.15. *For any deterministic* ZS *network,*

$$\mathcal{R}_{\mathsf{DZS}} \subseteq \bigcup_{r \leq \min(m_{11}, n_{11})} \overline{\mathcal{R}}_{\mathsf{DZS}}(r). \tag{5.34}$$

We will prove this lemma in Appendix D.2.

5.6 The Deterministic ZZ Network

In this section we prove Theorem 5.11. This is done in two parts, that provide the converse and achievability proofs.

5.6.1 The Outer Bound

In the following we will show that any achievable rate pair (R_1, R_2) satisfies constraints (DZZ-1)-(DZZ-6). The individual rate bounds can be directly obtained by the generalized cut-set bound introduced in [74], where the maximum flow of information through a cut in a linear deterministic network is upper bounded by the rank of the transition matrix from the sender part of the cut to its receiver part. We only present the proof of these bounds in Appendix D.3 for completeness.

Unlike the individual rate constraints, the sum-rate bounds in (DZZ-5) and (DZZ-6) are genie-aided bounds which are tighter that the cut-set bounds. In the following, we focus on these two bounds, and present their proofs in detail. Again we assume that there exists a coding scheme with block length ℓ which can be used to communicate at rates R_1 and R_2 over the network.

▷ (DZZ-5) $R_1 + R_2 \leq \max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ + n_{12}$

In order to prove this inequality we focus on the flow of information from the sources to the relays. The key idea here is to provide A with the information sent by B to D_1 as side information. In such condition, the information A has received about W_1 is stronger than the information available at D_1, and therefore A can decode W_1 since D_1 can as well. Once W_1 is decoded at A, it can determine the transmitted codeword from S_1. By removing the interference from S_1, A can also partially decode W_2.

More precisely, we can write

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(X_1^\ell, X_2^\ell; Y_1^{'\ell}, Y_2^{'\ell}) = H(Y_1^{'\ell}, Y_2^{'\ell}) \leq H(Y_1^{'\ell}, Y_2^{'\ell}, \Gamma_2^\ell) \\
&= H(Y_1^{'\ell}, \Gamma_2^\ell) + H(Y_2^{'\ell} | Y_1^{'\ell}, \Gamma_2^\ell) \\
&\leq H(Y_1^{'\ell}) + H(\Gamma_2^\ell) + H(Y_2^{'\ell} | Y_1^{'\ell}, \Gamma_2^\ell), \qquad (5.35)
\end{aligned}$$

where $\Gamma_2^\ell = \mathbf{N}_{12} X_2^{'\ell}$ is the part of the signal received at D_2 from B as in Figure 5.2.2. The first two terms are easily bounded by $\ell \max(m_{11}, m_{12})$ and ℓn_{12}, respectively. Deriving an upper bound for the last term is more involved.

Similar to Γ_2^ℓ, we define $\Gamma_1^\ell = \mathbf{M}_{12} X_2^\ell$, where we have

$$\begin{aligned} H(\Gamma_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) &= H(Y_1^{'\ell} - \mathbf{M}_{11} X_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) &\quad (5.36)\\ &\le H(X_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) \\ &\le H(W_1 | Y_1^{'\ell}, \Gamma_2^\ell) \\ &= H(W_1 | Y_1^{'\ell}, X_1^{'\ell}, \Gamma_2^\ell) \\ &\le H(W_1 | \mathbf{N}_{11} X_1^{'\ell} + \Gamma_2^\ell) \\ &= H(W_1 | Y_1^\ell) \le \ell\epsilon_\ell, &\quad (5.37) \end{aligned}$$

where $\epsilon_\ell \to 0$ as ℓ grows. We have used the invertibility property of the deterministic multiple access channel in (5.36), and (5.37) follows from the Fano's inequality, and the fact that D_1 can decode the message sent by S_1. Therefore, we have $H(\Gamma_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) \le \ell\epsilon_\ell$. Hence,

$$\begin{aligned} H(Y_2^{'\ell} | Y_1^{'\ell}, \Gamma_2^\ell) &\le H(Y_2^{'\ell}, \Gamma_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) \\ &= H(Y_2^{'\ell} | \Gamma_1^\ell, Y_1^{'\ell}, \Gamma_2) + H(\Gamma_1^\ell | Y_1^{'\ell}, \Gamma_2^\ell) \\ &\le H(Y_2^{'\ell} | \Gamma_1^\ell) + \ell\epsilon_\ell \\ &= H(\mathbf{M}_{22} X_2^\ell | \mathbf{M}_{12} X_2^\ell) + \ell\epsilon_\ell \\ &\le \ell(m_{22} - m_{12})^+ + \ell\epsilon_\ell. &\quad (5.38) \end{aligned}$$

Replacing the upper bounds for each term in (5.35), we get

$$R_1 + R_2 \le \max(m_{11}, m_{12}) + n_{12} + (m_{22} - m_{12})^+. \tag{5.39}$$

Remark 5.16. *It is worth mentioning that the cut-set bound for $\Omega_s = \{S_1, S_2\}$ and $\Omega_d = \{A, B, D_1, D_2\}$ gives us*

$$\ell(R_1 + R_2) \le \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}),$$

where $\mathbf{G}_{\Omega,\Omega^c}$ is the transfer matrix of the cut from the inputs $X_\Omega = (X_1, X_2)$ to $Y_{\Omega^c} = (Y_1', Y_2', Y_1, Y_2)$, given by

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{0} & \mathbf{M}_{22} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

which yields in

$$R_1 + R_2 \le \max(m_{11} + m_{22}, m_{12}). \tag{5.40}$$

Note that depending on the channel gains, (5.40) *can be looser than the genie-aided bound in* (DZZ-5)*. This is due to the different decodability requirements assumed in developing two bounds: the cut-set bound is derived based on the* *assumption that the nodes at Ω^c can cooperate to decode messages (W_1, W_2), while the genie-aided bound is more restricted, by posing an extra requirement that the relay node B should be able to decode W_2 based on its own received signal. Moreover, the maximum interference can be tolerated at A is upper bounded by n_{12}, the capacity of the cross link in the second layer.*

▷ (DZZ-6) $R_1 + R_2 \leq \max(n_{11}, n_{12}) + (n_{22} - n_{12})^+ + m_{12}$

The last inequality captures the maximum flow of information from the relays to the destinations. Intuitively, this inequality says that the number of interfering bits can get neutralized at D_1 cannot exceed the minimum of m_{12} and n_{12}. In order to make this intuition formal, we provide Γ_1^ℓ, the partial information about W_2 which is available at A, as side information for D_1. We then have

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(Y_1^\ell, Y_2^\ell; X_1^\ell, X_2^\ell) = H(Y_1^\ell, Y_2^\ell) \\
&\leq H(Y_1^\ell, Y_2^\ell, \Gamma_1^\ell) \\
&\leq H(Y_1^\ell) + H(\Gamma_1^\ell) + H(Y_2^\ell | Y_1^\ell, \Gamma_1^\ell).
\end{aligned}$$

Again, we can simply upper bound the first two terms by the rank of the corresponding matrices. In order to bound the last term, similar to the proof of (DZZ-5), we use the following bounding technique.

$$\begin{aligned}
H(\Gamma_2^\ell | Y_1^\ell, \Gamma_1^\ell) &= H(Y_1^\ell - \mathbf{N}_{11} X_1^{'\ell} | Y_1^\ell, \Gamma_1^\ell) \\
&\leq H(X_1^{'\ell} | Y_1^\ell, \Gamma_1^\ell) \\
&\leq H(Y_1^{'\ell} | Y_1^\ell, \Gamma_1^\ell) \\
&= H(\mathbf{M}_{11} X_1^\ell + \Gamma_1^\ell | Y_1^\ell, \Gamma_1^\ell) \\
&\leq H(X_1^\ell | Y_1^\ell, \Gamma_1^\ell) \\
&\leq H(X_1^\ell | Y_1^\ell) \\
&\leq H(W_1 | Y_1^\ell) \leq \ell \epsilon_\ell, \qquad (5.41)
\end{aligned}$$

where (5.41) follows from the Fano's inequality. This inequality can be used as

$$\begin{aligned}
H(Y_2^\ell | Y_1^\ell, \Gamma_1^\ell) &\leq H(Y_2^\ell, \Gamma_2^\ell | Y_1^\ell, \Gamma_1^\ell) \\
&= H(Y_2^\ell | \Gamma_2^\ell, Y_1^\ell, \Gamma_1^\ell) + H(\Gamma_2^\ell | Y_1^\ell, \Gamma_1^\ell) \\
&\leq H(Y_2^\ell | \Gamma_2^\ell) + \ell \epsilon_\ell \\
&= H(\mathbf{N}_{22} X_2^{'\ell} | \mathbf{N}_{12} X_2^{'\ell}) + \ell \epsilon_\ell \\
&\leq \ell (n_{22} - n_{12})^+ + \ell \epsilon_\ell. \qquad (5.42)
\end{aligned}$$

Therefore, we have

$$R_1 + R_2 \leq \max(n_{11}, n_{12}) + m_{12} + (n_{22} - n_{12})^+. \qquad (5.43)$$

Remark 5.17. *The cut-set bound for the cut* $\Omega_s = \{S_1, S_2, A, B\}$ *and* $\Omega_d = \{D_1, D_2\}$ *is given by*

$$R_1 + R_2 \leq \max(n_{11} + n_{22}, n_{12}). \tag{5.44}$$

Again, it can be shown that the genie-aided bound can be arbitrarily tighter than the cut-set bound for different channel gains.

These together with the proofs of the individual rate constraints presented in Appendix D.3 complete the proof of the converse part of Theorem 5.11.

5.6.2 The Achievability Proof

In this part we will show that all rate pairs satisfying inequalities (DZZ-1)-(DZZ-6) are achievable. In particular, we introduce a coding scheme which achieves such rates. Our coding strategy provides the interference neutralization at the destination. This is performed by splitting the messages into two parts, namely *private* and *functional* parts. The private sub-messages can be decoded at the relays, and forwarded to the destinations. The functional sub-message of the second source can be also decoded at B. However, A only receives a combination (`xor`) of the functional sub-messages, and cannot decode them. It only forwards such combination on proper (power) levels such that the interference caused by the functional sub-message of S_2 get neutralized over the second layer of the network, and D_1 can decode the sub-message of its interest.

Our analysis is based on characterizing the number of *pure* and *combined* bits can be sent through each layer of the network. In the following we focus on one layer of the network, and obtain an achievable rate region for these numbers. Next, we use this region to build the encoding scheme for the ZZ network, and obtain an achievable rate region, which matches with the outer bound.

Definition 5.18. *Consider a deterministic* Z *network with gains* (n_{11}, n_{12}, n_{22}). *as shown in Figure 5.13. Each of the transmitters has a set of information bits to transmit to the receivers. This set for* F_i *includes* Υ_i *private bits and* Υ_0 *functional bits, namely,* $\mathcal{W}_{i,P} = \{W_{i,P}(1), \dots, W_{i,P}(\Upsilon_i)\}$ *and* $\mathcal{W}_{i,N} = \{W_{i,N}(1), \dots, X_{i,N}(\Upsilon_0)\}$. *The second receiver wishes to receive all the private and functional bits of* F_2*, while the first receiver is interested in receiving the private bits of* F_1*, and the* `xor` *of the functional bits of* F_1 *and* F_2*. More precisely, denoting by* $\hat{\mathcal{W}}_i$ *the set of bits* G_i *is interested in, we have*

$$\hat{\mathcal{W}}_1 = \mathcal{W}_{1,P} \cup \{\tilde{W}_{1,N}(j) \triangleq W_{1,N}(j) \oplus W_{2,N}(j) : j = 1, \dots, \Upsilon_0\},$$
$$\hat{\mathcal{W}}_2 = \mathcal{W}_{2,P} \cup \mathcal{W}_{2,N}.$$

We term this network with the described decoding demands as deterministic Z*-neutralization network.*

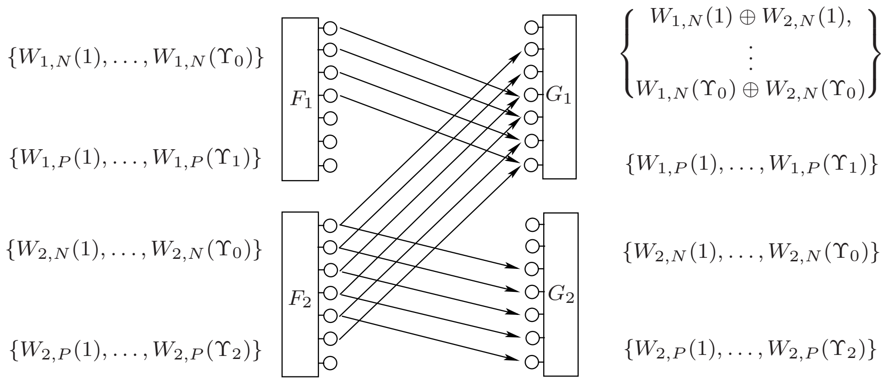


Figure 5.13: A deterministic Z-neutralization network with the message demands.

Our next goal is to characterize the set of achievable tuples $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ for the deterministic Z-neutralization network. The following lemma gives an achievable rate region for this network The proof of this lemma can be found in Appendix D.4.

Lemma 5.19. *Consider the deterministic* Z*-neutralization network defined in Definition 5.18 with channel gains* (n_{11}, n_{12}, n_{22}) *(see Figure 5.13). Any rate tuple* $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ *satisfying*

$$\Upsilon_0 \leq \lambda \triangleq \min\{n_{11}, n_{12}, n_{22}\}, \tag{5.45}$$

$$\Upsilon_0 + \Upsilon_1 \leq n_{11}, \tag{5.46}$$

$$\Upsilon_0 + \Upsilon_2 \leq n_{22}, \tag{5.47}$$

$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \mu \triangleq \max\{n_{11}, n_{12}, n_{22}, n_{11} + n_{22} - n_{12}\}. \tag{5.48}$$

is achievable for this network.

Now, having an achievable rate region for the deterministic Z-neutralization network, we are ready to present the coding scheme and analyze its rate region for the ZZ network.

Recall that the ZZ network consists of two cascaded Z networks. In the first layer, the source nodes split their message into private and functional parts. They can send these parts to the relays as long as their rates belong to the achievable rate region of the first layer given in Lemma 5.19. Once the relays receive these sub-messages, forward them to the destination nodes using the same scheme for the private and functional sub-messages. This can be done if the rate tuple for the sub-messages satisfy the corresponding inequalities for the second layer as well. Note that functional bits received at the destination are $\tilde{W}_{1,N}(j) \oplus W_{2,N}(j) = [W_{1,N}(j) \oplus W_{2,N}(j)] \oplus W_{2,N}(j) = W_{1,N}(j)$. Therefore, the interference of these bits get neutralized, and pure information bits will be received at the destination.

The achievable rate region of this scheme is given by

$$\begin{aligned}\overline{\mathcal{R}}_{\mathsf{DZZ}} = \Big\{(R_1,R_2) :& \exists \Upsilon_0, \Upsilon_1, \Upsilon_2 \geq 0,\\ & R_1 = \Upsilon_0 + \Upsilon_1,\\ & R_2 = \Upsilon_0 + \Upsilon_2,\\ & \Upsilon_0 \leq \min\{\lambda_m, \lambda_n\},\\ & \Upsilon_0 + \Upsilon_1 \leq \min\{m_{11}, n_{11}\},\\ & \Upsilon_0 + \Upsilon_2 \leq \min\{m_{22}, n_{22}\},\\ & \Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \min\{\mu_m, \mu_n\}\Big\} \end{aligned} \tag{5.49}$$

Here we used subscripts m and n to denote λ and μ parameters of the first and the second layer of the network, respectively. Applying Fourier-Motzkin elimination on this set to project it on the (R_1, R_2) plane, gives us the rate region claimed in the theorem.

5.7 The Deterministic Z-chain Network

In this section, we generalize the deterministic ZZ network, and consider a chain with arbitrary number of Z channels called *deterministic* Z*-chain network*, and characterize the region of the admissible rates under the deterministic model.

In order to obtain an outer bound, we first develop a genie-aided outer bound for the rate-region. We then show that this rate region is achievable using linear operations, where decoding a message of rate R is possible if and only if R non-interfered linearly independent equations describing the message are available at the receiver.

We *implicitly* show that the rate region obtained by the outer bound is achievable. That is, instead of proposing an explicit encoding scheme for the source and relay nodes, we show that there *exists* at least one scheme to achieve the capacity of the Z-chain network. More precisely, in our achievability proof, we use a new technique, called *analysis of pure and mixed equations*, where we keep track of the number of the equations involving bits of each of the interfering messages at the relay nodes in the different layers of the network. We show that among all possible encoding schemes at the relay nodes, there exists at least one which guarantee to provide the desired number of pure equations, to be able to decode the source message at the appropriate destination.

In the following, we first give a formal definition for the Z-chain network, the problem statement, and present the main result of this section. We then discuss the outer bound and the achievability analysis.

Consider the network shown in Figure 5.14, which is formed by cascading N consecutive Z channels. The transmitters S_1 and S_2 wish to communicate at rates R_1 and R_2 to the destination nodes D_1 and D_2, respectively.

The network is formed by N layers, where the k-th one connects the relay nodes A_{k-1} and B_{k-1} to A_k and B_k, for $k = 0, 2, \ldots, N$, where the relays in

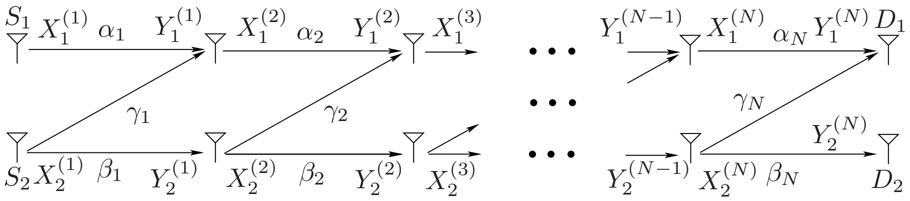


Figure 5.14: Transmission model: the Z-chain network.

layers zero and N are the source (transmitter) and receiver nodes, respectively. We use slightly different notation to denote the gains of the channel as follows. The k-th layer (hop) of the network is parametrized by the triple $\{(\alpha_k, \beta_k, \gamma_k)\}$, where α_k and β_k are the gains of the first and second direct links of the k-th layer of the network and γ_k is the gain of the cross link of the same layer.

We denote the inputs of the layer k by $X_{k,1}$ and $X_{k,2}$ and received vectors of this layer by $Y_{k,1}$ and $Y_{k,2}$. Transmission model in the k-th layer of the Z-chain network can be written as

$$\begin{aligned} Y_{k,1} &= N_{11}^{(k)} X_{k,1} + N_{12}^{(k)} X_{k,2}, \\ Y_{k,2} &= N_{22}^{(k)} X_{k,2}, \end{aligned} \tag{5.50}$$

where $N_{11}^{(k)} = \mathbf{J}^{p-\alpha_k}$, $N_{22}^{(k)} = \mathbf{J}^{p-\beta_k}$, and $N_{12}^{(k)} = \mathbf{J}^{p-\gamma_k}$.

The relay node A_k forms its encoded message for the next layer, $X_{k+1,1}$ as a function of its received signal $Y_{k,1}$, and similarly for B_k.

The rate pair (R_1, R_2) is called achievable if and only if, for some large enough ℓ, there exist codes of length ℓ to be used at the relays such that $W_1 \in \{1, \ldots, 2^{\ell R_1}\}$ and $W_2 \in \{1, \ldots, 2^{\ell R_2}\}$ can be transmitted to the destination nodes, respectively, with vanishing error probability. The following theorem gives a complete characterization for $\mathcal{R}_{\mathsf{Z-chain}}$, the set of all achievable rate pairs of the deterministic Z-chain network.

Theorem 5.20 (The capacity region of deterministic Z-chain network)**.** *Consider a deterministic Z-chain network with channel gains $\{(\alpha_k, \beta_k, \gamma_k)\}_{k=1}^N$. The capacity region of this network is the set of all rate pairs (R_1, R_2) which satisfy*

$$R_1 \le \alpha_k \qquad \forall k, \tag{5.51}$$

$$R_2 \le \beta_k \qquad \forall k, \tag{5.52}$$

$$R_1 + R_2 \le \Psi_k + \Phi_N - 2\gamma_k \qquad \forall k, \tag{5.53}$$

where

$$\Psi_k \triangleq \max(\alpha_k, \gamma_k) + \max(\beta_k, \gamma_k), \qquad k = 1, 2, \ldots, N, \tag{5.54}$$

and

$$\Phi_k \triangleq \sum_{i=1}^{k} \gamma_i, \qquad k = 1, 2, \ldots, N. \tag{5.55}$$

5.7.1 The Outer Bound

In this part we prove the optimality of the rate region introduced in Theorem 5.20. Similar to the proof of Theorem 5.11, the proof of the individual rate bounds follow the min-cut theorem. However, we need a more sophisticated argument to show the sum-rate bounds, similar to the genie-aided in Theorem 5.11.

We start with a rate pair $(R_1, R_2) \in \mathcal{R}_{\mathsf{Z-chain}}$ which can be achieved using a code of length ℓ. The transmitters encode the messages into sequences $X_{1,1}^{\ell} = (X_{1,1}(1), \ldots X_{1,1}(\ell))$ and $X_{1,2}^{\ell} = (X_{1,2}(1), \ldots X_{1,2}(\ell))$. Similarly, we use $X_{k,i}^{\ell}$ to denote the codeword vectors

$$\left(X_{k,i}((k-1)\ell + 1), X_{k,i}^{\ell}((k-1)\ell + 2), \ldots, X_{k,i}^{\ell}((k-1)\ell + \ell)\right),$$

for $i = 1, 2$ and $k = 0, 1, \ldots, N-1$. The term $(k-1)\ell$ in the time index is due to the layered structure of the network and the fact that it takes $(k-1)$ time blocks for messages (W_1, W_2) to get received at the k-th layer of the network. We also use bold-face symbols to denote ℓ copies of the respective matrices, as the transfer matrix applied over a codeword of length ℓ, *i.e.*, $\mathbf{N}_{ij}^{(k)} = I_\ell \otimes N_{ij}^{(k)}$.

First, consider the cut with partitions $\Omega = \{S_1, A_1, A_2, \ldots, A_k\}$ at the transmitting part, and $\Omega^c = \{S_2, A_{k+1}, \ldots, A_{N-1}, B_1, \ldots B_{N-1}, D_1, D_2\}$ at the receiver part. Note that the two parts of this cut are connected through the links with gains $\{\gamma_1, \ldots, \gamma_{k-1}, \alpha_k\}$ (the red cut in Figure 5.15). However, the cross links carry information from Ω^c to Ω, and there is only one direct link with gain α_k which carries information from Ω to Ω^c. Therefore, in order to show to the first bound in (5.51), we can write

$$\ell R_1 \le \operatorname{rank}\left(\mathbf{G}_{\Omega,\Omega^c}\right) = \operatorname{rank}\left(\mathbf{N}_{11}^{(k)}\right) = \ell \operatorname{rank}\left(N_{11}^{(k)}\right) = \ell\alpha_k. \tag{5.56}$$

Similarly, the bound on r_2 corresponds to the maximum flow of information through the cut with partitions

$$\Omega = \{S_1, S_2, A_1, \ldots, A_{N-1}, B_1, \ldots, B_{k-1}, D_1\}$$

and

$$\Omega^c = \{B_k, \ldots, B_{N-1}, D_2\}.$$

These two sets are connected though the links with gains $\{\beta_k, \gamma_{k+1}, \ldots, \gamma_N\}$ (the blue cut in Figure 5.15), where only the direct link carries information from Ω to Ω^c. Therefore, we have

$$\ell R_2 \le \operatorname{rank}\left(\mathbf{G}_{\Omega,\Omega^c}\right) = \operatorname{rank}\left(\mathbf{N}_{22}^{(k)}\right) = \ell \operatorname{rank}\left(N_{22}^{(k)}\right) = \ell\beta_k. \tag{5.57}$$

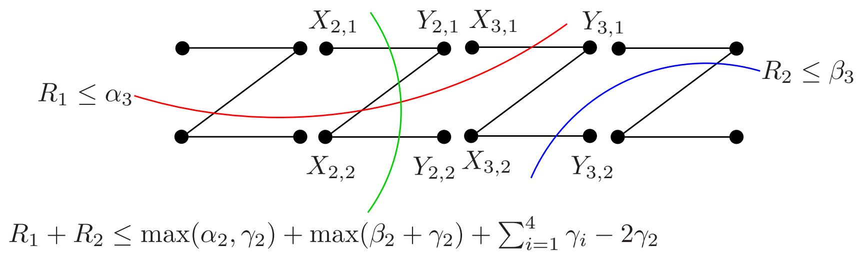


Figure 5.15: Three different kinds of cut-set bounds

The sum-rate bound is obtained by bounding the amount of information can be passed through the k-th layer of the network, namely, the edges $\{\alpha_k, \beta_k, \gamma_k\}$ (the green cut in Figure 5.15). In the following chain of inequalities we will use a new random variable which is the interference observed[7] by the relay node A_k, and defined as $\Gamma_k = N_{12}^{(k)} X_{k,2}$ for $k = 1, 2, \ldots, N$. We also use Γ_k^ℓ to denote a block of length ℓ of Γ_k.

This bound is essentially a genie-aided bound, where a genie provides the output of all the cross links except the k-th one $(\Gamma_1^\ell, \ldots, \Gamma_{k-1}^\ell, \Gamma_{k+1}^\ell, \ldots, \Gamma_N^\ell)$ at A_k. Intuitively, we capture the maximum interference neutralization, and argue that having such aid from the genie, A_k can decode the message W_1 if the destination node D_1 can decode it.

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(Y_{k,1}^\ell, Y_{k,2}^\ell; X_{k,1}^\ell, X_{k,2}^\ell) \\
&= H(Y_{k,1}^\ell, Y_{k,2}^\ell) \\
&\leq H(Y_{k,1}^\ell, Y_{k,2}^\ell, \Gamma_1^\ell, \ldots, \Gamma_{k-1}^\ell, \Gamma_{k+1}^\ell, \ldots, \Gamma_N^\ell) \\
&\leq H(Y_{k,1}^\ell) + H(\Gamma_1^\ell) + \cdots + H(\Gamma_{k-1}^\ell) + H(\Gamma_{k+1}^\ell) + \cdots + H(\Gamma_N^\ell) \\
&\quad + H(Y_{k,2}^\ell | Y_{k,1}^\ell, \Gamma_1^\ell, \ldots, \Gamma_{k-1}^\ell, \Gamma_{k+1}^\ell, \ldots, \Gamma_N^\ell). \qquad (5.58)
\end{aligned}$$

We define $\mathcal{T}_{<k} = (\Gamma_1^\ell, \ldots, \Gamma_{k-1}^\ell)$ and $\mathcal{T}_{>k} = (\Gamma_{k+1}, \ldots, \Gamma_N^\ell)$. Note that

$$\begin{aligned}
H(Y_{N,1}^\ell | Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) &\leq H(Y_{N,1}^\ell | Y_{k,1}^\ell, \mathcal{T}_{>k}) \\
&= H(Y_{N,1}^\ell - \Gamma_N^\ell | Y_{k,1}^\ell, \mathcal{T}_{>k}) \\
&= H(X_{N-1,1}^\ell | Y_{k,1}^\ell, \mathcal{T}_{>k}) \\
&\leq H(Y_{N-1,1}^\ell | Y_{k,1}^\ell, \mathcal{T}_{>k}) \\
&\vdots \\
&\leq H(Y_{k,1}^\ell | Y_{k,1}^\ell, \mathcal{T}_{>k}) = 0. \qquad (5.59)
\end{aligned}$$

[7]One can think of these random variables as the output of some auxiliary channels with the same behavior as the cross links.

Combining (5.59) with Fano's inequality, $H(W_1|Y_{N,1}^\ell) \le \ell\epsilon_\ell$, we have

$$\begin{aligned} H(W_1|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) &\le H(W_1, Y_{N,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(Y_{N,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) + H(W_1|Y_{N,1}^\ell, Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &\le \ell\epsilon_\ell, \end{aligned} \tag{5.60}$$

where $\epsilon_\ell \to 0$ as ℓ grows. Therefore,

$$\begin{aligned} H(X_{k,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) &\le H(Y_{k-1,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(Y_{k-1,1}^\ell - \Gamma_{k-1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(X_{k-1,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &\vdots \\ &\le H(X_{1,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &\le H(W_1|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \le \ell\epsilon_\ell. \end{aligned} \tag{5.61}$$

Hence,

$$\begin{aligned} H(\Gamma_k^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) &\le H(\Gamma_k^\ell, X_{k,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(X_{k,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) + H(\Gamma_k^\ell|Y_{k,1}^\ell, X_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(X_{k,1}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &\quad + H(Y_{k,1}^\ell - \mathbf{N}_{11}^{(k)} X_{k,1}^\ell|Y_{k,1}^\ell, X_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &\overset{(a)}{\le} \ell\epsilon_\ell, \end{aligned} \tag{5.62}$$

where we used (5.61) in (a). Finally, we get

$$\begin{aligned} H(Y_{k,2}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) &\le H(Y_{k,2}^\ell, \Gamma_k^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) \\ &= H(\Gamma_k^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \mathcal{T}_{>k}) + H(Y_{k,2}^\ell|Y_{k,1}^\ell, \mathcal{T}_{<k}, \Gamma_k^\ell, \mathcal{T}_{>k}) \\ &\le \ell\epsilon_\ell + H(Y_{k,2}^\ell|\Gamma_k^\ell) \\ &\le \ell(\beta_k - \gamma_k)^+ + \ell\epsilon_\ell. \end{aligned} \tag{5.63}$$

Replacing (5.63) in (5.58) we get

$$\ell(R_1 + R_2) \le \ell \max(\alpha_k, \gamma_k) + \sum_{i \ne k} \ell\gamma_i + \ell(\beta_k - \gamma_k)^+ + \ell\epsilon'_\ell$$

which yields in (5.53) after some simplifications.

Remark 5.21. *Note the bounds corresponding to the other cuts (e.g., bounds shown in Figure 5.16) are implied by (5.51)-(5.53), and do not further tighten the capacity region.*

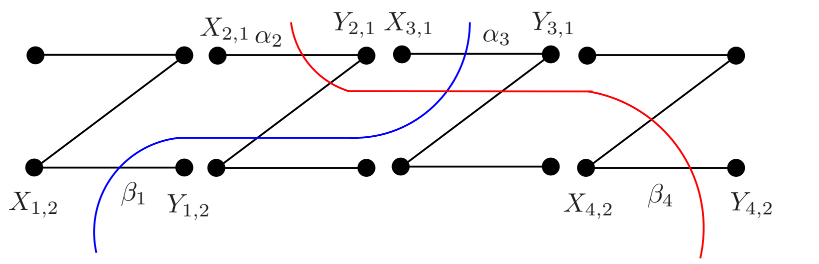


Figure 5.16: The dominated cuts in the Z-chain network.

5.7.2 The Achievability Proof

The goal of this section is to show that any rate pair $(R_1, R_2) \in \mathcal{R}_{\text{Z-chain}}$ is achievable. We will show that such rate pair is achievable using only linear operations, hence, the signal at any relay or destination node would be a linear combination of the input bits of $W_1 = [b_1(1), b_1(2), \ldots, b_1(R_1)]^T$ and $W_2 = [b_2(1), b_2(2), \ldots, b_2(R_2)]^T$, the binary representations of the input messages. It is clear that destinations can decode if and only if the nodes D_1 (D_2) can obtain exactly R_1 (R_2) *linearly independent equations* which only involve the unknown bits of W_1 (W_2) from the set of received equations. In order to show achievability, we use interference neutralization similar to the encoding scheme used in Section 5.6. Here the interfering signal is eliminated when mixed over the air, without necessarily decoding it.

We focus on a special class of encoding schemes, where the relay nodes B_k's, first decode the corresponding message W_2, and then encode it again and send *exactly* R_2 linearly independent equations describing W_2. The encoding scheme at B_k can be chosen such that the message received at A_{k+1} gets more interference, or (a part of) its interference gets neutralized. We choose R_2 nodes among the top β_k available nodes for transmission, opportunistically, such that the message can be decoded at B_{k+1} and the *desired* interfering situation happens at A_{k+1}.

Also the relay nodes A_k transmit *exactly* R_1 equations, where some of them may only involve bits from W_1 and the others involve bits of both W_1 and W_2. However, the equations are chosen such that the induced equations on each of W_1 and W_2 are linearly independent. More precisely, all the non-trivial (non-zero) bits sent by A_k are elements of the vector

$$X_{k,1} = \left[\begin{array}{c|c} U & Q \\ \hline V & 0 \end{array}\right] \left[\begin{array}{c} W_1 \\ \hline W_2 \end{array}\right] \tag{5.64}$$

where the matrices Q and $\begin{bmatrix} U^T & V^T \end{bmatrix}^T$ are full-rank[8].

[8]It is easy to show that such matrices always exist. We can start with R_1 linearly

Transmission of W_2 from S_2 to D_2 through B_k needs to only send linearly R_2 independent equations since no interference can affect the message. We need the following definition for our proof.

Definition 5.22. *A linear equation is called* pure *if it only involves the bits of W_1 as the unknown variables. For a given encoding scheme, let ρ_k denote the number of linearly independent pure equations received at the relay node A_k for $k = 1, \ldots, N-1$. Clearly, we have $\rho_0 = R_1$.*

In fact A_k has ρ_k linearly independent pure equations and $R_1 - \rho_k$ *mixed* equations involving the unknown bits of both W_1 and W_2. It may also have some equations which only involve unknown bits of W_2. Such equations can be used for interference suppression as well as interference neutralization as described in Section 5.3.

The value of ρ_k depends on both the number of pure equations in the previous layer, ρ_{k-1}, as well the encoding strategy used at the relay consisting nodes A_{k-1} and B_{k-1}. Therefore, even for a fixed ρ_{k-1}, different values for ρ_k can be obtained using different coding strategies. According to the above definition, ρ_N denotes the number of pure equations received at D_1. In the following we will study the evolution of the number of pure equations and show that if $(R_1, R_2) \in \mathcal{R}_{\mathsf{Z-chain}}$ then there *exist* coding strategies used at the relays such that one can obtain $\rho_N = R_1$ pure equations at D_1. It is clear that having R_1 linearly independent equations, D_1 can always solve the system of equations, and reconstruct the bits of W_1.

Definition 5.23. *Define $\mathcal{P}_k$ as the set of all possible number of pure equations at the k-th layer of the network for all possible encoding strategies. We define $M_k = \max \mathcal{P}_k$ and $m_k = \min \mathcal{P}_k$ as the largest and smallest elements of $\mathcal{P}_k$, respectively.*

It is clear that $M_k \leq R_1$ for all values of k, since the number of linearly independent equations cannot exceed the number of variables. Also note that we never send more than R_2 equations from B_{k-1}, and therefore the number of mixed equations cannot exceed R_2. Therefore, we have $R_1 - \rho_k \leq R_2$ and therefore $m_k \geq (R_1 - R_2)^+$.

The following example illustrates the concept of the number of pure equations and role of the encoding scheme.

Example 5.24. *Consider* Z*-chain network which is used for transmission at rate pair $(R_1, R_2) = (5, 3)$. Assume that $p = 5$, and the k-th layer of the network has parameters $\alpha_k = 5$, $\beta_k = 4$, and $\gamma_k = 2$, as shown in Figure 5.17. The relay node B_{k-1} receives R_2 linearly independent equations involving bits*

independent (possibly interfered) equations describing W_1, where those corrupted by interference form the Q part of the matrix. Then, if Q is not full-rank, one can remove one row of $\begin{bmatrix} U & Q \end{bmatrix}^T$, and add a row to V instead. This process can be repeated until all the remaining rows of Q are linearly independent.

of W_2, and therefore has access to $Y_{k-1,2}$. Assume that the relay node A_{k-1} has received $Y_{k-1,2}$ as

$$Y_{k-1,1} = \begin{bmatrix} b_1(1)+b_1(3)+b_1(4) \\ b_1(1)+b_1(2) \\ b_1(2)+b_2(1) \\ b_1(4)+b_1(5)+b_2(2)+b_2(3) \\ b_1(5)+b_2(2) \end{bmatrix}, \qquad Y_{k-1,2} = \begin{bmatrix} b_2(1) \\ b_2(2) \\ b_2(3) \\ 0 \\ 0 \end{bmatrix},$$

which means $\rho_{k-1} = 2$ is achieved.

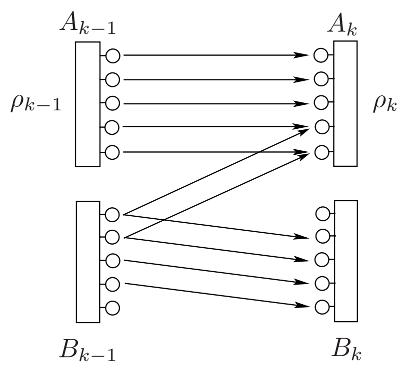


Figure 5.17: A single layer of the Z chain in Example 5.24.

The following different coding schemes used at A_{k-1} and B_{k-1} result in different values for ρ_k.

- *If the relay nodes encode their message and send the vectors*

$$X_{k,1} = \begin{bmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&1 \\ 0&0&0&0&1 \end{bmatrix} Y_{k-1,1}, \qquad X_{k,2} = \begin{bmatrix} 1&0&0&0&0 \\ 0&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&0&0 \end{bmatrix} Y_{k-1,2},$$

  then the received signals at the next layer would be $Y_{k,1} = [b_1(1)+b_1(3)+b_1(4), b_1(1)+b_1(2), b_1(2)+b_2(1), b_1(4)+b_2(1)+b_2(3), b_1(5)+b_2(2)]^T$ and $Y_{k,2} = [0, b_2(1), 0, b_2(2), b_2(3)]^T$. It is clear that ρ_k, the number of linearly independent equations involving only the bits of W_1, equals 2.

- *If the first relay simply forward the vector $X_{k,1} = y_{k-1,1}$, and the second sends an encoded message*

$$X_{k,2} = \begin{bmatrix} 0&1&1&0&0 \\ 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \end{bmatrix},$$

the relay nodes at the next layer will receive vectors $Y_{k,1} = [b_1(1)+b_1(3)+b_1(4), b_1(1)+b_1(2), b_1(2)+b_2(1), b_1(4)+b_1(5), b_1(5)+b_2(1)+b_2(2)]^T$ *and* $Y_{k,2} = [0, b_2(2)+b_2(3), b_2(1), b_2(2), 0]^T$*. Therefore we have* $\rho_k = 3$*.*

- *By re-encoding the received sequence at* A_{k-1} *and* B_{k-1} *and sending* $X_{k,1} = Y_{k-1,1}$ *and*

$$X_{k,2} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} Y_{k-1,2},$$

  the relay nodes A_k *and* B_k *will receive* $Y_{k,1} = [b_1(1)+b_1(3)+b_1(4), b_1(1)+b_1(2), b_1(2)+b_2(1), b_1(4)+b_1(5), b_1(5)+b_2(1)]^T$ *and* $Y_{k,2} = [0, b_2(2)+b_2(3), b_2(2), b_2(1)+b_2(2)+b_2(3), 0]^T$*, which corresponds to* $\rho_k = 4$*.*

It is clear through the above specific encoding strategies that if $3 \in \mathcal{P}_{k-1}$ *then* $\{2,3,4\} \subseteq \mathcal{P}_k$ *for these special channel parameters. However it is impossible to obtain* $\rho_k = 5$ *from* $\rho_{k-1} = 2$*, since among three mixed equation available at* A_{k-1}*, at most two of them can become pure by interference neutralization from the cross link.* $\square$

In the following we will investigate how the ρ_k changes from one layer of the network to the next one.

Evolution of the number of pure equations: Assume we are at the encoding part of the k-th layer of the network with parameters $(\alpha_k, \beta_k, \gamma_k)$, and have ρ_{k-1} pure and $R_1 - \rho_{k-1}$ mixed equations. We need to find m_k and M_k that can be achieved for the next layer. The following two lemmas determine the values of m_k and M_k in terms of the channel parameters.

Lemma 5.25. *Given* ρ_{k-1} *linearly independent equations at the relay node* A_{k-1}*, the minimum number of pure equations achievable at* A_k *is*

$$\min \rho_{k|\rho_{k-1}} = \max\{0, R_1 - R_2, \rho_{k-1} - \gamma_k\}.$$

We will present the proof of this Lemma in Appendix D.5.

Lemma 5.26. *If the relay node* A_{k-1} *sends* ρ_{k-1} *pure equations, then maximum achievable number of pure equations in the next layer's relay,* A_k *is*

$$\max \rho_{k|\rho_{k-1}} = \min\{R_1, \Psi_k - R_2 + R_1 - \gamma_k - \rho_{k-1}, \rho_{k-1} + \gamma_k\}.$$

The proof of this lemma can be found in Appendix D.6. The following theorem summarizes the discussion of this subsection. We will prove this theorem in Appendix D.7.

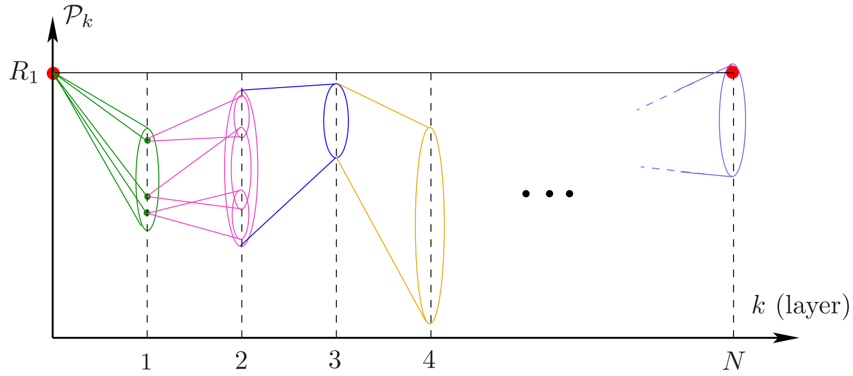


Figure 5.18: Evolution of the number of linearly independent pure equations describing W_1 in layers of the network. Each single point ρ_{k-1} can be mapped to a set of values for ρ_k, depending on the encoding scheme used at the relays.

Theorem 5.27. *The set of all achievable numbers of linearly independent pure equations at the k-th layer of the network is*

$$\mathcal{P}_k = \{p \in \mathbb{Z}^+ : m_k \le p \le M_k\} \tag{5.65}$$

where

$$m_k = \max\{0, R_1 - R_2, R_1 - \Phi_k\}, \qquad k = 1, 2, \ldots, N, \tag{5.66}$$

and M_k is obtained using the recurrence relations

$$M_k = \min\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \gamma_k + M_{k-1}\}, \tag{5.67}$$

for $k = 1, 2, \ldots, N$ and with the initial condition $M_0 = R_1$.

An achievable path for $\rho_N = R_1$: We obtained the set of achievable numbers of pure equations as a recurrence relation (see Figure 5.18). As mentioned before, showing that $R_1 \in \mathcal{P}_N$ means that there exists an encoding scheme which provides us R_1 linearly independent equations involving the bits of W_1, and hence, the decoder D_1 can solve the system of equations and decode the transmitted message. Since we do not have an explicit expression for $\mathcal{P}_N$, we cannot check if $R_1 \in \mathcal{P}_N$.

Instead, we implicitly show that for (R_1, R_2) satisfying the outer bound, $R_1 \in \mathcal{P}_N$, using the recursive form of the evolution of pure equations. Therefore, this shows that rate (R_1, R_2) is indeed achievable.

Let us define

$$\rho_k^* = \min\left(R_1, \min_{1\le t\le k} \{\Psi_t + \Phi_k - 2\gamma_t - R_2\}\right), \tag{5.68}$$

as a special number of pure equations at node A_k. Then we have the following lemma.

Lemma 5.28. *For any* Z*-chain network with arbitrary number of layers, we have*

$$\rho_k^* \in \mathcal{P}_k, \qquad k = 1, 2, \dots, N. \tag{5.69}$$

We will present the proof of this lemma in Appendix D.8. This lemma essentially gives a specific member of the set $\mathcal{P}_k$. In particular, this specific member for the last layer of the network, is exactly the number of linearly independent equations we D_1 needs to decode W_1. This is the message of the following lemma, which we prove just after it.

Lemma 5.29.

$$\rho_N^* = R_1. \tag{5.70}$$

Proof of Lemma 5.29. Note that by definition

$$\rho_N^* = \min\left(R_1, \min_{1 \le t \le N} \{\Psi_t - 2\gamma_t + \Phi_N - R_2\}\right).$$

It is clear that $\rho_N^* \le R_1$ In order to show the equality, one has to show that $R_1 \ge \Psi_\ell - 2\gamma_t + \Phi_N - R_2 \ge R_1$ for $1 \le t \le N$. But such inequality always holds, since (R_1, R_2) satisfy (5.53). □

Lemmas 5.28 and 5.29 together show that $R_1 \in \mathcal{P}_N$, *i.e.,* there exist encoding schemes used at the relays which can provide R_1 linearly independent pure equations for $A_N = D_1$. Then it is clear that D_1 can use such equations to solve for the bits of W_1.

It is worth mentioning that we used a new technique in this achievability proof. Instead of providing a specific coding scheme which supports the desired rate, we considered the set of all possible encoding schemes, and proved the *existence* of such strategy which guarantees any rate pair satisfying the outer bound. This can be thought of as an *implicit achievability argument.*

6 The Gaussian Relay-Interference Network

We studied some classes of interference-relay networks under linear deterministic model in Chapter 5. The main goal of this study is, in fact, to simplify the original Gaussian (noisy) network to a deterministic one, and get insights for both transmission strategies as well as bounding techniques which can be translated from the simplified network to the Gaussian one. In particular, we see that the proof techniques used in the outer bound, as well as the transmission schemes used in achievability for the Gaussian case are inspired by the corresponding deterministic network.

This path is illustrated in Figure 6.1. This approach consists of three steps: (i) simplify the model; (ii) obtain optimal solution for the simplified model; (iii) translate the optimal scheme and outer bounds back to the original model. Finally, we have to show that the gap between the performance of the translated scheme an the outer bound is small, which results in an approximate solution for the problem.

We have already studied the deterministic ZS and ZZ networks in Chapter 5, where we derived the exact capacity region for each, and demonstrated the optimal transmission schemes can be used to achieve these capacity regions. These characterizations will be translated into a universally approximate characterization for the (noisy) Gaussian network in this chapter. More importantly, the transmission strategies used for the deterministic networks have to be translated and adapted to the Gaussian networks.

Here, we focus on the real Gaussian noise and channel gain for simplicity, although all the presented results can be generalized to complex Gaussian parameters.

In this chapter, after a formal statement of the problem and introducing the notation, we present approximate capacity regions for the Gaussian ZS (GZS) and Gaussian ZZ (GZZ) networks in Section 6.2. This is done by introducing

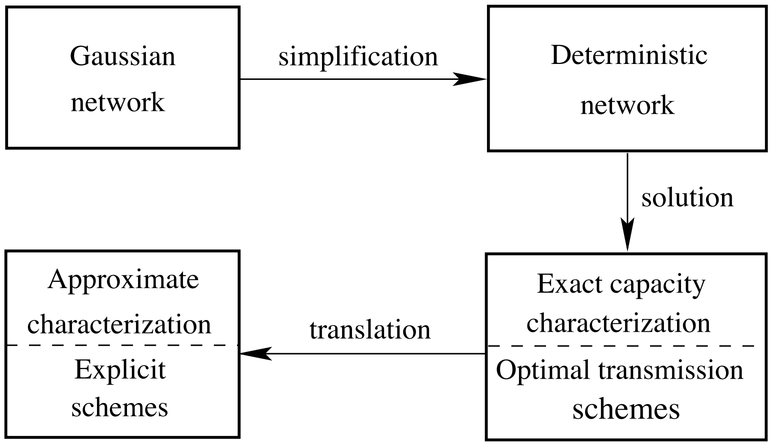


Figure 6.1: The deterministic approach to approximately solve the wireless problem.

inner and outer bound for the achievable rate regions, and proving a universal constant gap between the bounds.

We prove the approximation result for the GZS network in Section 6.4. The outer bound derived for the capacity of DZS network will be adapted to the Gaussian network. Here, the main difficulty is that unlike the deterministic case, it is not a-priori clear which input distribution optimizes the cut-set bound. Moreover, the network decomposition used to achieve the capacity of the DZS cannot directly applied to the Gaussian network. This can be done using message splitting, superposition coding and proper power allocation. We will use this technique to achieve an approximate capacity for the Gaussian ZS network in Section 6.3.

The outer bound for the capacity region of the GZZ network is provided based on genie-aided sum-rate bounds similar to that of the DZZ network. The proof of this is presented in Section 6.5. As mentioned earlier, the interference neutralization strategy used for the deterministic ZZ in Chapter 5, has a Gaussian counterpart, which is used to obtain the approximate capacity region. This technique is performed by using *lattice codes* in a Gaussian network. In order to illustrate this idea, we first present the technique and fundamental operating blocks in Section 6.3, and then utilize them in Section 6.5.

6.1 Problem Formulation

As mentioned before, our goal in this chapter is to derive approximate capacity characterizations for a class of 2-user relay-interference networks shown in Figure 5.1, which we call the XX network. In particular, we are interested in capacity characterization of XX networks with two weak cross links, which are termed as GZS and GZZ networks. We start by describing our notation for Gaussian channels.

Two transmitters, S_1 and S_2, encode their messages W_1 and W_2 of rates R_1 and R_2, respectively, and broadcast the obtained signals to the relay nodes, A and B. Denote the transmitted signals by x_1 and x_2, and the received signals at the relays by y_1' and y_2'. Then, we have

$$\begin{aligned} y_1'[t] &= \sqrt{g_{11}}x_1[t] + \sqrt{g_{12}}x_2[t] + z_1'[t], \\ y_2'[t] &= \sqrt{g_{21}}x_1[t] + \sqrt{g_{22}}x_2[t] + z_2'[t], \end{aligned} \tag{6.1}$$

where z_1', z_2' are unit-variance Gaussian noises, independent of each other and of x_1, x_2.

The relay nodes perform any (causal) processing on their received signal sequences $\{y_1'[t]\}$ and $\{y_2'[t]\}$ respectively, to obtain their transmitting signal sequences, $\{x_1'(t)\}$ and $\{x_2'(t)\}$. The received signals at the destination nodes can be written as

$$\begin{aligned} y_1[t] &= \sqrt{h_{11}}x_1'[t] + \sqrt{h_{12}}x_2'[t] + z_1[t] \\ y_2[t] &= \sqrt{h_{21}}x_1'[t] + \sqrt{h_{22}}x_2'[t] + z_2[t], \end{aligned} \tag{6.2}$$

where the z_1', z_2', z_1, and z_2 are independent zero-mean unit-variance noises, which are also independent of x_1 and x_2. There is a power constraint for each transmitted signal, that is, $\mathbb{E}[x_1^2] \leq 1$, $\mathbb{E}[x_2^2] \leq 1$, $\mathbb{E}[x_1'^2] \leq 1$ and $\mathbb{E}[x_2'^2] \leq 1$.

Each destination node D_i, $i = 1, 2$, is interested in decoding its message W_i, using its received signals $\{y_i[t]\}$. We define a rate pair (R_1, R_2) to be admissible if there exist a transmission scheme under which D_1 and D_2 can decode W_1 and W_2, respectively, with arbitrary small (average) error probability in the standard manner [3]. This would allow two end-to-end reliable unicast sessions at rates (R_1, R_2) for the source/destination pairs (S_1, D_1) and (S_2, D_2).

In particular, we are interested in sub-classes of the Gaussian XX network with two weak links. As mentioned before, if such weak links are in the same layer, then the network would be converted to an interference channel cascaded with a parallel line network, whose approximate capacity is well-studied in the literature [10, 15, 73]. On the other hand if the weak links lie in different layers of the network, the either have a ZS or ZZ network.

6.1.1 The Gaussian ZS Network

The ZS network is a special case of the interference-relay network defined in (6.1)-(6.2). In the ZS network one cross link in each layer has a negligible gain, and therefore does not cause interference, as illustrated in Figure 6.2. In particular, we assume $g_{21} = h_{12} = 0$. The resulting GZS network is shown in Figure 6.2.

6.1.2 The Gaussian ZZ Network

The ZZ network is another special configuration interference-relay network, wherein one cross link in each layer has zero gain. However, the difference is

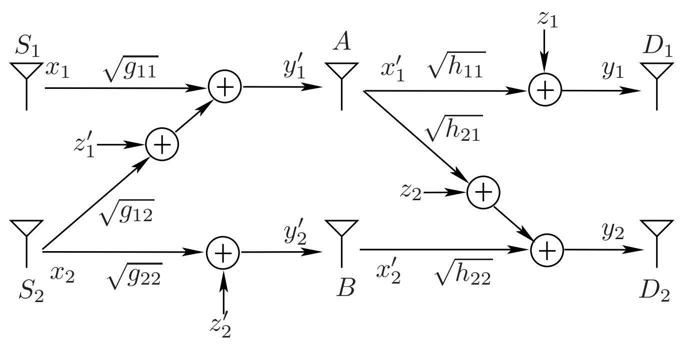


Figure 6.2: The Gaussian ZS network.

that, here the missing links are in parallel. In particular, we assume $g_{21} = h_{21} = 0$. The Gaussian ZZ networks is shown in Figure 6.3.

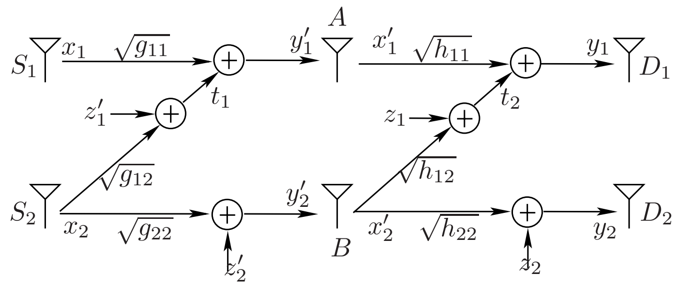


Figure 6.3: The Gaussian ZZ network.

6.2 Main Results

In this section we present the main results of this chapter, which is the approximate capacity characterization of the Gaussian ZS and ZZ interference-relay networks. The insights are obtained by analyzing the deterministic versions of these problems in Chapter 5 lead to these approximate characterizations. The achievability and outer bound results for the Gaussian cases are directly inspired by these results.

The coding strategies and fundamental communication blocks for the Gaussian problem are outlined in Section 6.3. The detailed analysis of these strategies and the corresponding outer bounds which lead to Theorems 6.2 and 6.4 are given in Appendix E.

6.2.1 The Gaussian ZS Network

Theorem 6.2 gives approximate (within 2 bits) characterizations for the capacity of the Gaussian ZS network illustrated in Figure 6.2.

Definition 6.1. *For a given* GZS *network with channel gains* $\{g_{11}, g_{12}, g_{22}\}$ *and* $\{h_{11}, h_{21}, h_{22}\}$*, let* $\underline{\mathcal{R}}_{GZS}$ *be the set of all rate pairs* (R_1, R_2) *which satisfying*

$$R_1 \leq \frac{1}{2}\log(1 + g_{11}) \tag{GZS-1}$$

$$R_2 \leq \frac{1}{2}\log(1 + g_{12} + g_{22}) \tag{GZS-2}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) \tag{GZS-3}$$

$$R_2 \leq \frac{1}{2}\log(1 + g_{12}) + \frac{1}{2}\log(1 + h_{22}) \tag{GZS-4}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{22}) + \frac{1}{2}\log(1 + h_{11} + h_{21}) \tag{GZS-5}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log(1 + h_{22}) \tag{GZS-6}$$

$$R_1 \leq \frac{1}{2}\log(1 + h_{11}) \tag{GZS-7}$$

$$R_2 \leq \frac{1}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}) \tag{GZS-8}$$

$$R_2 \leq \frac{1}{2}\log(1 + g_{22}) + \frac{1}{2}\log(1 + h_{21}) \tag{GZS-9}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}) + \frac{1}{2}\log\left(1 + \frac{h_{11}}{h_{21}}\right). \tag{GZS-10}$$

We also define $\overline{\mathcal{R}}_{\mathsf{GZS}}$ *as*

$$\overline{\mathcal{R}}_{\mathsf{GZS}} = \{(R_1, R_2) \in \mathbb{R}_+^2 : (R_1 + 1, R_2 + 1.5) \in \underline{\mathcal{R}}_{\mathsf{GZS}}\}. \tag{6.3}$$

Theorem 6.2 (An Approximate capacity region of Gaussian ZS network)**.** *The set* $\underline{\mathcal{R}}^{\mathsf{GZS}}$ *is an outer bound for the capacity region of the Gaussian* ZS *network. Moreover, for any* $(R_1, R_2) \in \underline{\mathcal{R}}_{\mathsf{GZS}}$*, there exists a transmission scheme with rates* $(R_1', R_2') = (R_1 - \delta_1, R_2 - \delta_2)$*, where* $\delta_1 = 1$ *and* $\delta_2 = 1.5$ *are universal constants, independent of the channel gain, and required rates. More precisely,*

$$\overline{\mathcal{R}}_{\mathsf{GZS}} \subseteq \mathcal{R}_{\mathsf{GZS}} \subseteq \underline{\mathcal{R}}_{\mathsf{GZS}}. \tag{6.4}$$

Similar to the deterministic case, the outer bound this theorem follows a fairly standard arguments based on reducing a multi-letter mutual information into single-letter forms by appropriately using decodability requirements at the different destinations. The details of these are given in Section 6.4.

The insight from the network decomposition observed in Section 5.5 leads to the idea of strategic *rate-splitting* and power allocation in the Gaussian channel. For the Gaussian coding scheme, we need to strategically partition the messages and allocate powers in order for the relays to partially decode appropriate messages and setup cooperation. The details of this strategy are outlined in Section 6.4.

6.2.2 The Gaussian ZZ Network

The Gaussian ZZ network is illustrated in Figure 6.3. Although superficially the ZS and ZZ networks may look similar, the subtle difference in the network connectivity, makes the two problems completely different, both in terms of capacity characterization, as well as transmission schemes. It will be shown that a new interference management scheme, which we term as interference neutralization, is needed to (approximately) achieve the capacity of this network. The most intuitive description for interference neutralization is to cancel interference over air without processing at the destinations. This scheme can be used whenever there are more than one path for interference to get received at a destination. We will explain it in more detail in Sections 6.5.1.

Theorem 6.4 gives an approximate (within 2 bits) characterizations for the capacity region of the deterministic and the Gaussian ZZ networks, respectively. Another new ingredient used here is needed a genie-aided outer bound that gives the (noisy) cross link of the first (or correspondingly second) layer to the destination (or correspondingly to the relay). This genie-aided bound allows us to develop outer bounds that are apparently tighter than the information-theoretic cut-set bounds by utilizing the decoding structure needed.

Definition 6.3. *For a* GZZ *network, with channel gains* $\{g_{11}, g_{12}, g_{22}\}$ *in the first, and* $\{h_{11}, h_{12}, h_{22}\}$ *in the second layer, define* $\underline{\mathcal{R}}_{\mathsf{GZZ}}$ *to be the set of all non-negative* (R_1, R_2) *which satisfy*

$$R_1 \leq \frac{1}{2}\log(1+g_{11}) \tag{GZZ-1}$$

$$R_2 \leq \frac{1}{2}\log(1+g_{22}) \tag{GZZ-2}$$

$$R_1 \leq \frac{1}{2}\log(1+h_{11}) \tag{GZZ-3}$$

$$R_2 \leq \frac{1}{2}\log(1+h_{22}) \tag{GZZ-4}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) + \frac{1}{2}\log(1+h_{12}), \tag{GZZ-5}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1+h_{11}+h_{12}) + \frac{1}{2}\log\left(1+\frac{h_{22}}{h_{12}}\right) + \frac{1}{2}\log(1+g_{12}). \tag{GZZ-6}$$

Moreover, define $\overline{\mathcal{R}}_{\mathsf{GZZ}}$ *as*

$$\overline{\mathcal{R}}_{\mathsf{GZZ}} = \{(R_1, R_2) \in \mathbb{R}_+^2 : (R_1 + \frac{1}{4}\log 96, R_2 + \frac{1}{4}\log 96) \in \underline{\mathcal{R}}_{\mathsf{GZZ}}\}. \tag{6.5}$$

Theorem 6.4 (An approximate capacity region of Gaussian ZZ network)**.** *Any achievable rate pair* (R_1, R_2) *for the Gaussian* ZZ *networks belongs to* $\underline{\mathcal{R}}_{\mathsf{GZZ}}$. *Moreover, for any rate pair* $(R_1, R_2) \in \underline{\mathcal{R}}_{\mathsf{GZZ}}$, *there exists an encoding scheme with rates* $(R_1', R_2') = (R_1 - \frac{7}{4}, R_2 - \frac{7}{4})$. *More precisely,*

$$\overline{\mathcal{R}}_{\mathsf{GZZ}} \subseteq \mathcal{R}_{\mathsf{GZZ}} \subseteq \underline{\mathcal{R}}_{\mathsf{GZZ}}. \tag{6.6}$$

6.3 Examples Illustrating Transmission Techniques

This section is devoted to providing the basic ideas of the coding schemes used in the Gaussian ZS and ZZ networks. We present how the transmission schemes used in the deterministic ZS and ZZ networks can be adapted to the Gaussian networks through some examples. Then we develop an outline of how to analyze these coding strategies.

We start with an example which illustrates the essence of approximation in a multiple unicast network.

Example 6.5 (Gaussian Z network)**.** *Consider the Gaussian* Z *network shown in Figure 6.4, with channel gains* $g_{11} \geq 1$, $g_{12} \geq 1$, *and* $g_{22} \geq 1$.

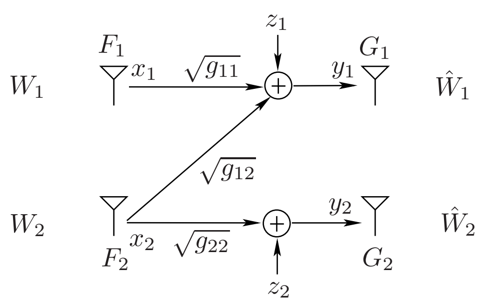


Figure 6.4: A Gaussian Z network.

The source nodes F_i *wishes to encode and send message* W_i *to the destination node* G_i, *for* $i = 1, 2$. *Denoting the rate of message* W_i *by* R_i, *an* *approximate capacity characterization for this network is given by*

$$\begin{aligned}\mathcal{R}_{\mathsf{Z}} = \Big\{(R_1, R_2) : & R_1 \leq \frac{1}{2}\log(1+g_{11}) \\ & R_2 \leq \frac{1}{2}\log(1+g_{22}) \\ & R_1 + R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right)\Big\}.\end{aligned} \tag{6.7}$$

It is easy to show that any achievable rate pair belongs to $\mathcal{R}_{\mathsf{Z}}$*, and hence* $\mathcal{R}_{\mathsf{Z}}$ *establishes an outer bound for the capacity region. Moreover, one can show that the rate pair* $(R_1 - \frac{1}{2}, R_2 - \frac{1}{2})$ *is achievable provided that* $(R_1, R_2) \in \mathcal{R}_{\mathsf{Z}}$*. The encoding strategy to achieve such rate pair involves message splitting and proper power allocation. We will discuss this in more details in Appendix E.1.*

Our next example shows how interference neutralization can be implemented in a Gaussian network using lattice codes.

Example 6.6 (Use of Lattice Codes to Implement Interference Neutralization over Gaussian ZZ Network)**.** *The idea of interference neutralization illustrated in Example 5.8 can be also used in Gaussian networks. In this case a group structured code, such as lattice code, is required to play the role of composition and decomposition of the signal and interference in two layers of the network. Consider the Gaussian* ZZ *network in Figure 6.3. We can use message splitting and interference neutralization to improve the achievable rate pairs of this network.*

Let the second source split its message into two parts as $W_2 = (W_2^{(N)}, W_2^{(P)})$*, namely, the functional (neutralization) and private parts, of rates* $R_{2,N} = R_1$ *and* $R_{2,P} = R_2 - R_1$*. Both transmitters use a common lattice code to encode* W_1 *and* $W_2^{(N)}$*, and map them into* $\mathbf{x}_1^{(N)}$ *and* $\mathbf{x}_2^{(N)}$*, respectively. The other message* $W_2^{(P)}$ *can be encoded to* $\mathbf{x}_2^{(P)}$ *using a random Gaussian code. We assume that both the lattice code and the random Gaussian code have average power equal to* 1*. Then, the transmitting signals would be a linear combination of the codewords with a proper power allocation,* i.e.,

$$\mathbf{x}_1 = \sqrt{\alpha_N}\mathbf{x}_1^{(N)}, \qquad \mathbf{x}_2 = \sqrt{\beta_N}\mathbf{x}_2^{(N)} + \sqrt{\beta_P}\mathbf{x}_2^{(P)}, \tag{6.8}$$

where the power allocation coefficients satisfy $\alpha_N \leq 1$ *and* $\beta_N + \beta_P \leq 1$*. The transmitters choose the power allocated to* $\mathbf{x}_1^{(N)}$ *to* $\mathbf{x}_2^{(N)}$ *in a way that they get received at A with the same power. In this way, their summation would be again a lattice code and can be decoded at A by treating* $\mathbf{x}_1^{(P)}$ *as noise. A similar strategy will be used for signaling at the relay for transmission in the second layer of the network. The only difference is that instead of sending* $\mathbf{x}_2^{(N)}$*, the relay node B sends* $-\mathbf{x}_2^{(N)}$*. Then, the lattice point observed at* D_1 *would be exactly* $\mathbf{x}_1^{(N)}$ *and it can find* W_1*. The other decoder can simply first reverse* $-\mathbf{x}_2^{(N)}$ *to* $\mathbf{x}_2^{(N)}$*, and then decode it. This is illustrated in Figure 6.5. It is worth mentioning that the relay decodes a function of the messages rather than the messages themselves.*

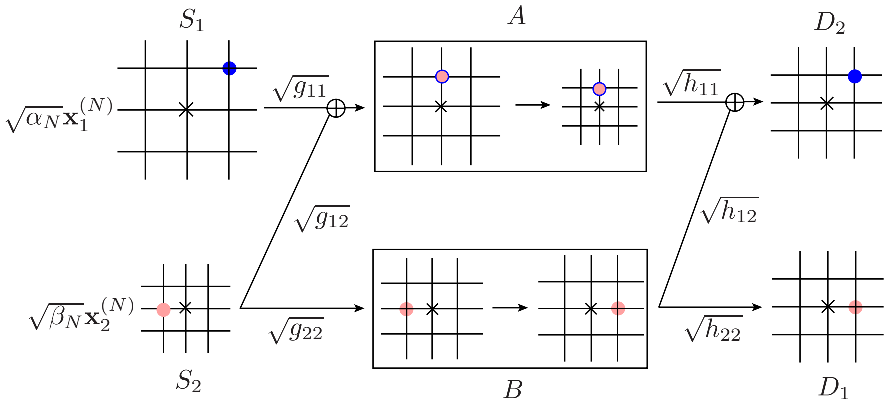


Figure 6.5: Using lattice codes for interference neutralization over a Gaussian ZZ network. The origin is specified by a cross "$\times$". Power allocated to the messages at the transmitters are chosen such that the two lattice points corresponding to $\mathbf{x}_1^{(N)}$ and $\mathbf{x}_2^{(N)}$ get received at B at the same power level, and their summation becomes a point on the scaled lattice. The same strategy is used by the relays. The relay node B also reverses its transmitting lattice point in order to neutralize the interference caused in the first layer of the network.

6.4 The Gaussian ZS Network

In this section we study the Gaussian ZS network in detail. We first prove the optimality of the rate region $\underline{\mathcal{R}}_{\mathsf{GZS}}$. We then present an encoding scheme whose achievable rate region is only constant bits away from the outer bound. These together lead to an approximate capacity characterization for the GZS network.

6.4.1 The Outer Bound

In the following we will prove each of the inequalities in (GZS-1)-(GZS-10), separately. We will use the notation as shown in Figure 6.6, and assume that the rate pair (R_1, R_2) can be achieved with small enough decoding error probability using a code of length ℓ.

Lemma 6.7. *Any achievable rate pair* (R_1, R_2) *satisfies*

$$\ell R_1 \leq I(x_1^\ell; y_1^\ell) + \ell\epsilon_\ell, \tag{6.9}$$

$$\ell R_2 \leq I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell, \tag{6.10}$$

$$\ell(R_1 + R_2) \leq I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\epsilon_\ell. \tag{6.11}$$

Note that $\epsilon_\ell \to 0$ *as* ℓ *grows.*

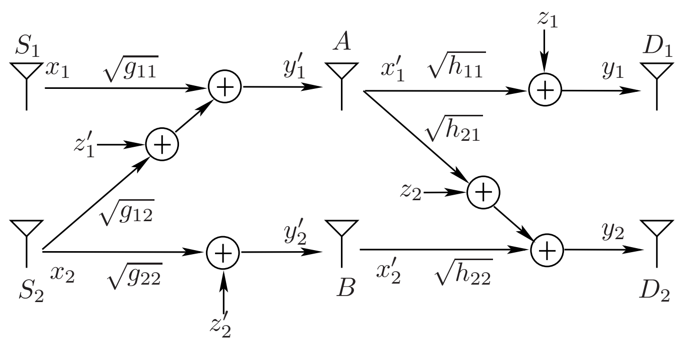


Figure 6.6: The Gaussian ZS network.

This result is a consequence of the Fano's lemma combined with the decodability requirements imposed by the problem, and its proof is given in Appendix E.2.

Most of the inequalities in (GZS-1)-(GZS-10) are cut-set type bounds, although the proof presented here is slightly different than the standard argument. However, the sum-rate bounds in (GZS-3) and (GZS-10) are different from the well known cut-set bounds. These two bounds are in general tighter than the cut values for the corresponding cuts. This is because the decoders are inherently allowed to cooperate in deriving a cut-set bound, while individual decoding abilities are imposed in this problem. In the following we first present the proofs of (GZS-3) and (GZS-10), which are more involved, and then prove the cut-set type bounds.

a) The proofs of non-cut-set type bounds

▷ (GZS-3): $R_1 + R_2 < \frac{1}{2}\log(1 + g_{11} + g_{12}) + \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right)$

We start with Lemma 6.7 for the sum-rate which implies

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{\leq} I(x_1^\ell, x_2^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\epsilon_\ell \\
&= I(x_1^\ell, x_2^\ell; y_1^{'\ell}) + I(x_1^\ell, x_2^\ell; y_2^{'\ell}|y_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1 + g_{11} + g_{12}) + h(y_2^{'\ell}|y_1^{'\ell}) - h(y_2^{'\ell}|y_1^{'\ell}, x_1^\ell, x_2^\ell) + \ell\epsilon_\ell,
\end{aligned} \tag{6.12}$$

where (a) follows from the data processing inequality. Now, note that

$$\begin{aligned}
h(y_2^{'\ell}, W_1|y_1^{'\ell}) &= h(y_2^{'\ell}|y_1^{'\ell}) + H(W_1|y_1^{'\ell}, y_2^{'\ell}) \\
&= H(W_1|y_1^{'\ell}) + h(y_2^{'\ell}|W_1, y_1^{'\ell}) \\
&\leq H(W_1|y_1^\ell) + h(y_2^{'\ell}|W_1, y_1^{'\ell}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
h(y_2^{'\ell}|y_1^{'\ell}) &\leq h(y_2^{'\ell}|W_1, y_1^{'\ell}) + \ell\epsilon_\ell \\
&\overset{(b)}{\leq} h(y_2^{'\ell}|x_1^\ell, y_1^{'\ell}) + \ell\epsilon_\ell \\
&\overset{(c)}{=} h(y_2^{'\ell}|x_1^\ell, \sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq h(\sqrt{g_{22}}x_2^\ell + z_2^{'\ell}|\sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\epsilon_\ell \\
&= h(\sqrt{g_{22}}x_2^\ell + z_2^{'\ell} - \frac{\sqrt{g_{22}}}{\sqrt{g_{12}}}(\sqrt{g_{12}}x_2^\ell + z_1^{'\ell})|\sqrt{g_{12}}x_2^\ell + z_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq h(z_2^{'\ell} - \frac{\sqrt{g_{22}}}{\sqrt{g_{12}}}z_1^{'\ell}) + \ell\epsilon_\ell \\
&= \frac{\ell}{2}\log(2\pi e)\left(1 + \frac{g_{22}}{g_{12}}\right) + \ell\epsilon_\ell, \qquad (6.13)
\end{aligned}$$

where (b) holds since x_1^ℓ is a function of W_1, and in (c) we used the invertibility property of the function $y_1^\ell = \sqrt{g_{11}}x_1^\ell + \sqrt{g_{12}}x_2^\ell + z_1^{'\ell}$. Replacing $h(y_2^{'\ell}|y_1^{'\ell})$ from (6.13) in (6.12), we get the desired bound.

▷ (GZS-10): $R_1 + R_2 < \frac{1}{2}\log\left(1 + \frac{h_{11}}{h_{21}}\right) + \frac{1}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}})$

The sum-rate can be upper bounded as in Lemma 6.7. Next, we have

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\epsilon_\ell \\
&\leq I(x_1^{'\ell}, x_2^{'\ell}; y_1^\ell, y_2^\ell) + \ell\epsilon_\ell \\
&= I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) + I(x_2^{'\ell}; y_1^\ell|y_2^\ell) + I(x_1^{'\ell}; y_1^\ell|x_2^{'\ell}, y_2^\ell) + \ell\epsilon_\ell. \qquad (6.14)
\end{aligned}$$

The first term in (6.14) can be simply upper bounded as

$$I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) \leq \frac{\ell}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}). \qquad (6.15)$$

In order to bound the second term, we can use the fact that W_2 can be decoded from y_2^ℓ, and write

$$\begin{aligned}
I(x_2^{'\ell}; y_1^\ell, W_2|y_2^\ell) &= I(x_2^{'\ell}; y_1^\ell|y_2^\ell) + I(x_2^{'\ell}; W_2|y_1^\ell, y_2^\ell) \\
&= I(x_2^{'\ell}; y_1^\ell|W_2, y_2^\ell) + I(x_2^{'\ell}; W_2|y_2^\ell) \\
&\leq I(x_2^{'\ell}; y_1^\ell|W_2, y_2^\ell) + H(W_2|y_2^\ell) \\
&\leq I(x_2^{'\ell}; y_1^\ell|y_2^\ell, W_2) + \ell\epsilon_\ell.
\end{aligned}$$

Therefore,

$$I(x_2^{'\ell}; y_1^\ell|y_2^\ell) \leq I(x_2^{'\ell}; y_1^\ell|y_2^\ell, W_2) + \ell\epsilon_\ell \leq I(y_2^{'\ell}; y_1^\ell|y_2^\ell, W_2) + \ell\epsilon_\ell = \ell\epsilon_\ell, \qquad (6.16)$$

where the second inequality follows from the fact that $x_2^{'\ell}$ is a function of $y_2^{'\ell}$, and (6.16) holds since $y_2^{'\ell}$ and y_1^ℓ are independent if W_2 is given.

Finally, we bound the last term as follows.

$$\begin{aligned}
I(x_1^{'\ell}; y_1^\ell | x_2^{'\ell}, y_2^\ell) &= I(x_1^{'\ell}; y_1^\ell | x_2^{'\ell}, \sqrt{h_{21}} x_1^{'\ell} + z_2^\ell) \\
&= h(\sqrt{h_{11}} x_1^{'\ell} + z_1^\ell | x_2^{'\ell}, \sqrt{h_{21}} x_1^{'\ell} + z_2^\ell) \\
&\quad - h(y_1^\ell | x_2^{'\ell}, \sqrt{h_{21}} x_1^{'\ell} + z_2^\ell, x_1^{'\ell}) \\
&\leq h(\sqrt{h_{11}} x_1^{'\ell} + z_1^\ell - \frac{\sqrt{h_{11}}}{\sqrt{h_{21}}}(\sqrt{h_{21}} x_1^{'\ell} + z_2^\ell)) - h(z_1^\ell) \\
&\leq \frac{\ell}{2} \log\left(1 + \frac{h_{11}}{h_{12}}\right). \qquad (6.17)
\end{aligned}$$

Replacing the bound derived for the three terms, (6.15), (6.16), and (6.17) in (6.14), we get the desired bound.

b) The proofs of cut-set type bounds

$\triangleright$ (GZS-1): $R_1 < \frac{1}{2}\log(1 + g_{11})$

We start by Lemma 6.7, and write

$$\begin{aligned}
\ell R_1 &= I(x_1^\ell; y_1^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{\leq} I(x_1^\ell; y_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq I(x_1^\ell; x_2^\ell, y_1^{'\ell}) + \ell\epsilon_\ell \\
&\overset{(b)}{=} I(x_1^\ell; x_2^\ell) + I(x_1^\ell; y_1^{'\ell} | x_2^\ell) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1 + g_{11}) + \ell\epsilon_\ell \qquad (6.18)
\end{aligned}$$

where (a) follows from the data-processing inequality for the Markov chain $x_1^\ell \leftrightarrow y_1^{'\ell} \leftrightarrow x_1^{'\ell} \leftrightarrow y_1^\ell$, and in (b) we used the fact that x_1^ℓ and x_2^ℓ are independent. It is worth mentioning that this inequality essentially bounds the maximum flow that can be transmitted through the cut $\Omega_s = \{S_1\}$ and $\Omega_d = \{S_2, A, B, D_1, D_2\}$.

$\triangleright$ (GZS-2): $R_2 < \frac{1}{2}\log(1 + g_{12} + g_{22})$

Again starting from Lemma 6.7, we have

$$\begin{aligned}
\ell R_2 &\leq I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{\leq} I(x_2^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\epsilon_\ell \\
&\leq I(x_2^\ell; x_1^\ell, y_1^{'\ell}, y_2^{'\ell}) + \ell\epsilon_\ell \\
&= I(x_2^\ell; x_1^\ell) + I(x_2^\ell; y_1^{'\ell}, y_2^{'\ell} | x_1^\ell) + \ell\epsilon_\ell \\
&= h(y_1^{'\ell}, y_2^{'\ell} | x_1^\ell) - h(y_1^{'\ell}, y_2^{'\ell} | x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\leq h(\sqrt{g_{12}} x_2^\ell + z_1^{'\ell}, \sqrt{g_{22}} x_2^\ell + z_2^{'\ell}) - h(z_1^{'\ell}, z_2^{'\ell}) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1 + g_{12} + g_{22}) + \ell\epsilon_\ell, \qquad (6.19)
\end{aligned}$$

where the data processing inequality implies (b) for the Markov chain $x_2^\ell \leftrightarrow (y_1'^\ell, y_2'^\ell) \leftrightarrow (x_1'^\ell, x_2'^\ell) \leftrightarrow y_1^\ell$. Note that this bound is essentially the cut-set bound for the cut $\Omega_s = \{S_2\}$ and $\Omega_d = \{S_1, A, B, D_1, D_2\}$.

▷ (GZS-4): $R_2 < \frac{1}{2}\log(1+g_{12}) + \frac{1}{2}\log(1+h_{22})$

Again we use Lemma 6.7 to upper bound R_2 as

$$\begin{aligned}
\ell R_2 &\le I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell \\
&\le I(x_2'^\ell, x_2^\ell; x_1^\ell, y_1'^\ell, y_2^\ell) + \ell\epsilon_\ell \\
&= I(x_2'^\ell, x_2^\ell; x_1^\ell) + I(x_2'^\ell, x_2^\ell; y_1'^\ell, y_2^\ell | x_1^\ell) + \ell\epsilon_\ell \\
&= I(x_2'^\ell, x_2^\ell; y_1'^\ell | x_1^\ell) + I(x_2'^\ell, x_2^\ell; y_2^\ell | x_1^\ell, y_1'^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{=} I(x_2^\ell; y_1'^\ell | x_1^\ell) + I(x_2'^\ell; y_1'^\ell | x_1^\ell, x_2^\ell) \\
&\quad + I(x_2'^\ell; y_2^\ell | x_1^\ell, y_1'^\ell) + I(x_2^\ell; y_2^\ell | x_1^\ell, y_1'^\ell, x_2'^\ell) + \ell\epsilon_\ell \\
&= I(x_2^\ell; y_1'^\ell | x_1^\ell) + I(x_2'^\ell; y_2^\ell | x_1^\ell, y_1'^\ell) + \ell\epsilon_\ell \\
&\le \frac{\ell}{2}\log(1+g_{12}) + \frac{\ell}{2}\log(1+h_{22}) + \ell\epsilon_\ell. \qquad (6.20)
\end{aligned}$$

Note that we used the fact that the second and fourth terms in (a) are zero. This follows from

$$\begin{aligned}
I(x_2'^\ell; y_1'^\ell | x_1^\ell, x_2^\ell) &\le I(x_2'^\ell; y_1'^\ell - \sqrt{g_{11}}x_1^\ell - \sqrt{g_{12}}x_2^\ell | x_1^\ell, x_2^\ell) \\
&= I(x_2'^\ell; z_1'^\ell | x_1^\ell, x_2^\ell) = 0,
\end{aligned}$$

and

$$\begin{aligned}
I(x_2^\ell; y_2^\ell | x_1^\ell, y_1'^\ell, x_2'^\ell) &\le I(x_2^\ell; y_2^\ell | x_1^\ell, x_1'^\ell, x_2'^\ell) \\
&\le I(x_2^\ell; y_2^\ell - \sqrt{h_{21}}x_1'^\ell - \sqrt{h_{22}}x_2'^\ell | x_1^\ell, x_1'^\ell, x_2'^\ell) \\
&\le I(x_2^\ell; z_2^\ell | x_1^\ell, x_1'^\ell, x_2'^\ell) = 0.
\end{aligned}$$

▷ (GZS-5): $R_1 + R_2 < \frac{1}{2}\log(1+g_{22}) + \frac{1}{2}\log(1+h_{11}+h_{21})$

We start from Lemma 6.7 and write

$$\begin{aligned}
\ell(R_1+R_2) &\le I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{\le} I(y_1^\ell, y_2^\ell; x_1'^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\le I(y_1^\ell, y_2^\ell, y_2'^\ell; x_1'^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&= I(y_2'^\ell; x_2^\ell) + I(y_2'^\ell; x_1'^\ell | x_2^\ell) + I(y_1^\ell, y_2^\ell; x_1'^\ell, x_2^\ell | y_2'^\ell) + \ell\epsilon_\ell \\
&= I(y_2'^\ell; x_2^\ell) + I(y_1^\ell, y_2^\ell; x_1'^\ell, x_2^\ell | x_2'^\ell) + \ell\epsilon_\ell \\
&= I(y_2'^\ell; x_2^\ell) + h(y_1^\ell, y_2^\ell | x_2'^\ell) - h(y_1^\ell, y_2^\ell | x_2^\ell, x_1'^\ell, x_2'^\ell) + \ell\epsilon_\ell \\
&\le I(y_2'^\ell; x_2^\ell) + h(\sqrt{h_{11}}x_1'^\ell + z_1^\ell, \sqrt{h_{21}}x_1'^\ell + z_2^\ell) - h(z_1^\ell, z_2^\ell) + \ell\epsilon_\ell
\end{aligned}$$

$$\leq \frac{\ell}{2}\log(1+g_{22}) + \frac{\ell}{2}\log(1+h_{11}+h_{21}) + \ell\epsilon_\ell \tag{6.21}$$

where (a) follows from the data processing inequality and the fact that given $x_1^{'}$, x_1 is independent of all the other transmitted/received signals. Note that this bound essentially captures the maximum flow of information through the cut $\Omega_s = \{S_1, S_2, A\}$ and $\Omega_d = \{B, D_1, D_2\}$.

$\triangleright$ (GZS-6): $\quad R_1 + R_2 < \frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log(1+h_{22})$

Similar to the previous bounds, we start from Lemma 6.7 and write

$$\begin{aligned}
R_1 + R_2 &\leq I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\overset{(a)}{\leq} I(y_1^{'\ell}, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\leq I(y_1^{'\ell}, y_2^\ell; x_1^\ell, x_2^\ell, x_2^{'\ell}) + \ell\epsilon_\ell \\
&\overset{(b)}{=} I(y_1^{'\ell}; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_2^{'\ell}|x_1^\ell, x_2^\ell) \\
&\quad + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + I(y_2^\ell; x_1^\ell, x_2^\ell|y_1^{'\ell}, x_2^{'\ell}) + \ell\epsilon_\ell \\
&= I(y_1^{'\ell}; x_1^\ell, x_2^\ell) + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+g_{11}+g_{12}) + I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) + \ell\epsilon_\ell.
\end{aligned} \tag{6.22}$$

Note that in (a) we used the data processing inequality. An argument similar to that is used in the proof of (GZS-4) shows that the second and fourth terms in (b) are zero. Now, we have

$$\begin{aligned}
I(y_2^\ell; x_2^{'\ell}|y_1^{'\ell}) &= h(y_2^\ell|y_1^{'\ell}) - h(y_2^\ell|x_2^{'\ell}, y_1^{'\ell}) \\
&\leq h(y_2^\ell|x_1^{'\ell}) - h(y_2^\ell|x_1^{'\ell}, x_2^{'\ell}, y_1^{'\ell}) \\
&= h(\sqrt{h_{22}}x_2^{'\ell} + z_2^\ell|x_1^{'\ell}) - h(z_2^\ell|x_1^{'\ell}, x_2^{'\ell}, y_1^{'\ell}) \\
&\leq h(\sqrt{h_{22}}x_2^{'\ell} + z_2^\ell) - h(z_2^\ell) \\
&\leq \frac{\ell}{2}\log(1+h_{22})
\end{aligned} \tag{6.23}$$

Finally, we obtain the desired bound by replacing (6.23) in (6.22). It is worth mentioning that this bound is the same as the cut-set bound for the cut $\Omega_s = \{S_1, S_2, B\}$ and $\Omega_d = \{A, D_1, D_2\}$.

$\triangleright$ (GZS-7): $\quad R_1 < \frac{1}{2}\log(1+h_{11})$

Using Lemma 6.7 and the data processing inequality, we can write

$$\ell R_1 \leq I(x_1^{'\ell}; y_1^\ell) + \ell\epsilon_\ell \leq I(x_1^{'\ell}; y_1^\ell) + \ell\epsilon_\ell \leq \frac{\ell}{2}\log(1+h_{11}) + \ell\epsilon_\ell. \tag{6.24}$$

$\triangleright$ (GZS-8) $\quad R_2 < \frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}})$

Starting from Lemma 6.7 and applying the data processing inequality for the Markov chain $x_2^\ell \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow y_2^\ell$, we have

$$\begin{aligned}\ell R_2 &\leq I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell \\ &\leq I(x_1^{'\ell}, x_2^{'\ell}; y_2^\ell) + \ell\epsilon_\ell \\ &\leq \frac{\ell}{2}\log(1 + h_{21} + h_{22} + 2\sqrt{h_{21}h_{22}}) + \ell\epsilon_\ell. \end{aligned} \tag{6.25}$$

Note that $x_1^{'\ell}$ and $x_2^{'\ell}$ are not independent. However, their variance is upper bounded by $(\sqrt{h_{21}} + \sqrt{h_{22}})^2$.

▷ (GZS-9): $R_2 < \frac{1}{2}\log(1 + g_{22}) + \frac{1}{2}\log(1 + h_{21})$

Consider the cut which partitions the network into $\Omega_s = \{S_1, S_2, A, D_1\}$ and $\Omega_d = \{B, D_2\}$. We have

$$\begin{aligned}\ell R_2 &\leq I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell \\ &\leq I(x_1^{'\ell}, x_2^\ell; y_2^{'\ell}, y_2^\ell) + \ell\epsilon_\ell \\ &\overset{(a)}{=} I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^{'\ell}|x_2^\ell) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}) + I(x_2^\ell; y_2^\ell|y_2^{'\ell}, x_1^{'\ell}) + \ell\epsilon_\ell \\ &= I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}) + \ell\epsilon_\ell \\ &= I(x_2^\ell; y_2^{'\ell}) + I(x_1^{'\ell}; y_2^\ell|y_2^{'\ell}, x_2^{'\ell}) + \ell\epsilon_\ell \\ &\leq \frac{\ell}{2}\log(1 + g_{22}) + \frac{\ell}{2}\log(1 + h_{21}) + \ell\epsilon_\ell. \end{aligned} \tag{6.26}$$

We again used an argument similar to that is used in proof of (GZS-4) to show that the second and fourth terms in (a) are zero.

This completes the proof of the outer bound in Theorem 6.2.

6.4.2 The Achievability Part

In this section we provide an encoding scheme for the Gaussian ZS network, and show that the rate region that can be achieved using this scheme is only a constant bit gap away from the outer bound. We distinguish the following two cases: (i) *large channel gains* where all the links have gain at least 1, and (ii) *small channel gains*, wherein there exists at least one link in the GZS network with gain less than 1, In the first case, more sophisticated schemes schemes are required to (approximately) achieve the capacity of the network. However, in the second case, the network can be approximated by a simpler network.

Large Channel Gains

In this part, we assume that all channel gains are at least 1, *i.e.*, $g_{ij} \geq 1$, and $h_{ij} \geq 1$. Note that if any of the gains is small, then either one of the rates is small (of the order of our constant bit gap), or the cross links are negligible. We will discuss these cases later.

We first give an overview of the encoding/decoding strategies we use to achieve the capacity of the network, and then analyze the performance of the proposed scheme, to show that it is within constant bit gap of the outer bound.

Transmission Strategies The coding strategy for the Gaussian ZS network is essentially a partial-decode-and-forward strategy, along with a strategic rate-splitting of the messages. Let the messages to be sent from S_1, S_2 be denoted by W_1, W_2 respectively (see Figure 6.2). We will break the ZS network into two cascaded interference channels, where we require particular messages to be decoded at the relays and forwarded to the destinations. The first stage is a Z interference channel, where the message W_2 is split into three parts: $\left(U_2^{(1)}, U_2^{(2)}, U_2^{(3)}\right)$. The intention of this strategic split is to allow the the node G_1 (which is relay A in the original ZS network) to decode $\left(U_1^{(1)}, U_2^{(1)}, U_2^{(2)}\right)$ and node G_2 (which is relay B in the original ZS network), to decode $\left(U_2^{(1)}, U_2^{(3)}\right)$. This is illustrated in Figure 6.7. Here, $U_2^{(1)}$ plays the role of a common message which can be decoded at both receivers, whereas $U_2^{(2)}$ and $U_2^{(3)}$ are the private messages for G_1 and G_2 respectively.

The next stage of the ZS network is a S interference channel depicted in Figure 6.8. Here we take the messages delivered and decoded by the Z interference channel of the first stage and further process them to ensure delivery of the desired messages to the destination. In particular, we further split the decoded messages from the first stage into several parts and require delivery of messages as shown in Figure 6.8. This splitting and delivery of appropriate pieces, finally ensures that W_1 and W_2 are decodable at the destinations. This is the encoding strategy in the ZS network. In the following lemmas, we give the rates at which messages at each stage can be delivered. Putting these lemmas together, we get the desired result given in Theorem 6.2. This will be discussed in detail in the rest of this section.

A formal statement of the argument above is given below.

Definition 6.8. *Consider the Gaussian Z network shown in Figure 6.7, where the first transmitter, F_1, has a messages $U_1^{(1)}$ of rate $\Upsilon_{1,1}$ and the second transmitter, F_2, has three independent messages $U_2^{(1)}$, $U_2^{(2)}$, and $U_2^{(3)}$, of rates $\Upsilon_{2,1}$, $\Upsilon_{2,2}$, and $\Upsilon_{2,3}$, respectively. The transmitters wish to encode their messages and send them to the transmitters such that G_1 be able to decode U_1, $U_2^{(1)}$ and $U_2^{(2)}$ and G_2 can decode $U_2^{(1)}$ and $U_2^{(3)}$, with arbitrary small error probability.*

The following lemma gives an achievable rate region for the Gaussian Z network defined in Definition 6.8. We will present its proof in Appendix E.3.

Lemma 6.9. *Consider a Gaussian Z interference network with channel parameters (g_{11}, g_{12}, g_{22}), and decoding requirements as shown in Figure 6.7. Denoting the rate of the sub-message $U_i^{(j)}$ by $\Upsilon_{i,j}$, any rate tuple $(\Upsilon_{1,1}, \Upsilon_{2,1}, \Upsilon_{2,2}, \Upsilon_{2,3})$ which satisfies*

$$\Upsilon_{1,1} \leq \left(\frac{1}{2}\log\left(1+g_{11}\right)-\frac{1}{2}\right)^+, \tag{6.27}$$

$$\Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1+\frac{g_{12}}{g_{22}}\right)-\frac{1}{2}\right)^+, \tag{6.28}$$

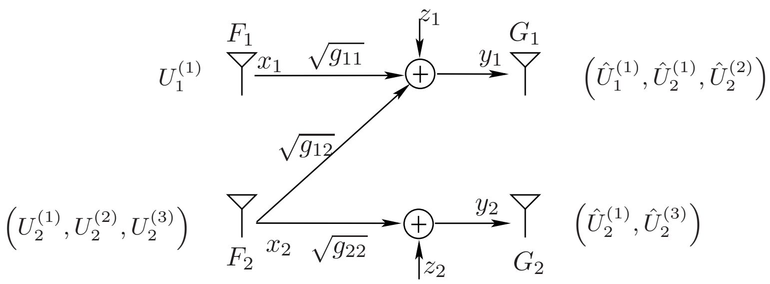


Figure 6.7: The Z interference channel with particular message requirements, captures the proposed coding scheme for the first layer of the Gaussian ZS network.

$$\Upsilon_{2,1} + \Upsilon_{2,2} \le \left(\frac{1}{2}\log(1+g_{12}) - \frac{1}{2}\right)^+, \tag{6.29}$$

$$\Upsilon_{1,1} + \Upsilon_{2,1} + \Upsilon_{2,2} \le \left(\frac{1}{2}\log(1+g_{11}+g_{12}) - \frac{1}{2}\right)^+, \tag{6.30}$$

$$\Upsilon_{2,3} \le \left(\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\right)^+, \tag{6.31}$$

$$\Upsilon_{2,1} + \Upsilon_{2,3} \le \left(\frac{1}{2}\log(1+g_{22}) - \frac{1}{2}\right)^+, \tag{6.32}$$

is achievable.

The next definition and corresponding lemma give an achievable rate region for the second layer of the ZS network, which is a S interference network depicted in Figure 6.8. Let us first formally define this interference network, with its message decodability requirements.

Definition 6.10. *Consider the Gaussian* S *network shown in Figure 6.8, where the first transmitter, F_1, has six independent messages $V_1^{(1)}$, $V_1^{(2)}$, $V_2^{(1)}$, $V_2^{(2)}$, $V_2^{(3)}$ and $V_2^{(4)}$ and the second transmitter has F_2 three independent messages $V_2^{(1)}$, $V_2^{(2)}$, and $V_2^{(5)}$. We denote by $Q_{i,j}$ the rate of message $V_i^{(j)}$. The transmitters wish to encode their messages and send them to the transmitters such that G_1 be able to decode $V_1^{(1)}$, $V_1^{(2)}$, $V_2^{(1)}$ and $V_2^{(3)}$ and G_2 can decode $V_1^{(1)}$, $V_2^{(1)}$, $V_2^{(2)}$, $V_2^{(3)}$, $V_2^{(4)}$ and $V_2^{(5)}$, with arbitrary small error probability.*

Lemma 6.11. *Consider the Gaussian* S *interference network with channel gains (h_{11}, h_{21}, h_{22}), and decoding requirements as shown in Figure 6.8, where $\Theta_{i,j}$ denotes the rate of message $V_i^{(j)}$. Any rate tuple $(\Theta_{1,1}, \Theta_{1,2}, \Theta_{2,1}, \Theta_{2,2}, \Theta_{2,3}, \Theta_{2,4}, \Theta_{2,5})$ which satisfies*

$$\Theta_{1,1} + \Theta_{1,2} + \Theta_{2,1} + \Theta_{2,3} \le \left(\frac{1}{2}\log(1+h_{11}) - \frac{1}{2}\right)^+, \tag{6.33}$$

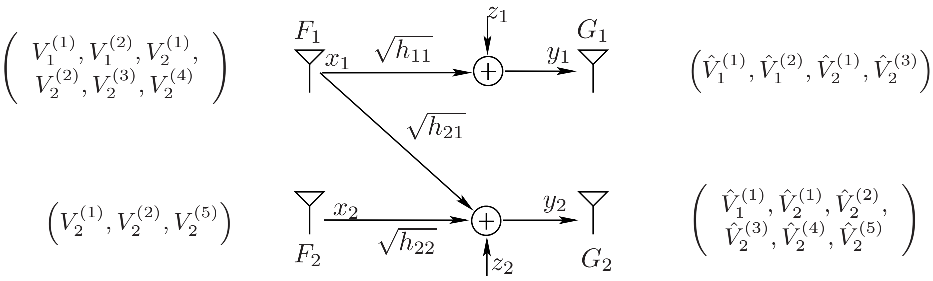


Figure 6.8: The S interference channel with particular message requirements, depicting the proposed coding strategy for the second layer of the Gaussian ZS network.

$$\Theta_{1,2} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{11}}{h_{12}}\right)-\frac{1}{2}\right)^+, \quad (6.34)$$

$$\Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{21}}{h_{11}}\right)-\frac{1}{2}\right)^+, \quad (6.35)$$

$$\Theta_{1,1}+\Theta_{2,3}+\Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+h_{21}\right)-\frac{1}{2}\right)^+, \quad (6.36)$$

$$\Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1+h_{22}\right)-\frac{1}{2}\right)^+, \quad (6.37)$$

$$\Theta_{1,1}+\Theta_{2,1}+\Theta_{2,2}+\Theta_{2,3}+\Theta_{2,4}+\Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1+h_{21}+h_{22}\right)-\frac{1}{2}\right)^+, \quad (6.38)$$

is achievable.

We present the proof of this lemma in Appendix E.4.

Performance Analysis The encoding scheme proposed for the Gaussian ZS network consists of two separate parts. We first split the message of the second source nodes as $W_1 = U_1^{(1)}$ and $W_2 = (U_2^{(1)}, U_2^{(2)}, U_2^{(3)})$, where $U_2^{(1)}$ can be decoded at both relay nodes A and B, and $U_2^{(2)}$ and $U_2^{(3)}$ can be decoded only at A and B, respectively (see Figure 6.7). Denoting the rate of message $W_i^{(j)}$ by $\Upsilon_{i,j}$, the following rate constraints are imposed by this message splitting

$$R_1 = \Upsilon_{1,1}, \quad (6.39)$$

$$R_2 = \Upsilon_{2,1}+\Upsilon_{2,2}+\Upsilon_{2,3}. \quad (6.40)$$

An achievable rate region for this message splitting is given in Lemma 6.9.

In the second layer of the network (see Figure 6.8), relay node A further splits its messages as follows: $W_1 = U_1^{(1)} = \left(V_1^{(1)}, V_1^{(2)}\right)$, $U_2^{(1)} = \left(V_2^{(1)}, V_2^{(2)}\right)$, and $U_2^{(2)} = \left(V_2^{(3)}, V_2^{(4)}\right)$. A similar message splitting is also performed at node B to obtain $U_2^{(1)} = \left(V_2^{(1)}, V_2^{(2)}\right)$ and $U_2^{(3)} = V_2^{(5)}$. This message splitting imposes the following rate equations

$$\Upsilon_{1,1} = \Theta_{1,1} + \Theta_{1,2}, \tag{6.41}$$
$$\Upsilon_{2,1} = \Theta_{2,1} + \Theta_{2,2}, \tag{6.42}$$
$$\Upsilon_{2,2} = \Theta_{2,3} + \Theta_{2,4}, \tag{6.43}$$
$$\Upsilon_{2,3} = \Theta_{2,5}, \tag{6.44}$$

where $\Theta_{i,j}$ denotes the rate of the message $V_i^{(j)}$. Next, the relay nodes have to convey the messages to the destination nodes such that D_1 can decode $V_1^{(1)}$, $V_1^{(2)}$, $V_2^{(1)}$ and $V_2^{(3)}$, and D_2 be able to decode $V_1^{(1)}$, $V_2^{(1)}$, $V_2^{(2)}$, $V_2^{(3)}$, $V_2^{(4)}$ and $V_2^{(5)}$. An achievable rate region for this transmission scenario is given in Lemma 6.11.

Putting the rate constraints in Lemma 6.9 and Lemma 6.11 together with the equations in (6.39)-(6.40) and (6.41)-(6.44), we obtain the following achievable rate region for the Gaussian ZS network.

$$\begin{aligned}
\overline{\mathcal{R}}_{\mathsf{GZS}}^{(1)} = \Big\{(R_1, R_2) : \exists \Upsilon_{1,1}, \Upsilon_{2,1}, \Upsilon_{2,2}, \Upsilon_{2,3},\\
&\Theta_{1,1}, \Theta_{1,2}, \Theta_{2,1}, \Theta_{2,2}, \Theta_{2,3}, \Theta_{2,4}, \Theta_{2,5} \geq 0,\\
&R_1 = \Upsilon_{1,1},\\
&R_2 = \Upsilon_{2,1} + \Upsilon_{2,2} + \Upsilon_{2,3},\\
&\Upsilon_{1,1} = \Theta_{1,1} + \Theta_{1,2},\\
&\Upsilon_{2,1} = \Theta_{2,1} + \Theta_{2,2},\\
&\Upsilon_{2,2} = \Theta_{2,3} + \Theta_{2,4},\\
&\Upsilon_{2,3} = \Theta_{2,5},\\
&\Upsilon_{1,1} \leq \left(\frac{1}{2}\log\left(1 + g_{11}\right) - \frac{1}{2}\right)^+,\\
&\Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1 + \frac{g_{12}}{g_{22}}\right) - \frac{1}{2}\right)^+,\\
&\Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1 + g_{12}\right) - \frac{1}{2}\right)^+,\\
&\Upsilon_{1,1} + \Upsilon_{2,1} + \Upsilon_{2,2} \leq \left(\frac{1}{2}\log\left(1 + g_{11} + g_{12}\right) - \frac{1}{2}\right)^+,\\
&\Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\right)^+,
\end{aligned}$$

$$
\begin{aligned}
&\Upsilon_{2,1} + \Upsilon_{2,3} \leq \left(\frac{1}{2}\log\left(1+g_{22}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{1,1} + \Theta_{1,2} + \Theta_{2,1} + \Theta_{2,3} \leq \left(\frac{1}{2}\log\left(1+h_{11}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{1,2} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{11}}{h_{12}}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+\frac{h_{21}}{h_{11}}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{1,1} + \Theta_{2,3} + \Theta_{2,4} \leq \left(\frac{1}{2}\log\left(1+h_{21}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{2,5} \leq \left(\frac{1}{2}\log\left(1+h_{22}\right) - \frac{1}{2}\right)^{+}, \\
&\Theta_{1,1} + \Theta_{2,1} + \Theta_{2,2} + \Theta_{2,3} + \Theta_{2,4} + \Theta_{2,5} \\
&\qquad\qquad \leq \left(\frac{1}{2}\log\left(1+h_{21}+h_{22}\right) - \frac{1}{2}\right)^{+} \Bigg\}.
\end{aligned}
$$

We apply the Fourier-Motzkin elimination on this region, to project it on the coordinated R_1 and R_2, and obtain the following rate region. After some simplifications, we get:

$$
\begin{aligned}
\overline{\mathcal{R}}_{\mathsf{GZS}}^{(1)} = \Bigg\{(R_1,R_2): &R_1 \leq \left(\frac{1}{2}\log\left(g_{11}\right) - \frac{1}{2}\right)^{+}, \\
&R_2 \leq \left(\frac{1}{2}\log\left(g_{12}+g_{22}\right) - \frac{1}{2}\right)^{+}, \\
&R_1 + R_2 \leq \left(\frac{1}{2}\log\left(g_{11}+g_{12}\right) + \frac{1}{2}\log\left(\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\right)^{+}, \\
&R_2 \leq \left(\frac{1}{2}\log\left(g_{12}\right) + \frac{1}{2}\log\left(h_{22}\right) - \frac{1}{2}\right)^{+}, \\
&R_1 + R_2 \leq \left(\frac{1}{2}\log\left(g_{22}\right) + \frac{1}{2}\log\left(h_{11}+h_{21}\right) - \frac{1}{2}\right)^{+}, \\
&R_1 + R_2 \leq \left(\frac{1}{2}\log\left(g_{11}+g_{12}\right) + \frac{1}{2}\log\left(h_{22}\right) - \frac{1}{2}\right)^{+}, \\
&R_1 \leq \left(\frac{1}{2}\log\left(h_{11}\right) - \frac{1}{2}\right)^{+}, \\
&R_2 \leq \left(\frac{1}{2}\log\left(h_{21}+h_{22}\right) - \frac{1}{2}\right)^{+}, \\
&R_2 \leq \left(\frac{1}{2}\log\left(g_{22}\right) + \frac{1}{2}\log\left(h_{21}\right) - \frac{1}{2}\right)^{+},
\end{aligned}
$$

$$R_1 + R_2 \le \left(\frac{1}{2}\log\left(h_{21} + h_{22}\right) + \frac{1}{2}\log\left(\frac{h_{11}}{h_{21}}\right) - \frac{1}{2}\right)^{+} \Big\}.$$

Note that this rate region is characterized by a set of constraints which are similar to the inequalities in the definition of $\underline{\mathcal{R}}_{\mathsf{GZS}}$, except for the additive constants, and the fact that $\log(1+x)$ is replaced by $\log(x)$. Note that since $x \ge 1$, we have

$$\frac{1}{2}\log(1+x) - \frac{1}{2}\log(x) \le \frac{1}{2}. \tag{6.45}$$

Hence, the difference between the RHS's of two sets of inequalities do not exceed 1 for R_1, and 3/2 for R_2 and $R_1 + R_2$. Therefore, for any rate pair $(R_1, R_2) \in \underline{\mathcal{R}}_{\mathsf{GZS}}$, we have $(R_1 - 1, R_2 - 1.5) \in \overline{\mathcal{R}}^{(1)}_{\mathsf{GZS}}$. Therefore, we have $\overline{\mathcal{R}}_{\mathsf{GZS}} \subseteq \overline{\mathcal{R}}^{(1)}_{\mathsf{GZS}} \subseteq \mathcal{R}_{\mathsf{GZS}}$. This completes the proof.

Small Channel Gains

We will show in this part that if any of the channel gains in the GZS network is small, then the outer bound in Theorem 6.2 is still within a constant bit gap of an achievable rate region. This argument is based on the analysis of a modified version of the network, in which all the links with gain smaller than 1 are removed. One can show that the capacity region of this modified network is within a constant gap from that of the original one. On the other hand, we can argue that the gap between the achievable rate pairs of the modified network and the outer bound in Theorem 6.2 is bounded by a constant. Therefore, we can conclude that if $(R_1, R_2) \in \mathcal{R}_{\mathsf{GZS}}$ then $(R_1 - \delta_1, R_2 - \delta_2)$ is achievable for the original network, where $\delta_1 = 1$ and $\delta_2 = 1.5$.

The main intuition behind this argument is the fact that since all the nodes are assumed to have power constraint equal to 1, the flow of information through a link with gain not exceeding 1 is upper bounded by $\frac{1}{2}\log(1+\mathsf{SNR}) \le \frac{1}{2}\log(1+1) = \frac{1}{2}$ bit. Therefore, by removing such links from the network, the achievable rates change by at most $\frac{1}{2}$ bit. On the other hand, the incoming signals over small channel gains may act as an interference on the original network, which cause a total noise power not exceeding 1. Therefore, by doubling the noise variances in the original network, we guarantee that capacity region of the modified network is always smaller than that of the original one.

The advantage of analyzing the modified network instead of the original one is that some of the links are removed in the modified network, which convert it to simpler network to analyze.

A precise analysis of the modified networks requires considering several cases separately. However, similar techniques and ideas will be used for all cases. In the following we present one illustrating example, and skip the details for the other cases.

Example 6.12. *Consider the Gaussian* ZS *network in Figure 6.2, and assume that $g_{12} = 0$. Therefore, the first layer of the network would be two parallel*

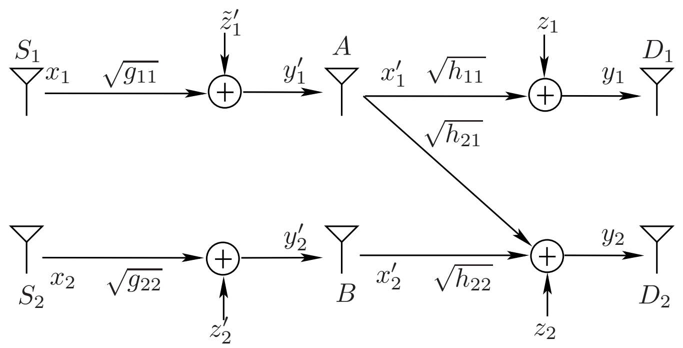


Figure 6.9: A modified GZS network obtained assuming $g_{12} = 0$.

links as shown in Figure 6.9, where $\mathbb{E}[\tilde{z}_1'^2] = 2$. *Moreover, the rate region in* (GZS-1)-(GZS-10) *will be reduced to*

$$R_1 \le \frac{1}{2}\log(1+g_{11}) \tag{6.46}$$

$$R_2 \le \frac{1}{2}\log(1+g_{12}) \tag{6.47}$$

$$R_1 \le \frac{1}{2}\log(1+h_{11}) \tag{6.48}$$

$$R_2 \le \frac{1}{2}\log(1+h_{22}) \tag{6.49}$$

$$R_1 + R_2 \le \frac{1}{2}\log(1+h_{21}+h_{22}+2\sqrt{h_{21}h_{22}}) + \frac{1}{2}\log\left(1+\frac{h_{11}}{h_{21}}\right). \tag{6.50}$$

The encoding strategy for this network is fairly simple. Let (R_1, R_2) *be a rate pair satisfying* (6.46)-(6.50). *The goal is to show that* $(R_1 - 1, R_2 - 1)$ *is achievable. Since* (R_1-1, R_2-1) *satisfies* (6.46) *and* (6.47), *transmission over the first layer of the network from the source nodes to the relays is simply done using random Gaussian codes. The relays decode the messages and forward them through the second layer of the network. The encoding at* A *is performed by splitting the message into two parts* $W_1 = \left(W_1^{(1)}, W_1^{(2)}\right)$, *where* $W_1^{(1)}$ *can be decoded at both* D_1 *and* D_2, *but* $W_1^{(2)}$ *is decodable only at* D_1. *Note that the region characterized by* (6.48)-(6.50) *is a two-dimensional polytopes with two corner-points. It is easy to show that each corner-point can be approximately achieved using a proper power allocation for the sub-message sent by* A.

6.5 The Gaussian ZZ Network

We will present the proof of Theorem 6.4 in this section. We start by showing that $\underline{\mathcal{R}}_{\mathsf{GZZ}}$ establishes an outer bound for the achievable rate region. We then present our encoding scheme which guarantees achieving any rate pair in $\overline{\mathcal{R}}_{\mathsf{GZZ}}$.

6.5.1 The Outer Bound

In the following we present the proof for each of the inequalities in (GZZ-1)-(GZZ-6), separately. We again present the Gaussian ZZ network in Figure 6.10, to clarify the notation used in the proof. In particular, we use two variables, which are the noisy signals received at A and D_1 through the cross links assuming the direct links were absent, namely,

$$\gamma_1 = \sqrt{g_{12}}x_2 + z_1',$$
$$\gamma_2 = \sqrt{h_{12}}x_2' + z_1.$$

Note that $y_1' = \sqrt{g_{11}}x_1 + \gamma_1$ and $y_1 = \sqrt{h_{11}}x_1' + \gamma_2$.

Suppose that the rate pair (R_1, R_2) is achieved with a small decoding error probability ϵ_ℓ using a code of length ℓ. The following chains of inequalities provide upper bounds on the individual rates as well as the sum-rate. We again use Lemma 6.7, which essentially captures the decodability requirements of the network.

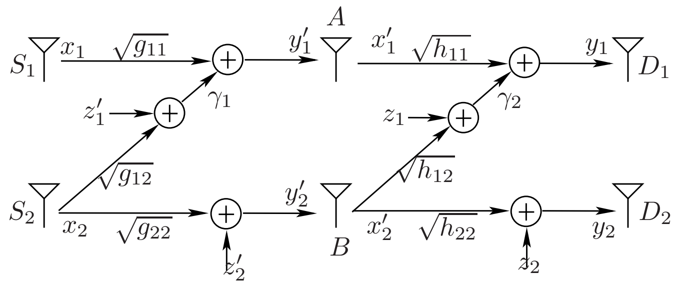


Figure 6.10: The Gaussian ZZ network.

The individual rate bounds in (GZZ-1)-(GZZ-4) have the same structure as the cut-set bound, although we derive them through a slightly different argument. However, the two sum-rate bounds in (GZZ-5) and (GZZ-6) are conceptually different than the cut-set bounds. These two bounds which are tighter than cut-set bounds are derived through a genie-aided argument; that is, we assume that the signal sent over the cross link of one layer is given by a genie to the receiver of the other layer (relay node A in layer 1 and destination node D_1 in layer 2). Therefore, we present the proofs of (GZZ-5) and (GZZ-6) first. The more standard cut-set type bounds are provided later for completeness.

a) The proof of the genie-aided bounds

$\triangleright$ (GZZ-5): $R_1+R_2 \leq \frac{1}{2}\log(1+g_{11}+g_{12})+\frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right)+\frac{1}{2}\log(1+h_{12})$

We start with the sum-rate inequality in Lemma 6.7, and write

$$\begin{aligned}\ell(R_1+R_2) &\leq I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\ &\leq I(y_1'^{\ell}, y_2'^{\ell}; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell\end{aligned}$$

$$
\begin{aligned}
&\leq I(y_1^{'\ell}, y_2^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_2^{'\ell}; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + \ell\epsilon_\ell \\
&\leq I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_2^{'\ell}, \gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + \ell\epsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&\quad + I(y_2^{'\ell}; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) + \ell\epsilon_\ell \\
&= I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) + I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) + I(y_2^{'\ell}; x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\
&\quad + I(y_2^{'\ell}; x_1^\ell | x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) + \ell\epsilon_\ell. \qquad (6.51)
\end{aligned}
$$

Each of the terms in (6.51) can be bounded as follows. In order to bound the first term, we can simply write

$$
\begin{aligned}
I(y_1^{'\ell}, \gamma_2^\ell; x_1^\ell, x_2^\ell) &= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell | \gamma_2^\ell) \\
&= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + h(y_1^{'\ell} | \gamma_2^\ell) - h(y_1^{'\ell} | x_1^\ell, x_2^\ell, \gamma_2^\ell) \\
&\overset{(a)}{\leq} I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + h(y_1^{'\ell}) - h(y_1^{'\ell} | x_1^\ell, x_2^\ell) \\
&= I(\gamma_2^\ell; x_1^\ell, x_2^\ell) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell) \\
&\overset{(b)}{\leq} I(\gamma_2^\ell; x_2^{'\ell}) + I(y_1^{'\ell}; x_1^\ell, x_2^\ell) \\
&= \frac{\ell}{2}\log(1 + h_{12}) + \frac{\ell}{2}\log(1 + g_{11} + g_{12}), \qquad (6.52)
\end{aligned}
$$

where in (a) we have used the fact that conditioning decreases the entropy, and the Markov chain $\gamma_2^\ell \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^{'\ell} \leftrightarrow (x_1^\ell, x_2^\ell) \leftrightarrow y_1^{'\ell}$. Also (b) follows from the same Markov chain.

The second term can be bounded as

$$
\begin{aligned}
I(\gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) &= I(y_1^{'\ell} - \gamma_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&= I(\sqrt{g_{11}} x_1^\ell; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&\leq I(W_1; x_1^\ell, x_2^\ell | y_1^{'\ell}, \gamma_2^\ell) \\
&\leq H(W_1 | y_1^{'\ell}, \gamma_2^\ell) \\
&\overset{(c)}{\leq} H(W_1 | y_1^\ell) \leq \ell\epsilon_\ell, \qquad (6.53)
\end{aligned}
$$

where (c) holds since $y_1^\ell = \sqrt{h_{11}} x_1^{'\ell} + \gamma_2^\ell = f_1(y_1^{'\ell}) + \gamma_2^\ell = f_2(y_1^{'\ell}, \gamma_2^\ell)$.

In order to bound the third term in (6.51) we can write

$$
\begin{aligned}
I(y_2^{'\ell}; x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) &= h(x_2^\ell | y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) - h(x_2^\ell | y_1^{'\ell}, y_2^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\
&\overset{(d)}{\leq} h(x_2^\ell | \gamma_1^\ell) - h(x_2^\ell | y_1^{'\ell}, y_2^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\
&= h(x_2^\ell | \gamma_1^\ell) - [h(x_2^\ell, t_1^\ell | y_1^{'\ell}, y_2^{'\ell}, \gamma_2^\ell) - h(t_1^\ell | y_1^{'\ell}, y_2^{'\ell}, \gamma_2^\ell)] \\
&\overset{(e)}{=} h(x_2^\ell | \gamma_1^\ell) - [h(x_2^\ell, t_1^\ell | y_1^{'\ell}, y_2^{'\ell}) - h(t_1^\ell | y_1^{'\ell}, y_2^{'\ell})]
\end{aligned}
$$

$$
\begin{aligned}
&= h(x_2^\ell|\gamma_1^\ell) - h(x_2^\ell|y_1^{'\ell}, y_2^{'\ell}, \gamma_1^\ell) \\
&\overset{(f)}{=} h(x_2^\ell|\gamma_1^\ell) - h(x_2^\ell|x_1^\ell, y_2^{'\ell}, \gamma_1^\ell) \\
&= h(x_2^\ell|\gamma_1^\ell) - [h(x_2^\ell|y_2^{'\ell}, \gamma_1^\ell) + h(x_1^\ell|x_2^\ell, y_2^{'\ell}, \gamma_1^\ell) - h(x_1^\ell|y_2^{'\ell}, \gamma_1^\ell)] \\
&\overset{(g)}{=} h(x_2^\ell|\gamma_1^\ell) - h(x_2^\ell|y_2^{'\ell}, \gamma_1^\ell) \\
&= I(y_2^{'\ell}; x_2^\ell|\gamma_1^\ell) \\
&= h(y_2^{'\ell}|\gamma_1^\ell) - h(y_2^{'\ell}|\gamma_1^\ell, x_2^\ell) \\
&\overset{(h)}{=} h(y_2^{'\ell}|\gamma_1^\ell) - h(y_2^{'\ell}|x_2^\ell) \\
&= h(y_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}\gamma_1^\ell|\gamma_1^\ell) - h(y_2^{'\ell} - \sqrt{g_{22}}x_2^\ell|x_2^\ell) \\
&= h(z_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}z_1^{'\ell}|\gamma_1^\ell) - h(z_2^{'\ell}|x_2^\ell) \\
&\overset{(i)}{\leq} h(z_2^{'\ell} - \sqrt{\frac{g_{22}}{g_{12}}}z_1^{'\ell}) - h(z_2^{'\ell}) \\
&= \frac{\ell}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right),
\end{aligned} \tag{6.54}
$$

where in (d) we have used the fact that conditioning reduces the differential entropy, and (e) holds due to the Markov chain $(x_2^\ell, t_1^\ell) \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow \gamma_2^\ell$. Then in (f) we replaced $(y_1^{'\ell}, \gamma_1^\ell)$ by $(x_1^\ell, \gamma_1^\ell)$ since there is an one-to-one map, $y_1^{'\ell} = \sqrt{g_{11}}x_1^\ell + t_1^\ell$, between these joint variables, and in (g) we used the fact that x_1^ℓ is independent of $(x_2^\ell, y_2^{'\ell}, t_1^\ell)$ to conclude $h(x_1^\ell|x_2^\ell, y_2^{'\ell}, \gamma_1^\ell) = h(x_1^\ell|y_2^{'\ell}, \gamma_1^\ell) = h(x)$. Also (h) holds due to the Markov chain $y_2^{'\ell} \leftrightarrow x_2^\ell \leftrightarrow \gamma_1^\ell$. Finally, (i) is true due to removing conditioning and the fact that z_2' is independent of x_2.

The last term in (6.51) is negligible since

$$
\begin{aligned}
I(y_2^{'\ell}; x_1^\ell|x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) &\leq I(y_2^{'\ell}; W_1|x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\
&\leq H(W_1|x_2^\ell, y_1^{'\ell}, \gamma_1^\ell, \gamma_2^\ell) \\
&\leq H(W_1|y_1^{'\ell}, \gamma_2^\ell) \\
&\overset{(j)}{\leq} H(W_1|y_1^\ell) \\
&\overset{(k)}{\leq} \ell\epsilon_\ell,
\end{aligned} \tag{6.55}
$$

where (j) is due to the fact that $y_1^\ell = \sqrt{h_{11}}x_1^{'\ell} + \gamma_2^\ell = f_1(y_1^{'\ell}) + \gamma_2^\ell = f_2(y_1^{'\ell}, \gamma_2^\ell)$ is a function of $(y_1^{'\ell}, \gamma_2^\ell)$, and (k) follows the Fano's inequality.

Replacing (6.52)–(6.55) in (6.51), we get the desires bound.

▷ (GZZ-6): $R_1 + R_2 \leq \frac{1}{2}\log(1 + h_{11} + h_{12}) + \frac{1}{2}\log\left(1 + \frac{h_{22}}{h_{12}}\right) + \frac{1}{2}\log(1 + g_{12})$

Before proving this inequality, we present a lemma which will be used in this proof. We will present the proof of this lemma later in Appendix E.5.

Lemma 6.13. *Let X_1 and X_2 be two (arbitrarily correlated) random variables with variance constraints $\mathbb{E}[X_1^2] = \sigma_1^2$ and $\mathbb{E}[X_2^2] = \sigma_2^2$, which form a Markov chain $X_1 \leftrightarrow \Gamma \leftrightarrow X_2$ for some random variable Γ. Also assume that Z is a zero-mean unit variance Gaussian random variable independent of X_1, X_2 and Γ. Then the conditional differential entropy of $Y = X_1 + X_2 + Z$ is upper bounded by*

$$h(Y|\Gamma) \leq \frac{1}{2}\log 2\pi e(1 + \sigma_1^2 + \sigma_2^2). \tag{6.56}$$

Now, in order to prove (GZZ-5), we start with Lemma 6.7.

$$\begin{aligned}
\ell(R_1 + R_2) &\leq I(y_1^\ell, y_2^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\leq I(y_1^\ell, y_2^\ell, \gamma_1^\ell; x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&= I(y_1^\ell, \gamma_1^\ell; x_1^\ell, x_2^\ell) + I(y_2^\ell; x_1^\ell, x_2^\ell | y_1^\ell, \gamma_1^\ell) + \ell\epsilon_\ell \\
&= I(\gamma_1^\ell; x_1^\ell, x_2^\ell) + I(y_1^\ell; x_1^\ell, x_2^\ell | \gamma_1^\ell) + I(y_2^\ell; x_1^\ell, x_2^\ell | y_1^\ell, \gamma_1^\ell) + \ell\epsilon_\ell.
\end{aligned} \tag{6.57}$$

Since γ_1^ℓ is independent of x_1^ℓ, the first term can be simply bounded as

$$\begin{aligned}
I(\gamma_1^\ell; x_1^\ell, x_2^\ell) &= I(\gamma_1^\ell; x_2^\ell) + I(t^\ell; x_1^\ell | x_2^\ell) \\
&= I(\gamma_1^\ell; x_2^\ell) + I(z_1^{'\ell}; x_1^\ell | x_2^\ell) \leq \frac{\ell}{2}\log(1 + g_{12}).
\end{aligned} \tag{6.58}$$

For the second term we can write

$$\begin{aligned}
I(y_1^\ell; x_1^\ell, x_2^\ell | \gamma_1^\ell) &= h(y_1^\ell | \gamma_1^\ell) - h(y_1^\ell | x_1^\ell, x_2^\ell, \gamma_1^\ell) \\
&\overset{(a)}{\leq} h(y_1^\ell | \gamma_1^\ell) - h(y_1^\ell | x_1^{'\ell}, x_2^{'\ell}) \\
&= h(y_1^\ell | \gamma_1^\ell) - h(z_1^\ell | x_1^{'\ell}, x_2^{'\ell}) \\
&= h(\sqrt{h_{11}} x_1^{'\ell} + \sqrt{h_{12}} x_2^{'\ell} + z_1^\ell | \gamma_1^\ell) - h(z_1^\ell) \\
&\overset{(b)}{\leq} \frac{\ell}{2}\log\Big(2\pi e(1 + h_{11} + h_{12})\Big) - \frac{\ell}{2}\log 2\pi e \\
&= \frac{\ell}{2}\log(1 + h_{11} + h_{12}),
\end{aligned} \tag{6.59}$$

where (a) follows from the Markov chain $y_1^\ell \leftrightarrow (x_1^{'\ell}, x_2^{'\ell}) \leftrightarrow (x_1^\ell, x_2^\ell, \gamma_1^\ell)$. In (b) we have used Lemma 6.13 for t_1^ℓ, $x_1^{'\ell}$ and $x_2^{'\ell}$ which form a Markov chain, since

$$\begin{aligned}
I(x_1^{'\ell}; x_2^{'\ell} | t_1^\ell) &\leq I(y_1^{'\ell}; y_2^{'\ell} | t_1^\ell) \\
&= I\Big(\frac{y_1^{'\ell} - t_1^\ell}{\sqrt{g_{11}}}; y_2^{'\ell} | t_1^\ell\Big) \\
&= I(x_1^\ell; y_2^{'\ell} | t_1^\ell)
\end{aligned}$$

$$= h(x_1^\ell|t_1^\ell) - h(x_1^\ell|t_1^\ell, y_2^{'\ell}) = 0.$$

The third term can be further upper bounded by

$$\begin{aligned}
I(y_2^\ell; x_1^\ell, x_2^\ell|y_1^\ell, \gamma_1^\ell) &= h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_1^\ell, x_2^\ell, y_1^\ell, \gamma_1^\ell) \\
&\overset{(c)}{\le} h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_2^{'\ell}) \\
&\overset{(d)}{\le} h(y_2^\ell|y_1^\ell, \gamma_1^\ell) - h(y_2^\ell|x_2^{'\ell}, x_1^{'\ell}, y_1^\ell, \gamma_1^\ell) \\
&= I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le I(y_2^\ell, \gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&= I(\gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) + I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell), \qquad (6.60)
\end{aligned}$$

where both (c) and (d) follow from the Markov chain $y_2^\ell \leftrightarrow x_2^{'\ell} \leftrightarrow (x_1^\ell, y_1^\ell, \gamma_1^\ell)$. Now, we have

$$\begin{aligned}
I(\gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) &= I(y_1^\ell - \gamma_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&= I(\sqrt{h_{11}}x_1^{'\ell}; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\overset{(e)}{\le} I(y_1^{'\ell}; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le I(y_1^{'\ell} - t_1^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&= I(\sqrt{g_{11}}x_1^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le I(W_1; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell) \\
&\le H(W_1|y_1^\ell) \le \ell\epsilon_\ell, \qquad (6.61)
\end{aligned}$$

where (e) follows from the fact that $x_1^{'\ell}$ is a function of $y_1^{'\ell}$. Finally,

$$\begin{aligned}
I(y_2^\ell; x_1^{'\ell}, x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) &= h(y_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell) - h(y_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&= h\Big(y_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}\gamma_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell\Big) - h(y_2^\ell - \sqrt{h_{22}}x_2^{'\ell}|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&= h\Big(z_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}z_1^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell\Big) - h(z_2^\ell|y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell}) \\
&\overset{(f)}{\le} h\Big(z_2^\ell - \sqrt{\frac{h_{22}}{h_{12}}}z_1^\ell\Big) - h(z_2^\ell) \\
&= \frac{\ell}{2}\log\left(1 + \frac{h_{22}}{h_{12}}\right). \qquad (6.62)
\end{aligned}$$

Here, in (f) we have used the fact that conditioning decreases the differential entropy, and the fact that z_2^ℓ is independent of $(y_1^\ell, \gamma_1^\ell, \gamma_2^\ell, x_1^{'\ell}, x_2^{'\ell})$. Replacing (6.58)– (6.62) in(6.57), we will obtain the desired inequality.

b) The proofs of cut-set type bounds

$\triangleright$ (GZZ-1): $R_1 \leq \frac{1}{2}\log(1+g_{11})$

The individual rate bound can be simply obtained from

$$\begin{aligned}
\ell R_1 &= I(x_1^\ell; y_1^\ell) + \ell\epsilon_\ell \\
&\leq I(x_1^\ell; y_1^{'\ell}, y_2^{'\ell}) + \ell\epsilon_\ell && (6.63)\\
&= I(x_1^\ell; y_1^{'\ell}|y_2^{'\ell}) + I(x_1^\ell; y_2^{'\ell}) + \ell\epsilon_\ell \\
&= h(y_1^{'\ell}|y_2^{'\ell}) - h(y_1^{'\ell}|x_1^\ell, y_2^{'\ell}) && (6.64)\\
&\leq h(y_1^{'\ell}|x_2^\ell) - h(z_1^{'\ell}|x_1^\ell, y_2^{'\ell}) \\
&= h(\sqrt{g_{11}}x_1^\ell + z_1^{'\ell}) - h(z_1^{'\ell}) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+g_{11}) + \ell\epsilon_\ell,
\end{aligned}$$

where (6.63) follows from the data processing inequality for the Markov chain $x_1^\ell \leftrightarrow (y_1^{'\ell}, y_2^{'\ell}) \leftrightarrow y_1^\ell$, and (6.64) follows from the Markov chain $y_1^{'\ell} \leftrightarrow x_2^\ell \leftrightarrow y_2^{'\ell}$). Note that $\epsilon_\ell \to 0$ as ℓ grows. It is worth mentioning that this bound is similar to the cut-set bound for the cut $\Omega_s = \{S_1\}$ and $\Omega_d = \{S_2, A, B, D_1, D_2\}$.

$\triangleright$ (GZZ-2): $R_2 \leq \frac{1}{2}\log(1+g_{22})$

For the second rate bound, we can start with Lemma 6.7 and write

$$\ell R_2 \leq I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell \leq I(x_2^\ell; y_2^{'\ell}) + \ell\epsilon_\ell \leq \frac{\ell}{2}\log(1+g_{22}) + \ell\epsilon_\ell,$$

where we have used the data processing inequality and the Markov chain $x_2^\ell \leftrightarrow y_2^{'\ell} \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^\ell$ in the second inequality. Note that this bound captures the maximum flow of information through the cut specified by $\Omega_s = \{S_2\}$ and $\Omega_d = \{S_1, A, B, D_1, D_2\}$.

$\triangleright$ (GZZ-3): $R_1 \leq \frac{1}{2}\log(1+h_{11})$ In order to prove this upper bound, we use the cut-set bound for the cut $\Omega_s = \{S_1, S_2, A, B, D_2\}$ and $\Omega_d = \{D_1\}$.

$$\begin{aligned}
\ell R_1 &\leq I(x_1^{'\ell}; y_1^\ell|x_2^{'\ell}) + \ell\epsilon_\ell \\
&= h(y_1^\ell|x_2^{'\ell}) - h(y_1^\ell|x_1^{'\ell}, x_2^{'\ell}) + \ell\epsilon_\ell \\
&= h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell|x_2^{'\ell}) - h(z_1^\ell|x_1^{'\ell}, x_2^{'\ell}) + \ell\epsilon_\ell \\
&\leq h(\sqrt{h_{11}}x_1^{'\ell} + z_1^\ell) - h(z_1^\ell) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2}\log(1+h_{11}) + \ell\epsilon_\ell.
\end{aligned}$$

$\triangleright$ (GZZ-4): $R_2 \leq \frac{1}{2}\log(1+h_{22})$ Starting from Lemma 6.7, we can write

$$\ell R_2 \leq I(x_2^{'\ell}; y_2^\ell) + \ell\epsilon_\ell \leq I(x_2^{'\ell}; y_2^\ell) + \ell\epsilon_\ell \leq \frac{1}{2}\log(1+h_{22}) + \ell\epsilon_\ell,$$

where the second inequality follows from the data processing inequality for the Markov chain $x_2^\ell \leftrightarrow y_2^{'\ell} \leftrightarrow x_2^{'\ell} \leftrightarrow y_2^\ell$.

This shows that the rate region $\underline{\mathcal{R}}_{\mathsf{GZZ}}$ is an outer bound for the achievable region region of the Gaussian ZZ network.

6.5.2 The Achievability Part

In this section we present an encoding/decoding scheme, and derive an achieve rate region for this strategy. We then show that the gap between the boundary of this achievable rate region and that of $\mathcal{R}_{\mathsf{GZZ}}$ is upper bounded by a constant.

Like in the Gaussian ZS network, we only consider the large channel gain case, where we assume that all the channel gains are lower bounded by 1. A similar argument to that we used for the ZS network shows that for small channel gain cases the network is reduced to a simple one and its achievability and gap analysis are fairly simple.

The encoding scheme needed for the ZZ network is more sophisticated than the ZS network. An additional component to strategic message splitting is that of interference neutralization. We used this technique to achieve the capacity of the deterministic ZZ network as illustrated in Example 5.8 in Section 5.3.

This technique can be applied to the Gaussian network as discussed in Example 6.6. This along with message splitting inspired by the network decomposition illustrated in Example 5.9 of Section 5.3 form the basis of the encoding scheme for the Gaussian ZZ network.

More formally, the interference that has to be neutralized, will be combined with the main message in the first layer according to some partially-invertible function. In the second layer the inverse of the function is applied on this combination and the other interference received through the cross link. The remaining parts of the interference has to be either decoded or treated as noise. The neutralization is implemented using lattice codes and the rate-splitting along with appropriate power allocation is also used.

We formally define a partially-invertible function and a Z-*neutralization* network in the following. The Gaussian ZZ network is essentially a cascade of two Z-*neutralization* networks. An achievable rate region for the Z-neutralization network is given in Lemma 6.16. This rate region will be later used to obtain an achievable rate region for the Gaussian ZZ network. We will analyze the performance of the Gaussian encoding/decoding in the following.

Definition 6.14. *Let $\mathcal{U}$ and $\mathcal{V}$ be two finite sets. A function $\phi(\cdot,\cdot)$ defined on $\mathcal{U}\times\mathcal{V}$ is called* partially-invertible, *if and only if having $\phi(u,v)$ and u, one can always reconstruct v for any $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Similarly, u can be obtained from $\phi(u,v)$ and v.*

An intuitive way of thinking about a partially-invertible $\phi(u,v)$ is the following. An arbitrary function defined on a finite sets $\mathcal{U}$ and $\mathcal{V}$ creates a table with rows corresponding to the elements of $\mathcal{U}$ and columns corresponding to the elements of $\mathcal{V}$, the each cell of the table consists the value assigned to its row and column by the function. A function will be partially-invertible, if and only if no two cells in the same column or row of its table be identical.

Note that summation over real numbers, and multiplication over non-zero numbers are two examples of partially-invertible functions. However, it is clear multiplication over real numbers is not partially-invertible, since $w = \phi(1,0) = \phi(2,0)$, and therefore having w and $v = 0$, u can be anything.

Definition 6.15. *Consider the* Z *network shown in Figure 6.11, which consists of a Gaussian broadcast channel from F_2 to the receivers and a Gaussian multiple access channel from F_1 and F_2 to G_1. A* Z-neutralization *network is a*

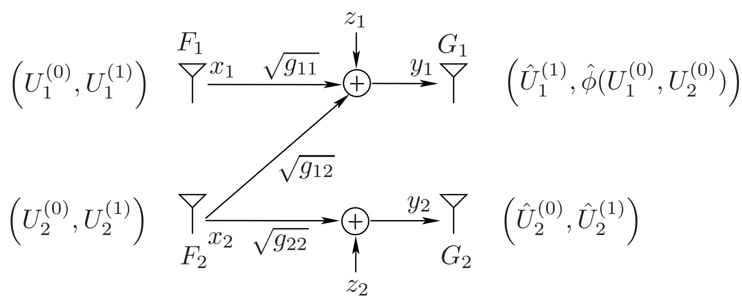


Figure 6.11: The Gaussian Z-neutralization channel.

Z *network, wherein the first source node has two messages $(U_1^{(0)}, U_1^{(1)})$ of rates Υ_0 and Υ_1, respectively. Similarly the second source observes two independent messages $(U_2^{(0)}, U_2^{(1)})$ of rates Υ_0 and Υ_2.*

The second receiver is interested in decoding $U_2^{(0)}$ and $U_2^{(1)}$, while the first destination wishes to decode $\phi(U_1^{(0)}, U_2^{(0)})$ and $U_1^{(1)}$, where $\phi(\cdot,\cdot)$ can be any arbitrary partially-invertible function. A rate tuple $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ is called achievable if the receivers can decode their messages with arbitrary small error probability.

Lemma 6.16. *Consider the* Z*-neutralization network defined Definition 6.15 with channel gains (g_{11}, g_{12}, g_{22}) (see Figure 6.11). Let*

$$\lambda \triangleq \min\{g_{11}, g_{12}, g_{22}\}, \tag{6.65}$$

and

$$\mu \triangleq \max\left\{g_{11}, g_{12}, g_{22}, \frac{g_{11}g_{22}}{g_{12}}\right\}. \tag{6.66}$$

Any rate tuple $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ satisfying

$$\Upsilon_0 \leq \left(\frac{1}{2}\log(\lambda) - \frac{1}{2}\right)^+, \tag{6.67}$$

$$\Upsilon_0 + \Upsilon_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \tag{6.68}$$

$$\Upsilon_0 + \Upsilon_2 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \tag{6.69}$$

$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \left(\frac{1}{2}\log(\mu) - \frac{3}{2}\right)^+, \tag{6.70}$$

is achievable.

The proof of this lemma is presented in Appendix E.6.

As mentioned before, we strategically split the messages and require functional reconstructions for some of them at the relay nodes to facilitate neutralization at the destinations. More precisely, in the first layer of the network, each source node splits its message into two parts, namely, "functional" and private parts, $W_1 = \left(U_1^{(0)}, U_1^{(1)}\right)$ and $W_2 = \left(U_2^{(0)}, U_2^{(1)}\right)$. The "functional" parts $U_1^{(0)}, U_2^{(0)}$ both have the same rates Υ_0. Both transmitters use a common lattice code to encode their functional sub-messages. Now the first layer encodes the message such that the first receiver (which is relay A in the original ZZ network) can decode $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$, and the second one (relay B in the original ZZ network) can decode $U_2^{(0)}$ and $U_2^{(1)}$. Lemma 6.16 gives the rates at which these can be sent reliably. The second stage operates in a manner similar to the first stage, by splitting the messages into functional and private parts. The first sender (relay A in the original network) uses $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$ as the private and functional parts and the other one (relay B) uses $U_2^{(1)}$ and $U_2^{(0)}$ as the private and functional parts.

The functional parts are sent appropriately, using a common lattice code in both stages. Let $\mathbf{x}_1^{(N)}$ and $\mathbf{x}_2^{(N)}$ be the lattice codewords, corresponding to $U_1^{(0)}$ and $U_2^{(0)}$, respectively. The power allocation In the first layer it is done so that two lattice points get received at A at the same power (see Figure 6.5). The group structure of the lattice code implies that the summation of two received lattice point, $\tilde{\mathbf{x}}^{(N)} = \mathbf{x}_1^{(N)} + \mathbf{x}_2^{(N)}$ is still a valid codeword, and can be decoded by A. The function $\phi(\cdot, \cdot)$ is in fact the decoded message from $\tilde{\mathbf{x}}^{(N)}$. In the second stage, relay node B, sends the inverse of the the received lattice point, that is $\mathbf{x}_2^{'(N)} = -\mathbf{x}_2^{(N)}$, while A forwards the sum lattice point, $\mathbf{x}_1^{'(N)} = \tilde{\mathbf{x}}^{(N)}$. Again these lattice points are scaled properly so that they get received at D_1 at the same power. Thus, their summation would be a lattice point and equals $\mathbf{x}_1^{'(N)} + \mathbf{x}_2^{'(N)} = (\mathbf{x}_1^{(N)} + \mathbf{x}_2^{(N)}) - \mathbf{x}_2^{(N)} = \mathbf{x}_1^{(N)}$, which will be decoded to $U_1^{(0)}$. The other destination D_2, receives $-\mathbf{x}_2^{(N)}$, finds its inverse $\mathbf{x}_2^{(N)}$, and finally decodes it to $U_2^{(0)}$.

We essentially use the result of Lemma 6.16 as an achievable rate region for the Z-neutralization network. We use notation (λ_g, μ_g) and (λ_h, μ_h) to distinguish between λ and μ parameters of the first and the second layers of the network.

In the first layer of the network, each source node splits its message into two parts, namely, functional and private parts, $W_1 = \left(U_1^{(0)}, U_1^{(1)}\right)$ and $W_2 = \left(U_2^{(0)}, U_2^{(1)}\right)$, where the functional parts, have the same rate, *i.e.,* $\Upsilon_{1,0} = \Upsilon_{2,0} = \Upsilon_0$. Both transmitters use a common lattice code to encode their functional sub-messages into $\mathbf{x}_{1,0} = \psi(U_1^{(0)})$ and $\mathbf{x}_{2,0} = \psi(U_2^{(0)})$, where ψ is the one-to-one encoding map induced by the lattice code. We define the partially-invertible function by

$$\phi\left(U_1^{(0)}, U_2^{(0)}\right) = \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)})\right) = \psi^{-1}(\mathbf{x}_{1,0} + \mathbf{x}_{2,0}). \tag{6.71}$$

We denote the rates of the private sub-messages by Υ_1 and Υ_2, where $\Upsilon_i = R_i - \Upsilon_0$, for $i = 1, 2$. The goal is to encode and forward messages to A and B in such a way that A can decode $U_1^{(1)}$ and $\phi(U_1^{(0)}, U_2^{(0)})$, and B can decode $U_2^{(0)}$ and $U_2^{(1)}$. Based on Lemma 6.16, this can be done provided that

$$\begin{aligned}
\Upsilon_0 &\leq \left(\frac{1}{2}\log(\lambda_g) - \frac{1}{2}\right)^+, \\
\Upsilon_0 + \Upsilon_1 &\leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \\
\Upsilon_0 + \Upsilon_2 &\leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \\
\Upsilon_0 + \Upsilon_1 + \Upsilon_2 &\leq \left(\frac{1}{2}\log(\mu_g) - \frac{3}{2}\right)^+.
\end{aligned}$$

The second layer of the network is another Z-neutralization network with transmitters A and B, and receivers D_1 and D_2. In this layer, we use $V_1^{(0)} = \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)})\right)$, $V_1^{(1)} = U_1^{(1)}$ as the functional and private messages of the first relay node, and $V_2^{(0)} = \psi^{-1}(-\mathbf{x}_{2,0}) = \psi^{-1}(-\psi(U_2^{(0)}))$ and $V_2^{(1)} = U_2^{(1)}$ for the functional and private messages of second relay. Denoting the corresponding rates by Θ_0, Θ_1, and Θ_2, we have

$$\Theta_i = \Upsilon_i, \qquad i = 0, 1, 2. \tag{6.72}$$

The goal is to encode and send these messages to the destinations, such that D_1 can decode $\phi(V_1^{(0)}, V_2^{(0)})$ and $V_1^{(1)}$, and D_2 can decode $V_2^{(0)}$ and $V_2^{(1)}$. Again we use the achievable rate region proposed in Lemma 6.16.

$$\begin{aligned}
\Theta_0 &\leq \left(\frac{1}{2}\log(\lambda_h) - \frac{1}{2}\right)^+, \\
\Theta_0 + \Theta_1 &\leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+, \\
\Theta_0 + \Theta_2 &\leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+, \\
\Theta_0 + \Theta_1 + \Theta_2 &\leq \left(\frac{1}{2}\log(\mu_h) - \frac{3}{2}\right)^+.
\end{aligned}$$

Note that the first destination observes $\phi(V_1^{(0)}, V_2^{(0)})$, which is equivalent to

$$\begin{aligned}\phi\left(V_1^{(0)}, V_2^{(0)}\right) &= \psi^{-1}\left(\psi(V_1^{(0)}) + \psi(V_2^{(0)})\right)\\ &= \psi^{-1}\left(\psi(U_1^{(0)}) + \psi(U_2^{(0)}) - \psi(U_2^{(0)})\right)\\ &= U_1^{(0)}. \end{aligned} \tag{6.73}$$

Therefore, combining it with $V_1^{(1)} = U_1^{(1)}$, the first destination node can decode W_1. The second destination node D_2 has $V_2^{(0)}$ and $V_2^{(1)} = U_2^{(1)}$, and can compute

$$\psi^{-1}\left(-\psi(V_2^{(0)})\right) = U_2^{(0)}, \tag{6.74}$$

and hence it decodes W_2.

It is clear that this scheme can reliably transmit the messages with rate pair in

$$\begin{aligned}\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}} = \Big\{(R_1, R_2) :\exists\ &\Upsilon_0, \Upsilon_1, \Upsilon_2, \Theta_0, \Theta_1, \Theta_2 \geq 0,\\ &R_1 = \Theta_0 + \Theta_1,\\ &R_2 = \Theta_0 + \Theta_2,\\ &\Theta_i = \Upsilon_i, \qquad i = 0,1,2,\\ &\Upsilon_0 \leq \left(\frac{1}{2}\log(\lambda_g) - \frac{1}{2}\right)^+,\\ &\Upsilon_0 + \Upsilon_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+,\\ &\Upsilon_0 + \Upsilon_2 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+,\\ &\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \left(\frac{1}{2}\log(\mu_g) - \frac{3}{2}\right)^+,\\ &\Theta_0 \leq \left(\frac{1}{2}\log(\lambda_h) - \frac{1}{2}\right)^+,\\ &\Theta_0 + \Theta_1 \leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+,\\ &\Theta_0 + \Theta_2 \leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+,\\ &\Theta_0 + \Theta_1 + \Theta_2 \leq \left(\frac{1}{2}\log(\mu_h) - \frac{3}{2}\right)^+ \Big\}. \end{aligned} \tag{6.75}$$

It only remains to apply Fourier-Motzkin elimination to project this region onto (R_1, R_2). This gives us

$$\begin{aligned}\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}} = \Big\{(R_1,R_2) : & R_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \\ & R_1 \leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_g) + \frac{1}{2}\log(\lambda_h) - \frac{3}{2}\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_h) + \frac{1}{2}\log(\lambda_g) - \frac{3}{2}\right)^+ \Big\}.\end{aligned}$$

Note that the RHS's of the sum-rate bounds depend on the order of the channel gains. For most of possible orderings, these two inequalities would be consequences of the individual rate bounds. For example, if $\lambda_h = h_{22}$, then the last bound is implied by the first and fourth bounds, since $\mu_g \geq g_{11}$. It can be shown in general that $\mathcal{R}^{\mathsf{GZZ}}_{ach}$ is equivalent to

$$\begin{aligned}\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}} = \Big\{(R_1,R_2) : & R_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+, \\ & R_1 \leq \left(\frac{1}{2}\log(g_{22}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(h_{11}) - 1\right)^+, \\ & R_2 \leq \left(\frac{1}{2}\log(h_{22}) - 1\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_g) + \frac{1}{2}\log(h_{12}) - \frac{3}{2}\right)^+, \\ & R_1 + R_2 \leq \left(\frac{1}{2}\log(\mu_h) + \frac{1}{2}\log(g_{12}) - \frac{3}{2}\right)^+ \Big\}.\end{aligned}$$

Note that the RHS's of the inequities in the characterization of $\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}}$ and that of $\underline{\mathcal{R}}_{\mathsf{GZZ}}$ in Definition 6.3 are not similar. However, we can show that the gap between the RHS of each inequality in $\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}}$ and its corresponding one in $\underline{\mathcal{R}}_{\mathsf{GZZ}}$ is bounded by a constant. For the individual rate inequalities, we have

$$\frac{1}{2}\log(1+x) \leq \frac{1}{2}\log(x) + \frac{1}{2} \tag{6.76}$$

for all $x \geq 1$. Moreover, for the sum-rate inequalities we can write

$$\begin{aligned}
&\frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) \\
&\qquad \leq \frac{1}{2}\log(3\max\{g_{11},g_{12}\}) + \frac{1}{2}\log\left(\frac{2\max\{g_{12},g_{22}\}}{g_{12}}\right) \\
&\qquad \leq \frac{1}{2}\log\left(\frac{\max\{g_{11},g_{12}\}\cdot\max\{g_{22},g_{12}\}}{g_{12}}\right) + \frac{1}{2}\log 6 \\
&\qquad \leq \frac{1}{2}\log(\mu_g) + \frac{1}{2}\log 6. \qquad (6.77)
\end{aligned}$$

for $g_{11} \geq 1$, $g_{12} \geq 1$, and $g_{22} \geq 1$. This implies

$$\begin{aligned}
&\left[\frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) + \frac{1}{2}\log(1+h_{12})\right] \\
&\qquad - \left[\frac{1}{2}\log(\mu_g) + \frac{1}{2}\log(h_{12}) - \frac{3}{2}\right]^+ \leq 2 + \frac{1}{2}\log 6. \qquad (6.78)
\end{aligned}$$

A similar argument holds for the other sum-rate bound.

Applying (6.76) and (6.78), we obtain the following achievable rate region, which is a subset of $\overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}}$.

$$\begin{aligned}
\overline{\mathcal{R}}_{\mathsf{GZZ}} = \Bigg\{(R_1,R_2) : & R_1 \leq \left(\frac{1}{2}\log(1+g_{11}) - \frac{3}{2}\right)^+, \\
& R_2 \leq \left(\frac{1}{2}\log(1+g_{22}) - \frac{3}{2}\right)^+, \\
& R_1 \leq \left(\frac{1}{2}\log(1+h_{11}) - \frac{3}{2}\right)^+, \\
& R_2 \leq \left(\frac{1}{2}\log(1+h_{22}) - \frac{3}{2}\right)^+, \\
& R_1+R_2 \leq \Bigg(\frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) \\
& \qquad + \frac{1}{2}\log(1+h_{12}) - 2 + \frac{1}{2}\log 6\Bigg]^+, \\
& R_1+R_2 \leq \Bigg[\frac{1}{2}\log(1+h_{11}+h_{12}) + \frac{1}{2}\log\left(1+\frac{h_{22}}{h_{12}}\right) \\
& \qquad + \frac{1}{2}\log(1+g_{12}) - 2 + \frac{1}{2}\log 6\Bigg]^+\Bigg\}
\end{aligned}$$

Therefore, we have

$$\overline{\mathcal{R}}_{\mathsf{GZZ}} \subseteq \overline{\mathcal{R}}^{(1)}_{\mathsf{GZZ}} \subseteq \mathcal{R}_{\mathsf{GZZ}}.$$

Since for any rate pair $(R_1, R_2) \in \underline{\mathcal{R}}_{\mathsf{GZZ}}$, the rate pair $(R_1 - \frac{1}{4}\log 96, R_2 - \frac{1}{4}\log 96)$ belongs to $\overline{\mathcal{R}}_{\mathsf{GZZ}}$. This argument leads to an approximate capacity characterization for the Gaussian ZZ network, with a gap bounded by $\frac{1}{4}\log 96 \simeq 1.65$ bits per user.

7 The Diamond Network with Adversarial Jammer

Wireless communication is inherently susceptible to malicious interference attempting to disrupt communications. An adversary can utilize the broadcast medium to insert disruptive signals. This problem has been well-studied for point-to-point communication, with early works in arbitrarily varying channels (AVC) [75]. This topic has not received significant attention in the context of wireless networks, where there are relay nodes to assist communication. In this chapter we formulate and study coding strategies in this problem for a simple relay network.

Here, we start with study of this problem in a deterministic context. That is, examining the impact of the adversarial node using a linear deterministic signal interaction on a simple wireless network, which we call *deterministic diamond network*, shown in Figure 7.2. Then we use the insights for both encoding scheme as well as bounding techniques obtained from the deterministic model, to build similar results for the *Gaussian diamond* network depicted in Figure 7.1.

The role of malicious jamming nodes in wired networks has received recent significant attention in network coded systems (see [76], [77] and references therein). However, the problem in wireless networks is quite different due to the signal interactions caused by the broadcast nature of the channel. We will utilize the fact that the disrupting signal transmitted by the adversary (see Figure 7.1) cause the received signals at the authenticated relays to be related to each other. We use this in order to neutralize the adversarial signal without separating it from the legitimate transmitted signal. The idea is that we utilize the "correlation" in the received signal at the relays to cancel part of the undesired adversarial signal. A similar idea is used in [66] for an amplify-forward relaying strategy to reduce the interference at the receiver, and assuming a *sum power constraint* at the relays allows to utilize a beam-forming strategy.

The main contributions of this chapter are the following. We formulate the problem of adversarial jamming for a wireless (diamond) network, and study it under the deterministic model as well as noisy Gaussian model, called the DDA and GDA networks, respectively. In particular, we show an exact characterization of reliable transmission rate for a diamond network with linear deterministic model [65]. The coding strategy for this case is based on ideas similar to that of the interference neutralization technique developed in [20], to reduce the effect of the jamming signal on the received signal at the destination. However, a fundamental difference here is that, the jamming signal does not have to describe a message from a given codebook, and can be any thing. Hence, instead of decoding a *functional* message, the relay nodes forward that part of the received signal which cannot be decoded at signal level. The power allocated to these two interfered part should be chosen such that the effective interference power at the destination be small. From this point of view, what happens in the diamond network is interference nulling, but not interference neutralization.

Moreover, we use insight obtained by analysis of the simplified problem in order to derive an approximate capacity characterization for the GDA network. We provide an outer bound as well as achievable strategies for this network. The coding scheme proposed here is based on a superposition coding at the source, and partially decoding the message and utilizing interference nulling at the relays,. This is shown that the gap between the performance of the proposed scheme and the any optimal strategy is bounded above by 4 bits/sample, which results in an approximate capacity characterization.

The deterministic version of this problem is studied for some regimes of parameters in a recent work [78]. However, here we generalize the deterministic characterization for arbitrary channel parameters. More importantly, we show that the achievable strategy inspired by the deterministic analysis is within a constant number of bits of the outer bound in the Gaussian case. A similar problem is also studied in [79], where the signal received at the relays are corrupted by the the *same* copies of an interfering signal. An important difference between [79] and our work is that they assume orthogonal bit pipes from relays to the destination, while we consider a noisy multiple access channel to the destination. The main focus of [79] is to characterize the scaling laws of the capacity (the pre-log as the power of the transmitter and the interference are taken to infinity) in this setting.

We describe the problem and main results in Section 7.1. The analysis for linear deterministic networks is given in Section 7.2. Section 7.3 develops the outer bound and the achievable strategy for the Gaussian case. The results presented here lead to several natural questions on the generalizations to arbitrary networks, multiple adversaries etc. These are topics of future work.

7.1 Problem Formulation and Main Results

Consider the diamond network in Figure 7.1, where the source wishes to send the message W reliably to the destination D. It encodes the message and broadcasts X^n to the relays B_1 and B_2 through the Gaussian channels. However, the relays receive interference from an adversarial node A, who wishes to jam the transmission by inserting noise to the system. The signal received by the relays can be written as

$$y_1[t] = \sqrt{f_1}x[t] + \sqrt{g_1}u[t] + z_1[t]$$
$$y_2[t] = \sqrt{f_2}x[t] + \sqrt{g_2}u[t] + z_2[t],$$

where x is the transmitted signal by the source, u is the interfering signal inserted by the jammer, and z_1 and z_2 are the additive white Gaussian noise with unit variance over each channel. The relay nodes perform any (causal) processing on their received signal sequences $\{y_1[t]\}$ and $\{y_2[t]\}$ respectively, to obtain their transmitting signal sequences, $\{x_1(t)\}$ and $\{x_2(t)\}$. The received signals at the destination nodes from the Gaussian multiple access channel can be written as

$$y[t] = \sqrt{h_1}x_1[t] + \sqrt{h_2}x_2[t] + z[t],$$

and wishes to decode M based on its received signal. We also assume equal power constraints for the source, relays, and adversarial node, that is, $\mathbb{E}[x^2] \leq 1$, $\mathbb{E}[x_1^2] \leq 1$, $\mathbb{E}[x_2^2] \leq 1$, and $\mathbb{E}[u^2] \leq 1$. We term this network the Gaussian diamond network with adversary (GDA).

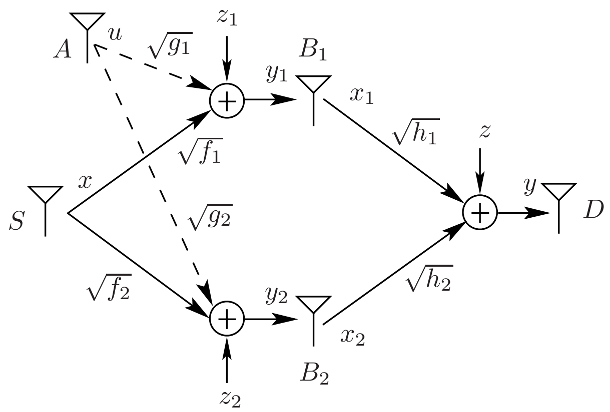


Figure 7.1: Transmission model; Source node S wishes to communicate to D, while the system is jammed by the adversarial node A.

The AVC problem in a point-to-point system is studied under two assumptions [75]: (i) there exists common randomness shared between the source and destination, unknown to adversary; This facilitates the use of *random* codebooks chosen using the common randomness; (ii) if there is no such a common randomness, and a *fixed* codebook is used for transmission. This notion of shared randomness plays no role in the analysis of the deterministic network. For the Gaussian problem, though we present the work for case (i), i.e., shared secret common randomness between source and the relays, these results can be easily extended to case (ii).

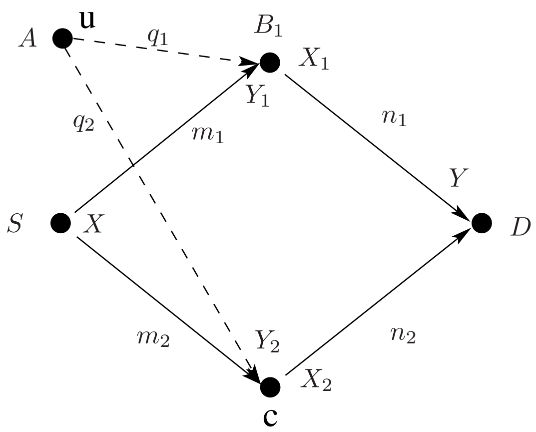


Figure 7.2: The deterministic model of the problem.

The deterministic counterpart of this problem is shown in Figure 7.2, where source node S wishes to communicate its message to the destination node D via the relay nodes B_1 and B_2. The channels from source to relays are modeled by M_1 and M_2 and from the relays to the destination are denoted by N_1 and N_2. We also use P_1 and P_2 to denote the transition matrix from the adversarial node to the relays. Similar lowercase letters are used to denote the rank of the matrices (channel gains). Therefore, the transmission over the first and second layers of the network can be respectively written as

$$Y_i[t] = M_i X[t] + Q_i U[t], \qquad i = 1, 2, \tag{7.1}$$

and

$$Y[t] = N_1 X_1[t] + N_2 X_2[t] \tag{7.2}$$

where here X, U, Y_1 and Y_2 are vectors of length p with elements form a finite field $\mathbb{F}$, and the channel matrices are powers of the shift matrix $\mathbf{J}$, *e.g.,* $M_1 = \mathbf{J}^{p-m_1}$, were $\mathbf{J}$ is defined in (5.3). We refer to this network by the deterministic diamond network with adversary (DDA) for the rest of this chapter.

The following theorem states the capacity characterization for this deterministic network.

Theorem 7.1 (The capacity of the deterministic adversarial diamond network)**.** *The capacity of the deterministic diamond network with adversarial node is given by*

$$\begin{aligned} C_{\mathsf{DDA}} = \min \Big\{ & \psi(m_1, m_2, q_1, q_2) - \max\{q_1, q_2\}, \\ & \max(n_1, n_2) \\ & (m_1 - q_1)^+ + n_2, \\ & (m_2 - q_2)^+ + n_1, \Big\}, \end{aligned} \tag{7.3}$$

where

$$\Psi(m_1, m_2, q_1, q_2) = \begin{cases} \max\{m_1, m_2, q_1, q_2\} & \textit{if } m_1 + q_2 = m_2 + q_1 \\ \max\{m_1 + q_2, m_2 + q_1\} & \textit{otherwise.} \end{cases} \tag{7.4}$$

We will use this result to devise a coding strategy which gives an approximate capacity characterization for the Gaussian network. This is given in the following theorem.

Theorem 7.2 (An approximate capacity for the Gaussian adversarial diamond network)**.** *For a given* GDA *network with parameters given in Figure 7.1, define*

$$\overline{C}_{\mathsf{GDA}} = \min \Big\{ \frac{1}{2}\log\left(1 + \frac{f_1 + f_2 + (\sqrt{f_1 g_2} - \sqrt{f_2 g_1})^2}{1 + g_1 + g_2}\right), \tag{7.5}$$

$$\frac{1}{2}\log\left(1 + (\sqrt{h_1} + \sqrt{h_2})^2\right), \tag{7.6}$$

$$\frac{1}{2}\log\left(1 + \frac{f_1}{1 + g_1}\right) + \frac{1}{2}\log(1 + h_2), \tag{7.7}$$

$$\frac{1}{2}\log\left(1 + \frac{f_2}{1 + g_2}\right) + \frac{1}{2}\log(1 + h_1). \Big\}. \tag{7.8}$$

Then the randomized capacity of the network C_{GDA} *satisfies*

$$\overline{C}_{\mathsf{GDA}} - 4 \leq C_{\mathsf{GDA}} \leq \overline{C}_{\mathsf{GDA}}. \tag{7.9}$$

A trivial sub-optimal scheme here is to treat the interference as independent noises, and follow the known schemes for noisy diamond relay network [65]. However, such noises are correlated since they are generated by the same jamming source. This correlation can be utilized to significantly improve the communication rate.

We will discuss and present the proof of these two theorems in the following section.

7.2 The DDA Network

We study a deterministic version of the diamond network with jammer problem in this section. In this model the randomness of the noise is ignored by considering the interference limited regime, and we focus on the signal interaction instead. The main goal is to prove Theorem 7.1.

a) The converse part

We first show that any achievable rate R satisfies $R \leq C_{\mathsf{DDA}}$. In order to do this, we show that R is upper bounded by all the four terms in the minimization in (7.3). Assume we use code of length ℓ, and W can be decoded from Y^ℓ with error probability ϵ_ℓ, where $\epsilon_\ell \to 0$ as ℓ grows. Therefore, using Fano's inequality, we have

$$H(X^\ell|Y^\ell) \leq H(W|Y^\ell) \leq \ell\epsilon_\ell. \tag{7.10}$$

We denote by G_X and G_U the transfer matrix from the adversarial node and the source to the relays, respectively. These matrices are defined as

$$G_X = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \qquad G_U = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}. \tag{7.11}$$

So, the signals received by the relays can be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = G_{X,U} \begin{bmatrix} X \\ U \end{bmatrix} \tag{7.12}$$

where $G_{X,U} = [G_X|G_U]$ is the transfer matrix from A and S to the relay nodes. We use bold face notation to denote ℓ copies of a channel matrix, which is the transfer matrix applied over a codeword of length ℓ, *e.g.*, $\mathbf{G}_{X,U} = I_\ell \otimes G_{X,U}$.

Note that the Markov chain $X^\ell \leftrightarrow (Y_1^\ell, Y_2^\ell) \leftrightarrow Y^\ell$ implies that the signals received at B_1 and B_2 are enough to decode the message. It is also clear that once B_1 and B_2 can decode X, they also know $\mathbf{G}_U U^\ell$. Hence, $H(X^\ell, \mathbf{G}_U U^\ell | Y_1^\ell, Y_2^\ell) \leq \ell\epsilon_\ell$. Therefore,

$$\begin{aligned} \ell R + H(\mathbf{G}_U U^\ell) &= H(X^\ell) + H(\mathbf{G}_U U^\ell) = H(X^\ell, \mathbf{G}_U U^\ell) \\ &\leq I(X^\ell, \mathbf{G}_U U^\ell; Y_1^\ell, Y_2^\ell) + \ell\epsilon_\ell \\ &\leq H(Y_1^\ell, Y_2^\ell) + \ell\epsilon_\ell = \operatorname{rank}(\mathbf{G}_{X,U}) + \ell\epsilon_\ell \\ &= \ell \operatorname{rank}(G_{X,U}) + \ell\epsilon_\ell \end{aligned} \tag{7.13}$$

It is clear that the adversary can choose U^ℓ such that $H(\mathbf{G}_U U^\ell) = \ell \operatorname{rank}(G_U) = \ell \max\{q_1, q_2\}$. Therefore, we have $R \leq \operatorname{rank}(G_{X,U}) - \max\{q_1, q_2\} + \epsilon_\ell$. We use the following lemma to evaluate $\operatorname{rank}(G_{X,U})$ which is proved in Appendix F.1.

Lemma 7.3. *Let G be a diagonal matrix of the form*

$$G = \begin{bmatrix} \mathbf{J}^{p-m_1} & \mathbf{J}^{p-q_1} \\ \mathbf{J}^{p-m_2} & \mathbf{J}^{p-q_2} \end{bmatrix}.$$

Then, we have $\text{rank}\,(G) = \Psi(m_1, m_2, q_1, q_2)$*, where the function* Ψ *is defined in* (7.4).

Having this lemma, and by evaluating $\text{rank}\,(G_{X,U})$ in (7.13), we get the desired inequality. Note that this bound essentially captures the maximum amount of information can be transmitted from through the cut with partitions $\Omega_1 = \{S, A\}$ and $\Omega^c = \{B_1, B_2, D\}$.

In order to show that $R \le (m_1 - q_1)^+ + n_2$, we first recall that decodability of W from Y^ℓ implies

$$\begin{aligned} H(X^\ell, \mathbf{Q}_1 U^\ell | Y_1^\ell, Y^\ell) &= H(X^\ell | Y_1^\ell, Y^\ell) + H(\mathbf{Q}_1 U^\ell | X^\ell, Y_1^\ell, Y^\ell) \\ &\le H(W | Y_1^\ell, Y^\ell) + H(Y_1^\ell - \mathbf{M}_1 X^\ell | X^\ell, Y_1^\ell, Y^\ell) \\ &\le \ell \epsilon_\ell. \end{aligned} \tag{7.14}$$

Hence,

$$\begin{aligned} \ell R + H(\mathbf{Q}_1 U^\ell) &= H(X^\ell) + H(\mathbf{Q}_1 U^\ell) \\ &\overset{(a)}{=} H(X^\ell, \mathbf{Q}_1 U^\ell) \\ &\le I(X^\ell, \mathbf{Q}_1 U^\ell; Y_1^\ell, Y^\ell) + \ell \epsilon_\ell \\ &\le H(Y_1^\ell, Y^\ell) + \ell \epsilon_\ell \\ &= H(Y_1^\ell) + H(Y^\ell | Y_1^\ell) + \ell \epsilon_\ell \\ &\overset{(b)}{\le} H(Y_1^\ell) + H(\mathbf{N}_2 X_2^\ell | Y_1^\ell) + \ell \epsilon_\ell \\ &\le H(Y_1^\ell) + H(\mathbf{N}_2 X_2^\ell) + \ell \epsilon_\ell \\ &\le \ell \max\{m_1, q_1\} + \ell n_2 + \ell \epsilon_\ell. \end{aligned} \tag{7.15}$$

where in (a) we used the assumption that the adversary does not know the message, and therefore, its interfering signal is independent of X^ℓ, and (b) holds since X_1^ℓ is a function of Y_1^ℓ. Combining (7.15) with the fact that the adversarial node can make $H(\mathbf{Q}_1 U^\ell)$ as large as ℓq_1, gives the second bound. It is worth mentioning that this bound essentially captures the maximum flow of information through the cut $\Omega = \{S, A, B_1\}$. The proof of the third bound is just repeating the same argument for a symmetric situation.

Finally, in order to proof the last inequality, we consider $\Omega = \{S, A, B_1, B_2\}$. It is clear that

$$\begin{aligned} \ell R &\le H(X^\ell) \le I(X^\ell; Y^\ell) + \ell \epsilon_\ell \\ &\le H(Y^\ell) + \ell \epsilon_\ell \le \ell \max\{n_1, n_2\} + \ell \epsilon_\ell. \end{aligned} \tag{7.16}$$

b) The achievability part

In the following we present an encoding scheme to achieve rates any $R \le C_{\mathsf{DDA}}$. The idea is provide enough number of linearly independent equations about the message codeword. Note that some of sub-nodes in B_1 and B_2 receive

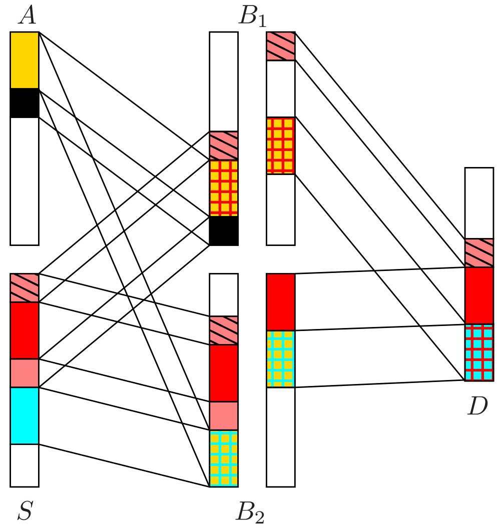


Figure 7.3: Transmission strategy to provide interference nulling.

the same bit from A, and therefore this interfering bit can get nulled [20] if they get forwarded and received on the same sub-node of D.

We split the message into two parts, which essentially leads to a network decomposition, based on the sub-levels of each node which are involved in transmitting each of the sub-messages. We first identify the maximum number of bits of the message that can be recovered by nulling the interference. Any sub-level which is involved in such nulling would belong the first network partition.

There are also possibly a subset of message bits received at the relays above the interference level, and therefore not corrupted with interference. These bits can be directly forwarded to the destination. This idea is illustrated in Figure 7.3.

We use $R^{(N)}$ and $R^{(P)}$ to denote the number of equations can be received at the destination using *interference nulling*, and forwarding *pure* signal bits from the relays, respectively.

We denote the levels of B_1 at the receiver side by $Y_{1,1}$ (for the highest) to $Y_{1,p}$ (for the lowest), and similarly for B_2. Define

$$\delta \triangleq \min\left\{q_1, q_2\right\} - \min\left\{(q_1 - m_1)^+, (q_2 - m_2)^+\right\}.$$

It is easy to show that $Y_{1,(p-(q_1-q_2)^+ +\kappa)}$ and $Y_{2,(p-(q_2-q_1)^+ +\kappa)}$ are corrupted by the same bit from A, for $\kappa = 0, \ldots, \delta - 1$. Moreover, at least one of $Y_{1,(p-(q_1-q_2)^+ +\kappa)}$ and $Y_{2,(p-(q_2-q_1)^+ +\kappa)}$ receive a bit from the source. Therefore, for each κ, if $Y_{1,(p-(q_1-q_2)^+ +\kappa)}$ and $Y_{2,(p-(q_2-q_1)^+ +\kappa)}$ can get forwarded on the same destination sub-node, the destination receives an equation about the source whose interference is nulled. It is worth mentioning that if $m_1 + q_2 = m_2 + q_1$, then the interference-nulled equations received at D are zero, since the message bits would be also nulled. On the other hand, the condition $m_1 + q_2 \neq m_2 + q_1$ guarantees that such received legitimate bits are different.

Each bit nulling utilizes one sub-link for B_1 to D, and one sub-link from B_2 to D. Hence, we can send up to

$$R^{(N)} \leq \nu \triangleq \begin{cases} 0 & \text{if } m_1 + q_2 = m_2 + q_1 \\ \min\{\delta, n_1, n_2\} & \text{otherwise.} \end{cases} \tag{7.17}$$

On the other hand, nodes B_1 and B_2, respectively, receive $(m_1 - q_1)^+$ and $(m_2 - q_2)^+$ bits from the source which are above the interference level, and therefore not corrupted. These bits can be forwarded to D through the remaining $(n_1 - \nu)$ and $(n_2 - \nu)$ links in the second layer of the network. However, these two set of bits have overlap, since they both are the upper level bits sent by A. Analysis of the number of non-interfered bits can be sent to the destination node, is equivalent to a study of another linear shift deterministic diamond network without adversary, where there are $(m_1 - q_1)^+$ and $(m_2 - q_2)^+$ links from S to the relays, and $(n_1 - \nu)$ and $(n_2 - \nu)$ links from the relays to the destination. The capacity of this network is easy to compute as in [65]. We get

$$\begin{aligned} R^{(P)} \leq \min\big\{ &\max\{(m_1 - q_1)^+, (m_2 - q_2)^+\}, (m_1 - q_1)^+ + (n_1 - \nu), \\ &(m_2 - q_2)^+ + (n_2 - \nu), \max\{n_2 - \nu, n_2 - \nu\}\big\}. \end{aligned} \tag{7.18}$$

It is easy to show that equations we receive using two method are linearly independent, and therefore any rate $R \leq R^{(N)} + R^{(P)}$ is achievable. Using some algebra and manipulations, one can show that adding the RHS's of (7.17) and (7.18), gives us the same bound claimed in the theorem. This is discussed in detail in Appendix F.2. □

In the following example we discuss this network decomposition idea in more detail.

Example 7.4. *Consider a diamond network with parameters $m_1 = 4$, $m_2 = 6$, $q_1 = 3$, $q_2 = 2$, $n_1 = 5$, and $n_2 = 4$. Also assume $p = 7$. Theorem 7.1 implies that $C = 5$. In the following we show how the destination can get 5 linearly independent equations about the bits transmitted by the source node. Denoting the source and interference bits by X_i and u_j, for $i = 1, \ldots, 6$, and $j = 1, 2, 3$, we have*

$$Y_1 = \begin{bmatrix} Y_1(1) \\ Y_1(2) \\ Y_1(3) \\ Y_1(4) \\ Y_1(5) \\ Y_1(6) \\ Y_1(7) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ X(1) \\ X(2) + U(1) \\ X(3) + U(2) \\ X(4) + U(3) \end{bmatrix}, \quad Y_2 = \begin{bmatrix} Y_2(1) \\ Y_2(2) \\ Y_2(3) \\ Y_2(4) \\ Y_2(5) \\ Y_2(6) \\ Y_2(7) \end{bmatrix} = \begin{bmatrix} 0 \\ X(1) \\ X(2) \\ X(3) \\ X(4) \\ X(5 + U(1) \\ X(6) + U(2) \end{bmatrix}.$$

Recall that we can null up to $\nu = \min\{\delta, n_1, n_2\} = 2$ *bits. This will be done by forwarding* $Y_1(5)$ *and* $Y_1(6)$ *by* B_1 *and* $Y_2(6)$ *and* $Y_2(7)$ *by* B_2 *over the lowest* 2 *links to* D. *The destination node will receive* $Y(6) = X(2) + X(5)$ *and* $Y(7) = X(3) + X(6)$ *on its lowest level.*

The relay node B_1 *has only one bit of non-corrupted signal,* X_1 *and node* B_2 *has four of them,* $X(1)$, $X(2)$, $X(3)$, *and* $X(4)$. *In order to send these bits,* B_1 *send* $X(1)$ *on its highest level, and* B_2 *forwards* $X(2)$ *and* $X(3)$ *on its highest levels. Note that, we cannot decode all the six bits, but obtain five linearly independently equations involving* $X(1), \ldots, X(6)$.

7.3 The GDA Network

In this section, we translate the deterministic case analysis in Section 7.2 into a universally approximate characterization for the (noisy) Gaussian network, and prove Theorem 7.2. In fact, we use the insights given by analysis of the network in deterministic model for both deriving an upper bound for the capacity, as well as proposing an (approximately) optimal encoding strategy which leads to a lower bound for the capacity within a constant bit gap from the upper bound.

As mentioned before, we assume that there exist a common randomness shared between all the legitimate nodes (source, relays, and destination), which is not available for the jammer. These nodes use this common randomness to randomly choose a codebook from a family of pre-determined codebooks, for each message. This allows us to assume that the adversary is not aware of the current codebook used at the time, although it knows the family of the codebooks. It can be shown that the number of random bits required for a secret choice of codebook, is asymptotically negligible [75] in comparison of the number of bits communicated over the network. We can use the technique of elimination of correlation developed in [75] to extend this result to the case when such shared randomness does not exist.

a) The Upper Bound

We imitate the same bounding techniques we used to prove Theorem 7.1. We first start with the cut $\Omega = \{S, A\}$. Recall that since we assume a shared randomness between the the legitimate nodes, the adversary does not know the currently used codebook, and therefor the worst distribution it can use to generate its signal is the Gaussian distribution [80].

$$\begin{aligned}\ell R &\leq I(y_1^\ell, y_2^\ell; x^\ell) + \ell\epsilon_\ell = h(y_1^\ell, y_2^\ell) - h(y_1^\ell, y_2^\ell | x^\ell) + \ell\epsilon_\ell \\ &\leq \frac{\ell}{2}\log\left(\frac{1 + g_1 + g_2 + f_1 + f_2 + (\sqrt{f_1 g_2} - \sqrt{f_2 g_1})^2}{1 + g_1 + g_2}\right) + \ell\epsilon_\ell. \qquad (7.19)\end{aligned}$$

The second bound is easy to derive.

$$\ell R \leq I(y_1^\ell, y^\ell; x^\ell) + \ell\epsilon_\ell$$

$$
\begin{aligned}
&= I(y_1^\ell; x^\ell) + I(y^\ell; x^\ell | y_1^\ell) + \ell\epsilon_\ell \\
&= I(y_1^\ell; x^\ell) + h(y^\ell | y_1^\ell) - h(y^\ell | x^\ell, y_1^\ell) + \ell\epsilon_\ell \\
&\leq I(y_1^\ell; x^\ell) + h(y^\ell | x_1^\ell) - h(y^\ell | x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \qquad (7.20) \\
&= I(y_1^\ell; x^\ell) + h(\sqrt{h_2} x_2^\ell + z^\ell | x_1^\ell) - h(y^\ell | x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\leq I(y_1^\ell; x^\ell) + h(\sqrt{h_2} x_2^\ell + z^\ell) - h(y^\ell | x_1^\ell, x_2^\ell) + \ell\epsilon_\ell \\
&\leq \frac{\ell}{2} \log\left(1 + \frac{f_1}{1+g_1}\right) + \frac{\ell}{2}\log(1+h_2) + \ell\epsilon_\ell. \qquad (7.21)
\end{aligned}
$$

where in (7.20) we used the fact that x_1^ℓ is a function of y_1^ℓ, and also the Markov chain $y^\ell \leftrightarrow (x_1^\ell, x_2^\ell) \leftrightarrow (x^\ell, y_1^\ell)$. The third bound can be proved by repeating the same argument for $I(y_2^\ell, y^\ell; x^\ell)$.

In order to show the last upper bound, we can write

$$
\begin{aligned}
\ell R &\leq I(y^\ell; x^\ell) + \ell\epsilon_\ell \\
&\leq I(y^\ell; x_1^\ell x_2^\ell) + \ell\epsilon_\ell \qquad (7.22) \\
&\leq \frac{\ell}{2}\log\left(1 + (\sqrt{h_1} + \sqrt{h_2})^2\right). \qquad (7.23)
\end{aligned}
$$

We again used the data processing inequality for the Markov chain $y^\ell \leftrightarrow (x_1^\ell, x_2^\ell) \leftrightarrow x^\ell$ in (7.22).

b) The Lower Bound The achievability scheme we propose here is induced by the strategy proposed for the deterministic model. It is based on *message splitting* and *superposition coding*. The power allocation should be performed such that the part of the message which is not corrupted by interference can be decoded at the relays. Moreover, the interfered part get forwarded to the destination such that the effective interference at the destination be small enough such that this part of the message can be decoded at the destination. In the following we only explain this idea in more details.

Without loss of generality, we assume that the relay B_1 is stronger than B_2, *i.e.,* $\mathsf{SINR}_1 \geq \mathsf{SINR}_2$, where $\mathsf{SINR}_i = f_i/(1+g_i)$. We first split the message W into three parts, namely, W_c, W_p, W_n, with rates R_c, R_p, and R_n. Our encoding and decoding strategy guarantees that the common message, W_c, can be decoded at both relays, while the private sub-message, W_p, can be only decoded at B_1. However, neither of the relays can decode the neutralization sub-message, W_n, and it can be only decoded at the destination, once the interference is nulled.

We use three random codebooks of rates R_c, R_p, and R_n, generated according to the Gaussian distribution with unit variance. The source maps its sub-messages to the codewords from corresponding codebooks, and obtains $\mathbf{x}_c$, $\mathbf{x}_p$, and $\mathbf{x}_n$. Then the signal transmitted by the source is formed as a superposition of the three codewords, using a proper power allocation,

$$
\mathbf{x} = \sqrt{\alpha_c}\mathbf{x}_c + \sqrt{\alpha_p}\mathbf{x}_p + \sqrt{\alpha_n}\mathbf{x}_n, \qquad (7.24)
$$

where the power allocation coefficients satisfy

$$\alpha_c + \alpha_p + \alpha_n \leq 1. \tag{7.25}$$

In particular, we choose $\alpha_n = \min(1, 1/\mathsf{SINR}_1)$, $\alpha_p = \min(1, 1/\mathsf{SINR}_2) - \alpha_n$, and $\alpha_c = 1 - \alpha_n - \alpha_p$, which clearly satisfy the power constraint at the transmitter.

The relay nodes receive

$$\begin{aligned}
\mathbf{y}_1 &= \sqrt{f_1 \alpha_c}\mathbf{x}_c + \sqrt{f_1 \alpha_p}\mathbf{x}_p + \sqrt{f_1 \alpha_n}\mathbf{x}_n + \sqrt{g_1}\mathbf{u} + \mathbf{z}_1 \\
\mathbf{y}_2 &= \sqrt{f_2 \alpha_c}\mathbf{x}_c + \sqrt{f_2 \alpha_p}\mathbf{x}_p + \sqrt{f_2 \alpha_n}\mathbf{x}_n + \sqrt{g_2}\mathbf{u} + \mathbf{z}_2.
\end{aligned} \tag{7.26}$$

Both nodes B_1 and B_2, first decodes $\mathbf{x}_c$ treating everything else as noise. Note that assuming randomized coding, this W_c can be decoded at B_1 and B_2 as long as $R_c < \overline{R}_{c,1}$ and $R_c < \overline{R}_{c,2}$, where

$$\begin{aligned}
\overline{R}_{c,1} &\triangleq \frac{1}{2}\log\left(\frac{1 + f_1 + g_1}{1 + f_1(\alpha_n + \alpha_p) + g_1}\right) \\
\overline{R}_{c,2} &\triangleq \frac{1}{2}\log\left(\frac{1 + f_2 + g_2}{1 + f_2(\alpha_n + \alpha_p) + g_2}\right) \\
&= \frac{1}{2}\log\left(\frac{1 + f_2 + g_2}{1 + \min(f_1, g_2 + 1) + g_2}\right) \\
&> \frac{1}{2}\log\left(\frac{1 + f_2 + g_2}{2(1 + g_2)}\right) \\
&= \frac{1}{2}\log\left(1 + \mathsf{SINR}_2\right) - \frac{1}{2}.
\end{aligned} \tag{7.27}$$

It is easy to show that $\overline{R}_{c,1} > \overline{R}_{c,2}$. Therefore, any common rate satisfying

$$R_c < \left(\frac{1}{2}\log\left(1 + \mathsf{SINR}_2\right) - \frac{1}{2}\right)^+ \tag{7.28}$$

is achievable for the relays.

Once $\mathbf{x}_c$ is decoded, B_1 can cancel it from its received signal, and decode $\mathbf{x}_p$ treating $\mathbf{x}_n$, $\mathbf{u}$ and $\mathbf{z}_1$ as noise. This can be done if and only if $R_p < \overline{R}_{p,1}$, where

$$\begin{aligned}
\overline{R}_{p,1} &\triangleq \frac{1}{2}\log\left(\frac{1 + f_1(\alpha_p + \alpha_n) + g_1}{1 + f_1\alpha_n + g_1}\right) \\
&> \frac{1}{2}\log\left(1 + \mathsf{SINR}_1\right) - \frac{1}{2}\log\left(1 + \mathsf{SINR}_2\right) - \frac{1}{2}.
\end{aligned} \tag{7.29}$$

Therefore, any private rate satisfying

$$R_p < \left(\frac{1}{2}\log\left(1 + \mathsf{SINR}_1\right) - \frac{1}{2}\log\left(1 + \mathsf{SINR}_2\right) - \frac{1}{2}\right)^+ \tag{7.30}$$

is achievable.

The operation at the relays in order to generate their transmitting signals is just to allocate proper power to their available different components. Note that the remaining (uncoded) parts of the signals will be also used for forming the transmitting signal. The relays nodes B_1 and X_2, send

$$\begin{aligned}
\mathbf{x}_1 &= \sqrt{\beta_c}\mathbf{x}_c + \sqrt{\beta_p}\mathbf{x}_p + \sqrt{\beta_n}\frac{\mathbf{y}_1 - \sqrt{f_1\alpha_c}\mathbf{x}_c - \sqrt{f_1\alpha_p}\mathbf{x}_p}{\sqrt{f_1\alpha_n + g_1 + 1}} \\
\mathbf{x}_2 &= \sqrt{\gamma_c}\mathbf{x}_c - \sqrt{\gamma_n}\frac{\mathbf{y}_2 - \sqrt{f_2\alpha_c}\mathbf{x}_c}{\sqrt{f_2\alpha_p + f_2\alpha_n + g_2 + 1}},
\end{aligned} \tag{7.31}$$

where again the power coefficients satisfy

$$\begin{aligned}
\beta_c + \beta_p + \beta_n &\le 1 \\
\gamma_c + \gamma_n &\le 1.
\end{aligned}$$

Finally the decoder receivers a noisy linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ over the multiple access channel. The signal received at the destination node can be written as

$$\begin{aligned}
\mathbf{y} &= \sqrt{h_1}\mathbf{x}_1 + \sqrt{h_2}\mathbf{x}_2 + \mathbf{z} \\
&= \left(\sqrt{h_1\beta_c} + \sqrt{h_2\gamma_c}\right)\mathbf{x}_c + \left(\sqrt{h_1\beta_p} - \sqrt{\frac{h_2\gamma_n f_2\alpha_p}{f_2(\alpha_p+\alpha_n)+g_2+1}}\right)\mathbf{x}_p \\
&\quad + \left(\sqrt{\frac{h_1\beta_n f_1}{f_1\alpha_n+g_1+1}} - \sqrt{\frac{h_2\gamma_n f_2}{f_2(\alpha_p+\alpha_n)+g_2+1}}\right)\sqrt{\alpha_n}\mathbf{x}_n \\
&\quad + \left(\sqrt{\frac{h_1\beta_n g_1}{f_1\alpha_n+g_1+1}} - \sqrt{\frac{h_2\gamma_n g_2}{f_2(\alpha_p+\alpha_n)+g_2+1}}\right)\mathbf{u} \\
&\quad + \left(\sqrt{\frac{h_1\beta_n}{f_1\alpha_n+g_1+1}}\mathbf{z}_1 - \sqrt{\frac{h_2\gamma_n}{f_2(\alpha_p+\alpha_n)+g_2+1}}\mathbf{z}_2 + \mathbf{z}\right).
\end{aligned} \tag{7.32}$$

In order to seek of simplicity, we denote the received power of $\mathbf{x}_c$, $\mathbf{x}_p$, $\mathbf{x}_n$ and $\mathbf{u}$ by P_c, P_p, P_n, and P_u, respectively. We also use N to denote the received noise power at the destination.

We can choose the power allocation coefficients arbitrarily. In particular, we can set them such that

$$\eta \triangleq \frac{h_1\beta_n(g_1+1)}{f_1\alpha_n+g_1+1} = \frac{h_2\gamma_n(g_2+1)}{f_2(\alpha_p+\alpha_n)+g_2+1}. \tag{7.33}$$

Note that

$$\eta < \frac{h_1\beta_n(g_1+1)}{g_1+1} = h_1\beta_n. \tag{7.34}$$

We also have $f_1\alpha_n \le f_1/\mathsf{SINR}_1 = g_1+1$, which implies

$$\eta \ge \frac{h_1\beta_n(g_1+1)}{2(g_1+1)} = h_1\beta_n/2. \tag{7.35}$$

Therefore, η is sandwiched by

$$h_1\beta_n/2 \le \eta < h_1\beta_n, \tag{7.36}$$

and similarly,

$$h_2\gamma_n/2 \le \eta < h_2\gamma_n. \tag{7.37}$$

Using this new notation, we can rewrite the power of neutralization part and the jamming signal as

$$\begin{aligned} P_n &= \alpha_n\left(\sqrt{\frac{h_1\beta_n f_1}{f_1\alpha_n+g_1+1}} - \sqrt{\frac{h_2\gamma_n f_2}{f_2(\alpha_p+\alpha_n)+g_2+1}}\right)^2 \\ &= \alpha_n\eta\left(\sqrt{\mathsf{SINR}_1} - \sqrt{\mathsf{SINR}_2}\right)^2, \end{aligned} \tag{7.38}$$

and

$$\begin{aligned} P_u &= \left(\sqrt{\frac{h_1\beta_n g_1}{f_1\alpha_n+g_1+1}} - \sqrt{\frac{h_2\gamma_n g_2}{f_2(\alpha_p+\alpha_n)+g_2+1}}\right)^2 \\ &= \eta\left(\sqrt{\frac{g_1}{g_1+1}} - \sqrt{\frac{g_2}{g_2+1}}\right)^2. \end{aligned} \tag{7.39}$$

Similarly, the total power of the noise would be

$$N = \frac{h_1\beta_n}{f_1\alpha_n+g_1+1} + \frac{h_2\gamma_n}{f_2(\alpha_p+\alpha_n)+g_2+1} + 1 = \eta\left(\frac{1}{g_1+1} + \frac{1}{g_2+1}\right) + 1. \tag{7.40}$$

The first task of the destination is to jointly decode W_c and W_p, treating $\mathbf{x}_n$ and $\mathbf{u}$ as noise. This can be done as long as

$$\begin{aligned} R_c &\le \frac{1}{\ell} I(\mathbf{y};\mathbf{x}_c|\mathbf{x}_p), \\ R_p &\le \frac{1}{\ell} I(\mathbf{y};\mathbf{x}_p|\mathbf{x}_c), \\ R_c+R_p &\le \frac{1}{\ell} I(\mathbf{y};\mathbf{x}_c\mathbf{x}_p). \end{aligned}$$

The total noise and interference for decoding $\mathbf{x}_c$ and $\mathbf{x}_p$ is upper bounded by

$$P_n+P_u+N \le (\sqrt{h_1\beta_n} + \sqrt{h_2\gamma_n})^2 + 1 < 8\eta + 1, \tag{7.41}$$

where we used (7.36) and (7.37) in the last inequality. Therefore, the decodability conditions for $\mathbf{x}_c$ and $\mathbf{x}_p$ can be further tightened, and rewritten as

$$R_p \leq \frac{1}{2}\log\left(\frac{h_1+1}{8\eta+1}\right) \tag{7.42}$$

$$R_c + R_p \leq \frac{1}{2}\log\left(\frac{h_1+h_2+1}{8\eta+1}\right). \tag{7.43}$$

Combining (7.28), (7.30), (7.42), and (7.43), gives the following achievable rate.

$$\begin{aligned} R_c + R_p \leq \min\Big\{ & \frac{1}{2}\log(1+\mathsf{SINR}_1) - 1, \\ & \frac{1}{2}\log(1+\mathsf{SINR}_2) + \frac{1}{2}\log(1+h_1) - \frac{1}{2}\log(1+8\eta) - \frac{1}{2}, \\ & \frac{1}{2}\log(1+h_1+h_2) - \frac{1}{2}\log(1+8\eta)\Big\}. \end{aligned} \tag{7.44}$$

One $\mathbf{x}_c$ and $\mathbf{x}_p$ are decoded, the destination node can remove them and decode $\mathbf{x}_n$. In order to do this, it has to treat $\mathbf{u}$ as noise. Note that, $\mathbf{x}_n$ is decodable as long as $R_n \leq \overline{R}_n$, where

$$\begin{aligned} \overline{R}_n &\triangleq \frac{1}{2}\log\left(1+\frac{P_n}{P_u+N}\right) \\ &= \frac{1}{2}\log\left(1+\frac{\alpha_n\eta\left(\sqrt{\mathsf{SINR}_1}-\sqrt{\mathsf{SINR}_2}\right)^2}{\eta\left(\sqrt{\frac{g_1}{g_1+1}}-\sqrt{\frac{g_2}{g_2+1}}\right)^2+\eta(\frac{1}{g_1+1}+\frac{1}{g_2+1})+1}\right) \end{aligned} \tag{7.45}$$

It is easy to show that

$$\left(\sqrt{\frac{g_1}{g_1+1}}-\sqrt{\frac{g_2}{g_2+1}}\right)^2 < \frac{1}{\min(g_1,g_2)+1} \tag{7.46}$$

and

$$\frac{1}{g_1+1}+\frac{1}{g_2+1} \leq \frac{2}{\min(g_1,g_2)+1}. \tag{7.47}$$

Therefore, we have

$$\overline{R}_n > \frac{1}{2}\log\left(1+\frac{\alpha_n\eta\left(\sqrt{\mathsf{SINR}_1}-\sqrt{\mathsf{SINR}_2}\right)^2}{\eta\left(\frac{3}{\min(g_1,g_2)+1}\right)+1}\right). \tag{7.48}$$

We can further show that $\overline{R}_n$ is lower bounded by

$$\min\left\{\frac{1}{2}\log\left(1+\frac{\alpha_n(\min(g_1,g_2)+1)\left(\sqrt{\mathsf{SINR}_1}-\sqrt{\mathsf{SINR}_2}\right)^2}{3}\right), \frac{1}{2}\log\left(1+\frac{\eta}{2}\right)\right\}. \tag{7.49}$$

In the next step, we can prove

$$(\min(g_1, g_2) + 1)\left(\sqrt{\mathsf{SINR}_1} - \sqrt{\mathsf{SINR}_2}\right)^2 \geq \frac{\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{3(g_1 + g_2 + 1)}. \tag{7.50}$$

Hence,

$$\overline{R}_n \geq \min\left\{\frac{1}{2}\log\left(1 + \frac{\alpha_n\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{9(g_1 + g_2 + 1)}\right), \frac{1}{2}\log(1 + 8\eta) - 2\right\}. \tag{7.51}$$

Summing up (7.44) and (7.51), we get the total achievable rate. It is worth mentioning that

$$\begin{aligned}
\frac{1}{2}\log(1+\mathsf{SINR}_1) &+ \frac{1}{2}\log\left(1 + \frac{\alpha_n\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{9}\right) \\
&\geq \frac{1}{2}\log\left(1 + \mathsf{SINR}_1 + \frac{\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{9(g_1 + g_2 + 1)}\right) \\
&\geq \frac{1}{2}\log\left(1 + 2\mathsf{SINR}_1 + \frac{\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{g_1 + g_2 + 1}\right) - \frac{1}{2}\log 18 \\
&\geq \frac{1}{2}\log\left(1 + \mathsf{SINR}_1 + \mathsf{SINR}_2 + \frac{\left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{g_1 + g_2 + 1}\right) - \frac{1}{2}\log 18 \\
&\geq \frac{1}{2}\log\left(1 + \frac{f_1 + f_2 + \left(\sqrt{f_1 g_2} - \sqrt{f_2 g_1}\right)^2}{g_1 + g_2 + 1}\right) - \frac{1}{2}\log 18
\end{aligned}$$

where in the first inequality we used the fact that

$$\begin{aligned}
(1 + \mathsf{SINR}_1)\alpha_n &= (1 + \mathsf{SINR}_1)\min\left(1, \frac{1}{\mathsf{SINR}_1}\right) \\
&= \min\left(1 + \mathsf{SINR}_1, 1 + \frac{1}{\mathsf{SINR}_1}\right) \geq 1.
\end{aligned} \tag{7.52}$$

$\square$

8 Discussion and Future Work

We focused on the approximation in network information theory problems in this thesis. While a complete and exact solution is not available for most of the problems in multi-user information theory, we can make progress by seeking approximate solutions with a guarantee on the gap from the optimal solution. This approach, in its general setting, suggests to simplify the problem, obtain optimal solution for the simplified problem, and translate optimal schemes as well as bounding techniques to the main problem. In fact, the insight obtained by analysis of the simplified problem can be used as a guideline in treating the original problem.

This approach yields interesting results in various long standing problems in network information theory, both in lossy data compression (*e.g.,* see Chapters 3 and 4, as well as [36]), and wireless communication (*e.g.,* see Chapter 6, and [15, 17]). Based on these, we hope that this technique can lead to meaningful results in other difficult communication problems. A more unified connection between such difficult problems and their simplified counterparts, obtained by different kinds of determination, is expected in near future.

8.1 Approximation in Lossy Data Compression

In the first part of the thesis is dedicated to provide an approximate characterization for the achievable rate region of the multiple description problem. We provided approximate characterizations of the individual-description rate-distortion function, as well as the achievable rate region, for the Gaussian SMD problem. This is done by combining two inter-connected parts: the derivation of an outer bound, and careful analysis of achievability schemes to generate inner bounds. Either component alone will not be able to provide these results. The new lower bound is obtained by generalizing Ozarow's well-known technique, and expanding the probability space of the original problem by many random variables with special structure among them.

The symmetric multi-level diversity (SMLD) coding problem, which can be understood as a lossless counterpart of the SMD problem, shed tremendous light on the geometric structure of the SMD rate-distortion region. We use the lossless SMLD coding rate region as a polytopic template for both inner and outer bounds for the SMD rate-distortion region.

A similar analysis for the general Gaussian multiple description problem (without posing symmetric distortion constraints), required understanding the asymmetric lossless counterpart. We formulated the asymmetric multilevel diversity coding problem, an asymmetric counterpart for the symmetric version of the problem. A complete characterization of the achievable rate region is provided for the three description case.

The optimal encoding scheme consists partitioning independent source sequences and applying linear network coding (binary `xor`) on the partitioned subsequences. It turns out that using such a strategy of jointly encoding the independent data streams is crucial, and the outer bound is not achievable without using it, in contrast to the symmetric problem, in which the source-separation coding is known to be optimal.

Using the intuition gained through analysis of the AMLD problem, we studied the Gaussian asymmetric three description problem. Inner and outer bounds for the achievable rate region are provided, and the difference between them is shown to be bounded by small universal constants.

Though the general asymmetric Gaussian MD rate distortion region is hard to characterize, it is satisfying to see that a simple coding architecture is almost optimal. With the increasing complexity of a source coding problem being considered in information theory literature, the complexity of its lossless counterpart increases as well, and it can become an increasingly dominant component of the overall problem. In this context, our work can be understood as an attempt to make explicit connection between the lossless source coding problem and its lossy counterpart.

8.2 Approximation in Multiple Unicast Wireless Networks

Interference management is perhaps the most fundamental open problem in wireless networks. The recent progress in (approximate) characterization of the interference channel capacity and the utility of the deterministic approach inspired the questions studied in the second part of the thesis.

Even though the interference-relay networks studied in this work were special, they revealed several new features needed for information transmission. In particular, the interference neutralization and network flow decomposition techniques were uncovered through the study of ZZ and ZS networks. We also saw the importance of using structured lattice codes for interference neutralization. Moreover, we believe that the neutralization technique is robust to channel uncertainties and one could get partial neutralization in such situations. This is a topic of ongoing work on this topic. We also believe that the outer bounding techniques developed in this work could have more general applicability in the wireless multiple-unicast problem.

8.3 Extensions

The main goal of studying various approximation results in network information theory is to bring them in a common framework and unify this approach. We hope this technique can lead to meaningful results in other difficult communication problems.

In both data compression and wireless network contexts, the connection between the original problem and the simplified one provides useful insight to the coding scheme and outer bounding proof technique. We expect that the connection between the deterministic (lossless) problems and their non-deterministic (lossy) counterparts can be used on other information theoretic problems, and the approach of using the former as a guideline in treating the latter to be a fruitful path.

In particular, the lossy distributed source coding is one of the most important problems in data compression which has been open for many years. Recent attempts in [35, 36] shows partial success in approximate characterization of the rate region, for some specific class of correlation between the sources. In these works, both ideas of transmitting bits above the distortion level (as in the MD problem) and neutralization of interference without post-processing (as in ZZ network) are utilized to approach the upper bound of the rate region. However, it is not clear if these are enough to arrive at an approximation for an arbitrarily correlated set of sources. A deep understanding of this problem can be one interesting direction for further studies in this context.

A deterministic perspective to analyze larger networks allows us to have a broader view of the significant phenomena happening in a network, instead of struggling with small effect complicated issues. Such an overview and insight obtained by analysis of the deterministic networks, can potentially be used to discover new schemes for transmission over such networks. The long time impact of this analysis can lead to performance improvement in various communication networks, such as cellular networks and ad-hoc networks. The relay-interference network is the simplest model studied in this context. The two-unicast problem in arbitrary layered wireless networks would be a natural next step arising out of our work, in order to analyze complex networks.

Another application of a deterministic study of information theoretic problems is to quantify how a simple scheme performs if applied in a difficult problem. In a recent work by Tian *et. al.* [81] it is shown that a single source-channel separation architecture in networks is optimal for the distributed network source coding problem and the general network unicast problem. Moreover, the separation approach is also approximately optimal for the general network multicast problem under the "difference" distortion measure, that is the distortion between the original and reconstructed source is only a function of their difference. These results are given in terms of rate-distortion trade-off, however, clearly they also hold for lossless coding. These results provide strong theoretical justification for using the separation approach in these scenarios.

Appendix for Chapter 2

A

A.1 Proof of Lemma 2.9

Note that $i \leq j \leq \mathscr{L}(\boldsymbol{v})$. Therefore, the decoding requirement for the decoder with access to $\boldsymbol{v}$ implies that the reconstructed sequence $\hat{U}_j^n(\boldsymbol{v})$ equals to U_j^n with high probability. Then

$$\begin{aligned}
H(\Gamma_{\boldsymbol{v}}|U_i^n) &\overset{(a)}{=} H(\Gamma_{\boldsymbol{v}}, \hat{U}_j^n(\boldsymbol{v})|U_i^n) \\
&= H(\Gamma_{\boldsymbol{v}}, U_j^n, \hat{U}_j^n(\boldsymbol{v})|U_i^n) - H(U_j^n|\Gamma_{\boldsymbol{v}}, \hat{U}_j^n(\boldsymbol{v}), U_i^n) \\
&\geq H(\Gamma_{\boldsymbol{v}}, U_j^n|U_i^n) - H(U_j^n|\hat{U}_j^n(\boldsymbol{v})) \\
&\overset{(b)}{=} H(\Gamma_{\boldsymbol{v}}|U_j^n) + H(U_j^n|U_i^n) - H(U_j^n|\hat{U}_j^n(\boldsymbol{v})), \qquad \text{(A.1)}
\end{aligned}$$

where (a) holds since $\hat{U}_j^n(\boldsymbol{v})$ is function of $\Gamma_{\boldsymbol{v}}$, for $j \leq \mathscr{L}(\boldsymbol{v})$, and (b) is due to the fact that U_i^n is a subsequence of U_j^n for $j \geq i$. The underlying independent distribution of U_i^n and U_j^n implies $H(U_j^n|U_i^n) = n(H_j - H_i)$. The last term in (A.1) can be upper bounded using the Fano's inequality [3] as

$$H(U_j^n|\hat{U}_j^n(\boldsymbol{v})) \leq h_b(P_e) + P_e \log(|\mathcal{U}_j^n| - 1) \leq 1 + ncP_e \qquad \text{(A.2)}$$

where $P_e = \Pr(\hat{U}_j^n(\boldsymbol{v}) \neq U_j^n)$ which tends to zero as n grows, $h_b(p)$ defined as

$$h_B(p) = -p\log_2(p) - (1-p)\log_2(1-p)$$

is the binary entropy function, and $c = \log|\mathcal{U}_j|$ is a constant. The proof is complete by setting $\delta_n = \frac{1}{n} + cP_e$. □

A.2 List of Corner points for Different Ordering in AMLD

In this section we list all of the corner points of the 3-description AMLD achievable rate region for different orderings and regimes. This would be useful to figure out the proper achievable scheme for each corner point, which is essentially similar to those presented in Section 2.4. For each ordering and regime listed in Table 2.2, we present a table, in which each row represents the intersecting hyper-planes, and the coordinates of the rate triple for one corner point.

It is shown that the supporting hyper-planes of the rate regions are characterized by (AMLD-1)-(AMLD-5). Note that each of (AMLD-1), (AMLD-2), and (AMLD-3) represent three constraints for different permutations of i, j, and k. Hence, the rate region would be characterized by 11 hyper-planes. To sake of brevity, we refer to these hyper-planes as in Table A.1. However, depending on the underlying ordering and regime, some of the constraint might be dominated by the others, and do not contribute in the characterization of the corner points.

Table A.1: Short-hand notations for the hyper-planes in 3-description AMLD.

Short-cut	Constraint
1	(AMLD-1) for $i = 1$
2	(AMLD-1) for $i = 2$
3	(AMLD-1) for $i = 3$
4	(AMLD-2) for $i = 1,\ j = 2$
5	(AMLD-2) for $i = 1,\ j = 3$
6	(AMLD-2) for $i = 2,\ j = 3$
7	(AMLD-3) for $i = 1$
8	(AMLD-3) for $i = 2$
9	(AMLD-3) for $i = 3$
10	(AMLD-4)
11	(AMLD-5)

A.2.1 Ordering $\mathscr{L}_1$: $\mathscr{L}_1(100) < \mathscr{L}_1(010) < \mathscr{L}_1(001) < \mathscr{L}_1(110) < \mathscr{L}_1(101) < \mathscr{L}_1(011) < \mathscr{L}_1(111)$

Table A.I.1: Ordering: $\mathscr{L}_1$, Regime: $h_3 \geq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4+h_6+h_7$	H_5
$\langle 2,4,8\rangle$	$H_1+h_3+h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1+h_3+h_4+h_7$	H_2	H_6
$\langle 3,5,10\rangle$	$H_1+h_4+h_5$	$H_3+h_6+h_7$	H_3
$\langle 3,6,10\rangle$	$H_1+h_3+h_7$	$H_2+h_4+h_5+h_6$	H_3
$\langle 4,7,10\rangle$	H_1+h_4	H_3	$H_3+h_5+h_6+h_7$
$\langle 4,8,10\rangle$	H_1+h_3	H_2+h_4	$H_3+h_5+h_6+h_7$
$\langle 5,7,10\rangle$	H_1+h_4	$H_3+h_6+h_7$	H_3+h_5
$\langle 6,8,10\rangle$	$H_1+h_3+h_7$	H_2+h_4	$H_3+h_5+h_6$

Table A.I.2: Ordering $\mathscr{L}_1$, Regime: $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4+h_6+h_7$	H_5
$\langle 2,4,8\rangle$	$H_1+h_3+h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1+h_3+h_4+h_7$	H_2	H_6
$\langle 3,5,9\rangle$	$H_1+h_4+h_5$	$H_2+h_4+h_5+h_6+h_7$	H_3
$\langle 3,6,9\rangle$	$H_1+h_4+h_5+h_7$	$H_2+h_4+h_5+h_6$	H_3
$\langle 4,7,10\rangle$	H_1+h_4	H_3	$H_3+h_5+h_6+h_7$
$\langle 4,8,10\rangle$	H_1+h_3	H_2+h_4	$H_3+h_5+h_6+h_7$
$\langle 5,7,10\rangle$	H_1+h_4	$H_3+h_6+h_7$	H_3+h_5
$\langle 5,9,10\rangle$	H_1+h_3	$H_3+h_6+h_7$	$H_2+h_4+h_5$
$\langle 6,8,10\rangle$	$H_1+h_3+h_7$	H_2+h_4	$H_3+h_5+h_6$
$\langle 6,9,10\rangle$	$H_1+h_3+h_7$	H_3+h_6	$H_2+h_4+h_5$

Table A.I.3: Ordering $\mathscr{L}_1$, Regime: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4+h_6+h_7$	H_5
$\langle 2,4,8\rangle$	$H_1+h_3+h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1+h_3+h_4+h_7$	H_2	H_6
$\langle 3,5,9\rangle$	$H_1+h_4+h_5$	$H_2+h_4+h_5+h_6+h_7$	H_3
$\langle 3,6,9\rangle$	$H_1+h_4+h_5+h_7$	$H_2+h_4+h_5+h_6$	H_3
$\langle 4,7,8,11\rangle$	$H_1+\frac{h_3+h_4}{2}$	$H_2+\frac{h_3+h_4}{2}$	$H_2+\frac{h_3+h_4}{2}+h_5+h_6+h_7$
$\langle 5,7,9,11\rangle$	$H_1+\frac{h_3+h_4}{2}$	$H_2+\frac{h_3+h_4}{2}+h_6+h_7$	$H_2+\frac{h_3+h_4}{2}+h_5$
$\langle 6,8,11\rangle$	$H_1+\frac{h_3+h_4}{2}+h_7$	$H_2+\frac{h_3+h_4}{2}$	$H_2+\frac{h_3+h_4}{2}+h_5+h_6$
$\langle 6,9,11\rangle$	$H_1+\frac{h_3+h_4}{2}+h_7$	$H_2+\frac{h_3+h_4}{2}+h_6$	$H_2+\frac{h_3+h_4}{2}+h_5$

A.2.2 Ordering $\mathscr{L}_2$: $\mathscr{L}_2(100) < \mathscr{L}_2(010) < \mathscr{L}_2(001) < \mathscr{L}_2(110) < \mathscr{L}_2(011) < \mathscr{L}_2(101) < \mathscr{L}_2(111)$

Table A.II.1: Ordering $\mathscr{L}_2$, Regime: $h_3 \geq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4 + h_7$	H_6
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_6 + h_7$	H_2	H_5
$\langle 3,5,10\rangle$	$H_1 + h_4 + h_5 + h_6$	$H_3 + h_7$	H_3
$\langle 3,6,10\rangle$	$H_1 + h_3 + h_6 + h_7$	$H_2 + h_4 + h_5$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4$	H_3	$H_3 + h_5 + h_6 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3$	$H_2 + h_4$	$H_3 + h_5 + h_6 + h_7$
$\langle 5,7,10\rangle$	$H_1 + h_4$	$H_3 + h_7$	$H_3 + h_5 + h_6$
$\langle 6,8,10\rangle$	$H_1 + h_3 + h_6 + h_7$	$H_2 + h_4$	$H_3 + h_5$

Table A.II.2: Ordering $\mathscr{L}_2$, Regime: $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4 + h_7$	H_6
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_6 + h_7$	H_2	H_5
$\langle 3,5,9\rangle$	$H_1 + h_4 + h_5 + h_6$	$H_2 + h_4 + h_5 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_5 + h_6 + h_7$	$H_2 + h_4 + h_5$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4$	H_3	$H_3 + h_5 + h_6 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3$	$H_2 + h_4$	$H_3 + h_5 + h_6 + h_7$
$\langle 5,7,10\rangle$	$H_1 + h_4$	$H_3 + h_7$	$H_3 + h_5 + h_6$
$\langle 5,9,10\rangle$	$H_1 + h_3 + h_6$	$H_3 + h_7$	$H_2 + h_4 + h_5$
$\langle 6,8,10\rangle$	$H_1 + h_3 + h_6 + h_7$	$H_2 + h_4$	$H_3 + h_5$
$\langle 6,9,10\rangle$	$H_1 + h_3 + h_6 + h_7$	H_3	$H_2 + h_4 + h_5$

Table A.II.3: Ordering $\mathscr{L}_2$, Regimes: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_4	H_7
$\langle 1,5,7\rangle$	H_1	$H_4 + h_7$	H_6
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_6 + h_7$	H_2	H_5
$\langle 3,5,9\rangle$	$H_1 + h_4 + h_5 + h_6$	$H_2 + h_4 + h_5 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_5 + h_6 + h_7$	$H_2 + h_4 + h_5$	H_3
$\langle 4,7,8,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7$
$\langle 5,7,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6$
$\langle 5,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_6$	$H_2 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_5$
$\langle 6,8,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+\frac{1}{2}h_4}{2} + h_5$

A.2.3 Ordering $\mathscr{L}_3$: $\mathscr{L}_3(100) < \mathscr{L}_3(010) < \mathscr{L}_3(001) < \mathscr{L}_3(101) < \mathscr{L}_3(110) < \mathscr{L}_3(011) < \mathscr{L}_3(111)$

Table A.III.1: Ordering $\mathscr{L}_3$, Regimes: $h_3 \geq h_4 + h_5$ and $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_5	$H_4 + h_6 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_4
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_7$	H_2	H_6
$\langle 3,5,7,10\rangle$	$H_1 + h_4$	$H_3 + h_5 + h_6 + h_7$	H_3
$\langle 3,6,10\rangle$	$H_1 + h_3 + h_7$	$H_2 + h_4 + h_5 + h_6$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4$	$H_3 + h_5$	$H_3 + h_6 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3$	$H_2 + h_4 + h_5$	$H_3 + h_6 + h_7$
$\langle 6,8,10\rangle$	$H_1 + h_3 + h_7$	$H_2 + h_4 + h_5$	$H_3 + h_6$

Table A.III.2: Ordering $\mathscr{L}_3$, Regime: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_5	$H_4 + h_6 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_4
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5$	H_2	H_7
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_7$	H_2	H_6
$\langle 3,5,9\rangle$	$H_1 + h_4$	$H_2 + h_4 + h_5 + h_6 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_7$	$H_2 + h_4 + h_5 + h_6$	H_3
$\langle 4,7,8,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_6 + h_7$
$\langle 5,7,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$
$\langle 6,8,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_6$
$\langle 6,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6$	$H_2 + \frac{h_3+h_4}{2}$

A.2.4 Ordering $\mathscr{L}_4$: $\mathscr{L}_4(100) < \mathscr{L}_4(010) < \mathscr{L}_4(001) < \mathscr{L}_4(101) < \mathscr{L}_4(011) < \mathscr{L}_4(110) < \mathscr{L}_4(111)$

Table A.IV.1: Ordering $\mathscr{L}_4$, Regimes: $h_3 \geq h_4 + h_5$ and $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_6	$H_4 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_4
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6$	H_2	$H_5 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_5
$\langle 3,5,7,10\rangle$	$H_1 + h_4$	$H_3 + h_5 + h_6 + h_7$	H_3
$\langle 3,6,8,10\rangle$	$H_1 + h_3 + h_6 + h_7$	$H_2 + h_4 + h_5$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4$	$H_3 + h_5 + h_6$	$H_3 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3 + h_6$	$H_2 + h_4 + h_5$	$H_3 + h_7$

Table A.IV.2: Ordering $\mathscr{L}_4$, Regime: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_6	$H_4 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_4
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6$	H_2	$H_5 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_5
$\langle 3,5,9\rangle$	$H_1 + h_4$	$H_2 + h_4 + h_5 + h_6 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_6 + h_7$	$H_2 + h_4 + h_5$	H_3
$\langle 4,7,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6$	$H_2 + \frac{h_3+h_4}{2} + h_7$
$\langle 4,8,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_6$	$H_2 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_7$
$\langle 5,7,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$
$\langle 6,8,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2}$

A.2.5 Ordering $\mathscr{L}_5$: $\mathscr{L}_5(100) < \mathscr{L}_5(010) < \mathscr{L}_5(001) < \mathscr{L}_5(011) < \mathscr{L}_5(110) < \mathscr{L}_5(101) < \mathscr{L}_5(111)$

Table A.V.1: Ordering $\mathscr{L}_5$, Regimes: $h_3 \geq h_4 + h_5$ and $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_5	H_7
$\langle 1,5,7\rangle$	H_1	$H_5 + h_7$	H_6
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5$	H_2	$H_4 + h_6 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_4
$\langle 3,5,10\rangle$	$H_1 + h_4 + h_5 + h_6$	$H_3 + h_7$	H_3
$\langle 3,6,8,10\rangle$	$H_1 + h_3 + h_5 + h_6 + h_7$	$H_2 + h_4$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4 + h_5$	H_3	$H_3 + h_6 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3 + h_5$	$H_2 + h_4$	$H_3 + h_6 + h_7$
$\langle 5,7,10\rangle$	$H_1 + h_4 + h_5$	$H_3 + h_7$	$H_3 + h_6$

Table A.V.2: Ordering $\mathscr{L}_5$, Regime: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_5	H_7
$\langle 1,5,7\rangle$	H_1	$H_5 + h_7$	H_6
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5$	H_2	$H_4 + h_6 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_4
$\langle 3,5,9\rangle$	$H_1 + h_4 + h_5 + h_6$	$H_2 + h_4 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_5 + h_6 + h_7$	$H_2 + h_4$	H_3
$\langle 4,7,8,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_6 + h_7$
$\langle 5,7,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2} + h_6$
$\langle 5,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5 + h_6$	$H_2 + \frac{h_3+h_4}{2} + h_7$	$H_2 + \frac{h_3+h_4}{2}$
$\langle 6,8,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2}$

A.2.6 Ordering $\mathscr{L}_6$: $\mathscr{L}_6(100) < \mathscr{L}_6(010) < \mathscr{L}_6(001) < \mathscr{L}_6(011) < \mathscr{L}_6(101) < \mathscr{L}_6(110) < \mathscr{L}_6(111)$

Table A.VI.1: Ordering $\mathscr{L}_6$, Regimes: $h_3 \geq h_4 + h_5$ and $h_4 \leq h_3 \leq h_4 + h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_6	$H_5 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_5
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6$	H_2	$H_4 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_4
$\langle 3,5,7,10\rangle$	$H_1 + h_4 + h_5$	$H_3 + h_6 + h_7$	H_3
$\langle 3,6,8,10\rangle$	$H_1 + h_3 + h_5 + h_6 + h_7$	$H_2 + h_4$	H_3
$\langle 4,7,10\rangle$	$H_1 + h_4 + h_5$	$H_3 + h_6$	$H_3 + h_7$
$\langle 4,8,10\rangle$	$H_1 + h_3 + h_5 + h_6$	$H_2 + h_4$	$H_3 + h_7$

Table A.VI.2: Ordering $\mathscr{L}_6$, Regime: $h_3 \leq h_4$

Planes	R_1	R_2	R_3
$\langle 1,4,7\rangle$	H_1	H_6	$H_5 + h_7$
$\langle 1,5,7\rangle$	H_1	H_7	H_5
$\langle 2,4,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6$	H_2	$H_4 + h_7$
$\langle 2,6,8\rangle$	$H_1 + h_3 + h_4 + h_5 + h_6 + h_7$	H_2	H_4
$\langle 3,5,9\rangle$	$H_1 + h_4 + h_5$	$H_2 + h_4 + h_6 + h_7$	H_3
$\langle 3,6,9\rangle$	$H_1 + h_4 + h_5 + h_6 + h_7$	$H_2 + h_4$	H_3
$\langle 4,7,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_6$	$H_2 + \frac{h_3+h_4}{2} + h_7$
$\langle 4,8,11\rangle$	$H_1 + \frac{V}{2} + h_5 + h_6$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2} + h_7$
$\langle 5,7,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5$	$H_2 + \frac{h_3+h_4}{2} + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$
$\langle 6,8,9,11\rangle$	$H_1 + \frac{h_3+h_4}{2} + h_5 + h_6 + h_7$	$H_2 + \frac{h_3+h_4}{2}$	$H_2 + \frac{h_3+h_4}{2}$

A.2.7 Ordering $\mathscr{L}_7$: $\mathscr{L}_7(100) < \mathscr{L}_7(010) < \mathscr{L}_7(110) < \mathscr{L}_7(001) < \mathscr{L}_7(101) < \mathscr{L}_7(011) < \mathscr{L}_7(111)$

Table A.VII.1: Ordering $\mathscr{L}_7$, Regime: $h_3 \geq h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7,10 \rangle$	H_1	H_3	H_7
$\langle 1,5,7,10 \rangle$	H_1	$H_3 + h_6 + h_7$	H_5
$\langle 2,4,8,10 \rangle$	$H_1 + h_3$	H_2	H_7
$\langle 2,6,8,10 \rangle$	$H_1 + h_3 + h_7$	H_2	H_6
$\langle 3,5,10 \rangle$	$H_1 + h_5$	$H_3 + h_6 + h_7$	H_4
$\langle 3,6,10 \rangle$	$H_1 + h_3 + h_7$	$H_2 + h_5 + h_6$	H_4

Table A.VII.2: Ordering $\mathscr{L}_7$, Regimes: $h_3 \leq h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7,10 \rangle$	H_1	H_3	H_7
$\langle 1,5,7,10 \rangle$	H_1	$H_3 + h_6 + h_7$	H_5
$\langle 2,4,8,10 \rangle$	$H_1 + h_3$	H_2	H_7
$\langle 2,6,8,10 \rangle$	$H_1 + h_3 + h_7$	H_2	H_6
$\langle 3,5,9 \rangle$	$H_1 + h_5$	$H_2 + h_5 + h_6 + h_7$	H_4
$\langle 3,6,9 \rangle$	$H_1 + h_5 + h_7$	$H_2 + h_5 + h_6$	H_4
$\langle 5,9,10 \rangle$	$H_1 + h_3$	$H_3 + h_6 + h_7$	$H_2 + h_4 + h_5$
$\langle 6,9,10 \rangle$	$H_1 + h_3 + h_7$	$H_3 + h_6$	$H_2 + h_4 + h_5$

A.2.8 Ordering $\mathscr{L}_8$: $\mathscr{L}_8(100) < \mathscr{L}_8(010) < \mathscr{L}_8(110) < \mathscr{L}_8(001) < \mathscr{L}_8(011) < \mathscr{L}_8(101) < \mathscr{L}_8(111)$

Table A.VIII.1: Ordering $\mathscr{L}_8$, Regime: $h_3 \geq h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7,10\rangle$	H_1	H_3	H_7
$\langle 1,5,7,10\rangle$	H_1	$H_3 + h_7$	H_6
$\langle 2,4,8,10\rangle$	$H_1 + h_3$	H_2	H_7
$\langle 2,6,8,10\rangle$	$H_1 + h_3 + h_6 + h_7$	H_2	H_5
$\langle 3,5,10\rangle$	$H_1 + h_5 + h_6$	$H_3 + h_7$	H_4
$\langle 3,6,10\rangle$	$H_1 + h_3 + h_6 + h_7$	$H_2 + h_5$	H_4

Table A.VIII.2: Ordering $\mathscr{L}_8$, Regimes: $h_3 \leq h_5$

Planes	R_1	R_2	R_3
$\langle 1,4,7,10\rangle$	H_1	H_3	H_7
$\langle 1,5,7,10\rangle$	H_1	$H_3 + h_7$	H_6
$\langle 2,4,8,10\rangle$	$H_1 + h_3$	H_2	H_7
$\langle 2,6,8,10\rangle$	$H_1 + h_3 + h_6 + h_7$	H_2	H_5
$\langle 3,5,9\rangle$	$H_1 + h_5 + h_6$	$H_2 + h_5 + h_7$	H_4
$\langle 3,6,9\rangle$	$H_1 + h_5 + h_6 + h_7$	$H_2 + h_5$	H_4
$\langle 5,9,10\rangle$	$H_1 + h_3 + h_6$	$H_3 + h_7$	$H_2 + h_4 + h_5$
$\langle 6,9,10\rangle$	$H_1 + h_3 + h_6 + h_7$	H_3	$H_2 + h_4 + h_5$

Appendix for Chapter 3

B

B.1 Proof of Lemma 3.12

Lemma. *For any given non-increasing distortion vector* $\boldsymbol{D}$*, its enhanced distortion vector* $\boldsymbol{D^*}$ *defined in* (3.9) *satisfies the following properties.*

- *It imposes more stringent distortion constraints,* i.e., $D^*_\alpha \le D_\alpha$, *for* $\alpha = 1, 2, \ldots, K$.
- *It is a non-increasing vector,* i.e., $D^*_1 \ge D^*_2 \ge \cdots \ge D^*_K$.
- *It induces a non-increasing* Φ *sequence,* i.e., $\Phi_1(D^*_1) \ge \Phi_2(D^*_2) \ge \cdots \ge \Phi_K(D^*_K)$.

Proof. Note that the condition $\Phi_\alpha(D_\alpha) > \Phi_{\alpha-1}(D^*_{\alpha-1})$ can be written as $D_\alpha > \frac{(\alpha-1)D^*_{\alpha-1}}{\alpha - D^*_{\alpha-1}}$. Moreover, the function $\Phi_\alpha(D)$ is an increasing function for $\alpha = 1, \ldots, K$ and $D \in (0, 1)$. Therefore, (3.9) can be rewritten as

$$\begin{aligned} D^*_1 &= D_1, \\ D^*_\alpha &= \min\left\{\frac{(\alpha-1)D^*_{\alpha-1}}{\alpha - D^*_{\alpha-1}}, D_\alpha\right\}, \quad \alpha = 2, 3, \ldots, K, \end{aligned} \tag{B.1}$$

which directly implies $D^*_\alpha \le D_\alpha \le 1$. Also, since $D^*_{\alpha-1} \le 1$, we have

$$D^*_\alpha \le \frac{(\alpha-1)D^*_{\alpha-1}}{\alpha - D^*_{\alpha-1}} \le D^*_{\alpha-1}, \tag{B.2}$$

which proves the second property. In order to show the last property, we again use monotonicity of $\Phi_\alpha(D)$ and (B.2), and write

$$\Phi_{\alpha-1}(D^*_{\alpha-1}) = \frac{(\alpha-1)D^*_{\alpha-1}}{1 - D^*_{\alpha-1}} = \Phi_\alpha\left(\frac{(\alpha-1)D^*_{\alpha-1}}{\alpha - D^*_{\alpha-1}}\right) \ge \Phi_\alpha(D^*_\alpha).$$

□

B.2 Proof of Lemma 3.30

Lemma. *Let $\boldsymbol{D}^*$ be the enhanced distortion vector associated with any valid distortion vector $\boldsymbol{D}$, then the random variables defined by (3.33), (3.34), (3.35) and (3.36) exist, and they satisfy*

$$\mathbb{E}\left[X - \mathbb{E}(X|Y_{|\boldsymbol{v}|,j}; j : v_j = 1)\right]^2 = D^*_{|\boldsymbol{v}|}, \qquad |\boldsymbol{v}| = 1, 2, \ldots, K-1,$$
$$\mathbb{E}\left[X - \mathbb{E}(X|\{Y_{K-1,j}; j \in \mathcal{I}_K\}, Y_K)\right]^2 = D^*_K.$$

Proof. First note that $\Phi_\alpha(D^*_\alpha) \leq \Phi_{\alpha-1}(D^*_{\alpha-1})$ for any enhanced distortion vector due to Lemma 3.12. , and it is then clear that (3.35) and (3.36) indeed specify a valid set of non-negative variances, and therefore, $Y_{\alpha,k}$ variables exist. In order to verify the condition in (3.37), notice that linear minimum mean square-error estimation of X based on $\{Y_{|\boldsymbol{v}|,j}; j : v_j = 1\}$ is

$$\hat{X} = \sum_{j:v_j=1} \beta_j Y_{|\boldsymbol{v}|,j}, \tag{B.3}$$

where β_j is the estimation coefficient for the j-th observation, and symmetry of the $Y_{|\boldsymbol{v}|,j}$'s implies $\beta_j = \beta_{j'} = \beta$ for $j : v_j = 1$ and $j' : v_{j'} = 1$. Moreover, we have

$$\mathbb{E}[(X - \hat{X})Y_{|\boldsymbol{v}|,j}] = 0, \qquad j : v_j = 1$$

which together with $\mathbb{E}[Y_{|\boldsymbol{v}|,j}Y_{|\boldsymbol{v}|,j'}] = 1$ and $\mathbb{E}[Y^2_{|\boldsymbol{v}|,j}] = 1 + \Phi_\alpha(D^*_{|\boldsymbol{v}|})$ imply $\beta(|\boldsymbol{v}| + \Phi_\alpha(D^*_{|\boldsymbol{v}|})) = 1$. Hence, the estimation error would be

$$\mathbb{E}[(X - \hat{X})]^2 = 1 - \beta|\boldsymbol{v}| = D^*_{|\boldsymbol{v}|}.$$

Finally, the condition in (3.38), we first recall from definition of Y_K in (3.34) that

$$\mathbb{E}[Y_K Y_{K-1,j}] = 0, \qquad j \in \mathcal{I}_K,$$
$$\mathbb{E}[Y_K X] = \frac{\Phi_{K-1}(D^*_{K-1})}{K + \Phi_{K-1}(D^*_{K-1})},$$
$$\mathbb{E}[Y_K^2] = \frac{\Phi_{K-1}(D^*_{K-1})}{K + \Phi_{K-1}(D^*_{K-1})} + \tilde{\sigma}^2_K. \tag{B.4}$$

Then, it is easy to show that

$$\mathbb{E}[X|\{Y_{K-1,j}; j \in \mathcal{I}_K\}, Y_K] = \frac{1}{K + \Phi_{K-1}(D^*_{K-1})} \sum_{j=1}^{K} Y_{K-1,j} + \frac{(1 - D^*_K)[\Phi_{K-1}(D^*_{K-1}) - \Phi_K(D^*_K)]}{\Phi_{K-1}(D^*_{K-1})} Y_K,$$

which together with (B.4) give us the estimation error claimed in the lemma. □

B.3 Proof of Lemma 3.32

Define $Z_a = N_a + N_b$ and $Z_b = N_b$. To prove the first statement, we consider the following chain of inequalities

$$\begin{aligned}
I(\Gamma_{\boldsymbol{v}}; Y_a^n) &= nh(Y_a) - h(Y_a^n|\Gamma_{\boldsymbol{v}}) \\
&= nh(Y_a) - h(X^n + Z_a^n|\Gamma_{\boldsymbol{v}}) \\
&= nh(Y_a) - h(X^n + Z_a^n - \hat{X}_{\boldsymbol{v}}^n|\Gamma_{\boldsymbol{v}}) \\
&\overset{(a)}{\geq} nh(Y_a) - h(X^n + Z_a^n - \hat{X}_{\boldsymbol{v}}^n) \\
&\overset{(b)}{\geq} nh(Y_a) - \sum_{i=1}^{n} h[X(i) + Z_a(i) - \hat{X}_{\boldsymbol{v}}(i)] \\
&\overset{(c)}{\geq} nh(Y_a) - \sum_{i=1}^{n} \frac{1}{2} \log \left\{ (2\pi e)\mathbb{E}[(X(i) + Z_a(i) - \hat{X}_{\boldsymbol{v}}(i))^2] \right\} \\
&= nh(Y_a) - \sum_{i=1}^{n} \frac{1}{2} \log \left[(2\pi e)(\mathbb{E}d(X(i), \hat{X}_{\boldsymbol{v}}(i)) + d_a) \right],
\end{aligned}$$

where $\hat{X}_{\boldsymbol{v}}^n$ is the reconstruction with descriptions $\Gamma_{\boldsymbol{v}} = \{\Gamma_i; i : v_i = 1\}$, and its i-th position is denoted as $\hat{X}_{\boldsymbol{v}}(i)$. The inequality (a) is because conditioning reduces entropy, (b) is because of the chain rule for differential entropy and the fact that conditioning reduces entropy, and (c) is because Gaussian distribution maximizes the differential entropy for a given second moment. Since $\log(\cdot)$ is a concave function, we have

$$\sum_{i=1}^{n} \frac{1}{2} \log \left[(2\pi e)(\mathbb{E}d(X(i), \hat{X}\boldsymbol{v}(i)) + d_a) \right] \leq \frac{n}{2} \log \left[2\pi e \left(\mathbb{E}d(X^n, \hat{X}_{\boldsymbol{v}}^n) + d_a \right) \right].$$

It follows

$$\begin{aligned}
I(\Gamma_{\boldsymbol{v}}; Y_a^n) &\geq nh(Y_a) - \frac{n}{2} \log \left[2\pi e \left(\mathbb{E}d(X^n, \hat{X}_{\boldsymbol{v}}^n) + d_a \right) \right] \\
&\geq nh(Y_a) - \frac{n}{2} \log \left[2\pi e (D_{|\boldsymbol{v}|} + d_a) \right] \\
&= \frac{n}{2} \log \frac{1 + d_a}{D_{|\boldsymbol{v}|} + d_a},
\end{aligned}$$

which is the first claim in the lemma.

To prove the second claim, we write the following

$$I(\Gamma_{\boldsymbol{v}}; Y_b^n) - I(\Gamma_{\boldsymbol{v}}; Y_a^n) = nh(Y_b) - nh(Y_a) + h(Y_a^n|\Gamma_{\boldsymbol{v}}) - h(Y_b^n|\Gamma_{\boldsymbol{v}}).$$

For the latter two terms, we have

$$\begin{aligned}
h(Y_a^n|\Gamma_{\boldsymbol{v}}) - h(Y_b^n|\Gamma_{\boldsymbol{v}}) &\overset{(a)}{=} h(Y_a^n|\Gamma_{\boldsymbol{v}}) - h(Y_b^n|N_a^n, \Gamma_{\boldsymbol{v}}) \\
&\overset{(b)}{=} h(Y_a^n|\Gamma_{\boldsymbol{v}}) - h(Y_a^n|N_a^n, \Gamma_{\boldsymbol{v}}) \\
&= I(Y_a^n; N_a^n|\Gamma_{\boldsymbol{v}}),
\end{aligned}$$

where (a) is because N_a^n is independent of Y_b^n and $\Gamma_{\boldsymbol{v}}$; (b) is by the definition of Y_a. Continuing the chain of inequalities, we have

$$\begin{aligned}
I(Y_a^n; N_a^n|\Gamma_{\boldsymbol{v}}) &\overset{(a)}{=} h(N_a^n) - h(N_a^n|X^n + N_a^n + N_b^n, \Gamma_{\boldsymbol{v}}) \\
&= h(N_a^n) - h(N_a^n|X^n + N_b^n + N_a^n, \hat{X}_{\boldsymbol{v}}^n, \Gamma_{\boldsymbol{v}}) \\
&\overset{(b)}{\geq} h(N_a^n) - h(N_a^n|X^n - \hat{X}_{\boldsymbol{v}}^n + N_a^n + N_b^n) \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} \Big[h(N_a(i)) - h\Big(N_a(i)|X(i) - \hat{X}_{\boldsymbol{v}}(i) + N_a(i) + N_b(i)\Big)\Big] \\
&= \sum_{i=1}^{n} I\Big(N_a(i); X(i) - \hat{X}_{\boldsymbol{v}}(i) + N_b(i) + N_a(i)\Big) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} \frac{1}{2} \log \frac{\mathbb{E}d(X(i), \hat{X}_{\boldsymbol{v}}(i)) + d_a}{\mathbb{E}d(X(i), \hat{X}_{\boldsymbol{v}}(i)) + d_b} \\
&\overset{(e)}{\geq} \frac{n}{2} \log \frac{D_{|\boldsymbol{v}|} + d_a}{D_{|\boldsymbol{v}|} + d_b},
\end{aligned}$$

where (a) is because N_a is independent of $\Gamma_{\boldsymbol{v}}$; (b) is because conditioning reduces entropy; (c) is by applying the chain rule, and the facts that N_a^n is an i.i.d. sequence and conditioning reduces entropy; (d) is by applying the mutual information game result (see page 263, [3], as well as [80]) that Gaussian noise is the worst additive noise under a variance constraint, and taking $N_a(i)$ as channel input; finally (e) is due to the convexity and monotonicity of $\log \frac{x+d_a}{x+d_b}$ in $x \in (0, \infty)$ when $d_a \geq d_b \geq 0$. This completes the proof for the second claim.

A similar line of argument was used in [28] to derive a sum rate lower bound for a system with two levels of distortion constraints. However, Lemma 3.32 generalizes that result since there exists only one auxiliary random variable in the setting of [28], but there are two auxiliary random variables Y_a and Y_b in the current setting. □

B.4 Proof of Corollary 3.33

To facilitate discussion, define the following index set of loose constraints

$$C_L = \{\alpha : D_\alpha^* < D_\alpha\},$$

and it follows that $C_L^c = \mathcal{I}_K \backslash C_L$; note that $1 \in C_L^c$. For a given $\alpha \in C_L$, define $N(\alpha)$ as the index of lower neighboring distortion constraint to α that is not loose, *i.e.*,

$$N(\alpha) = \max_{\substack{k<\alpha \\ k \in C_L^c}} k.$$

We may distinguish the following two cases for clarity.

Case I. First consider the case when the distortion vector is given such that it satisfies the conditions

$$\Phi_{\alpha-1}(D_{\alpha-1}) \geq \Phi_\alpha(D_\alpha), \tag{B.5}$$

for all $\alpha = 2, 3, ..., K$, where we take $D_0 \triangleq 1$. Note this implies $D_\alpha = D^*_\alpha$, for $\alpha = 1, 2, ..., K$, and hence $C_L = \emptyset$. In this case we choose $\check{d}_\alpha = \Phi_\alpha(D_\alpha)$, for $\alpha = 1, 2, ..., K-1$, which is clearly valid. Evaluating $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$ for $\boldsymbol{d} = \check{\boldsymbol{d}}$, we have

$$\begin{aligned}
\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \check{\boldsymbol{d}}) &= \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{(1+\check{d}_\alpha)(D_\alpha+\check{d}_{\alpha-1})}{(1+\check{d}_{\alpha-1})(D_\alpha+\check{d}_\alpha)} \\
&= \frac{1}{2}\sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\log\left[\frac{D_{\alpha-1}}{D_\alpha}\frac{1+(\alpha-1)D_\alpha}{1+(\alpha-2)D_{\alpha-1}}\frac{\alpha-1-D_\alpha+\frac{D_\alpha}{D_{\alpha-1}}}{\alpha+1-D_\alpha}\right] \\
&\quad + \frac{1}{2}\log\frac{1}{D_1(2-D_1)} + \frac{1}{2}\log\left[\frac{D_{K-1}}{D_K}\frac{K-1-D_K+\frac{D_K}{D_{K-1}}}{1+(K-2)D_{K-1}}\right] \\
&= \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} + \frac{1}{2}\sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\log\left[\frac{1+(\alpha-1)D_\alpha}{1+(\alpha-2)D_{\alpha-1}}\frac{\alpha-1-D_\alpha+\frac{D_\alpha}{D_{\alpha-1}}}{\alpha+1-D_\alpha}\right] \\
&\quad + \frac{1}{2}\log\frac{1}{(2-D_1)} + \frac{1}{2}\log\left[\frac{K-1-D_K+\frac{D_K}{D_{K-1}}}{1+(K-2)D_{K-1}}\right] \\
&= \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} + \frac{1}{2}\log\frac{1}{(2-D_1)} \\
&\quad + \frac{1}{2}\sum_{\alpha=2}^{K-1}[\frac{1}{\alpha}-\frac{1}{\alpha+1}]\log(1+(\alpha-1)D_\alpha) \\
&\quad + \frac{1}{2}\sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\log\frac{\alpha-1-D_\alpha+\frac{D_\alpha}{D_{\alpha-1}}}{\alpha+1-D_\alpha} + \frac{1}{2}\log\left(K-1-D_K+\frac{D_K}{D_{K-1}}\right) \\
&\overset{(a)}{\geq} \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} + \frac{1}{2}\log\frac{1}{(2-D_1)} + \frac{1}{2}\sum_{\alpha=2}^{K-1}\frac{1}{\alpha}\log\frac{\alpha-1}{\alpha+1-D_\alpha} \\
&\quad + \frac{1}{2}\log(K-1) \\
&= \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log(\alpha-D_{\alpha-1}) + \frac{K}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1) \\
&\geq \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D_{\alpha-1}}{D_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log\alpha + \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1) \\
&= \frac{1}{2}\sum_{\alpha=1}^{K}\frac{1}{\alpha}\log\frac{D^*_{\alpha-1}}{D^*_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha-1}\log\alpha + \frac{1}{2}\sum_{\alpha=2}^{K}\frac{1}{\alpha}\log(\alpha-1),
\end{aligned} \tag{B.6}$$

where in (a) we used $\frac{D_\alpha}{D_{\alpha-1}} \geq D_\alpha$, and omitted the third term which is positive. Thus if (B.5) holds, we have

$$R_{\mathsf{SMD}}(\boldsymbol{D}) \geq \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \check{\boldsymbol{d}}) \geq \check{\underline{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*). \tag{B.7}$$

It is also clear that since $\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}) \geq \check{\underline{R}}_{\mathsf{SMD}}(\boldsymbol{D}^*)$, since it is optimized for $\boldsymbol{d}$.

Case II. For the case when (B.5) does not hold, then we choose $\check{d}_\alpha = \Phi_\alpha(D_\alpha)$, for $\alpha \in C_L^c$ as before; however for any $\alpha \in C_L$, we choose $\check{d}_\alpha = \Phi_{N(\alpha)}(D_{N(\alpha)})$. Note that with such a choice, we have $d_\alpha = d_{\alpha-1}$ for $\alpha \in C_L$, and therefore,

$$\log \frac{(1+d_\alpha)(D_\alpha + d_{\alpha-1})}{(1+d_{\alpha-1})(D_\alpha + d_\alpha)} = 0, \qquad \alpha \in C_L.$$

If we replace D_α with D_α^* in the left hand side of the above equation, the equality still holds. Moreover, it is easy to verify that this choice of $\check{d}$ is equivalent to choosing

$$d_\alpha = \Phi_\alpha(D_\alpha^*), \qquad \alpha \in \mathcal{I}_{K-1}.$$

Thus by using this particular choice of $(\check{d}_1, \check{d}_2, ..., \check{d}_{K-1})$, we have

$$\begin{aligned} \underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \check{\boldsymbol{d}}) &= \frac{1}{2}\sum_{\alpha=1}^{K} \frac{1}{\alpha} \log \frac{(1+\check{d}_\alpha)(D_\alpha + \check{d}_{\alpha-1})}{(1+\check{d}_{\alpha-1})(D_\alpha + \check{d}_\alpha)} \\ &= \frac{1}{2}\sum_{\alpha=1}^{K} \frac{1}{\alpha} \log \frac{(1+\check{d}_\alpha)(D_\alpha^* + \check{d}_{\alpha-1})}{(1+\check{d}_{\alpha-1})(D_\alpha^* + \check{d}_\alpha)}, \end{aligned}$$

and the exact same derivation holds as in the case when (B.5) holds, with D_α^* replacing D_α. This completes the proof. □

B.5 Proof of Theorem 3.41

We first give another characterization for $\hat{\mathcal{R}}_{\mathsf{SMD}}^{\mathsf{PPR}}(\boldsymbol{Y})$.

Definition B.1. *Let $\tilde{\mathcal{R}}_{\mathsf{SMD}}^{\mathsf{PPR}}(\boldsymbol{Y})$ be the set of all $\boldsymbol{R} \geq 0$ such that for all $\boldsymbol{A} \in \mathbb{R}_+^K$ but $\boldsymbol{A} \neq 0$*

$$\boldsymbol{A} \cdot \boldsymbol{R} \geq \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \tilde{H}_\alpha(\boldsymbol{Y}), \tag{B.8}$$

where $f_\alpha(\boldsymbol{A})$ is defined in (3.5) of Section 3.2, and $\tilde{H}_\alpha(\boldsymbol{Y})$ is defined in (3.3) and (3.4).

The following theorem extracted from [30].

Theorem B.2 ([30] Theorem 2)**.**
$$\tilde{\mathcal{R}}_{\mathsf{SMD}}^{\mathsf{PPR}}(\boldsymbol{Y}) = \hat{\mathcal{R}}_{\mathsf{SMD}}^{\mathsf{PPR}}(\boldsymbol{Y}).$$

Remark B.3. *In the definition of $\tilde{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})$, the requirement that $\boldsymbol{R} \geq 0$ can be safely removed without loss of generality when $\tilde{H}_\alpha(\boldsymbol{Y}) \geq 0$. To see this, let $\boldsymbol{A} = (1, 0, ..., 0)$, then (B.8) reduces to $R_1 \geq \tilde{H}_1(\boldsymbol{Y})$ by applying Lemma 3.11.*

In order to prove Theorem 3.41, consider the following. For a fixed set of (generalized symmetric) random variables $\{\{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K\}$, since both $\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$ and $\hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ are convex, they can be characterized by the supporting hyper-planes. As such if we can prove that for any $\boldsymbol{A} \in \mathbb{R}^K_+$ and $\boldsymbol{A} \neq 0$,

$$\min_{\boldsymbol{R} \in \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})} \boldsymbol{A} \cdot \boldsymbol{R} \leq \min_{\boldsymbol{R} \in \hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})} \boldsymbol{A} \cdot \boldsymbol{R}, \tag{B.9}$$

holds, then it follows that $\hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y}) \subseteq \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$, and hence, $\hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})$ is an achievable region.

By Theorem B.2, we have

$$\min_{\boldsymbol{R} \in \hat{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{Y})} \boldsymbol{A} \cdot \boldsymbol{R} = \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \tilde{H}_\alpha(\boldsymbol{Y}). \tag{B.10}$$

Thus it suffices to prove that for any $\boldsymbol{A} \in \mathbb{R}^K_+$ and $\boldsymbol{A} \neq 0$, there always exists a rate vector in the achievable rate region that satisfy (B.10) with equality, *i.e.*, there exists $\boldsymbol{R} \in \mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$ such that

$$\boldsymbol{A} \cdot \boldsymbol{R} = \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \tilde{H}_\alpha(\boldsymbol{Y}). \tag{B.11}$$

This would imply (B.9), which further implies the result of Theorem 3.41.

In order to show (B.11), first fix a set of (generalized symmetric) random variables $\{\{Y_{\alpha,k}; \alpha \in \mathcal{I}_{K-1}, k \in \mathcal{I}_K\}, Y_K\}$. Then, for a given $\boldsymbol{A} \geq 0$, let l_α be the non-negative integer defined in Lemma 3.7 for the α-level. For any $\alpha \in \mathcal{I}_K$, set

$$R_{\alpha,k} = \begin{cases} 0 & \text{if} \quad 1 \leq k \leq l_\alpha \\ \frac{\tilde{H}_\alpha}{\alpha - l_\alpha} & \text{if} \quad l_\alpha + 1 \leq k \leq K. \end{cases} \tag{B.12}$$

In fact, the rate $R_{\alpha,k}$ will be the rate assigned to the α-th layer for the k-th description. Denote $(R_{\alpha,1}, R_{\alpha,2}, ..., R_{\alpha,K})$ by $\boldsymbol{R_\alpha}$. It is clear from the original PPR multilayer scheme [26] that if each of the description has rate approximately $\tilde{H}_\alpha/\alpha$ at the α-th level, then any of the α descriptions can guarantees decoding with high probability. However, because the first l_α descriptions are not given any rate for the α-th layer in (B.12), this can not be achieved directly without proper coding.

The generalized coding scheme is by combining the original PPR multilayer scheme with proper MDS channel codes. The PPR multilayer scheme is still used as the main encoding step, and let us denote the codeword (the output index written in a large enough appropriate alphabet) for the α-th level for description k as $C_{\alpha,k}$, for $\alpha \in \mathcal{I}_K$. A post-coding packaging step is now added at the α-th layer as follows. The last $K - l_\alpha$ codeword indices are written in the descriptions as in the original scheme. Each of the first l_α codeword indices $C_{\alpha,k}$, $k = 1, 2, ..., l_\alpha$ is encoded by a $(K - l_\alpha, \alpha - l_\alpha)$ MDS code, and each of the resulting codeword (index) is written into one of the last $K - l_\alpha$ description. This results in an additional rate $\tilde{H}_\alpha/\alpha(\alpha - l_\alpha)$ in each description. Note that since $l_\alpha \leq \alpha - 1$, the above MDS code rate is always well defined. It is clear that the rate of the k-th description, $k > l_\alpha$, for the α-th layer is

$$R_{\alpha,k} = \frac{\tilde{H}_\alpha}{\alpha} + \frac{\tilde{H}_\alpha}{\alpha(\alpha - l_\alpha)} l_\alpha = \frac{\tilde{H}_\alpha}{\alpha - l_\alpha},$$

as we claimed.

Consider a decoder $\boldsymbol{v}$ with $|\boldsymbol{v}| = k$ at which the set of descriptions $\Gamma_{\boldsymbol{v}}$ is available, and let $\alpha \in \mathcal{I}_k$ be a specific level. The pre-decoding unpackaging procedure is as follows. Define $\underline{\boldsymbol{v}}$ and $\overline{\boldsymbol{v}}$ as

$$\underline{v}_i = \begin{cases} v_i & \text{if } i \leq l_\alpha \\ 0 & \text{if } i > l_\alpha, \end{cases}$$

and

$$\overline{v}_i = \begin{cases} 0 & \text{if } i \leq l_\alpha \\ v_i & \text{if } i > l_\alpha, \end{cases}$$

It is clear that $\boldsymbol{v} = \underline{\boldsymbol{v}} + \overline{\boldsymbol{v}}$, and $k = |\underline{\boldsymbol{v}}| + |\overline{\boldsymbol{v}}|$.

The descriptions $\Gamma_{\overline{\boldsymbol{v}}}$ clearly can recover their respective codewords, *i.e.*, $C_{\alpha,i}$ for i such that $\overline{v}_i = 1$. However, since $|\underline{\boldsymbol{v}}| \leq l_\alpha$, we have also $|\overline{\boldsymbol{v}}| = k - |\underline{\boldsymbol{v}}| \geq \alpha - |\underline{\boldsymbol{v}}| \geq \alpha - l_\alpha$ pieces of the MDS encoded $C_{\alpha,i}$ for $i \in \mathcal{I}_{l_\alpha}$, which can be correctly decoded by the property of the MDS code. Therefore, we can recover all corresponding codewords $C_{\alpha,i}$ for all i satisfying $v_i = 1$. This holds true for all $\alpha = 1, 2, \ldots, k$, and then the main decoding step in the PPR multilayer scheme can be applied.

We remark here that the decoding can be easily improved, because if $|\underline{\boldsymbol{v}}| < l_\alpha$, there is additional information that the main decoding step is not utilizing. However the above simple procedure suffices for proving the current theorem.

It remains to show (B.11) is true with the given rate vector, the proof of which follows closely the step in [30] for the proof of Theorem B.2. Let $\{c(\boldsymbol{v})\}$ be an optimal α-resolution for $\boldsymbol{A}$. We have

$$\boldsymbol{A} \cdot \boldsymbol{R_\alpha} = \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})(\boldsymbol{v} \cdot \boldsymbol{R_\alpha}) + (\boldsymbol{A} - \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})\boldsymbol{v}) \cdot \boldsymbol{R_\alpha}.$$

By Lemma 3.6, for any $\boldsymbol{v}$ where $|\boldsymbol{v}| = \alpha$ such that $c_\alpha(\boldsymbol{v}) > 0$, $v_i = 1$ for $i = 1, 2, \ldots, l_\alpha$; moreover, exactly $\alpha - l_\alpha$ of the remaining components are 1's. Since the first l_α components of $\boldsymbol{R_\alpha}$ are 0's, and the remaining components are equal, we have

$$\boldsymbol{v} \cdot \boldsymbol{R_\alpha} = (\alpha - l_\alpha)\frac{\tilde{H}_\alpha}{\alpha - l_\alpha} = \tilde{H}_\alpha, \qquad \text{for } \boldsymbol{v} : c_\alpha(\boldsymbol{v}) > 0.$$

It follows that

$$\sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})(\boldsymbol{v} \cdot \boldsymbol{R_\alpha}) = \tilde{H}_\alpha \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v}) = f_\alpha(\boldsymbol{A})\tilde{H}_\alpha.$$

Since the vector

$$\boldsymbol{A} - \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})\boldsymbol{v} = \boldsymbol{A} - \breve{\boldsymbol{A}}$$

has zeros in the last $K - l_\alpha$ components, and $\boldsymbol{R_\alpha}$ has zeros in the complement positions, we have

$$(\boldsymbol{A} - \sum_{\boldsymbol{v} \in \Omega_K^\alpha} c_\alpha(\boldsymbol{v})\boldsymbol{v}) \cdot \boldsymbol{R_\alpha} = 0.$$

It follows

$$\boldsymbol{A} \cdot \boldsymbol{R_\alpha} = f_\alpha(\boldsymbol{A})\tilde{H}_\alpha.$$

Summing over $\alpha \in \mathcal{I}_K$ now completes the proof since $R_i = \sum_{\alpha=1}^{K} R_{\alpha,i}$. $\square$

B.6 Proof of Corollary 3.44

Corollary. *For the Gaussian source, $\underline{\check{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ establishes an outer bound for the* SMD *rate-distortion region,* i.e.,

$$\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}) \subseteq \underline{\check{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*}). \tag{B.13}$$

Proof. The proof follows that for Corollary 3.33, however we directly use $\boldsymbol{D^*}$ to replace $\boldsymbol{D}$. Let $\check{d}_\alpha = \Phi_\alpha(D_\alpha^*)$ for $\alpha = 1, 2, ..., K-1$. Let $\mathbf{R} = (R_1, \ldots, R_K) \in$

$\mathcal{R}_{\mathsf{SMD}}(D)$ be an achievable rate tuple. We start with the result of Theorem 3.23.

$$
\begin{aligned}
\sum_{i=1}^{K} A_i R_i &\geq \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{(1+\check{d}_\alpha)(D_\alpha+\check{d}_{\alpha-1})}{(1+\check{d}_{\alpha-1})(D_\alpha+\check{d}_\alpha)} \\
&\overset{(a)}{\geq} \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{(1+\check{d}_\alpha)(D^*_\alpha+\check{d}_{\alpha-1})}{(1+\check{d}_{\alpha-1})(D^*_\alpha+\check{d}_\alpha)} \\
&= \frac{1}{2}\sum_{\alpha=2}^{K-1} f_\alpha(\boldsymbol{A}) \left[\log \frac{D^*_{\alpha-1}}{D^*_\alpha} + \log \frac{1+(\alpha-1)D^*_\alpha}{1+(\alpha-2)D^*_{\alpha-1}} + \log \frac{\alpha-1-D^*_\alpha+\frac{D^*_\alpha}{D^*_{\alpha-1}}}{1+\alpha-D^*_\alpha}\right] \\
&\quad + \frac{1}{2} f_1(\boldsymbol{A}) \log \frac{1}{D^*_1(2-D^*_1)} + \frac{1}{2} f_K(\boldsymbol{A}) \log \left[\frac{D^*_{K-1}}{D^*_K} \frac{K-1-D^*_K+\frac{D^*_K}{D^*_{K-1}}}{1+(K-2)D^*_{K-1}}\right] \\
&\overset{(b)}{\geq} \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha} + \frac{1}{2}\sum_{\alpha=2}^{K-1}\Big[f_{\alpha-1}(\boldsymbol{A}) - f_\alpha(\boldsymbol{A})\Big] \log(1+(\alpha-1)D^*_\alpha) \\
&\quad - \frac{1}{2}\sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D^*_{\alpha-1}) + \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha-1) \\
&\overset{(c)}{\geq} \frac{1}{2}\sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha} - \frac{1}{2}\sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D^*_{\alpha-1}) \\
&\quad + \frac{1}{2}\sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha-1), \qquad \text{(B.14)}
\end{aligned}
$$

where (a) follows from the fat that $(x+\check{d}_{\alpha-1})/(x+\check{d}_\alpha)$ is an increasing function and the fact that $D_\alpha \geq D^*_\alpha$ due to Lemma 3.12; (b) is true because $\frac{D^*_\alpha}{D^*_{\alpha-1}} \geq D^*_\alpha$; and (c) we omitted the second term, because Lemma 3.10 implies $f_\alpha(\boldsymbol{A}) \leq \frac{\alpha-1}{\alpha} f_{\alpha-1}(\boldsymbol{A}) \leq f_{\alpha-1}(\boldsymbol{A})$. This completes the proof. □

B.7 Proof of Corollary 3.45

The key idea is the following: though we have an uncountable number of supporting hyper-planes to characterize $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$, if there exists a set $\mathcal{S}_R \subseteq \check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ with finite number of elements, such that for each $\boldsymbol{A}$, inequality (3.50) can be satisfied with equality for some element in $\mathcal{S}_R$, then $\check{\underline{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D^*})$ is a polytope. In the following we show the existence of such a finite set.

Since $\log(\alpha - D^*_{\alpha-1}) \geq 0$ for $\alpha \geq 2$, we can construct a set of independent fictitious sources $U_1, U_2, ..., U_K$, such that

$$
H(U_\alpha) = \frac{1}{2}\log(\alpha+1-D^*_\alpha), \quad \alpha = 1, 2, ..., K-1,
$$

and $H(U_K) = 0$. The MLD coding rate region for this K-source can be equivalently given in two forms, as implied by Theorem B.2, with $\tilde{H}_\alpha(\boldsymbol{Y})$ replaced by $H(U_\alpha)$. Since the rate region of this MLD coding problem is clearly a polytope, there exists a finite set of rate vectors, denoted as $\mathcal{S}_r$, such that for any $\boldsymbol{A}$, there exists at least one rate vector $(r_1, r_2, ..., r_K) \in \mathcal{S}_r$, such that

$$\sum_{i=1}^{K} A_i r_i = \frac{1}{2} \sum_{\alpha=2}^{K} f_{\alpha-1}(\boldsymbol{A}) \log(\alpha - D^*_{\alpha-1})$$

Now define $\grave{R}_i = R_i + r_i$, $i = 1, 2, ..., K$, and consequently (3.50) reduces to the condition that

$$\sum_{i=1}^{K} A_i \grave{R}_i \geq \frac{1}{2} \sum_{\alpha=1}^{K} f_\alpha(\boldsymbol{A}) \log \frac{D^*_{\alpha-1}}{D^*_\alpha} + \frac{1}{2} \sum_{\alpha=2}^{K} f_\alpha(\boldsymbol{A}) \log(\alpha - 1), \qquad \text{(B.15)}$$

We can again define a set of fictitious independent sources $W_1, W_2, ..., W_K$, such that

$$H(W_\alpha) = \log \frac{D^*_{\alpha-1}}{D^*_\alpha} + \log(\alpha - 1), \quad \alpha = 2, 3, ..., K,$$

and

$$H(W_1) = \log \frac{1}{D^*_\alpha}.$$

Now we would like to apply Theorem B.2 to assert (B.15) is in fact a characterization of the MLD coding rate region for this source, however one technicality has to be addressed first. Recall that $\boldsymbol{R}$ is not constrained to be non-negative, because otherwise $\grave{\boldsymbol{R}}$ must satisfy the additional constraint $\grave{\boldsymbol{R}} \geq \boldsymbol{r}$, and Theorem B.2 can not be applied directly. However, by relaxing $\boldsymbol{R}$ to allow negative component, $\grave{\boldsymbol{R}}$ may have non-positive components, which will render Theorem B.2 not applicable without the fact given in the remark immediately after Theorem B.2. With that remark, now by applying Theorem B.2, we see that (B.15) is indeed a characterization of the MLD coding rate region for this set of fictitious sources.

Since the MLD coding rate region is a polytope, there exists a finite set of rate vectors $\mathcal{S}_{\grave{R}}$ such that for any $\boldsymbol{A}$, there exists at least one rate vector $(\grave{R}_1, \grave{R}_2, ..., \grave{R}_1) \in \mathcal{S}_{\grave{R}}$, such that (B.15) is satisfied with equality. Since both $\mathcal{S}_r$ and $\mathcal{S}_{\grave{R}}$ are finite, it follows that there exists a finite set S_R, such that for any $\boldsymbol{A}$, there exists at least one vector $\boldsymbol{R} = \grave{\boldsymbol{R}} - \boldsymbol{r} \in \mathcal{S}_R$ satisfying (3.50) with equality. This subsequently implies that the set $\underline{\check{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D}^*)$ is a polytope, which completes the proof.

Remark B.4. *It is worth mentioning that* $\underline{\check{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D}^*) \subseteq \mathbb{R}^K_+$*. This is true since for the choice of* $\boldsymbol{A} = \mathbb{1}_i(K)$*, we have* $f_1(\boldsymbol{A}) = 1$ *and* $f_\alpha(\boldsymbol{A}) = 0$*, for* $\alpha = 1, 2, \ldots, K$*. Therefore, the inequality in* (3.50) *will be translated to*

$$R_i \geq \frac{1}{2} \log \frac{1}{D^*_1} - \frac{1}{2} \log(2 - D^*_1) = \frac{1}{2} \log \frac{1}{D^*_1(2 - D^*_1)} \geq 0,$$

which implies that all the elements of any rate tuple in $\underline{\check{\mathcal{R}}}_{\mathsf{SMD}}(\boldsymbol{D}^*)$ *are non-negative.* □

B.8 Proof of Theorem 3.46

We pick up the story from (3.44) for the lower bound and rewrite it slightly differently.

$$\begin{aligned} R_{\mathsf{SMD}}(\boldsymbol{D}) + \epsilon &= \frac{1}{K}\sum_{i=1}^{K}(R_i + \epsilon) \\ &\geq \sum_{\alpha=1}^{K-1} \frac{1}{n\alpha\binom{K}{\alpha}} \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} \Big[I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - I(\Gamma_{\boldsymbol{v}}; Y_{\alpha-1}^n) \Big] \\ &\quad + \frac{1}{nK}\left[I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) - I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \right], \end{aligned} \tag{B.16}$$

where now the random variables Y_α, $\alpha = 1, 2, ..., K-1$ are defined as in (3.31) and (3.32), and for simplicity we define $Y_0 = 0$, *i.e.*, a constant.

Next we consider the upper bound $\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D})$, (*c.f.* Theorem 3.36), using the same set of random variables Y_α, $\alpha = 1, 2, ..., K$ as above

$$\begin{aligned} \overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) &= \sum_{\alpha=1}^{K} \frac{1}{\alpha} I(X; Y_\alpha | Y_{\alpha-1}) \\ &= \sum_{\alpha=1}^{K} \frac{1}{\alpha} [I(X; Y_\alpha) - I(X; Y_{\alpha-1})]. \end{aligned} \tag{B.17}$$

Note we have used that fact that $X \leftrightarrow Y_\alpha \leftrightarrow Y_{\alpha-1}$ is a Markov chain for any $\alpha \in \mathcal{I}_K$. The auxiliary random variables used in the lower and upper bounds are in fact the same, and it is clear that this is a valid choice in deriving the lower bound by definition.

Thus we can now bound the difference between the upper and lower bound on the symmetric individual-description rate using (B.16) and (B.17) as follows

$$\begin{aligned} &\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) - R_{\mathsf{SMD}}(\boldsymbol{D}) \\ &\leq \sum_{\alpha=1}^{K} \frac{1}{\alpha} I(X; Y_\alpha | Y_{\alpha-1}) - \frac{1}{nK}\left[I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) - I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \right] \\ &\quad - \sum_{\alpha=1}^{K-1} \left[\frac{1}{n\alpha\binom{K}{\alpha}} \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} [I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) - I(\Gamma_{\boldsymbol{v}}; Y_{\alpha-1}^n)] \right] \\ &= \sum_{\alpha=1}^{K-1} \left\{ \frac{1}{\alpha\binom{K}{\alpha}} \sum_{\boldsymbol{v}:|\boldsymbol{v}|=\alpha} \left[I(X; Y_\alpha) - I(X; Y_{\alpha-1}) - \frac{1}{n} I(\Gamma_{\boldsymbol{v}}; Y_\alpha^n) + \frac{1}{n} I(\Gamma_{\boldsymbol{v}}; Y_{\alpha-1}^n) \right] \right\} \\ &\quad + \frac{1}{K}\Big\{ I(X; Y_K) - I(X; Y_{K-1}) \\ &\qquad - \frac{1}{n} I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) + \frac{1}{n} I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \Big\}. \end{aligned} \tag{B.18}$$

Now, consider an arbitrary $\alpha \in \mathcal{I}_{K-1}$, and an arbitrary $\boldsymbol{v}$ such that $|\boldsymbol{v}| = \alpha$, it follows that

$$\begin{aligned}
I(X;Y_K)-&I(X;Y_{K-1}) - \frac{1}{n}I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) + \frac{1}{n}I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \\
&= h(Y_\alpha) - h(Y_\alpha|X) - h(Y_{\alpha-1}) + h(Y_{\alpha-1}|X) \\
&\quad - \frac{1}{n}[h(Y_\alpha^n) - h(Y_\alpha^n|\Gamma_{\boldsymbol{v}})] + \frac{1}{n}[h(Y_{\alpha-1}^n) - h(Y_{\alpha-1}^n|\Gamma_{\boldsymbol{v}})] \\
&\overset{(a)}{=} - h(Z_\alpha) + h(Z_{\alpha-1}) + \frac{1}{n}h(Y_\alpha^n|\Gamma_{\boldsymbol{v}}) - \frac{1}{n}h(Y_{\alpha-1}^n|\Gamma_{\boldsymbol{v}}) \\
&\overset{(b)}{=} \frac{1}{2}\log\frac{d_{\alpha-1}}{d_\alpha} - \frac{1}{n}I(Y_{\alpha-1}^n; N_{\alpha-1}^n|\Gamma_{\boldsymbol{v}}) \\
&\overset{(c)}{\leq} \frac{1}{2}\log\left[\frac{d_{\alpha-1}}{d_\alpha}\frac{D_\alpha + d_\alpha}{D_\alpha + d_{\alpha-1}}\right] \\
&= \frac{1}{2}\log\frac{2 - D_\alpha}{1 - D_\alpha + \frac{D_\alpha}{D_{\alpha-1}}} \overset{(d)}{\leq} \frac{1}{2}
\end{aligned}$$

where (a) is due to the fact that Y_α^n and $Y_{\alpha-1}^n$ are i.i.d. sequences, (b) holds since $Y_{\alpha-1} = Y_\alpha + N_{\alpha-1}$, and in (c) we used the bounding technique used in the proof of Lemma 3.32, the equality

$$d_\alpha = \sum_{i=\alpha}^{K}\sigma_i^2 = \frac{D_\alpha}{1 - D_\alpha}, \quad \alpha = 1, 2, ..., K-1, \tag{B.19}$$

which is implied by (3.32). Finally, in (d) we used the fact $D_\alpha - \frac{D_\alpha}{D_{\alpha-1}} \leq 0$ and $D_\alpha \geq 0$.

The last term in (B.18) can be bounded similarly by noticing $I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) \geq I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_K^n)$. We can write

$$\begin{aligned}
I(X;Y_K) &- I(X;Y_{K-1}) - \frac{1}{n}I(\{\Gamma_i; i \in \mathcal{I}_K\}; X^n) + \frac{1}{n}I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \\
&\leq I(X;Y_K) - I(X;Y_{K-1}) - \frac{1}{n}I(\{\Gamma_i; i \in \mathcal{I}_K\}; Y_K^n) + \frac{1}{n}I(\Gamma_i; i \in \mathcal{I}_K\}; Y_{K-1}^n) \\
&= - h(Z_K) + h(Z_{K-1}) + \frac{1}{n}h(Y_K^n|\{\Gamma_i; i \in \mathcal{I}_K\}) - \frac{1}{n}h(Y_{K-1}^n|\{\Gamma_i; i \in \mathcal{I}_K\}) \\
&\overset{(d)}{\leq} \frac{1}{2}\log\left[\frac{d_{K-1}}{\sigma_K^2}\frac{D_K + \sigma_K^2}{D_K + d_{K-1}}\right] \leq \frac{1}{2}
\end{aligned} \tag{B.20}$$

where again we used Lemma 3.32 in (d), and the last step is due to the fact that $\sigma_K^2 = \frac{D_K}{1-D_K}$.

Now, summarize all the bounds and replacing (B.19) and (B.20) in (B.18) we have

$$\overline{R}'_{\mathsf{SMD}}(\boldsymbol{D}) - R_{\mathsf{SMD}}(\boldsymbol{D}) \leq \sum_{\alpha=1}^{K}\frac{1}{2\alpha},$$

which completes the proof. □

Appendix for Chapter 4

C

C.1 Proof of Lemma 4.1

Lemma. *For a given distortion vector* $\boldsymbol{D}$*, define* $\tilde{\boldsymbol{D}}$ *as* $\tilde{\boldsymbol{D}} = (\tilde{D}_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K)$*, where*

$$\tilde{D}_{\boldsymbol{v}} = \min_{\boldsymbol{u}:\boldsymbol{u}\leq\boldsymbol{v}} D_{\boldsymbol{u}}.$$

Then $\mathcal{R}_{\mathsf{AMD}}(\tilde{\boldsymbol{D}}) = \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$.

Proof. It is clear that $\tilde{D}_{\boldsymbol{v}} \leq D_{\boldsymbol{v}}$ for all $\boldsymbol{v} \in \Omega_K$, and therefore $\mathcal{R}_{\mathsf{AMD}}(\tilde{\boldsymbol{D}}) \subseteq \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$. So, it remains to prove $\mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D}) \subseteq \mathcal{R}_{\mathsf{AMD}}(\tilde{\boldsymbol{D}})$. Let $\boldsymbol{R} \in \mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$ be an achievable rate tuple for $\boldsymbol{D}$, and $(n; M_i; \Delta_{\boldsymbol{v}})$ be a code for a given ϵ which achieves the distortion constraints $\boldsymbol{D}$, with encoding functions $\{F_i; i \in \mathcal{I}_K\}$ and decoding functions $\{G_{\boldsymbol{v}}; \boldsymbol{v} \in \Omega_K\}$. We can easily modify the decoding functions and obtain a code which satisfies $\tilde{\boldsymbol{D}}$. By the definition of $\tilde{\boldsymbol{D}}$, for all $\boldsymbol{v}$ we have $\tilde{D}_{\boldsymbol{v}} = D_{\tilde{\boldsymbol{v}}}$, where

$$\tilde{\boldsymbol{v}} \triangleq \arg \min_{\boldsymbol{u}:\boldsymbol{u}\leq\boldsymbol{v}} D_{\boldsymbol{u}}.$$

Define

$$\tilde{X}^n_{\boldsymbol{v}} = \tilde{G}_{\boldsymbol{v}}(F_j(X^n); j : v_j = 1) \triangleq \hat{X}^n_{\tilde{\boldsymbol{v}}}.$$

Obviously,

$$\mathbb{E}d(X^n, \tilde{X}^n_{\boldsymbol{v}}) = \mathbb{E}d(X^n, \hat{X}^n_{\tilde{\boldsymbol{v}}}) \leq D_{\tilde{\boldsymbol{v}}} + \epsilon = \tilde{D}_{\boldsymbol{v}} + \epsilon.$$

Thus the similar code with the modified decoding functions satisfies the constraint tuple $\tilde{\boldsymbol{D}}$, and therefore $\boldsymbol{R} \in \mathcal{R}_{\mathsf{AMD}}(\tilde{\boldsymbol{D}})$. $\square$

Appendix for Chapter 5

D

D.1 Cut-set Type Bounds for the Capacity of the DZS Network

We only proved the optimality of the sum-rate bounds in (DZS-3) and (DZS-10) in Subsection 5.5.1. As mentioned before, the other bounds follow from the generalized cut-set bound [74]. In the following we present the proof of the remaining inequalities in Theorem 5.10. Let rate pair (R_1, R_2) is achievable using a code of length ℓ. So, the n-length multi-letter signal interactions can be written using ℓ copies of the original channel matrices, and denoted by bold letters, *e.g.*, $\mathbf{M}_{11} = I_\ell \otimes M_{11}$.

- (DZS-1) $R_1 \leq m_{11}$: Consider the cut which partition the network into $\Omega = \{S_1\}$ and $\Omega^c = \{S_2, A, B, D_1, D_2\}$. We denote the received signal at Ω^c by $Y_{\Omega^c} = (Y_1', Y_2', Y_1, Y_2)$. It can be written in terms of the transmitted signal $X_\Omega = X_1$ as follows.

$$\begin{bmatrix} Y_1'^{\ell} \\ Y_2'^{\ell} \\ Y_1^{\ell} \\ Y_2^{\ell} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{M}_{11} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{G}_{\Omega,\Omega^c}} X_1^{\ell} + \begin{bmatrix} \mathbf{M}_{12} & \mathbf{0} & \mathbf{0} \\ \mathbf{M}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix} \begin{bmatrix} X_2^{\ell} \\ X_1'^{\ell} \\ X_2'^{\ell} \end{bmatrix}.$$

Hence, the cut-set upper bound gives us

$$\ell R_1 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{M}_{11}) = \ell \operatorname{rank}(M_{11}) = \ell m_{11}. \qquad \text{(D.1)}$$

Dividing both sides of (D.1) by ℓ, we get the bound in (DZS-1).

- (DZS-2) $R_2 \le \max(m_{12}, m_{22})$: Similar to the proof of (DZS-1), the transition matrix for the cut specified by $\Omega = \{S_2\}$ and $\Omega^c = \{S_1, A, B, D_1, D_2\}$ with input and output signals $X_\Omega = X_2$ and $Y_{\Omega^c} = (Y_1', Y_2', Y_1, Y_2)$ can be written as

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{M}_{12} \\ \mathbf{M}_{22} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Therefore,

$$\ell R_2 \le \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \ell \operatorname{rank}\begin{bmatrix} M_{12} \\ M_{22} \end{bmatrix} = \ell \max(m_{12}, m_{22}). \tag{D.2}$$

- (DZS-4) $R_2 \le m_{12} + n_{22}$: Consider the cut with partitions $\Omega = \{S_2, B\}$ and $\Omega^c = \{S_1, A, D_1, D_2\}$ with input signal $X_\Omega = (X_2, X_2')$ and $Y_{\Omega^c} = Y_1', Y_1, Y_2$. The transfer matrix for this cut can is

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{M}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{22} \end{bmatrix}.$$

Hence, the flow of information through this cut be bounded by

$$\begin{aligned} \ell R_2 \le \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) &= \ell \operatorname{rank}\begin{bmatrix} M_{12} & 0 \\ 0 & N_{22} \end{bmatrix} \\ &= \ell\left(\operatorname{rank}(M_{12}) + \operatorname{rank}(N_{22})\right) = \ell(m_{12} + n_{22}). \end{aligned} \tag{D.3}$$

- (DZS-5) $R_1 + R_2 \le m_{22} + \max(n_{11}, n_{21})$: This bound corresponds to the cut $\Omega = \{S_1, S_2, A\}$ and $\Omega^c = \{B, D_1, D_2\}$. The transition matrix from the input of the cut $X_\Omega = (X_1, X_2, X_1')$ to its output $Y_{\Omega^c} = (Y_2', Y_1, Y_2)$ can be written as

$$\begin{bmatrix} Y_2'^{\ell} \\ Y_1^{\ell} \\ Y_2^{\ell} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{0} & \mathbf{M}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_{11} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_{21} \end{bmatrix}}_{\mathbf{G}_{\Omega,\Omega^c}} \begin{bmatrix} X_1^{\ell} \\ X_2^{\ell} \\ X_1'^{\ell} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{N}_{22} \end{bmatrix} X_2'^{\ell}.$$

Therefore, from [74] we have

$$\begin{aligned} \ell(R_1 + R_2) \le \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) &= \operatorname{rank}(\mathbf{M}_{22}) + \operatorname{rank}\begin{bmatrix} \mathbf{N}_{11} \\ \mathbf{N}_{21} \end{bmatrix} \\ &= \ell m_{22} + \ell \max(n_{11}, n_{21}). \end{aligned} \tag{D.4}$$

- (DZS-6) $R_1 + R_2 \le \max(m_{11}, m_{12}) + n_{22}$: The sum-rate can be also bounded by the flow of information through the cut $\Omega = \{S_1, S_2, B\}$

and $\Omega^c = \{A, D_1, D_2\}$. The transfer matrix between the input signals $X_\Omega = (X_1, X_2, X_2')$ and the outputs $Y_{\Omega^c} = (Y_1', Y_1, Y_2)$ can be written as

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_{22} \end{bmatrix}.$$

Therefore, the maximum flow of information through this cut can is

$$\begin{aligned} \ell(R_1 + R_2) &\leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \end{bmatrix} + \operatorname{rank}(\mathbf{N}_{22}) \\ &= \ell \max(m_{11}, m_{12}) + \ell n_{22}. \end{aligned} \tag{D.5}$$

- (DZS-7) $R_1 \leq n_{11}$: The rate of the first message can be bounded by the cut-set value of $\Omega = \{S_1, S_2, A, B, D_2\}$ and $\Omega^c = \{D_1\}$. The transfer matrix or this cut from $X_\Omega = (X_1, X_2, X_1', X_2')$ to $Y_{\Omega^c} = Y_1$ is

  $$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N}_{11} \end{bmatrix},$$

  hence,

  $$\ell R_1 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{N}_{11}) = \ell n_{11}. \tag{D.6}$$

- (DZS-8) $R_2 \leq \max(n_{21}, n_{22})$: This inequality corresponds to the information flow through the cut $\Omega = \{S_1, S_2, A, B, D_1\}$ and $\Omega^c = \{D_2\}$, with $X_\Omega = (X_1, X_2, X_1', X_2')$ and $Y_{\Omega^c} = Y_2$. We have

  $$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix}.$$

  Thus,

  $$\ell R_2 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}\begin{bmatrix} \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix} = \ell \max(n_{21}, n_{22}). \tag{D.7}$$

- (DZS-9) $R_2 \leq m_{22} + n_{21}$: Consider the cut which partitions the network to $\Omega = \{S_1, S_2, A, D_1\}$ and $\Omega^c = \{B, D_2\}$. The transfer matrix of this cut between $X_\Omega = (X_1, X_2, X_1')$ and $Y_{\Omega^c} = (Y_2', Y_2)$ would be

  $$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{M}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_{21} \end{bmatrix}.$$

  Therefore, the flow of information through this cut is bounded by

  $$\ell R_2 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{M}_{22}) + \operatorname{rank}(\mathbf{N}_{21}) = \ell(m_{22} + n_{21}). \tag{D.8}$$

$\square$

D.2 Proof of Lemma 5.15

Lemma. *For any deterministic* ZS *network,*

$$\mathcal{R}_{\mathsf{DZS}} \subseteq \bigcup_{r \leq \min(m_{11}, n_{11})} \overline{\mathcal{R}}_{\mathsf{DZS}}(r). \tag{D.9}$$

Proof. Let $(R_1, R_2) \in \mathcal{R}_{\mathsf{DZS}}$ be an arbitrary rate pair which satisfies (DZS-1)-(DZS-10). In particular $R_1 \leq \min\{m_{11}, n_{11}\}$. We claim that $(R_1, R_2) \in \mathcal{R}_{\mathsf{DZS},1}(\kappa) \times \mathcal{R}_{\mathsf{DZS},2}(\kappa)$ for $\kappa = R_1$, and therefore (R_1, R_2) is achievable using network decomposition. In order to do this we have to show that any R_2 satisfying (DZS-1)-(DZS-10), fulfills the constraints in the definition of $\mathcal{R}_{\mathsf{DZS},2}(R_1)$.

Using (DZS-2) and (DZS-3), we have

$$\begin{aligned}
R_2 &\leq \min\Big\{\max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ - R_1, \max(m_{12}, m_{22})\Big\} \\
&= \min\Big\{\max(m_{11}, m_{12}) - R_1, m_{12}\Big\} + (m_{22} - m_{12})^+ \\
&= m'_{12}(R_1) + (m_{22} - m_{12})^+ \\
&\leq m'_{12}(R_1) + (m'_{22}(R_1) - m'_{12}(R_1))^+ && \text{(D.10)} \\
&= \max(m'_{12}(R_1), m'_{22}(R_1)), && \text{(D.11)}
\end{aligned}$$

where in (D.10) we have used the fact that

$$(\min(a, b) - \min(c, d))^+ \geq \min\left((a - c)^+, (b - d)^+\right).$$

Moreover, since R_2 satisfies (DZS-3), (DZS-5), and (DZS-8), we have

$$\begin{aligned}
R_2 &\leq \min\Big\{\max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ - R_1, \\
&\qquad m_{22} + \max(n_{11}, n_{21}) - R_1, m_{22} + n_{21}\Big\} \\
&\leq \min\Big\{\max(m_{11}, m_{12}) + (m_{22} - m_{12})^+ - R_1, m_{22}\Big\} \\
&\quad + \min\Big\{\max(n_{11}, n_{21}) - R_1, n_{21}\Big\} && \text{(D.12)} \\
&= m'_{22}(R_1) + n'_{21}(R_1), && \text{(D.13)}
\end{aligned}$$

where (D.12) holds since

$$\min(a, b) + \min(c, d) \geq \min(a, b + c, b + d),$$

for non-negative a, b, c, and d.

In order to show that the third constraint is satisfied, we can start with (DZS-4), (DZS-6), and (DZS-10).

$$\begin{aligned}
R_2 \leq \min\Big\{&\max(m_{11}, m_{12}) + n_{22} - R_1, m_{12} + n_{22}, \\
&\max(n_{11}, n_{21}) + (n_{22} - n_{21})^+ - R_1\Big\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \min\Big\{ \max(m_{11}, m_{12}) + -R_1, m_{12} \Big\} \\
&\quad + \min\Big\{ \max(n_{11}, n_{21}) + (n_{22} - n_{21})^+ - R_1, n_{22} \Big\} \\
&= m'_{12}(r_1) + n'_{22}(r_1). \qquad \text{(D.14)}
\end{aligned}
$$

Finally, using (DZS-8) and (DZS-10), we have

$$
\begin{aligned}
R_2 &\leq \min\Big\{ \max(n_{11}, n_{21}) + (n_{22} - n_{21})^+ - r_1, \max(n_{21}, n_{22}) \Big\} \\
&= \min\Big\{ \max(n_{11}, n_{21}) - r_1, n_{21} \Big\} + (n_{22} - n_{21})^+ \\
&= n'_{21}(r_1) + (n_{22} - n_{21})^+ \\
&\leq n'_{21}(r_1) + (n'_{22}(r_1) - n'_{21}(r_1))^+ \\
&= \max(n'_{21}(r_1), n'_{22}(r_1)). \qquad \text{(D.15)}
\end{aligned}
$$

Putting inequalities in (D.11) and (D.13)-(D.15) together shows that $R_2 \in \mathcal{R}_{\mathsf{DZS},2}(R_1)$, and completes the proof. □

D.3 Cut-set Type Bounds for the Capacity of the DZZ Network

As mentioned in Section 5.6.1, the individual rate constraints in Theorem 5.11 are straight-forwardly follow from the the generalized cut-set bound [74]. In this part we we present the proof of (DZZ-1)-(DZZ-4), which together with the sum-rate bounds proved in Section 5.6.1 show the optimality of the rate region define in Theorem 5.11 for the deterministic ZZ network. We again start with $(R_1, R_2) \in \mathcal{R}_{\mathsf{DZZ}}$ and assume that this rate pair is achievable using a code of length ℓ. Let the

- (DZZ-1) $R_1 \leq m_{11}$: Consider the cut which partitions the network into $\Omega = \{S_1\}$ and $\Omega^c = \{S_2, A, B, D_1, D_2\}$. We have

$$
\begin{bmatrix} Y_1'^{\ell} \\ Y_2'^{\ell} \\ Y_1^{\ell} \\ Y_2^{\ell} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{M}_{11} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{\mathbf{G}_{\Omega,\Omega^c}} X_1^{\ell} + \begin{bmatrix} \mathbf{M}_{12} & \mathbf{0} & \mathbf{0} \\ \mathbf{M}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_{22} \end{bmatrix} \begin{bmatrix} X_2^{\ell} \\ X_1'^{\ell} \\ X_2'^{\ell} \end{bmatrix}.
$$

  Therefore, the rate of the first message can be upper bounded by

$$
\ell R_1 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{M}_{11}) = \ell m_{11}. \qquad \text{(D.16)}
$$

- (DZZ-2) $R_2 \leq m_{22}$: Similarly for the cut specified by $\Omega = \{S_1, S_2, A, D_1\}$ and $\Omega^c = \{B, D_2\}$, and input and out signals $X_\Omega = (X_1, X_2, X_1')$ and

$Y_{\Omega^c} = (Y_1', Y_2)$, the transfer matrix would be simply

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{M}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Hence,

$$\ell R_2 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{M}_{22}) = \ell m_{22}. \tag{D.17}$$

- (DZZ-3) $R_1 \leq n_{11}$: Consider the cut with partitions $\Omega = \{S_1, A\}$ and $\Omega^c = \{S_2, B, D_1, D_2\}$. The transition matrix for this cut can be written as

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_{11} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

which upper bounds the rate of the first message by

$$\ell R_1 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{N}_{11}) = \ell n_{11}. \tag{D.18}$$

- (DZZ-4) $R_2 \leq n_{22}$: Similarly for $\Omega = \{S_1, S_2, A, B, D_1\}$ and $\Omega^c = \{D_2\}$, we have

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N}_{22} \end{bmatrix},$$

and hence,

$$\ell R_2 \leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \operatorname{rank}(\mathbf{N}_{22}) = \ell n_{22}. \tag{D.19}$$

□

D.4 Proof of Lemma 5.19

Lemma. *Consider the deterministic* Z*-neutralization network defined in Definition 5.18 with channel gains* (n_{11}, n_{12}, n_{22}) *(see Figure 5.13). Any rate tuple* $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ *satisfying*

$$\Upsilon_0 \leq \lambda \triangleq \min\{n_{11}, n_{12}, n_{22}\}, \tag{D.20}$$
$$\Upsilon_0 + \Upsilon_1 \leq n_{11}, \tag{D.21}$$
$$\Upsilon_0 + \Upsilon_2 \leq n_{22}, \tag{D.22}$$
$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \mu \triangleq \max\{n_{11}, n_{12}, n_{22}, n_{11} + n_{22} - n_{12}\}. \tag{D.23}$$

is achievable for this network.

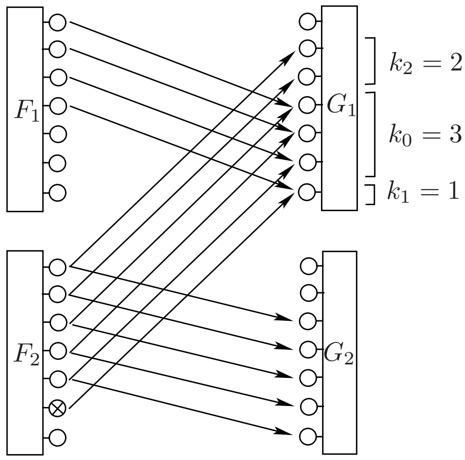


Figure D.1: A deterministic Z-neutralization network. The upper 2 sub-nodes in G_1 are only connected to F_2, and therefore $k_2 = 2$. The next 3 sub-nodes receive information from both F_1 and F_2, and hence $k_0 = 3$. Although the lowest sub-node is also connected to both transmitters, it only receives information from F_1 since F_2 keeps silent on its sub-nodes below n_{22}.

Proof. The coding strategy we present here is based a network decomposition, where the sub-nodes and the links of the deterministic Z-interference network are partitioned into two disjoint sets. We analyze the rate region of each network, and derive an achievable rate region for the original network based on this analysis.

We just point out here that in this coding strategy, the second sender F_2, never sends a bit on a sub-node which is not received at G_2, even if $n_{12} > n_{22}$.

The first partition of the network $\mathcal{N}_1$, consists of those sub-nodes in G_1 which are connected to one of the top m_{11} sub-nodes of F_1 *and* one of the top m_{22} sub-nodes of F_2. All the sub-nodes in the network which are related to (see Definition 5.14) any of these sub-nodes also belong to the first network partition. The remaining nodes and link form the second part of the network $\mathcal{N}_2$. It is clear that these two networks are node-disjoint, and do not cause interference on each other.

We first characterize the number sub-nodes in G_1 which belong to $\mathcal{N}_1$, by determining whether each of them can receive a bit from F_1, F_2, or both of them. We denote the number of levels in G_1 which are only connected to a transmitting level in F_1 by k_1. Similarly, the number of those only connected to a a transmitting level (the top $\min(n_{12}, n_{22})$) in F_2 by k_2. Finally, k_0 denotes the number of levels which are connected to transmitting levels of both F_1 and F_2 (see Figure D.1).

First, we derive k_0. Enumerate the levels of G_1 from 1 (for the highest) to q (for the lowest). Let j be the index of a sub-node in G_1 belong to $\mathcal{N}_1$, *i.e.,* it receives bits from both F_1 and F_2. Its neighbors in F_1 and F_2 (if there is any) are indexed by $j + n_{11} - q$ and $j + n_{12} - q$, respectively. Therefore, j belongs to $\mathcal{N}_1$ if and only if $1 \leq j + n_{11} - q \leq n_{11}$ and $1 \leq j + n_{12} - q \leq \min(n_{12}, n_{22})$. Therefore, the number of such sub-nodes is given by

$$\begin{aligned} k_0 &= [\min\{q, q - n_{12} + n_{22}\} - \max\{q - n_{11}, q - n_{12}\}]^+ \\ &= \min\{n_{11}, n_{12}, n_{22}, (n_{11} + n_{22} - n_{12})^+\}. \end{aligned} \tag{D.24}$$

It is clear from the definition of k_0 that the remaining $n_{11} - k_0$ lowest levels of G_1 are only connected to sub-nodes of F_1, and hence, $k_1 = n_{11} - k_0$. Similarly, $\min\{n_{12}, n_{22}\}$ sub-nodes in G_1 are receiving information from F_2, where k_0 of them are also connected to F_1. Therefore, the remaining sub-nodes are only connected to G_2. Thus, $k_2 = \min\{n_{12}, n_{22}\} - k_0$.

We partition the network into two parts: The first part consists of the k_0 sub-nodes of G_1 connected to both F_1 and F_2, and sub-nodes connected to them. The remaining sub-nodes form the second partition of the network. We characterize the achievable tuples for each, denoted by (Q'_0, Q'_1, Q'_2) and (Q''_0, Q''_1, Q''_2), respectively. The fact that these two partitions are isolated allows us to conclude that the summation of such achievable tuples is also achievable for the original network.

Consider the first partition of the network. It is clear that any of the k_0 levels of G_1 connected to both F_1 and F_2 and can be used to communicate a functional bit, since G_1 naturally receives the `xor` of the transmitting bits. On the other hand, such sub-node can be used to communicate one private bit from any of F_1 or F_2 to G_1 by keeping the other one silent. Therefore, any rate tuple satisfying

$$Q'_0 + Q'_1 + Q'_2 \leq k_0 \tag{D.25}$$

is achievable.

The non-interfered links of the second partition of the network can be used to send private bits from the transmitters to G_1 simultaneously. Moreover, each transmitter can use one of its non-interfering sub-nodes to send a functional bit to G_1, and then, G_1 computes their `xor`, after receiving them separately. This can provide up to $\min\{k_1, k_2\}$ new functional bits for G_1. Moreover, the lower $(n_{22} - n_{12})^+$ sub-nodes of F_2 which are connected to G_2 but not to G_1 can be used to send private bits to G_2 without causing any interference at G_1.

Hence, this strategy can transmit any rate tuple satisfying

$$\begin{aligned} Q''_0 &\leq \min\{k_1, k_2\}, \\ Q''_0 + Q''_1 &\leq k_1, \\ Q''_0 + Q''_2 &\leq k_2 + (n_{22} - n_{12})^+. \end{aligned} \tag{D.26}$$

Summing up the rates achieved on each partition of the network, we have arrive at $Q_i = Q'_i + Q''_i$ for $i = 0, 1, 2$, where (Q'_0, Q'_1, Q'_2)'s and (Q''_0, Q''_1, Q''_2)

satisfy (D.25) and (D.26), respectively. It only remains to apply the Fourier-Motzkin elimination to project the rate region on the (Q_0, Q_1, Q_2) space. This gives us

$$\begin{aligned} Q_0 &\le k_0 + \min\{k_1, k_2\}, \\ Q_0 + Q_1 &\le k_0 + k_1, \\ Q_0 + Q_2 &\le k_0 + k_2 + (n_{22} - n_{12})^+, \\ Q_0 + Q_1 + Q_2 &\le k_0 + k_1 + k_2 + (n_{22} - n_{12})^+. \end{aligned} \tag{D.27}$$

Some simple manipulations show that the RHS's of the inequalities in (D.27) are the same as that claimed in the lemma. □

D.5 Proof of Lemma 5.25

Lemma. *Given ρ_{k-1} linearly independent equations at the relay node A_{k-1}, the minimum number of pure equations achievable at A_k is*

$$\min \rho_{k|\rho_{k-1}} = \max\{0, R_1 - R_2, \rho_{k-1} - \gamma_k\}.$$

Proof. Recall that ρ_k denotes the number of pure linearly independent equations can be received at node A_k using some encoding scheme. Minimizing the number of pure equations is equivalent to maximizing the number of mixed equations. A mixed equations at A_k can be obtained by either receiving a mixed equation from A_{k-1} whose interference is not neutralized by the new interference, or combination of a pure equation from A_{k-1} and an interference from B_{k-1}. Note that we have $R_1 - \rho_{k-1}$ mixed equations, and among the ρ_{k-1} pure equations at most γ_k of them can become mixed in the next layer. This can be done by sending such pure equations on the lowest sub-nodes of A_{k-1} such that they get interfered at A_k by the bits transmitted from B_{k-1}. It is also important that related sub-nodes of B_{k-1} transmit equations independent of the interfering equations at A_{k-1}.

On the other hand, since W_2 has only R_2 bits, at most R_2 equations can be affected by the interference. Therefore, the maximum number of mixed equations would be

$$\begin{aligned} R_1 - \min \rho_k &\le \min\left\{R_1 - \rho_{k-1} + \min\{\rho_{k-1}, \gamma_k\}, R_2\right\} \\ &= \min\{R_1, R_1 - \rho_{k-1} + \gamma_k . R_2\}. \end{aligned}$$

Hence, given ρ_{k-1} pure equations at stage $k-1$, the number of pure equations at stage k in lower bounded by

$$\min \rho_{k|\rho_{k-1}} \ge \max\{0, R_1 - R_2, \rho_{k-1} - \gamma_k\}.$$

It remains to show that this number of pure equations is also achievable. It is, however, clear that depending on the minimizing term of the expression, the corresponding encoding scheme discussed above provides the desired number of linearly independent pure equations. □

D.6 Proof of Lemma 5.26

Lemma. *If the relay node A_{k-1} sends ρ_{k-1} pure equations, then the maximum achievable number of pure equations in the next layer's relay, A_k is*

$$\max \rho_{k|\rho_{k-1}} = \min\{R_1, \Psi_k - R_2 + R_1 - \gamma_k - \rho_{k-1}, \rho_{k-1} + \gamma_k\}.$$

Proof. A pure equation at the next layer A_k can be obtained by either receiving a pure equation from A_{k-1} at a sub-node which is not affected by the message from B_{k-1}, or a mixed message from A_{k-1} whose interference is neutralized[1] by another equation received from B_{k-1}. We denote the number of these two sets of equations by $\rho_k(\mathrm{P})$ and $\rho_k(\mathrm{N})$, respectively.

We first enumerate the first kind of such messages. The maximum number of sub-nodes in A_k which can receive message from A_{k-1} and have not occupied by the signal received from B_{k-1} can be found as illustrated in Figure D.2. The relay node B_{k-1} chooses R_2 sub-nodes among its top β_k sub-nodes to transmit its message to B_k, and at least $R_2 - (\beta_k - \gamma_k)^+$ of them would be among the top γ_k sub-nodes, whose message will be also observed by A_k. Among them, at most $(\gamma_k - \alpha_k)$ are out of the range of A_{k-1}, but the remaining will have overlap with the sub-nodes in range of A_{k-1}. Therefore, at least $[R_2 - (\beta_k - \gamma_k)^+ - (\gamma_k - \alpha_k)^+]^+$ sub-nodes among the α_k sub-nodes of A_k get interfered. Hence, the maximum number of pure equations of the first kind would be the minimum of the number of available pure equations, and the number of non-occupied sub-nodes, which equals to

$$\begin{aligned}
\rho_k(\mathrm{P}) &= \min\left\{\rho_{k-1}, \alpha_k - \left[R_2 - (\beta_k - \gamma_k)^+ - (\gamma_k - \alpha_k)^+\right]^+\right\}\\
&= \min\left\{\rho_{k-1}, \alpha_k - [R_2 - \max(\alpha_k, \gamma_k) + \alpha_k - \max(\beta_k, \gamma_k) + \gamma_k]^+\right\}\\
&= \min\left\{\rho_{k-1}, \alpha_k - [R_2 + \alpha_k + \gamma_k - \Psi_k]^+\right\} \qquad \text{(D.28)}
\end{aligned}$$

On the other hand, we have $R_1 - \rho_{k-1}$ mixed equations, where at most γ_k of them can be neutralized by the message from B_{k-1}. Thus, the maximum number of the second class of pure equations at the k-th layer would be

$$\rho_k(\mathrm{N}) = \min\{R_1 - \rho_{k-1}, \gamma_k\}. \qquad \text{(D.29)}$$

By adding up (D.28) and (D.29) we have

$$\begin{aligned}
\max \rho_{k|\rho_{k-1}} &= \min\left\{\rho_{k-1}, \alpha_k - [R_2 + \alpha_k + \gamma_k - \Psi_k]^+\right\} + \min\{R_1 - \rho_{k-1}, \gamma_k\}\\
&\overset{(a)}{=} \min\left\{\rho_{k-1}, \alpha_k - [R_2 + \alpha_k + \gamma_k - \Psi_k]\right\} + \min\{R_1 - \rho_{k-1}, \gamma_k\}\\
&= \min\{R_1, \Psi_k - R_2 + R_1 - \gamma_k - \rho_{k-1}, \rho_{k-1} + \gamma_k, \Psi_k - R_2\}
\end{aligned}$$

[1] This is done by pre-coding at B_{k-1}. Since it can decode the whole vector $\mathbf{X}_2$, it can encode it again such that a desired number of mixed equations get neutralized.

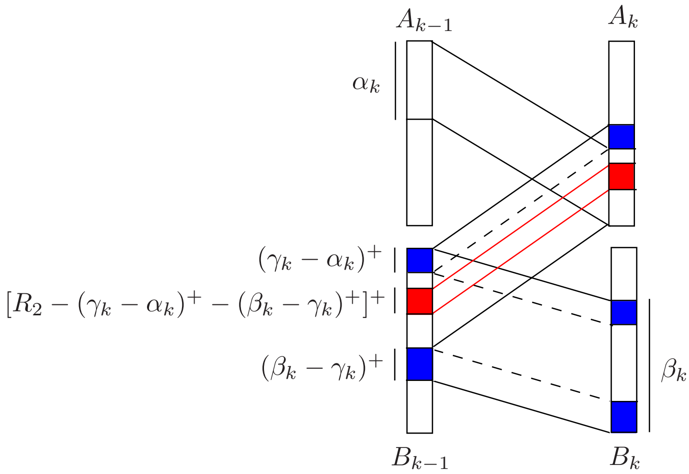


Figure D.2: The maximum number of nodes in A_k which are not affected by interference.

$$\stackrel{(b)}{=} \min\{R_1, \Psi_k - R_2 + R_1 - \gamma_k - \rho_{k-1}, \rho_{k-1} + \gamma_k\} \tag{D.30}$$

where in (a) we have used the fact that $\rho_{k-1} \leq R_1 \leq \alpha_k$, and (b) is due to the inequality $\Psi_k \geq \alpha_k + \beta_k \geq R_1 + R_2$ which shows that dropping the last term does not change the minimum.

Note that this value is always achievable by choosing the transmitting nodes of B_{k-1} and the equations they send properly, such that the required number of nodes A_k do not affected by interference and a specific number of the rest get neutralized. □

D.7 Proof of Theorem 5.27

Lemmas 5.25 and 5.26 determine the minimum and maximum achievable ρ_k provided that the relay node A_{k-1} sends ρ_{k-1} pure equations. A similar argument shows that the extreme values in both lemmas are in fact achievable. Not surprisingly, it can be shown that if $\rho_k = u_1$ and $\rho_k = u_2$ are achievable for $u_1 < u_2$, then any integer $u \in [u_1, u_2]$ is also achievable.

The minimum and maximum values obtained in Lemmas 5.25 and 5.26 depend on ρ_{k-1}. However depending on the required ρ_k at the k-layer, one can choose any $\rho_{k-1} \in \mathcal{P}_{k-1}$ for encoding at A_{k-1}. We will prove (5.66) using induction over k. For $k = 1$, the claim is just rewriting Lemma 5.25 since $m_0 = R_1$. Assuming (5.66) for $k-1$, we have

$$\begin{aligned} m_k &= \min_{\rho_{k-1} \in \mathcal{P}_{k-1}} \max\{0, R_1 - R_2, \rho_{k-1} - \gamma_k\} \\ &= \max\{0, R_1 - R_2, m_{k-1} - \gamma_k\} \end{aligned}$$

$$= \max\{0, R_1 - R_2, \max\{0, R_1 - R_2, R_1 - \Phi_{k-1}\} - \gamma_k\}$$
$$= \max\{0, R_1 - R_2, R_1 - \Phi_k\}.$$

Similarly, using (D.30) M_k can be obtained as the maximum achievable ρ_k is the maximum value that the RHS of (D.30) can get for different value of ρ_{k-1}.

$$\begin{aligned} M_k &= \max_{\rho_{k-1} \in \mathcal{P}_{k-1}} \min\{R_1, R_1 + \Psi_k - R_2 - \gamma_k - \rho_{k-1}, \gamma_k + \rho_{k-1}\} \\ &= \min\{R_1, R_1 + \Psi_k - R_2 - \gamma_k - m_{k-1}, \gamma_k + M_{k-1}\}. \end{aligned} \tag{D.31}$$

Replacing m_k from (5.66), we get

$$\begin{aligned} &R_1 + \Psi_k - R_2 - \gamma_k - m_{k-1} \\ &\quad = R_1 + \Psi_k - R_2 - \gamma_k - \max\{0, R_1 - R_2, R_1 - \Phi_{k-1}\} \\ &\quad = R_1 + \Psi_k - R_2 - \gamma_k + \min\{0, R_2 - R_1, \Phi_{k-1} - R_1\} \\ &\quad = \min\{R_1 + \Psi_k - R_2 - \gamma_k, \Psi_k - \gamma_k, \Psi_k + \Phi_k - 2\gamma_k - R_2\} \end{aligned}$$

where the first two terms are not less than R_1 and do not affect the minimization in (D.31). Therefore,

$$M_k = \min\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \gamma_k + M_{k-1}\} \tag{D.32}$$

However, solving the last recursive relation and evaluating M_k is not easy and we leave it as an optimization expression. □

D.8 Proof of Lemma 5.28

Lemma. *For any Z-chain network with arbitrary number of layers, we have*

$$\rho_k^* \in \mathcal{P}_k, \qquad k = 1, 2, \ldots, N. \tag{D.33}$$

Proof. In order to prove the lemma, we need to show $m_k \leq \rho_k^* \leq M_k$. The first inequality is straight forward and shown as follows.

$$\begin{aligned} \rho_k^* - m_k &= \min\left\{R_1, \min_{1\le\ell\le k}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\right\} - \max\{0, R_1 - R_2, R_1 - \Phi_k\} \\ &= \min\left\{R_1, \min_{1\le\ell\le k}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\right\} + \min\{0, R_2 - R_1, \Phi_k - R_1\} \\ &= \min\left\{R_1, R_2, \Phi_k, \min_{1\le\ell\le k}\{\Psi_\ell + 2\Phi_k - 2\gamma_\ell - R_1 - R_2\},\right. \\ &\qquad \left. \min_{1\le\ell\le k}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_1\}, \min_{1\le\ell\le k}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\right\} \\ &\geq 0 \end{aligned}$$

where the last inequality follows from the facts that $R_1 \leq \alpha_\ell$ and $R_2 \leq \beta_\ell$ for $\ell = 1, \ldots, N$.

To show the second inequality we use induction over k, namely, we show that $\rho_k^* \leq M_k$ provided that $\rho_{k-1}^* \leq M_{k-1}$. For $k = 1$, is claim is trivial by just comparing ρ_1^* and M_1 in (5.67). Assuming $\rho_{k-1}^* \leq M_{k-1}$ and using (5.67), we have

$$
\begin{aligned}
M_k &\geq \min\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \gamma_k + \rho_{k-1}^*\} \\
&= \min\Bigg\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \\
&\qquad\qquad \gamma_k + \min\Bigg\{R_1, \min_{1\leq\ell\leq k-1}\{\Psi_\ell + \Phi_{k-1} - 2\gamma_\ell - R_2\}\Bigg\}\Bigg\} \\
&= \min\Bigg\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \\
&\qquad\qquad \min\Bigg(R_1 + \gamma_k, \min_{1\leq\ell\leq k-1}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\Bigg)\Bigg\} \\
&= \min\Bigg\{R_1, \Psi_k + \Phi_k - 2\gamma_k - R_2, \min_{1\leq\ell\leq k-1}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\Bigg\} \\
&= \min\Bigg\{R_1, \min_{1\leq\ell\leq k}\{\Psi_\ell + \Phi_k - 2\gamma_\ell - R_2\}\Bigg\} \\
&= \rho_k^*.
\end{aligned}
$$

which shows that ρ_k^* does not exceed M_k. Hence $\rho_k^* \in \mathcal{P}_k$.

□

Appendix for Chapter 6

E

E.1 Discussion of Example 6.5

The converse proof is fairly simple and follows from a similar argument we used to prove (GZS-1), (GZS-2), and (GZS-3) in Appendix 6.4.

In the following we will present an encoding strategy which guarantees to achieve rate pair $(R_1 - \frac{1}{2}, R_2 - \frac{1}{2})$, provided that $(R_1, R_2) \in \mathcal{R}^{\mathsf{Z}}$. This gives us an approximate capacity characterization for the Gaussian Z network. In order to do this, we consider the following two cases.

Case A: $g_{12} \geq g_{22}$:

Assume (R_1, R_2) be an achievable rate pair. Then, the first receiver G_1 is able to decode W_1 sent at rate R_1, and remove the signal associated to W_1 from its received signal. The remaining signal provides a higher SNR to decode W_2 than the signal received at G_2. Therefore, in this particular regime, the first receive would be able to decode both messages. Hence, we have a Gaussian multiple access channel from F_1 and F_2 to G_1, combined with a line network from F_2 to G_2. Therefore, the intersection of the rate regions of the Gaussian MAC and the line networks is simply achievable. That is

$$\begin{aligned}\mathcal{R}^{\mathsf{Z}}_{\mathrm{ach},A} = & \left\{ (R_1, R_2) : R_1 \leq \frac{1}{2}\log\left(1 + g_{11}\right), R_2 \leq \frac{1}{2}\log\left(1 + g_{12}\right), \right. \\ & \left. \qquad\qquad R_1 + R_2 \leq \frac{1}{2}\log\left(1 + g_{11} + g_{12}\right) \right\} \\ & \bigcap \\ & \left\{ (R_1, R_2) : R_2 \leq \frac{1}{2}\log\left(1 + g_{22}\right) \right\}\end{aligned}$$

$$= \left\{ (R_1, R_2) : R_1 \le \frac{1}{2} \log (1 + g_{11}), R_2 \le \frac{1}{2} \log (1 + g_{22}), \right.$$

$$\left. R_1 + R_2 \le \frac{1}{2} \log (1 + g_{11} + g_{12}) \right\}. \tag{E.1}$$

Note that the individual rate bounds in $\mathcal{R}^{\mathsf{Z}}$ and $\mathcal{R}^{\mathsf{Z}}_{ach,A}$ are the same. Moreover, the difference between the sum rate bounds is bounded by

$$\frac{1}{2} \log \left(1 + \frac{g_{22}}{g_{12}}\right) \le \frac{1}{2} \log (1 + 1) = \frac{1}{2}. \tag{E.2}$$

Therefore, the gap between each boundary point of $\mathcal{R}^{\mathsf{Z}}$ and $\mathcal{R}^{\mathsf{Z}}_{ach,A}$ is at most $\frac{1}{2}$ bit.

Case B: $g_{12} \le g_{22}$:

The encoding scheme we use for this case is similar to Han-Kobayashi's scheme for 2-user interference channel. We first split the second message W_2 into the common and private parts, $W_2 = (W_2^c, W_2^p)$, with rates R_2^c and R_2^p, respectively, where W_2^c can be decoded at both receivers and W_2^p is only decodable at G_2. Sub-messages W_1, W_2^c, and W_2^p are encoded by corresponding randomly generated Gaussian codes to $\mathbf{x}_1$, $\mathbf{x}_2^c$ and $\mathbf{x}_2^p$, and the resulting codewords are sent over the channel.

We allocate $\alpha_p = 1/g_{12}$ fraction of the transmission power available at F_2 to W_2^p, and the remaining power $\alpha_c = 1 - \alpha_p$ is allocated to W_2^c. Therefore, we have

$$\mathbf{x}_2 = \sqrt{\alpha_c}\mathbf{x}_2^p + \sqrt{\alpha_c}\mathbf{x}_2^p.$$

The first receiver, G_1, decodes W_1 and W_2^c treating W_2^p as noise. Therefore, the effective noise power received at G_1 would be $\mathbb{E}[\sqrt{g_{12}\alpha_p}x_p + z_1]^2 = 2$. According to the capacity region of Gaussian multiple access channel, this can be done provided that

$$\begin{array}{rl} R_1 & \le \frac{1}{2} \log \left(1 + \frac{g_{11}}{2}\right), \\ R_2^c & \le \frac{1}{2} \log \left(\frac{1+g_{12}}{2}\right), \\ R_1 + R_2^c & \le \frac{1}{2} \log \left(\frac{1+g_{11}+g_{12}}{2}\right). \end{array} \tag{E.3}$$

The second decoder first decodes W_2^c treating W_2^p as noise. It then removes the corresponding codeword from the received signal, and decodes W_2^p. This can be done as long as

$$\begin{array}{rl} R_2^c & \le \frac{1}{2} \log \left(\frac{1+g_{22}}{1+g_{22}/g_{12}}\right), \\ R_2^p & \le \frac{1}{2} \log \left(1 + \frac{g_{22}}{g_{12}}\right). \end{array} \tag{E.4}$$

Note that we have two upper bounds for R_2^c. However, it is easy to show that $\frac{1+g_{22}}{1+g_{22}/g_{12}} \ge \frac{1+g_{12}}{2}$, for $1 \le g_{12} \le g_{22}$, and therefore, the first bound dominates the second one. Using Fourier-Motzkin elimination to write the achievable region in terms of R_1 and $R_2 = R_2^c + R_2^p$, and after some simplification, we get that the region

$$\begin{aligned}\mathcal{R}^{\mathsf{Z}}_{ach,B} = \Big\{(R_1,R_2) : & R_1 \le \frac{1}{2}\log(1+g_{11}) - \frac{1}{2},\\ & R_2 \le \frac{1}{2}\log(1+g_{22}) - \frac{1}{2},\\ & R_1 + R_2 \le \frac{1}{2}\log(1+g_{11}+g_{12}) + \frac{1}{2}\log\left(1+\frac{g_{22}}{g_{12}}\right) - \frac{1}{2}\Big\}.\end{aligned}$$

is achievable. Therefore, if $(R_1, R_2) \in \mathcal{R}^{\mathsf{Z}}$, then $(R_1 - \frac{1}{2}, R_2 - \frac{1}{2})$ is achievable. □

E.2 Proof of Lemma 6.7

Lemma. *Any achievable rate pair (R_1, R_2) satisfies*

$$\begin{aligned}\ell R_1 &\le I(x_1^\ell; y_1^\ell) + \ell\epsilon_\ell,\\ \ell R_2 &\le I(x_2^\ell; y_2^\ell) + \ell\epsilon_\ell,\\ \ell(R_1+R_2) &\le I(x_1^\ell, x_2^\ell; y_1^\ell, y_2^\ell) + \ell\epsilon_\ell.\end{aligned}$$

Note that $\epsilon_\ell \to 0$ as ℓ grows.

Proof. As mentioned before, we will use the Fano's inequality in order to prove this lemma. We have

$$\begin{aligned}\ell R_1 = H(W_1) &= I(W_1; y_1^\ell) + H(W_1|y_1^\ell)\\ &\le I(W_1; y_1^\ell) + \ell e_\ell \qquad \text{(E.5)}\\ &\le I(x_1^\ell; y_1^\ell) + \ell e_\ell, \qquad \text{(E.6)}\end{aligned}$$

where (E.5) is implied by the Fano's inequality, and in (E.6) we used the data processing inequality for the Markov chain $W_1 \leftrightarrow x_1^\ell \leftrightarrow y_1^\ell$. Note that where $\epsilon_\ell \to 0$ as ℓ grows. The proofs of the other two inequalities follow the same lines, and we skip them to sake of brevity. □

E.3 Proof of Lemma 6.9

The following achievability scheme simply uses superposition encoding of sub-messages at F_2, and a successively decode and cancel strategy at G_1 and G_2. We use a random codebook with a proper number of codewords, generated according to a zero-mean unit-variance Gaussian distribution for each message. A proper power allocation for the messages at the transmitters allow the decoders to apply a decode and cancel strategy. We denote the codeword corresponding to the message $U_i^{(j)}$ by $\mathbf{x}_{i,j}$, and the power allocated to this message by $\alpha_{i,j}$.

The available power at F_2 can be arbitrarily allocated to its sub-messages. In particular, we choose the power coefficients so that they satisfy $\alpha_{2,2} \leq 1/g_{22}$, $\alpha_{2,3} \leq 1/g_{12}$, and $\alpha_{2,1} = 1 - \alpha_{2,2} - \alpha_{2,3}$. In the decoding part, G_1 and G_2 treat $U_2^{(3)}$ and $U_2^{(2)}$, respectively, as noise. Therefore, the total noise at G_1 and G_2 would be $\tilde{\mathbf{z}}_1 = \sqrt{g_{12}\alpha_{2,3}}\mathbf{x}_{2,3} + \mathbf{z}_1$ and $\tilde{\mathbf{z}}_2 = \sqrt{g_{22}\alpha_{2,2}}\mathbf{x}_{2,2} + \mathbf{z}_2$. However, the effective noise power cannot exceed 2 since $\mathbb{E}[g_{12}\alpha_{2,3}+1] \leq 2$ and $\mathbb{E}[g_{22}\alpha_{2,2}+1] \leq 2$.

The receiver F_1 observes a Gaussian multiple access channel (with noise power upper bounded by 2), where $U_1^{(1)}$ is sent by one user, and $(U_2^{(1)}, U_2^{(2)})$ is sent by the other user. The bounds in (6.27)-(6.30) guarantee that these rates are achievable over the multiple access channel.

On the other hand, the channel from F_2 to G_2 is Gaussian point-to-point channel with modified additive noise. Therefore, any total rate not exceeding its capacity can be reliably transmitted. This is condition is fulfilled here since $\Upsilon_{2,1} + \Upsilon_{2,3}$ satisfies (6.32). Finally, the bound on the power allocated to $U_2^{(3)}$ upper bounds its rate as in (6.31). □

E.4 Proof of Lemma 6.11

Again, the achievability scheme we propose for the Gaussian S interference network (illustrated in Figure 6.8) is based on superposition coding, and a successively decode and cancel decoding strategy, such that the requirements of the problem are fulfilled. A proper power allocation is required to guarantee achievability of the rate tuples mentioned in this lemma.

Note that G_1 does not decode $V_2^{(2)}$ and $V_2^{(4)}$, and treats them as noise. We choose the total fraction of power allocated to $V_2^{(2)}$ and $V_2^{(4)}$ to be at most $1/h_{11}$, that is $\alpha_{2,2} + \alpha_{2,4} \leq 1/h_{11}$. Therefore, the total noise power received at G_1 is upper bounded as $\mathbb{E}[h_{11}(\alpha_{2,2} + \alpha_{2,4}) + 1] \leq 2$.

Similarly, $V_1^{(2)}$ is treated as noise at G_2. By bounding the fraction of power allocated to this sub-message, we can upper bound the effective noise power observed at G_2 by $\mathbb{E}[h_{11}(\alpha_{2,2} + \alpha_{2,4}) + 1] \leq 2$.

The point-to-point Gaussian channel from F_1 to G_1 can support any sum-rate below its capacity as in (6.33). Moreover, $\Theta_{1,2}$ is bounded above since its allocated power does not exceed $1/h_{12}$.

On the other hand, we have a Gaussian multiple access channel from F_1 and F_2 to G_2, with total noise power not exceeding 2. The bounds in (6.35)-(6.38) guarantee that the desired rates belong to the capacity region of this channel, and therefore they are achievable. We skip the details of power allocation here, but we point out that the achievability of the region is a consequence of the Gaussian multiple access rate region achievability. □

E.5 Proof of Lemma 6.13

Note that Z is independent of everything else, and X_1 and X_2 are conditionally independent. Without loss of generality we can also assume that $\mu_i(\gamma) = \mathbb{E}[X_i|\Gamma = \gamma] = 0$ for $\forall\gamma$ (otherwise, for any given $\Gamma = \gamma$, we can shift X_i by $\mu_i(\gamma)$, while the entropy does not change). Let $\mathbb{E}[X_i^2|\Gamma = \gamma] = \sigma_i^2(\gamma)$ for $i = 1, 2$. Therefore the conditional variance of Y can be bounded as

$$\mathbb{E}[Y^2|\Gamma = \gamma] = \mathbb{E}[(X_1 + X_2 + Z)^2|\Gamma = \gamma] = \sigma_1^2(\gamma) + \sigma_2^2(\gamma) + 1. \tag{E.7}$$

Therefore,

$$\begin{aligned} h(Y|\Gamma) = \mathbb{E}_\Gamma[h(Y|\Gamma = \gamma)] &= \mathbb{E}_T[h(X_1 + X_2 + Z|\Gamma = \gamma)] \\ &\le \mathbb{E}_\Gamma[\log 2\pi e(\sigma_1^2(\gamma) + \sigma_2^2(\gamma) + 1)] &\text{(E.8)} \\ &\le \log 2\pi e(\mathbb{E}_\Gamma[\sigma_1^2(\gamma) + \sigma_2^2(\gamma) + 1]) &\text{(E.9)} \\ &= \log 2\pi e(\sigma_1^2 + \sigma_2^2 + 1), &\text{(E.10)} \end{aligned}$$

where in (E.8) we have used the fact that Gaussian random variable has the maximum differential entropy among all random variables with the same variance, and (E.9) follows from the concavity of the function $\log(\cdot)$. Finally, (E.10) is just the tower property, $\mathbb{E}_\Gamma[\mathbb{E}[X_i^2|\Gamma]] = \mathbb{E}[X_i^2]$. $\square$

E.6 Proof of Lemma 6.16

In this part we show that any rate tuple satisfying (6.67)-(6.70) is achievable. We represent the channel model again for clarity in Figure E.1.

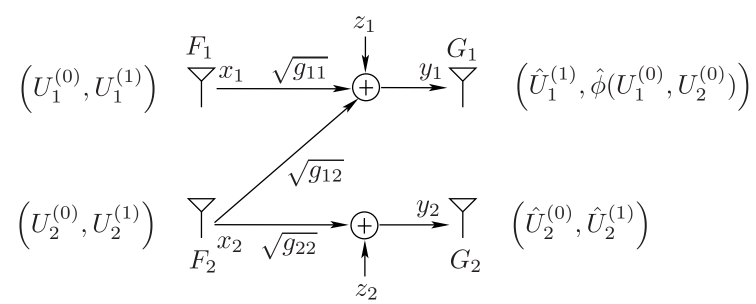


Figure E.1: The Gaussian Z-neutralization channel.

The main idea of this proof can be summarized as follows.

- Use a common codebook with group structure, such as lattice codes, for $W_1^{(0)}$ and $W_2^{(0)}$, which maps them to $\mathbf{x}_{1,0}$ and $\mathbf{x}_{2,0}$, *i.e.*, $\psi(\Big(W_i^{(0)}\Big)) = \mathbf{x}_{i,0}$, for $i = 1, 2$.

- Choose a proper power allocation for $\mathbf{x}_{1,0}$ and $\mathbf{x}_{2,0}$ such that they get received at G_1 at the same power level; More precisely, denoting their power allocation by α_0 and β_0, they should satisfy $g_{11}\alpha_0 = g_{12}\beta_0$. This condition guarantees that the two lattice points get scaled by the same factor, and therefore the result is still a lattice point on the scaled lattice and can be decoded as long as enough signal to noise ratio is provided.

- Use random Gaussian codebooks to encode the private sub-messages to $\mathbf{x}_{1,1}$ and $\mathbf{x}_{2,1}$, and use proper power allocation, α_1 and β_1, *i.e.*, $\mathbb{E}[x_{1,1}^2] \leq \alpha_1$, and $\mathbb{E}[x_{2,1}^2] \leq \beta_1$.

The first receiver G_1 needs to decode the partial-invertible ϕ which we define as

$$\begin{aligned}\phi\left(W_1^{(0)}, W_2^{(0)}\right) &= \psi^{-1}\left(\psi\left(W_1^{(0)}\right) + \psi\left(W_1^{(0)}\right)\right)\\ &= \psi^{-1}\left(\mathbf{x}_{1,0} + \mathbf{x}_{2,0}\right)\end{aligned}$$

where ψ is the one-to-one encoding function which maps the functional messages to the common lattice codebook. Note that the group structure of the code implies that $\mathbf{x}_{1,0} + \mathbf{x}_{2,0}$ is still a valid codeword. It is easy to check that this function is partial-invertible.

Let us define

$$\eta \triangleq \min\left\{g_{11}, g_{12}, g_{22}, \frac{g_{11}g_{22}}{g_{12}}\right\}, \tag{E.11}$$

and recall the definitions of λ and μ from Lemma 6.16:

$$\lambda = \min\{g_{11}, g_{12}, g_{22}\},$$

and

$$\mu = \max\left\{g_{11}, g_{12}, g_{22}, \frac{g_{11}g_{22}}{g_{12}}\right\}.$$

Depending on the minimizer in η, we identify four cases. In each case, the achievable rate region is a polytopes, with a certain number of corner points. It suffices to show the achievability only for the cornet points, since a standard time-sharing argument guarantees achievability for the rest of the region.

The proof details for each corner point includes message splitting, and power allocation for sub-messages such that the decoders be able to decode corresponding messages. In the following we describe this strategy in details for the case where $\eta = g_{11}$. The extension of this method for other cases is straight-forward, and therefore we skip it here to sake of brevity.

Case I. $\eta = g_{11}$ It is clear from the definition of η that in this case $g_{11} \leq g_{12} \leq g_{22}$, and therefore $\lambda = g_{11}$ and $\mu = g_{22}$. Hence, the characterization of the region in the lemma reduces to the set of all non-negative rate tuples $(\Upsilon_0, \Upsilon_1, \Upsilon_2)$ satisfying

$$\Upsilon_0 + \Upsilon_1 \leq \left(\frac{1}{2}\log(g_{11}) - 1\right)^+,$$
$$\Upsilon_0 + \Upsilon_1 + \Upsilon_2 \leq \left(\frac{1}{2}\log(g_{22}) - \frac{3}{2}\right)^+.$$

This rate region is depicted in Figure E.2. It suffices to show that the corner points A, B and C are achievable, since the points D and E are degenerated from B and C, respectively.

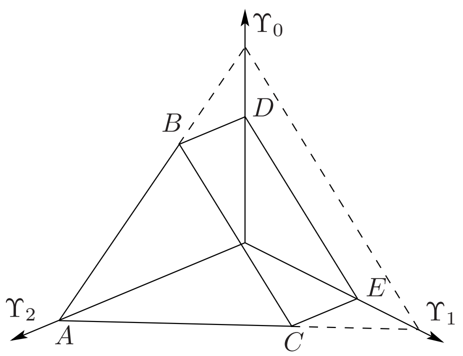


Figure E.2: Achievable rate region of the Z-neutralization network when $\eta = g_{11}$.

- A, with coordinates $(\Upsilon_0, \Upsilon_1, \Upsilon_2) = \left(0, 0, \left(\frac{1}{2}\log(g_{22}) - \frac{3}{2}\right)^+\right)$: The encoding strategy for this corner point is fairly simple. The second transmitter uses all its available power to send $W_2^{(1)}$, while the first transmitter keeps silent. That is, $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = \mathbf{x}_{2,1}$. The first decoder has nothing to decode, and the second one can decode $\mathbf{x}_2$ from $\mathbf{y}_2$ as long as $\Upsilon_2 \leq \frac{1}{2}\log(1 + g_{22})$. It is clear that in particular $\Upsilon_2 = \left(\frac{1}{2}\log(g_{22}) - \frac{3}{2}\right)^+$ is achievable.

- B: The rate triple for this point is given by $\Upsilon_0 = \left(\frac{1}{2}\log(g_{11}) - 1\right)^+$, $\Upsilon_1 = 0$, and $\Upsilon_2 = \left(\frac{1}{2}\log(g_{22}) - \frac{3}{2}\right)^+ - \left(\frac{1}{2}\log(g_{11}) - 1\right)^+$. In order to achieve this rate triple, the first encoder sends its lattice codeword with power allocation $\alpha_0 = (g_{11} - 1)/g_{11}$. The second encoder splits its private message into $W_2^{(1)} = (W_2^{(1,1)}, W_2^{(1,2)})$ of rates $\Upsilon_{2,1}$ and $\Upsilon_{2,2}$ where $\Upsilon_2 = \Upsilon_{2,1} + \Upsilon_{2,2}$. Then it sends

$$\mathbf{x}_2 = \sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + \sqrt{\beta_0}\mathbf{x}_{2,0} + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}$$

where the power allocation coefficients are fixed to be $\beta_{1,2} = 1/g_{12}$, $\beta_0 = (g_{11} - 1)/g_{12}$, and $\beta_{1,1} = 1 - \beta_0 - \beta_{1,2}$. The signal received at the destinations are

$$\begin{aligned}
\mathbf{y}_1 &= \sqrt{g_{11}}\mathbf{x}_1 + \sqrt{g_{12}}\mathbf{x}_2 + \mathbf{z}_1, \\
&= \sqrt{g_{11}}[\sqrt{\alpha_0}\mathbf{x}_{1,0}] + \sqrt{g_{12}}[\sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + \sqrt{\beta_0}\mathbf{x}_{2,0} + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}] + \mathbf{z}_1, \\
&= \sqrt{g_{12} - g_{11}}\mathbf{x}_{2,1,1} + \sqrt{g_{11} - 1}[\mathbf{x}_{1,0} + \mathbf{x}_{2,0}] + \mathbf{x}_{2,1,2} + \mathbf{z}_1, \qquad \text{(E.12)}\\
\mathbf{y}_2 &= \sqrt{g_{22}}\mathbf{x}_2 + \mathbf{z}_2, \\
&= \sqrt{g_{22}}[\sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + \sqrt{\beta_0}\mathbf{x}_{2,0} + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}] + \mathbf{z}_2 \\
&= \sqrt{\frac{g_{22}(g_{12} - g_{11})}{g_{12}}}\mathbf{x}_{2,1,1} + \sqrt{\frac{g_{22}(g_{11} - 1)}{g_{12}}}\mathbf{x}_{2,0} + \sqrt{\frac{g_{22}}{g_{12}}}\mathbf{x}_{2,1,2} + \mathbf{z}_2. \qquad \text{(E.13)}
\end{aligned}$$

The first node decode and cancel $\mathbf{x}_{2,1,1}$, $\tilde{\mathbf{x}}_0 = \mathbf{x}_{1,0} + \mathbf{x}_{2,0}$, and $\mathbf{x}_{2,1,2}$ in order, while the second one performs the same decoding for $\mathbf{x}_{2,1,1}$, $\mathbf{x}_{2,0}$, and $\mathbf{x}_{2,1,2}$. Note that $\mathbf{x}_{2,1,1}$ is decodable at G_1 and G_2 if $\Upsilon_{2,1}$ does not exceed

$$\frac{1}{2}\log\left(\frac{g_{12}+1}{2g_{11}}\right) > \frac{1}{2}\log\left(\frac{g_{12}}{g_{11}}\right) - \frac{1}{2}, \qquad \text{(E.14)}$$

and

$$\frac{1}{2}\log\left(\frac{g_{22}+1}{g_{11}g_{22}/g_{12}+1}\right) > \frac{1}{2}\log\left(\frac{g_{12}}{g_{11}}\right) - \frac{1}{2}, \qquad \text{(E.15)}$$

respectively. Therefore $\Upsilon_{2,1} = \frac{1}{2}\log\left(\frac{g_{12}}{g_{11}}\right) - \frac{1}{2}$ is achievable.

Once $\mathbf{x}_{2,1,1}$ is decoded and removed from the received signal, G_1 has to decode $\tilde{\mathbf{x}}_0$, which can be done as long as

$$\Upsilon_0 \leq \frac{1}{2}\log\left(\frac{g_{11}+1}{2}\right). \qquad \text{(E.16)}$$

On the other hand, Υ_0 is upper bounded as

$$\Upsilon_0 \leq \frac{1}{2}\log\left(\frac{g_{11}g_{22}/g_{12}+1}{g_{22}/g_{12}+1}\right), \qquad \text{(E.17)}$$

since $\mathbf{x}_{2,0}$ should be decodable at G_2. Clearly $\Upsilon_0 = \frac{1}{2}\log(g_{11}) - 1$ satisfies both (E.16) and (E.17), and therefore, is achievable.

The last step is to decode $\mathbf{x}_{2,1,2}$ at G_2. This can be done if $\Upsilon_{2,2}$ is upper bounded by

$$\frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right). \qquad \text{(E.18)}$$

Again, it is easy to show that $\Upsilon_{2,2} = \frac{1}{2}\log(g_{22}/g_{12})$ satisfies this bound. Hence, the total rate of

$$\Upsilon_2 = \Upsilon_{2,1} + \Upsilon_{2,2} = \frac{1}{2}\log\left(\frac{g_{12}}{g_{11}}\right) - \frac{1}{2} + \frac{1}{2}\log(g_{22}/g_{12})$$

is achievable, which proves achievability of the cornet point B.

- C: This point is characterized by $\Upsilon_0 = 0$, $\Upsilon_1 = \left(\frac{1}{2}\log(g_{11}) - 1\right)^+$, and $\Upsilon_2 \left(\frac{1}{2}\log(g_{22}) - \frac{3}{2}\right)^+ - \left(\frac{1}{2}\log(g_{11}) - 1\right)^+$. For this rate tuple, the rate of the functional message is zero, and we have to deal with a usual Z channel. The second transmitter splits its private message similar to that of corner point B. The transmission power is distributed between among the sub-message as $\alpha_0 = 0$, $\alpha_1 = 1$, $\beta_{1,2} = 1/g_{12}$, $\beta_0 = 0$, and $\beta_{1,1} = 1 - \beta_{1,2}$. The received signals at G_1 and G_2 would be

$$\begin{aligned}
\mathbf{y}_1 &= \sqrt{g_{11}}\mathbf{x}_1 + \sqrt{g_{12}}\mathbf{x}_2 + \mathbf{z}_1, \\
&= \sqrt{g_{11}}[\sqrt{\alpha_1}\mathbf{x}_{1,1}] + \sqrt{g_{12}}[\sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}] + \mathbf{z}_1, \\
&= \sqrt{g_{12} - 1}\mathbf{x}_{2,1,1} + \sqrt{g_{11}}\mathbf{x}_{1,1} + \mathbf{x}_{2,1,2} + \mathbf{z}_1, && \text{(E.19)} \\
\mathbf{y}_2 &= \sqrt{g_{22}}\mathbf{x}_2 + \mathbf{z}_2, \\
&= \sqrt{g_{22}}[\sqrt{\beta_{1,1}}\mathbf{x}_{2,1,1} + + \sqrt{\beta_{1,2}}\mathbf{x}_{2,1,2}] + \mathbf{z}_2 \\
&= \sqrt{\frac{g_{22}(g_{12} - 1)}{g_{12}}}\mathbf{x}_{2,1,1} + \sqrt{\frac{g_{22}}{g_{12}}}\mathbf{x}_{2,1,2} + \mathbf{z}_2. && \text{(E.20)}
\end{aligned}$$

The first sub-message of W_2, $\mathbf{x}_{2,1,1}$ has to be decoded at both G_1 and G_2, Therefore its rate is upper bounded by

$$\Upsilon_{2,1} \le \frac{1}{2}\log\left(\frac{g_{12} + g_{11} + 1}{g_{11} + 2}\right),$$

and

$$\Upsilon_{2,1} \le \frac{1}{2}\log\left(\frac{g_{22} + 1}{g_{22}/g_{12} + 1}\right).$$

Hence, $\Upsilon_{2,1} = \frac{1}{2}\log(g_{12}/g_{11}) - \frac{1}{2}$ satisfies both inequalities since we assumed $g_{11} \le g_{12} \le g_{22}$. Then G_1 removes $\mathbf{x}_{2,1,1}$, and decodes $\mathbf{x}_{1,1}$, treating $\mathbf{x}_{2,1,2}$ as noise. Therefore, any rate satisfying

$$\Upsilon_1 \le \frac{1}{2}\log\left(\frac{g_{11} + 2}{2}\right),$$

and in particular, $\Upsilon_1 = \frac{1}{2}\log(g_{11}) - 1$ is achievable. It only remains to decode $\mathbf{x}_{2,1,2}$ at G_2, which can be done, provided that

$$\Upsilon_{2,2} \le \frac{1}{2}\log\left(1 + \frac{g_{22}}{g_{12}}\right).$$

Hence, $\Upsilon_{2,2} = \frac{1}{2}\log(g_{22}/g_{12})$ is a valid choice, which yields in the total rate of

$$\begin{aligned}\Upsilon_2 = \Upsilon_{2,1} + \Upsilon_{2,2} &= \frac{1}{2}\log(g_{22}/g_{12}) + \frac{1}{2}\log(g_{12}/g_{11}) - \frac{1}{2}\\ &= \frac{1}{2}\log(g_{22}) - \frac{1}{2}\log(g_{11}) - \frac{1}{2},\end{aligned}$$

which implies the achievability of the rate point C.

□

Appendix for Chapter 7

F

F.1 Proof of Lemma 7.3

Consider a block matrix with shift matrix block as

$$G = \begin{bmatrix} \mathbf{J}^{p-m_1} & \mathbf{J}^{p-q_1} \\ \mathbf{J}^{p-m_2} & \mathbf{J}^{p-q_2} \end{bmatrix}.$$

We can, without loss of generality assume that $m_1 = \max\{m_1, m_2, q_1, q_2\} = p$. Let r denote the rank of this matrix, and our goal is to prove that $r = \Psi(m_1, m_2, q_1, q_2)$, defined in (7.4). Therefore, $\mathbf{G}$ can be written as a product of two full-rank matrices as

$$G = G_1 G_2 = \begin{bmatrix} \mathbf{I} & \mathbf{J}^{m_1-q_1} \\ \mathbf{J}^{m_1-m_2} & \mathbf{J}^{m_1-q_2} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{J}^{m_1-m_2} & G_1' \end{bmatrix}}_{r} \left.\begin{bmatrix} \mathbf{I} & \mathbf{J}^{m_1-q_1} \\ \mathbf{0} & G_2' \end{bmatrix}\right\} r$$

where $\mathbf{I}$ is the identity matrix of size $p \times p$ and $\mathbf{0}$ is the zero matrix of size $p \times (r-p)$. Moreover, G_1 and G_2 are full-rank matrices which satisfy

$$\mathbf{J}^{2m_1-m_2-q_1} + G_1 G_2 = \mathbf{J}^{m_1-q_2}.$$

It is clear that the fact that G_1 is full-rank, implies G_1' is also full-rank, and moreover, the first p columns and the last $(r-p)$ columns of $\mathbf{G}_1$ are linearly independent. Therefore, $r = \operatorname{rank}(G_1) = p + \operatorname{rank}(G_1')$.

Note that if $m_1 + q_2 = m_2 + q_2$, then $2m_1 - m_2 - q_1 = m_1 - q_2$, and $G_1 = G_2 = \mathbf{0}$. Recall that G_1 and G_2 are full-rank. So, they should have zero columns, and hence $r = m_1$, in this case.

On the other hand, if $m_1 + q_2 \neq m_2 + q_2$, then $G_1' G_2' = \mathbf{J}^{m_1-q_2} - \mathbf{J}^{2m_1-m_2-q_1}$, and

$$\operatorname{rank}(G_1' G_2') = \max\{p - (m_1 - q_2), m_1 - (2m_1 - m_2 - q_1)\}.$$

Since G'_1 and G'_2 are also full-rank, each of them is of the same rank. Therefore,

$$r = p + \text{rank}\,(G'_1) = m_1 + \max\{q_2, m_2 + q_1 - m_1\} = \max\{m_1 + q_2, m_2 + q_1\}.$$

□

F.2 Evaluation of $R^{(N)} + R^{(P)}$

Note that combining (7.17) and (7.18), we conclude any rate satisfying

$$\begin{aligned} R \le \nu + \min\big\{ &\max\{(m_1 - q_1)^+, (m_2 - q_2)^+\}, (m_1 - q_1)^+ + (n_1 - \nu), \\ &(m_2 - q_2)^+ + (n_2 - \nu), \max\{n_2 - \nu, n_2 - \nu\}\big\} \end{aligned} \tag{F.1}$$

is achievable.

First assume $m_1 + q_2 = m_2 + q_2$. In this case, from (7.17) we have $\nu = 0$. Therefore, (F.1) can be written as

$$R \le \min\big\{(m_1 - q_1)^+, (m_1 - q_1)^+ + n_1, (m_2 - q_2)^+ + n_2, \max\{n_2, n_2\}\big\}.$$

It remains to show that $(m_1 - q_1)^+ = \max\{m_1, m_2, q_1, q_2\} - \max\{q_1, q_2\}$, which is straight-forward, since we assumed $m_1 + q_2 = m_2 + q_1$.

In the second case, where $m_1 + q_2 \ne m_2 + q_2$, we have $\nu = \min\{n_1, n_2, \delta\}$. Therefore, (F.1) can be written as

$$\begin{aligned} R &\le \min\big\{ \max\{(m_1 - q_1)^+, (m_2 - q_2)^+\} + \nu, (m_1 - q_1)^+ + n_1, \\ &\qquad (m_2 - q_2)^+ + n_2, \max\{n_2, n_2\}\big\} \\ &= \min\big\{ \max\{(m_1 - q_1)^+, (m_2 - q_2)^+\} + \min\{\delta, n_1, n_2\}, (m_1 - q_1)^+ + n_1, \\ &\qquad (m_2 - q_2)^+ + n_2, \max\{n_2, n_2\}\big\} \\ &= \min\big\{ \max\{(m_1 - q_1)^+, (m_2 - q_2)^+\} + \delta, (m_1 - q_1)^+ + n_1, \\ &\qquad (m_2 - q_2)^+ + n_2, \max\{n_2, n_2\}\big\} \end{aligned} \tag{F.2}$$

where the last equality holds, since if $\delta = n_1$ then the first expression is always greater than or equal to the second term, and does not appear in the minimization result. Similarly if $\delta = n_2$. Now, without loss of generality, we can assume that $m_1 + q_2 > m_2 + q_1$, and conclude $(m_1 - q_1)^+ \ge (m_2 - q_2)^+$. Hence,

$$\begin{aligned} &\delta + \max\{(m_1 - q_1)^+, (m_2 - q_2)^+\} \\ &\quad = \min\{q_1, q_2\} - \min\{(q_1 - m_1)^+, (q_2 - m_2)^+\} \\ &\qquad + \max\{(m_1 - q_1)^+, (m_2 - q_2)^+\} \\ &\quad = (q_1 + q_2 - \max\{q_1, q_2\}) - (q_1 - m_1)^+ + (m_1 - q_1)^+ \\ &\quad = (q_1 + q_2 - \max\{q_1, q_2\}) - (\max\{q_1, m_1\} - m_1) + (\max\{q_1, m_1\} - q_1) \\ &\quad = q_1 + q_2 - \max\{q_1, q_2\} + m_1 - q_1 \\ &\quad = m_1 + q_2 - \max\{q_1, q_2\} \\ &\quad = \max\{m_1 + q_2, m_2, q_1\} - \max\{q_1, q_2\}. \end{aligned} \tag{F.3}$$

Replacing this in the (F.3) in (F.2) gives us the desired result. □

Notation, Symbols, and Abbreviations

$\triangleq$	definition
$\mathcal{I}_N$	$\{1, 2, \ldots, N\}$; Subsection 2.1.1
Ω_K^α	$\{\boldsymbol{v} \in \{0,1\}^K : \lvert\boldsymbol{v}\rvert = \alpha\}$; Subsection 2.1.1
$\lvert\boldsymbol{v}\rvert$	Hamming weight of the vector $\boldsymbol{v}$
$d_H(\cdot, \cdot)$	Hamming distance between two vectors
$\prod$	set product
$\mathbb{E}$	expectation operator
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers
$\mathbb{Z}$	set of integer numbers
$\mathbb{Z}_+$	set of positive integer numbers
$\boldsymbol{u} \leq \boldsymbol{v}$	coordinate-wise inequality for vectors $\boldsymbol{u}$ and $\boldsymbol{v}$; see Definition 2.2
$\mathbb{1}_j(K)$	indicator for position j in a K-dimensional space; see Def. 2.4
$\mathscr{L}$	ordering level; see Definition 2.5
$\mathcal{N}(\mu, \sigma^2)$	the Gaussian distribution with mean μ and variance σ^2
$\pi(\cdot)$	permutation
$\succ$	covering for optimal α-resolutions; see Defi. 3.8
$A_{\min}$	minimum component of a vector
A_{sum}	summation of the components of a vector
conv	convex hull operator
$D^\star$	enhanced distortion vector; see (3.9)
MLD	multi-level diversity coding
SMLD	symmetric multi-level diversity coding
$\mathcal{R}_{\mathsf{SMLD}}(\boldsymbol{H})$	achievable rate region of the SMLD problem; see Thm. 2.1
AMLD	asymmetric multi-level diversity coding
$\mathcal{R}_{\mathsf{AMLD}}(\boldsymbol{H})$	achievable rate region of the AMLD problem; see Thm. 2.7
$\mathcal{R}_{\mathsf{AMLD}}^{\mathscr{L}}(\boldsymbol{H})$	achievable rate region of the AMLD problem for a specific order
MD	multiple description coding
SMD	symmetric multiple description coding
$\mathcal{R}_{\mathsf{SMD}}(\boldsymbol{D})$	admissible rate region of the SMD problem; see (3.1)
$\overline{\mathcal{R}}_{\mathsf{SMD}}^{\mathsf{SR}}(\boldsymbol{D})$	inner bound for $\mathcal{R}_{\mathsf{SMD}}$ based on successive refinement

$\overline{\mathcal{R}}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D})$	inner bound for $\mathcal{R}_{\mathsf{SMD}}$ based on PPR scheme
$\underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D})$	outer bound for $\mathcal{R}_{\mathsf{SMD}}$
$\underline{\mathcal{R}}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$	parametric outer bound for $\mathcal{R}_{\mathsf{SMD}}$
$R_{\mathsf{SMD}}(\boldsymbol{D})$	symmetric individual-description rate-distortion; see (3.2)
$\overline{R}^{\mathsf{SR}}_{\mathsf{SMD}}(\boldsymbol{D})$	upper bound for R_{SMD} based on successive refinement
$\overline{R}^{\mathsf{PPR}}_{\mathsf{SMD}}(\boldsymbol{D})$	upper bound for R_{SMD} based on PPR scheme
$\underline{R}_{\mathsf{SMD}}(\boldsymbol{D})$	lower bound for R_{SMD}
$\underline{R}_{\mathsf{SMD}}(\boldsymbol{D}, \boldsymbol{d})$	parametric lower bound for R_{SMD}
$\mathcal{R}_{\mathsf{AMD}}(\boldsymbol{D})$	admissible rate region of the AMD problem
$\overline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$	inner bound for $\mathcal{R}_{\mathsf{AMD}}$
$\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D})$	outer bound for $\mathcal{R}_{\mathsf{AMD}}$
$\underline{\mathcal{R}}_{\mathsf{AMD}}(\boldsymbol{D}, \boldsymbol{d})$	parametric outer bound for $\mathcal{R}_{\mathsf{AMD}}$
SID − RD	symmetric individual description rate-distortion
PPR	Puri, Pradhan and Ramchandran coding scheme
SR	successive refinement
SR − MLD	successive refinement with multi-level diversity coding
SR − ULP	successive refinement with unequal loss protection coding
MDS	maximum distance separable
AMD	asymmetric multiple description coding
MSB	most significant bit
BC	broadcast channel
MAC	multiple access channel
XX	the XX network
Z	the Z network
S	the S network
ZS	the ZS network
ZZ	the ZZ network
DZS	the deterministic ZS network
DZZ	the deterministic ZZ network
GZS	the Gaussian ZS network
GZZ	the Gaussian ZZ network
Z − chain	the deterministic Z-chain network
C	capacity
C^{det}	capacity of the deterministic network
C^{g}	capacity of the Gaussian network
Ω	a cut in a network
$\Lambda(\mathcal{S}; \mathcal{D})$	the set of cuts which separate $\mathcal{S}$ and $\mathcal{D}$
$\mathbf{G}_{\Omega,\Omega^c}$	the transfer matrix of cut Ω
$\mathcal{R}_{\mathsf{DZS}}$	achievable rate region of DZS network
$\mathcal{R}_{\mathsf{DZZ}}$	achievable rate region of DZZ network
$\underline{\mathcal{R}}_{\mathsf{GZS}}$	outer bound for the achievable rate region of GZS network
$\overline{\mathcal{R}}_{\mathsf{GZS}}$	inner bound for the achievable rate region of GZS network
$\underline{\mathcal{R}}_{\mathsf{GZZ}}$	outer bound for the achievable rate region of GZZ network
$\overline{\mathcal{R}}_{\mathsf{GZZ}}$	inner bound for the achievable rate region of GZZ network

$(x)^+$	positive part of x
SNR	signal to noise ratio
INR	interference to noise ratio
SINR	signal to interference and noise ratio
AVC	arbitrarily varying channel
GDA	the Gaussian diamond network with adversarial jammer
DDA	the deterministic diamond network with adversarial jammer
C_{DDA}	capacity of DDA network
$\overline{C}_{\mathsf{GDA}}$	upper bound on the capacity of GDA network

Bibliography

[1] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Academic Press, Inc. Orlando, FL, USA, 1982.

[2] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J*, vol. 27, pp. 379–423, 1948.

[3] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 2006.

[4] V. V. Vazirani, *Approximation Algorithms*. New York: Springer-Verlag, 2001.

[5] D. Tse, "It is easier to approximate," Seoul, Korea, June 2009.

[6] R. Ahlswede, "Multi-way communication channels," in *IEEE International Symposium on Information Theory (ISIT)*, Tsahkadsor, Armenia, USSR, September 1971, pp. 23–52.

[7] H. Liao, "Multiple Access Channels." 1972.

[8] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[9] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[10] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, pp. 49–60, January 1981.

[11] A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[12] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1912–1923, 1997.

[13] P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, 1973.

[14] R. Gallager, “Capacity and coding for degraded broadcast channels,” *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.

[15] R. H. Etkin, D. Tse, , and H. Wang, “Gaussian interference channel capacity to within one bit,” *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[16] E. Telatar and D. N. C. Tse, “Bounds on the capacity region for a class of interference channels,” in *IEEE international Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 2871–2874.

[17] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *to appear in IEEE Transactions on Information Theory*, 2010.

[18] P. Gupta and P. Kumar, “The capacity of wireless networks,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[19] A. Ozgur, O. Leveque, and D. Tse, “Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks,” *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[20] S. Mohajer, S. N. Diggavi, C. Fragouli, and D. N. C. Tse, “Transmission techniques for relay-interference networks,” in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2008.

[21] ——, “Approximate capacity characterization for a class of gaussian relay-interference wireless networks,” 2010, to be submitted to IEEE Transactions on Information Theory.

[22] L. Ozarow, “On a source-coding problem with two channels and three receivers,” *Bell Syst. Tech. J*, vol. 59, no. 10, pp. 1909–1921, 1980.

[23] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” *IEEE Transactions on Information Theory*, vol. 33, no. 4, pp. 502–521, 1987.

[24] J. Chen, C. Tian, S. Hemami, and T. Berger, “Multiple description quantization via Gram-Schmidt orthogonalization,” *IEEE Transaction on Information Theory*, vol. 52, no. 12, pp. 5197–5217, 2006.

[25] S. S. Pradhan, R. Puri, and K. Ramchandran, “n-channel symmetric multiple descriptions - part i: (n, k) source-channel erasure codes,” *IEEE Transaction on Information Theory*, vol. 50, no. 1, pp. 47–61, 2004.

[26] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions - part ii: an achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1377–1392, 2005.

[27] R. Venkataramani, G. Kramer, and V. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, 2003.

[28] H. Wang and P. Viswanath, "Vector gaussian multiple description with individual and central receivers," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2133–2153, 2007.

[29] ——, "Vector gaussian multiple description with two levels of receivers," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 401–410, 2009.

[30] R. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 609–621, 1999.

[31] C. Tian, S. Mohajer, and S. Diggavi, "Approximating the gaussian multiple description rate region under symmetric distortion constraints," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3869–3891, Aug. 2009.

[32] S. Mohajer, C. Tian, and S. Diggavi, "Asymmetric multi-level diversity coding," in *Proceedings of Data Compression Conference*, Snowbird, USA, 2008, pp. 412–421.

[33] ——, "Asymmetric multilevel diversity coding and asymmetric gaussian multiple descriptions," 2009, accepted for publication in IEEE Transactions on Information Theory, Available from http://arxiv.org/pdf/0911.2346.

[34] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.

[35] M. Maddah-Ali and D. Tse, "Approximating the rate-distortion region of the distributed source coding for three jointly Gaussian tree-structured sources," in *IEEE international Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009, pp. 1468–1472.

[36] M. Maddah-Ali and D. N. C. Tse, "Interference Neutralization in Distributed Lossy Source Coding," in *IEEE international Symposium on Information Theory (ISIT)*, Austin, USA, June 2010, pp. 166–170.

[37] U. Erez and R. Zamir, "Achieving log (1+ SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[38] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[39] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication Over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE Trans. Information Theory*, vol. 54, no. 8, pp. 3457–3470, August 2008.

[40] V. Cadambe and S. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Trans. Information Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.

[41] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of an one-sided Gaussian interference channel," in *Proc. 45th Annu. Allerton Conf. Communications, Control, and Computing*, Sep 2007, pp. 791–801.

[42] J. Roche, R. Yeung, and K. Hau, "Symmetrical multilevel diversity coding," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1059–1064, 1997.

[43] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1737–1744, 1996.

[44] R. Yeung, "Multilevel diversity coding with distortion," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 412–422, 1995.

[45] R. Singleton, "Maximum distance q-nary codes," *IEEE Transactions on Information Theory*, vol. 10, no. 2, pp. 116–118, 1964.

[46] I. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, pp. 300–304, 1960.

[47] J. Roche, "Distributed information storage," *Ph.D. dissertation*, 1992.

[48] K. Hau, "Multilevel diversity coding with independent data streams," *Master of Philosophy Thesis, Department of Information Engineering, The Chinese University of Hong Kong*, 1995.

[49] A. Goldman, "Resolution and separation theorems for polyhedral convex sets," *Ann. of Math. Studies*, vol. 38, pp. 41–51, 1956.

[50] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[51] C. Tian, S. Mohajer, and S. Diggavi, "On the symmetric gaussian multiple description rate-distortion function," in *Proceedings of Data Compression Conference*, Snowbird, USA, 2008, pp. 402–411.

[52] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Transaction on Information Theory*, vol. 54, no. 5, pp. 1919–1937, 2008.

[53] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3778–3788, 2006.

[54] V. Koshelev, "Hierarchical coding of discrete sources," *Problemy peredachi informatsii*, vol. 16, no. 3, pp. 31–49, 1980.

[55] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 299–308, 1981.

[56] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 269–275, 1991.

[57] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 253–259, 1994.

[58] R. Zamir and R. Yeung, "Multilevel diversity coding via successive refinement," in *Proc. 1997 IEEE International Symposium on Information Theory*, 1997, p. 265.

[59] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 819–828, 2000.

[60] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[61] L. Lastras-Montano and V. Castelli, "Near sufficiency of random coding for two descriptions," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 681–695, 2006.

[62] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 3991–4005, 2009.

[63] C. Tian, S. Mohajer, and S. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints," in *IEEE International Symposium on Information Theory, 2008. ISIT 2008*, 2008, pp. 1413–1417.

[64] A. Schrijver, *Theory of Linear and Integer Programming*. New York: Wiley, 1998.

[65] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[66] K. Gomadam and S. A. Jafar, "The effect of noise correlation in amplifyand-forward relay networks," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 731–745, Feb. 2009.

[67] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3413–3424, Aug. 2009.

[68] V. Prabhakaran and P. Viswanath, "Interference channel with destination cooperation," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

[69] C. Suh and D. Tse, "Symmetric feedback capacity of the gaussian interference channel to within one bit," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, June 2009.

[70] G. Bresler and D. Tse, "The Two-User Gaussian Interference Channel: A Deterministic View," *European Transactions in Telecommunications*, vol. 19, pp. 333–354, June 2008.

[71] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge Univ Pr, 2005.

[72] A. Amaudruz and C. Fragouli, "Combinatorial algorithms for wireless information flow," in *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms.* Society for Industrial and Applied Mathematics, 2009, pp. 555–564.

[73] G. Bresler and D. Tse, "The Two-User Gaussian Interference Channel: A Deterministic View," *European Transactions in Telecommunications*, vol. 19, no. 4, pp. 333–354, 2008.

[74] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow," in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[75] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 44, pp. 159–175, 1978.

[76] R. Koetter and F. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 3579–3591, Aug. 2008.

[77] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Médard, and M. Effros, "Resilient network coding in the presence of byzantine adversaries," *IEEE Transactions on Information Theory*, vol. 54, pp. 2596–2603, June 2008.

[78] S. M. H. T. Yazdi and M. R. Aref, "The capacity of a class of linear deterministic networks," 2010, available from http://arxiv.org/pdf/1001.2164.

[79] A. Sanderovich, M. Peleg, and S. Shamai (Shitz), "Scaling Laws and Techniques in Decentralized Processing of Interfered Gaussian Channels," in *International Zurich Seminar on Communications, Zurich, Switzerland*, 2008.

[80] S. Diggavi and T. Cover, "The worst additive noise under a covariance constraint," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, 2001.

[81] C. Tian, J. Chen, S. Diggavi, and S. Shamai, "Optimality and Approximate Optimality of Source-Channel Separation in Networks," 2010, submitted to IEEE Transaction on Information Theory, Available from http://arxiv.org/pdf/1004.2648.

Curriculum Vitae – Soheil Mohajer

Education

- **École Polytechnique Fédérale de Lausanne (EPFL)**, Lausanne, Switzerland

  PhD. in Computer and Communication Sciences

  Research topic: Network Informatioy Theory

  Oct. 2005 – Sep. 2010

- **École Polytechnique Fédérale de Lausanne (EPFL)**, Lausanne, Switzerland

  M.Sc. in Communication Systems

  Research topic: Rate-Distortion Theory

  Oct. 2004 – Sep. 2005

- **Sharif University of Technology**, Tehran, Iran

  B.Sc. in Electrical Engineering

  Sep. 1999 – Aug. 2004

Journal Publications

[J1] C. Tian, S. Mohajer and S. N. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints", *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3869–3891, Aug. 2009.

[J2] S. Mohajer, C. Tian, and S. N. Diggavi, "Asymmetric Gaussian multiple descriptions and asymmetric multilevel diversity coding," accepted for publication in *IEEE Transactions on Information Theory.*

[J3] S. Mohajer, S. N. Diggavi, C. Fragouli and D. N. C. Tse, "Approximate capacity of a class of relay-interference networks", submitted to *IEEE Transactions on Information Theory.*

[J4] M. Jafari, S. Mohajer, C. Fragouli, and S. N. Diggavi, "On the capacity of non-coherent network coding", submitted to *IEEE Transactions on Information Theory*.

[J5] A. Hormati, A. Karbasi, S. Mohajer, and M. Vetterli, "An estimation theoretic approach for sparsity pattern recovery in the noisy setting," submitted to *IEEE Transactions on Information Theory*.

[J6] S. Mohajer, P. Pakzad, and A. Kakhbod, "Tight bounds on the redundancy of Huffman codes," submitted to *IEEE Transactions on Information Theory*.

[J7] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing", submitted to *IEEE Transactions on Information Theory*.

Conference Publications

[C1] S. Mohajer, S. N. Diggavi, "On the Gaussian relay channel with adversarial noise inserter," *IEEE Information Theory Workshop*, Dublin, Ireland, September 2010.

[C2] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing", *IEEE International Symposium on Information Theory*, Austin, USA, June 2010.

[C3] S. Mohajer, C. Tian, and S. N. Diggavi, "On source transmission over deterministic relay networks", *IEEE Information Theory Workshop*, Cairo, Egypt, Jan. 2010.

[C4] S. Mohajer, S. N. Diggavi, and D. N. C. Tse, "Approximate capacity of a class of Gaussian relay-interference networks", *IEEE International Symposium on Information Theory*, Seoul, South Korea, July 2009.

[C5] M. Jafari, S. Mohajer, C. Fragouli, and S. N. Diggavi, "On the Capacity of Non-Coherent Network Coding", *IEEE International Symposium on Information Theory*, Seoul, South Korea, July 2009.

[C6] A. Karbasi, A. Hormati, S. Mohajer and M. Vetterli, "Support recovery in compressed sensing: An estimation theoretic approach", *IEEE International Symposium on Information Theory*, Seoul, South Korea, July 2009.

[C7] S. Mohajer, S. N. Diggavi, C. Fragouli and D. N. C. Tse, " Capacity of deterministic Z-chain relay-interference network", *IEEE Information Theory Workshop*, Volos, Greece, June 2009.

[C8] S. Mohajer, S. N. Diggavi, "A deterministic approach to wireless network error correction," *IEEE Information Theory Workshop*, Volos, Greece, June 2009.

[C9] S. Mohajer, M. Jafari, S. N. Diggavi, and C. Fragouli, "On the capacity of multisource non-coherent network coding," *IEEE Information Theory Workshop*, Volos, Greece, June 2009.

[C10] S. Mohajer, S. N. Diggavi, C. Fragouli and D. N. C. Tse, " Transmission techniques for relay-interference networks", *Allerton Conference on Communication, Control, and Computing*, Illinois, USA, September 2008.

[C11] S. Mohajer, C. Tian and S. N. Diggavi, "Asymmetric Gaussian multiple descriptions and asymmetric multilevel diversity coding," *IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[C12] C. Tian, S. Mohajer, and S. N. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints," *IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[C13] S. Mohajer, C. Tian, and S. N. Diggavi, "Asymmetric multi-level diversity coding," *IEEE Data Compression Conference*, Snowbird, Utah, USA, March 2008.

[C14] C. Tian, S. Mohajer and S. N. Diggavi, "On the symmetric Gaussian multiple description rate-distortion function," *IEEE Data Compression Conference*, Snowbird, Utah, USA, March 2008.

[C15] S. Mohajer, A. Kakhbod, "Tight bounds on the AUH codes", *42nd Annual Conference on Information Sciences and Systems, CISS*, Princeton, USA, March 2008.

[C16] C. Tian, S. Mohajer, and S. N. Diggavi, "On the Gaussian K-description problem under symmetric distortion constraints," *Information Theory and Applications Workshop*, San Diego, USA, January 2008.

[C17] S. Mohajer and M. A. Shokrollahi, "Raptor codes with fast hard decision decoding algorithms," *IEEE Information Theory Workshop*, pp. 56–60, Chengdu, China, October 2006.

[C18] S. Mohajer, P. Pakzad, and A. Kakhbod, "Tight bounds on the redundancy of Huffman codes," *IEEE Information Theory Workshop*, Punta del Este, Uruguay, March 2006.