Title
Towards Low-complexity Relay Selection and Scheduling in Wireless Networks

Author
Essa, Yahya Hussain Ezzeldin

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Towards Low-complexity Relay Selection and Scheduling in Wireless Networks

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical and Computer Engineering

by

Yahya Hussain Ezzeldin M. A. Essa

2020

ABSTRACT OF THE DISSERTATION

Towards Low-complexity Relay Selection and Scheduling in Wireless Networks

by

Yahya Hussain Ezzeldin M. A. Essa
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2020
Professor Christina Fragouli, Chair

Next-generation wireless networks are expected to usher in a revolution of communication systems in terms of the network capabilities, technologies used, and serviced applications. A key aim in these networks is to expand available bandwidth and enable multi-gigabit emerging applications that range from ultra-high-definition video to autonomous vehicle platoons and Industry 4.0 low latency communication. The focus of this dissertation is to study two technologies that are envisioned to drive the physical layer in the next generation of networks - multi-hop relay selection and mmWave transmission - in terms of fundamental theoretical guarantees on performance and the efficiency of operation and scheduling.

We study the problem of relay selection for routing in a wireless full-duplex relay network. We derive fundamental worst-case guarantees on the performance of routing in wireless relay networks operating in full-duplex with arbitrary topology. This extends the state of the art in terms of these fundamental guarantees which were previously only studied for the diamond network topology. These worst-case guarantees provide guidance for heuristic and optimal routing approaches in full-duplex wireless networks.

Second, we study the problem of efficiently selecting a subset of relays in a diamond net-

work operating in half-duplex and the fundamental guarantees that follow from the selection of the best relay subset. We prove fundamental guarantees on the retained capacity when a subset of $N-1$ relays are retained out of N relays in a diamond topology, For selecting $k < N-1$ relays, we develop efficient selecting algorithms with theoretical guarantees.

Next, we show that selecting the simple route with the highest approximate capacity with a half-duplex network with arbitrary topology is in general an NP-hard problem (unlike its full-duplex counterpart). Additionally, we provide sufficient properties on the network topology that ensure that finding the best route can be done in polynomial-time in the number of nodes in the network. These results are enabled by a closed-form expression that we derive for the approximate capacity of a half-duplex network that enabled the study of its NP-hardness as well as deriving an efficient routing algorithm.

Finally, we study the impact of directional communication which is envisioned in mmWave communication networks. To tackle this, we propose a new information-theoretic model for multi-hop wireless networks referred to as "1-2-1 network". In a 1-2-1 network, at any time instance, two nodes can communicate only if they point beams (which need to be scheduled) at each other, while if they do not point beams at each other, no signal can be exchanged. We used this model to approximate the Shannon capacity for mmWave networks with arbitrary topology operating with full-duplex or half-duplex mmWave nodes. Additionally, we develop a provably optimal polynomial-time algorithm to compute the approximate capacity and an optimal beam-orientation schedule in full-duplex and half-duplex mmWave networks.

The dissertation of Yahya Hussain Ezzeldin M. A. Essa is approved.

Paulo Tabuada

Lieven Vandenberghe

Suhas Diggavi

Christina Fragouli, Committee Chair

University of California, Los Angeles

2020

To my mother, sister and late father ...

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I cannot picture the finish line to this Ph.D. journey ever being reached without the support from the people who helped shape who I am before and during this journey. Before that, however, I would like to thank Allah for giving me the strength, patience and guidance throughout my life in general and during my Ph.D. journey in particular.

First, I would like to sincerely thank my advisor, Professor Christina Fragouli. Christina has always been a source of continuous support and guidance during my Ph.D. journey. Christina has constantly provided me with insightful information and intuitions about different aspects in research, which on several occasions led to breakthroughs during research struggles. She has also introduced me to several of my collaborators within and outside UCLA, which aided in shaping the direction of my research during my Ph.D. But above all, I always felt that Christina genuinely had my best interest at heart; never shying away from support during my Ph.D. and in planning of my future steps. I have learned a lot of valuable things from Christina that will affect my academic career in a positive way.

I have been lucky to work with Prof. Suhas Diggavi (UCLA) on a number of research projects. I have enjoyed all my technical and non-technical discussions with him and I am very grateful for his continued support and advice regarding my career. I have also enjoyed the opportunity to collaborate with Prof. Daniela Tuninetti (University of Illinois at Chicago) on multiple papers. During this collaboration, I have learned a lot from Prof. Tuninetti contributions and comments. I would also like to thank Prof. Giuseppe Caire (Technical University of Berlin) for the very insightful technical and non-technical discussions we shared during our collaboration. In addition to being an expert information theorist, Prof. Caire also provided a practical perspective on millimeter communication networks, which shaped great aspects of our collaborative research. Furthermore, I have had the opportunity to collaborate with Prof. Paulo Tabuada (UCLA) and have gained valuable insight from his important contributions and comments.

Within our research lab, ARNI, I would like to thank all my collaborators for the delightful discussions and exceptional research experience. Among those, I would like to dedicate thanks to my lab-mates and friends, Dr. Gaurav Agarwal, Sundara Rajan and Osama Hanna for their continuous help and support. I would also like to dedicate special thanks to my longtime friend and collaborator and previous lab member, Prof. Martina Cardone. I am grateful for all the technical insight you taught me as well as your continued support and friendship beyond our technical collaboration. Furthermore, I would like to thank my previous lab mates, Prof. Linqi Song, Dr. Ayan Sengupta and Dr. Iris Safaka. I am grateful for all the (technical and non-technical) help and all the great moments we had together throughout the years.

Next, I would also like to extend my gratitude to my mentors at Alexandria University, Prof. Ahmed Sultan, Prof. Moustafa Youssef, in addition to my advisor (and big brother), Prof. Karim Seddik, for all that I learned from them and for their support.

I would like to thank all my friends I got to know at UCLA, in particular, Dr. Ahmed Hareedy, Dr. Ahmed Alaa, Dr. Omar Hussein and Dr. Moustafa Al-Zantout and many others. I will always cherish our time together at UCLA and how your presence and camaraderie have made the journey smoother. I would like to dedicate a special "thank you" to my longtime friend Dr. Mohammed Karmoose. It is not often that one finds a friend that shares almost every aspect of your life for an extended period of time: working together in Egypt, being apartment mates as well as lab mates at UCLA. In addition to our longterm friendship, you have also always been an inspiration to me and someone that I continuously turn to for advise. I hope that our friendship continues for many years to come.

Finally, my love and sincere gratitude for my father Hussain Ezzeldin, my mother Amal Sharkas and my sister Rahma Ezzeldin. Even though my father passed away long before my Ph.D. journey, the core values and beliefs as well as the love for science he instilled in me as a child still drive me to this day. To my mother and sister, I am blessed for your existence in my life. You have provided me with the strength to continue during the hardest moments

of this journey, and you were there with me at every step on the way. For that I am forever grateful.

VITA

2011	B.S. (Electrical and Electronic Engineering), Alexandria University, Egypt
2011-2014	Teaching Assistant, Electrical and Electronic Engineering, Alexandria University, Egypt
2014	M.S. (Electrical Engineering), Alexandria University, Egypt
2014	Joined Ph.D. program at the ECE Department, UCLA
2016	Teaching Assistant, ECE Department, UCLA
2016	Advanced to Ph.D candidacy
2018	Machine Learning Platform Engineer, Intel.
2019-2020	Dissertation Year Fellow, UCLA.

SELECTED PUBLICATIONS

Y. H. Ezzeldin, M. Cardone, C. Fragouli and D. Tuninetti, "Efficiently finding simple schedules in Gaussian half-duplex relay line networks," 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, 2017, pp. 471-475, doi: 10.1109/ISIT.2017.8006572.

Y. H. Ezzeldin, M. Karmoose and C. Fragouli, "Communication vs distributed computation: An alternative trade-off curve," 2017 IEEE Information Theory Workshop (ITW), Kaohsiung, 2017, pp. 279-283, doi: 10.1109/ITW.2017.8278011.

G. K. Agarwal, **Y. H. Ezzeldin**, M. Cardone and C. Fragouli, "Secure Communication over 1-2-1 Networks," 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, 2018, pp. 196-200, doi: 10.1109/ISIT.2018.8437607.

Y. H. Ezzeldin, M. Cardone, C. Fragouli and D. Tuninetti, "Network Simplification in Half-Duplex: Building on Submodularity," in IEEE Transactions on Information Theory, vol. 65, no. 10, pp. 6801-6818, Oct. 2019, doi:10.1109/TIT.2019.2923994.

Y. H. Ezzeldin, M. Cardone, C. Fragouli and G. Caire, "Polynomial-time Capacity Calculation and Scheduling for Half-Duplex 1-2-1 Networks," 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019, pp. 460-464, doi: 10.1109/ISIT.2019.8849671.

O. A. Hanna, **Y. H. Ezzeldin**, T. Sadjadpour, C. Fragouli and S. Diggavi, "On Distributed Quantization for Classification," in IEEE Journal on Selected Areas in Information Theory, vol. 1, no. 1, pp. 237-249, May 2020, doi: 10.1109/JSAIT.2020.2986467.

Y. H. Ezzeldin, M. Cardone, C. Fragouli and D. Tuninetti, "The Approximate Capacity of Half-Duplex Line Networks," in IEEE Transactions on Information Theory, vol. 66, no. 7, pp. 4449-4467, July 2020, doi: 10.1109/TIT.2020.2998184.

Y. H. Ezzeldin, A. Sengupta and C. Fragouli, "Wireless Network Simplification: The Performance of Routing," in IEEE Transactions on Information Theory, vol. 66, no. 9, pp. 5703-5711, Sept. 2020, doi: 10.1109/TIT.2020.3010990.

Y. H. Ezzeldin, M. Cardone, C. Fragouli and G. Caire, "Gaussian 1-2-1 Networks: Capacity Results for mmWave Communications," in IEEE Transactions on Information Theory, doi: 10.1109/TIT.2020.3034569.

CHAPTER 1

Introduction

Wireless communication systems have become extremely prevalent in our everyday lives over the last two decades. Over the same period, these systems have experienced a huge evolution and growth, going from a few kilobits of data rate in earlier generation networks to multiple gigabits of information in currently deployed 4G systems. These growing wireless systems have supported a simultaneous revolution of wireless mobile data over the past years, where high data rates are no longer expected only in home/office environments but also on the move in a user's hand and pocket. However, this wireless data revolution is expected to bring about an increasing demand for extremely high data rates in the future as a staggering number of devices come online in the upcoming years for a myriad of different applications, such as Internet of Things, vehicular communication, virtual reality, etc.

To support this foreseen explosion in connected devices requiring ubiquitous high data rates, there is a consensus that shifting towards the mmWave spectrum is a key enabling factor in 5G in order to support the extremely high data rates requirements for next-generation applications. This transition towards the higher frequency mmWave spectrum brings along a new bag of challenges and opportunities compared to communication networks (particularly cellular) that have been designed, used and operated in the "beachfront spectrum" - the rather slim range of microwave frequencies from several hundred MHz to a few GHz - in the past two decades.

Communication over the mmWave spectrum suffers from extreme propagation properties including high path loss, atmospheric absorption, low penetration through objects and lim-

ited diffraction ability around obstacles, which in contrast are more manageable in the lower microwave frequencies ranges. These properties make it extremely prohibitive to use wireless communication only as the last stage of communication as in modern cellular communications networks (from base stations to users) and local area networks (from access points to users). In particular, a wireless service provider can no longer rely solely on the powerful wired high-capacity links as the infrastructure of the network. To tackle propagation challenges in mmWave communication, two key cornerstone solutions are proposed: *wireless multi-hop relaying* and *high-gain directional antennas*, which can be used to get around obstacles and combat path loss, respectively. The aim of the present dissertation is to conduct a theoretic study on these two network solutions for next-generation networks and contribute to the understanding of their fundamental operational limits in terms of performance and computational efficiency in wireless networks.

Multi-hop networking is by no means a unique premise for wireless systems; it is after all the driving force for our shared Internet infrastructure. However, the broadcast nature of the wireless medium allows for unique cooperation opportunities between wireless nodes. By allowing such cooperation, the performance of wireless systems can be significantly enhanced. Cooperation comes at the expense of more sophisticated transmission schemes needed to optimally operate the network. It is as a consequence then that in large coverage wireless networks operating with low frequencies, such as current cellular networks, multi-hop communication was reserved to the wired backhaul and wireless can be used for the last hop due to penetrative power of a wireless signal at such frequencies. In mmWave communication, wireless multi-hop becomes a necessity either due to propagation considerations or due to new application scenarios such as in vehicular networks. Small cells can expand capacity and improve coverage, whereas wired backhaul deployment might be expensive and complicated. This can be alleviated by using wireless relaying to provide a backhaul to interconnect small cells dynamically, for example like the highly ambitious Terragraph Facebook project [Nor19]. In this case, it becomes of critical importance to smartly and efficiently

select which nodes should cooperate in order to transfer information across the network. For instance, we would like to have low-complexity selection algorithms which allow us to select a subset of the nodes in the network to *relay* communication between a source and a destination while guaranteeing a target rate. Therefore, we need to understand: *what performance guarantees we should expect when a subset of nodes (relays) cooperate for wireless communication in a multi-hop network?*

The use of highly directional antennas in communication networks can also be dated back to terrestrial microwave links used for low rate backhaul connections. These directional antennas required line of sight alignment and were fixated in place on communication towers. In contrast in 5G networks, directional antennas are envisioned to be *steerable* either to be able to direct transmission towards a moving end user or to schedule the flow of information towards different regions in the wireless backhaul. This new steering and scheduling functionality naturally raises the question of computational efficiency. In particular, given that we want to maximize the throughput for a stream of data in the network: *how often does the beam configuration in the network need to change to operate the network optimally? how computationally expensive is it to find an optimal beam schedule?*

In this dissertation, we develop fundamental guarantees on the performance of a subset of nodes in a wireless relay network in the context of the *network simplification* framework, which provides worst-case guarantees on the fraction of the network capacity that can be retained using a subset of the network. Additionally, we study the complexity of selecting the subset of nodes within a wireless relay network that achieve these guarantees, where we provide efficient polynomial time algorithms for their selection or prove their equivalence to NP-hard selection problems.

To study mmWave communication and its operational properties derived from directional communication, we propose an information-theoretic network model, namely the Gaussian 1-2-1 network that captures the inherent directivity and scheduling of mmWave networks due to beam steering. Using this model, we establish an understanding of the complexity of

scheduling beam orientations as well as inherent network simplification properties of networks constructed using highly directional communication.

In the following, we give a background review of the state of the literature on multi-hop mmWave networks before discussing the key contributions of this dissertation.

1.1 Background

Studies of mmWave communication have focused on profiling the distribution of Signal-to-Interference-plus-Noise Ratio (SINR) in random environments both in cellular and ad-hoc network settings [DFP17, BAH14, TBH16]. mmWave communication in a cellular setting was studied in [DFP17, BAH14] using stochastic geometry to characterize the performance through SINR distributions in a random environment. A similar characterization for an ad-hoc connection between a pair of nodes has also been studied in [TBH16]. These previous works consider communication over a single-hop either between ad-hoc nodes or in a cellular system between a base station and a user equipment. The effectiveness of multi-hop relaying for routing in mmWave networks in random environments has been studied in [LA15,DTF17] where it was shown that multi-hop relaying through line-of-sight connections can improve the network connectivity. In [YPK03], scaling laws for networks with directional antennas were studied using the setting introduced in [GK00] for omnidirectional communication. These results look at order arguments for multiple unicast sessions through routing and do not explore fundamental bounds such as the information theoretical capacity.

Shannon capacity characterization for the classical Gaussian relay network with isotropic transmissions is a long standing open problem, both in FD and HD modes of operation. Several schemes [ADT11], [OD13], [LKE11], [LKK14] have been shown to achieve a rate that is a constant gap (i.e., a value that only depends on the number of nodes and is independent of the channel parameters and operating SNR) away from the Shannon capacity. This is accomplished by showing that the achieved rate is a constant gap away from the well-known

cut-set upper bound [CE79] on the Shannon capacity. For general network topologies, the constant gap is fundamentally linear in the number of nodes N in the network [CO15, WO15], although for some specific FD network topologies, the constant gap can be shown to be sublinear [SWF12, CO12].

For classical Gaussian networks with N relays operating in HD, the approximate capacity characterization is challenging as it requires an additional optimization to schedule the relays over 2^N listen/transmit configuration states. Thus, although the approximate capacity can be computed in polynomial time in the network size for the FD mode [PE14], in HD such a result is known to hold only for a few specific classes of layered networks [EPS14], [JEC19]. Furthermore, although there have been several works that characterize the complexity of the structure of the optimal schedule for Gaussian HD wireless relay networks [BFO16], [CTK14], the problem of efficiently (i.e., in polynomial time in the number of nodes) efficiently finding the schedule optimal for the approximate capacity for any general number of relays N has been very elusive.

In networks with non-isotropic transmissions, a scheduling component naturally arises in optimizing the achieved rates. In [Ari84], it has been shown that the multi-access problem in ad-hoc wireless networks is NP-hard. In [JX06], scheduling in wireless networks under interference constraints has been studied and an approach for scheduling was proposed that is guaranteed to converge to the optimal point. Unfortunately the proposed approach runs in exponential time in the number of links in the network. For spread spectrum networks, it has been proved in [HS88] that link-based transmissions where the restriction is relaxed to disallowing a node to converse with more than one node, can be scheduled in polynomial time in the number of nodes. In [THM12, CCH07], scheduling of network coded flows was studied by modeling broadcast instances as hyperedges in a hypergraph model of the wireless network. Optimal scheduling in the previous model was shown to be possible in polynomial time for a class of networks having claw-free conflict graphs [KM17]. Computing a schedule for broadcast transmissions has also been studied for crossbar switches [SMK07] and was

shown to reduce to the fractional weighted graph coloring problem, which is NP-hard in general.

Different from the aforementioned thread of research, where the main objective is to provide an operating schedule for a network when all the N nodes/relays are active, another consideration in wireless networks is understanding what fraction can be guaranteed when only a subset of $k \leq N$ relays out of N is utilized. This line of work, as aforementioned, is referred to as *network simplification* and it was first studied in [NOF14] for classical Gaussian FD diamond networks. Specifically, [NOF14] showed that in any N-relay classical Gaussian FD diamond network, there always exists a subnetwork of k relays that retains at least $\frac{k}{k+1}$ of the unicast approximate capacity of the full network. This guarantee was shown to be tight, i.e., there exist N-relay Gaussian FD diamond networks for which the best k-relay subnetwork (i.e., the one with the largest approximate capacity) achieves this fraction of the full network approximate FD capacity. Although showing exciting guarantees for the diamond network, there has not been any work that provides similar guarantees for networks of general topology until the onset of the research in this dissertation.

1.2 Summary of Contributions

The focus of this dissertation is to study the problem of efficient relay selection and scheduling in multi-hop relay networks both in the absence and presence of high-gain directional antennas. This work differs from other information-theoretic studies in that we are focused on low-complexity algorithm for scheduling and relay selection and in proving fundamental bounds on these algorithms. Our key contributions in this dissertation can be summarized as follows:

1. We prove fundamental worst-case guarantees on the retained capacity by the best route in a Gaussian full-duplex relay network with arbitrary topology. This extends and generalizes previously known results in the literature for selecting a subset in relays that

was only limited to the diamond network topology. Additionally, this complements the body of work on efficient route selection algorithms in wireless networks by providing fundamental bounds for their performance.

2. We study the capacity of Gaussian half-duplex wireless networks and show that for half-duplex network routes: (a) The capacity can be approximated to within a constant gap by a closed-form expression that can be efficiently computed in linear time similar to its full-duplex counterpart; (b) The schedule that achieves this approximate capacity requires only a linear number of active network states. We also show the surprising result that, even with the similarities between the expression for capacity in full-duplex and half-duplex, routing in half-duplex is inherently an NP-hard problem in general. For a subclass of networks, with some topological constraints, we develop a polynomial time algorithm for selecting the best route in the network.

3. To model the inherent directivity property of mmWave networks and the need to schedule beam orientations, we propose a new information-theoretic model for multi-hop wireless networks referred to as "1-2-1 network". At any time instance in a 1-2-1 network, two nodes can communicate only if they point beams (which need to be scheduled) at each other, while if they do not point beams at each other, no signal can be exchanged or interference can be generated. We use this model to: (a) Characterize the Shannon unicast and multicast capacities, up to a constant gap, for mmWave networks with arbitrary topology operating with full-duplex or half-duplex mmWave nodes; (b) Develop a provably optimal polynomial time algorithm to compute the approximate capacity and an optimal beam-orientation schedule in full-duplex and half-duplex mmWave networks; (c) Characterize fundamental guarantees on operating a subset of relays in a mmWave network as a fraction of the capability of the full mmWave network.

1.3 Outline of Dissertation

The results in this dissertation are organized in what follows into two parts as we study relay selection and scheduling in multi-hop relay networks both in the absence and in presence of high-gain directional antennas, respectively. In **Part I (Chapter 2-4)**, we study multi-hop relaying in wireless networks operating with omnidirectional antennas, while **Part II (Chapter 5-8)** focuses on modeling the abstract effects of highly directional antennas on the overall multi-hop network capacity as well as on relay selection guarantees and network scheduling. We detail the results and problems studied in each chapter below.

1.3.1 Part I : Physical Layer Cooperation

In Chapters 2 and 3, we establish fundamental worst-case guarantees for selecting a subset of relays in a Gaussian full-duplex and half-duplex relay network, respectively. In particular, in **Chapter 2**, we show that selecting a single route within a full-duplex network with arbitrary topology can in the worst-case achieve a fraction of the approximate network capacity that is inversely proportional with the number of nodes in the network. This is in contrast with preliminary results on the problem for the diamond network topology [NOF14] which showed a substantial guaranteed fraction of 1/2 independent of the number of nodes in the network. In **Chapter 3**, we study the problem of selecting relays in a Gaussian half-duplex diamond network and provide fundamental worst-case guarantees on the achieved rates as a fraction of the approximate network capacity.

Chapter 4 studies the efficient scheduling of transmissions in half-duplex line networks (routes), where we establish both a linear-time scheduling algorithm (matching a fundamental existential result for scheduling in half-duplex networks [CTK14]) as well as simple expression for the approximate capacity of the half-duplex line network that mirrors its full-duplex counterpart. By efficient scheduling, we refer to scheduling that can be performed in polynomial time in the network size. In the same chapter, we also show that although both

capacity expressions for full-duplex and half-duplex routes are closely related, finding the best half-duplex route in a network is a fundamentally NP-hard problem. This is in striking contrast with the finding the best full-duplex route can be done efficiently in polynomial time by an array of dynamic programming algorithms such as Dijkstra's algorithm [Dij59].

1.3.2 Part II: Modeling mmWave Transmission

In Part II, we present and study a class of networks that we term 1-2-1 networks that offer a simple yet informative model for mmWave networks that capture the use of high-gain directional antennas and multi-hop communication. To establish a communication link, both the mmWave transmitter and receiver employ antenna arrays that they electronically steer to direct their beams towards each other - we term this a 1-2-1 link, as both nodes need to focus their beams to face each other for the link to be active. Thus, in 1-2-1 networks, instead of broadcasting or interference, we have coordinated steering of transmit and receive beams to activate different links at each time.

Chapter 5 presents the Gaussian 1-2-1 network model and establishes its approximate network capacity both for unicast and multicast traffic in full-duplex and half-duplex modes of operation. In **Chapter 6**, we analyze the derived approximate capacity expression for the full-duplex mode of operation and show that, although it has an exponential number of variables, it can be computed in polynomial time in the number of nodes. Furthermore, we design a polynomial time algorithm to compute a schedule for beam steering that is optimal for the approximate multicast capacity. **Chapter 7** extends the previous study to the Gaussian half-duplex 1-2-1 network and establishes similar results in terms of efficient scheduling and computation. As a result of the different nature of beam orientation in full-duplex and half-duplex, these two modes of operation require a different machinery to develop the result. For instance, in full-duplex, we leverage equivalence between linear programs and the decomposition of stochastic matrices to design the polynomial time algorithm for scheduling. In the half-duplex case, the core technical component consists of

showing that a polynomial time separation oracle exists for our approximate capacity representation, by using algorithm tools such as perfect matching polytopes, Gomory-Hu trees and an algorithmic implementation of Caratheodory's theorem.

In **Chapter 8**, we discuss the limit of abstraction of the 1-2-1 model by relaxing the ideal high-gain directional antenna patterns to imperfect antennas patterns that exhibit side-lobe leakage. We characterize the gap between the approximate capacities of the imperfect and ideal 1-2-1 models for the same channel coefficients and transmit power. We show that, under some conditions, this gap only depends on the number of nodes, i.e., under these conditions, the ideal 1-2-1 network model presented in Chapter 5 provides a very good approximation for the network properties seen in networks with imperfect antenna patterns.

Parts of this dissertation are published in [ESF16, ECF18, ECF19b, ECF20b, ECF20a, ESF20].

Part I

Physical Layer Cooperation

CHAPTER 2

Network Simplification in Full-Duplex Networks: The Performance of Routing

In this chapter, we explore the *network simplification* problem for Gaussian full-duplex relay networks with arbitrary topology. Particularly, given an N-relay Gaussian full-duplex network, the network simplification problem seeks to find fundamental guarantees on the capacity of the best subnetwork, among a particular class of subnetworks, as a fraction of the full-network capacity. The focus of this work is the case when the selected subnetwork class is a route from the source to the destination. The results in this paper show that routing can, in the worst case, result in an unbounded gap from the network capacity - or reversely, physical layer cooperation can offer unbounded gains over routing.

2.1 Introduction

In this chapter, we consider a wireless network where a source wishes to communicate with a destination using the help of a large number of wireless full-duplex relay nodes. The wireless network simplification problem asks the following: what fraction of the capacity of the full network is guaranteed to be retained, when operating only a subset of network relays?

This problem was first studied by the authors in [NOF14] in the context of Gaussian full-duplex diamond networks[1]. The importance of this problem stems from the several

[1] An N-relay diamond network is a two-hop network where the source communicates with the destination through N non-interfering relays.

benefits achieved by operating a subset of network relays. For example, operating all the available relays in the network (in order to achieve rates close to the Shannon capacity) requires complex communication schemes, and might incur a significant cost in terms of consumed power. In contrast, operating a simpler subnetwork (for instance a route) can provide potential solutions in term of these limitations. As a result, it is of interest to understand how the capacity of the selected subnetwork relates to the full network capacity.

In this chapter, we find the solution to the network simplification problem for N-relay Gaussian full-duplex networks with arbitrary topology when the selected subset of relays form a path from the source to the destination. Our focus on this case is motivated by the prevalent use of routing as a key enabler to the deployment of wireless sensor networks [ZLZ14,KAH17], and as a key component in next-generation networks such as device-to-device networks and ultra dense cellular networks [GTM16].

Studying the network simplification problem for full-duplex networks with arbitrary topologies is more challenging as compared to the diamond network [NOF14]. This is due to the more complex algebraic representation of cuts in networks with arbitrary topologies as compared to a simple (nearly fixed) structure when the problem is specialized to only diamond networks. We elaborate more on these differences when we present our result in Section 2.3.

In this chapter, we seek to understand how much of the approximate capacity of an N-relay full-duplex network can be achieved by selecting the best route from the source to the destination (the route that has the largest capacity). In particular, our aim is to develop a worst-case performance guarantee in terms of the fraction achievable by the best route. We seek a universal guarantee that holds independently of the channel coefficients and/or the operating SNR. Our main contributions in this chapter can be summarized as follows:

1. We prove that selecting the best route in the network can in the worst-case achieve a fraction of the approximate capacity of the network that is inversely proportional

with the number of nodes in the network - or reversely, physical layer cooperation can offer gains over routing that grow linearly with the number of nodes in the network. Particularly, we show that in any N-relay Gaussian full-duplex network, there always exists a route that achieves at least a fraction $\frac{1}{\lfloor N/2 \rfloor + 1}$ of the approximate capacity of the network. Furthermore, we prove that that this fractional guarantee is tight, i.e., there exist networks in which the achievable fraction is exactly equal to the capacity of the best route.

2. At the heart of our proof of the aforementioned result, we deal with the problem of analyzing how the capacities of subsets (in terms of antennas) of a MIMO channel behave with respect to the capacity of the full MIMO channel. We derive a simplification result for the Gaussian MIMO channel In particular, we show that for any $n_t \times n_r$ Gaussian MIMO channels, the best $k_t \times k_r$ subchannel approximately achieves at least a fraction $\frac{\min\{k_t,k_r\}}{\min\{n_t,n_r\}}$ of the full MIMO channel capacity, independently of the channel coefficients and/or operating SNR. Moreover, this fractional guarantee is fundamental, i.e., it is the largest fraction of the MIMO channel capacity that can be guaranteed universally across all channel configurations. To the best of our knowledge, this is the first result that provides a fundamental worst-case guarantee on antenna selection for MIMO channels. Beyond its utilization in the proofs of our main results, this property might be of independent interest to characterize the performance of antenna selection algorithms in MIMO channels [JV09, GGP03, MWC05].

Chapter Organization. Section 2.2 describes the N-relay Gaussian full-duplex network and its approximate capacity expression. Section 2.3 discusses the main result in the chapter which gives a universal guarantee on the achievable rate by the best route in a Gaussian full-duplex relay network with arbitrary topology. Section 2.4 derives our main result. Section 2.5 proves a simplification result for MIMO channels with i.i.d inputs which is used as a main ingredient in the proof in Section 2.4. Section 2.6 concludes the discussion in the chapter. Some of the proofs are delegated to the appendices.

2.2 System Model and Preliminaries

Throughout the chapter, we denote with $[a:b]$ the set of integers from a to b, where $b \geq a$. We consider a Gaussian relay network where the *Source* (S) wants to communicate with the *Destination* (D) through the help of N relays operating in full-duplex. The set of all nodes in the network is denoted by $\mathcal{V}$. Nodes in $\mathcal{V}$ are indexed with the integers $[0:N+1]$ where the Source and Destination are indexed by 0 and $N+1$, respectively.

At any time t, the received signal $Y_j[t]$ at node j is a function of the transmitted signals from all other nodes in the network (except D), i.e.,

$$Y_j[t] = \sum_{\substack{i=0,\\ i\neq j}}^{N} h_{ji} X_i[t] + W_j[t], \quad \forall j \in [1:N+1], \tag{2.1}$$

where: (i) $X_i[t]$ is the transmitted signal from the i-th node at time t; (ii) the additive white Gaussian noise $W_j[t] \sim \mathcal{CN}(0,1)$ at j-th node is independent of the inputs, as well as of the noise terms at the other nodes; (iii) the channel gain between nodes i and j is denoted by $h_{ji} \in \mathbb{C}$ and is assumed to be constant for the whole transmission duration and hence known to all nodes. Transmitted signals from each network node satisfy an average power constraint $\mathbb{E}[|X_i|^2] \leq 1 \quad \forall i \in [0:N]$.

The exact Shannon capacity C of the network described in (2.1) is not known in general. However, in [ADT11] the authors showed that it is within a constant gap $G = O(N)$ from the cut-set upper bound evaluated with i.i.d Gaussian input distributions. This evaluation of the cut-set bound is given by

$$\overline{\mathsf{C}} \triangleq \min_{\substack{\Omega\in 2^{\mathcal{V}},\\ 0\in\Omega, N+1\in\Omega^c}} \overline{\mathsf{C}}(\Omega, \mathcal{V}), \tag{2.2}$$

where

$$\overline{\mathsf{C}}(\Omega, \mathcal{V}) \triangleq \log\det\left(\mathbf{I} + \mathbf{H}_\Omega \mathbf{H}_\Omega{}^\dagger\right). \tag{2.3}$$

The matrix $\mathbf{H}_\Omega$ represents a MIMO channel matrix from transmitting nodes in Ω to receiving nodes in $\Omega^c = \mathcal{V}\backslash\Omega$. We refer to $\Omega \subseteq \mathcal{V}$ as a *"cut"* in the network. In the rest of the chapter,

we work with the approximate capacity $\overline{\mathsf{C}}$ in place of the network capacity to prove our results.

In an N-relay Gaussian network, we denote the capacity of the point-to-point channel between node i and node j as $R_{i\to j} \triangleq \log\left(1+|h_{ji}|^2\right),\ \forall i,j \in [0:N+1]$. A path (route) $\mathcal{P}$ in an N-relay Gaussian full-duplex network is defined by a sequence of $|\mathcal{P}|+1$ non-repeating nodes $\{v_0, v_1, \ldots, v_{|\mathcal{P}|}\}$, where $v_0 = 0$, $v_{|\mathcal{P}|} = N+1$ and $v_i \in [1:N], \forall i \in [1:|\mathcal{P}|-1]$. The path $\mathcal{P}$, therefore, defines a line network from S to D induced by the links connecting nodes v_{i-1} and v_i for $i \in [1:|\mathcal{P}|]$. The capacity of the path $\mathcal{P}$ is denoted by $\mathsf{C}_{\mathcal{P}}$ and is known to be equal to

$$\mathsf{C}_{\mathcal{P}} = \min_{0\le i\le |\mathcal{P}|-1} R_{v_i\to v_{i+1}}, \tag{2.4}$$

and can be achieved through the Decode-And-Forward scheme [EK11].

2.3 Main Results

The main result of this chapter is a universal guarantee on the retained capacity when selecting the best route in a Gaussian full-duplex network with arbitrary topology. The result is stated below in Theorem 2.3.1.

Theorem 2.3.1. In any N-relay Gaussian full-duplex network with approximate capacity $\overline{\mathsf{C}}$, there exists a path $\mathcal{P}$ (line network) with capacity $\mathsf{C}_{\mathcal{P}}$ that satisfies

$$\mathsf{C}_{\mathcal{P}} \ge \underbrace{\frac{1}{\lfloor N/2\rfloor+1}}_{\alpha(N)}\overline{\mathsf{C}} - 2\log\left(\frac{N+2}{2}\right). \tag{2.5}$$

Moreover, there exists a class of networks with N relays such that for all paths $\mathcal{P}$,

$$\mathsf{C}_{\mathcal{P}} \le \frac{1}{\lfloor N/2\rfloor+1}\overline{\mathsf{C}}. \tag{2.6}$$

Proof Sketch: We formally prove Theorem 2.3.1 in Section 2.4. The tightness of the results is proved by providing network constructions where the bound in (2.6) is satisfied. The lower bound in (2.5) is proved by contradiction. At its core, the proof of the lower bound in Theorem 2.3.1 relies on the fact that the approximate capacity $\overline{\mathsf{C}}$ in (2.2) is given by the minimum capacity among 2^N MIMO channels with i.i.d inputs, which are defined by the 2^N network cuts. We will develop a set of upper bounds on the approximate capacity $\overline{\mathsf{C}}$ by upper bounding the capacity of each of these MIMO channels in terms of its best SISO subchannel. Afterwards, we show that if the result in Theorem 2.3.1 is false, then at least one of the aforementioned upper bounds is violated. Towards developing such upper bounds on the approximate capacity, we make use of the second main result of the chapter which gives a universal upper bound bound on the capacity of an $n_t \times n_r$ MIMO channel with i.i.d inputs using its best $k_t \times k_r$ subchannel. The result is stated in the following theorem.

Theorem 2.3.2. For any $n_t \times n_r$ Gaussian MIMO channel with i.i.d inputs and capacity C_{n_t,n_r}, the best $k_t \times k_r$ subchannel has a capacity $\mathsf{C}^{\star}_{k_t,k_r}$ such that

$$\mathsf{C}^{\star}_{k_t,k_r} \geq \frac{\min(k_t,k_r)}{\min(n_t,n_r)}\, \mathsf{C}_{n_t,n_r} - \log\left(\binom{n_t}{k_t}\binom{n_r}{k_r}\right). \tag{2.7}$$

Moreover, this bound is tight up to a constant gap, i.e., there exist $n_t \times n_r$ channels for which

$$\mathsf{C}^{\star}_{k_t,k_r} \leq \frac{\min(k_t,k_r)}{\min(n_t,n_r)}\, \mathsf{C}_{n_t,n_r}.$$

Before delving into the proofs of Theorem 2.3.1 and Theorem 2.3.2 in the following sections, we remark on the results in the two theorems and how they relate to results in the literature.

Remark 2.3.1. While the result in Theorem 2.3.1 is true in general for arbitrary topologies, we can get tighter characterizations for the ratio $\alpha(N)$ if we are interested in specific classes of network topologies. For instance, the result in [NOF14, Theorem 1] proves that, in diamond networks, we have $\alpha(N) = \frac{1}{2}$. In particular, using the same machinery that is employed in

this chapter to prove Theorem 2.3.1, we can prove that, for layered networks with L layers and $N_L \geq 2$ relays, the ratio $\alpha(N)$ in Theorem 2.3.1 is replaced with $\alpha(L, N_L)$ below [ESF16]

$$\alpha(L, N_L) = \begin{cases} \dfrac{2}{(L-1)N_L + 4}, & \text{number of layers } L \text{ is odd} \\ \dfrac{2}{LN_L + 2}, & \text{number of layers } L \text{ is even,} \end{cases} \tag{2.8}$$

which subsumes the aforementioned result proved in [NOF14] for diamond networks.

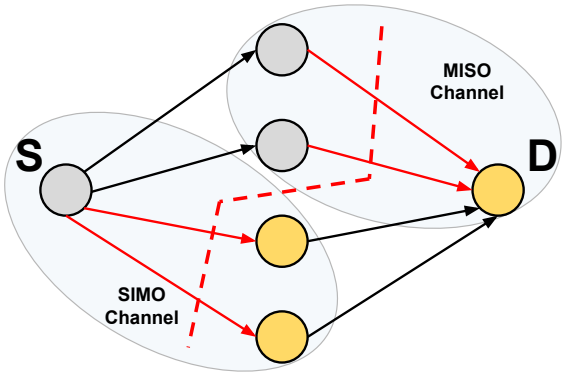


Figure 2.1: Example of a cut in a diamond full-duplex diamond network.

Remark 2.3.2. Theorem 2.3.1 and the ratio in (2.8) highlight that, the independence of the ratio from the number of nodes in the network is a unique property of diamond networks. This is due to the fact that, in a diamond network, a cut can be represented as the sum of a SIMO and a MISO orthogonal channels (which each have a minimum channel dimension of 1) as seen in Fig. 2.1. In contrast, in an arbitrary topology network, a cut can be represented by MIMO channels of different minimum dimensions, and thus the approximate capacity using physical layer cooperation can be potentially much higher as compared to a single route in the network. Our result in Theorem 2.3.1 gives a universal characterization of the ratio between these two values in the worst case.

Remark 2.3.3. Note that for arbitrary topology networks, the smallest relevant subnetwork to consider is a route from source to destination. If the problem is presented instead as selecting k relays in N-relay similar to [NOF14], then the worst-case guarantee given by the

simplification problem is trivially zero for all $k < N$. To illustrate this, consider an N-relay line network that has capacity unity. In this example, by selecting any subset of $k < N$ relays, the capacity we have for the subnetwork is zero as the source and destination are now disconnected. On another note, if we expand our interest to multiple routes, then through the same example, we would arrive at the same worst-case guarantees since there is only one available route in the network.

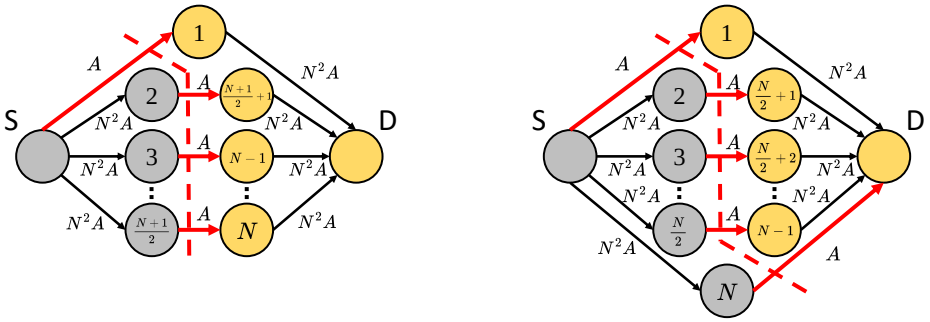
Remark 2.3.4. Although the lower bound in Theorem 2.3.2 is stated for MIMO channels with i.i.d inputs, the result also holds (with a different constant gap) for Gaussian MIMO channels without the i.i.d inputs restriction. This directly follows by noting that, for an $n_t \times n_r$ MIMO channel, the gap between rates achieved by the optimal (waterfilling) solution and the i.i.d equal power allocation solution is at most $n_t(1 + \frac{1}{e})$, independently of the channel coefficients and/or SNR. This gap between the two solutions was proved in [ADT11, Appendix F]. To the best of our knowledge, this is the first tight lower bound on the capacity of a $k_t \times k_r$ MIMO subchannel in terms of the full $n_t \times n_r$ MIMO channel capacity that can be applied to any MIMO subchannel dimensions.

2.4 Proof of Theorem 2.3.1

We start by showing that the guarantee stated in Theorem 2.3.1 is fundamental. To show this, we present constructions for N-relay Gaussian networks in which the guarantee is tight up to a constant gap. Afterwards, we prove the lower bound stated in the theorem.

2.4.1 Tightness of the Bound in Theorem 2.3.1

Consider the two network constructions shown in Fig. 2.2. Note that in both constructions, all the non-highlighted links (in black color) have point-to-point link capacities equal to N^2A for $A > 0$. Thus, any cut that includes a non-highlighted link will have a value of at least N^2A. It is not difficult to see that Fig. 2.2 shows the only cut that does contain a non-



(a) N odd. (b) N even.

Figure 2.2: N-relay network constructions where the capacity that every route achieves is upper bounded by $\frac{1}{1+N/2}$ of the approximate capacity of the full network. Edge labels indicate the capacity of the corresponding links. Red links and node colors highlight the minimum cut in each network construction.

highlighted link for both odd and even number of relays N, and that the value of the shown cut is $(\lfloor N/2 \rfloor + 1)\, A$ in both cases. Therefore, the illustrated cut is the minimum cut in the network, and we have that $\overline{\mathsf{C}} = (\lfloor N/2 \rfloor + 1)\, A$. Additionally, any path from the source to the destination contains a link with a point-to-point capacity equal to A. As a result, for all paths $\mathcal{P}$, we have that $\mathsf{C}_{\mathcal{P}} \leq A = \frac{1}{\lfloor N/2 \rfloor + 1} \overline{\mathsf{C}}$, which proves (2.6) in Theorem 2.3.1.

2.4.2 Proof of the Lower Bound in Theorem 2.3.1

From Section 2.2, the approximate capacity of an N-relay Gaussian network is given by

$$\overline{\mathsf{C}} \triangleq \min_{\substack{\Omega \in 2^{\mathcal{V}}, \\ 0 \in \Omega, N+1 \in \Omega^c}} \overline{\mathsf{C}}(\Omega, \mathcal{V}), \tag{2.9}$$

where $\mathcal{V}$ is the set of all nodes in the network, and

$$\overline{\mathsf{C}}(\Omega, \mathcal{V}) \triangleq \log \det \left(\mathbf{I} + \mathbf{H}_{\Omega} \mathbf{H}_{\Omega}^{\dagger} \right). \tag{2.10}$$

$\overline{\mathsf{C}}(\Omega, \mathcal{V})$ is equal to the capacity of the Gaussian MIMO channel between nodes in Ω and Ω^c, using i.i.d inputs.

Our proof of the lower bound in Theorem 2.3.1 can be summarized in two steps. First,

we derive an upper bound on $\overline{\mathsf{C}}$ by bounding each of the expressions $\overline{\mathsf{C}}(\Omega, \mathcal{V})$ in (2.9) using Theorem 2.3.2. Afterwards, we show that if the lower bound in Theorem 2.3.1 is assumed to be false, then there is a contradiction in the upper bound derived in the first step.

Upper bound: By applying Theorem 2.3.2 with $k_t = k_r = 1$, we can upper bound $\overline{\mathsf{C}}$ as follows

$$\begin{aligned}
\overline{\mathsf{C}} &\leq \overline{\mathsf{C}}(\Omega, \mathcal{V}), && \forall \Omega \subseteq \mathcal{V}, 0 \in \Omega, N+1 \in \Omega^c \\
&\overset{(a)}{\leq} \min(|\Omega|, |\Omega^c|) \max_{i \in \Omega, j \in \Omega^c} R_{i \to j} + \min(|\Omega|, |\Omega^c|) \log\left(|\Omega||\Omega^c|\right), && \forall \Omega \subseteq \mathcal{V}, 0 \in \Omega, N+1 \in \Omega^c \\
&\overset{(b)}{\leq} \left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) \max_{i \in \Omega, j \in \Omega^c} R_{i \to j} + 2\left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) \log\left(\frac{N+2}{2}\right), && \forall \Omega \subseteq \mathcal{V}, 0 \in \Omega, N+1 \in \Omega^c,
\end{aligned} \tag{2.11}$$

where: (a) follows from Theorem 2.3.2 and the fact that $\overline{\mathsf{C}}(\Omega, \mathcal{V})$ represents a MIMO channel with i.i.d inputs; (b) follows since $\max\limits_{\Omega} \min(|\Omega|, |\Omega^c|) = \lfloor N/2 \rfloor + 1$ and $\max\limits_{\Omega} \log(|\Omega||\Omega^c|) \leq 2\log((N+2)/2)$.

Contradiction: We can now prove Theorem 2.3.1 by contradiction using the bound in (2.11). Assume that the lower bound stated in Theorem 2.3.1 is false. Thus, for all paths $\mathcal{P}$ in the network, we have

$$\mathsf{C}_{\mathcal{P}} < \frac{1}{\lfloor N/2 \rfloor + 1} \overline{\mathsf{C}} - 2\log\left(\frac{N+2}{2}\right). \tag{2.12}$$

Now, let $\mathcal{B}$ be the set of links that have a capacity strictly less than the RHS of (2.12). The inequality in (2.12) implies that every path $\mathcal{P}$ contains at least one link that belongs to $\mathcal{B}$ (recall that the capacity of a path is equal to its minimum link capacity). Therefore, removal of the links in $\mathcal{B}$ disconnects the source and destination. As a result, we can construct a cut $\Omega_{\mathcal{B}}$ that is compromised completely of links from $\mathcal{B}$, and we would have that

$$\max_{i \in \Omega_{\mathcal{B}}, j \in \Omega_{\mathcal{B}}^c} R_{i \to j} < \frac{1}{\lfloor N/2 \rfloor + 1} \overline{\mathsf{C}} - 2\log\left(\frac{N+2}{2}\right). \tag{2.13}$$

Given the upper bound in (2.13), we would get a contradiction by applying (2.11) for $\Omega_{\mathcal{B}}$ as follows

$$\overline{\mathsf{C}} \leq \left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) \max_{i \in \Omega_{\mathcal{B}}, j \in \Omega_{\mathcal{B}}^c} R_{i \to j} + 2\left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) \log\left(\frac{N+2}{2}\right)$$

$$< \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right) \left[\frac{1}{\lfloor N/2 \rfloor + 1} \overline{\mathsf{C}} - 2 \log \left(\frac{N+2}{2} \right) \right] + 2 \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right) \log \left(\frac{N+2}{2} \right) = \overline{\mathsf{C}}.$$

Therefore, the lower bound in Theorem 2.3.1 cannot be false, which concludes our proof.

2.5 Proof of Theorem 2.3.2 : A Simplification Result for MIMO Channels

2.5.1 Tightness of the Bound

To prove that the bound in Theorem 2.3.2 is tight (to within a constant gap), it suffices to provide a MIMO channel for which the best $k_t \times k_r$ subchannel has a capacity that is exactly the fraction in Theorem 2.3.2. Towards this end, consider the $n \times n$ parallel MIMO channel with the parallel point-to-point link capacities equal to $A > 0$. The capacity $\mathsf{C}_{n,n}$ of this MIMO channel is nA. For any $k_t, k_r \leq n$, it is not difficult to see that a $k_t \times k_r$ MIMO subchannel can at most capture $\min(k_t, k_r)$ parallel links. Therefore, we have that $\mathsf{C}^{\star}_{k_t,k_r} = \min(k_t, k_r)A$, which proves that

$$\mathsf{C}^{\star}_{k_t,k_r} \leq \frac{\min(k_t, k_r)}{\min(n_t, n_r)} \mathsf{C}_{n_t,n_r},$$

for some network constructions. This proves the tightness statement in Theorem 2.3.2.

2.5.2 Proof of the Lower Bound in Theorem 2.3.2

In this subsection, we derive the lower bound in Theorem 2.3.2. Towards this end, we first state three special cases of the intended bound, and then show how these bounds can be combined to prove the lower bound in Theorem 2.3.2. The aforementioned three bounds are stated in Lemma 2.5.1 and are proved in Appendix 2.8.1.

Lemma 2.5.1. For any $n_t \times n_r$ Gaussian MIMO channel with i.i.d inputs and capacity C_{n_t,n_r} where $n_t \leq n_r$, the best $k_t \times k_r$ MIMO subchannel has a capacity $\mathsf{C}^{\star}_{n_t,k_r}$ that satisfies the following

(a) For $k_t = n_t$ and $k_r \leq n_t \leq n_r$,

$$\mathsf{C}^{\star}_{k_t,k_r} \geq \frac{k_r}{n_t}\ \mathsf{C}_{n_t,n_r} - \log\left(\frac{\binom{n_r}{k_r}}{\binom{n_t}{k_r}}\right). \tag{2.14a}$$

(b) For $k_t = n_t$ and $n_t < k_r \leq n_r$,

$$\mathsf{C}_{n_t,n_r} \geq\ \mathsf{C}^{\star}_{n_t,k_r} \geq\ \mathsf{C}_{n_t,n_r} - \log\left(\frac{\binom{n_r}{k_r}}{\binom{n_r-n_t}{k_r-n_t}}\right). \tag{2.14b}$$

(c) For $k_t \leq n_t$ and $k_r \leq n_r$,

$$\mathsf{C}^{\star}_{k_t,k_r} \geq \frac{k_t \cdot k_r}{n_t \cdot n_r}\ \mathsf{C}_{n_t,n_r} \tag{2.14c}$$

Proof. The proof relies on a result that relates the determinants of the principal submatrices of a positive semidefinite matrix to its eigenvalues. The proof is delegated to Appendix 2.8.1. □

Remark 2.5.1. Although Lemma 2.5.1 is stated for the case where $n_t \leq n_r$, it can also be applied to the opposite case by considering the reciprocal MIMO channel or appealing to Sylvester's determinant identity.

Remark 2.5.2. For MIMO channels with dimensions satisfying constraints in Lemma 2.5.1, the guarantee in Lemma 2.5.1(c) gives a looser fractional guarantee than the one in (a) and (b). However, a key limitation of the bounds in Lemma 2.5.1(a)-(b), is that they allow reduction only in the channel side that contains the larger number of elements; for instance, to only decrease the number of receivers when $n_r \geq n_t$. In contrast, Lemma 2.5.1(c) can be applied without any constraints on the MIMO channel dimensions.

We are now ready to derive the bound on $\mathsf{C}^{\star}_{k_t,k_r}$ in Theorem 2.3.2 for any chosen dimension $k_t \times k_r$ by sequentially applying the bounds in Lemma 2.5.1. Without loss of generality, we assume that $n_t \leq n_r$ since otherwise, we can consider the reciprocal channel. The proof roughly goes as follows: from the $n_t \times n_r$ channel, we can create an $n_t \times k_r$ subchannel such

that $\mathsf{C}^{\star}_{n_t,k_r} \geq \frac{\min(k_r,n_t)}{n_t}\ \mathsf{C}_{n_t,n_r} - G_1$, by keeping only the best k_r receiver antennas; next, from this $n_t \times k_r$ channel, we can get a $k_t \times k_r$ subchannel such that

$$\mathsf{C}^{\star}_{k_t,k_r} \geq \frac{\min(k_t,k_r)}{\min(n_t,k_r)}\ \mathsf{C}^{\star}_{n_t,k_r} - G_2 \geq \frac{\min(k_t,k_r)}{\min(n_t,n_r)}\ \mathsf{C}_{n_t,n_r} - G_1 - G_2.$$

Formally, the constants G_1 and G_2 and the sequence of applying the bounds (2.14) in Lemma 2.5.1 are captured by the following three cases:

1. For $k_t \leq k_r \leq n_t \leq n_r$:

$$\begin{aligned}
\mathsf{C}^{\star}_{k_t,k_r} &\overset{(a)}{\geq} \frac{k_t}{k_r}\ \mathsf{C}^{\star}_{n_t,k_r} - \log\left(\frac{\binom{n_t}{k_t}}{\binom{k_r}{k_t}}\right) \\
&\overset{(b)}{\geq} \frac{k_r}{n_t}\frac{k_t}{k_r}\mathsf{C}_{n_t,n_r} - \frac{k_t}{k_r}\log\left(\frac{\binom{n_r}{k_r}}{\binom{n_t}{k_r}}\right) - \log\left(\frac{\binom{n_t}{k_t}}{\binom{k_r}{k_t}}\right) \\
&\geq \frac{k_t}{n_t}\ \mathsf{C}_{n_t,n_r} - \log\left(\binom{n_t}{k_t}\right) - \log\left(\binom{n_r}{k_r}\right),
\end{aligned}$$

   where: (a) follows by applying (2.14a) on the reciprocal of the $n_t \times k_r$ MIMO channel; (b) applies (2.14a) to relate $\mathsf{C}^{\star}_{n_t,k_r}$ to C_{n_t,n_r}.

2. For $k_r \leq k_t \leq n_t \leq n_r$:

$$\begin{aligned}
\mathsf{C}^{\star}_{k_t,k_r} &\overset{(c)}{\geq} \mathsf{C}^{\star}_{n_t,k_r} - \log\left(\frac{\binom{n_t}{k_t}}{\binom{n_t-k_r}{k_t-k_r}}\right) \\
&\overset{(d)}{\geq} \frac{k_r}{n_t}\ \mathsf{C}_{n_t,n_r} - \log\left(\frac{\binom{n_t}{k_t}}{\binom{n_t-k_r}{k_t-k_r}}\right) - \log\left(\frac{\binom{n_r}{k_r}}{\binom{n_t}{k_r}}\right) \\
&\geq \frac{k_r}{n_t}\ \mathsf{C}_{n_t,n_r} - \log\left(\binom{n_t}{k_t}\right) - \log\left(\binom{n_r}{k_r}\right),
\end{aligned}$$

   where: (c) relates C_{k_t,k_r} to C_{n_t,k_r} using the bound in (2.14b); relation (d) follows by applying (2.14a) to the $n_t \times n_r$ MIMO channel.

3. For $k_t \leq n_t \leq k_r \leq n_r$:

$$\mathsf{C}^{\star}_{k_t,k_r} \overset{(e)}{\geq} \frac{k_t}{n_t}\ \mathsf{C}^{\star}_{n_t,k_r}$$

$$\overset{(f)}{\geq} \frac{k_t}{n_t}\, \mathsf{C}_{n_t,n_r} - \log\left(\frac{\binom{n_r}{k_r}}{\binom{n_r-n_t}{k_r-n_t}}\right)$$
$$\geq \frac{k_t}{n_t}\, \mathsf{C}_{n_t,n_r} - \log\left(\binom{n_t}{k_t}\right) - \log\left(\binom{n_r}{k_r}\right),$$

where (e) follows by applying (2.14c) to select a $k_t \times k_r$ subchannel from the $n_t \times k_r$ MIMO channel; (f) follows from (2.14b).

By combining the three aforementioned cases, we have the lower bound stated in Theorem 2.3.2. This concludes the proof of Theorem 2.3.2.

2.6 Conclusion

In the chapter, we investigated the network simplification problem in an N-relay Gaussian full-duplex network with arbitrary topology. We proved that there always exists a route in the network that retains at least $\frac{1}{\lfloor N/2 \rfloor + 1}$ of the approximate capacity of the full network and that this represents the best worst-case fraction guarantee for the rate achievable by routing. This was proved by showing that if this result is false, then an upper bound to the min-cut expression of the approximate capacity is contradictory. This upper bound on the approximate capacity was obtained by leveraging a simplification result for MIMO channels that was derived here and which might be of independent interest for other applications.

2.7 Related Work

For the Gaussian full-duplex relay networks, the capacity characterization is not known in general. The tightest known universal upper bound for the capacity is the information-theoretic cut-set upper bound [CG79]. In [ADT11], the authors first showed that the network can achieve a rate that is a constant gap approximation of the Shannon capacity of the network through the quantize-map-and-forward relaying strategy. In particular, [ADT11] shows that restricting the optimization of the cut-set bound over independent channel input

distributions approximates the Shannon capacity to within a constant additive gap, i.e., a gap that is independent of the channel coefficients and/or the operating SNR. Similar following strategies [LKE11] and [LKK14] have been shown to achieve a similar result. In this chapter, we refer to this constant gap approximation of the capacity as the *approximate capacity* of the network. One class of Gaussian full-duplex relay networks for which the capacity is explicitly characterized are the line networks, where the relays are arranged in a cascaded manner forming a path from the source to the destination. For this path (route), the capacity is achieved through the decode-and-forward scheme [EK11], and is equal to the value of the bottleneck point-to-point link capacity along the path.

In the thread of work on wireless network simplification, [NOF14] studied the problem for the Gaussian full-duplex diamond network and provided universal capacity guarantees for selecting k-relay subnetworks. They showed that selecting k out of N relays in the network is always guaranteed to achieve at least $\frac{k}{k+1}$ of the full network capacity, to within a constant gap. When applied to a single route selection, [NOF14] states that a route in a diamond relay network can always approximately achieve half of the capacity of the diamond network. The work in [NEO11] extended the result in [NOF14] for some scenarios of the Gaussian full-duplex diamond network with multiple antennas at the nodes. As a scheme-specific performance guarantee (as opposed to guaranteeing capacity fractions), the work of [AJC12] proved upper bounds on multiplicative and additive gaps for relay selection based on the amplify-and-forward scheme, primarily for diamond full-duplex networks. In [AA16], the authors characterized the performance of network simplification (in terms of achievable secure capacity) for layered full-duplex relay networks operating using amplify-and-forward in the presence of an eavesdropper.

From a system design point of view, there is a rich literature of routing protocols for general wireless ad-hoc networks [PR99], [JM96], [CJ03], [PB94] as well as specialized routing protocols for wireless sensor networks [KAH17, ZLZ14]. The goal of this rich literature of routing protocols is to pick a route from the source to the destination while reducing the

network overhead associated with discovering and maintaining this route. However, to the best of our knowledge, there has not been any work that compares the rate achieved by a route, contrast to physical layer cooperation between relay nodes, in the worst-case prior to the first work on wireless network simplification [NOF14]. The notion of studying worst-case capacity results has been considered in wireless sensor networks, in the context of finding sensor topologies that result in the worst many-to-one rate [Mos07]. In this setting, physical layer cooperation is not employed, and communication needs to be scheduled based on either protocol-level constraints or physical-level constraints. In contrast, the focus in the network simplification problem is on the worst-case ratio between the unicast capacity of the best subnetwork and the unicast capacity of the full network (that uses physical layer cooperation).

To prove our main result in this chapter, we first set out to prove a universal guarantee on antenna selection in MIMO channels with i.i.d inputs, i.e., channel inputs that are independent and identically distributed. A number of efficient algorithms have been proposed in the literature for antenna selection in a MIMO channel [GGP03, MWC05, SN04]. Some antenna selection algorithms [JV09] have been proved to be optimal in the diversity-multiplexing tradeoff sense. In contrast to these results, our result does not provide an algorithm for antenna selection but instead, provides a tight worst-case performance for any optimal antenna selection algorithm independently of the SNR and channel coefficients.

2.8 Appendices

2.8.1 Proof of Lemma 2.5.1

Let $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ be the channel matrix for the $n_t \times n_r$ MIMO channel with $n_t \leq n_r$. Without loss of generality, we assume that each transmitter uses unity power, i.e., $P = 1$. Otherwise, we can rewrite $\mathsf{C}_{n_t,n_r} = \log\det(\mathbf{I} + P\mathbf{H}\mathbf{H}^\dagger)$ as $\log\det(\mathbf{I} + \tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger)$ where $\tilde{\mathbf{H}} = \sqrt{P}\mathbf{H}$. Define $\mathbf{F} \triangleq \mathbf{I} + \mathbf{H}\mathbf{H}^\dagger$ and denote the eigenvalues of $\mathbf{F}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_r} \geq 1$.

By selecting a MIMO subchannel of size $n_t \times k_r$ as in Lemma 2.5.1(a)-(b), we are selecting a *principal submatrix* of $\mathbf{F}$ by only keeping the rows and columns indexed by the k_r selected receiver indices. Therefore, our bounds in Lemma 2.5.1 will be proved by relating the maximum determinant of a $k_r \times k_r$ submatrix of the Hermitian matrix $\mathbf{F}$ with the determinant of the full $n_r \times n_r$ matrix $\mathbf{F}$. To do this, we employ the following result that relates the determinants of all principal submatrices of a positive semidefinite matrix to the eigenvalues of the full matrix.

Lemma 2.8.1. Let $\mathbf{A} \in \mathbb{C}^{n_r \times n_r}$ be a positive semidefinite matrix with eigenvalues $\{\beta_i\}_{i=1}^{n_r}$. For a subset $\Lambda \subseteq [1 : n_r]$, define $\mathbf{A}_\Lambda$ to be a principal submatrix of $\mathbf{A}$ that is constructed by only keeping the rows and columns of $\mathbf{A}$ indexed by Λ. Then, for all $k_r \in [1 : n_r]$, we have that

$$\sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \det(\mathbf{A}_\Lambda) = \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j \in \Lambda} \beta_j. \tag{2.15}$$

Proof. The proof of Lemma 2.8.1 is delegated to Appendix B. □

In the following subsections, we show how to prove each of the bounds in Lemma 2.5.1 using (2.15).

2.8.1.1 Proof of Lemma 2.5.1(a)

The assumption for Lemma 2.5.1(a) is that $k_r \leq n_t \leq n_r$. Recall that, $\mathbf{F} = \mathbf{I} + \mathbf{H}\mathbf{H}^\dagger$ is a positive definite matrix. Thus, we can apply Lemma 2.8.1 to the matrix $\mathbf{F}$. Furthermore, the expression in (2.15) can be lower bounded when $k_r \leq n_t$ to be

$$\sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \det(\mathbf{F}_\Lambda) = \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j \in \Lambda} \lambda_j$$

$$\overset{(a)}{\geq} \binom{n_t}{k_r} \sum_{\substack{\Lambda \subseteq [1:n_t], \\ |\Lambda| = k_r}} \frac{1}{\binom{n_t}{k_r}} \prod_{j \in \Lambda} \lambda_j \overset{(b)}{\geq} \binom{n_t}{k_r} \prod_{\substack{\Lambda \subseteq [1:n_t], \\ |\Lambda| = k_r}} \left(\prod_{j \in \Lambda} \lambda_j \right)^{\binom{n_t}{k_r}^{-1}}$$

$$= \binom{n_t}{k_r} \left(\prod_{i=1}^{n_t} \lambda_i \right)^{\binom{n_t-1}{k_r-1}\binom{n_t}{k_r}^{-1}} \overset{(c)}{=} \binom{n_t}{k_r} \left(\prod_{i=1}^{n_r} \lambda_i \right)^{\frac{k_r}{n_t}}, \tag{2.16}$$

where: (a) follows by considering only k_r-tuples of the eigenvalues $\{\lambda_i\}_{i=1}^{n_t}$; (b) follows from the AM-GM inequality; (c) follows by the simplification of the exponent and the fact that $\lambda_i = 1$ for $i \in \{n_t + 1, \dots n_r\}$. By averaging the LHS of (2.16), we have that

$$\frac{1}{\binom{n_r}{k_r}} \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \det(\mathbf{F}_\Lambda) \geq \frac{\binom{n_t}{k_r}}{\binom{n_r}{k_r}} \left(\prod_{i=1}^{n_r} \lambda_i \right)^{\frac{k_r}{n_t}},$$

This implies that there exists a selection Λ_s of k_r receivers such that, the matrix $\mathbf{F}_{\Lambda_s} = \mathbf{I} + \mathbf{H}_{\Lambda_s} \mathbf{H}_{\Lambda_s}{}^\dagger$ satisfies

$$\log\det(\mathbf{F}_{\Lambda_s}) \geq \log \left(\frac{\binom{n_t}{k_r}}{\binom{n_r}{k_r}} \left(\prod_{i=1}^{n_r} \lambda_i \right)^{\frac{k_r}{n_t}} \right) = \frac{k_r}{n_t} \log\det(\mathbf{F}) - \log \left(\frac{\binom{n_r}{k_r}}{\binom{n_t}{k_r}} \right).$$

Since $\mathsf{C}_{n_t,n_r} = \log\det(\mathbf{F})$ and the capacity of the best $n_t \times k_r$ subchannel $\mathsf{C}^\star_{n_t,k_r} \geq \log\det(\mathbf{F}_{\Lambda_s})$, then we arrive at the bound given in Lemma 2.5.1(a).

2.8.1.2 Proof of Lemma 2.5.1(b)

The assumption for this part of Lemma 2.5.1 is that $n_t \leq k_r \leq n_r$. Since $k_r \geq n_t$, then there exists at least one set Λ with cardinality k_r such that $[1 : n_t] \subseteq \Lambda \subseteq [1 : n_r]$. Therefore, we can get a lower bound from applying (2.15) in Lemma 2.8.1 as follows

$$\begin{aligned}
\frac{1}{\binom{n_r}{k_r}} \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \det(\mathbf{F}_\Lambda) &= \frac{1}{\binom{n_r}{k_r}} \sum_{\substack{\Lambda \subseteq [1:n_r], \, j \in \Lambda \\ |\Lambda| = k_r}} \prod \lambda_j \\
&\geq \frac{1}{\binom{n_r}{k_r}} \sum_{\substack{[1:n_t] \subseteq \Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j \in \Lambda} \lambda_j \overset{(a)}{\geq} \frac{1}{\binom{n_r}{k_r}} \sum_{\substack{[1:n_t] \subseteq \Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j=1}^{n_t} \lambda_j \\
&\overset{(b)}{=} \frac{\binom{n_r - n_t}{k_r - n_t}}{\binom{n_r}{k_r}} \left(\prod_{j=1}^{n_t} \lambda_j \right) \overset{(c)}{=} \frac{\binom{n_r - n_t}{k_r - n_t}}{\binom{n_r}{k_r}} \left(\prod_{j=1}^{n_r} \lambda_j \right),
\end{aligned} \tag{2.17}$$

where: (a) follows since $\lambda_i \geq 1,\ \forall i \in [1:n_r]$; (b) follows since there are $\binom{n_r-n_t}{k_r-n_t}$ Λ sets such that, $|\Lambda| = k_r$ and $[1:n_t] \subseteq \Lambda \subseteq [1:n_r]$; (c) follows since $\lambda_i = 1, \forall i \in [n_t+1:n_r]$.

Similar to the conclusion in proof of Lemma 2.5.1(a) in Appendix 2.8.1.1, the average relation in (2.17) implies the lower bound in Lemma 2.5.1(b). Additionally, by the interlacing property of eigenvalues [Tho68], we know that $\mathsf{C}^{\star}_{n_t,k_r} \leq \mathsf{C}_{n_t,n_r}$. This concludes the proof of Lemma 2.5.1(b).

2.8.1.3 Proof of Lemma 2.5.1(c)

To prove Lemma 2.5.1(c), it suffices to prove the statement for the following two simple cases:

$$\text{1) For } k_t = n_t, k_r = n_r - 1, \mathsf{C}^{\star}_{n_t,n_r-1} \geq \frac{n_r-1}{n_r}\mathsf{C}_{n_t,n_r}, \tag{2.18}$$

$$\text{2) For } k_t = n_t - 1, k_r = n_r, \mathsf{C}^{\star}_{n_t-1,n_r} \geq \frac{n_t-1}{n_t}\mathsf{C}_{n_t,n_r}. \tag{2.19}$$

Using the two statements in (2.18) and (2.19), we can reduce an $n_t \times n_r$ system to a $k_t \times k_r$ system by systematically removing one receiver or one transmitter at a time from a system to get that

$$\begin{aligned}\mathsf{C}_{n_t,n_r} \leq& \frac{n_t}{n_t-1}\mathsf{C}^{\star}_{n_t-1,n_r} \leq \frac{n_t}{n_t-1}\frac{n_t-1}{n_t-2}\mathsf{C}^{\star}_{n_t-2,n_r} = \frac{n_t}{n_t-2}\mathsf{C}^{\star}_{n_t-2,n_r}\\ &\leq \cdots \leq \frac{n_t}{k_t}\mathsf{C}^{\star}_{k_t,n_r} \leq \frac{n_t}{k_t}\frac{n_r}{n_r-1}\mathsf{C}^{\star}_{k_t,n_r-1} \leq \frac{n_t}{k_t}\frac{n_r}{k_r}\mathsf{C}^{\star}_{k_t,k_r}.\end{aligned}$$

To conclude the proof of Lemma 2.5.1(c), we will now prove the statement in (2.18). The proof of (2.19) would directly follow by applying (2.18) on the reciprocal MIMO channel.

By substituting $k_r = n_r - 1$ in (2.15), we have that

$$\frac{1}{n_r}\sum_{\substack{\Lambda\subseteq[1:n_r],\\ |\Lambda|=n_r-1}} \det(\mathbf{F}_\Lambda) = \frac{1}{n_r}\sum_{\substack{\Lambda\subseteq[1:n_r],\\ |\Lambda|=n_r-1}}\prod_{j\in\Lambda}\lambda_j \overset{(a)}{\geq} \prod_{\substack{\Lambda\subseteq[1:n_r],\\ |\Lambda|=n_r-1}}\left(\prod_{j\in\Lambda}\lambda_j\right)^{\frac{1}{n_r}} = \left(\prod_{i=1}^{n_r}\lambda\right)^{\frac{n_r-1}{n_r}}, \tag{2.20}$$

where, (a) follows from the AM-GM inequality. Note that, the averaging in (2.20) implies that there exists a submatrix $\mathbf{F}_{\Lambda_s}$ given by subchannel matrix $\mathbf{H}_{\Lambda_s} \in \mathbf{C}^{n_t\times(n_r-1)}$ with a deter-

minant that is lower bounded by the RHS in (2.20). Taking the logarithm of $\det(\mathbf{F}_{\Lambda_s})$ and the RHS of (2.20) proves the statement in (2.18). This concludes the proof of Lemma 2.5.1(c).

2.8.2 Proof of Lemma 2.8.1

The proof of Lemma 2.8.1 is a consequence of the following property [Tho68] that relates characteristic polynomials of the principal submatrices of a Hermitian matrix to the characteristic polynomial of the full matrix.

Property 2.8.1. [Tho68] Let $\mathbf{A}$ be an $n_r \times n_r$ Hermitian matrix. For a subset $\Lambda \subseteq [1 : n_r]$, define $\mathbf{A}_\Lambda$ to be a principal submatrix of $\mathbf{A}$, constructed by only keeping the rows and columns of $\mathbf{A}$ indexed by Λ. Denote with $\rho(\lambda)$ and $\rho_\Lambda(\lambda)$ the characteristic polynomials of $\mathbf{A}$ and $\mathbf{A}_\Lambda$, respectively. Then the following property holds for all $k_r \in [1 : n_r]$

$$(n_r - k_r)! \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \rho_\Lambda(\lambda) = \rho^{(n_r - k_r)}(\lambda), \tag{2.21}$$

where $\rho^{(j)}(\lambda)$ is the j-th derivative of $\rho(\lambda)$ with respect to λ.

Let $[x^m]g(x)$ be the coefficient of x^m in the polynomial $g(x)$. Then by inspecting the coefficient of λ^0 in both sides of (2.21), we get that the following sequence of implications

$$\begin{aligned}
& [\lambda^0] \left((n_r - k_r)! \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \rho_\Lambda(\lambda) \right) = [\lambda^0] \rho^{(n_r - k_r)}(\lambda) \\
\implies \quad & (n_r - k_r)! \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} [\lambda^0] \rho_\Lambda(\lambda) = [\lambda^0] \rho^{(n_r - k_r)}(\lambda) \\
\implies \quad & (n_r - k_r)! \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} [\lambda^0] \rho_\Lambda(\lambda) = (n_r - k_r)! [\lambda^{n_r - k_r}] \rho(\lambda) \\
\overset{(a)}{\implies} \quad & \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} (-1)^{k_r} \left| [\lambda^0] \rho_\Lambda(\lambda) \right| = (-1)^{k_r} \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j \in \Lambda} \lambda_j
\end{aligned}$$

$$\implies \qquad \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} |\det(\mathbf{A}_\Lambda)| = \sum_{\substack{\Lambda \subseteq [1:n_r], \\ |\Lambda| = k_r}} \prod_{j \in \Lambda} \lambda_j, \tag{2.22}$$

where in (a) the RHS follows from the fact that $\rho(\lambda) = \prod_{i=1}^{n_r}(\lambda - \lambda_i)$, with $\{\lambda_i\}_{i=1}^{n_r}$ being the eigenvalues of $\mathbf{A}$. The statement in (2.22) is almost our goal statement in (2.15) if we remove the modulus in the LHS. To do this, we make use of the fact that, in Lemma 2.15, the matrix $\mathbf{A}$ is positive semidefinite, i.e., its minimum eigenvalue is non-negative. Thus, by the interlacing property of eigenvalues [Tho68], the minimum eigenvalue of each of the submatrices $\mathbf{A}_\Lambda$ (and by consequence its determinant) is also non-negative. This concludes the proof of Lemma 2.8.1.

CHAPTER 3

Network Simplification in Half-Duplex Networks

In this chapter, we shift our study of the network simplification problem to half-duplex networks. In particular, we focus on the half-duplex diamond networks with N relays and develop efficient approaches to select a subset of k relays within the network. Additionally, we provide fundamental bounds on the worst-case fraction guarantee when selecting $k = N - 1$ relays out of N relays.

3.1 Introduction

In this chapter, we extend the study our study on network simplification in full-duplex network (Chapter 2) to Gaussian Half-Duplex (HD)diamond networks with N relays. Our study in HD networks is motivated by the fact that currently employed relays operate in HD, unless sufficient isolation between the antennas can be guaranteed or different bands are used for transmission and reception. Additionally, as recently announced in 3GPP Rel-13, HD is also expected to be employed in next generation Internet of Things networks to enable low-cost communication modules for short-distance and infrequent data transmissions.

Studying the network simplification problem is more challenging when networks operate in HD compared to full-duplex. This is due to the intrinsic combinatorial nature of approximate capacity characterization in HD relay networks.

Similar to Gaussian full-duplex networks, the capacity characterization of the additive white Gaussian noise relay network is a long-standing open problem. The tightest upper

bound on the capacity is the well-known cut-set bound [CE79]. A number of schemes have been proposed [LKE11], [ADT11], [OD13], [LKK14] that achieve a rate approximating the Shannon capacity of the Gaussian HD network within a constant gap. The tightest refinement of the approximate capacity constant gap for Gaussian HD networks is $1.96(N+2)$ bits/sec [CTK14][1].

In general, the evaluation of the approximate capacity of Gaussian HD relay networks is more challenging than the full-duplex counterpart since it requires an optimization over the 2^N listen/transmit states. We refer to the states that suffice to characterize the approximate capacity by *active* states. In [CTK15] the authors proved a surprising result, which was first conjectured in [BOF12]: for a class of HD relay networks, which includes Gaussian networks, the simplest optimal schedule, i.e., the schedule that uses the least number of active states, operates with at most $N+1$ states out of 2^N. This result generalizes those in [BMK14], [BF14b] and [BFO16], valid only for Gaussian HD diamond networks with limited network sizes. However, even though we understand that a schedule with at most $N+1$ active states exists, it is not yet known if we can find these states efficiently for relay networks with arbitrary topology. A similar thread of research [BSF14] has focused on deriving approximate capacity guarantees when each relay operates with its optimal schedule - computed as if the other relays were not there - and is allowed to switch multiple times between listen and transmit modes of operation. For approximate capacity evaluation, the authors in [EPS14] proposed an approach that, for certain network topologies – such as the line network and a specific class of layered networks – outputs the approximate capacity in polynomial time in the number of relays. This result is quite promising, but it relies on the simplified topology of certain classes of relay networks.

Different from the aforementioned lines of research, where the main objective is to provide

[1]The constant gap in [CTK14] was derived by using the approach first proposed in [Kra04]. The work in [Kra04] showed that HD relay networks can be studied within the framework of their full-duplex counterparts, by expressing the channel inputs and outputs as functions of the states of the relays. In particular, it was observed that information can be conveyed by randomly switching the relay between transmit and receive modes. However, this only improves the capacity by a constant, at most 1 bit per relay.

a low-complexity characterization of the network approximate capacity, in this chapter, we seek to understand what fraction of the network approximate capacity can be guaranteed when only a subset of $k \in [1 : N]$ relays is operated. From the result in [NOF14], it directly follows that in Gaussian HD diamond networks, by selecting k relays, one can always retain at least a fraction $\frac{k}{2(k+1)}$ of the approximate HD capacity of the whole network. This is accomplished by operating the k relays (selected as in full-duplex) in only 2 states of equal duration: in the first phase, all the k relays listen and in the second phase, all the k relays transmit. Although providing a performance guarantee, this result might be too conservative. This is indeed confirmed by the result in [BF14a] where it was proved that, in any Gaussian HD diamond network, there always exists a subnetwork of $k = 2$ relays that retains at least half of the approximate capacity of the full network. The selected two relays are restricted to operate in complementary fashion, i.e., when one relay transmits, the other listens and vice versa. In this chapter, we do not restrict the selected k relays to operate only in certain states as in [BF14a], which leads to better performance guarantees in terms of retained fraction of the approximate capacity.

Contributions. In this chapter we seek to understand how much of the approximate HD capacity one can retain by smartly selecting a subset of k relays in a Gaussian HD diamond network with a fixed number of relays N. In particular, our goal is to provide a worst-case performance guarantee in terms of retained fraction that holds universally, i.e., independently of the values of the channel parameters. Our main contributions can be summarized as follows:

1. We first derive properties of Gaussian diamond networks and submodular functions, which provide a combinatorial handle on the network simplification problem in Gaussian HD diamond networks. For instance, we prove a result that we refer to as the *partition lemma*, which states that if we partition the network into multiple subnetworks such that each relay belongs to only one of such subnetworks, then the approximate capacity of the full network is upper bounded by the sum of the approximate capacities

of the subnetworks. Beyond their utilization in the proofs of our main results, these properties might be of independent interest.

2. We analyze a straightforward algorithm to select a subnetwork of $k = N - 1$ relays, which operates all the relays except the *worst* one. We say that, among the N relays, the i-th relay is the *worst* if it has the smallest single-relay approximate capacity, i.e., if the maximum HD flow that can be routed through it is less than or equal to the other $N - 1$ flows through each of the remaining $N - 1$ relays. We prove that the algorithm outputs, in $O(N)$ time, a subnetwork whose approximate HD capacity is at least half of the approximate HD capacity of the whole network. We also show that this fraction guarantee is tight if we know only the single-relay approximate capacities, i.e., there exists a class of Gaussian HD diamond networks with N relays where, by removing the worst relay, the remaining $(N-1)$-relay subnetwork has an approximate capacity that is half of the approximate capacity of the full network. This guarantee might be too conservative and indeed a smarter choice leads to a better performance, as described in the next point. However, an appealing feature of this algorithm is that it only requires the knowledge of the N single-relay approximate capacities.

3. We prove that, in any N-relay Gaussian HD diamond network, there always exists a subnetwork of $k = N-1$ relays that retains at least a fraction $\frac{N-1}{N}$ of the approximate capacity of the full network. We also show that this fraction of $\frac{N-1}{N}$ is tight. This result significantly improves over the fraction of half guaranteed by the algorithm described in the previous point. Moreover, this guarantee is fundamental, i.e., it is the largest fraction that can be ensured when $N - 1$ relays are selected. In addition, we show a surprising result: any optimal schedule for the approximate capacity of the full network provides a deterministic construction for a schedule for each of the $(N-1)$-relay subnetworks, such that at least one of these subnetworks retains a rate that is greater than or equal to the worst performance guarantee of $\frac{N-1}{N}$ of the approximate capacity of the full network. This leads to a complexity reduction in the scheduling

problem; in fact, it implies that, in order to select an $(N-1)$-relay subnetwork that retains a fraction $\frac{N-1}{N}$ of the approximate capacity of the full network, there is no need to compute the optimal schedule for each of the N subnetworks. It suffices to compute an optimal schedule of the full network.

4. We generalize the results described in the previous two points to generic values of $k \in [1:N]$. In particular, we show that: (i) the straightforward algorithm that removes the $N-k$ worst relays and runs in $O(N\log(N))$, ensures that the selected k-relay subnetwork has an approximate capacity that is at least $2^{-(N-k)}$ of the approximate capacity of the original network with N relays; (ii) a fraction $\frac{k}{N}$ of the approximate capacity of the full network can always be retained by selecting k relays and operating them with an optimal schedule of the full network. However, this last worst-case fraction guarantee does not appear to be tight. This result suggests that, when $k < N-1$, forcing the k-relay subnetworks to operate with the optimal schedule of the full network may be suboptimal.

5. We find significant differences between the wireless simplification problem for HD and full-duplex networks. For instance: (i) in HD, when $k \in \{1,2\}$ relays are selected, the fraction of the retained approximate capacity depends on N and decreases as N increases; (ii) the worst-case networks in HD and full-duplex are not necessarily the same; (iii) the best k-relay subnetworks in HD and full-duplex might be different. These results show that full-duplex and HD relay networks have a different nature. This might be due to the fact that in HD the schedule plays a crucial role and hence removing some of the relays can change the schedule at which the selected subnetwork should be optimally operated.

Chapter Organization. Section 3.2 describes the N-relay Gaussian HD diamond network and summarizes known capacity results. Section 3.3 derives properties of submodular functions and diamond networks. Section 3.4 studies the performance (in terms of retained

fraction) of a simple algorithm that selects $k \in [1:N]$ relays out of the N possible ones, by removing the worst $N-k$ relays. In particular, Section 3.4 first considers the case $k = N-1$ and then generalizes the result to any $k \in [1:N]$. Section 3.5 provides a fundamental guarantee (in terms of retained fraction) when $N-1$ relays are selected out of the N possible ones. Section 3.5 also generalizes the lower bound on the fraction guarantee for $k = N-1$ to general $k \in [1:N]$. Finally, Section 3.6 discusses some implications of the presented results, highlights differences between the selection performances in HD and full-duplex networks and concludes the chapter. Some of the proofs can be found in the Appendix.

Notation. In the rest of the chapter, we use the following notation convention. We denote with $[a:b]$ the set of integers from a to $b \geq a$. Y^j is a vector of length j with components $(Y_1, \ldots, Y_j)$, $|z|$ is the component-wise absolute value of the vector z and z^T is the transpose of the vector z. For two sets $\mathcal{A}_1, \mathcal{A}_2$, $\mathcal{A}_1 \subseteq \mathcal{A}_2$ indicates that $\mathcal{A}_1$ is a subset of $\mathcal{A}_2$, $\mathcal{A}_1 \cup \mathcal{A}_2$ represents the union of $\mathcal{A}_1$ and $\mathcal{A}_2$, $\mathcal{A}_1 \cap \mathcal{A}_2$ represents the intersection of $\mathcal{A}_1$ and $\mathcal{A}_2$ and $\mathcal{A}_1 \backslash \mathcal{A}_2$ is the set of elements that belong to $\mathcal{A}_1$ but not to $\mathcal{A}_2$. With $|\mathcal{A}|$ we indicate the cardinality of $\mathcal{A}$, $\emptyset$ is the empty set and $\mathbb{E}[\cdot]$ indicates the expected value. For all $x \in \mathbb{R}$, the ceiling and floor functions are denoted by $\lceil x \rceil$ and $\lfloor x \rfloor$, respectively. The L_1-norm of a vector λ is represented by $\|\lambda\|_1$. Table 3.1 summarizes and defines quantities that are frequently used throughout the chapter.

3.2 System Model and Known Results

We consider the Gaussian HD diamond network $\mathcal{N}_{[1:N]}$ in Fig. 3.1 where a source node (node 0) wishes to communicate with a destination (node $N+1$) through N non-interfering relays operating in HD. The channel gains are assumed to be constant for the whole transmission duration and hence known to all nodes. Let $X_{i,t} \in \mathbb{C}$ denote the signal transmitted by node i, $\forall i \in [0:N]$, at time instant t. Similarly, let $Y_{i,t} \in \mathbb{C}$ denote the observation received by node i, $\forall i \in [1:N+1]$, at time instant t. For each channel use, the source has

Table 3.1: Quantities of interest used throughout Chapter 3.

Quantity	**Definition**
$\mathcal{N}_{\mathcal{K}}$	Network which contains only the relays in $\mathcal{K} \subseteq [1:N]$
$\bar{\mathcal{N}}_i$	$\mathcal{N}_{[1:N]\setminus\{i\}}$
$\mathsf{C}_{\mathcal{N}_{\mathcal{K}}}$	Half-duplex Shannon capacity of $\mathcal{N}_{\mathcal{K}}$
$\overline{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}$	Cut-set bound for $\mathcal{N}_{\mathcal{K}}$ in HD
$\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}$	Approximate HD capacity of $\mathcal{N}_{\mathcal{K}}$
$\mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}}$	Half-duplex approximate rate of $\mathcal{N}_{\mathcal{K}}$ when operated with the schedule λ
$\mathsf{C}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{K}}}$	Full-duplex capacity of $\mathcal{N}_{\mathcal{K}}$
$\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{K}}}$	approximate full-duplex capacity of $\mathcal{N}_{\mathcal{K}}$

a message W uniformly distributed over $\left[1 : 2^{KR}\right]$ for the destination, where $K \in \mathbb{N}$ denotes the codeword length and $R \in \mathbb{R}_+$ is the transmission rate in bits per channel use. At time $t \in [1:K]$: (i) the source maps the message W into a channel input through a codebook mapping $X_{0,t}(W) : [1 : 2^{RK}] \to \mathbb{C}$; (ii) the i-th relay, with $i \in [1:N]$, if in transmission mode of operation, maps its past channel observations $Y_i^{t-1} \in \mathbb{C}^{t-1}$ into a channel input symbol $X_{i,t}\left(Y_i^{t-1}\right) \in \mathbb{C}$, where $Y_i^{t-1} = (Y_{i,1}, Y_{i,2}, \cdots, Y_{i,t-1})$. At time K, the destination outputs an estimate $\widehat{W}$ of the message based on all its channel observations Y_{N+1}^K. A rate R is said to be ϵ-achievable if there exists a sequence of codes indexed by the block length K such that $\mathbb{P}\left[W \neq \widehat{W}\right] \leq \epsilon$ for any $\epsilon > 0$. The capacity is the largest nonnegative rate that is ϵ-achievable for $\epsilon \in (0,1)$.

The single-antenna static Gaussian HD diamond network $\mathcal{N}_{[1:N]}$, shown in Fig. 3.1, is defined by the input/output relationship

$$Y_{i,t} = (1 - S_{i,t})\left[h_{si}X_{0,t} + Z_{i,t}\right], \quad \forall i \in [1:N], \tag{3.1a}$$

$$Y_{N+1,t} = \sum_{i=1}^{N} S_{i,t}h_{id}X_{i,t} + Z_{N+1,t}, \tag{3.1b}$$

where: (i) $S_{i,t}$ is the binary random variable that represents the state of the i-th relay at

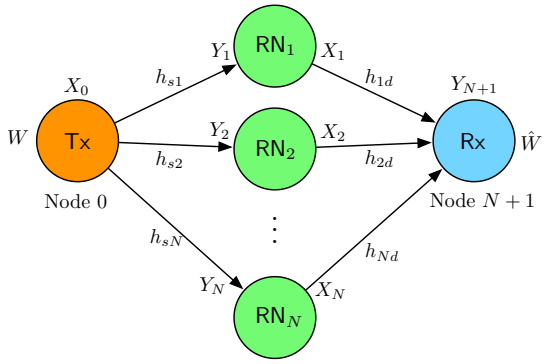


Figure 3.1: Gaussian diamond network with N relays.

time t, i.e., when $S_{i,t} = 0$ the i-th relay is receiving while when $S_{i,t} = 1$ the i-th relay is transmitting; (ii) the channel inputs are subject to a unitary average power constraint, i.e., $\mathbb{E}\left[|X_{k,t}|^2\right] \leq 1,\ \forall k \in [0:N]$; (iii) $(h_{si}, h_{id}) \in \mathbb{C}^2$ represent the channel coefficients from the source to the i-th relay and from the i-th relay to the destination, respectively. Without loss of generality, SNR is assumed to be incorporated in the channel coefficients; (iv) $Z_{i,t}$, $i \in [1:N+1]$, indicates the additive white Gaussian noise at the i-th node; noises are assumed to be independent and identically distributed as $\mathcal{CN}(0,1)$. In the remainder of the chapter, we drop the dependence on the time t since the channel is discrete memoryless. We denote with ℓ_i and r_i the individual link capacities, namely

$$\ell_i := \log\left(1 + |h_{si}|^2\right), \quad \forall i \in [1:N], \tag{3.2a}$$

$$r_i := \log\left(1 + |h_{id}|^2\right), \quad \forall i \in [1:N]. \tag{3.2b}$$

The capacity of the Gaussian HD relay network is not known in general. The best known upper-bound is the cut-set bound [CE79], formulated as

$$\overline{\mathsf{C}}_{\mathcal{N}_{[1:N]}} = \max_{\mathbb{P}_{\{X_i,S_i\}}(\cdot)} \min_{\mathcal{A}\subseteq[1:N]} I(X_{\mathcal{A}}, S_{\mathcal{A}}; Y_{\mathcal{A}^c}|X_{\mathcal{A}^c}, S_{\mathcal{A}^c}), \tag{3.3}$$

where $\mathbb{P}_{\{X_i,S_i\}}(\cdot)$ is the joint distribution on all channel inputs X_i and scheduling states S_i in the network. The distribution that maximizes the cut-set expression in (3.3) is not known in general. For the particular case of a two-hop HD line network, where a source communicates

to a destination by hopping information through one relay, it has been shown in [ZJS15] that the optimal input distribution is a mixture between continuous and discrete distributions.

The notion of approximate capacity in Gaussian relay networks provides an expression that is guaranteed to be at most an additive constant gap away from the Shannon capacity. In particular, let $\mathsf{C}_{\mathcal{N}_{[1:N]}}(\mathbf{h})$ be the Shannon capacity of the network $\mathcal{N}_{[1:N]}$ with channel coefficients $\mathbf{h} = (h_{s1}, h_{d1}, h_{s2}, h_{d2}, \cdots, h_{sN}, h_{Nd})$. Formally, we say that an expression $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}(\mathbf{h})$ is a constant gap approximation of the Shannon capacity $\mathsf{C}_{\mathcal{N}_{[1:N]}}(\mathbf{h})$, if there exists a value (additive gap) a_N, that does not depend on the channel coefficients - and by our assumption on the SNR - such that

$$\left|\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}(\mathbf{h}) - \mathsf{C}_{\mathcal{N}_{[1:N]}}(\mathbf{h})\right| \leq a_N. \tag{3.4}$$

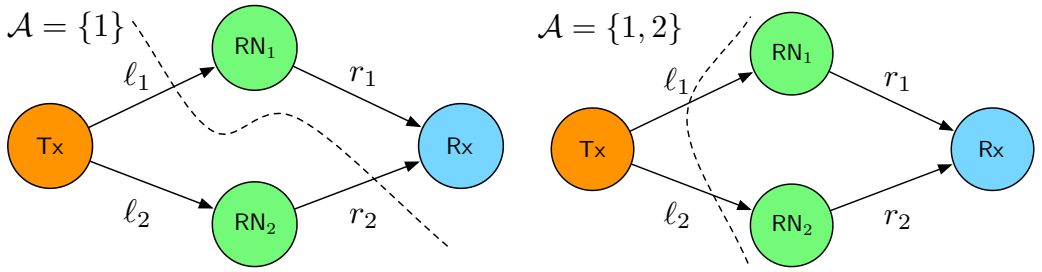
The best known gap is $a_N = 1.96(N+2)$ by [CTK14]. For brevity, in the remainder of the chapter, we drop the channel coefficients arguments to the values $\mathsf{C}_{\mathcal{N}_{[1:N]}}$, $\overline{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$ and $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$.

In this chapter, we focus on an approximation of $\mathsf{C}_{[1:N]}$ proved in [LKE11], [ADT11], [OD13], [LKK14] which uses fixed schedules, as well as, independent and identically distributed Gaussians with zero mean and unit variance as channel inputs $\{X_i\}_{i=0}^{N}$. This approximate capacity, which we term the approximate capacity, for Gaussian HD diamond networks is defined as

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} = \max_{\lambda \in \Lambda} \min_{\mathcal{A} \subseteq [1:N]} \sum_{s \in [0:1]^N} \lambda_s \left(\max_{i \in \mathcal{L}_s \cap \mathcal{A}} \ell_i + \max_{i \in \mathcal{T}_s \cap \mathcal{A}^c} r_i \right), \tag{3.5}$$

where: (i) a state s is expanded as $s = (s_1, s_2, s_3, ..., s_N)$, with $S_i = s_i$ representing the binary state of node i. Therefore, s in (3.5) ranges over all possible 2^N states of the network; (ii) λ denotes a schedule of the network which defines, for all $s \in [0:1]^N$, a fraction of time λ_s for which the network operates with the state s; (iii) $\Lambda = \{\lambda : \lambda \in \mathbb{R}_+^{2^N},\ \|\lambda\|_1 = 1\}$ is the set of all possible listen/transmit schedules λ; (iv) $\mathcal{L}_s$ (respectively, $\mathcal{T}_s$) represents the set of indices of relays listening (respectively, transmitting) in the state $s \in [0:1]^N$. Note that among the relays 'on the side of the destination', indexed by $\mathcal{A}$, only those in receive mode matter. Similarly, among the relays 'on the side of the source' (in (3.5) indexed by $\mathcal{A}^c = [1:N] \backslash \mathcal{A}$)

only those in transmit mode matter. For illustration, Fig. 3.2 depicts two different cuts in a 2-relay network and shows the 'source-side' and 'destination-side' in both cases.



(a) (b)

Figure 3.2: Two cuts in a 2-relay diamond network.

For the particular case of $N = 1$, the approximate capacity in (3.5) becomes

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{1\}}} = \frac{\ell_1 r_1}{\ell_1 + r_1} \tag{3.6}$$

and when $N = 2$ the authors in [BMK14] derived $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:2]}}$ in (3.5) in closed form.

Remark 3.2.1. The additive constant gap $a_N = O(N)$ between the Shannon capacity and the approximate capacity in (3.5) is universal and independent of the channel coefficients $\mathbf{h}$ and hence of the link capacities ℓ_i, r_i. Therefore, it is possible to construct networks where the approximate capacity grows exponentially with N, while the gap remains linear in N. For example, consider an N-relay diamond network with all point-to-point link capacities equal to b. For this particular network, the approximate capacity – obtained by applying (3.5) – is b. If we let $b \to \infty$ (by increasing the SNR or by having larger channel coefficients), then the approximate capacity grows unbounded while the gap is still $O(N)$. This conclusion is true for any approximate capacity (not necessarily the used in this chapter), since from (3.4) it is not difficult to see that any two constant approximations of $\mathsf{C}_{\mathcal{N}_{[1:N]}}$ are at most a constant gap away from each other.

In what follows we say that the subnetwork $\mathcal{N}_{\mathcal{K}}$ with $\mathcal{K} \subseteq [1 : N]$ operates with a 'natural' schedule derived from the schedule λ of $\mathcal{N}_{[1:N]}$ if the schedule of $\mathcal{N}_{\mathcal{K}}$ is constructed directly from λ, as better explained through the following example.

Example. Consider a Gaussian HD diamond network $\mathcal{N}_{[1:N]}$ with $N = 3$. Let

$$\lambda = [\lambda_{000}\ ,\ \lambda_{001}\ ,\ \lambda_{010}\ ,\ \lambda_{011}\ ,\ \lambda_{100}\ ,\ \lambda_{101}\ ,\ \lambda_{110}\ ,\ \lambda_{111}]^T$$

be a schedule for $\mathcal{N}_{[1:3]}$. Denote with $\lambda^{(\mathcal{N}_{\{2,3\}})}$ (respectively, $\lambda^{(\mathcal{N}_{\{2\}})}$) the schedule that is derived naturally from λ for the subnetwork $\mathcal{N}_{\{2,3\}}$ (respectively, $\mathcal{N}_{\{2\}}$). With this, we have

$$\lambda^{(\mathcal{N}_{\{2,3\}})} = [\lambda_{000} + \lambda_{100}\ ,\ \lambda_{001} + \lambda_{101}\ ,\ \lambda_{010} + \lambda_{110}\ ,\ \lambda_{011} + \lambda_{111}]^T$$

and similarly we get

$$\lambda^{(\mathcal{N}_{\{2\}})} = [\lambda_{000} + \lambda_{001} + \lambda_{100} + \lambda_{101}\ ,\ \ \lambda_{010} + \lambda_{011} + \lambda_{110} + \lambda_{111}]^T.$$

Thus, from the expression in (3.5), the approximate rate $\mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}}$ of a subnetwork (for example $\mathcal{N}_{\{2,3\}}$) when operating with the 'natural' schedule derived from λ is

$$\begin{aligned}
\mathsf{R}^{\lambda}_{\mathcal{N}_{\{2,3\}}} &= \min_{\mathcal{A}\subseteq\{2,3\}} \sum_{s\in[0:1]^2} \lambda_s^{(\mathcal{N}_{\{2,3\}})} \left(\max_{i\in\mathcal{A}} \ell'_{i,s} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,s} \right) \\
&= \min_{\mathcal{A}\subseteq\{2,3\}} \Bigg[\lambda_{00}^{(\mathcal{N}_{\{2,3\}})} \left(\max_{i\in\mathcal{A}} \ell'_{i,00} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,00} \right) + \lambda_{01}^{(\mathcal{N}_{\{2,3\}})} \left(\max_{i\in\mathcal{A}} \ell'_{i,01} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,01} \right) \\
&\qquad + \lambda_{10}^{(\mathcal{N}_{\{2,3\}})} \left(\max_{i\in\mathcal{A}} \ell'_{i,10} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,10} \right) + \lambda_{11}^{(\mathcal{N}_{\{2,3\}})} \left(\max_{i\in\mathcal{A}} \ell'_{i,11} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,11} \right) \Bigg] \\
&\overset{(a)}{=} \min_{\mathcal{A}\subseteq\{2,3\}} \Bigg[(\lambda_{000}+\lambda_{100}) \left(\max_{i\in\mathcal{A}} \ell'_{i,00} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,00} \right) \\
&\qquad + (\lambda_{001}+\lambda_{101}) \left(\max_{i\in\mathcal{A}} \ell'_{i,01} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,01} \right) \\
&\qquad + (\lambda_{010}+\lambda_{110}) \left(\max_{i\in\mathcal{A}} \ell'_{i,10} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,10} \right) \\
&\qquad + (\lambda_{011}+\lambda_{111}) \left(\max_{i\in\mathcal{A}} \ell'_{i,11} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,11} \right) \Bigg] \\
&= \min_{\mathcal{A}\subseteq\{2,3\}} \sum_{s\in[0:1]^3} \lambda_s \left(\max_{i\in\mathcal{A}} \ell'_{i,s} + \max_{i\in\{2,3\}\setminus\mathcal{A}} r'_{i,s} \right), \qquad (3.7)
\end{aligned}$$

where

$$\ell'_{i,s} = \begin{cases} \ell_i & \text{if } i \in \mathcal{L}_s \\ 0 & \text{otherwise} \end{cases}, \qquad r'_{i,s} = \begin{cases} r_i & \text{if } i \in \mathcal{T}_s \\ 0 & \text{otherwise} \end{cases}, \qquad (3.8)$$

and where the equality in (a) follows by using the construction of $\lambda^{(\mathcal{N}_{\{2,3\}})}$.

3.3 Diamond Networks and Submodularity Properties

In this section we derive and discuss some properties of diamond networks and submodular functions, which represent the main ingredient in the proof of our main results. It is worth noting that, beyond their utilization in the proofs, these properties might be of independent interest.

3.3.1 A Partition Lemma for Diamond Networks

The first result that we derive provides an upper bound on the approximate HD rate that can be achieved by the full network. This upper bound is stated in the following lemma – which we refer to as the *partition lemma* – whose proof can be found in Appendix 3.7.1.

Lemma 3.3.1 (Partition lemma). Let λ be a schedule for the N-relay Gaussian HD diamond network $\mathcal{N}_{[1:N]}$. Then, for any $\mathcal{K} \subseteq [1:N]$, we have

$$\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}} \leq \mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}} + \mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]\setminus\mathcal{K}}}, \tag{3.9}$$

where for a subnetwork $\mathcal{N}_{\mathcal{K}}$ of $\mathcal{N}_{[1:N]}$, $\mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}}$ represents the HD approximate rate when operated with the deterministic schedule constructed from λ.

The partition lemma result has the following two consequences:

1. Let $\lambda^{\star}$ be an optimal schedule for the approximate capacity of the full network $\mathcal{N}_{[1:N]}$, i.e., $\mathsf{R}^{\lambda^{\star}}_{\mathcal{N}_{[1:N]}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. Since the 'natural' schedule constructed from $\lambda^{\star}$ might not be the optimal one for the approximate capacity of the subnetworks $\mathcal{N}_{\mathcal{K}}$ and $\mathcal{N}_{[1:N]\setminus\mathcal{K}}$, then $\mathsf{R}^{\lambda^{\star}}_{\mathcal{N}_{\mathcal{K}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}$ and similarly $\mathsf{R}^{\lambda^{\star}}_{\mathcal{N}_{[1:N]\setminus\mathcal{K}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]\setminus\mathcal{K}}}$. Hence, the partition lemma straightforwardly implies that

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]\setminus\mathcal{K}}}. \tag{3.10}$$

   By applying the partition lemma, we obtain different bounds on the approximate capacity of the network. For example, consider $\mathcal{N}_{[1:N]}$ with $N = 3$, then the result

implied by the partition lemma gives the following four bounds on the approximate capacity

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:3]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{1,2\}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{3\}}}, \quad \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:3]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{1,3\}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{2\}}},$$
$$\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:3]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{2,3\}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{1\}}}, \quad \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:3]}} \leq \sum_{i=1}^{3} \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}.$$

2. The partition lemma relates to the following question studied in [HEJ10, JEH11]: Can removing a single edge of capacity δ, reduce the capacity region of a network along each dimension, by more than δ? The answer to this question is known only in a few specific cases. The partition lemma implies a negative answer to this question for the approximate capacity of Gaussian HD diamond networks. In particular, without loss of generality, let $\delta = \ell_i$, for some $i \in [1:N]$. Then, from (3.10), we have

$$\begin{aligned}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq \widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} \implies \widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i} &\geq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} - \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} \\ &\overset{\text{(a)}}{\geq} \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} - \min\{\delta, r_i\} \geq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} - \delta,\end{aligned}$$

where $\bar{\mathcal{N}}_i = \mathcal{N}_{[1:N]\setminus\{i\}}$ and the inequality in (a) follows since $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} \leq \mathsf{C}^{\text{FD}}_{\mathcal{N}_{\{i\}}} = \min\{\delta, r_i\}$. Here, $\mathsf{C}^{\text{FD}}_{\mathcal{N}_{\{i\}}}$ denotes the full-duplex capacity of the relay subnetwork that uses only relay i.

3.3.2 Submodular Functions and Cut Properties

We now derive a property of submodular functions, which we next leverage to prove a property on cuts in diamond networks.

Definition 3.3.1. For a finite set Ω, let $f : 2^{\Omega} \to \mathbb{R}$ be a set function defined on Ω. The set function f is submodular if

$$\forall \mathcal{A}, \mathcal{B} \subseteq \Omega, \quad f(\mathcal{A}) + f(\mathcal{B}) \geq f(\mathcal{A} \cup \mathcal{B}) + f(\mathcal{A} \cap \mathcal{B}). \tag{3.11}$$

Building on the definition in (3.11), we now prove a property for a general submodular function.

Lemma 3.3.2. Let f be a submodular set function defined on Ω. Then, for any group of n sets $\mathcal{A}_i \subseteq \Omega$, $i \in [1:n]$,

$$\sum_{i=1}^{n} f\left(\mathcal{A}_i\right) \geq \sum_{j=1}^{n} f\left(\mathcal{E}_j^{(n)}\right),$$

where $\mathcal{E}_j^{(n)}$ is the set of elements that appear in at least j sets $\mathcal{A}_i, i \in [1:n]$.

Proof. The proof relies on the definition of submodular functions and on some set-theoretic properties. The detailed proof can be found in Appendix 3.7.2. □

To better understand what Lemma 3.3.2 implies, consider the following example.

Example. Let $\Omega = [1:7]$ and consider the subsets $\mathcal{A}_1 = \{1,2,5,7\},\ \mathcal{A}_2 = \{4,5\},\ \mathcal{A}_3 = \{2,4,5,6\}$. Lemma 3.3.2 proves that, for a submodular set function f defined over Ω, we get

$$f(\overbrace{\{1,2,5,7\}}^{\mathcal{A}_1}) + f(\overbrace{\{4,5\}}^{\mathcal{A}_2}) + f(\overbrace{\{2,4,5,6\}}^{\mathcal{A}_3}) \geq f(\underbrace{\{1,2,4,5,6,7\}}_{\mathcal{E}_1^{(3)}}) + f(\underbrace{\{2,4,5\}}_{\mathcal{E}_2^{(3)}}) + f(\underbrace{\{5\}}_{\mathcal{E}_3^{(3)}}). \tag{3.12}$$

Now, as an example, consider $f(\mathcal{A}) = \max_{i \in \mathcal{A}} \{i\}$ for $\mathcal{A} \subseteq \Omega$, which is a submodular set function. By evaluating both sides of (3.12) for our example function, we get

$$\sum_{i=1}^{3} f(\mathcal{A}_i) = 7 + 5 + 6 = 18, \quad \sum_{i=1}^{3} f\left(\mathcal{E}_j^{(3)}\right) = 7 + 5 + 5 = 17$$
$$\implies \sum_{i=1}^{3} f(\mathcal{A}_i) \geq \sum_{i=1}^{3} f\left(\mathcal{E}_j^{(3)}\right) .$$

Next, we use the result on submodular functions in Lemma 3.3.2 to prove the following result for Gaussian diamond networks.

Lemma 3.3.3. Consider an N-relay Gaussian diamond network $\mathcal{N}_{[1:N]}$. Then, for any collection of sets $\mathcal{A}_i \subseteq [1:N] \backslash \{i\}$, there exists a collection of $(N-1)$ sets $\mathcal{A}_{\mathrm{F}_j} \subseteq [1:N]$, with $j \in [1:N-1]$ such that

$$\sum_{j=1}^{N} \left(\max_{i \in \mathcal{A}_j} \ell_i + \max_{i \in ([1:N] \backslash \{j\}) \backslash \mathcal{A}_j} r_i \right) \geq \sum_{j=1}^{N-1} \left(\max_{i \in \mathcal{A}_{\mathrm{F}_j}} \ell_i + \max_{i \in [1:N] \backslash \mathcal{A}_{\mathrm{F}_j}} r_i \right). \tag{3.13}$$

Particularly, $\mathcal{A}_{\mathrm{F}_j}$ is the set of elements that appear in at least j sets $\mathcal{A}_i$, $i \in [1:N]$ and therefore, it does not depend on the values of (ℓ_i, r_i).

Proof. The proof, which is based on the result in Lemma 3.3.2 and on simple counting arguments, can be found in Appendix 3.7.3. □

We next provide a simple example that better explains the implication of Lemma 3.3.3.

Example. Consider a 3-relay Gaussian diamond network $\mathcal{N}_{[1:3]}$. With this, we have $\bar{\mathcal{N}}_1 = \mathcal{N}_{\{2,3\}}$, $\bar{\mathcal{N}}_2 = \mathcal{N}_{\{1,3\}}$ and $\bar{\mathcal{N}}_3 = \mathcal{N}_{\{1,2\}}$. Now for the subnetwork $\bar{\mathcal{N}}_i$ consider the following possible cut $\mathcal{A}_i$: (i) $\mathcal{A}_1 = \emptyset$ (i.e., in $\bar{\mathcal{N}}_1$ relays 2 and 3 are 'on the side of the source'); (ii) $\mathcal{A}_2 = \{3\}$ (i.e., in $\bar{\mathcal{N}}_2$ relay 1 is 'on the side of the source' and relay 3 is 'on the side of the destination'); (iii) (i) $\mathcal{A}_3 = \{1, 2\}$ (i.e., in $\bar{\mathcal{N}}_3$ relays 1 and 2 are 'on the side of the destination'). With this, by evaluating the left-hand side of (3.13), we obtain

$$\begin{aligned}\sum_{j=1}^{3} \left(\max_{i \in \mathcal{A}_j} \ell_i + \max_{i \in ([1:3] \setminus \{j\}) \setminus \mathcal{A}_j} r_i \right) &= \max_{i \in \{2,3\}} r_i + \ell_3 + r_1 + \max_{i \in \{1,2\}} \ell_i \\ &\geq \max_{i \in [1:3]} \ell_i + \max_{i \in [1:3]} r_i \\ &= \sum_{j=1}^{2} \left(\max_{i \in \mathcal{A}_{\mathrm{F}_j}} \ell_i + \max_{i \in [1:3] \setminus \mathcal{A}_{\mathrm{F}_j}} r_i \right),\end{aligned}$$

where we let $\mathcal{A}_{\mathrm{F}_1} = [1:3]$ and $\mathcal{A}_{\mathrm{F}_2} = \emptyset$. In this example, we considered a specific choice of $\mathcal{A}_i, i \in [1:3]$ in $\bar{\mathcal{N}}_i$. By repeating the same reasoning, it is possible to show that, for any of the $2^{N(N-1)} = 4^3$ possible combinations of cuts $\mathcal{A}_i$, there always exist two cuts $\mathcal{A}_{\mathrm{F}_j}, j \in [1:2]$ in the full network $\mathcal{N}_{[1:3]}$ – whose construction is provided in Lemma 3.3.3 – such that (3.13) holds.

Before concluding this section and going into the technical details of how to use these results to prove our main results, we state a couple of remarks.

Remark 3.3.1. By considering the specific values of the link capacities (ℓ_i, r_i) in a given network, we could prove the inequality in Lemma 3.3.3 with a different construction than the

one provided in the lemma. For illustration, consider the following example with $N = 3$. Let $(\ell_1, \ell_2, \ell_3) = (3, 4, 11)$ and $(r_1, r_2, r_3) = (6, 8, 6)$. Additionally, let $\mathcal{A}_1 = \{2\}$, $\mathcal{A}_2 = \{1\}$ and $\mathcal{A}_3 = \{1, 2\}$. For the given network, it is not difficult to verify that for $\mathcal{A}'_{\mathrm{F}_1} = \mathcal{A}'_{\mathrm{F}_2} = \{1, 2, 3\}$, we have

$$\sum_{j=1}^{3} \left(\max_{i \in \mathcal{A}_j} \ell_i + \max_{i \in [1:3] \setminus \{j\} \setminus \mathcal{A}_j} r_i \right) = (4 + 6) + (3 + 6) + 4 = 23$$

$$\geq \left(\max_{i \in \mathcal{A}'_{\mathrm{F}_1}} \ell_i + \max_{i \in [1:3] \setminus \mathcal{A}'_{\mathrm{F}_1}} r_i \right) + \left(\max_{i \in \mathcal{A}'_{\mathrm{F}_2}} \ell_i + \max_{i \in [1:3] \setminus \mathcal{A}'_{\mathrm{F}_2}} r_i \right) = 11 + 11 = 22.$$

However, if we change the value of r_3 to 4, then the same construction of $\mathcal{A}'_{\mathrm{F}_1}, \mathcal{A}'_{\mathrm{F}_2}$ violates the inequality since we would get $19 \ngeq 22$. In contrast, the construction in Lemma 3 does not depend on the values of ℓ_i, r_i. In fact, for the given example we would have $\mathcal{A}_{\mathrm{F}_1} = \{1, 2\}$ and $\mathcal{A}_{\mathrm{F}_2} = \{1, 2\}$, which would satisfy the inequality (3.13) for any possible configurations of ℓ_i, r_i.

The key property of the construction presented in Lemma 3.3.3 is that it is independent of (ℓ_i, r_i). This is of critical importance when we consider HD cuts, as we will see in Section 3.5 when we prove Theorem 3.5.1.

Remark 3.3.2. If the network and its subnetworks operate in full-duplex, then Lemma 3.3.3 directly relates cuts of the subnetworks $\bar{\mathcal{N}}_i$ to cuts of the full network $\mathcal{N}_{[1:N]}$ (see also the example above). Furthermore, by choosing $\mathcal{A}_i$ to be the minimum full-duplex cut of the subnetwork $\bar{\mathcal{N}}_i$, we get

$$N \max_{i \in [1:N]} \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\bar{\mathcal{N}}_i} \geq \sum_{i=1}^{N} \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\bar{\mathcal{N}}_i} \geq \sum_{j=1}^{N-1} \left(\max_{i \in \mathcal{A}_{\mathrm{F}j}} \ell_i + \max_{i \in [1:N] \setminus \mathcal{A}_{\mathrm{F}j}} r_i \right) \geq (N-1) \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}.$$

This is a different way of proving the result in [NOF14, Theorem 1] for $k = N - 1$.

3.4 A Simple Selection Algorithm

In this section, we investigate the performance of a simple algorithm that selects a subnetwork of $k = N - 1$ relays, in terms of the retained fraction of the approximate capacity

of the full network. In particular, the algorithm computes the N single-relay approximate capacities and removes the worst relay, i.e., the one with the smallest single-relay approximate capacity. Since computing the single-relay approximate capacities in a Gaussian HD diamond network with N relays requires $O(N)$ operations, this algorithm runs in linear time in the number of relays and outputs an $(N-1)$-relay subnetwork that retains at least half of the approximate capacity of the full network. Moreover, if only the single-relay approximate capacities $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}, \forall i \in [1:N]$ are known, i.e., the individual point-to-point link capacities are not available, then the guarantee aforementioned is tight. The main result of this section is summarized in the following theorem.

Theorem 3.4.1. Consider a Gaussian HD diamond network $\mathcal{N}_{[1:N]}$. Then, there always exists $i \in [1:N]$ such that we can guarantee at least $\widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$, where $\bar{\mathcal{N}}_i = \mathcal{N}_{[1:N]\setminus\{i\}}$. Moreover, if only the single-relay approximate capacities are known, then this bound is tight.

Proof. We argue the lower bound in Theorem 3.4.1 by contradiction. Without loss of generality, let $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{N\}}} \leq \min_{i\in[1:N]} \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}$, i.e., the N-th relay is the worst. Assume that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-1]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. From the implication of the partition lemma in (3.10), we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-1]}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{N\}}} \geq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. This property, together with the assumption that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-1]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$, implies that $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{N\}}} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. However, since the relay number N has the lowest approximate capacity, then $\forall j \in [1:N-1]$, $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{j\}}} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. Therefore, we finally have the following contradiction

$$\forall j \in [1:N-1], \quad \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{j\}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-1]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}.$$

This concludes the proof of the lower bound in Theorem 3.4.1.

To prove that the bound in Theorem 3.4.1 is indeed tight it suffices to provide a network construction where having the knowledge of only the single-relay approximate capacities does not guarantee that a subnetwork $\mathcal{N}_{\mathcal{K}}$ of $N-1$ relays, with $\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}$ strictly greater than $\frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$, can be chosen deterministically. For $N \geq 2$, let

$$\ell_i = \frac{1}{2}, \quad r_i = b, \qquad \forall i \in [1:N-1], \tag{3.14a}$$

$$\ell_N = b, \quad r_N = \frac{1}{2}, \tag{3.14b}$$

where $b \to \infty$. Note that for the network construction in (3.14) we have: (i) $\forall i \in [1 : N]$, $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} = \frac{1}{2}$ and (ii) the approximate HD capacity of the full network is $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} = 1$. We now want to remove the worst relay based only on the knowledge of the single-relay approximate capacities. Since these are all equal, then one can choose to remove one relay uniformly at random. If the N-th relay is removed, then the remaining network has an approximate capacity of $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-1]}} = \frac{1}{2}$, which shows that the lower bound in Theorem 3.4.1 is indeed tight if the choice of which relay to remove is based only on the single-relay approximate capacities. □

The tightness argument in Theorem 3.4.1 implies that, for an algorithm that removes the worst relay - by only computing the single-relay approximate capacities - no higher worst-case guarantee can be provided. However, this result is pretty conservative. In fact, with reference to the specific network construction in (3.14), if we are allowed to select $N-1$ relays based on the approximate capacities of the 2-relay subnetworks, then we would never remove the N-th relay. This is because any 2-relay subnetwork which involves the N-th relay has an approximate capacity of $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{N,i\}}} = 1 = \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}, \forall i \in [1 : N-1]$. This simple example suggests that a smarter choice (compared to the one based on removing the worst relay) of which $N-1$ relays to select might lead to a higher worst-case retained fraction, compared to the $\frac{1}{2}$ in Theorem 3.4.1. In the next section, we will formally prove that this observation is indeed true. Before concluding this section, we next generalize the lower bound in Theorem 3.4.1 to generic values of $k \in [1 : N]$.

3.4.1 The General Case $k \in [1 : N]$

We now generalize the lower bound in Theorem 3.4.1 when $k \in [1 : N]$. Towards this end, we consider an algorithm that removes the worst $N-k$ relays, i.e., those with the lowest single-relay approximate capacities, from the network of N relays. The algorithm first computes

the single-relay approximate capacities – which requires $O(N)$ operations. It then orders the relays in descending order based on their single-relay approximate capacities, i.e., in this new ordering the first relay is the one for which $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{1\}}} \geq \max_{i\in[2:N]} \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}$, the second relay is the one for which $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{2\}}} \geq \max_{i\in[3:N]} \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}$ and so on till the N-th relay for which $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{N\}}} = \min_{i\in[1:N]} \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}$; this step requires $O(N\log(N))$ operations. Finally, the algorithm discards the last $N-k$ relays. In other words, the algorithm runs in $O(N\log(N))$ and outputs a k-relay subnetwork whose performance guarantee is provided in the following lemma.

Lemma 3.4.1. Consider a Gaussian HD diamond network $\mathcal{N}_{[1:N]}$ where the relays are ordered in descending order based on their single-relay approximate capacities. By operating only the relays in $\mathcal{N}_{[1:k]}$, we can always guarantee at least $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:k]}} \geq 2^{-(N-k)}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$.

Proof. Clearly, for the case $k = N-1$ the lower bound in Lemma 3.4.1 is equivalent to the one in Theorem 3.4.1. We now argue the lower bound in Lemma 3.4.1 by contradiction. Without loss of generality, assume that instead of removing the last $N-k$ relays all together (recall that relays are ordered in descending order based on their single-relay approximate capacities), we remove them in $N-k$ steps, i.e., at step $i \in [1:N-k]$ we remove the relay number $N-i+1$. Assume that at step i we have that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$. From (3.10), we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i]}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{N-i+1\}}} \geq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$. This property, together with the assumption that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$, implies that $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{N-i+1\}}} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$. However, since the relay number $N-i+1$ has the lowest approximate HD capacity at step i, then $\forall j \in [1:N-i]$, $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{j\}}} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$. Therefore, we finally have the following contradiction

$$\forall j \in [1:N-i], \quad \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{j\}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i]}} < \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}.$$

Thus, $\forall i \in [1:N-k]$, we have that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i]}} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N-i+1]}}$. By recursively applying this expression $(N-k)$ times we are left with a k-relay subnetwork that retains an approximate capacity $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:k]}} \geq 2^{-(N-k)}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. This concludes the proof. □

3.5 A Fundamental Guarantee for Selecting $N-1$ Relays

In this section we derive a fundamental guarantee (in terms of retained fraction) when $N-1$ relays are selected out of the N possible ones. We assert that this guarantee is fundamental because it represents the highest worst-case fraction that can be guaranteed when $N-1$ relays are selected, independently of the actual values of the channel parameters. In particular, our main result is stated in the following theorem.

Theorem 3.5.1. For any N-relay Gaussian HD diamond network $\mathcal{N}_{[1:N]}$, there always exists a subnetwork $\mathcal{N}_{\mathcal{K}}$, with $|\mathcal{K}| = N-1$, that retains at least $\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}} \geq \frac{N-1}{N}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. Moreover, this bound is tight.

Proof. In order to derive the lower bound in Theorem 3.5.1, we first state the following lemma, whose proof is based on Lemma 3.3.3 and is delegated to Appendix 3.7.4.

Lemma 3.5.1. Consider an arbitrary N-relay Gaussian HD diamond network $\mathcal{N}_{[1:N]}$ operated with the schedule λ and let $\bar{\mathcal{N}}_i = \mathcal{N}_{[1:N]\setminus\{i\}}$ be the subnetwork of $N-1$ relays constructed by removing relay i. Then,

$$\sum_{i=1}^{N} \mathsf{R}^{\lambda}_{\bar{\mathcal{N}}_i} \geq (N-1)\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}, \tag{3.15}$$

where $\forall \mathcal{K} \subseteq [1:N]$ and $\mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}}$ is the HD rate for the network $\mathcal{N}_{\mathcal{K}}$ when operated with the deterministic schedule constructed from λ.

The lower bound in Theorem 3.5.1 is a direct consequence of Lemma 3.5.1 as explained in what follows. Let $\lambda^{\star}$ be an optimal schedule for the full network $\mathcal{N}_{[1:N]}$, i.e., $\mathsf{R}^{\lambda^{\star}}_{\mathcal{N}_{[1:N]}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. Since the 'natural' schedule constructed from $\lambda^{\star}$ might not be the optimal one for the subnetwork $\bar{\mathcal{N}}_i$, then clearly we have $\mathsf{R}^{\lambda^{\star}}_{\bar{\mathcal{N}}_i} \leq \widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i}, \forall i \in [1:N]$. Using the result in Lemma 3.5.1 with $\lambda^{\star}$, we get

$$(N-1)\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq \sum_{i=1}^{N} \mathsf{R}^{\lambda^{\star}}_{\bar{\mathcal{N}}_i} \leq \sum_{i=1}^{N} \widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i} \leq N \max_{i\in[1:N]} \widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i}.$$

Let $i^\star = \arg\max \left\{ \widetilde{\mathsf{C}}_{\mathcal{N}_i} \right\}$. Then, by setting $\mathcal{K} = [1:N] \backslash \{i^\star\}$, we have that

$$\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}} \geq \frac{N-1}{N} \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}.$$

This completes the proof of the lower bound in Theorem 3.5.1.

To prove that the ratio in Theorem 3.5.1 is tight, it suffices to provide an example of an N-relay network where the best (i.e., the one with the largest approximate capacity) subnetwork of $N-1$ relays retains an approximate capacity, which is exactly the fraction of the full network approximate capacity in Theorem 3.5.1. Towards this end, consider the following structure:

$$\ell_i = \ell_{\lfloor \frac{N}{2} \rfloor + i} = \frac{2i}{N}, \; i \in \left[1 : \left\lfloor \frac{N}{2} \right\rfloor \right], \tag{3.16a}$$

$$r_i = r_{\lfloor \frac{N}{2} \rfloor + i} = \frac{N - 2i + 2}{N}, \; i \in \left[1 : \left\lfloor \frac{N}{2} \right\rfloor \right], \tag{3.16b}$$

$$\text{if } N \text{ is odd: } \; \ell_N = b, \; r_N = \frac{1}{N}, \tag{3.16c}$$

where $b \to \infty$. Fig. 3.3 gives a representation of $\max\limits_{\mathcal{K} \subseteq [1:N]} \dfrac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}}$ for $N \in [2:10]$ with $|\mathcal{K}| = N-1$. From Fig. 3.3 we observe that $\max\limits_{\mathcal{K} \subseteq [1:N]} \dfrac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \dfrac{N-1}{N}$. This completes the proof. □

Before concluding this section, we highlight some results, which are direct consequences of Lemma 3.5.1 and Theorem 3.5.1.

Remark 3.5.1. Theorem 3.5.1 provides a performance guarantee that significantly improves over the one in Theorem 3.4.1. In fact, for high values of N, Theorem 3.5.1 ensures that we can approach $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$ by operating only $N-1$ relays, which is twice the guarantee of $\frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$ (independent of the value of N) provided by Theorem 3.4.1.

Remark 3.5.2. The result in Theorem 3.5.1 implies that, for any N-relay Gaussian HD diamond network, smartly removing one relay can reduce the approximate HD capacity of

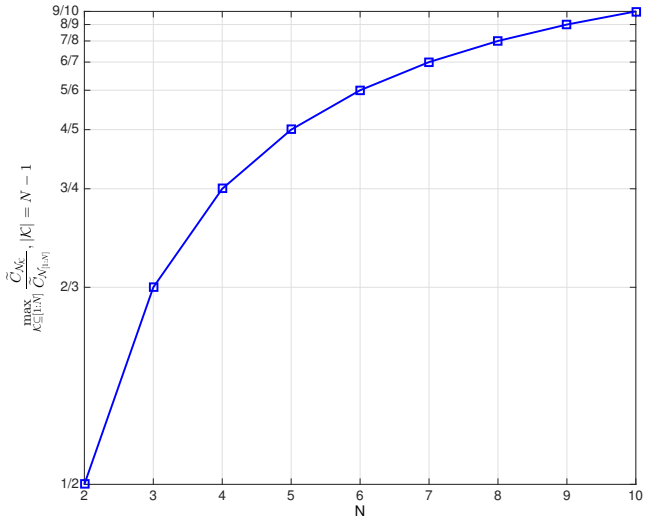


Figure 3.3: $\max_{\mathcal{K}\subseteq[1:N]} \frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}}$ with $|\mathcal{K}| = N-1$ for the network in (3.16) for $N \in [2:10]$.

the network by at most $\frac{1}{N}$ of the full network approximate capacity. We also highlight that the removed relay may not be the worst relay since in this case, as proved in Theorem 3.4.1, we can guarantee only $\widetilde{\mathsf{C}}_{\bar{\mathcal{N}}_i} \geq \frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$, where $i \in [1:N]$ is the index of the worst relay. However, for the specific network in (3.16) the full network has an approximate capacity of $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} = 1$ (see Appendix 3.7.5 for the detailed computation) and all the $(N-1)$-relay subnetworks have an approximate capacity of $\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}} = \frac{N-1}{N}$, $\forall \mathcal{K} \subseteq [1:N]$, $|\mathcal{K}| = N-1$. Hence, for this particular network, by removing *any* of the relays (i.e., the best or the worst), we always retain $\frac{N-1}{N}$ of the approximate capacity of the full network.

Corollary 3.5.1. Let $\lambda^\star$ be an optimal schedule for the approximate capacity of the full network $\mathcal{N}_{[1:N]}$, then:

1. For any N-relay Gaussian HD diamond network, there exists a subnetwork $\mathcal{N}_\mathcal{K}$, with $|\mathcal{K}| = N-1$, such that, when operated with $\lambda^\star$, it satisfies that

$$\mathsf{R}^{\lambda^\star}_{\mathcal{N}_\mathcal{K}} \geq \frac{N-1}{N}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}.$$

2. There exist N-relay Gaussian HD diamond networks where $\lambda^\star$ can be used to naturally construct the optimal schedule for the approximate capacity of each subnetwork of $N-1$ relays (see for example, the network in (3.16)).

Remark 3.5.3. Corollary 3.5.1 implies that, to select a subnetwork of $N-1$ relays that guarantees the performance in Theorem 3.5.1, it is sufficient to know an optimal schedule $\lambda^\star$ for the approximate capacity of the whole network $\mathcal{N}_{[1:N]}$. In other words, by knowing $\lambda^\star$, there is no need to compute the optimal schedules for the approximate capacity of each of the N subnetworks. This implies that, if $\lambda^\star$ can be used to construct a 'natural' schedule for all $\mathcal{N}_\mathcal{K}$, with $|\mathcal{K}| = N-1$, in polynomial time, then a subnetwork $\mathcal{N}_\mathcal{K}$ that satisfies the guarantee in Theorem 3.5.1 can be discovered in polynomial time.

We next leverage the result in Theorem 3.5.1 to derive a lower bound for $k \in [1:N]$.

3.5.1 The General Case $k \in [1:N]$

In this subsection we generalize the lower bound derived in Theorem 3.5.1 when $k \in [1:N]$. In particular, our result is stated in the following lemma.

Lemma 3.5.2. Consider an arbitrary N-relay Gaussian HD diamond network $\mathcal{N}_{[1:N]}$ operated with the schedule λ. There always exists a subnetwork $\mathcal{N}_\mathcal{K}$ with $|\mathcal{K}| = k \in [1:N]$ that, when operated with the 'natural' schedule derived from λ, retains an approximate rate $\mathsf{R}^\lambda_{\mathcal{N}_\mathcal{K}}$ such that $\mathsf{R}^\lambda_{\mathcal{N}_\mathcal{K}} \geq \frac{k}{N}\mathsf{R}^\lambda_{\mathcal{N}_{[1:N]}}$.

Proof. We recursively apply the result in Lemma 3.5.1. We again let λ be a schedule (not necessarily optimal) of the full N-relay network $\mathcal{N}_{[1:N]}$. With this we obtain

$$\exists\, i_1 \in [1:N] \text{ such that for } \mathcal{S}^{(1)} = \mathcal{N}_{[1:N]\setminus\{i_1\}}: \tag{3.17a}$$

$$\mathsf{R}^\lambda_{\mathcal{S}^{(1)}} \geq \frac{N-1}{N}\mathsf{R}^\lambda_{\mathcal{N}_{[1:N]}}, \tag{3.17b}$$

$$\exists\, i_2 \in \mathcal{S}^{(1)} = \mathcal{N}_{[1:N]\setminus\{i_1\}} \text{ such that for } \mathcal{S}^{(2)} = \mathcal{N}_{[1:N]\setminus\{i_{[1:2]}\}}:$$

$$\mathsf{R}^{\lambda}_{\mathcal{S}^{(2)}} \geq \frac{N-2}{N-1}\mathsf{R}^{\lambda}_{\mathcal{S}^{(1)}} \overset{(3.17b)}{\geq} \frac{N-2}{N}\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}, \tag{3.17c}$$

$\exists\, i_3 \in \mathcal{S}^{(2)} = \mathcal{N}_{[1:N]\setminus\{i_{[1:2]}\}}$ such that for $\mathcal{S}^{(3)} = \mathcal{N}_{[1:N]\setminus\{i_{[1:3]}\}}$:

$$\mathsf{R}^{\lambda}_{\mathcal{S}^{(3)}} \geq \frac{N-3}{N-2}\mathsf{R}^{\lambda}_{\mathcal{S}^{(2)}} \overset{(3.17c)}{\geq} \frac{N-3}{N}\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}, \tag{3.17d}$$

$$\vdots$$

$\exists\, i_{N-k} \in \mathcal{S}^{(N-k-1)}$ such that for $\mathcal{S}^{(N-k)} = \mathcal{N}_{[1:N]\setminus\{i_{[1:N-k]}\}}$:

$$\mathsf{R}^{\lambda}_{\mathcal{S}^{(N-k)}} \geq \frac{k}{k+1}\mathsf{R}^{\lambda}_{\mathcal{S}^{(N-k-1)}} \geq \frac{k}{N}\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}, \tag{3.17e}$$

which, since $\mathcal{S}^{(N-k)}$ contains k relays, completes the proof. □

Remark 3.5.4. Let $\lambda^\star$ be an optimal schedule for the approximate capacity of the full network $\mathcal{N}_{[1:N]}$, i.e., $\mathsf{R}^{\lambda^\star}_{\mathcal{N}_{[1:N]}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. Since the 'natural' schedule constructed from $\lambda^\star$ might not be the optimal one for the approximate capacity of the subnetwork $\mathcal{N}_{\mathcal{K}}$, i.e., $\mathsf{R}^{\lambda^\star}_{\mathcal{N}_{\mathcal{K}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}$, then Lemma 3.5.2 provides a different bound from the one in [BF14a] and from the $\frac{k}{2(k+1)}$ that is readily obtained from the result in [NOF14]. These bounds can be combined as

$$\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} \geq \begin{cases} \max\left\{\frac{1}{N}\,,\frac{1}{4}\right\}, & k=1 \\ \max\left\{\frac{k}{N}\,,\frac{1}{2}\right\}, & N \geq k \geq 2 \end{cases}. \tag{3.18}$$

From (3.18), we can see that in some cases (particularly when $k > N/2$), the new bound in Lemma 3.5.2 gives a better guarantee than those available in the literature. Clearly, when $k = N-1$ the lower bound in (3.18) is equivalent to the one in Theorem 3.5.1. However, the lower bound in Lemma 3.5.2 is not tight for general $k \in [1:N]$. Deriving tighter lower bounds is an interesting open problem, which is object of current investigation. For instance, for the case $k = 1$, numerically we could not find network examples for which the fraction guarantee is less than $\frac{N}{4(N-1)}$.

Remark 3.5.5. The proof of Lemma 3.5.2 provides the blueprint for an algorithm that selects a subnetwork of k relays that satisfies the guarantee in the lemma. The algorithm operates iteratively as follows. On the first iteration, given a network $\mathcal{N}^{(0)} = \mathcal{N}_{[1:N]}$ with

N relays and an operating schedule λ, we find a subnetwork $\mathcal{N}^{(1)}$ with $N-1$ relays such that $\mathcal{N}^{(1)}$, when operated with the 'natural' schedule derived from λ, satisfies the bound in Lemma 3.5.2 for $k = N-1$. We can repeat the previous iteration $(N-k)$ times where on iteration i, we remove one relay to select a subnetwork $\mathcal{N}^{(i)}$ such that

$$\mathsf{R}^{\lambda}_{\mathcal{N}^{(i)}} \geq \frac{N-i}{N-i+1}\mathsf{R}^{\lambda}_{\mathcal{N}^{(i-1)}}.$$

It is clear that after $(N-k)$ iterations, we have a subnetwork $\mathcal{N}^{(N-k)}$ that contains exactly k relays and for which

$$\mathsf{R}^{\lambda}_{\mathcal{N}^{(N-k)}} \geq \frac{k}{N}\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}.$$

In [PE14] the authors showed that the problem of computing the approximate capacity of a Gaussian full-duplex relay network can be cast as a minimization problem of a submodular function, which can be solved in polynomial time. Therefore, if the fixed schedule λ at which $\mathcal{N}_{[1:N]}$ is operated can be used to construct a 'natural' schedule for $\mathcal{N}^{(1)}$ in polynomial time, then the algorithm described above runs in polynomial time and provides the fraction guarantee in Lemma 3.5.2.

3.6 Discussion and Conclusions

In this section, we discuss some implications of the results derived in the previous sections and highlight differences between the selection performances in HD and full-duplex diamond networks. We believe that the reason for this different behavior is that in HD the schedule plays a key role, i.e., removing some of the relays can change the optimal schedule for the approximate capacity of the remaining network.

1) In HD the guarantee on $\max_{\mathcal{K}\subseteq[1:N]} \frac{\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}}$ for $|\mathcal{K}| = k \in [1:2]$ decreases as N increases. We here show that in HD, for the case $|\mathcal{K}| = k \in [1:2]$, the worst-case fraction $\max_{\mathcal{K}\subseteq[1:N]} \frac{\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}}$ depends on N and decreases as N increases. This represents a surprising

difference with respect to full-duplex – where the worst case ratio for a fixed value of k does not depend on N – and shows that full-duplex and HD relay networks have a different nature. In particular, from the result in Theorem 3.5.1 for $|\mathcal{K}| = k \in [1:2]$ and $N = k+1$, we have $\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} \geq \frac{k}{k+1}$ as in full-duplex [NOF14, Theorem 1]. However, in the regime $N \gg 1$, these values reduce to $\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} \geq \frac{1}{4}$ for $k = 1$ and to $\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} \geq \frac{1}{2}$ for $k = 2$. Notice that these values coincide with the lower bounds: (i) of $\frac{k}{2(k+1)}$ for $k = 1$, which is readily obtained from the result in [NOF14] by letting the selected relay listen for half of the time and transmit for the other half of the time; (ii) derived in [BF14a] for the case $k = 2$, where the 2 selected relays operate in a complementary fashion. In particular, we have

Theorem 3.6.1. There exist Gaussian HD diamond networks for which, when $N \gg 1$, the best subnetwork $\mathcal{N}_\mathcal{K}$ gives

$$\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \begin{cases} \frac{1}{4} & |\mathcal{K}| = 1, \\ \frac{1}{2} & |\mathcal{K}| = 2. \end{cases} \tag{3.19}$$

Proof. Consider the network in (3.16). The best subnetwork $\mathcal{N}_\mathcal{K}$ with $|\mathcal{K}| = 1$ satisfies

$$\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \frac{N+2}{4N},$$

which for $N \gg 1$ gives $\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \frac{1}{4}$, while the best subnetwork $\mathcal{N}_\mathcal{K}$ with $|\mathcal{K}| = 2$ relays satisfies

$$\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \frac{N+2}{2N},$$

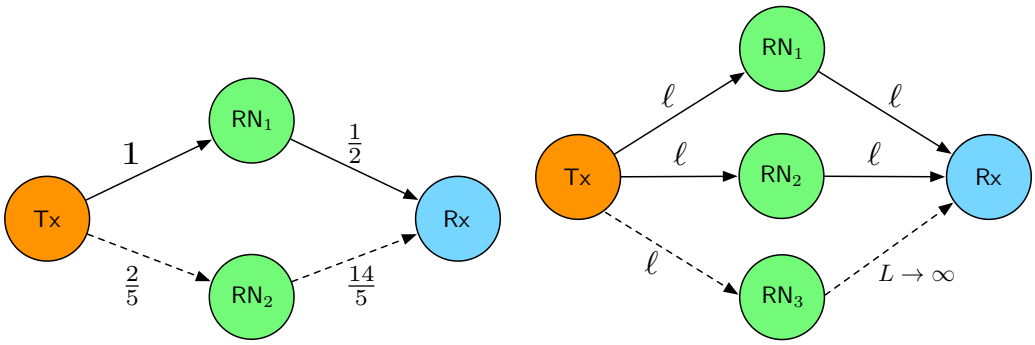
which for $N \gg 1$ gives $\frac{\widetilde{\mathsf{C}}_{\mathcal{N}_\mathcal{K}}}{\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}} = \frac{1}{2}$. We refer the reader to Appendix 3.7.5 for a detailed computation of these values. □

Table 3.2 summarizes the results presented above for $k \in [1:2]$. In particular, Table 3.2 shows that in full-duplex, the worst-case fraction guarantee does not depend on N, while in HD the worst-case guarantee decreases as N increases.

Table 3.2: Worst-case fraction guarantees in full-duplex and HD for $k \in [1:2]$.

	Full-Duplex	**Half-Duplex**
$N = k+1$	$\frac{k}{k+1}$	$\frac{k}{k+1}$
$N \gg 1$	$\frac{k}{k+1}$	$\begin{cases} 1/4 & k=1, \\ 1/2 & k=2 \end{cases}$

2) The best HD and full-duplex subnetworks are not necessarily the same. We next provide a couple of examples where we show that the best relay in HD and in full-duplex might not be necessarily the same. As a first example, consider the Gaussian 2-relay diamond network depicted in Fig. 3.4(a). It is not difficult to see that if the relays operate in full-duplex, then the first relay is the best and it achieves Shannon capacity $\mathsf{C}^{\mathrm{FD}}_{\mathcal{N}_{\{1\}}} = \frac{1}{2}$, while if the relays operate in HD then the second relay is the best giving $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{2\}}} = \frac{2/5 \times 14/5}{2/5+14/5} = 7/20$ compared to $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{1\}}} = \frac{1\times 1/2}{1+1/2} = 1/3$ given by the first relay.



(a) (b)

Figure 3.4: Network examples where the best full-duplex and HD subnetworks are different. Link labels represent the point-to-point link capacities and the best HD single-relay subnetwork is marked with dashed links.

As a second example consider the Gaussian 3-relay diamond network shown in Fig. 3.4(b). When the $N = 3$ relays operate in full-duplex, they all have the same single-relay Shannon capacity given by $\mathsf{C}^{\mathrm{FD}}_{\mathcal{N}_{\{i\}}} = \ell, \forall i \in [1:3]$. This means that, by selecting any of the relays at random, we get the same performance guarantee. Differently, when the $N = 3$ relays operate

in HD, the third relay is strictly better giving $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{3\}}} = \ell$ compared to $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} = \ell/2$, $\forall i \in [1:2]$. These two simple examples suggest that, when the relays operate in HD, choosing the best subnetwork based on the full-duplex capacities might not be a smart choice. For instance, in the second example if we select either the first or the second relay which is optimal in full-duplex, we would incur a loss of 50% in the approximate capacity compared to selecting the third relay.

3) Worst-case networks in HD and full-duplex are not necessarily the same. Consider the network example in (3.16) and suppose that we want to select $N-1$ relays. We already showed (see Section 3.5) that, by selecting any $(N-1)$-relay subnetwork $\mathcal{N}_{\mathcal{K}}$ with $|\mathcal{K}| = N-1$, we get $\widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}} = \frac{N-1}{N}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$, i.e., the network in (3.16), when operated in HD, represents a worst-case scenario. Now, suppose that we operate the network in (3.16) in full-duplex. Then, it is not difficult to see that there always exists an $(N-1)$-relay subnetwork $\mathcal{N}_{\mathcal{K}}$ with $|\mathcal{K}| = N-1$, that guarantees $\widetilde{\mathsf{C}}^{\text{FD}}_{\mathcal{N}_{\mathcal{K}}} = \widetilde{\mathsf{C}}^{\text{FD}}_{\mathcal{N}_{[1:N]}}$, which is greater than the worst-case ratio of $\frac{N-1}{N}$ proved in [NOF14, Theorem 1]. This suggests that tight network examples for HD with general values of k and N might not be the same as those in full-duplex; this adds an extra degree of complication in the study of the network simplification problem in HD since the approximate capacity in HD (because of the required optimization over the 2^N listen/transmit states) cannot be computed directly as in the full-duplex counterpart.

In this chapter, we investigated the network simplification problem in an N-relay Gaussian HD diamond network. We proved that there always exists a subnetwork of $k = N-1$ relays that retains at least a fraction $\frac{N-1}{N}$ of the approximate capacity of the full network. This result was derived by showing that any optimal schedule for the approximate capacity of the full network can be used by at least one of the N subnetworks of $k = N-1$ relays to satisfy the worst performance guarantee. Moreover, we provided an example of a class of Gaussian HD diamond networks for which this fraction is tight. Then, by leveraging the results obtained for $k = N-1$, we derived lower bounds on the fraction guarantee for general $k \in [1:N]$, which are tighter than currently available bounds when $k > \frac{N}{2}$. Finally,

we showed that, when we select $k = 1$ or $k = 2$ relays, the fraction guarantee decreases as N increases; this is a surprising difference between the network simplification problem in HD and full-duplex. These results were obtained by leveraging properties of submodular functions and diamond networks that were derived here and that might be of independent interest for other applications.

3.7 Appendices

3.7.1 Proof of Lemma 3.3.1 (Partition lemma)

In order to prove the result in the partition lemma, we make use of the following lemma, valid for Gaussian full-duplex diamond networks.

Lemma 3.7.1. For any Gaussian full-duplex diamond network $\mathcal{N}_{[1:N]}$ and $\mathcal{K} \subseteq [1:N]$, we have that

$$\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}} \leq \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{K}}} + \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]\backslash\mathcal{K}}}, \tag{3.20}$$

where

$$\begin{aligned}
\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}} &= \min_{\mathcal{A}_{\mathrm{F}} \subseteq [1:N]} \left\{ \max_{i \in \mathcal{A}_{\mathrm{F}}} \ell_i + \max_{i \in \mathcal{N}_{[1:N]} \backslash \mathcal{A}_{\mathrm{F}}} r_i \right\} \\
\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{M}_j}} &= \min_{\mathcal{A}_j \subseteq \mathcal{M}_j} \left\{ \max_{i \in \mathcal{A}_j} \ell_i + \max_{i \in \mathcal{M}_j \backslash \mathcal{A}_j} r_i \right\}, \qquad \forall j \in [1:2],
\end{aligned}$$

with $\mathcal{M}_1 = \mathcal{K}$ and $\mathcal{M}_2 = [1:N]\backslash\mathcal{K}$.

Proof. For any two sets $\mathcal{A}_1$ and $\mathcal{A}_2$, where $\mathcal{A}_i \subseteq \mathcal{M}_i, \forall i \in [1:2]$, we have that

$$\begin{aligned}
&\max_{i \in \mathcal{A}_1} \ell_i + \max_{i \in \mathcal{A}_2} \ell_i + \max_{i \in \mathcal{M}_1 \backslash \mathcal{A}_1} r_i + \max_{i \in \mathcal{M}_2 \backslash \mathcal{A}_2} r_i \\
&\geq \max_{i \in \mathcal{A}_1 \cup \mathcal{A}_2} \ell_i + \max_{i \in (\mathcal{M}_1 \backslash \mathcal{A}_1) \cup (\mathcal{M}_2 \backslash \mathcal{A}_2)} r_i \\
&\overset{(a)}{=} \max_{i \in \mathcal{A}_1 \cup \mathcal{A}_2} \ell_i + \max_{i \in (\mathcal{M}_1 \cup \mathcal{M}_2) \backslash (\mathcal{A}_1 \cup \mathcal{A}_2)} r_i
\end{aligned}$$

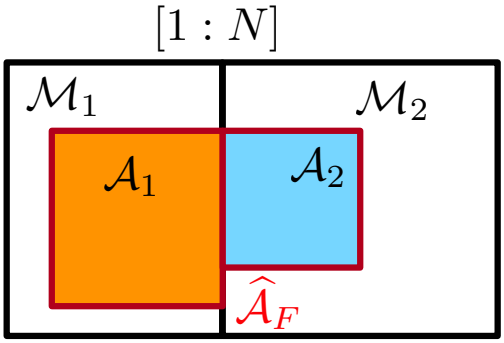


Figure 3.5: Illustration of the relationship between sets in (3.22).

$$\begin{aligned} &\overset{\text{(b)}}{=} \max_{i\in\widehat{\mathcal{A}}_{\mathrm{F}}} \ell_i + \max_{i\in[1:N]\setminus\widehat{\mathcal{A}}_{\mathrm{F}}} r_i \\ &\geq \min_{\mathcal{A}_{\mathrm{F}}\subseteq[1:N]} \left\{ \max_{i\in\mathcal{A}_{\mathrm{F}}} \ell_i + \max_{i\in[1:N]\setminus\mathcal{A}_{\mathrm{F}}} r_i \right\} = \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}. \end{aligned} \tag{3.21}$$

The equality in (b) follows by defining $\widehat{\mathcal{A}}_{\mathrm{F}} = \mathcal{A}_1 \cup \mathcal{A}_2$ and from the fact that $\mathcal{M}_1 \cup \mathcal{M}_2 = [1:N]$. The equality in (a) appeals to the following property (recall that $\mathcal{M}_1$ and $\mathcal{M}_2$ are disjoint and $\mathcal{A}_i \subseteq \mathcal{M}_i, i \in [1:2]$)

$$(\mathcal{M}_1\backslash\mathcal{A}_1) \cup (\mathcal{M}_2\backslash\mathcal{A}_2) \overset{\text{(c)}}{=} (\mathcal{M}_1\setminus(\mathcal{A}_1\cup\mathcal{A}_2)) \cup (\mathcal{M}_2\setminus(\mathcal{A}_1\cup\mathcal{A}_2)) \tag{3.22}$$

$$\overset{\text{(d)}}{=} (\mathcal{M}_1\cup\mathcal{M}_2)\setminus(\mathcal{A}_1\cup\mathcal{A}_2), \tag{3.23}$$

where the equality in (c) follows since $\mathcal{M}_1 \cap \mathcal{A}_2 = \emptyset$ and $\mathcal{M}_2 \cap \mathcal{A}_1 = \emptyset$ and the equality in (d) follows since $(\mathcal{B}\backslash\mathcal{A}) \cup (\mathcal{C}\backslash\mathcal{A}) = (\mathcal{B}\cup\mathcal{C})\backslash\mathcal{A}$. An illustration of the relationship in (3.22) is shown in Fig. 3.5. The result in (3.21) is valid $\forall\mathcal{A}_1 \subseteq \mathcal{M}_1$ and $\forall\mathcal{A}_2 \subseteq \mathcal{M}_2$, hence also for the minimum cuts of the networks $\mathcal{N}_{\mathcal{M}_1}$ and $\mathcal{N}_{\mathcal{M}_2}$, i.e.,

$$\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{M}_1}} + \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{M}_2}} = \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{\mathcal{K}}} + \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]\setminus\mathcal{K}}} \geq \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}.$$

□

We now show how the result in Lemma 3.7.1, valid for Gaussian full-duplex diamond networks, extends to the HD case. For a given schedule λ of the full network $\mathcal{N}_{[1:N]}$, from (3.7)

we have that

$$\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}} = \min_{\mathcal{A}_{\mathrm{F}} \subseteq [1:N]} \sum_{s \in [0:1]^N} \lambda_s \left(\max_{i \in \mathcal{A}_{\mathrm{F}}} \ell'_{i,s} + \max_{i \in [1:N] \setminus \mathcal{A}_{\mathrm{F}}} r'_{i,s} \right),$$

where $\ell'_{i,s}$ and $r'_{i,s}$ are defined in (3.8). From the result in (3.21), $\forall \mathcal{A}_1 \subseteq \mathcal{M}_1$ and $\forall \mathcal{A}_2 \subseteq \mathcal{M}_2$, with $\mathcal{M}_1 = \mathcal{K}$ and $\mathcal{M}_2 = [1:N] \backslash \mathcal{K}$, we have that

$$\begin{aligned} &\sum_{s \in [0:1]^N} \lambda_s \left[\max_{i \in \mathcal{A}_1} \ell'_{i,s} + \max_{i \in \mathcal{A}_2} \ell'_{i,s} + \max_{i \in \mathcal{M}_1 \setminus \mathcal{A}_1} r'_{i,s} + \max_{i \in \mathcal{M}_2 \setminus \mathcal{A}_2} r'_{i,s} \right] \\ \geq &\sum_{s \in [0:1]^N} \lambda_s \left(\max_{i \in \widehat{\mathcal{A}}_{\mathrm{F}}} \ell'_{i,s} + \max_{i \in [1:N] \setminus \widehat{\mathcal{A}}_{\mathrm{F}}} r'_{i,s} \right) \geq \mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}, \end{aligned}$$

where $\widehat{\mathcal{A}}_{\mathrm{F}} = \mathcal{A}_1 \cup \mathcal{A}_2$. This implies

$$\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}} \leq \mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{M}_1}} + \mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{M}_2}} = \mathsf{R}^{\lambda}_{\mathcal{N}_{\mathcal{K}}} + \mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N] \setminus \mathcal{K}}}.$$

This concludes the proof of Lemma 3.3.1.

3.7.2 Proof of Lemma 3.3.2

Let f be a submodular set function defined on Ω (see Definition 3.3.1). We want to prove that for any collection of n sets $\mathcal{A}_i \subseteq \Omega$,

$$\sum_{i=1}^{n} f\left(\mathcal{A}_i\right) \geq \sum_{j=1}^{n} f\left(\mathcal{E}_j^{(n)}\right),$$

where $\mathcal{E}_j^{(n)}$ is the set of elements that appear in at least j sets $\mathcal{A}_i, i \in [1:n]$. The proof is by induction. For the base case (i.e., $n = 1$) we clearly have that $f(\mathcal{A}_1) = f\left(\mathcal{E}_1^{(1)}\right)$. For the proof of the induction step, we prove and use the following property of submodular functions.

Property 3.7.1. Let f be a submodular function. Then, $\forall n > 0$ and $0 \leq k < n$,

$$f\left(\bigcup_{\substack{\mathcal{I} \subseteq [1:n] \\ |\mathcal{I}| = k}} \left(\mathcal{A}_{n+1} \bigcap_{i \in \mathcal{I}} \mathcal{A}_i \right) \right) + f\left(\bigcup_{\substack{\mathcal{I} \subseteq [1:n] \\ |\mathcal{I}| = k+1}} \left(\bigcap_{i \in \mathcal{I}} \mathcal{A}_i \right) \right)$$

$$\geq f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right). \tag{3.24}$$

We now use Property 3.7.1, whose proof can be found at the end of this appendix, to prove the induction step. Assume that for some $n>0$, we have that

$$\sum_{i=1}^{n} f\left(\mathcal{A}_i\right)\geq\sum_{j=1}^{n} f\left(\mathcal{E}_j^{(n)}\right). \tag{3.25}$$

Our goal is to prove that

$$\sum_{i=1}^{n+1} f\left(\mathcal{A}_i\right)\geq\sum_{j=1}^{n+1} f\left(\mathcal{E}_j^{(n+1)}\right).$$

From (3.25), by adding the positive quantity $f\left(\mathcal{A}_{n+1}\right)$ to both sides of the inequality, we have that

$$\sum_{i=1}^{n} f\left(\mathcal{A}_i\right)+f\left(\mathcal{A}_{n+1}\right)\geq\sum_{j=1}^{n} f\left(\mathcal{E}_j^{(n)}\right)+f\left(\mathcal{A}_{n+1}\right),$$

which can be equivalently rewritten as

$$\begin{aligned}\sum_{i=1}^{n} f\left(\mathcal{A}_i\right)+f\left(\mathcal{A}_{n+1}\right)&\geq f\left(\mathcal{A}_{n+1}\right)+\sum_{k=0}^{n-1}\underbrace{f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)}_{f\left(\mathcal{E}_{k+1}^{(n)}\right)}\\&=f\left(\mathcal{A}_{n+1}\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+\sum_{k=1}^{n-1} f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right).\end{aligned}$$

The final step in the proof follows by inductively applying Property 3.7.1 on the underlined terms with the appropriate k as shown in what follows,

$$\underline{f\left(\mathcal{A}_{n+1}\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)}+\sum_{k=1}^{n-1} f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)$$

$$
\begin{aligned}
&\overset{(k=0)}{\geq} f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+\underline{f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=1}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=2}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)}\\
&\quad+\sum_{k=2}^{n-1}f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)\\
&\overset{(k=1)}{\geq}\sum_{\ell=1}^{2}f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=\ell}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+\underline{f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=2}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=3}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)}\\
&\quad+\sum_{k=3}^{n-1}f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)\\
&\qquad\qquad\vdots\\
&\overset{(k=n-1)}{\geq}\sum_{\ell=1}^{n}f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=\ell}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=n}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)\\
&=\sum_{\ell=1}^{n+1}f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=\ell}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)=\sum_{j=1}^{n+1}f\left(\mathcal{E}_j^{(n+1)}\right).
\end{aligned}
$$

This concludes the proof of Lemma 3.3.2.

3.7.2.1 Proof of Property 3.7.1

By using properties of submodular functions and set operations we have

$$
f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)+f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)
$$

$$
\begin{aligned}
&\overset{(a)}{\geq} f\left(\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k}}\left(\mathcal{A}_{n+1}\bigcap\mathcal{A}_i\right)\right]\bigcup\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right]\right)\\
&\quad + f\left(\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k}}\left(\mathcal{A}_{n+1}\bigcap\mathcal{A}_i\right)\right]\bigcap\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right]\right)\\
&\overset{(b)}{=} f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right) + f\left(\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k}}\left(\mathcal{A}_{n+1}\bigcap\mathcal{A}_i\right)\right]\bigcap\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right]\right)\\
&\overset{(c)}{=} \underbrace{f\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n+1]\\|\mathcal{I}|=k+1}}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\right)}_{T_1} + f\left(\underbrace{\mathcal{A}_{n+1}\bigcap\left[\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right]\bigcap\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right]}_{\mathcal{S}}\right),
\end{aligned}
$$

where: (i) the inequality in (a) follows from the definition of submodular function (see Definition 3.3.1); (ii) the equality in (b) follows by combining the union in the first term of the inequality in (a); (iii) the equality in (c) follows from the distributive property of intersection over unions. Note that T_1 is already the first term we need in the inequality. To arrive at the second term, we shall prove that

$$
\mathcal{S} = \bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\mathcal{A}_{n+1}\bigcap\mathcal{A}_i\right). \tag{3.26}
$$

Towards this end, notice that the distributive property of intersection over unions gives

$$
\begin{aligned}
&\mathcal{A}_{n+1}\bigcap\left[\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right]\bigcap\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right]\\
&\quad = \mathcal{A}_{n+1}\bigcap\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left[\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\bigcap\left(\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right)\right].
\end{aligned} \tag{3.27}
$$

Now note that $\forall \mathcal{I} \subseteq [1:n]$ with $|\mathcal{I}| = k+1$, $\exists\ \mathcal{J}_{\mathcal{I}} \subset \mathcal{I}$ with $|\mathcal{J}_{\mathcal{I}}| = k$. This observation implies that, for each $\mathcal{I}$, we have

$$\left(\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcap\left(\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right)=\left(\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcap\left(\left(\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcup\left(\bigcup_{\substack{\mathcal{L}\subseteq[1:n]\\\mathcal{L}\neq\mathcal{J}_{\mathcal{I}}\\|\mathcal{L}|=k}}\bigcap_{\ell\in\mathcal{L}}\mathcal{A}_\ell\right)\right)\overset{(c)}{=}\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i,$$

where the equality in (c) follows since $\mathcal{U}\cap(\mathcal{U}\cup\mathcal{V})=\mathcal{U}$. As a consequence, for each $\mathcal{I}$, we have

$$\begin{aligned}\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\bigcap\left(\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right)&=\left(\bigcap_{i\in\mathcal{I}\backslash\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcap\left(\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcap\left(\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right)\\&=\left(\bigcap_{i\in\mathcal{I}\backslash\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)\bigcap\left(\bigcap_{i\in\mathcal{J}_{\mathcal{I}}}\mathcal{A}_i\right)=\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right).\end{aligned}\tag{3.28}$$

Finally, by applying (3.28) for each $\mathcal{I}$ in (3.27), we get

$$\begin{aligned}\mathcal{S}&=\mathcal{A}_{n+1}\bigcap\left[\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right]\bigcap\left[\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right]\\&=\mathcal{A}_{n+1}\bigcap\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left[\left(\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)\bigcap\left(\bigcup_{\substack{\mathcal{J}\subseteq[1:n]\\|\mathcal{J}|=k}}\bigcap_{j\in\mathcal{J}}\mathcal{A}_j\right)\right]\\&=\mathcal{A}_{n+1}\bigcap\left(\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right)=\bigcup_{\substack{\mathcal{I}\subseteq[1:n]\\|\mathcal{I}|=k+1}}\left(\mathcal{A}_{n+1}\bigcap_{i\in\mathcal{I}}\mathcal{A}_i\right),\end{aligned}$$

where the last equality follows by using the distributive property of intersection over unions. This proves (3.26) hence concluding the proof of Property 3.7.1.

3.7.3 Proof of Lemma 3.3.3

From the statement of Lemma 3.3.3, recall that $\mathcal{A}_i \subseteq [1:N]\backslash\{i\}$. Throughout the proof, we let $\mathcal{B}_i = ([1:N]\backslash\{i\})\backslash\mathcal{A}_i$, $\forall i \in [1:N]$, $f(\mathcal{A}) = \max_{i\in\mathcal{A}} \ell_i$ and $g(\mathcal{A}) = \max_{i\in\mathcal{A}} r_i$, with

$\mathcal{A} \subseteq [1:N]$. It is not difficult to see that f and g are submodular functions. As a result, we have

$$\begin{aligned}
&\sum_{j=1}^{N} \left(\max_{i \in \mathcal{A}_j} \ell_i + \max_{i \in ([1:N] \backslash \{j\}) \backslash \mathcal{A}_j} r_i \right) \\
&= \sum_{j=1}^{N} \left[f\left(\mathcal{A}_j\right) + g\left(\mathcal{B}_j\right) \right] \\
&\overset{\text{(a)}}{\geq} \sum_{j=1}^{N} \left[f\left(\mathcal{E}_j^{(N)}\right) + g\left(\mathcal{F}_j^{(N)}\right) \right] \\
&\overset{\text{(b)}}{=} \sum_{j=1}^{N-1} \left[f\left(\mathcal{E}_j^{(N)}\right) + g\left(\mathcal{F}_j^{(N)}\right) \right] \\
&\overset{\text{(c)}}{=} \sum_{j=1}^{N-1} \left[f\left(\mathcal{E}_j^{(N)}\right) + g\left(\mathcal{F}_{N-j}^{(N)}\right) \right], \qquad (3.29)
\end{aligned}$$

where: (i) the inequality in (a) follows from Lemma 3.3.2 with $\mathcal{E}_j^{(N)}$ (respectively, $\mathcal{F}_j^{(N)}$) being the set of elements that appear in at least j sets $\mathcal{A}_i,\ i \in [1:N]$ (respectively, $\mathcal{B}_i$); (ii) the equality in (b) follows because $\mathcal{E}_N^{(N)} = \mathcal{F}_N^{(N)} = \emptyset$ since $\bigcap_{i=1}^{N}([1:N]\backslash\{i\}) = \emptyset$; (iii) the equality in (c) follows by simply reordering the sum.

Fix an element $i \in [1:N]$. By definition of sets $\mathcal{A}_i$, $\mathcal{B}_i$, it holds that $i \notin \mathcal{A}_i$ and $i \notin \mathcal{B}_i$. Furthermore, for any $j \in [1:N]$ with $j \neq i$, we have $i \in \mathcal{A}_j$ if and only if $i \notin \mathcal{B}_j$. Indeed, if $i \in \mathcal{A}_j$, then $i \notin \mathcal{B}_j$ since $\mathcal{A}_j$ and $\mathcal{B}_j$ are by definition disjoint; if $i \notin \mathcal{A}_j$ then, since $i \neq j$, we have $i \in ([1:N]\backslash\{j\})\backslash\mathcal{A}_j = \mathcal{B}_j$. Thus, the element i belongs to exactly $(N-1)$ sets among the $2N$ sets in the collection of all $\mathcal{A}_j$, $\mathcal{B}_j$ sets for $j \in [1:N]$. We now claim that $[1:N]\backslash\mathcal{E}_j^{(N)} = \mathcal{F}_{N-j}^{(N)},\ j \in [1:N-1]$. Consider an element $x \in [1:N]$; then:

1. Let $x \in \mathcal{E}_j^{(N)}$, i.e., x appears in at least j sets $\mathcal{A}_i$. Since x appears exactly $(N-1)$ times in $\mathcal{A}_i$ and $\mathcal{B}_i$, this means that x appears in at most $(N-1)-j$ sets $\mathcal{B}_i$, i.e., $x \notin \mathcal{F}_{N-j}^{(N)}$. In other words, $x \in [1:N]\backslash\mathcal{F}_{N-j}^{(N)}$. Since this is true $\forall x \in \mathcal{E}_j^{(N)}$, it implies that $\mathcal{E}_j^{(N)} \subseteq [1:N]\backslash\mathcal{F}_{N-j}^{(N)}$ and as a result $[1:N]\backslash\mathcal{E}_j^{(N)} \supseteq \mathcal{F}_{N-j}^{(N)}$.

2. Let $x \notin \mathcal{E}_j^{(N)}$, i.e., x appears in at most $(j-1)$ sets $\mathcal{A}_i$; since x in total appears exactly

$(N-1)$ times in $\mathcal{A}_i$ and $\mathcal{B}_i$, this means that x appears in at least $(N-1)-(j-1)$ sets $\mathcal{B}_i$, i.e., $x \in \mathcal{F}_{N-j}^{(N)}$. Since this is true $\forall x \in [1:N]\backslash\mathcal{E}_j^{(N)}$, it implies that $[1:N]\backslash\mathcal{E}_j^{(N)} \subseteq \mathcal{F}_{N-j}^{(N)}$.

The points in 1) and 2) imply that $[1:N]\backslash\mathcal{E}_j^{(N)} = \mathcal{F}_{N-j}^{(N)},\ \forall j \in [1:N-1]$. Applying this equality into (3.29), we obtain

$$\sum_{j=1}^{N}\left(\max_{i\in\mathcal{A}_j}\ell_i + \max_{i\in([1:N]\backslash\{j\})\backslash\mathcal{A}_j} r_i\right) \geq \sum_{j=1}^{N-1}\left[f\left(\mathcal{E}_j^{(N)}\right) + g\left(\mathcal{F}_{N-j}^{(N)}\right)\right]$$
$$= \sum_{j=1}^{N-1}\left[f\left(\mathcal{E}_j^{(N)}\right) + g\left([1:N]\backslash\mathcal{E}_j^{(N)}\right)\right] = \sum_{j=1}^{N-1}\left(\max_{i\in\mathcal{A}_{\mathrm{F}_j}}\ell_i + \max_{i\in[1:N]\backslash\mathcal{A}_{\mathrm{F}_j}} r_i\right),$$

where we let $\mathcal{A}_{\mathrm{F}_j} = \mathcal{E}_j^{(N)}$. Since throughout the proof we made no assumptions on the values of $(\ell_i,\ r_i)$, then the sets $\mathcal{A}_{\mathrm{F}_j}$ do not depend on the values of (ℓ_i, r_i). This concludes the proof of Lemma 3.3.3.

3.7.4 Proof of Lemma 3.5.1

Let λ be a schedule (non necessarily optimal) of the full network $\mathcal{N}_{[1:N]}$ with N relays. Denote by $\mathcal{A}_j^\star$ the minimum cut of the network $\bar{\mathcal{N}}_j$ when operated with the 'natural' schedule constructed from λ. Then, by following the same steps as in the example in Section 3.2, from (3.7) we obtain

$$\sum_{i=1}^{N}\mathsf{R}_{\bar{\mathcal{N}}_i}^{\lambda} = \sum_{s\in[0:1]^N}\lambda_s\left[\sum_{j=1}^{N}\left(\max_{i\in\mathcal{A}_j^\star}\ell'_{i,s} + \max_{i\in([1:N]\backslash\{j\})\backslash\mathcal{A}_j^\star} r'_{i,s}\right)\right],$$

where $\ell'_{i,s}$ and $r'_{i,s}$ are defined in (3.8). From the result in Lemma 3.3.3 we know that $\exists\left\{\mathcal{A}_{\mathrm{F}_j}\right\},\ j\in[1:N-1]$, such that for each $s\in[0:1]^N$:

$$\sum_{j=1}^{N}\left(\max_{i\in\mathcal{A}_j^\star}\ell'_{i,s} + \max_{i\in([1:N]\backslash\{j\})\backslash\mathcal{A}_j^\star} r'_{i,s}\right) \geq \sum_{j=1}^{N-1}\left(\max_{i\in\mathcal{A}_{\mathrm{F}_j}}\ell'_{i,s} + \max_{i\in[1:N]\backslash\mathcal{A}_{\mathrm{F}_j}} r'_{i,s}\right),$$

where $\mathcal{A}_{\mathrm{F}_j}\subseteq[1:N],\ \forall j\in[1:N-1]$. Additionally, from Lemma 3.3.3 we have that $\mathcal{A}_{\mathrm{F}_j}$ is independent of $(\ell'_{i,s}, r'_{i,s})$ and is therefore independent of (ℓ_i, r_i) and of the state s. Hence

$$\sum_{i=1}^{N}\mathsf{R}_{\bar{\mathcal{N}}_i}^{\lambda} = \sum_{s\in[0:1]^N}\lambda_s\left[\sum_{j=1}^{N}\left(\max_{i\in\mathcal{A}_j^\star}\ell'_{i,s} + \max_{i\in([1:N]\backslash\{j\})\backslash\mathcal{A}_j^\star} r'_{i,s}\right)\right]$$

$$
\begin{aligned}
&\geq \sum_{s\in[0:1]^N} \lambda_s \left[\sum_{j=1}^{N-1} \left(\max_{i\in\mathcal{A}_{\mathrm{F}_j}} \ell'_{i,s} + \max_{i\in[1:N]\setminus\mathcal{A}_{\mathrm{F}_j}} r'_{i,s}\right)\right] \\
&= \sum_{j=1}^{N-1} \sum_{s\in[0:1]^N} \lambda_s \left(\max_{i\in\mathcal{A}_{\mathrm{F}_j}} \ell'_{i,s} + \max_{i\in[1:N]\setminus\mathcal{A}_{\mathrm{F}_j}} r'_{i,s}\right) \\
&\geq (N-1) \min_{\mathcal{A}\subseteq[1:N]} \left\{\sum_{s\in[0:1]^N} \lambda_s \left(\max_{i\in\mathcal{A}} \ell'_{i,s} + \max_{i\in[1:N]\setminus\mathcal{A}} r'_{i,s}\right)\right\} \\
&= (N-1)\mathsf{R}^{\lambda}_{\mathcal{N}_{[1:N]}}.
\end{aligned}
$$

This completes the proof of Lemma 3.5.1.

3.7.5 Detailed Analysis for the Network in (3.16)

In this section, we analyze in details the network in (3.16). We start by deriving an upper bound and a lower bound on the approximate capacity $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$ for the network described in (3.16) and show they are both equal to one, hence proving $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} = 1$. A trivial upper bound on $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$ is given by $\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}$, i.e., $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq \widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}$. It is not difficult to see that, for the network in (3.16), $\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}} = 1$, which implies $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \leq 1$.

We now derive a lower bound on $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}$. We start by considering even values for N. Let the network in (3.16) operate only in 2 states with the same duration, namely,

$$
\lambda_{\underbrace{00\ldots0}_{\frac{N}{2}}\underbrace{11\ldots1}_{\frac{N}{2}}} = \lambda_{\underbrace{11\ldots1}_{\frac{N}{2}}\underbrace{00\ldots0}_{\frac{N}{2}}} = \frac{1}{2}.
$$

In other words, half of the time the first $\frac{N}{2}$ relays listen, while the remaining $\frac{N}{2}$ relays transmit and half of the time the opposite occurs. Let $R^{\mathrm{E}}_{\mathcal{N}_{[1:N]}}$ be the corresponding approximate rate; clearly we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}} \geq R^{\mathrm{E}}_{\mathcal{N}_{[1:N]}}$. Let $\{\mathcal{M}_1, \mathcal{M}_2\}$ be a partition of $[1:N]$, where $\mathcal{M}_1 = \left[1:\frac{N}{2}\right]$. With this, we have

$$
\begin{aligned}
R^{\mathrm{E}}_{\mathcal{N}_{[1:N]}} &= \min_{\mathcal{A}\subseteq[1:N]} \left\{\frac{1}{2}\max_{i\in\mathcal{A}\cap\mathcal{M}_1} \ell_i + \frac{1}{2}\max_{i\in\mathcal{A}^c\cap\mathcal{M}_2} r_i + \frac{1}{2}\max_{i\in\mathcal{A}\cap\mathcal{M}_2} \ell_i + \frac{1}{2}\max_{i\in\mathcal{A}^c\cap\mathcal{M}_1} r_i\right\} \\
&= \frac{1}{2}\min_{\mathcal{A}\subseteq[1:N]} \left\{\left[\max_{i\in\mathcal{A}\cap\mathcal{M}_1} \ell_i + \max_{i\in\mathcal{A}^c\cap\mathcal{M}_1} r_i\right] + \left[\max_{i\in\mathcal{A}\cap\mathcal{M}_2} \ell_i + \max_{i\in\mathcal{A}^c\cap\mathcal{M}_2} r_i\right]\right\}
\end{aligned}
$$

$$\geq \frac{1}{2}\left[\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap\mathcal{M}_1}\ell_i+\max_{i\in\mathcal{A}^c\cap\mathcal{M}_1}r_i\right\}+\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap\mathcal{M}_2}\ell_i+\max_{i\in\mathcal{A}^c\cap\mathcal{M}_2}r_i\right\}\right]$$
$$=\frac{1}{2}\left(\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}+\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}\right)=1.$$

Hence, for even values of N, we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}\geq 1$, which together with the upper bound $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}\leq 1$, implies $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}=1$. We now consider odd values for N. Let the network in (3.16) operate only in 2 states with the same duration, namely

$$\lambda_{\underbrace{00\ldots0}_{\frac{N-1}{2}}\underbrace{11\ldots1}_{\frac{N-1}{2}}\underbrace{1}_{1}}=\lambda_{\underbrace{11\ldots1}_{\frac{N-1}{2}}\underbrace{00\ldots0}_{\frac{N-1}{2}}\underbrace{1}_{1}}=\frac{1}{2}.$$

In other words, the N-th relay is always transmitting, while half of the time the first $\frac{N-1}{2}$ relays listen, while the remaining $\frac{N-1}{2}$ relays transmit and half of the time the opposite occurs. Let $R^{\mathrm{O}}_{\mathcal{N}_{[1:N]}}$ be the corresponding approximate rate; clearly we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}\geq R^{\mathrm{O}}_{\mathcal{N}_{[1:N]}}$. Let $\mathcal{M}_1=\left[1:\frac{N-1}{2}\right]$ and $\mathcal{M}_2=\left[\frac{N+1}{2}:N-1\right]$. With this, we have

$$\begin{aligned}
R^{\mathrm{O}}_{\mathcal{N}_{[1:N]}}&=\min_{\mathcal{A}\subseteq[1:N]}\left\{\frac{1}{2}\max_{i\in\mathcal{A}\cap\mathcal{M}_1}\ell_i+\frac{1}{2}\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_2\cup\{N\})}r_i+\frac{1}{2}\max_{i\in\mathcal{A}\cap\mathcal{M}_2}\ell_i+\frac{1}{2}\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_1\cup\{N\})}r_i\right\}\\
&=\frac{1}{2}\min_{\mathcal{A}\subseteq[1:N]}\left\{\left[\max_{i\in\mathcal{A}\cap\mathcal{M}_1}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_1\cup\{N\})}r_i\right]+\left[\max_{i\in\mathcal{A}\cap\mathcal{M}_2}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_2\cup\{N\})}r_i\right]\right\}\\
&\geq\frac{1}{2}\left[\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap\mathcal{M}_1}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_1\cup\{N\})}r_i\right\}+\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap\mathcal{M}_2}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_2\cup\{N\})}r_i\right\}\right]\\
&\overset{\text{(a)}}{=}\frac{1}{2}\left[\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap(\mathcal{M}_1\cup\{N\})}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_1\cup\{N\})}r_i\right\}\right.\\
&\quad\left.+\min_{\mathcal{A}\subseteq[1:N]}\left\{\max_{i\in\mathcal{A}\cap(\mathcal{M}_2\cup\{N\})}\ell_i+\max_{i\in\mathcal{A}^c\cap(\mathcal{M}_2\cup\{N\})}r_i\right\}\right]=\frac{1}{2}\left(\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}+\widetilde{\mathsf{C}}^{\mathrm{FD}}_{\mathcal{N}_{[1:N]}}\right)=1,
\end{aligned}$$

where the equality in (a) follows since the N-th relay is never in the minimum cut $\mathcal{A}$ as otherwise the approximate capacity would be infinity (since from (3.16) we have $\ell_N\to\infty$). Hence, also for odd values of N we have $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}\geq 1$, which together with the upper bound $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}\leq 1$, implies $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}=1$. This concludes the proof that $\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}}=1$ for the network in (3.16).

Now, assume that $N=4t-2$, where $t\in\mathbb{N}\backslash\{0\}$ and with this suppose we want to select the best subnetwork $\mathcal{N}_{\mathcal{K}}$ with $|\mathcal{K}|=1$ in the network $\mathcal{N}_{[1:N]}$ in (3.16), i.e., we want to select

the best relay. From (3.6) we obtain that the single-relay approximate capacity of the i-th relay with $i \in \left[1 : \left\lfloor \frac{N}{2} \right\rfloor\right]$ is given by

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{\{\lfloor N/2 \rfloor + i\}}} = \frac{\ell_i r_i}{\ell_i + r_i} = \frac{2i\,(N - 2i + 2)}{N\,(N+2)}, \tag{3.30a}$$

$$\text{if } N \text{ is odd: } \widetilde{\mathsf{C}}_{\mathcal{N}_{\{N\}}} = \frac{1}{N}. \tag{3.30b}$$

It is not difficult to see that the expression of $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}}$ in (3.30) achieves its maximum value for

$$i^\star = \frac{N+2}{4}, \tag{3.31}$$

for which

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}} = \frac{2\frac{N+2}{4}\left(N - 2\frac{N+2}{4} + 2\right)}{N\,(N+2)} = \frac{\frac{N+2}{2}\left(\frac{2N+4}{4}\right)}{N(N+2)} = \frac{N+2}{4N} = \frac{t}{4t-2}, \tag{3.32}$$

which for $t \to \infty$ gives

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}} = \frac{1}{4} \implies \widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}} = \frac{1}{4}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}},\ |\mathcal{K}| = 1.$$

Now, for the same network, suppose we want to select the best subnetwork $\mathcal{N}_{\mathcal{K}}$ with $|\mathcal{K}| = 2$, i.e., we want to select the best 2-relay subnetwork. Clearly from the partition lemma, if we select relays number $i \in [1 : N]$ and $j \in [1 : N]$ with $i \neq j$ a trivial upper bound on the approximate capacity $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i,j\}}}$ is given by

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i,j\}}} \leq \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i\}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{j\}}} \leq 2\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}}.$$

Consider relays number $i^\star$ and $j^\star = i^\star + \frac{N}{2}$, where $i^\star$ is defined in (3.31). By substituting $j^\star$ into (3.16) we obtain

$$\ell_{i^\star} = \ell_{j^\star} = r_{i^\star} = r_{j^\star} = \frac{2^{\frac{N+2}{4}}}{N},$$

which implies $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{\{j^\star\}}}$, where $\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}}$ is defined in (3.32) and from [BMK14] we have

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star, j^\star\}}} = \widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}} + \widetilde{\mathsf{C}}_{\mathcal{N}_{\{j^\star\}}} = 2\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star\}}} = \frac{t}{2t-1}, \tag{3.33}$$

which for $t \to \infty$ gives

$$\widetilde{\mathsf{C}}_{\mathcal{N}_{\{i^\star,j^\star\}}}=\frac{1}{2} \implies \widetilde{\mathsf{C}}_{\mathcal{N}_{\mathcal{K}}}=\frac{1}{2}\widetilde{\mathsf{C}}_{\mathcal{N}_{[1:N]}},\ |\mathcal{K}|=2.$$

So, the network in (3.16), for $N = 4t - 2$, where $t \in \mathbb{N}\backslash\{0\}$, represents an example for the network described in the statement of Theorem 3.6.1. This concludes the proof of Theorem 3.6.1.

CHAPTER 4

The Approximate Capacity of Half-Duplex Line Networks

In this chapter we focus on the capacity of Half-Duplex (HD) line networks (routes), and the complexity of finding the route with the best capacity that connects a source and a destination within a larger relay network. First, we show that the approximate capacity (optimal up to a constant additive gap that only depends on the number of nodes in the network) of an HD N-relay line network equals half the minimum of the harmonic means of the point-to-point link capacities of each two consecutive links in the path. We then prove that the $N+1$ listen/transmit states (out of the 2^N possible ones) sufficient to characterize the approximate capacity can be found in linear time. In the second part of the chapter, we show that the problem of finding the path that has the largest HD approximate capacity in a network that can be represented as a graph is NP-hard. We show, however, that if the number of cycles in the network is polynomial in the number of nodes, then a polynomial time algorithm can indeed be designed.

4.1 Introduction

In recent years, promising advances have been made in designing Full-Duplex (FD) transceivers [DSA14, ESZ16]. However, the proposed FD designs still require complex self-interference cancellation techniques. Due to this, in the near future it is envisioned that nodes will continue to operate in Half-Duplex (HD) mode in order to enable low-cost communica-

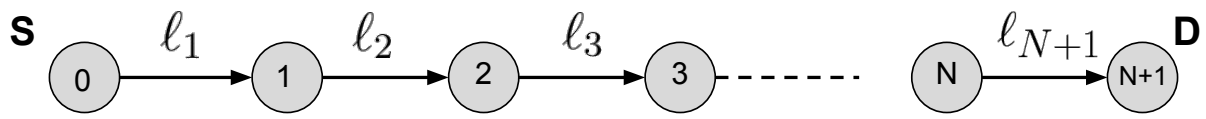


Figure 4.1: The line network $\mathcal{R}$ with N relays, source S and destination D.

tions – as recently announced, for example, in 3GPP Rel-13 [WLA17]. A widespread approach to route information from a source node to a destination node in an HD network is to find the path with the largest FD capacity and then operate the path in HD mode [AHR04, DAB05, BMJ98]. This approach is used because of the simple nature of the FD capacity expression, which is given by the minimum of the point-to-point link capacities in the path, and thus can be distributively computed in linear time. However, selecting a route based on its FD capacity may be suboptimal if then the nodes in the selected path are operated in HD mode.

In this chapter we investigate the problem of routing in HD networks. We address the three following fundamental questions: (i) Can a closed-form expression of the capacity (or of an approximation of it) of an N-relay HD line network be derived with the same promising features of the FD counterpart? (ii) Can the linear number of active listen/transmit configuration states sufficient to characterize the HD capacity (or an approximation of it) be found efficiently in polynomial time? (iii) Does there exist a low-complexity algorithm that finds the route in a network with the largest HD capacity (or an approximation of it)?

Our main contributions can be summarized as follows:

1. We derive the HD approximate capacity of the N-relay line network $\mathcal{R}$ (shown in Fig. 4.1) in closed form and show it is given by half of the minimum of the harmonic means of the point-to-point capacities of each two consecutive links in the path, that is

$$\mathsf{C}_{\mathcal{R}} = \min_{i\in[1:N]} \left\{ \frac{\ell_i \ell_{i+1}}{\ell_i + \ell_{i+1}} \right\}, \tag{4.1}$$

where ℓ_i is the point-to-point capacity of the link between node $i-1$ and node i. This approximate capacity expression has the same appealing features of the FD counterpart, i.e., it can be evaluated in linear time and distributively computed among the nodes in the path. To the best of our knowledge, this represents the first approximate capacity characterization in closed form for a class of HD relay networks with general number of relays.

2. We prove that, with only the knowledge of the network topology (i.e., that the N relays are arranged in a line), the cardinality of the smallest search space of states over which a schedule that achieves the approximate capacity can be found is exponential in N. In other words, to reduce the cardinality of this search space to be polynomial in the number of relays, it is crucial to leverage the strength of the channel parameters.

3. We design an algorithm that allows to compute a simple schedule (i.e., with at most $N+1$ active states) that achieves the approximate capacity of the N-relay HD line network with complexity $O(N)$. This result sheds light on how to operate a class of HD relay networks close to the capacity with the minimum number of state switches. Moreover, to the best of our knowledge, this is the first result that provides an efficient way to find a simple schedule optimal for approximate capacity.

4. We prove that the problem of finding the route with the largest HD approximate capacity in a relay network is NP-hard in general. Our proof is based on a reduction from the 3SAT problem [Kar72], which is a special case of the well-known SAT problem, where the goal is to determine the satisfiability of a Boolean formula. The NP hardness of HD routing represents a surprising difference from FD, for which polynomial time algorithms exist to discover the path with the largest FD capacity, such as Dijkstra's algorithm [Dij59]. Intuitively, the NP-hardness in HD routing stems from the necessity to avoid cycles in the network while discovering an HD path, which is not necessary when finding a FD path.

Table 4.1: Quantities of interest used throughout Chapter 4.

Quantity	**Definition**
$\mathcal{G}$	Digraph representing a relay network
$\mathcal{R}$	Line network
$\mathcal{L}_{\mathcal{G}}$	Line digraph of the digraph $\mathcal{G}$
$\mathcal{P}$	Path between two nodes
$\ell_{i,j}$	Point-to-point link capacity from node v_i to node v_j in $\mathcal{G}$
ℓ_i	Point-to-point link capacity from node v_{i-1} to node v_i in $\mathcal{R}$
$\mathsf{C}_{\mathcal{R}}$ (resp. $\mathsf{C}_{\mathcal{P}}$)	Approximate HD capacity of $\mathcal{R}$ (resp. $\mathcal{P}$)
$\mathsf{C}_{\mathcal{R}}^{\lambda}$	Achievable HD rate for $\mathcal{R}$ with deterministic schedule λ
$\mathsf{C}_{\mathcal{R}}^{\mathrm{FD}}$ (resp. $\mathsf{C}_{\mathcal{P}}^{\mathrm{FD}}$)	FD capacity of $\mathcal{R}$ (resp. $\mathcal{P}$)

5. We show that, if the number of cycles in the network is polynomial in the total number of nodes, then a polynomial time algorithm that discovers the path with the largest HD approximate capacity can be designed. Thus, this represents a sufficient condition for which HD routing can be efficiently solved. A relevant class of relay networks for which this holds is the one of layered networks where the relays are arranged over L layers of relays and a relay in a layer can only communicate to the relays in the next layer.

Chapter Organization. Section 4.2 illustrates the setting of our problem, describes known capacity results for HD line networks and simplifies the approximate capacity expression for HD line networks. Section 4.3 presents our main results and discusses their implications. Section 4.4 proves the NP-hardness of finding the route with the largest HD approximate capacity in a relay network. Section 4.5 describes special network classes for which a polynomial time algorithm for finding the route with the largest HD approximate capacity exists. Section 4.6 concludes the chapter. Some of the proofs are delegated to the Appendix.

4.2 Half-Duplex Network Model

Table 4.1 summarizes and defines quantities that are used throughout the chapter.

We consider an HD relay network represented by the directed graph $\mathcal{G}$ where $\mathcal{V}(\mathcal{G})$ and $\mathcal{E}(\mathcal{G})$ are the set of vertices (communication nodes) and the set of edges (point-to-point links) in $\mathcal{G}$, respectively. The point-to-point links between nodes in the network are assumed to be non-interfering discrete memoryless channels. An edge connecting vertex v_i to vertex v_j where $v_i, v_j \in \mathcal{V}(\mathcal{G})$ is denoted by $e_{i,j}$. For each edge $e_{i,j} \in \mathcal{E}(\mathcal{G})$, we represent its point-to-point link capacity with $\ell_{i,j} > 0$. Over the graph $\mathcal{G}$ with $N+2$ vertices, information flows from a source node $S \in \mathcal{V}(\mathcal{G})$ (denoted by v_0) to a destination node $D \in \mathcal{V}(\mathcal{G})$ (denoted by v_{N+1}) with the help of the remaining N relay nodes. Each node in $\mathcal{G}$ operates in HD, i.e., it cannot transmit and receive simultaneously.

A relay network is called a *line network* if its vertices are arranged in a path (or a route) forming a cascade of non-interfering discrete memoryless channels[1]. The input/output relation for the line network (denoted by $\mathcal{R}$) with N relays can therefore be defined through the conditional distribution

$$p\left(Y_1,\ldots,Y_{N+1}\middle|(X_0,S_0),(X_1,S_1),\ldots,(X_N,S_N),S_{N+1}\right)=\prod_{i=0}^{N}p\left(Y_{i+1}\middle|(X_i,S_i),S_{i+1}\right), \quad (4.2)$$

where: (i) X_i (respectively, Y_i) denotes the channel input (respectively, output) at node v_i; (ii) S_i is the binary random variable which represents the state of node v_i, i.e., if $S_i = 0$ then node v_i is receiving, while if $S_i = 1$ then node v_i is transmitting; notice that $S_0 = 1$ (i.e., the source always transmits) and $S_{N+1} = 0$ (i.e., the destination always receives). The line network in (4.2) is a cascade of N discrete memoryless channels and as a result is physically degraded [Are81]. Thus, the capacity of the line network $\mathcal{R}$ is given by the cut-set bound that can be achieved by decoding transmissions from node i at node $i+1$ before encoding them for

[1]A line network consists of a cascade of noisy channels, which make it a physically degraded channel. Therefore, a node has a "cleaner" view of information with respect to a node that is next in the line and thus, replacing directed edges with undirected ones does not increase the capacity.

transmission further in the network as would typically happen with routing. This is exactly how the standard decode-and-forward relaying scheme operates over the line network. In particular, the capacity is given by the cut-set bound as

$$\mathsf{C}_{\mathcal{R}}^{(\mathrm{cs})} = \max_{p(X_0,\{X_i,S_i\}_{i=1}^N)} \min_{\mathcal{A}\subseteq[1:N]} I(Y_{N+1},\{Y_i\}_{i\in\mathcal{A}}; X_0,\{X_i,S_i\}_{i\in\mathcal{A}^c}|\{X_i,S_i\}_{i\in\mathcal{A}}), \tag{4.3}$$

where $\mathcal{A}$ represents the set of relays on the destination side of the cut, and $\mathcal{A}^c = [1{:}N]\backslash\mathcal{A}$. However, it is not clear what is the optimal distribution of $\{(X_i,S_i)\}_{i=0}^N$ needed to characterize the capacity of the HD line network $\mathcal{R}$ in (4.3). The capacity of the HD line network $\mathcal{R}$ described in (4.2) can however be approximated to within a constant gap $\mathrm{GAP} = N$ by using deterministic schedules. In particular, to obtain this constant gap approximation we upper bound the cut-set bound in (4.3) as

$$\begin{aligned}
\mathsf{C}_{\mathcal{R}}^{(\mathrm{cs})} &= \max_{p(X_0,\{X_i,S_i\}_{i=1}^N)} \min_{\mathcal{A}\subseteq[1:N]} I(Y_{N+1},\{Y_i\}_{i\in\mathcal{A}}; X_0,\{X_i,S_i\}_{i\in\mathcal{A}^c}|\{X_i,S_i\}_{i\in\mathcal{A}}) \\
&\overset{(a)}{\leq} \max_{p(X_0,\{X_i,S_i\}_{i=1}^N)} \min_{\mathcal{A}\subseteq[1:N]} I(Y_{N+1},\{Y_i\}_{i\in\mathcal{A}}; X_0,\{X_i\}_{i\in\mathcal{A}^c}|\{X_i\}_{i\in\mathcal{A}},\{S_i\}_{i=1}^N) + H(\{S_i\}_{i=1}^N) \\
&\overset{(b)}{\leq} \underbrace{\max_{p(\{X_i\}_{i=0}^N|\{S_i\})p(\{S_i\}_{i=1}^N))} \min_{\mathcal{A}\subseteq[1:N]} I(Y_{N+1},\{Y_i\}_{i\in\mathcal{A}}; X_0,\{X_i\}_{i\in\mathcal{A}^c}|\{X_i\}_{i\in\mathcal{A}},\{S_i\}_{i=1}^N)}_{\mathsf{C}_{\mathcal{R}}} + \underbrace{N}_{\mathrm{GAP}},
\end{aligned} \tag{4.4}$$

where: (a) follows from the chain rule of the mutual information and by the fact that, for a discrete random variable, the entropy is a non-negative quantity (the mutual information can hence be upper bounded by the entropy); and (b) follows by upper bounding the entropy of a discrete random variable by the logarithm of its support.

This constant gap capacity approximation, which is derived using similar arguments as in [ADT11, OD13, CTK14] for Gaussian HD relay networks, follows since a binary random variable (i.e., a relay state) can only improve the capacity by at most 1 bit. Hence, since we have N relays, the gap is at most equal to N. The first term $\mathsf{C}_{\mathcal{R}}$ in (4.4) which uses fixed schedules (i.e., the exact values of $\{S_i\}$ are in the conditioning of the mutual information term) is what we refer to as the *approximate capacity*. By using decode-and-forward with an

optimal product input distribution and deterministic schedules, the approximate capacity can be achieved and is expressed by

$$\mathsf{C}_{\mathcal{R}} = \max_{\lambda \in \Lambda} \min_{\mathcal{A} \subseteq [1:N]} \sum_{s \in [0:1]^N} \lambda_s \sum_{\substack{i \in \{N+1\} \cup \{\mathcal{T}_s^c \cap \mathcal{A}\} \\ i-1 \in \{0\} \cup \{\mathcal{T}_s \cap \mathcal{A}^c\}}} \ell_i, \tag{4.5}$$

where: (i) the schedule $\lambda \in \mathbb{R}^{2^N}$ determines the fraction of time the network operates in each of the states $s \in [0:1]^N$, i.e., $\lambda_s = \Pr\left(\{S_i\}_{i=1}^N = s\right)$; (ii) $\Lambda = \{\lambda : \lambda \in \mathbb{R}^{2^N},\ \lambda \geq 0,\ \sum_{s \in [0:1]^N} \lambda_s = 1\}$ is the set of all possible schedules; (iii) $\mathcal{T}_s$ (respectively, $\mathcal{T}_s^c = [1:N] \backslash \mathcal{T}_s$) represents the set of indices of relays transmitting (respectively, receiving) in the state $s \in [0:1]^N$; (iv) for ease of notation, we set $\ell_i = \ell_{i-1,i}$ to denote the point-to-point capacity of the link from v_{i-1} to v_i. We can equivalently write the expression in (4.5) as

$$\mathsf{C}_{\mathcal{R}} = \max_{\lambda \in \Lambda} \min_{\mathcal{A} \subseteq [1:N]} \sum_{s \in [0:1]^N} \lambda_s \sum_{\substack{i \in \{N+1\} \cup \mathcal{A} \\ i-1 \in \{0\} \cup \mathcal{A}^c}} \hat{\ell}_i^{(s)}, \tag{4.6}$$

where

$$\hat{\ell}_i^{(s)} := \begin{cases} \ell_i, & \text{if } i \in \mathcal{T}_s^c \cup \{N+1\} \text{ and } i-1 \in \mathcal{T}_s \cup \{0\} \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

Similarly, we denote with $\mathsf{C}_{\mathcal{R}}^{\lambda}$, the HD rate achieved by the line network $\mathcal{R}$ when operated with the deterministic schedule λ, i.e.,

$$\mathsf{C}_{\mathcal{R}}^{\lambda} = \min_{\mathcal{A} \subseteq [1:N]} \sum_{s \in [0:1]^N} \lambda_s \sum_{\substack{i \in \{N+1\} \cup \mathcal{A} \\ i-1 \in \{0\} \cup \mathcal{A}^c}} \hat{\ell}_i^{(s)}. \tag{4.8}$$

Note that, for all possible schedules λ, $\mathsf{C}_{\mathcal{R}}^{\lambda} \leq \mathsf{C}_{\mathcal{R}}$.

Definition 4.2.1 (Simple Schedule)**.** We say that a schedule $\lambda \in \mathbb{R}^{2^N}$ is *simple* if the number of active states, i.e., states s such that $\lambda_s > 0$ is at most $N+1$. In other words, λ is simple if $\|\lambda\|_0 \leq N+1$ (with $\|\lambda\|_0$ being the L0 norm of the vector λ). In [CTK16, Theorem 1], it was shown that for any Gaussian HD relay network with arbitrary topology, there always exists a simple schedule that is optimal for the approximate capacity.

4.2.1 Fundamental Cuts in HD Line Networks

In this subsection, we prove that for the HD line network in (4.2), we can compute $\mathsf{C}_{\mathcal{R}}$ in (4.6) by considering only $N+1$ cuts (out of the 2^N possible ones), which are the same that one would need to consider if the network was operating in FD.

For the line network $\mathcal{R}$, when all the N relays operate in FD, the FD capacity is given by

$$\mathsf{C}_{\mathcal{R}}^{\mathrm{FD}} = \min_{\mathcal{A}\subseteq[1:N]} \sum_{\substack{i\in\{N+1\}\cup\mathcal{A},\\ i-1\in\{0\}\cup\mathcal{A}^c}} \ell_i = \min_{i\in[1:N+1]} \{\ell_i\}\,, \tag{4.9}$$

that is, without *explicit* knowledge of the values of ℓ_i or their ordering, the number of cuts over which we need to optimize (see $\mathsf{C}_{\mathcal{R}}^{\mathrm{FD}}$ in (4.9)) is $N+1$. We refer to these cuts as *fundamental*. When states or cuts are referred to as fundamental of a certain type (e.g., maximum, minimum), we mean that they form the smallest set of that type that only depends on the network topology (i.e., relays are arranged in a line) and is independent of the actual values of the point-to-point link capacities. Let $\mathscr{F}$ denote the set of these fundamental cuts (which are of the form $\mathcal{A} = [i:N], i \in [1:N]$ or $\mathcal{A} = \emptyset$). For any cut $\mathcal{A}$ of the network

$$\sum_{\substack{i\in\{N+1\}\cup F(\mathcal{A}),\\ i-1\in\{0\}\cup F(\mathcal{A})^c}} \ell_i \le \sum_{\substack{i\in\{N+1\}\cup\mathcal{A},\\ i-1\in\{0\}\cup\mathcal{A}^c}} \ell_i \quad \text{for some } F(\mathcal{A}) \in \mathscr{F}. \tag{4.10}$$

Furthermore, the function $F(\cdot)$ in (4.10) does not depend on the values of ℓ_i.

We next prove that the fundamental cuts in HD equal those in (4.9) for FD. Consider a deterministic schedule λ. Then, by using (4.10) for the inner summation in (4.8), for each $s \in [0:1]^N$ we have

$$\sum_{\substack{i\in\{N+1\}\cup F(\mathcal{A}),\\ i-1\in\{0\}\cup F(\mathcal{A})^c}} \hat{\ell}_i^{(s)} \le \sum_{\substack{i\in\{N+1\}\cup\mathcal{A},\\ i-1\in\{0\}\cup\mathcal{A}^c}} \hat{\ell}_i^{(s)}. \tag{4.11}$$

Thus, we can simplify (4.8) as

$$\begin{aligned}\mathsf{C}_{\mathcal{R}}^{\lambda} &= \min_{\mathcal{A}\subseteq[1:N]} \sum_{s\in[0:1]^N} \lambda_s \sum_{\substack{i\in\{N+1\}\cup\mathcal{A}\\ i-1\in\{0\}\cup\mathcal{A}^c}} \hat{\ell}_i^{(s)} \\ &= \min_{\mathcal{A}\in\mathscr{F}} \sum_{s\in[0:1]^N} \lambda_s \sum_{\substack{i\in\{N+1\}\cup\mathcal{A}\\ i-1\in\{0\}\cup\mathcal{A}^c}} \hat{\ell}_i^{(s)} = \min_{i\in[1:N+1]} \left(\sum_{s\in\mathcal{S}_i} \lambda_s\right) \ell_i, \end{aligned} \tag{4.12}$$

where

$$\mathcal{S}_i = \{s \in [0:1]^N | i \in \{N+1\} \cup \mathcal{T}_s^c,\ i-1 \in \{0\} \cup \mathcal{T}_s\}. \tag{4.13}$$

The set $\mathcal{S}_i \subseteq [0:1]^N$ represents the collection of states that activate the i-th link. For illustration, for a network with $N = 3$ we have

$$\begin{aligned}\mathcal{S}_1 &= \{000,\ 001,\ 010,\ 011\},\\ \mathcal{S}_2 &= \{100,\ 101\},\\ \mathcal{S}_3 &= \{010,\ 110\},\\ \mathcal{S}_4 &= \{001,\ 011,\ 101,\ 111\}.\end{aligned}$$

Using the same arguments as in (4.12), we can similarly simplify the expression of $\mathsf{C}_{\mathcal{R}}$ in (4.6). Thus, the result presented in this section explicitly provides the $N+1$ cuts (out of the 2^N possible ones) over which it is sufficient to minimize in order to obtain $\mathsf{C}_{\mathcal{R}}$ in (4.6).

4.3 Main Results and Discussion

In this section, we present our main results and discuss their implications. Our first main result, stated in Theorem 4.3.1 is two-fold: (i) it provides a closed-form expression for the approximate capacity of the HD line network that can be evaluated in linear time, and (ii) it shows the existence of a polynomial time algorithm that outputs a simple schedule optimal for approximate capacity.

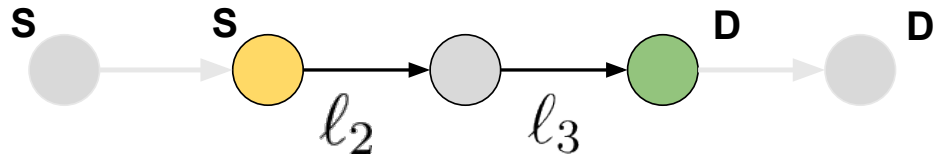


Figure 4.2: Upper bound on $\mathsf{C}_{\mathcal{R}}$ for $i = 2$.

Theorem 4.3.1. *For the N-relay HD line network $\mathcal{R}$ described in (4.2), a simple schedule (i.e., with at most $N+1$ active states) optimal for approximate capacity can be obtained in $O(N)$ time and the approximate capacity $\mathsf{C}_{\mathcal{R}}$ in (4.6) is given by (4.1).*

Proof. It is not difficult to argue that the right-hand side of (4.1) is an upper bound on $\mathsf{C}_{\mathcal{R}}$. This can be seen by assuming that, for a given $i \in [1:N]$, node v_{i-1} perfectly cooperates with node v_0 and node v_{i+1} perfectly cooperates with node v_{N+1} (see also Fig. 4.2 for an illustrative example with $i = 2$ and $N = 3$). Clearly, the HD approximate capacity of this new line network (equivalent to a single relay line network) is an upper bound on $\mathsf{C}_{\mathcal{R}}$ and is given by $\max\limits_{0\leq\beta\leq 1} \min\{(1-t_i)\ell_i\ , t_i\ell_{i+1}\} = \frac{\ell_i\ \ell_{i+1}}{\ell_i+\ell_{i+1}}$, which is achieved by setting $t_i = \frac{\ell_i}{\ell_i+\ell_{i+1}}$. Since this is true for all $i \in [1:N]$, then $\mathsf{C}_{\mathcal{R}}$ is less than or equal to the right-hand side of (4.1).

To prove the achievability of (4.1), we assign a duration of time – denoted as TX_i – for each relay $i \in [1:N]$ to be transmitting (and hence listening in the remaining time). The transmit period TX_i assigned to relay i is parameterized by t_i and is given by the following period assignment

$$\mathrm{TX}_i = \begin{cases} [0, t_i], & i \text{ is even}, \\ [1-t_i, 1], & i \text{ is odd}, \end{cases} \tag{4.14}$$

where $t_i = \dfrac{\ell_i}{\ell_i + \ell_{i+1}}, \forall i \in [1:N]$. We denote the time spent by relay i listening as $\mathrm{RX}_i = [0,1]\backslash\mathrm{TX}_i$.

It is not difficult to see that, with this time allocation, the network changes its state in at most N points in time given by the values $t_i, \forall i \in [1:N]$. Thus the proposed schedule has at most $N+1$ states. Furthermore, the schedule can be created in $O(N)$. What remains to show is that the proposed schedule achieves the rate given in (4.1). Towards this end, we need to compute the duration of time $\gamma_i, i \in [1:N]$, for which the link of capacity ℓ_i is active. This can be computed as follows

$$\gamma_i = |\mathrm{TX}_{i-1} \cap \mathrm{RX}_i| \overset{(a)}{=} \begin{cases} \frac{\ell_2}{\ell_1+\ell_2}, & i = 1, \\ \min\left(\frac{\ell_{i+1}}{\ell_i+\ell_{i+1}}, \frac{\ell_{i-1}}{\ell_i+\ell_{i-1}}\right), & i \in [2:N], \\ \frac{\ell_N}{\ell_{N+1}+\ell_N}, & i = N+1, \end{cases} \tag{4.15}$$

where (a) follows by: (i) computing the size of the intersection $\mathrm{TX}_{i-1} \cap \mathrm{RX}_i$, (ii) using the ranges given in (4.14), and (iii) the fact that the source is always transmitting and the destination is always listening. The rate achieved by the given schedule is hence equal to $\mathsf{C}^{\lambda}_{\mathcal{R}} = \min_{i=[1:N+1]}\{\gamma_i \ell_i\}$, which gives the result in (4.1). This concludes the proof of Theorem 4.3.1. □

Remark 4.3.1. Although the described achievable scheme used for the proof of Theorem 4.3.1 yields a rate equal to the approximate capacity in (4.1), it can be over-using non-bottleneck links in the network. In particular, as described in the proof of Theorem 4.3.1, the i-th link has an effective scheduled rate of $\min\{\frac{\ell_i \ell_{i-1}}{\ell_i+\ell_{i-1}}, \frac{\ell_i \ell_{i+1}}{\ell_i+\ell_{i+1}}\}$, i.e., the effective used rate can be strictly greater than the network approximate capacity. In [ECF17], we proposed an alternative approach for scheduling the relays which is based on edge-coloring and we proved that it also achieves the approximate capacity in (4.1) by using the link of capacity ℓ_j with an effective scheduled rate of $\min_{j\in[1:N]}\{\frac{\ell_j \ell_{j+1}}{\ell_j+\ell_{j+1}}\}$. In other words, the edge-coloring algorithm proposed in [ECF17] ensures that each link is active only for a duration of time that suffices to achieve the approximate capacity. This property of edge-coloring has been recently leveraged to develop queue-aware scheduling schemes [SC20] (assuming that the queues at the nodes are not infinite) that achieve rates that approach the approximate capacity. □

In what follows, we highlight some remarks to motivate the need to search for a simple schedule for the line network and to explain why our search and schedule presented in the proof of Theorem 4.3.1 cannot be simplified a priori.

Remark 4.3.2. [Are two active states sufficient for approximate capacity characterization?] Consider a line network with one relay. For this network, the schedule that achieves the approximate capacity has only two active states, which activate the links alternatively. Intuitively, one might think that this would extend to arbitrary number of relays. For example, for a network with $N = 3$, can we achieve the approximate capacity by only considering the listen/transmit states $s_1 = 010$ and $s_2 = 101$? The answer to this question is negative as we illustrate through the following example with $N = 3$ and

$$\ell_1 = 2r,\ \ell_2 = 2r,\ \ell_3 = 3r,\ \ell_4 = r, \tag{4.16}$$

where $r > 0$. By considering only the two aforementioned states, we can achieve a rate of $r\frac{2}{3}$. However, by applying the expression in (4.1), we get that the HD approximate capacity for this network is $r\frac{3}{4}$. □

Remark 4.3.3. [Can we a priori limit our search over a polynomial number of states?] For the FD line network, we can a priori limit our search for the minimum cut over $N + 1$ cuts (instead of 2^N). This reduction in the number of cuts is also possible for the HD line network as we proved in Section 4.2.1. This fact raises the question whether we can also a priori reduce the search space for the active states to a polynomial set (instead of 2^N). This is not possible as we state in the theorem below, which is proved in Appendix 4.8.1. This result might be due to the fact that the capacity expression in (4.1) depends on the harmonic mean between two consecutive links. Hence, different from FD, changing the order of the point-to-point link capacities, might also change the value of the approximate capacity. □

Theorem 4.3.2. *With only the knowledge that N relays are arranged in a line, the cardinality of the smallest search space of states over which a schedule optimal for approximate capacity can be found is $\Omega(2^{N/3})$.*

Remark 4.3.4. Theorem 4.3.1 has two promising consequences:

1. The HD approximate capacity of the N-relay line network can be computed in $O(N)$ time. This improves on the result in [EPS14], where the approximate capacity can be found in polynomial time (but not linear in the worst case) by solving a linear program with $O(N)$ variables.

2. The HD approximate capacity in Theorem 4.3.1 can be computed in a distributive way as follows. Each relay $i \in [1:N]$ computes the quantity

$$m_i = \min\left\{\frac{\ell_i\ \ell_{i+1}}{\ell_i + \ell_{i+1}}, m_{i-1}\right\},$$

where $m_0 = \infty$, and sends it to relay $i+1$. With this, at the end we have $m_N = \mathsf{C}_{\mathcal{R}}$. In other words, for HD approximate capacity computation, it is only required that each relay knows the capacity of its incoming and outgoing links. □

Remark 4.3.5. The properties discussed in Remark 4.3.4 are the same appealing properties that advocate for the use of the FD capacity in routing protocols. Thus, it is interesting to understand whether routing based on the FD capacities would also give the path with the largest HD approximate capacity or instead routing using the HD approximate capacity expression in (4.1) would yield different routes. Indeed, it turns out that the FD capacity approach is suboptimal as shown by the example in Fig. 4.3. By applying the expression in (4.1), we find that the best HD route (within the blue box in Fig. 4.3) has an HD approximate capacity of 13.04, which is 30% higher than the HD approximate capacity of the best FD route (within the red box in Fig. 4.3, with FD capacity of 20 but HD approximate capacity of 10). With best FD (respectively, HD) route we refer to the path that has the largest FD (respectively, HD approximate) capacity. □

The example illustrated in Fig. 4.3 shows that in general it is suboptimal to find the best HD path by using as optimization metric the FD capacity of the path. In fact, as shown in [ECF19b], there exist networks for which routing based on the FD capacities yields a

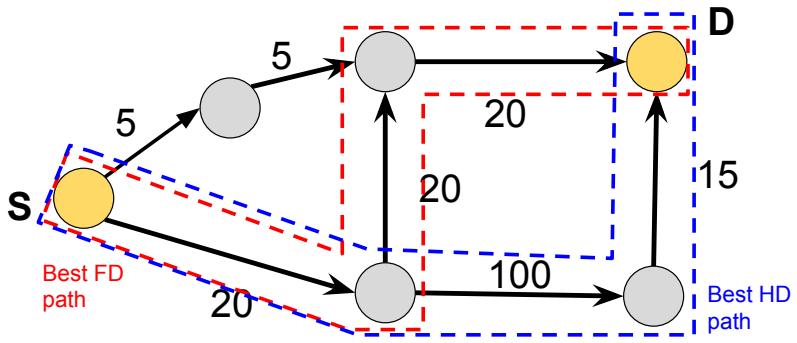


Figure 4.3: Example where the best FD and HD paths are different. Edge labels represent the point-to-point link capacities of the edges.

route with HD approximate capacity equal to half that of the best HD route. This observation naturally suggests the question: **Does there exist an efficient (polynomial time) algorithm that finds the route in a network with the largest HD approximate capacity?** We address this question in the following theorem.

Theorem 4.3.3. *For a relay network in the class described in Section 4.2, the problem of finding the best HD path is NP-hard.*

Intuitively, we can attribute the hardness stated in Theorem 4.3.3 to the fact that we need to keep track of whether the discovered paths contain cycles or not, unlike the FD counterpart (a more detailed discussion regarding this aspect can be found in Section 4.4). This observation suggests that, if the number of cycles in a network is regulated, then we can find the simple path (i.e., a path that contains no cycles) with the largest HD approximate capacity in polynomial time, as formalized in the lemma below.

Lemma 4.3.1. *If the number of cycles in a relay network with N relays (described by the digraph $\mathcal{G}$) is at most polynomial in N (i.e., $O(N^\alpha)$ for some constant α), then we can find the simple path with the largest HD approximate capacity in polynomial time, particularly in $O((N^\alpha+1)(|\mathcal{E}(\mathcal{G})|\log|\mathcal{E}(\mathcal{G})|+|\mathcal{E}(\mathcal{G})|d))$, where d is the maximum vertex degree in $\mathcal{G}$. This*

holds even when we do not have an a priori knowledge of the location of the cycles in the network.

As a network example for which Lemma 4.3.1 applies, we can study the layered network where the relays are arranged as M relays per layer over L layers of relays (in total, we have $N = ML$ relays). Every relay can only communicate with the relays in the following layer. It is not difficult to see that for this particular network, the number of cycles in the graph is equal to zero, i.e., $N^{\alpha} = 0$. In addition, the maximum degree d of a vertex is $O(M)$ and the number of edges in the network is $\Theta(LM^2)$. By substituting these values in the expression in Lemma 4.3.1, we get that the complexity of finding a simple path with the largest HD approximate capacity in a layered network is given by

$$\begin{aligned} &O((N^{\alpha}+1)(|\mathcal{E}(\mathcal{G})|\log|\mathcal{E}(\mathcal{G})| + |\mathcal{E}(\mathcal{G})|d)) \\ &\quad = O(LM^2 \log LM^2 + LM^2 M) \\ &\quad = O(LM^2 \log L + 2LM^2 \log M + LM^3) \\ &\quad = O(LM^2 \log L + LM^3). \end{aligned}$$

4.4 HD Routing is NP-hard

For a network represented by the directed graph $\mathcal{G}$, a path $\mathcal{P} = v_{k_1} - v_{k_2} - \ldots - v_{k_{m+1}}$ of length m in $\mathcal{G}$ is a sequence of vertices $v_{k_i} \in \mathcal{V}(\mathcal{G}), \forall i \in [1:m+1]$. An *S-D simple path* in $\mathcal{G}$ is a path for which $v_{k_1} = v_0 = S$ and $v_{k_m+1} = v_{N+1} = D$ and all $m+1$ vertices in $\mathcal{P}$ are distinct, i.e., there are no cycles in $\mathcal{P}$. From Theorem 4.3.1, the HD approximate capacity of the S-D simple path $\mathcal{P}$ is given by

$$\mathsf{C}_{\mathcal{P}} = \min_{i \in [2:m]} \left\{ \frac{\ell_{k_{i-1},k_i}\, \ell_{k_i,k_{i+1}}}{\ell_{k_{i-1},k_i} + \ell_{k_i,k_{i+1}}} \right\}. \tag{4.17}$$

Recall that ℓ_{k_{i-1},k_i} represents the link capacity of the edge from node $v_{k_{i-1}} \in \mathcal{P}$ to node $v_{k_i} \in \mathcal{P}$.

In this section, our goal is to prove Theorem 4.3.3, i.e., the problem of finding the best HD route in a network is NP-hard. Towards this end, we start by showing that, if we want to find the path $\mathcal{P}$ with the largest value of $\mathsf{C}_{\mathcal{P}}$ in (4.17), then we need to restrict our search over *simple* paths.

4.4.1 Non-simple Paths are Misleading in HD

Practically, a communication route through a network is expected to be a simple path, i.e, a path that contains no cycles. This is due to the fact that for a non-simple path, e.g., $\mathcal{P}_{\text{cyclic}} = S - v_1 - v_2 - \cdots - v_m - v_2 - D$, we know that – from the degraded nature of the network – the information sent from v_m to v_2 is a noisy version of the information that is already available at v_2 (since v_2 appeared earlier in the path). Thus, for the simple path $\mathcal{P}_{\text{simple}} = S - v_1 - v_2 - D$, we fundamentally have that

$$\mathsf{C}_{\mathcal{P}_{\text{cyclic}}} \leq \mathsf{C}_{\mathcal{P}_{\text{simple}}}. \tag{4.18}$$

This observation is true for both FD and HD paths in the network and therefore the best path (in FD or HD) is naturally a simple path. When routing using the FD capacities (to select the best FD route), this observation turns out to be just a technicality since the expression for the FD capacity already exhibits the fundamental property described in (4.18). Particularly, we have that $\mathcal{E}(\mathcal{P}_{\text{simple}}) \subseteq \mathcal{E}(\mathcal{P}_{\text{cyclic}})$, which directly implies that

$$\mathsf{C}^{\text{FD}}_{\mathcal{P}_{\text{cyclic}}} = \min_{e_{i,j} \in \mathcal{E}(\mathcal{P}_{\text{cyclic}})} \{\ell_{i,j}\} \leq \min_{e_{i,j} \in \mathcal{E}(\mathcal{P}_{\text{simple}})} \{\ell_{i,j}\} = \mathsf{C}^{\text{FD}}_{\mathcal{P}_{\text{simple}}}.$$

Thus, an algorithm that selects a route in FD can end up with either type of paths (simple or cyclic). If the path is cyclic, then we can prune it to get a simple path while ensuring that pruning can only improve the computed capacity.

Differently, for HD routing, it is very important to restrict ourselves to searching over simple paths as the HD approximate capacity expression in (4.17) only applies to simple paths. Furthermore, applying the expression in (4.17) to a path with a cycle can actually

increase the approximate capacity (in contradiction to the fundamental property in (4.18)). To illustrate this, consider the network example shown in Fig. 4.4. From Fig. 4.4, we now focus on the two paths: the simple path $\mathcal{P}_1 = S - v_1 - D$ and the non-simple path $\mathcal{P}_2 = S - v_1 - v_2 - v_3 - v_1 - D$. Note that $\mathcal{P}_1$ is a simple path and $\mathcal{P}_2$ is a cyclic extension of $\mathcal{P}_1$ by adding the cycle $v_1 - v_2 - v_3 - v_1$. If we apply the expression in (4.17) on both paths, we get the value equal to 7.5 for $\mathcal{P}_1$ and for $\mathcal{P}_2$ we get 13.05. Thus, if an algorithm is allowed to output non-simple paths, then it would output the path $\mathcal{P}_2$ even though we know fundamentally that $\mathsf{C}_{\mathcal{P}_1} \geq \mathsf{C}_{\mathcal{P}_2}$. This is the first major problem that arises when we allow an algorithm to consider non-simple paths based on the expression in (4.17). The second problem arises when we observe that $\mathcal{P}_3 = S - v_4 - D$ in Fig. 4.4 is actually the best HD simple path from S to D. However, since applying (4.17) for $\mathcal{P}_2$ yields 13.05, which is larger than what we get for $\mathcal{P}_3$ (i.e., 10), then the algorithm will output a non-simple path $\mathcal{P}_2$ which when pruned does not yield the best HD path. Thus, an algorithm designed with the goal to find the best HD path needs to be aware of the type of paths that it processes. In other words, we can no longer rely on pruning non-simple paths that an algorithm outputs as these in HD can mislead the algorithm into not selecting the best HD path as illustrated in this example.

As a consequence of the above discussion, in the rest of the section, we focus on the problem of finding the simple (i.e., acyclic) path with the largest HD approximate capacity.

4.4.2 Finding the Best HD Simple Path is NP-hard

Our goal in this subsection is to prove that the *search* problem of finding the S-D simple path with the largest HD approximate capacity in a relay network (represented by the digraph $\mathcal{G}$) is NP-hard. Towards proving this, we first show that the related decision problem "HD-Path", which is defined below, is NP-complete.

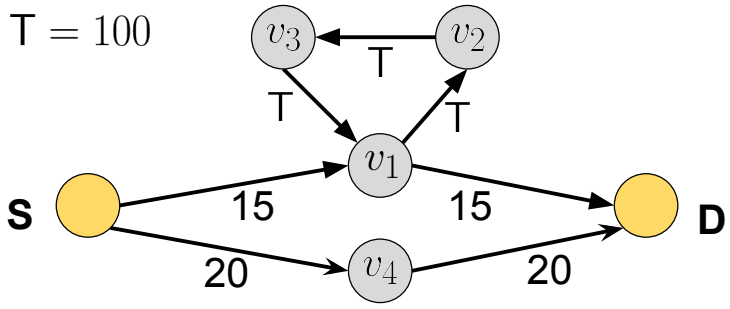


Figure 4.4: A network example in which a non-simple path can appear to have a larger HD approximate capacity than its simple subpath.

Definition 4.4.1 (HD-Path problem)**.** Given a directed graph $\mathcal{G}$ and a scalar $\mathsf{Z} > 0$, determine whether there exists an S-D simple path in $\mathcal{G}$ with an HD approximate capacity greater than or equal to Z.

Since the decision problem defined above can be reduced in polynomial time to finding the $S - D$ simple path with the largest HD approximate capacity, then by proving the NP-completeness of the decision problem in Definition 4.4.1, we also prove that the search problem is NP-hard.

The HD-Path problem is NP because, given a guess for a path, we can verify in polynomial time whether it is simple (i.e., no repeated vertices) and whether its HD approximate capacity is greater than or equal to Z by simply evaluating the expression in (4.17).

To prove the NP-completeness of the HD-Path problem, we now show that the classical 3SAT problem (which is NP-complete) [Kar72] can be reduced in polynomial time to the HD-Path decision problem in Definition 4.4.1. For the 3SAT problem, we are given a boolean expression B in 3-conjunctive normal form,

$$\begin{aligned} B(x_1, x_2, \ldots, x_n) =& (p_{11} \vee p_{12} \vee p_{13}) \wedge (p_{21} \vee p_{22} \vee p_{23}) \\ & \wedge \cdots \wedge (p_{m1} \vee p_{m2} \vee p_{m3}), \end{aligned} \tag{4.19}$$

where: (i) B is a conjunction of m clauses $\{C_1, C_2, \ldots, C_m\}$, each containing a disjunction of

three literals and (ii) a literal p_{ij} is either a boolean variable x_k or its negation $\bar{x}_k$ for some $k \in [1:n]$. The boolean expression B is *satisfiable* if the variables $x_{[1:n]}$ can be assigned boolean values so that B is true. The 3SAT problem answers the question: *Is the given B satisfiable?* We next prove the main result of this section through the following lemma.

Lemma 4.4.1. *A polynomial time reduction exists from the* 3SAT *problem to the* HD-Path *problem.*

Proof. To prove the claim, we create a sequence of graphs based on the boolean statement B given to the 3SAT problem. In each of these graphs, we show that the existence of a satisfying assignment for B is equivalent to a particular feature in the graph. Finally, we construct a network where the feature equivalent to a satisfying assignment of B is the existence of a simple path with HD approximate capacity greater than or equal to Z. In particular, our proof follows four steps of graph constructions, which are explained in detail in what follows.

Running example. To illustrate these four steps we use the following boolean expression as a running example,

$$B = (\bar{x}_1 \vee x_2 \vee x_3) \wedge (x_4 \vee x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_3 \vee \bar{x}_5), \tag{4.20}$$

where, with the notation in (4.19), the literals are assigned as

$$(p_{11}, p_{12}, p_{13}) = (\bar{x}_1, x_2, x_3), \tag{4.21a}$$

$$(p_{21}, p_{22}, p_{23}) = (x_4, x_1, \bar{x}_2), \tag{4.21b}$$

$$(p_{31}, p_{32}, p_{33}) = (\bar{x}_1, x_3, \bar{x}_5). \tag{4.21c}$$

Step 1. Assume that the boolean expression B is made of m clauses. For each clause $C_i, i \in [1:m]$ in B, construct a gadget digraph $\mathcal{G}_i$ with vertices $\mathcal{V}(\mathcal{G}_i) = \{t_i, v_{i1}, v_{i2}, v_{i3}, r_i\}$ and edges $\mathcal{E}(\mathcal{G}_i) = \bigcup_{j=1}^{3} \left\{e_{t_i,v_{ij}}, e_{v_{ij},r_i}\right\}$. Now we connect the gadget graphs $\mathcal{G}_i, i \in [1:m]$, by adding directed edges $e_{r_i,t_{i+1}}$, $\forall i \in [1:m-1]$. Finally, we introduce a source vertex S and

a destination vertex D and the directed edges e_{S,t_1} and $e_{r_m,D}$. We denote this new graph construction by $\mathcal{G}_B$. Note that each vertex v_{ij} in $\mathcal{G}_B$ represents a literal p_{ij} in the boolean expression B. We call a pair of vertices $(v_{ij}, v_{k\ell})$ in $\mathcal{G}_B$, with $i < k$, as *forbidden* if $p_{ij} = \overline{p_{k\ell}}$ in B.

Let $\mathcal{F}$ be the set of all such forbidden non-ordered pairs. Consider an S-D path $\mathcal{P} = S - t_1 - v_{1\ell_1} - r_1 - t_2 - \cdots - v_{m\ell_m} - r_m - D$ in the graph $\mathcal{G}_B$ that contains at most one vertex from any forbidden pair in $\mathcal{F}$. Using the indexes characterizing the path $\mathcal{P}$, if we set the literals $p_{i\ell_i}$ to be true $\forall i \in [1:m]$, then this is a valid assignment (since, by our definition, $\mathcal{P}$ avoids all forbidden pairs in $\mathcal{F}$). Additionally, since we set one literal to be true in each clause, then this assignment satisfies B. Hence the existence of a path $\mathcal{P}$ in $\mathcal{G}_B$ that avoids forbidden pairs implies that B is satisfiable. Similarly, we can show that if B is satisfiable, then we can construct a path that avoids forbidden pairs in $\mathcal{G}_B$ using any assignment that satisfies B.

Running example. The boolean expression in (4.20) has $m = 3$ clauses. Hence, we construct 3 gadget digraphs that are connected to form $\mathcal{G}_B$ as represented in Fig. 4.5. Since each vertex $v_{ij}, i \in [1:m],\ j \in [1:3]$, in $\mathcal{G}_B$ represents a literal p_{ij} in the boolean expression in (4.20) (i.e., $p_{ij} = v_{ij}$) and the literals are assigned as described in (4.21), then the set of forbidden pairs is given by

$$\mathcal{F} = \{(v_{11}, v_{22}), (v_{12}, v_{23}), (v_{22}, v_{31})\} \tag{4.22}$$

as also shown in Fig. 4.5.

Step 2. Next we modify the set of forbidden pairs $\mathcal{F}$ and the graph $\mathcal{G}_B$ such that each vertex appears at most once in $\mathcal{F}$. For each vertex v_{ij} that appears in at least one forbidden pair of $\mathcal{F}$, define $\mathcal{V}_{\mathcal{F}}(v_{ij}) = \{v_{i'j'} \in \mathcal{V}(\mathcal{G}_B) | (v_{ij}, v_{i'j'}) \in \mathcal{F} \text{ or } (v_{i'j'}, v_{ij}) \in \mathcal{F}\}$. Then, for each $\mathcal{V}_{\mathcal{F}}(v_{ij})$, we create $|\mathcal{V}_{\mathcal{F}}(v_{ij})|$ vertices and we label them as $v_{ij,k\ell}$, $\forall v_{k\ell} \in \mathcal{V}_{\mathcal{F}}(v_{ij})$. We finally replace the vertex v_{ij} in $\mathcal{G}_B$ with a path connecting the vertices $v_{ij,k\ell}$, $\forall v_{k\ell} \in \mathcal{V}_{\mathcal{F}}(v_{ij})$. We denote this new graph as $\mathcal{G}_B^{\circ}$. The new set of forbidden pairs $\mathcal{F}^{\circ}$ is defined based on the

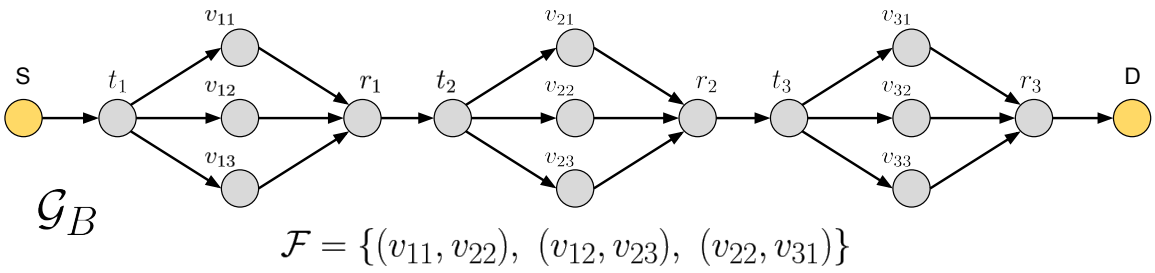


Figure 4.5: Graph $\mathcal{G}_B$ and set of forbidden pairs $\mathcal{F}$ for the boolean expression in (4.20).

set $\mathcal{F}$ as $\mathcal{F}^\circ = \{\,(v_{ij,k\ell}, v_{k\ell,ij}) |\ \ (v_{ij}, v_{k\ell}) \in \mathcal{F}\}$. Note that, for this new set of forbidden pairs, each vertex in $\mathcal{G}_B^\circ$ appears in at most one forbidden pair. Let $\mathcal{V}_{\mathcal{F}^\circ}$ be the set of vertices that appear in $\mathcal{F}^\circ$. Then $\forall v_{ij,kl} \in \mathcal{V}_{\mathcal{F}^\circ}$, we replace $v_{ij,kl}$ with a path that consists of three vertices. In particular, for any vertex $v_{ij,k\ell} \in \mathcal{V}_{\mathcal{F}^\circ}$, we replace it with a directed path $a_{ij,kl} - v_{ij,kl} - b_{ij,kl}$. We call this new graph $\mathcal{G}_B^\star$ and the forbidden pair set $\mathcal{F}^\star = \mathcal{F}^\circ$. The introduced vertices $a_{ij,k\ell}$ and $b_{ij,k\ell}$ are called *a-type* and *b-type* vertices, respectively.

Similar to our earlier argument for $\mathcal{G}_B$, note that a path in $\mathcal{G}_B^\star$ that avoids forbidden pairs in $\mathcal{F}^\star$ gives a valid satisfying assignment for the boolean argument B. In the reverse direction, if we have an assignment that satisfies B, then by taking one true literal from each clause C_i, $i \in [1:m]$, we can choose $t_i - r_i$ paths that avoid forbidden pairs. By connecting these paths together, we get an S-D path in $\mathcal{G}_B^\star$ that avoids forbidden pairs.

Running example. For our running example, given the set of forbidden pairs $\mathcal{F}$ in (4.22), we have

$$\mathcal{V}_{\mathcal{F}}(v_{11}) = \{v_{22}\} \implies v_{11} \leftarrow v_{11,22}, \quad \mathcal{V}_{\mathcal{F}}(v_{22}) = \{v_{11}, v_{31}\} \implies v_{22} \leftarrow v_{22,11} - v_{22,31},$$
$$\mathcal{V}_{\mathcal{F}}(v_{12}) = \{v_{23}\} \implies v_{12} \leftarrow v_{12,23}, \quad \mathcal{V}_{\mathcal{F}}(v_{23}) = \{v_{12}\} \implies v_{23} \leftarrow v_{23,12},$$
$$\mathcal{V}_{\mathcal{F}}(v_{31}) = \{v_{22}\} \implies v_{31} \leftarrow v_{31,22},$$

where $y \leftarrow \mathcal{Y}$ indicates that in $\mathcal{G}_B^\circ$ the vertex y is replaced by the path $\mathcal{Y}$. The set of forbidden pairs $\mathcal{F}^\circ$ is then given by

$$\mathcal{F}^\circ = \{(v_{11,22}, v_{22,11}), (v_{22,31}, v_{31,22}), (v_{12,23}, v_{23,12})\} \tag{4.23}$$

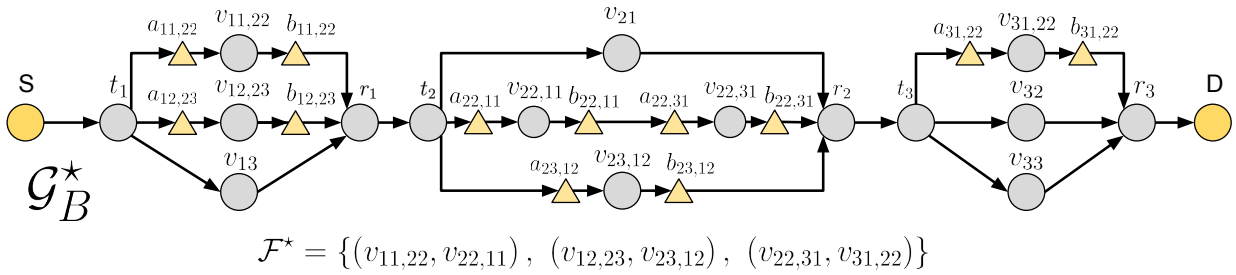


Figure 4.6: $\mathcal{G}_B^\star$ and set of forbidden pairs $\mathcal{F}^\star$. The graph $\mathcal{G}_B^\star$ is constructed from $\mathcal{G}_B$.

and hence $\mathcal{V}_{\mathcal{F}^\circ} = \{v_{11,22}, v_{22,11}, v_{22,31}, v_{31,22}, v_{12,23}, v_{23,12}\}$. Given this, we can now construct the graph $\mathcal{G}_B^\star$ by replacing any vertex inside $\mathcal{V}_{\mathcal{F}^\circ}$ as follows

$$
\begin{aligned}
v_{11,22} &\leftarrow a_{11,22} - v_{11,22} - b_{11,22}, & v_{22,11} &\leftarrow a_{22,11} - v_{22,11} - b_{22,11},\\
v_{22,31} &\leftarrow a_{22,31} - v_{22,31} - b_{22,31}, & v_{31,22} &\leftarrow a_{31,22} - v_{31,22} - b_{31,22},\\
v_{12,23} &\leftarrow a_{12,23} - v_{12,23} - b_{12,23}, & v_{23,12} &\leftarrow a_{23,12} - v_{23,12} - b_{23,12},
\end{aligned}
$$

as shown in Fig. 4.6 (in order to better visualize the 'evolution' from $\mathcal{G}_B$ to $\mathcal{G}_B^\star$, in Fig. 4.6 we also report again $\mathcal{G}_B$ of Fig. 4.5). Furthermore, we have $\mathcal{F}^\circ = \mathcal{F}^\star$, where $\mathcal{F}^\circ$ is defined in (4.23).

Step 3. Our next step is to modify $\mathcal{G}_B^\star$ to incorporate $\mathcal{F}^\star$ directly into the structure of the graph. For each $(v_{ij,k\ell}, v_{k\ell,ij}) \in \mathcal{F}^\star$ introduce a new vertex $f_{ij,k\ell}$ to replace $v_{ij,k\ell}$ and $v_{k\ell,ij}$. All edges that were incident from (to) $v_{ij,k\ell}$ and $v_{k\ell,ij}$ are now incident from (to) $f_{ij,k\ell}$. We call these newly introduced vertices as *f-type* vertices and denote this new graph as $\mathcal{G}_B^\bullet$. Note that in $\mathcal{G}_B^\bullet$, we now have incident edges from $a_{ij,k\ell}$ and $a_{k\ell,ij}$ to $f_{ij,k\ell}$ and edges incident from $f_{ij,k\ell}$ to vertices $b_{ij,k\ell}$ and $b_{k\ell,ij}$. A path in $\mathcal{G}_B^\star$ that avoids forbidden pairs in $\mathcal{F}^\star$ gives a path in $\mathcal{G}_B^\bullet$ that follows the following rules:

1. **Rule 1:** If any *f-type* vertex is visited, then it is visited at most once;

2. **Rule 2:** If an *f-type* vertex is visited then the preceding *a-type* vertex and the following *b-type* vertex both share the same index (i.e., we do not have $a_{ij,k\ell} - f_{ij,k\ell} - b_{k\ell,ij}$ or

$a_{k\ell,ij} - f_{ij,k\ell} - b_{ij,k\ell}$ as a subpath of our path in $\mathcal{G}_B^\bullet$).

It is not difficult to see that an $S - D$ path in $\mathcal{G}_B^\bullet$ that abides to the two aforementioned rules represents a feasible path that avoids forbidden pairs $\mathcal{F}^\star$ in $\mathcal{G}_B^\star$. Specifically, this can be seen by treating the subpath $(a_{ij,k\ell} - f_{ij,k\ell} - b_{ij,k\ell})$ in $\mathcal{G}_B^\bullet$ as the subpath $(a_{ij,k\ell} - v_{ij,k\ell} - b_{ij,k\ell})$ in $\mathcal{G}_B^\star$ and similarly $(a_{k\ell,ij} - f_{ij,k\ell} - b_{k\ell,ij})$ for $(a_{k\ell,ij} - v_{k\ell,ij} - b_{k\ell,ij})$. In other words, the problem of finding a path in $\mathcal{G}_B^\star$ that avoids forbidden pairs in $\mathcal{F}^\star$ is equivalent to finding a path in $\mathcal{G}_B^\bullet$ that satisfies Rule 1 and Rule 2.

Running example. For our running example, the graph $\mathcal{G}_B^\bullet$ is shown in Fig. 4.7. In particular, $\mathcal{G}_B^\bullet$ is constructed from $\mathcal{G}_B^\star$ in Fig. 4.6, where each $v_{ij,k\ell} \in \mathcal{V}(\mathcal{G}_B^\star)$ and $v_{k\ell,ij} \in \mathcal{V}(\mathcal{G}_B^\star)$ such that $(v_{ij,k\ell}, v_{k\ell,ij}) \in \mathcal{F}^\star$, with $\mathcal{F}^\star$ being defined in (4.23), is now replaced by $f_{ij,k\ell}$ in $\mathcal{G}_B^\bullet$, which is connected to the other nodes as explained above. In order to better visualize the 'evolution' from $\mathcal{G}_B$ to $\mathcal{G}_B^\bullet$, we also report again $\mathcal{G}_B$ of Fig. 4.5 and $\mathcal{G}_B^\star$ of Fig. 4.6.

Step 4. Our next step is to modify $\mathcal{G}_B^\bullet$ by introducing edge capacities. For any edge $e \in \mathcal{E}(\mathcal{G}_B^\bullet)$ that is not incident from or to an *f-type* vertex, we set the capacity of that edge to be $3\mathsf{Z}$. For an *f-type* vertex $f_{ij,k\ell}$, let g_1 and h_1 be the link capacities of the edges incident to it from $a_{ij,k\ell}$ and incident from it to $b_{ij,k\ell}$, respectively. Similarly, let g_2 and h_2 be the link capacities of the edges incident from $a_{k\ell,ij}$ and to $b_{k\ell,ij}$, respectively. Then, we set these capacities as

$$g_1 = h_2 = 1.5\mathsf{Z}, \quad g_2 = h_1 = 3\mathsf{Z}.$$

We now need to show that finding a path satisfying Rules 1 and 2 is equivalent to finding a simple path in $\mathcal{G}_B^\bullet$ with HD approximate capacity greater than or equal to Z. It is not difficult to see that a path that follows Rules 1 and 2 is simple and has an HD approximate capacity greater than or equal to Z (by avoiding subpaths $a_{ij,k\ell} - f_{ij,k\ell} - b_{k\ell,ij}$). This is due to the following fact. Consider a path that satisfies Rules 1 and 2. Then, for any two consecutive links in the considered path, at most one of the two links has a capacity of $1.5\mathsf{Z}$

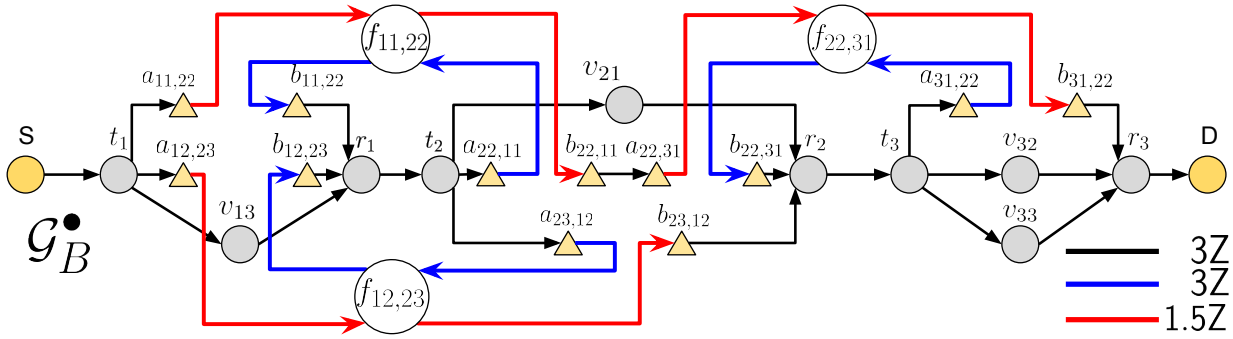


Figure 4.7: $\mathcal{G}_B^\bullet$ and the associated edge capacities. The graph $\mathcal{G}_B^\bullet$ is constructed from $\mathcal{G}_B^\star$.

(shown in red in Fig. 4.7), i.e., at least one link has a capacity of 3Z. Thus, a lower bound on the approximate capacity of the considered path is given by $(3\mathsf{Z} \times 1.5\mathsf{Z})/(3\mathsf{Z} + 1.5\mathsf{Z}) = \mathsf{Z}$.

To finally prove the equivalence, we now need to show that a simple path with capacity greater than or equal to Z satisfies Rules 1 and 2. Note that Rule 1 is inherently satisfied since the path is simple (i.e., it visits any vertex at most once). For Rule 2, we next argue that both subpaths are avoided by contradiction.

Assume that the simple path selected contains a subpath of the form $a_{ij,k\ell} - f_{ij,k\ell} - b_{k\ell,ij}$. By our construction of the edge capacities, both the edges $e_{a_{ij,k\ell},f_{ij,k\ell}}$ and $e_{f_{ij,k\ell},b_{k\ell,ij}}$ have a capacity equal to 1.5Z. This gives us a contradiction since half of the harmonic mean between the capacities of these two consecutive edges is equal to 0.75Z. Since the HD approximate capacity of a path is the minimum of half of the harmonic means of its consecutive edges, then the selected path cannot have an HD approximate capacity greater than or equal to Z, which leads to a contraction. Thus, a subpath $a_{ij,k\ell} - f_{ij,k\ell} - b_{k\ell,ij}$ is always avoided. We now need to prove that also the path $a_{k\ell,ij} - f_{ij,k\ell} - b_{ij,k\ell}$ is always avoided. Towards this end, assume that the simple path selected with HD approximate capacity greater than or equal to Z contains (for some i', j', k' and ℓ') a subpath of the form $a_{k'\ell',i'j'} - f_{i'j',k'\ell'} - b_{i'j',k'\ell'}$. Note that, as per our construction in the graph $\mathcal{G}_B^\bullet$, we have that $i' < k'$. Let $i^\star$ be the smallest index i' for which such a subpath exists in our selected path. Since for the subpath in question we have that $i^\star < k'$, then to reach $a_{k'\ell',i^\star j'}$ from S, we have already visited $r_{i^\star}$

earlier in the path. However, to move from $b_{i^\star j',k'\ell'}$ to D (after the subpath in question), we need to pass through $r_{i^\star}$ once more. Clearly, since the path is simple, this leads to a contradiction. Thus, a subpath $a_{k\ell,ij} - f_{ij,k\ell} - b_{ij,k\ell}$ is also always avoided.

This completes the proof that a simple path with capacity greater than or equal to Z satisfies Rule 2. Therefore, finding a path satisfying Rules 1 and 2 is equivalent to finding a simple path in $\mathcal{G}_B^\bullet$ with HD approximate capacity greater than or equal to Z. The second statement is an instance of the HD-Path problem in Definition 4.4.1.

Note that in each of the four graph constructions described earlier, we construct one graph from the other using a polynomial number of operations. Thus, this proves by construction that there exists a polynomial reduction from the 3SAT problem to the HD-Path problem. This concludes the proof of Lemma 4.4.1 and hence the proof of Theorem 4.3.3.

Running example. For our running example, the assignment of the edge capacities is shown in $\mathcal{G}_B^\bullet$ in Fig. 4.7, where *black* and *blue* edges have a capacity of $3\mathsf{Z}$ and *red* edges have $1.5\mathsf{Z}$. Fig. 4.7 also shows the evolution of $\mathcal{G}_B$ up to $\mathcal{G}_B^\bullet$.

□

4.5 Some Instances with a Polynomial-time Solution

In this section, we discuss a special class of networks for which a polynomial time algorithm exists to find a simple path with the largest HD approximate capacity. In particular, we focus on networks where the number of cycles is polynomial, i.e., the number of cycles is at most N^α for some constant $\alpha > 0$, where $N+2$ is the total number of nodes in the network. Our approach is based on relating paths in a network (described by the digraph $\mathcal{G}$) to paths in the line digraph of $\mathcal{G}$ denoted as $\mathcal{L}_\mathcal{G}$. We describe the relation in the next subsection and then present an algorithm that finds the best HD simple path in polynomial time for the aforementioned class of networks.

4.5.1 The Line Digraph Perspective to the Best HD Path Problem

The line digraph of a digraph $\mathcal{G}$ is defined as follows.

Definition 4.5.1 (Line digraph $\mathcal{L}_{\mathcal{G}}$)**.** For a given digraph $\mathcal{G}$, its line digraph $\mathcal{L}_{\mathcal{G}}$ is a digraph defined by the set of vertices $\mathcal{V}(\mathcal{L}_{\mathcal{G}})$ and the set of directed edges $\mathcal{E}(\mathcal{L}_{\mathcal{G}})$. The set $\mathcal{V}(\mathcal{L}_{\mathcal{G}})$ is defined as $\mathcal{V}(\mathcal{L}_{\mathcal{G}}) = \{v_{ij} | e_{i,j} \in \mathcal{E}(\mathcal{G})\}$ where $e_{i,j}$ is the directed edge from vertex v_i to vertex v_j. The set of edges $\mathcal{E}(\mathcal{L}_{\mathcal{G}})$ is defined as $\mathcal{E}(\mathcal{L}_{\mathcal{G}}) = \{e_{ij,k\ell} | k = j,\ v_{ij}, v_{k\ell} \in \mathcal{V}(\mathcal{L}_{\mathcal{G}})\}$.

An illustration of a digraph and its associated line digraph is shown in Fig. 4.8. We can make the following two observations on how simple HD paths are represented in the line digraph.

1) HD paths in $\mathcal{G}$ are equivalent to FD paths in $\mathcal{L}_{\mathcal{G}}$. Note that a path $\mathcal{P}$ in a network $\mathcal{G}$ can be equivalently defined as the sequence of its adjacent edges (instead of vertices), i.e., we can equivalently write the path $\mathcal{P} = v_{k_1} - v_{k_2} - \cdots - v_{k_m}$ in $\mathcal{G}$ as $\mathcal{P} = e_{k_1,k_2} - e_{k_2,k_3} - \cdots - e_{k_{m-1},k_m}$. Given this and from the definition of the line digraph $\mathcal{L}_{\mathcal{G}}$, the path $\mathcal{P}$ in $\mathcal{G}$ is equivalent to the path $\mathcal{P}_{\mathcal{L}} = v_{k_1k_2} - v_{k_2k_3} \cdots - v_{k_{m-1}k_m}$ in $\mathcal{L}_{\mathcal{G}}$. For each edge $e_{ij,jk} \in \mathcal{E}(\mathcal{L}_{\mathcal{G}})$, we define the capacity for the edge $e_{ij,jk}$ as

$$c_{\mathcal{L}}(e_{ij,jk}) = \frac{\ell_{i,j}\ \ell_{j,k}}{\ell_{i,j} + \ell_{j,k}}, \tag{4.24}$$

where $\ell_{i,j}$ is the point-to-point link capacity of the edge (link) $e_{i,j}$ in $\mathcal{G}$. Thus, we have that the FD capacity of the path $\mathcal{P}_{\mathcal{L}}$ in $\mathcal{L}_{\mathcal{G}}$ is given by

$$\begin{aligned} \mathsf{C}^{\mathrm{FD}}_{\mathcal{P}_{\mathcal{L}}} &= \min_{e_{ij,jk} \in \mathcal{E}(\mathcal{P}_{\mathcal{L}})} \{c_{\mathcal{L}}(e_{ij,jk})\} \\ &= \min_{e_{ij,jk} \in \mathcal{E}(\mathcal{P}_{\mathcal{L}})} \left\{ \frac{\ell_{i,j}\ \ell_{j,k}}{\ell_{i,j} + \ell_{j,k}} \right\} = \mathsf{C}_{\mathcal{P}}, \end{aligned} \tag{4.25}$$

where $\mathsf{C}_{\mathcal{P}}$ is defined in (4.17). From (4.25) and our previous discussion, we can conclude that, to find the path with the largest HD approximate capacity in the network described by the digraph $\mathcal{G}$, we can first find the path in $\mathcal{L}_{\mathcal{G}}$ that has the largest FD capacity (where

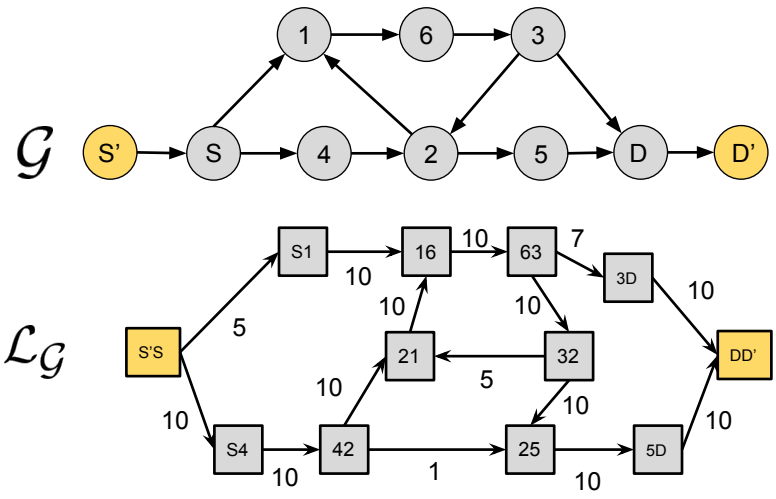


Figure 4.8: An example of a digraph $\mathcal{G}$ with its corresponding line digraph $\mathcal{L}_\mathcal{G}$. For ease of notation, indexes ij instead of v_{ij} are used and the edge capacities are only shown on $\mathcal{L}_\mathcal{G}$.

the link capacities in $\mathcal{L}_\mathcal{G}$ are defined as in (4.24)) and then map this path in $\mathcal{L}_\mathcal{G}$ into its equivalent in $\mathcal{G}$.

2) Simple paths in $\mathcal{G}$ are equivalent to simple chordless paths in $\mathcal{L}_\mathcal{G}$. We start by defining *chordal* and *chordless* paths in digraphs.

Definition 4.5.2 (Chordal and chordless paths)**.** A path in the digraph $\mathcal{G}'$ is *chordal* if there exists an edge $e \in \mathcal{E}(\mathcal{G}')$ such that its endpoints are two non-consecutive vertices in the path. A path that is not chordal is called *chordless*.

For example, with reference to Fig. 4.8, the path $S' - S - v_4 - v_2 - v_1 - v_6 - v_3 - D - D'$ is a chordal path in $\mathcal{G}$ since $e_{3,2} \in \mathcal{E}(\mathcal{G})$ and the vertices v_3 and v_2 belong to the path but are non-consecutive. Thus, $e_{3,2}$ is a chord for this path in $\mathcal{G}$. A similar reasoning holds for $e_{S,1}$.

Consider a cyclic path $\mathcal{P}_{\text{cycle}}$ in $\mathcal{G}$. This implies that some vertex $v_k \in \mathcal{P}_{\text{cycle}}$ appears at least twice in the path. Denote with v_{q_1} the node following v_k in its first appearance in $\mathcal{P}_{\text{cycle}}$ and with v_{q_2} the node preceding v_k in its second appearance in the path $\mathcal{P}_{\text{cycle}}$. Then, if we

write the line digraph equivalence of $\mathcal{P}_{\text{cycle}}$, we have

$$\mathcal{P}_{\mathcal{L}_{\text{cycle}}} = \cdots - v_{kq_1} - \cdots - v_{q_2k} - \ldots \ .$$

From the construction of $\mathcal{E}(\mathcal{L}_{\mathcal{G}})$ in Definition 4.5.1, we see that the edge $e_{q_2k,kq_1} \in \mathcal{E}(\mathcal{L}_{\mathcal{G}})$, which implies that $\mathcal{P}_{\mathcal{L}_{\text{cycle}}}$ is chordal. Differently, for a simple path $\mathcal{P}_{\text{simple}}$, any vertex $v_k \in \mathcal{P}_{\text{simple}}$ appears only once. Thus, in the line digraph equivalent path $\mathcal{P}_{\mathcal{L}_{\text{simple}}}$, the index k appears only in two consecutive vertices, which implies that $\mathcal{P}_{\mathcal{L}_{\text{simple}}}$ is chordless. This shows the equivalence described in our observation between simple paths in $\mathcal{G}$ and simple chordless paths in $\mathcal{L}_{\mathcal{G}}$.

Given the two observations above, we can now equivalently describe our HD routing problem on the line digraph as follows: *Can we find the chordless simple path in $\mathcal{L}_{\mathcal{G}}$ that has the largest FD capacity?*

4.5.2 An Algorithm on the Line Digraph $\mathcal{L}_{\mathcal{G}}$

The goal of the algorithm described in this section is to find the chordless simple path in $\mathcal{L}_{\mathcal{G}}$ that has the largest FD capacity. The algorithm described here is a modification of the result proposed in [AL09] for selecting shortest paths while avoiding forbidden subpaths in undirected graphs. The result in [AL09] needs to apply special care when fixing forbidden subpaths in a graph, due to the general unstructured nature of the forbidden set. In contrast, in our setting we will leverage the structured nature of our forbidden subpaths (chordal paths) and our line digraph $\mathcal{L}_{\mathcal{G}}$ to reduce the number of steps when breaking down a discovered chordal path (presented later in Step 3 of the algorithm). In particular, we make use of the fact that the first chord (corresponding to a chordal subpath) encountered within a selected path in the line digraph represents the smallest cycle encountered along the selected path in the original graph $\mathcal{G}$. Thus, by eliminating this subpath, we are sure that the number of remaining chordal paths (present on other paths or larger chordal paths on the same path) has not increased.

To start, we first modify our given network (described by $\mathcal{G}$) so that the source S and the destination D have at most degree one. In particular, we modify the digraph $\mathcal{G}$ by adding two new nodes (namely, S' and D') that are connected only to S and D with edges $e_{S',S}$ and $e_{D,D'}$ (similar to Fig. 4.8). These two added edges have point-to-point capacities equal to $X \to \infty$. Denote this new digraph by $\mathcal{G}'$ and create the line digraph associated with $\mathcal{G}'$ and denote it by $\mathcal{L}_{\mathcal{G}}^{(0)}$. In $\mathcal{L}_{\mathcal{G}}^{(0)}$, we now consider the node $v_{S'S}$ as our source and the node $v_{DD'}$ as our intended destination.

The algorithm is based on incrementally applying Dijkstra's algorithm [Dij59]. We first try to find the best FD path from $v_{S'S}$ to $v_{DD'}$ in $\mathcal{L}_{\mathcal{G}}^{(i)}$ by running Dijkstra's algorithm. Note that Dijkstra's algorithm returns a spanning tree rooted at $v_{S'S}$ that describes the best FD path from $v_{S'S}$ to each vertex v' in $\mathcal{L}_{\mathcal{G}}^{(i)}$. We denote the tree from our initial run as $\mathcal{T}_0$. From this point, the algorithm iterates (until termination) over four main steps described as follows (starting with $i = 0$).

Step 1. Given the line digraph $\mathcal{L}_{\mathcal{G}}^{(i)}$ and an existing best FD path spanning tree $\mathcal{T}_i$, check whether the path $\mathcal{P}_{\mathcal{L}}^{(i)}$ from $v_{S'S}$ to $v_{DD'}$ defined by $\mathcal{T}_i$ is chordless. If it is chordless, terminate the algorithm since we have found the chordless path from $v_{S'S}$ to $v_{DD'}$ with the largest FD capacity. Otherwise, if it is not chordless, then proceed to Step 2.

Running example. We use the line digraph from Fig. 4.8 as our $\mathcal{L}_{\mathcal{G}}^{(0)}$. Then, for $i = 0$, we have the spanning tree $\mathcal{T}_0$ (from Dijkstra's algorithm) and the selected path $\mathcal{P}_{\mathcal{L}}^{(0)}$ as shown in Fig. 4.9. The path $\mathcal{P}_{\mathcal{L}}^{(0)}$ is chordal since $e_{32,21} \in \mathcal{L}_{\mathcal{G}}^{(0)}$ and $e_{42,25} \in \mathcal{L}_{\mathcal{G}}^{(0)}$.

Step 2. Let $\mathcal{C}_{\mathcal{P}}^{(i)}$ be the set of edges in $\mathcal{L}_{\mathcal{G}}^{(i)}$ that are chords for the path $\mathcal{P}_{\mathcal{L}}^{(i)}$ from $v_{S'S}$ to $v_{DD'}$ discussed in the earlier step. Let $\mathcal{C}_{\mathcal{P},\text{first}}^{(i)} \in \mathcal{C}_{\mathcal{P}}^{(i)}$ be the first chord that appears along the path $\mathcal{P}_{\mathcal{L}}^{(i)}$. We denote the endpoints of $\mathcal{C}_{\mathcal{P},\text{first}}^{(i)}$ as $v_{k_1k_2}$ and $v_{k_mk_{m+1}}$, where $v_{k_1k_2}$ is the vertex that among the two appears earlier in the path $\mathcal{P}_{\mathcal{L}}^{(i)}$ and where m is the length of the subpath $\mathcal{P}_{\text{to-fix}}^{(i)}$ of $\mathcal{P}_{\mathcal{L}}^{(i)}$ connecting the two endpoints, i.e., we now have a path $\mathcal{P}_{\text{to-fix}}^{(i)} = v_{k_1k_2} - v_{k_2k_3} - \cdots - v_{k_mk_{m+1}}$. Notice that, with this, we have $k_{m+1} = k_1$.

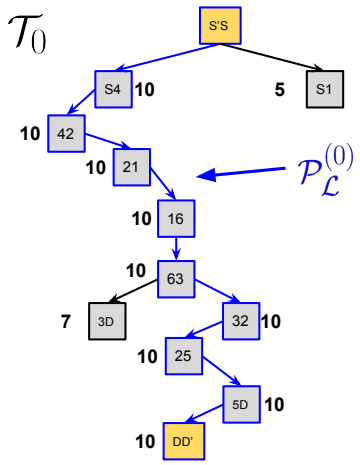


Figure 4.9: Tree $\mathcal{T}_0$ for $\mathcal{L}_{\mathcal{G}}^{(0)} = \mathcal{L}_{\mathcal{G}}$ in Fig. 4.8 (indexes ij instead of v_{ij} are used for ease of notation). Boldface numbers represent the FD capacity with which a node can be reached from SS' using $\mathcal{T}_0$. The highlighted path is the route selected from this tree $\mathcal{T}_0$ from $S'S$ to DD'.

Running example. For our running example and $i = 0$, we can see from Fig. 4.8 and Fig. 4.9 that the set of chords for $\mathcal{P}_{\mathcal{L}}^{(0)}$ is $\mathcal{C}_{\mathcal{P}}^{(0)} = \{e_{32,21}, e_{42,25}\}$. The selected chord $\mathcal{C}_{\mathcal{P},\text{first}}^{(0)}$ is $e_{32,21}$ because its effect on the path concludes earlier than $e_{42,25}$. Hence, we have $\mathcal{P}_{\text{to-fix}}^{(0)} = v_{21} - v_{16} - v_{63} - v_{32}$, which is of length $m = 4$.

Step 3. We now introduce new vertices to the graph $\mathcal{L}_{\mathcal{G}}^{(i)}$ by replicating every intermediate vertex in $\mathcal{P}_{\text{to-fix}}^{(i)}$. In particular, we introduce a replica vertex $v_{k_i' k_{i+1}'}$ for $v_{k_i k_{i+1}}$ where $i \in [2 : m-1]$. We connect these replicas of vertices to each other in the same way their corresponding originals are connected in $\mathcal{P}_{\text{to-fix}}^{(i)}$, i.e., we include the edge $e_{k_i' k_{i+1}', k_{i+1}' k_{i+2}'} \forall i \in [2 : m-1]$ with the same edge capacity as $e_{k_i k_{i+1}, k_{i+1} k_{i+2}}$.

Then, for every $v_{i'j'} \in \mathcal{V}(\mathcal{L}_{\mathcal{G}}^{(i)}) \backslash \mathcal{V}(\mathcal{P}_{\text{to-fix}}^{(i)})$ such that $e_{i'j', k_i k_{i+1}} \in \mathcal{E}(\mathcal{L}_{\mathcal{G}}^{(i)})$, we add an edge that connects $v_{i'j'}$ to the replica vertex of $v_{k_i k_{i+1}}$, i.e., we add the edge $e_{i'j', k_i' k_{i+1}'}$ (with the same edge capacity as $e_{i'j', k_i k_{i+1}}$). In other words, every vertex in $\mathcal{L}_{\mathcal{G}}^{(i)}$ that is not in $\mathcal{P}_{\text{to-fix}}^{(i)}$ and has an edge incident on an intermediate vertex $v_{k_i k_{i+1}}, i \in [2 : m-1]$, of $\mathcal{P}_{\text{to-fix}}^{(i)}$ now has a similar (replicated) edge incident on the replica $v_{k_i' k_{i+1}'}$ of $v_{k_i k_{i+1}}$. Note that at this point:

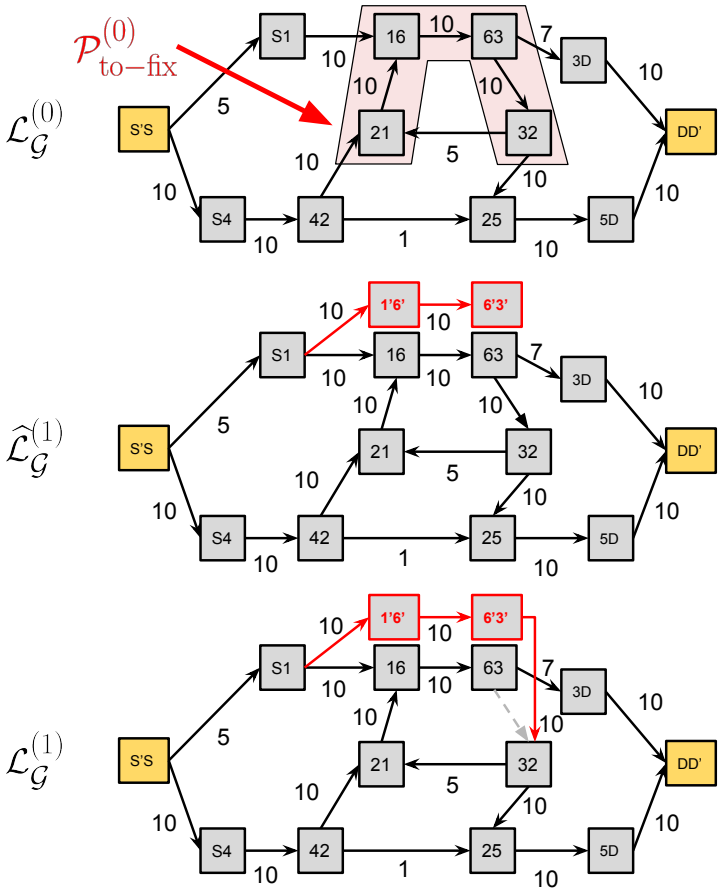


Figure 4.10: $\mathcal{L}^{(0)}_{\mathcal{G}}$ from Fig. 4.8 and the corresponding $\widehat{\mathcal{L}}^{(1)}_{\mathcal{G}}$ and $\mathcal{L}^{(1)}_{\mathcal{G}}$. The replica vertices and the added edges are shown in red while the deleted edges are dashed.

(i) the original vertices in $\mathcal{P}^{(i)}_{\text{to-fix}}$ still form a chordal path in $\mathcal{L}^{(i)}_{\mathcal{G}}$ and (ii) the replica vertices have every possible incident connection their original vertices had except connections to the two endpoint vertices of $\mathcal{P}^{(i)}_{\text{to-fix}}$. We denote the digraph at this point as $\widehat{\mathcal{L}}^{(i+1)}_{\mathcal{G}}$.

Now, our last change is to modify how the two endpoints of the path $\mathcal{P}^{(i)}_{\text{to-fix}}$ in $\widehat{\mathcal{L}}^{(i+1)}_{\mathcal{G}}$ connect to the intermediate vertices of the path and their replicas. We do this by adding the edge $e_{k'_{m-1}k'_m,k_mk_{m+1}}$ that connects the last replicated vertex $v_{k'_{m-1}k'_m}$ to the endpoint $v_{k_mk_{m+1}}$of $\mathcal{P}^{(i)}_{\text{to-fix}}$ and by removing the edge $e_{k_{m-1}k_m,k_mk_{m+1}}$ that connected the original last intermediate vertex to the endpoint. In particular, the new edge $e_{k'_{m-1}k'_m,k_mk_{m+1}}$ has the same capacity as

$e_{k_{m-1}k_m,k_mk_{m+1}}$ that was removed. Denote this new digraph as $\mathcal{L}_{\mathcal{G}}^{(i+1)}$. Note that in this new digraph $\mathcal{L}_{\mathcal{G}}^{(i+1)}$, the path $\mathcal{P}_{\text{to-fix}}^{(i)}$ does not exist anymore, while all the other chordless paths have stayed the same. This is due to the fact that we consider the first encountered chord and therefore vertex replication applied in this step either leaves larger chordal subpaths unaffected or eliminates them. This structure is the key step that allowed us to perform a simpler form of vertex replication as compared to [AL09]. Thus, we have successfully eliminated a cycle (chordal path) that appeared in the digraph before by replicating vertices and deleting edges.

Running example. For our running example and iteration $i = 0$, recall that the chordal path that we would like to fix is given by $\mathcal{P}_{\text{to-fix}}^{(0)} = v_{21} - v_{16} - v_{63} - v_{32}$ (see Fig. 4.10), where there is a chord due to v_{21} and v_{32}. Only the intermediate vertices of $\mathcal{P}_{\text{to-fix}}^{(0)}$, v_{16} and v_{63} are replicated, while the endpoints v_{21} and v_{32} are unchanged. To generate $\mathcal{L}_{\mathcal{G}}^{(1)}$, we first create $\widehat{\mathcal{L}}_{\mathcal{G}}^{(1)}$ by replicating the intermediate nodes v_{16} and v_{63} (denoted as $v_{1'6'}$ and $v_{6'3'}$) and all incident edges on them that are not part of $\mathcal{P}_{\text{to-fix}}^{(0)}$. This is shown in Fig. 4.10. In this case, the only such edge is $e_{S1,16}$ which is replicated by introducing edge $e_{S1,1'6'}$ with the same capacity. To arrive at $\mathcal{L}_{\mathcal{G}}^{(1)}$, we finally remove the last edge in $\mathcal{P}_{\text{to-fix}}^{(0)}$ that connects v_{63} to v_{32} and replace it with an edge connecting $v_{6'3'}$ to v_{32}. In this case, the chordal path $\mathcal{P}_{\text{to-fix}}^{(0)}$ is eliminated (by removing $e_{63,32}$), while all other paths of the type $\cdots - v_{21} - v_{16} - v_{63} - \cdots$ are still available from the remaining part of $\mathcal{P}_{\text{to-fix}}^{(0)}$. Additionally any path that would have used $v_{16} - v_{63} - v_{32}$ (for example $v_{S1} - v_{16} - v_{63} - v_{32}$) is now served by a replica path through the sequence of vertices $v_{S1} - v_{1'6'} - v_{6'3'} - v_{32}$. Thus, we have removed the chordal path $\mathcal{P}_{\text{to-fix}}^{(0)}$ and kept all the other possible paths unchanged or replaced them with a replica. The new generated digraphs $\widehat{\mathcal{L}}_{\mathcal{G}}^{(1)}$ and $\mathcal{L}_{\mathcal{G}}^{(1)}$ are shown in Fig. 4.10.

Step 4. In the fourth step, our goal is to create the spanning tree $\mathcal{T}_{i+1}$ of the best FD paths associated with the digraph $\mathcal{L}_{\mathcal{G}}^{(i+1)}$. To ensure termination of the algorithm, a condition for this construction is that $\mathcal{T}_{i+1}$ should be made as similar as possible to $\mathcal{T}_i$ [AL09]. To do so, we run Dijkstra's algorithm to find $\mathcal{T}_{i+1}$ but we start at an intermediate stage in the algorithm,

since we already know part of the spanning tree from $\mathcal{T}_i$. In particular, we do the following procedure. Recall our definition of $\mathcal{P}^{(i)}_{\text{to-fix}}$ and its endpoint $v_{k_m k_{m+1}}$ in Step 2. Define $\mathcal{V}^{(i+1)}_{\text{redo}}$ to be the set of vertices for which we need to find a new best FD path. In particular, define $\mathcal{V}^{(i+1)}_{\text{redo}}$ as the union of: (i) the set of all replica vertices introduced in $\mathcal{L}^{(i+1)}_{\mathcal{G}}$, (ii) the set of descendant vertices of $v_{k_m k_{m+1}}$ in $\mathcal{T}_i$, and (iii) the vertex $v_{k_m k_{m+1}}$. For any vertex $v \notin \mathcal{V}^{(i+1)}_{\text{redo}}$, the path connecting $v_{SS'}$ to v in $\mathcal{T}_i$ does not pass through $\mathcal{P}^{(i)}_{\text{to-fix}}$. As a result, we can copy this part of $\mathcal{T}_i$ to $\mathcal{T}_{i+1}$ without loss of generality. Clearly, replica vertices never existed before $\mathcal{L}^{(i+1)}_{\mathcal{G}}$ so there is no known path for them in $\mathcal{T}_i$. Similarly, the path from $v_{S'S}$ to $v_{k_m k_{m+1}}$ (and its descendants) passes through $\mathcal{P}^{(i)}_{\text{to-fix}}$, thus, we need to find a new route for them now that the chordal path has been removed from the graph. Also it is not difficult to see that any $v' \notin \mathcal{V}^{(i+1)}_{\text{redo}}$ will not be a descendant of v, $\forall v \in \mathcal{V}^{(i+1)}_{\text{redo}}$ as this would contradict the need to find a new path for some vertex in $\mathcal{V}^{(i+1)}_{\text{redo}}$.

As per our discussion above, we find the rest of $\mathcal{T}_{i+1}$ by initializing an intermediate point in the Dijkstra's algorithm and continue the execution of the algorithm from there. In particular, we start from the point where $\forall v \notin \mathcal{V}^{(i+1)}_{\text{redo}}$ have been expanded (and thus appear in $\mathcal{T}_{i+1}$ with the same path as in $\mathcal{T}_i$). We denote the intermediate version of $\mathcal{T}_{i+1}$ at this point as $\mathcal{T}'_{i+1}$, which is a pruned version of the tree $\mathcal{T}_i$. Note that, at any iteration of the classical Dijkstra's algorithm, a yet to be expanded vertex v has a best so-far path from $v_{S'S}$ of FD capacity $c'(v)$. This achievable FD capacity at an unexpanded vertex v is based on the maximum capacity achieved by each of the neighbor vertices that have already been expanded and added to the spanning tree $\mathcal{T}'_{i+1}$ as well as the capacities of incident edges from those neighbor vertices to the vertex v. We denote the capacity of a neighbor vertex v' that was already expanded as $\hat{c}^{\mathcal{T}_{i+1}}_{\mathcal{L}}(v')$. We now note that the point from which we are going to start Dijkstra's algorithm is when the set of unexpanded vertices is $\mathcal{V}^{(i+1)}_{\text{redo}}$ and the vertices in $\mathcal{T}'_{i+1}$ form the complement set $\mathcal{V}^{(i+1)^c}_{\text{redo}}$. Thus, for the vertices still unexpanded (i.e., those in $\mathcal{V}^{(i+1)}_{\text{redo}}$), the capacities currently achievable at them at this stage of the algorithm

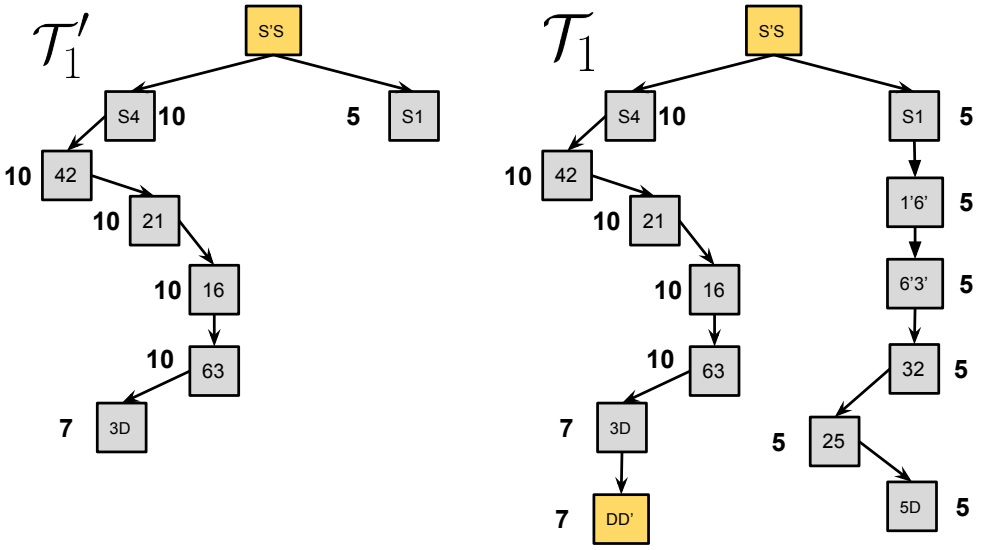


Figure 4.11: Spanning trees $\mathcal{T}_1'$ and $\mathcal{T}_1$ for $\mathcal{L}_{\mathcal{G}}^{(1)}$ in Fig. 4.10. Boldface numbers represent the FD capacity with which a node can be reached from SS' using each tree.

are initialized as

$$\widehat{c}_{\mathcal{L}}(v) = \max_{v' \notin \mathcal{V}_{\text{redo}}^{(i+1)}} \min\left\{c_{\mathcal{L}}^{\mathcal{T}_i}(v),\ c_{\mathcal{L}}(e_{v',v})\right\}.$$

Now that we have the initialization of Dijkstra's algorithm to the state that we want, we run the standard routine of the algorithm to continue expanding the vertices in $\mathcal{V}_{\text{redo}}^{(i+1)}$. When all the vertices have been expanded, we get the final tree $\mathcal{T}_{i+1}$.

Running example. For our running example and $i = 0$, the tree $\mathcal{T}_1'$ (which is a subset of $\mathcal{T}_0$) and the new generated tree $\mathcal{T}_1$ for $\mathcal{L}_{\mathcal{G}}^{(1)}$ are shown in Fig. 4.11. It is worth noting that the spanning tree $\mathcal{T}_1$ in Fig. 4.11 has the path $\mathcal{P}_{\mathcal{L}}^{(1)} = v_{SS'} - v_{S4} - v_{42} - v_{21} - v_{16} - v_{63} - v_{3D} - v_{DD'}$ of capacity $\mathsf{C}^{\text{FD}}_{\mathcal{P}_{\mathcal{L}}^{(1)}} = 7$ that is chordless. Hence the algorithm returns this path and terminates (see Step 1).

It is important to note that, from the replication procedure we do in Step 3, we add a number of replica vertices equal to the length of $\mathcal{P}_{\text{to-fix}}^{(i)}$ minus two (since we do not replicate the endpoints). Moreover, in addition to the replica vertices, only one endpoint of $\mathcal{P}_{\text{to-fix}}^{(i)}$ is

a member of $\mathcal{V}_{\text{redo}}^{(i+1)}$ (i.e., the vertex $v_{k_m k_{m+1}}$). As a result $\left|\mathcal{V}_{\text{redo}}^{(i)}\right| \leq |\mathcal{V}(\mathcal{L}_{\mathcal{G}})|, \forall i$. Thus, the size of the network that Dijkstra's algorithm processes in Step 4 does not increase from one iteration to the next. This implies that Step 4 of the algorithm has a complexity that is at most $O(V_{\mathcal{L}_{\mathcal{G}}} \log V_{\mathcal{L}_{\mathcal{G}}} + E_{\mathcal{L}_{\mathcal{G}}})$, where $V_{\mathcal{L}_{\mathcal{G}}} = |\mathcal{V}(\mathcal{L}_{\mathcal{G}})|$ and $E_{\mathcal{L}_{\mathcal{G}}} = |\mathcal{E}(\mathcal{L}_{\mathcal{G}})|$. The time complexity of Steps 1, 2 and 3 is linear in $V_{\mathcal{L}_{\mathcal{G}}}$ and $E_{\mathcal{L}_{\mathcal{G}}}$. Let $K_{\mathcal{G}}$ be the number of cycles in $\mathcal{G}$. From the observation in Section 4.5.1, this is equal to the number of chordal paths in $\mathcal{L}_{\mathcal{G}}$. Since in each iteration over the four steps, we eliminate one chordal path, then for a line graph with $K_{\mathcal{G}}$ chordal paths, we make at most $K_{\mathcal{G}}$ iterations. As a result, the complexity of the described algorithm for finding the simple chordless path with the largest FD capacity in $\mathcal{L}_{\mathcal{G}}$ is $O\left((K_{\mathcal{G}}+1)(V_{\mathcal{L}_{\mathcal{G}}} \log V_{\mathcal{L}_{\mathcal{G}}} + E_{\mathcal{L}_{\mathcal{G}}})\right)$.

Note that the number of vertices in $\mathcal{L}_{\mathcal{G}}$ is equal to the number of edges in $\mathcal{G}$ and the number of edges in $\mathcal{L}_{\mathcal{G}}$ is upper bounded by the number of edges in $\mathcal{G}$ multiplied by the maximum vertex degree d. Additionally, the complexity of constructing a line digraph $\mathcal{L}_{\mathcal{G}}$ from a digraph $\mathcal{G}$ is of order $O(|\mathcal{E}(\mathcal{G})|d)$. Thus, the problem of finding the simple path in $\mathcal{G}$ with the largest HD approximate capacity is equivalent to creating the line digraph $\mathcal{L}_{\mathcal{G}}$ with FD capacities and then finding the chordless path with the largest FD capacity in that line digraph $\mathcal{L}_{\mathcal{G}}$. The computational complexity of this procedure is $O\left(|\mathcal{E}(\mathcal{G})|d + (K_{\mathcal{G}}+1)(V_{\mathcal{L}_{\mathcal{G}}} \log V_{\mathcal{L}_{\mathcal{G}}} + E_{\mathcal{L}_{\mathcal{G}}})\right) = O\left((K_{\mathcal{G}}+1)(|\mathcal{E}(\mathcal{G})| \log |\mathcal{E}(\mathcal{G})| + |\mathcal{E}(\mathcal{G})|d)\right)$. Now if we let $K_{\mathcal{G}} = O(N^{\alpha})$, we get the expression in Lemma 4.3.1 which concludes the proof.

4.6 Conclusion

In this chapter, we studied the problem of characterizing the HD approximate capacity of the N-relay HD line network and investigated the HD routing problem in networks. Towards this end, our first main result was the closed-form characterization of the HD approximate capacity for an N-relay line network (i.e., a path) as a function of the link capacities. We then developed a polynomial time algorithm for finding a simple schedule (one with at most

$N+1$ active states out of the 2^N possible ones) that achieves the HD approximate capacity of the N-relay line network. To the best of our knowledge, this is the first work which provides a closed-form expression for the approximate capacity of an HD relay network with general number of relays and designs an efficient algorithm to find a simple schedule which achieves it.

By leveraging the derived closed-form expression for the HD approximate capacity, we then proved that finding the path from the source to the destination with the largest HD approximate capacity is NP-hard in general. This represents a surprising result and it is fundamentally different from the FD counterpart, since the path with the largest FD capacity can always be discovered in polynomial time. Finally, we showed that, if the number of cycles inside the network is polynomial in the number of nodes, then a polynomial time algorithm exists to find the path with the largest HD approximate capacity.

4.7 Related Work

A route/path connecting a source node to a destination node through N relays is an N-relay line network. Since the line network is a physically degraded relay channel, its capacity is known to be given by the cut-set upper bound [Are81] and achieved by decode-and-forward. While for the FD case the capacity can be expressed in an elegant and simple form as the minimum among the point-to-point link capacities in the line network, a similar result for the HD case is not yet available, to the best of our knowledge. The main reason for this is due to the fact that in HD the channel input at each relay is also characterized by the state (either listen or transmit) of that particular relay [Kra04]. This allows to transmit further information from the source to the destination by switching among the 2^N possible listen/transmit states that can occur inside the network. Given this, it is not clear what is the optimal input distribution that maximizes the cut-set upper bound. Results in [ADT11, OD13, CTK14] generalized an observation in [Kra04], by showing that the HD

capacity of a Gaussian relay network can be approximated to within an additive constant gap (i.e., which is independent on the channel parameters and only depends on the number of relays N) by the cut-set upper bound evaluated with a deterministic schedule independent of the transmitted and received signals and with independent inputs. Throughout the chapter, we refer to this as the *approximate capacity*.

For the class of relay networks defined in [CTK16, Theorem 1] (which includes the practically relevant Gaussian noise case), the evaluation of the approximate capacity of an HD relay network can be cast as an optimization problem over 2^N cuts of the network (as in FD) each of which is a function of 2^N listen/transmit configuration states. As N increases, this evaluation, as well as determining an optimal schedule, become computationally expensive. In [CTK16] the authors proved the conjecture posed in [BFO16] in the context of Gaussian diamond networks[2], which states that at most $N + 1$ states are sufficient for approximate capacity characterization for a class of HD N-relay networks, which includes the Gaussian noise case. A schedule with at most $N + 1$ active states is referred to as *simple*. This result is promising since it implies that the network can be operated close to its capacity with a limited number of state switches. However, to the best of our knowledge, it is not clear yet if such simple schedules can be found with low-complexity algorithms. The authors in [OMJ12] designed an iterative algorithm to determine a schedule optimal for the approximate capacity when the relays employ decode-and-forward. In [EPS14], the authors proposed a node-grouping technique that provides polynomial time algorithms to compute the approximate capacity of certain classes of Gaussian HD relay networks that include the line network as special case. While the results in [OMJ12, EPS14] show that the approximate capacity can be obtained efficiently in polynomial time for special network topologies by solving a linear program (LP), it is not clear how to efficiently construct the schedule that achieves this approximate capacity. In contrast, in this work we derive the approximate

[2]An N-relay diamond network is a relay network topology where the source can communicate with the destination only through N non-interfering relays.

capacity of a general memoryless HD line network in closed form and we design an efficient algorithm, which outputs in $O(N)$ time a simple schedule that achieves it.

In [LHK12], the exact capacity region is derived for a line network where the source node, and possibly multiple relays along the way, have messages intended to the destination. The capacity in [LHK12] is given as a maximization over the augmented alphabet used for the channel inputs (the alphabet used in the transmitted signal plus an additional random variable denoting that a node is listening). Additionally, the optimization in [LHK12] does not explicitly give the distribution for the channel inputs and does not provide an explicit schedule for the network. In particular, given this formulation, it is possible to use the schedule to send additional information; this follows since the schedule can now be considered part of the distribution of the channel input [Kra04]. In contrast, in our work, we consider a noisy line network model, where the point-to-point link capacities can be different from one link to the next (implying that different alphabets are used for transmission over each link). Furthermore, we upper bound the amount of information that can be transfered through the schedule by a quantity that only depends on N, and we look at the approximate capacity, where a fixed schedule is selected based on the network parameters but is not part of the codebook. This allows us to develop a closed-form expression for the approximate capacity.

Given the derived closed-form expression for the approximate capacity of an HD line network, we then analyze the problem of HD routing, which consists of efficiently discovering the path with the largest HD approximate capacity in a relay network. As a result of its widespread use in currently deployed wireless networks, routing has been extensively studied in the literature. For instance, a line of work [PR99, CJ03, JM96] focused on finding a route between the source and the destination by assuming that the point-to-point link capacities can only take values of 0 or 1. Under this assumption, the route selection based on the FD point-to-point link capacities is optimal. Finding the route with the largest FD capacity is equivalent to the problem of finding the widest-path between a pair of vertices in a graph [Pol60]. This can be efficiently solved by adapting any algorithm that finds the

shortest path between a pair of vertices in a graph (e.g., Dijkstra's algorithm [Dij59]). For routing with multi-rate link capacities, several heuristic metrics were proposed to enhance the selection of routes in ad-hoc wireless networks [AHR04,DAB05]. In contrast with this set of works, we are interested in selecting the route with the largest HD approximate capacity, by also trying to address the fundamental complexity of finding such a route.

4.8 Appendices

4.8.1 Proof of Theorem 4.3.2

We here prove Theorem 4.3.2 by proving the following relations.

1. We first prove that the set of fundamental states in an HD line network is equivalent to the set of fundamental maximum cuts in a FD line network.

2. We next show that the problem of finding the set of fundamental maximum cuts for an N-relay FD line network is equivalent to the problem of finding subsets of non-consecutive integers in $[1:N]$.

3. Finally, we prove that the number of subsets of non-consecutive integers in $[1:N]$ is at least exponential in N.

4.8.1.1 Set of Fundamental Maximum Cuts

In Section 4.2.1 we proved that we can compute the approximate capacity $\mathsf{C}_{\mathcal{R}}$ in (4.6) by considering only $N+1$ cuts, which are the same that one would need to consider if the network was operating in FD. These $N+1$ cuts are "fundamental", i.e., they do not depend on the values of the point-to-point link capacities. This implies that we can write (4.6) as

the LP

$$\begin{aligned} \mathsf{C}_{\mathcal{R}} = \quad & \text{maximize} \quad x \\ & \text{subject to} \quad \mathbf{1}_{N+1} x \leq \mathbf{A}\lambda \\ & \text{and} \quad \mathbf{1}_{2^N}^T \lambda = 1,\ \lambda \geq \mathbf{0}_{2^N},\ x \geq 0, \end{aligned} \tag{4.26a}$$

where $\mathbf{A} \in \mathbb{R}^{(N+1)\times 2^N}$ has non-negative entries

$$[\mathbf{A}]_{i,j} = \hat{\ell}_i^{(j)}, \tag{4.26b}$$

where: (i) $i \in [1 : N+1],\ j \in [1 : 2^N]$; (ii) $\hat{\ell}_i^{(j)}$ is defined in (4.7). Clearly, the LP in (4.26) is feasible. The dual of the LP in (4.26) is given by

$$\begin{aligned} \mathsf{C}_{\mathcal{R}} = \quad & \text{minimize} \quad y \\ & \text{subject to} \quad \mathbf{1}_{2^N} y \geq \mathbf{A}^T \mathbf{v} \\ & \text{and} \quad \mathbf{1}_{N+1}^T \mathbf{v} \geq 1,\ \mathbf{v} \geq \mathbf{0}_{N+1}, \end{aligned} \tag{4.27}$$

where $\mathbf{A}$ is defined in (4.26b). Since the LP in (4.27) is a minimization and the entries of $\mathbf{A}$ are non-negative, then it is not difficult to see that, for all optimal solutions of (4.27), we have $\mathbf{1}_{N+1}^T \mathbf{v} = 1$. As a result, an optimal solution of (4.27) is a solution of

$$\begin{aligned} \mathsf{C}_{\mathcal{R}} = \quad & \text{minimize} \quad y \\ & \text{subject to} \quad \mathbf{1}_{2^N} y \geq \mathbf{A}^T \mathbf{v} \\ & \text{and} \quad \mathbf{1}_{N+1}^T \mathbf{v} = 1,\ \mathbf{v} \geq \mathbf{0}_{N+1}. \end{aligned} \tag{4.28}$$

Since in the LP in (4.27) we are seeking to minimize the objective function, this implies that at least one of the constraints of the type $\mathbf{1}_{2^N} y \geq \mathbf{A}^T \mathbf{v}$ (i.e., the maximum) is satisfied with equality. We can interpret (4.28) as the problem of finding the least maximum FD cut among a class of line networks $\mathcal{R}_{\mathbf{V}}$ derived from the original network $\mathcal{R}$, where $\mathbf{V} = \{\mathbf{v} \in \mathbb{R}^{N+1} |\ \mathbf{v} \geq \mathbf{0}, \sum_{i \in [1:N+1] v_i} = 1\}$,where v_i is the i-th element of $\mathbf{v}$. For each $\mathbf{v} \in \mathbf{V}$, we define a line network $\mathcal{R}_{\mathbf{v}} \in \mathcal{R}_{\mathbf{V}}$, where the point-to-point link capacities are modified by $\mathbf{v}$ as $\ell_i^{(v)} = \ell_i v_i$.

Let $\mathscr{F}_{\mathcal{M}}$ be the set of fundamental maximum cuts in a FD line network, i.e., the smallest set of cuts over which we need to search for the maximum cut in FD without explicit knowledge of the values of the link capacities or their ordering. Since $\mathscr{F}_{\mathcal{M}}$ is the set of fundamental maximum cuts, then it contains a maximum cut for any FD line network. As a result, it also contains the least maximum FD cut among the class of line networks $\mathcal{R}_{\mathbf{V}}$. With this, the rows of $\mathbf{A}^T$ (constraints in (4.28)) not corresponding to $\mathscr{F}_{\mathcal{M}}$ are redundant and can be ignored when trying to find an optimal solution in (4.28). As a consequence of strong duality, the dual multipliers (the states λ_s in (4.26)) corresponding to the fundamental maximum cuts in $\mathscr{F}_{\mathcal{M}}$ are sufficient to find a schedule optimal for approximate capacity. We now prove that, without any knowledge of the link capacities, we need to consider the network states associated to every element of $\mathscr{F}_{\mathcal{M}}$, i.e., considering only the network states corresponding to a subset of $\mathscr{F}_{\mathcal{M}}$ is not sufficient to achieve the approximate capacity. To prove that, it suffices to provide a network example, where for each $\mathcal{A} \in \mathscr{F}_{\mathcal{M}}$ the state $s_{\mathcal{A}^c} = \mathbb{1}_{\mathcal{A}^c}$ is the unique optimal schedule, i.e., $\lambda_{s_{\mathcal{A}^c}} = 1$. For an arbitrary $\mathcal{A} \in \mathscr{F}_{\mathcal{M}}$, define the line network with the link capacities

$$\ell_i = \begin{cases} 1 & \text{if} \quad i \in \mathcal{M}_{\mathcal{A}} \\ X \to \infty & \text{otherwise} \end{cases},$$

where

$$\mathcal{M}_{\mathcal{A}} = \left\{ i \in [1{:}N{+}1] \,\middle|\, i \in \mathcal{A} \cup \{N{+}1\},\ i-1 \in \mathcal{A}^c \cup \{0\} \right\}.$$

From the previous network construction, it is not difficult to see that the unique optimal schedule (one for which $\mathsf{C}_{\mathcal{R}} = \mathsf{C}_{\mathcal{R}}^{\mathrm{FD}}$) is $s_{\mathcal{A}^c} = \mathbb{1}_{\mathcal{A}^c}$, i.e., $\lambda_{s_{\mathcal{A}^c}} = 1$. Thus, for this particular network construction, the state $s_{\mathcal{A}^c}$ is necessary and hence we cannot further reduce the sufficient set to a subset of $\mathscr{F}_{\mathcal{M}}$, i.e., we need to consider the network states corresponding to every element of $\mathscr{F}_{\mathcal{M}}$.

This result implies that, to find the smallest set of states over which we should search for an optimal schedule for approximate capacity, we should find the set of maximum cuts in FD

and then consider their dual multipliers in (4.26). In what follows, we focus on estimating the cardinality of the set of fundamental maximum cuts $\mathscr{F}_{\mathcal{M}}$ in a FD line network, which, as shown above, gives the cardinality of the smallest search space for an optimal schedule.

4.8.1.2 Finding the Set of Possible Maximum Cuts through an Equivalent Problem

We start by introducing some definitions, which will be used in the rest of this section.

Definition 4.8.1. For a set of consecutive integers $[a{:}b]$, we call $\mathcal{H}$ a "punctured" subset of $[a{:}b]$ if $\forall i, j \in \mathcal{H}$ with $i \neq j$, we have $|i-j|>1$, i.e., $\mathcal{H}$ contains non-consecutive integers of $[a{:}b]$.

Definition 4.8.2. We call $\mathcal{H}$ a *"primitive punctured"* subset of $[a{:}b]$ if $\mathcal{H}$ is punctured in $[a{:}b]$ and $\forall i \in [a{:}b]\backslash\mathcal{H}$, $\mathcal{H} \cup \{i\}$ is not a punctured set, i.e., $\mathcal{H}$ is not a subset of any other punctured subset of $[a{:}b]$. We denote by $\mathcal{P}(a, b)$ the collection of all primitive punctured subsets of $[a{:}b]$.

We now use the two above definitions to state the following lemma, which is proved in the rest of this section.

Lemma 4.8.1. *The problem of finding the set of possible maximum cuts for a FD line network is equivalent to the problem of finding $\mathcal{P}(1, N+1)$, i.e., the collection of primitive punctured subsets of $[1:N+1]$.*

Proof. We start by defining two problems, namely P_1 and P_2, which are important for the rest of the proof:

$$\mathsf{P}_1: \qquad \max_{\mathcal{A}\subseteq[1:N]} g_1(\mathcal{A}) = \sum_{\substack{i\in\mathcal{A}\cup\{N+1\}\\ i-1\in\mathcal{A}^c\cup\{0\}}} \ell_i, \tag{4.29a}$$

$$\mathsf{P}_2: \qquad \max_{\substack{\mathcal{B}\subseteq[1:N+1]\\ \mathcal{B} \text{ is punctured}}} g_2(\mathcal{B}) = \sum_{i\in\mathcal{B}} \ell_i. \tag{4.29b}$$

Note that P_1 is the problem of finding the maximum FD cut in an N-relay line network. To relate the solutions of P_1 and P_2, we make use of the following definition.

Definition 4.8.3. Given a problem P, we denote with $\text{suf}(\mathsf{P})$ the smallest set of feasible solutions among which an optimal solution can be found for any instance of the problem P.

The proof is organized as follows:

1. **Step 1:** We prove that P_1 and P_2 are equivalent; as a consequence, there exists a function f such that $\text{suf}(\mathsf{P}_1) = f(\text{suf}(\mathsf{P}_2))$.
2. **Step 2:** Next we prove that $\text{suf}(\mathsf{P}_2) \subseteq \mathcal{P}(1, N+1)$, which implies that
$$\text{suf}(\mathsf{P}_1) \subseteq f(\mathcal{P}(1, N+1)).$$
3. **Step 3:** The previous step implies that the set $\mathcal{M}$ of possible maximum cuts is a subset of $f(\mathcal{P}(1, N+1))$. We finally prove that $\mathcal{M} = f(\mathcal{P}(1, N+1))$.

Once proved, these steps imply that we can map the problem of finding the set of possible maximum cuts for a FD line network to the problem of finding $\mathcal{P}(1, N+1)$. We prove these three steps in Appendix 4.8.2. □

Example. Consider the FD line network with $N = 7$. To find the set of possible maximum cuts, according to Lemma 4.8.1, we need to find $\mathcal{P}(1, 8)$, which is given by

$$\mathcal{P}(1,8) = \Big\{ \{1,4,7\}, \{2,4,7\}, \{2,5,7\}, \{2,5,8\}, \{1,3,5,7\}, \\ \{1,3,6,8\}, \{1,4,6,8\}, \{2,4,6,8\}, \{1,3,5,8\} \Big\}.$$

It turns out that we can retrieve the candidate maximum cuts $\mathcal{A}_i$ from $\mathcal{P}(1,8)$ as follows:

$$\mathcal{A}_i = \mathcal{H}_i \backslash \{8\},\ \mathcal{H}_i \in \mathcal{P}(1,8), \quad \forall i \in [1 : |\mathcal{P}(1,8)|].$$

To conclude the proof of Theorem 4.3.2, we need to understand how the size of $\mathcal{P}(1, N+1)$ grows with N, which is the goal of the following subsection.

4.8.1.3 The Size of the Collection of Primitive Punctured Subsets

Here, we prove that the size of the collection of primitive punctured subsets of $[1 : N+1]$ grows exponentially in N. In particular, we prove the following lemma.

Lemma 4.8.2. *Let $T(N)$ be the number of primitive punctured subsets of $[1 : N]$. Then, for all $N \geq 4$, we have the following relation,*

$$T(N) = T(N-2) + T(N-3).$$

The proof of the above lemma can be found in Appendix 4.8.3.

Remark 4.8.1. The result in Lemma 4.8.2 suggests that $T(N)$ grows exponentially fast. This can be proved by observing the following lower bound on $T(N)$:

$$T(N) = T(N-2) + T(N-3) \geq 2T(N-3), \tag{4.30}$$

where the inequality is a consequence of the fact that $T(N-2) \geq T(N-3)$. By recursive application of the bound in (4.30), we have that

$$\begin{aligned} T(N) &\geq 2T(N-3) \geq 2\left(2T(N-6)\right) \\ &= 4T(N-6) \geq 4\left(2T(N-9)\right) \\ &= \cdots . \end{aligned}$$

Thus, we have that $T(N) \geq 2^k T(N-3k)$, $\forall k \in [1 : \lfloor N/3 \rfloor]$. By choosing $k = \lfloor N/3 \rfloor - 1$, then for all $N \geq 4$, we have that

$$\begin{aligned} T(N) &\geq 2^{\lfloor \frac{N}{3} \rfloor - 1} T(N - 3\lfloor N/3 \rfloor + 3) \\ &\geq 2^{\frac{N}{3} - 2} T(N - 3\lfloor N/3 \rfloor + 3) \\ &\geq 2^{\frac{N}{3} - 2} T(3) = \frac{T(3)}{4} 2^{N/3} \geq \frac{T(1)}{4} 2^{N/3}, \end{aligned}$$

where the last two inequalities follow from the fact that $T(\cdot)$ is a non-decreasing function. The bound proved above implies that $T(N) = \Omega(2^{N/3})$. $\square$

Since the number of candidate active states is equal to the number of candidate maximum cuts in FD (see the discussion in Appendix 4.8.1.1) and this is equal to the number of primitive punctured subsets of $[1:N+1]$ (see Lemma 4.8.1), then the number of candidate active states grows as $\Omega(2^{N/3})$. This concludes the proof of Theorem 4.3.2.

Remark 4.8.2. Using the recurrence relation in Lemma 4.8.2, it is not difficult to prove that $T(N) = \Theta(\beta^N)$ where β is the unique real root of the polynomial $x^3 - x - 1 = 0$, i.e., $x = 1.325$. □

4.8.2 Proof of Lemma 4.8.1

We here prove each of the three steps highlighted in the proof of Lemma 4.8.1.

Step 1. We first start by proving that any feasible solution for P_1 in (4.29a) can be transformed into a feasible solution for P_2 in (4.29b) with the same value for the objective function, i.e., $\forall \mathcal{A} \subseteq [1:N]$,

$$\exists \text{ punctured } \mathcal{B}_{\mathcal{A}} \in [1:N+1], \text{ s.t. } g_1(\mathcal{A}) = g_2(\mathcal{B}_{\mathcal{A}}).$$

To show this, for $\mathcal{A} \subseteq [1:N]$, we simply define $\mathcal{B}_{\mathcal{A}}$ as

$$\mathcal{B}_{\mathcal{A}} = \left\{ i \in [1{:}N{+}1] \middle| i \in \mathcal{A} \cup \{N{+}1\}, i{-}1 \in \mathcal{A}^c \cup \{0\} \right\}. \tag{4.31}$$

It is clear that $\mathcal{B}_{\mathcal{A}}$ is a punctured set as $\forall i \in \mathcal{B}_{\mathcal{A}},\ i-1 \notin \mathcal{B}_{\mathcal{A}}$. Additionally, (4.31) directly gives us the desired relation as

$$g_1(\mathcal{A}) = \sum_{\substack{i \in \mathcal{A} \cup \{N+1\} \\ i-1 \in \mathcal{A}^c \cup \{0\}}} \ell_i = \sum_{i \in \mathcal{B}_{\mathcal{A}}} \ell_i = g_2(\mathcal{B}_{\mathcal{A}}). \tag{4.32}$$

What remains to prove now is that any feasible solution $\mathcal{B}$ for P_2 gives a feasible solution $\mathcal{A}_{\mathcal{B}}$ for P_1 and $g_1(\mathcal{A}_{\mathcal{B}}) = g_2(\mathcal{B})$. For a punctured subset $\mathcal{B}$ of $[1:N+1]$, let

$$\mathcal{A}_{\mathcal{B}} = f_{\mathcal{A}\mathcal{B}}(\mathcal{B}) = \underbrace{\left\{ i \in [1:N] \middle| \ i > \sup(\mathcal{B}) \right\}}_{\mathcal{A}_{\text{tail}}} \cup \underbrace{\mathcal{B} \backslash \{N+1\}}_{\mathcal{A}_{\text{main}}}. \tag{4.33}$$

It is not difficult to see that, by applying the transformation in (4.31) on $\mathcal{A}_{\mathcal{B}}$, we get back $\mathcal{B}$, i.e., $\mathcal{B}_{\mathcal{A}_{\mathcal{B}}} = \mathcal{B}$. This is due to the fact that applying (4.31) removes $\mathcal{A}_{\text{tail}}$ which is composed of a consecutive number of integers while keeping $\mathcal{A}_{\text{main}}$ which, since $\mathcal{B}$ is punctured, is also punctured. Given this, we can directly see from (4.32) that $g_1(\mathcal{A}_B) = g_2(\mathcal{B}_{\mathcal{A}_{\mathcal{B}}}) = g_2(\mathcal{B})$. This concludes the proof of Step 1.

Step 2. We prove this step by showing that, if there exists an optimal solution $\mathcal{B}^\star$ for P_2 that is not primitive, then there also exists a primitive punctured set $\mathcal{B}'$ such that $g_2(\mathcal{B}^\star) = g_2(\mathcal{B}')$. Since $\mathcal{B}^\star$ is not a primitive punctured set, then there exists another punctured set $\mathcal{B}'$ such that $\mathcal{B}^\star \subset \mathcal{B}'$ and

$$g_2(\mathcal{B}^\star) = \sum_{i \in \mathcal{B}^\star} \ell_i \leq \sum_{i \in \mathcal{B}'} \ell_i = g_2(\mathcal{B}').$$

If we take the largest such $\mathcal{B}'$, we end up with a primitive punctured set. However, by definition (i.e., since $\mathcal{B}^\star$ is an optimal solution) we have that $\forall \mathcal{B}$ punctured, $g_2(\mathcal{B}) \leq g_2(\mathcal{B}^\star)$. This shows that $g_2(\mathcal{B}^\star) = g_2(\mathcal{B}')$ and therefore, $\text{suf}\,(\mathsf{P}_2) \subseteq \mathcal{P}(1, N+1)$. This concludes the proof of Step 2.

Step 3. In the first two steps, we proved that P_1 and P_2 are equivalent and that $\text{suf}\,(\mathsf{P}_2) \subseteq \mathcal{P}(1, N+1)$. This implies that $\text{suf}\,(\mathsf{P}_1) \subseteq f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$, where $f_{\mathcal{AB}}(\cdot)$ is defined in (4.33). We here prove that $\text{suf}\,(\mathsf{P}_1) = f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$. Consider an arbitrary set $\mathcal{A} \in f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$. To prove that $\mathcal{A} \in \text{suf}\,(\mathsf{P}_1)$, it suffices to provide a network (an instance of P_1) for which $\mathcal{A}$ is the unique maximizer of P_1. Towards this end, for the selected $\mathcal{A}$, we define $\mathcal{B}_{\mathcal{A}}$ as in (4.31). We know that $\mathcal{B}_{\mathcal{A}}$ is a primitive punctured set and $g_1(\mathcal{A}) = g_2(\mathcal{B}_{\mathcal{A}})$. Now consider the network with link capacities

$$\ell_i = \begin{cases} 1 & \text{if } i \in \mathcal{B}_{\mathcal{A}} \\ 0 & \text{otherwise} \end{cases}.$$

For this network, it is not difficult to see that $g_2(\mathcal{B}) = |\mathcal{B} \cap \mathcal{B}_{\mathcal{A}}|$, for any punctured set $\mathcal{B}$. We now want to show that $\forall \mathcal{A}' \in f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right) \setminus \mathcal{A}$, we have $g_1(\mathcal{A}') < g_1(\mathcal{A})$. Let $\mathcal{B}_{\mathcal{A}'}$

be defined as in (4.31). Again, from the proof of the previous steps the set $\mathcal{B}_{\mathcal{A}'}$ is primitive punctured and $g_1(\mathcal{A}') = g_2(\mathcal{B}_{\mathcal{A}'})$. Moreover, since $\mathcal{B}_{\mathcal{A}'}$ and $\mathcal{B}_{\mathcal{A}}$ are both primitive we have that $\mathcal{B}_{\mathcal{A}'} \cap \mathcal{B}_{\mathcal{A}} \subset \mathcal{B}_{\mathcal{A}}$. Thus, we obtain

$$\begin{aligned} g_2(\mathcal{B}_{\mathcal{A}'}) =& |\mathcal{B}_{\mathcal{A}'} \cap \mathcal{B}_{\mathcal{A}}| < |\mathcal{B}_{\mathcal{A}}| = g_2(\mathcal{B}_{\mathcal{A}}) \\ &\implies g_1(\mathcal{A}') < g_1(\mathcal{A}). \end{aligned}$$

Since this is true for any arbitrary $\mathcal{A} \in f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$, then it is true $\forall \mathcal{A} \in f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$. This implies that each element in $f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$ is a unique maximum cut for some network construction. Therefore, without any information about the link capacities ℓ_i, we cannot further reduce the set of possible maximum cuts and thus we have $\mathrm{suf}\left(\mathsf{P}_1\right) = f_{\mathcal{AB}}\left(\mathcal{P}(1, N+1)\right)$. This concludes the proof of Step 3 and hence the proof of Lemma 4.8.1.

4.8.3 Proof of Lemma 4.8.2

To compute the size of $\mathcal{P}(a, b)$, it is helpful to first prove some properties of $\mathcal{P}(a, b)$ and primitive punctured subsets that will help throughout the proof.

Property 4.8.1. *Let $\mathcal{H}$ be a primitive punctured subset of $[a : b]$, then* $\min\{\mathcal{H}\} \leq a + 1$.

Proof. We prove this result by contradiction. Assume that for some primitive punctured set $\mathcal{H}$, we have $\min\{\mathcal{H}\} \geq a + 2$. This implies that $\mathcal{H} \subset [a + 2 : b]$. Let $\hat{\mathcal{H}} = \mathcal{H} \cup \{a\}$. Since $\mathcal{H}$ is a punctured set, then $\hat{\mathcal{H}}$ is also a punctured set because $\forall i \in \mathcal{H},\ |a - i| > 1$. But since $\mathcal{H} \subset \hat{\mathcal{H}}$, then $\mathcal{H}$ is not a primitive punctured set, which is a contradiction. □

Property 4.8.1 implies that, for a primitive punctured subset of $[a : b]$, the minimum element is either a or $a + 1$. Therefore, we can write $\mathcal{P}(a, b)$ as

$$\mathcal{P}(a, b) = \mathcal{P}_1(a, b) \ \uplus \ \mathcal{P}_2(a, b),$$

where $\mathcal{P}_1(a, b)$ (respectively, $\mathcal{P}_2(a, b)$) is the collection of primitive punctured sets with minimum element a (respectively, $a + 1$). Clearly, $\mathcal{P}_1$ and $\mathcal{P}_2$ are disjoint (we use $\uplus$ to indicate that the union is over disjoint sets). Next, we prove some properties of $\mathcal{P}_1(a, b)$ and $\mathcal{P}_2(a, b)$.

Property 4.8.2. $\mathcal{P}_2(a,b) = \mathcal{P}_1(a+1,b)$.

Proof. Let $\mathcal{H}$ be a primitive punctured subset of $[a+1:b]$ that contains the element $a+1$. $\mathcal{H}$ is also a primitive punctured subset of $[a:b]$. This follows since we cannot add $\{a\}$ to $\mathcal{H}$ to get a larger set of non-consecutive elements. Therefore, $\mathcal{H} \in \mathcal{P}_1(a+1,b) \implies \mathcal{H} \in \mathcal{P}_2(a,b)$. The reverse implication is straightforward since, by definition, $\mathcal{P}_2(a,b)$ is a primitive punctured subset which contains the element $a+1$. □

For the next property, we need to define a new operation on the collection of sets. For a collection of sets $\mathcal{Q}$, let $\{i\} \sqcup \mathcal{Q} = \{\{i\} \cup \mathcal{H} \mid \mathcal{H} \in \mathcal{Q}\}$. We then have the following property.

Property 4.8.3. $\mathcal{P}_1(a,b) = \{a\} \sqcup \mathcal{P}(a+2,b)$.

Proof. Let $\mathcal{H}$ be a primitive punctured subset of $[a+2:b]$ and define $\hat{\mathcal{H}} = \{a\} \cup \mathcal{H}$. Since $\mathcal{H}$ is a primitive punctured subset of $[a+2:b]$, this means that $\nexists i \in [a+2:b] \backslash \mathcal{H}$ such that $\{i\} \cup \mathcal{H}$ is a punctured sequence of $[a+2:b]$. This implies that $\nexists i \in [a:b] \backslash [\mathcal{H} \cup \{i\}]$ such that $\{i\} \cup \hat{\mathcal{H}}$ is a punctured sequence of $[a:b]$. Therefore $\hat{\mathcal{H}}$ is a primitive punctured sequence of $[a:b]$, i.e., $\hat{\mathcal{H}} \in \mathcal{P}_1(a,b)$. To prove the reverse, consider $\widetilde{\mathcal{H}} \in \mathcal{P}_1(a,b)$. We need to prove that $\hat{\mathcal{H}} = \widetilde{\mathcal{H}} \backslash \{a\}$ is a primitive punctured subset of $[a+2:b]$. Note that the definition of primitive subset of $[a:b]$ implies that $\forall i \in [a+2:b] \backslash \widetilde{\mathcal{H}}$, $\widetilde{\mathcal{H}} \cup \{i\}$ is not a punctured set. Since $a \notin [a+2:b]$, this implies that $\forall i \in [a+2:b] \backslash \hat{\mathcal{H}}$, $\widetilde{\mathcal{H}} \cup \{i\}$ is not a punctured set. Now note that since $\widetilde{\mathcal{H}} \in \mathcal{P}_1(a,b)$ then $a+1 \notin \widetilde{\mathcal{H}}$. Therefore, $\forall i \in [a+2:b]$ removing the element a from $\widetilde{\mathcal{H}} \cup \{i\}$ does not make it a punctured set. We therefore conclude that, $\forall i \in [a+2:b] \backslash \hat{\mathcal{H}}$, $\hat{\mathcal{H}} \cup \{i\}$ is not a punctured set and as a result $\hat{\mathcal{H}} = \widetilde{\mathcal{H}} \backslash \{a\}$ is a primitive punctured subset of $[a+2:b]$. □

We now have all the necessary tools to prove Lemma 4.8.2. We obtain

$$\begin{aligned} \mathcal{P}(1,N) &= \mathcal{P}_1(1,N) \uplus \mathcal{P}_2(1,N) \\ &\overset{(a)}{=} \mathcal{P}_1(1,N) \uplus \mathcal{P}_1(2,N) \end{aligned}$$

$$\overset{\text{(b)}}{=} \Big[\{1\} \sqcup \mathcal{P}(3, N)\Big] \ \uplus \ \Big[\{2\} \sqcup \mathcal{P}(4, N)\Big],$$

where the equality in (a) follows from Property 4.8.2 and the equality in (b) follows from Property 4.8.3. Now note that

$$\begin{aligned}\Big|\,[\{i\} \sqcup \mathcal{P}(a, N)]\,\Big| = \Big|\mathcal{P}(a, N)\Big| &= \Big|\mathcal{P}(1, N - a + 1)\Big| \\ &= T(N - a + 1),\end{aligned}$$

since the number of sets in each collection remains the same. Therefore, we have

$$\begin{aligned}T(N) &= |\mathcal{P}(1, N)| \\ &= \Big|\,[\{1\} \sqcup \mathcal{P}(3, N)] \ \uplus \ [\{2\} \sqcup \mathcal{P}(4, N)]\,\Big| \\ &= \Big|\,[\{1\} \sqcup \mathcal{P}(3, N)]\,\Big| + \Big|\,[\{2\} \sqcup \mathcal{P}(4, N)]\,\Big| \\ &= T(N - 2) + T(N - 3).\end{aligned}$$

This concludes the proof of Lemma 4.8.2.

Part II

Modeling mmWave Transmission

CHAPTER 5

The Gaussian 1-2-1 Network Model

In this chapter, we propose a new model for wireless relay networks referred to as "1-2-1 network", where two nodes can communicate only if they point "beams" at each other, otherwise no signal can be exchanged or interference can be generated. This model is motivated by millimeter wave communications where, due to the high path loss, a link between two nodes can exist only if beamforming gain at both sides is established, while in the absence of beamforming gain the signal is received well below the thermal noise floor. In this chapter, we present the 1-2-1 model and develop a constant gap approximation for its multicast capacity, i.e., a characterization of the network multicast capacity to within an additive gap, which only depends on the number of nodes and is independent of the channel coefficients and operating SNR. Afterwards, the relation between the approximate multicast capacity and minimum unicast capacity is explored in full-duplex 1-2-1 networks and shown to be dependent on the network structure and the number of destinations, unlike in classical wireless (i.e., without 1-2-1 constraints) full-duplex networks.

5.1 Introduction

Millimeter Wave (mmWave) communications are expected to play a vital role in 5G mobile communications and beyond, expanding the available spectrum and enabling multi-gigabit services such as next generation business solutions, virtual and augmented reality applications, and autonomous vehicle and drone communication [Pap16,Hea16,Bro,CVG16,MSK16, RRE14]. Although for *single-hop* mmWave networks several works have examined channel

modeling, performance bounds and algorithms, as far as we know, for *multi-hop* mmWave networks, fundamental performance bounds such as the information theoretical capacity have not been well explored.

In this chapter, we present a class of networks that we term 1-2-1 networks that offer a simple yet informative model for mmWave networks. The inherent characteristic of mmWave communications that our model captures is directivity: mmWave requires beamforming with narrow beams to compensate for high path loss incurred by isotropic transmission. To establish a communication link, both the mmWave transmitter and receiver employ antenna arrays that they electronically steer to direct their beams towards each other - we term this a 1-2-1 link, as both nodes need to focus their beams to face each other for the link to be active. Thus, in 1-2-1 networks, instead of broadcasting or interference, we have coordinated steering of transmit and receive beams to activate different links at each time. An example of a relay network with $N = 5$ relays is shown in Fig. 5.1, where two different states for the configuration of the transmit/receive beams are depicted and the resulting activated links are highlighted. Note that in this figure, the transmit and receive beams at each relay can be simultaneously active, and hence the relays operate in *Full-Duplex* (FD) mode. Differently, if at any point on time, either the transmit beam or the receive beam can be active, then the relays operate in *Half-Duplex* (HD). The existence of multiple operating states in FD and HD gives rise to a *scheduling* aspect: the network needs to dictate how the nodes should align their beams and for what fraction of time they should keep this alignment.

We present unicast and multicast capacity approximation results, as well as scheduling algorithms for the proposed FD and HD 1-2-1 network class[1]. The capacity characterization of a relay network with scheduling aspects is challenging due to the added complexity of optimizing the schedule over an exponential number of network states.

It is indeed not clear if such an optimal schedule can be found efficiently (i.e., in poly-

[1]Unicast can be treated as a special case of multicast when the number of destinations is one.

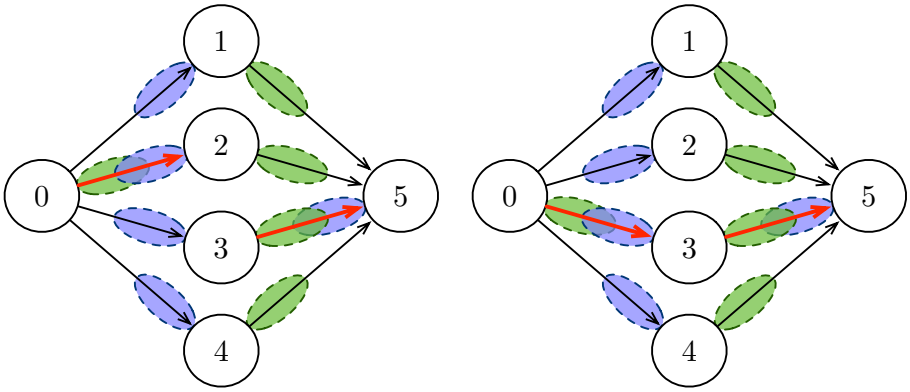


Figure 5.1: 1-2-1 network with $N = 5$ relays and two states.

nomial time in the network size) even for capacity approximation. Studying the complexity of scheduling 1-2-1 networks will be the topic of the next two chapters. In this chapter, we seek to develop an understanding for the operational properties for mmWave networks by studying an abstract information-theoretic model that captures key characteristics of their operation. Our main contributions towards this understanding are summarized as follows:

1. We propose an information-theoretic network model, namely the *Gaussian 1-2-1 network* that captures the inherent directivity and scheduling of mmWave networks due to beam steering. The proposed model considers both FD and HD modes of operation at the network nodes.

2. For the proposed network model, we derive a constant gap approximation of the multicast capacity when a single source node aims to communicate the same message to an arbitrary set of destination nodes. In particular, the derived expression encompasses the scheduling required to optimally operate the network by optimizing the fraction of time a beam orientation configuration state should be active. We show that the approximate multicast capacity, both for FD and HD, can be expressed as a linear program (LP), which has a polynomial number (in the number of nodes) of constraints, but an exponential number of variables.

3. We show that the derived expression for the approximate multicast capacity highlights key properties of Gaussian 1-2-1 networks that distinguish them from classical Gaussian wireless relay networks. In particular, we study the relation between the approximate multicast capacity and the minimum approximate unicast capacity in a Gaussian FD 1-2-1 network. Unlike in classical networks, where the ratio between the two quantities is universally equal to one, we show that for the FD 1-2-1 network, the ratio depends on the number of destinations as well as graph-theoretic properties of the network.

Chapter Organization. Section 5.2 describes the Gaussian 1-2-1 network model and Section 5.3 derives a constant gap approximation of its multicast capacity for FD and HD modes of operation. Section 5.4 shows that finding an optimal schedule for a Gaussian FD 1-2-1 network can be done in polynomial time. Section 5.5 derives the worst case ratio between multicast and unicast approximate capacities in Gaussian FD 1-2-1 networks. Section 5.6 concludes the chapter.

5.2 System Model

With $[n_1 : n_2]$ we denote the set of integers from n_1 to $n_2 \geq n_1$; $\emptyset$ is the empty set; $\mathbb{1}_P$ is the indicator function; 0^N is the all-zero vector of length N; $|A|$ is the absolute value of A when A is a scalar, and the cardinality when A is a set. For a matrix A, the notation $A_{R,C}$ indicates the submatrix of A where only the rows indexed by the set R and the columns indexed by the set C are retained.

We consider a Gaussian *1-2-1* network with $N+1$ nodes and multicast traffic. Node 0 is the source node and it wishes to communicate a common message to a set of destinations indexed by the set $\mathcal{D} \subseteq [1 : N]$. The remaining nodes $[1 : N] \backslash \mathcal{D}$ are relays that assist the communication between the source and the set of destinations. We assume that nodes that belong to $\mathcal{D}$ can act as relays for each other, i.e., if $i \in \mathcal{D}$, then node i can also be transmitting.

In a 1-2-1 network, at any particular time, a node can only direct (beamform) its transmissions towards at most another node. Similarly, a node can only receive transmissions from at most another node (to which its receive beam points towards). Thus, each node $i \in [0:N]$ in the network is characterized by two states, namely $S_{i,t}$ and $S_{i,r}$ that represent the node towards which node i is beamforming its transmissions and the node towards which node i is pointing its receive beam, respectively. In particular, $\forall i \in [0:N]$, we have that

$$\begin{aligned} S_{i,t} &\subseteq [1:N]\backslash\{i\}, \quad |S_{i,t}| \leq 1, \\ S_{i,r} &\subseteq [0:N]\backslash\{i\}, \quad |S_{i,r}| \leq 1, \end{aligned} \tag{5.1a}$$

where $S_{0,r} = \emptyset$ since the source node only transmits. We consider two modes of operation at the nodes, namely FD and HD. In FD, node $i \in [1:N]$ can be simultaneously receiving and transmitting, i.e., we can have both $S_{i,t} \neq \emptyset$ and $S_{i,r} \neq \emptyset$. In HD, node $i \in [1:N]$ can either receive or transmit, i.e., if $S_{i,t} \neq \emptyset$, then $S_{i,r} = \emptyset$ and vice versa. In particular, $\forall i \in [1:N]$, we have that

$$|S_{i,t}| + |S_{i,r}| \leq \begin{cases} 2 & \text{if nodes operate in FD} \\ 1 & \text{if nodes operate in HD} \end{cases}. \tag{5.1b}$$

We can now write the memoryless channel model for this Gaussian 1-2-1 network. We have that $\forall j \in [1:N]$

$$Y_j = Z_j + \sum_{i \in [0:N]\backslash\{j\}} h_{ji} \mathbb{1}_{\{i \in S_{j,r},\ j \in S_{i,t}\}} X_i, \tag{5.2}$$

where: (i) $S_{i,t}$ and $S_{i,r}$ are defined in (5.1); (ii) X_i (respectively, Y_i) denotes the channel input (respectively, output) at node i; (iii) $h_{ji} \in \mathbb{C}$ represents the complex channel coefficient from node i to node j; the channel coefficients are assumed to be constant for the whole transmission duration and known by the network; (iv) the channel inputs are subject to an individual power constraint, i.e., $\mathbb{E}[|X_k|^2] \leq P,\ k \in [0:N]$; (v) $Z_j,\ j \in [1:N]$ indicates the additive white Gaussian noise at the j-th node; noises across the network are assumed to be independent and identically distributed as $\mathcal{CN}(0,1)$.

Remark 5.2.1. Although we believe that 1-2-1 networks capture the essence of mmWave communications and enable to build useful insights on near-optimal information flow algorithms, we recognize that this model makes a number of simplifying assumptions that include: 1) we assume no interference among communication links (a reasonable assumption for relays spaced further apart than the beam width), and 2) we do not take into account the overhead of channel knowledge, and of beam-steering.

5.3 Multicast Approximate Capacity

In this section, we prove an approximation of the multicast capacity for the Gaussian 1-2-1 network model in (5.2). The multicast capacity[2] $\mathsf{C}^{\text{multi}}$ of the network defined in (5.2) is not known in general. However, a constant gap approximation of the multicast capacity can be developed as we show in this section.

The notion of a capacity approximation for a Gaussian network provides an expression that is guaranteed to be at most an additive constant gap away from the multicast capacity. In particular, let $\mathsf{C}^{\text{multi}}(\mathbf{h})$ be the multicast capacity of the N-node Gaussian 1-2-1 network with channel coefficients $\mathbf{h} = \{h_{ji} | i \in [0:N], j \in [1:N] \backslash \{i\}\}$. Then, an expression $\widetilde{\mathsf{C}}_{\text{multi}}(\mathbf{h})$ is said to be a constant gap approximation of the multicast capacity $\mathsf{C}^{\text{multi}}(\mathbf{h})$, if there exists a value (additive gap) $\mathsf{GAP}(N)$, that only depends on N and is independent of the channel coefficients or SNR such that

$$\left| \widetilde{\mathsf{C}}^{\text{multi}}(\mathbf{h}) - \mathsf{C}^{\text{multi}}(\mathbf{h}) \right| \leq \mathsf{GAP}(N). \tag{5.3}$$

For brevity, in the remainder of the chapter, we drop the explicit dependence of $\mathsf{C}^{\text{multi}}$ on the channel coefficients $\mathbf{h}$.

In order to develop such an expression, we first need to rewrite an equivalent expression for the channel model in (5.2) such that the switching parameters can be naturally incorporated

[2]We use standard definitions for codes, achievable rates and capacity [CT12].

into the channel input-output relation. Using an approach inspired by a similar analysis in [Kra04], the channel model in (5.2) can be modified to incorporate the state variables in the channel inputs. In particular, let the vector $\widehat{X}_i = (S_i, \overline{X}_i)$ be the input to the channel at node $i \in [0:N]$, where: (i) $S_i = (S_{i,t}, S_{i,r})$ with $S_{i,t}$ and $S_{i,r}$ being defined in (5.1), and (ii) $\overline{X}_i \in \mathbb{C}^N$, with elements $\overline{X}_i(k)$ defined as

$$\overline{X}_i(k) = X_i \mathbb{1}_{\{k \in S_{i,t}\}}. \tag{5.4}$$

In other words, $\overline{X}_i$ as a vector is a function of $S_{i,t}$ and the original input of the channel X_i. When node i is not transmitting, i.e., $S_{i,t} = \emptyset$, then $\overline{X}_i = 0^N$. It is not hard to see that the power constraint on X_i extends to $\overline{X}_i$ since at most one single index appears in the vector (recall that $|S_{i,t}| \leq 1$). Using this new channel input $\widehat{X}_i$, we can now equivalently rewrite the channel model in (5.2) as

$$Y_j = \begin{cases} h_{jS_{j,r}} \overline{X}_{S_{j,r}}(j) + Z_j & \text{if } |S_{j,r}| = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{5.5}$$

Using this equivalent modified representation, we can now show that the multicast capacity $\mathsf{C}^{\text{multi}}$ of the network defined in (5.4) and (5.5) can be approximated to within a constant-gap as stated in Theorem 5.3.1 below.

Theorem 5.3.1. *The multicast capacity* $\mathsf{C}^{\text{multi}}$ *of the network defined in* (5.4) *and* (5.5) *can be lower and upper bounded as*

$$\mathsf{C}^{\text{multi}}_{\text{cs,iid}} \leq \mathsf{C}^{\text{multi}} \leq \mathsf{C}^{\text{multi}}_{\text{cs,iid}} + \mathsf{GAP}(N), \tag{5.6a}$$

$$\mathsf{C}^{\text{multi}}_{\text{cs,iid}} = \max_{\substack{\lambda_s : \lambda_s \geq 0 \\ \sum_s \lambda_s = 1}} \min_{d \in \mathcal{D}} \underbrace{\min_{\substack{\Omega \subseteq [0:N]: 0 \in \Omega, \\ d \in \Omega^c}} \sum_{\substack{(i,j): i \in \Omega, \\ j \in \Omega^c}} \left(\sum_{\substack{s: \\ j \in s_{i,t}, \\ i \in s_{j,r}}} \lambda_s \right) \ell_{j,i}}_{\mathsf{C}_{\text{cs,iid}}(d,\lambda)}, \tag{5.6b}$$

$$\ell_{j,i} = \log\left(1 + P\,|h_{ji}|^2\right), \tag{5.6c}$$

$$\begin{aligned}\mathsf{GAP}(N) &= \mathsf{G}_1(N) + \mathsf{G}_2(N) + \mathsf{G}_3(N) \\ &= (N+1)\log e + \log(N+1) + N\log(\kappa), \end{aligned} \tag{5.6d}$$

where: (i) according to the definition (5.3)*,* $\mathsf{C}^{\text{multi}}_{\text{cs,iid}}$ *is a constant gap approximation of* $\mathsf{C}^{\text{multi}}$*; (ii)* $\Omega^c = [0:N]\backslash\Omega$*; (iii)* $\lambda_s = \mathbb{P}(S_{[0:N]} = s)$ *is the joint distribution of the states, where* s *enumerates the possible network states* $S_{[0:N]}$*; (iv)* $\mathsf{C}_{\text{cs,iid}}(d,\lambda)$ *is the achievable rate towards destination* $d \in \mathcal{D}$ *using* $\lambda = \{\lambda_s | s \in \mathcal{S}\}$ *where* $\mathcal{S}$ *is the set of all possible feasible network states; (v)* κ *is the number of different values that the state of a single relay/destination node can take, and is defined as*

$$\kappa = \begin{cases} N(N+1) & \textit{if nodes operate in FD} \\ 2N & \textit{if nodes operate in HD} \end{cases}. \tag{5.6e}$$

Proof. The proof relies on upper bounding the cut-set bound of the input-output relations in (5.4) and (5.5) by using independent inputs to the channel (hence the iid subscript). Moreover, the expression in (5.6b) is achieved through deterministic schedules, which do not depend on the transmitted message. The proof details and machinery are delegated to Appendix 5.8. □

Remark 5.3.1. The approximate capacity in (5.6b) seeks to maximize the minimum source-destination information flow. This can be computed as the minimum approximate capacity across all network cuts from the source to destination d, as captured by the minimization over the cuts Ω in (5.6b). Moreover, the nodes receive/transmit beams can be steered so as to maximize the information flow, and this is the role of the maximization over the λ_s variables in (5.6b). Note that in (5.6b) each link capacity is multiplied by the sum of the time fractions for which that link is active: since a link can be activated by more than one network state, we need to consider the total time duration a link is active.

Remark 5.3.2. The variable λ_s can be interpreted as the fraction of time for which the state configuration s is active. The collection of λ_s, $\forall s$ forms a joint probability distribution over

the state variables, i.e., $\sum_s \lambda_s = 1$. We refer to a particular feasible collection $\lambda = \{\lambda_s | s \in \mathcal{S}\}$ as a *schedule*. Note that sometimes, we drop the enumeration of $s \in \mathcal{S}$ and for brevity, we use $\lambda = \{\lambda_s\}$.

Remark 5.3.3. The variable GAP in (5.6d) only depends on the number of nodes N and represents the maximum loss incurred by using independent inputs and deterministic schedules at the nodes. In particular, G_1 represents the beamforming loss due to the use of independent inputs, while G_2 (respectively, G_3) accounts for the loss incurred by using a fixed schedule at the source (respectively, at the relays/destinations), as we explain next. Note that from (5.4), the input at the i-th node is also characterized by the state random variable $S_{i,t}$ (which indicates to which node – if any – node i is transmitting). Therefore, information can be conveyed from the source to the destinations by randomly switching between these states. However, as first highlighted in [Kra04] in the context of the HD relay channel, this random switch can only improve the capacity by a constant, whose maximum value equals the logarithm of the cardinality of the support of the state random variable. It therefore follows that the capacity can be approximated to within this constant by using a fixed/deterministic schedule at the nodes. In particular, for the source the cardinality of the support of its state random variable equals $N+1$ (since the source can only be transmitting to at most one node). Differently, the cardinality of the support of the state random variable at the remaining nodes depends on the mode of operation (either FD or HD) and is given by (5.6e).

In other words, $\mathsf{C}^{\text{multi}}_{\text{cs,iid}}$ in (5.6) – which can be achieved using quantize-map-and-forward as in [OD13] or noisy network coding as in [LKE11] – is a constant additive gap away from the multicast capacity $\mathsf{C}^{\text{multi}}$ of the network defined in (5.4) and (5.5). Thus, in the rest of the chapter we analyze $\mathsf{C}^{\text{multi}}_{\text{cs,iid}}$, which we refer to as the *approximate* multicast capacity for the Gaussian 1-2-1 network.

Note that the inner minimization in (5.6b) is the min-cut over a graph with source node

0, destination node d and edge capacities $\ell_{j,i}^{(s)}$, where

$$\ell_{j,i}^{(s)} = \left(\sum_{\substack{s: \\ j \in s_{i,t},\ i \in s_{j,r}}} \lambda_s \right) \ell_{j,i}. \tag{5.7}$$

Thus, for a fixed $d \in \mathcal{D}$ and a fixed schedule $\lambda = \{\lambda_s\}$, we can replace $\mathsf{C}_{\text{cs,iid}}(d, \lambda)$ in (5.6b) with its equivalent max-flow formulation, and we obtain

$$\begin{aligned} \text{P0} : \mathsf{C}_{\text{cs,iid}}^{\text{multi}} = \max_{\substack{\lambda_s : \lambda_s \geq 0 \\ \sum_s \lambda_s = 1}} \min_{d \in \mathcal{D}} \max_{\{F_{d,j}^{(d)}\}} \sum_{j \in [0:N] \setminus d} F_{d,j}^{(d)} \\ 0 \leq F_{j,i}^{(d)} \leq \ell_{j,i}^{(s)} \quad \forall (i,j) \in [0:N] \times [1:N], d \in \mathcal{D} \\ \sum_{j \in [1:N] \setminus \{i\}} F_{j,i}^{(d)} = \sum_{k \in [0:N] \setminus \{i\}} F_{i,k}^{(d)} \quad \forall i \in [1:N] \setminus \{d\}, d \in \mathcal{D}, \end{aligned} \tag{5.8}$$

where $F_{j,i}^{(d)}$ is the information flow from node i to node j when the destination is node d.

We can reorganize the max-min-max optimization in P0 to be a max-min optimization as we can see in the following lemma.

Lemma 5.3.1. *The LP* P0 *defined in* (5.8) *is equivalent to the following LP*

$$\begin{aligned} \text{P1} : \mathsf{C}_{\text{cs,iid}}^{\text{multi}} = \max_{\lambda_s, \mathbf{F}} \min_{d \in \mathcal{D}} \left\{ \sum_{j \in [0:N] \setminus d} F_{d,j}^{(d)} \right\} \\ (\text{P1}a) \quad 0 \leq F_{j,i}^{(d)} \leq \ell_{j,i}^{(s)} \qquad \forall (i,j) \in [0:N] \times [1:N], d \in \mathcal{D} \\ (\text{P1}b) \quad \sum_{j \in [1:N] \setminus \{i\}} F_{j,i}^{(d)} = \sum_{k \in [0:N] \setminus \{i\}} F_{i,k}^{(d)} \quad \forall i \in [1:N] \setminus \{d\}, d \in \mathcal{D} \\ (\text{P1}c) \quad \sum_s \lambda_s = 1, \\ (\text{P1}d) \quad \lambda_s \geq 0 \qquad \forall s, \end{aligned} \tag{5.9}$$

where $\mathbf{F} = \bigcup_{d \in \mathcal{D}} \{F_{d,j}^{(d)}\}$.

Proof. Note that, if we can exchange the inner min-max with max-min in the optimization problem P0 in (5.8), then we have proved its equivalence to P1. We start by noting that, by

the max-min inequality, we have

$$\min_{d\in\mathcal{D}} \max_{\mathbf{F}} \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)} \geq \max_{\mathbf{F}} \min_{d\in\mathcal{D}} \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)}. \tag{5.10}$$

Thus, to prove equality between the two sides of (5.10) (and consequentially the equivalence between P0 and P1), we need to prove that

$$\min_{d\in\mathcal{D}} \max_{\mathbf{F}} \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)} \leq \max_{\mathbf{F}} \min_{d\in\mathcal{D}} \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)}. \tag{5.11}$$

This comes as a direct consequence of the result in the following Lemma 5.3.2 by considering $\mathcal{I}=\mathcal{D}$, $x=\mathbf{F}$, $x_i=\{F_{i,j}^{(i)}\}$ and $f_i(x)=\sum_{j\in[0:N]\setminus i} F_{i,j}^{(i)}$. This proves Lemma 5.3.1. □

Lemma 5.3.2. *Let $\mathcal{I}$ be a discrete set, $\{\mathcal{X}_i\}_{i\in\mathcal{I}}$ be a collection of closed sets indexed by $\mathcal{I}$ and $\mathcal{X}=\prod_{i\in\mathcal{I}}\mathcal{X}_i$. Consider the set of bounded functions $\{f_i(\cdot)\}_{i\in\mathcal{I}}$, $f_i:\mathcal{X}\to\mathbb{R}$ such that $f_i(\cdot)$ depends only on $\mathcal{X}_i$. Then, we have*

$$\min_{i\in\mathcal{I}} \max_{x\in\mathcal{X}} f_i(x) \leq \max_{x\in\mathcal{X}} \min_{i\in\mathcal{I}} f_i(x). \tag{5.12}$$

Proof. For any $x\in\mathcal{X}$, we can write it as $x=[x_1,x_2,\cdots,x_{|\mathcal{I}|}]$, such that $x_i\in\mathcal{X}_i$. Since the value of $f_i(\cdot)$ depends only on x_i, then we can define $x_i^\star$ as the value of x_i that maximizes f_i. Thus, for $x^\star=[x_1^\star,x_2^\star,\cdots,x_{|\mathcal{I}|}^\star]\in\mathcal{X}$, we have that

$$\max_{x\in\mathcal{X}} f_i(x) = f_i(x^\star).$$

Furthermore, we have that

$$\max_{x\in\mathcal{X}} \min_{i\in\mathcal{I}} f_i(x) \overset{(a)}{\geq} \min_{i\in\mathcal{I}} f_i(x^\star) = \min_{i\in\mathcal{I}} \max_{x\in\mathcal{X}} f_i(x), \tag{5.13}$$

where the inequality in (a) follows from the fact that we are considering a particular $x\in\mathcal{X}$. This proves Lemma 5.3.2. □

Remark 5.3.4. Note that from (5.6b) and (5.9), the approximate multicast capacity has the same general expression both for Gaussian FD and HD 1-2-1 networks. The main difference

lies in the admissible network states in FD and HD. In FD, admissible states allow for two adjacent links (i.e., links with a common node) to be simultaneously active as long as a node is not playing the same role in both links (e.g., cannot be the transmitter in both links). In contrast, HD admissible states do not allow any adjacent links to be simultaneously active.

The expression of the approximate multicast capacity as the linear program P1 in (5.9) has a number of constraints that is polynomial in the number of nodes N. However, the number of variables (particularly, the number of network states) is exponential, which makes P1 computationally expensive to be solved directly. In what follows, we show how - in the FD mode of operation - this linear program can be solved in polynomial time to compute the approximate multicast capacity and a schedule λ that is optimal for the approximate capacity. Tackling the same problem in the HD mode of operation is more involved than the FD case, and as a result will be studied in the next chapter.

5.4 Scheduling of Gaussian FD 1-2-1 Networks

In this section, we investigate the problem of finding the approximate multicast capacity and an optimal schedule for the Gaussian FD 1-2-1 network. Towards this end, we develop an approach to solving the linear program P1 in (5.9) in polynomial time in the number of nodes for the FD mode of operation. In particular, our main result is given by the following theorem.

Theorem 5.4.1. *For the N-relay Gaussian FD 1-2-1 network, we have:*

1. *The approximate multicast capacity $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ can be found in polynomial time in N;*
2. *An optimal schedule for the approximate capacity $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ can be found in polynomial time in N.*

The main approach to prove Theorem 5.4.1 consists of developing an equivalent formulation for the LP P1 in (5.9) for Gaussian FD 1-2-1 networks. In particular, the equivalent

formulation is given by the following theorem.

Theorem 5.4.2. *For any N-relay Gaussian FD 1-2-1 network, we have that* P1 *in* (5.9) *is equivalent to the LP below*

$$\mathrm{P2}_{\mathrm{FD}}: \mathsf{C}_{\mathrm{cs,iid}}^{\mathrm{multi}} = \max_{\boldsymbol{\alpha},\mathbf{F}} \min_{d\in\mathcal{D}} \left\{ \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)} \right\}$$

$$\begin{aligned}
&(\mathrm{P2}a) \quad 0 \le F_{j,i}^{(d)} \le \alpha_{ji}\ell_{j,i} && \forall (i,j)\in[0:N]\times[1:N], d\in\mathcal{D},\\
&(\mathrm{P2}b) \quad \sum_{j\in[1:N]\setminus\{i\}} F_{j,i}^{(d)} = \sum_{k\in[0:N]\setminus\{i\}} F_{i,k}^{(d)} && \forall i\in[1{:}N]\setminus\{d\}, d\in\mathcal{D}\\
&(\mathrm{P2}c) \quad \sum_{j\in[1:N]\setminus\{i\}} \alpha_{ji} \le 1 && \forall i\in[0:N], \qquad (5.14)\\
&(\mathrm{P2}d) \quad \sum_{i\in[0:N]\setminus\{j\}} \alpha_{ji} \le 1 && \forall j\in[1:N],\\
&(\mathrm{P2}e) \quad \alpha_{ji} \ge 0 && \forall (i,j)\in[0:N]\times[1:N],
\end{aligned}$$

where $F_{j,i}^{(d)}$ represents the data flow through the link of capacity $\ell_{j,i}$ intended for destination d, α_{ji} represents the fraction of time for which the link is active and $\alpha = \{\alpha_{ji}|\ \forall i\in[0:N], j\in[1:N]$ and $i\neq j\}$.

Remark 5.4.1. Before delving into the proof of Theorem 5.4.2, we highlight interesting consequences of the LP $\mathrm{P2}_{\mathrm{FD}}$.

(a) Since $\mathrm{P2}_{\mathrm{FD}}$ has a polynomial number of variables and constraints in N, then Theorem 5.4.1 Part (a) is a direct consequence of the structure of $\mathrm{P2}_{\mathrm{FD}}$.

(b) If the mapping from an optimal point in $\mathrm{P2}_{\mathrm{FD}}$ to an optimal point in P1 can be done in polynomial time, then we have a polynomial time algorithm to find the optimal schedule for the approximate multicast capacity in the Gaussian FD 1-2-1 network. This can be done by first solving $\mathrm{P2}_{\mathrm{FD}}$ in polynomial time and then mapping its optimal solution in polynomial time to an optimal schedule in P1. This would prove Theorem 5.4.1 Part (b). In what follows, we indeed show that such a mapping from an optimal point in $\mathrm{P2}_{\mathrm{FD}}$ to P1 can be done by a construction that is polynomial in N.

We next prove Theorem 5.4.2 by showing that the LPs P1 and P2_{FD} are indeed equivalent for Gaussian FD 1-2-1 networks. Moreover, we also show that the mapping from an optimal point in P2_{FD} to P1 can be performed in polynomial time, hence proving Theorem 5.4.1 Part (b). Note that in P1 and P2_{FD}, the variables $F_{j,i}^{(d)}$ are the same, and hence we only need to find the mapping between $\{\lambda_s\}$ and $\{\alpha_{ji}\}$.

$\underline{\text{P1} \to \text{P2}_{\text{FD}}.}$ Given a feasible point in P1 we define

$$\alpha_{ji} = \sum_{\substack{s: \\ j \in s_{i,t},\ i \in s_{j,r}}} \lambda_s.$$

Using this definition, we have that

$$\text{(P1}a\text{)} \ : \forall (i,j) \quad F_{j,i}^{(d)} \leq \left(\sum_{\substack{s: \\ j \in s_{i,t},\ i \in s_{j,r}}} \lambda_s \right) \ell_{j,i} = \alpha_{ji} \ell_{j,i} \qquad \implies \text{(P2}a\text{)}$$

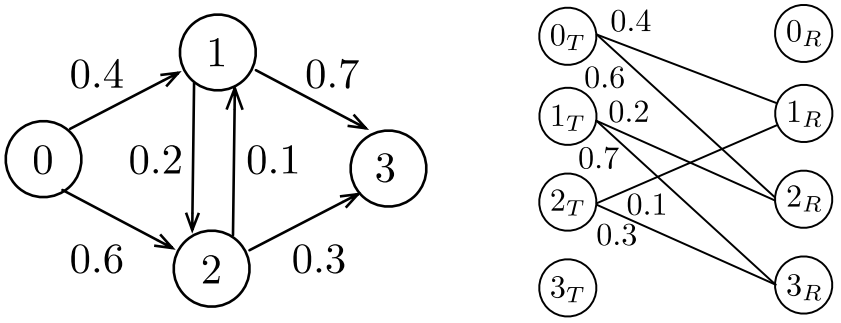
$$\text{(P1}c\text{)} \ : \forall i \in [0:N] \quad \sum_{j \in [1:N] \setminus \{i\}} \alpha_{ji} = \sum_{j \in [1:N] \setminus \{i\}} \sum_{\substack{s: \\ j \in s_{i,t},\ i \in s_{j,r}}} \lambda_s \leq \sum_s \lambda_s = 1 \qquad \implies \text{(P2}c\text{)}$$

$$\text{(P1}c\text{)} \ : \forall j \in [1:N] \quad \sum_{i \in [0:N] \setminus \{j\}} \alpha_{ji} = \sum_{i \in [0:N] \setminus \{j\}} \sum_{\substack{s: \\ j \in s_{i,t},\ i \in s_{j,r}}} \lambda_s \leq \sum_s \lambda_s = 1 \qquad \implies \text{(P2}d\text{)}.$$

In addition, since the variables $F_{j,i}^{(d)}$ are not changed in the mapping then the new mapped point in P2_{FD} has the same objective value as the original point in P1.

$\underline{\text{P2}_{\text{FD}} \to \text{P1}.}$ Given a feasible point in P2_{FD} we would like to construct a set of λ_s that represent the state activation times in the Gaussian FD 1-2-1 network, such that, they collectively activate each link (i,j) for at least the fraction dictated by α_{ji}. If this is satisfied, then the same flow variables $F_{j,i}^{(d)}$ from P2_{FD} can be used in P1 and thus achieve the same objective function value in P1.

To map P2_{FD} to P1, we introduce a visualization for P2_{FD} in terms of *bipartite* graphs. In particular, we divide each node $i \in [0:N]$ in the network into two vertices (i_T and i_R) representing the transmitting and receiving functions of the node; note that $0_R = \emptyset$ since the source (node 0) is always transmitting. This gives us the complete bipartite graph



(a) Example Gaussian FD 1-2-1 network with link activation times λ_{ji}.

(b) Constructed bipartite graph G_B.

Figure 5.2: An example of the construction of a bipartite graph based on the link activations.

$G_B = (T_G, R_G, E_G)$, where the vertices T_G (respectively, R_G) are the transmitting modules of our nodes (respectively, R_G collects our receiving modules), and we have a weighted edge $(i_T, j_R) \in E_G$ with weight α_{ji}, $\forall (i,j) \in [0:N] \times [0:N]$. Note that we assume $\alpha_{0i} = 0$, $\forall i \in [0:N]$. Fig. 5.2(b) shows the construction of G_B for the network example in Fig. 5.2(a). With such a representation, we can now write a feasible point in $\mathrm{P2_{FD}}$ as a weighted adjacency matrix L of the graph G_B. For instance, the weighted adjacency matrix L of the graph G_B in Fig. 5.2(b) is given by

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0.1 & 0 \\ 0.6 & 0.2 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0 \end{bmatrix}.$$

From the constraints (P2c)-(P2e), the matrix L inherites the following properties

$$\begin{aligned} \forall (i,j) \in [1:N+1]^2,\ [L]_{ji} \geq 0, \\ \forall i \in [1:N+1],\ \sum_{j \in [1:N+1]} [L]_{ji} \leq 1, \\ \forall j \in [1:N+1],\ \sum_{i \in [1:N+1]} [L]_{ji} \leq 1, \end{aligned} \tag{5.15}$$

$$\sum_{i\in[1:N+1]} [L]_{ii} = 0, \quad \text{(node } i \text{ cannot transmit to itself)},$$

$$\sum_{i\in[1:N+1]} [L]_{1i} = 0, \quad \text{(source does not receive)}.$$

It is not difficult to see that a valid state s in P1 corresponds to some matching[3] $M_s \subseteq E_G$ in the bipartite graph G_B. On the other hand, a matching M_s in G_B corresponds to a state s in P1 if $M_s \cap U = \emptyset$, where the set of rejected edges U is defined as

$$U = \{(i,j)|i = j \text{ or } i = 1\}. \tag{5.16}$$

In particular, the first condition defining U in (5.16) requires that edges connecting i_T to i_R are rejected, and the second condition requires that edges connecting to 0_R (the receiver of the source) are rejected. We refer to a matching M_s in G_B, such that $M_s \cap U = \emptyset$ as a *valid state matching*.

We now divert our attention to the set of perfect matchings[4] in G_B, which is represented by the set of *permutation* matrices of size $(N+1)\times(N+1)$. For a perfect matching $\widehat{M}_p$, we define $M_{s(p)}$ to be the largest size valid state matching such that $M_{s(p)} \subseteq M_p$. We note that $s(p)$ is unique and can be obtained by systematically removing the edges in $M_p \cap U$.

Remark 5.4.2. Note that, for all edges $(i,j) \in U$ defined in (5.16), we have that $[L]_{j+1,i+1} = 0$ as defined in (5.15).

Without loss generality, we can represent a matching M_s by an $(N+1)\times(N+1)$ matrix A_s, such that $[A_s]_{ji} = 1$ if $(i,j) \in M_s$ and zero otherwise. Thus, the mapping that we need from an optimal point in P2_{FD} to P1 can be performed by efficiently finding a set of K valid state matchings $\{M_{s_k}\}_{k=1}^K$ such that

$$L \leq \sum_{k\in[1:K]} \lambda_{s_k} A_{s_k}, \ \sum_{k\in[1:K]} \lambda_{s_k} = 1, \ \lambda_{s_k} \geq 0. \tag{5.17}$$

[3]A matching is a set of graph edges such that each vertex (node) in the graph has at most one edge from the set connected to it [BM11].

[4]A perfect matching is a matching such that all vertices in the graph are connected to one edge in the matching set [BM11].

Instead of doing the decomposition above, we aim at finding a set of K permutation matrices $\{\Pi_p\}_{p=1}^{K}$, each representing a perfect matching in G_B, that satisfy

$$L \leq \sum_{p\in[1:K]} \varphi_p \Pi_p, \quad \sum_{p\in[1:K]} \varphi_p = 1, \ \varphi_p \geq 0. \tag{5.18}$$

Note that (5.18) directly implies (5.17) by using $s_k = s(p)$ as the index of A_{s_k} and $\lambda_{s(p)} = \varphi_p$, $\forall p \in [1:K]$. This is due to the fact that $A_{s(p)}$ and Π_p only differ in the positions indexed by the edges in U (see (5.16)), but using Remark 5.4.2, all such positions in the matrix L are zero. The non-negativity of $A_{s(p)}$ and $\lambda_{s(p)}$ in addition to the aforementioned observation implies (5.17). Now, the explicit problem to solve is the existence of an efficient approach to perform the decomposition in (5.18), i.e., can we efficiently find a set of K permutation matrices $\{\Pi_p\}_{p=1}^{K}$, each representing a perfect matching, that satisfy (5.18)? In particular, in order to construct an optimal schedule for Gaussian FD 1-2-1 networks in polynomial time, we are interested in a polynomial time approach to find these Π_p matrices.

To answer the question above, we observe that the first three properties of L in (5.15) are the definition of a *doubly-substochastic matrix*. The existence of a convex decomposition for a doubly stochastic matrix (summation inequalities in (5.15) satisfied with equality) into permutation matrices is a classic result in mathematics by Birkhoff [Bir46] and a polynomial time algorithm in the matrix size exists to find this decomposition [Dul55]. The result in [CCH99] extends the algorithm in [Dul55] to doubly substochastic matrices. Thus, for our $L \in \mathbb{R}_{+}^{N+1\times N+1}$, a set of permutation matrices satisfying (5.18) can be found in polynomial time in N. The algorithm runs in $O(N^{4.5})$ and outputs N^2+1 permutation matrices. This proves the existence of a mapping from P2_{FD} to P1 that can be done in polynomial time, thus concluding the proof of Theorem 5.4.2 and Theorem 5.4.1 Part (b) for Gaussian FD 1-2-1 networks.

5.5 Multicast vs Minimum Unicast Approximate Capacity in Gaussian FD 1-2-1 Networks

In this section, we explore the relation between the approximate multicast and unicast capacities of Gaussian FD 1-2-1 networks. Towards this end, we focus on characterizing the ratio between the approximate multicast capacity and the minimum approximate unicast capacity of Gaussian FD 1-2-1 networks. We define the approximate unicast capacity for destination $d \in \mathcal{D}$ as the FD approximate multicast capacity when d is the only destination in the network. Thus, from $\mathrm{P2_{FD}}$ in Theorem 5.4.2, the approximate unicast capacity $\mathsf{C}^d_{\mathrm{cs,iid}}$ for destination $d \in \mathcal{D}$ is

$$
\begin{aligned}
\mathrm{P4}_d : \mathsf{C}^d_{\mathrm{cs,iid}} = \max_{\mathbf{F},\ \lambda} & \left\{ \sum_{j \in [0:N]\setminus d} F^{(d)}_{d,j} \right\} \\
& 0 \le F^{(d)}_{j,i} \le \alpha^{(d)}_{ji} \ell_{j,i} \qquad \forall (i,j) \in [0:N] \times [1:N] \\
& \sum_{j \in [1:N]\setminus\{i\}} F^{(d)}_{j,i} = \sum_{k \in [0:N]\setminus\{i\}} F^{(d)}_{i,k} \qquad \forall i \in [1:N] \\
& \sum_{j \in [1:N]\setminus\{i\}} \alpha^{(d)}_{ji} \le 1, \qquad \forall i \in [0:N] \\
& \sum_{k \in [0:N]\setminus\{i\}} \alpha^{(d)}_{ik} \le 1, \qquad \forall i \in [1:N] \\
& \alpha^{(d)}_{ji} \ge 0 \qquad \forall (i,j) \in [0:N] \times [1:N].
\end{aligned} \tag{5.19}
$$

Note that the superscript (d) is fixed throughout the LP and is included to specialize the variables used in computing $\mathsf{C}^d_{\mathrm{cs,iid}}$.

We now define the following parameters for our network:

- C_{mu}: minimum unicast approximate capacity from the source to the destinations, i.e.,

$$\mathsf{C}_{\mathrm{mu}} = \min_{d \in \mathcal{D}} \mathsf{C}^d_{\mathrm{cs,iid}}; \tag{5.20}$$

- Δ^+: maximum number of incoming links (with non-zero capacity) to a node in the 1-2-1 network;

- Δ^-: maximum number of outgoing links (with non-zero capacity) from a node in the 1-2-1 network.

With these definitions, we can now state the following theorem that relates C_{mu} and $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ for Gaussian FD 1-2-1 networks.

Theorem 5.5.1. *For a Gaussian FD 1-2-1 network with destination set $\mathcal{D}$, we have that*

$$\mathsf{C}_{\text{cs,iid}}^{\text{multi}} \geq \frac{1}{\min\left\{|\mathcal{D}|,\ \max\{\Delta^+, \Delta^-\}\right\}} \mathsf{C}_{\text{mu}}, \tag{5.21}$$

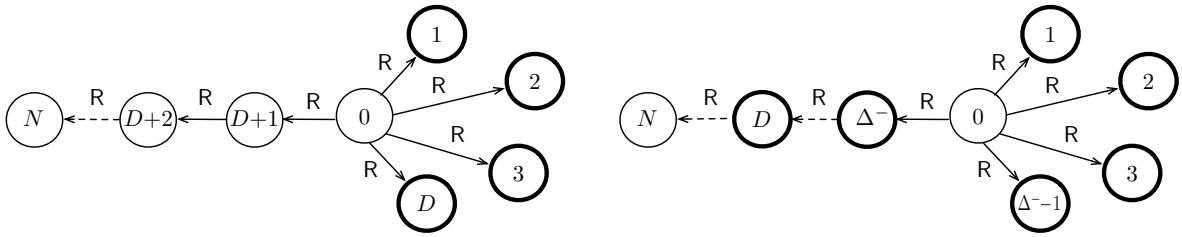
where C_{mu} is defined in (5.20)*. Furthermore, there exists a class of networks for which this ratio is tight.*

Proof. Without loss of generality, the destination nodes are indexed by $\{1, 2, \cdots, D\}$, with $D = |\mathcal{D}|$. The key intuition behind the worst-case ratio in Theorem 5.5.1 is that, when the destinations are spread out in different places in the network (e.g., in Fig. 5.3(a) and Fig. 5.3(b)), the network scheduling needs to balance the amount of traffic to be delivered to each destination. Thus, because of this, the approximate multicast capacity decreases. In what follows, we formalize this notion by considering two different cases, namely $|\mathcal{D}| \leq \max\{\Delta^+, \Delta^-\}$ and $|\mathcal{D}| > \max\{\Delta^+, \Delta^-\}$, respectively. In each of the two cases, we show that there exists a feasible schedule (in terms of link activation times) in (5.14) that achieves the bound in (5.21). Moreover, we also present network examples for which the ratio guarantee in (5.21) is indeed tight.

Case 1: $|\mathcal{D}| \leq \max\{\Delta^+, \Delta^-\}$. In this particular case, $\forall d \in \mathcal{D}$, let $\{\alpha_{ji}^{(d)\star}\}$ be an optimal schedule in P4_d for the approximate unicast capacity from the source to destination d. We can define a feasible schedule for the LP P2_{FD} in (5.14) as

$$\alpha'_{ji} = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \alpha_{ji}^{(d)\star} \qquad \forall (i,j) \in [0:N] \times [1:N].$$

In other words, for multicast traffic, we timeshare the network with the optimal schedule for each of the destinations $d \in \mathcal{D}$. Let $\{F_{j,i}^{\prime(d)}\}$ be the optimal flow variables that maximize



(a) $|\mathcal{D}| \leq \max\{\Delta^+, \Delta^-\}$. (b) $|\mathcal{D}| > \max\{\Delta^+, \Delta^-\}$.

Figure 5.3: Networks with tight ratio in Theorem 5.21. The $D = |\mathcal{D}|$ destinations are shown with bold border.

the objective function for the fixed schedule $\{\alpha'_{ji}\}$ in $\mathrm{P2_{FD}}$. By the timesharing argument, it is not difficult to see that for all destinations, we have that the evaluation of the objective function in $\mathrm{P2_{FD}}$ using this timesharing schedule gives that

$$\sum_{j\in[0:N]\setminus d} F'^{(d)}_{d,j} \geq \frac{1}{|\mathcal{D}|} \sum_{j\in[0:N]\setminus d} F^{(d)\star}_{d,j} = \frac{1}{|\mathcal{D}|}\mathsf{C}^{d}_{\mathrm{cs,iid}} \quad \forall d \in \mathcal{D},$$

where $\{F^{(d)\star}_{d,j}\}$ are optimal for $\mathrm{P4}_d$. Since the computed $\{F'^{(d)}_{j,i}\}$ and $\{\alpha'_{ji}\}$ are feasible in the LP $\mathrm{P2_{FD}}$, then we have the desired ratio, i.e.,

$$\mathsf{C}^{\mathrm{multi}}_{\mathrm{cs,iid}} \geq \min_{d\in\mathcal{D}} \sum_{j\in[0:N]\setminus d} F'^{(d)}_{d,j} \geq \min_{d\in\mathcal{D}} \frac{1}{|\mathcal{D}|}\mathsf{C}^{d}_{\mathrm{cs,iid}} = \frac{1}{|\mathcal{D}|}\mathsf{C}_{\mathrm{mu}}.$$

To show that the bound is tight, consider the network topology illustrated in Fig. 5.3(a). It is not difficult to see that, for this particular network, $\mathsf{C}_{\mathrm{mu}} = \mathsf{R}$. For the multicast approximate capacity $\mathsf{C}^{\mathrm{multi}}_{\mathrm{cs,iid}}$, the source has to timeshare between the $|\mathcal{D}|$ destinations to achieve a rate of $\mathsf{R}/|\mathcal{D}|$. Thus,

$$\mathsf{C}^{\mathrm{multi}}_{\mathrm{cs,iid}} = \frac{1}{|\mathcal{D}|}\mathsf{C}_{\mathrm{mu}}.$$

This concludes the proof for the first case.

Case 2: $|\mathcal{D}| > \max\{\Delta^+, \Delta^-\}$. In this particular case, we define the multicast schedule as

$$\alpha'_{ji} = \frac{1}{\max\{\Delta^+, \Delta^-\}}\mathbb{1}_{\{\ell_{j,i}>0\}} \qquad \forall (i,j) \in [0:N]\times[1:N].$$

To show that this schedule is feasible in the LP $\mathrm{P2_{FD}}$, we note that for every node i in the network, we have that

$$\begin{aligned}\sum_{k\in[0:N]\setminus\{i\}}\alpha'_{ik} &= \sum_{k\in[0:N]\setminus\{i\}}\frac{1}{\max\{\Delta^+,\Delta^-\}}\mathbb{1}_{\{\ell_{i,k}>0\}}\\ &=\frac{1}{\max\{\Delta^+,\Delta^-\}}\sum_{k\in[0:N]\setminus\{i\}}\mathbb{1}_{\{\ell_{i,k}>0\}}\\ &=\frac{1}{\max\{\Delta^+,\Delta^-\}}\sum_{k\in N^+(i)}1=\frac{|N^+(i)|}{\max\{\Delta^+,\Delta^-\}}\overset{(a)}{\leq}1,\end{aligned}\tag{5.22}$$

where: (i) $N^+(i)=\{k\in[0:N]|\ell_{ik}>0\}$ is the set of neighboring nodes to i that have incoming edges into i with non-zero point-to-point link capacity; (ii) the inequality in (a) follows from the definition of Δ^+ that ensures that $|N^+(i)|\leq\Delta^+$, $\forall i\in[1:N]$. Using similar arguments, we can also show that

$$\sum_{j\in[1:N]\setminus\{i\}}\alpha'_{ji}\leq 1.$$

The analysis above proves that the constructed schedule α'_{ji} is feasible, i.e., it satisfies the constraints in $(2c)-(2e)$ in the LP $\mathrm{P2_{FD}}$. By fixing and substituting $\{\alpha'_{ji}\}$ in $\mathrm{P2_{FD}}$, we can now compute the objective value for the proposed schedule through this LP

$$\begin{aligned}&\mathsf{R}'_{\text{multicast}}=\max_{\mathbf{F}}\ \min_{d\in\mathcal{D}}\left\{\sum_{j\in[0:N]\setminus d}F^{(d)}_{d,j}\right\}\\ &0\leq F^{(d)}_{j,i}\leq\frac{1}{\max\{\Delta^+,\Delta^-\}}\ell_{j,i}\quad\forall(i,j)\in[0{:}N]\times[1{:}N],d\in\mathcal{D}\\ &\sum_{j\in[1:N]\setminus\{i\}}F^{(d)}_{j,i}=\sum_{k\in[0:N]\setminus\{i\}}F^{(d)}_{i,k}\quad\forall i\in[1{:}N],d\in\mathcal{D}.\end{aligned}\tag{5.23}$$

Note that the only variables in the LP in (5.23) are the flow variables $\{F^{(d)}_{j,i}\}$. Thus, in (5.23) we are effectively computing the multicast capacity in a wired network with link capacities

$$\ell'_{j,i}=\frac{1}{\max\{\Delta^+,\Delta^-\}}\ell_{j,i},\tag{5.24}$$

and hence

$$\mathsf{R}'_{\text{multicast}}=\frac{1}{\max\{\Delta^+,\Delta^-\}}\min_{d\in\mathcal{D}}\left\{\mathsf{C}^{(\text{wired})}_d\right\},\tag{5.25}$$

where $\mathsf{C}_d^{(\text{wired})}$ is the unicast capacity to destination d when we consider a wired network with the link capacities as in (5.24). In other words, the network has orthogonal links that can be activated for 100% of the time. Thus, it is not difficult to see that $\mathsf{C}_d \leq \mathsf{C}_d^{(\text{wired})}$, $\forall d \in \mathcal{D}$ and we have that

$$\begin{aligned}\mathsf{C}_{\text{cs,iid}}^{\text{multi}} \geq \mathsf{R}'_{\text{multicast}} &= \frac{1}{\max\{\Delta^+, \Delta^-\}} \min_{d\in\mathcal{D}} \left\{\mathsf{C}_d^{(\text{wired})}\right\} \\ &\geq \frac{1}{\max\{\Delta^+, \Delta^-\}} \min_{d\in\mathcal{D}} \{\mathsf{C}_d\} = \frac{1}{\max\{\Delta^+,\Delta^-\}} \mathsf{C}_{\text{mu}},\end{aligned}$$

which proves the lower bound in the second case. To show that the bound is indeed tight in this case, consider the network shown in Figure 5.3(b). For this particular case, it is not difficult to see that the unicast approximate capacity to each of the destinations is R. Furthermore, in multicast, the source needs to switch (equally) between the Δ^- different paths connected to it to serve the different destinations. Thus, we have

$$\mathsf{C}_{\text{cs,iid}}^{\text{multi}} = \frac{1}{\max\{\Delta^+, \Delta^-\}} \mathsf{C}_{\text{mu}}.$$

This concludes the proof of Theorem 5.5.1. □

Remark 5.5.1. The result in Theorem 5.5.1 highlights a fundamental difference of Gaussian FD 1-2-1 networks from classical Gaussian FD wireless relay networks. In particular, in classical wireless relay networks, the ratio between the approximate multicast capacity and the minimum approximate unicast capacity is universally equal to one. Differently, when 1-2-1 constraints are included, Theorem 5.5.1 shows that the ratio depends on the number of destinations as well as graph-theoretic properties of the network.

5.6 Conclusion

In this chapter, we introduced the Gaussian 1-2-1 network, an abstract information-theoretic model that captures the high directivity aspect of mmWave communication. For both FD and HD modes of operation, we showed that a constant additive gap approximation exists

for the multicast capacity. Based on the derived approximate capacity, we showed that operating a Gaussian FD 1-2-1 network in multicast can be potentially more costly compared to classical wireless relay networks as the ratio between the approximate multicast capacity and the minimum unicast capacity is not unity and can be affected by the network structure and the number of destinations.

5.7 Related Work

Studies of mmWave communication have focused on profiling the distribution of Signal-to-Interference-plus-Noise Ratio (SINR) in random environments both in cellular and ad-hoc network settings [DFP17, BAH14, TBH16]. mmWave communication in a cellular setting was studied in [DFP17, BAH14] using stochastic geometry to characterize the performance through SINR distributions in a random environment. A similar characterization for an ad-hoc connection between a pair of nodes has also been studied in [TBH16]. These previous works consider communication over a single-hop either between ad-hoc nodes or in a cellular system between a base station and a user equipment. The effectiveness of multi-hop relaying for routing in mmWave networks in random environments has been studied in [LA15,DTF17] where it was shown that multi-hop relaying through line-of-sight connections can improve the network connectivity. In [YPK03], scaling laws for networks with directional antennas were studied using the setting introduced in [GK00] for omnidirectional communication. These results look at order arguments for multiple unicast sessions through routing and do not explore fundamental bounds such as the information theoretical capacity.

Shannon capacity characterization for the classical[5] Gaussian relay network with isotropic transmissions is a long standing open problem, both in FD and HD modes of operation. Several schemes [ADT11], [OD13], [LKE11], [LKK14] have been shown to achieve a rate that

[5]The term "classical" is used throughout the chapter to describe wireless relay networks with isotropic antennas and transmissions. This term is adopted to differentiate such scenarios from the one analyzed in this chapter, where we have 1-2-1 constraints.

is a constant gap (i.e., a value that only depends on the number of nodes and is independent of the channel parameters and operating SNR) away from the Shannon capacity. This is accomplished by showing that the achieved rate is a constant gap away from the well-known cut-set upper bound [CE79] on the Shannon capacity. For general network topologies, the constant gap is fundamentally linear in the number of nodes N in the network [CO15, WO15], although for some specific FD network topologies, the constant gap can be shown to be sublinear [SWF12, CO12].

For classical Gaussian networks with N relays operating in HD, the approximate capacity characterization is challenging as it requires an additional optimization to schedule the relays over 2^N listen/transmit configuration states. Thus, although the approximate capacity can be computed in polynomial time in the network size for the FD mode [PE14], in HD such a result is known to hold only for a few special cases such as line networks [ECF17] and specific classes of layered networks [EPS14], [JEC19]. Furthermore, although there have been several works that characterize the complexity of the structure of the optimal schedule for Gaussian HD wireless relay networks [BFO16], [CTK14], the problem of efficiently (i.e., in polynomial time in the number of nodes) finding the schedule optimal for the approximate capacity for any general number of relays N has only been solved for Gaussian line networks [ECF17].

In networks with non-isotropic transmissions, a scheduling component naturally arises in optimizing the achieved rates. In [Ari84], it has been shown that the multi-access problem in ad-hoc wireless networks is NP-hard. In [JX06], scheduling in wireless networks under interference constraints has been studied and an approach for scheduling was proposed that is guaranteed to converge to the optimal point. Unfortunately the proposed approach runs in exponential time in the number of links in the network. For spread spectrum networks, it has been proved in [HS88] that link-based transmissions where the restriction is relaxed to disallowing a node to converse with more than one node, can be scheduled in polynomial time in the number of nodes. In [THM12, CCH07], scheduling of network coded flows was studied by modeling broadcast instances as hyperedges in a hypergraph model of the wireless

network. Optimal scheduling in the previous model was shown to be possible in polynomial time for a class of networks having claw-free conflict graphs [KM17]. Computing a schedule for broadcast transmissions has also been studied for crossbar switches [SMK07] and was shown to reduce to the fractional weighted graph coloring problem, which is NP-hard in general.

Different from the aforementioned thread of research, where the main objective is to provide an operating schedule for a network when all the N nodes/relays are active, another consideration in wireless networks is understanding what fraction can be guaranteed when only a subset of $k \leq N$ relays out of N is utilized. This line of work is usually referred to as *network simplification* and it was first studied in [NOF14] for classical Gaussian FD diamond networks[6]. Specifically, [NOF14] showed that in any N-relay classical Gaussian FD diamond network, there always exists a subnetwork of k relays that retains at least $\frac{k}{k+1}$ of the unicast approximate capacity of the full network. This guarantee was shown to be tight, i.e., there exist N-relay Gaussian FD diamond networks for which the best k-relay subnetwork (i.e., the one with the largest approximate capacity) achieves this fraction of the full network approximate FD capacity. In [ESF16], the authors explored a more general topology, namely the classical Gaussian FD layered network and proved a worst-case fraction guarantee for selecting the best path in the network. The problem was also studied for classical Gaussian HD diamond networks in [CEF19] where it was shown that operating k relays that are carefully picked can always retain $\frac{k}{N}$ of the unicast approximate capacity. Tightness was shown for this guarantee for the case $k = N - 1$.

[6] An N-relay diamond network is a relay network topology where the source can communicate with a destination only through N non-interfering relays.

5.8 Appendix: Constant Gap Multicast Capacity Approximation for the Gaussian 1-2-1 Network

The memoryless model of the channel allows to upper bound the channel multicast capacity $\mathsf{C}^{\text{multi}}$ using the cut-set upper bound $\mathsf{C}_{\text{cs}}^{\text{multi}}$ as

$$
\begin{aligned}
\mathsf{C}_{\text{cs}}^{\text{multi}} &= \max_{\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} I(\widehat{X}_\Omega; Y_{\Omega^c}|\widehat{X}_{\Omega^c}) \\
&= \max_{\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} I(S_\Omega, \overline{X}_\Omega; Y_{\Omega^c}|S_{\Omega^c}, \overline{X}_{\Omega^c}) \\
&= \max_{\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} I(\overline{X}_\Omega; Y_{\Omega^c}|S_\Omega, S_{\Omega^c}, \overline{X}_{\Omega^c}) + I(S_\Omega; Y_{\Omega^c}|S_{\Omega^c}, \overline{X}_{\Omega^c}) \\
&\leq \max_{\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} I(\overline{X}_\Omega; Y_{\Omega^c}|S_{[0:N]}, \overline{X}_{\Omega^c}) + H(S_\Omega) \\
&\overset{(a)}{\leq} \max_{\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} I(\overline{X}_\Omega; Y_{\Omega^c}|S_{[0:N]}, \overline{X}_{\Omega^c}) + \log(N+1) + N\log(\kappa) \\
&\overset{(b)}{=} \max_{\mathbb{P}_{\{S_i\}}(\cdot)} \max_{\mathbb{P}_{\{\overline{X}_i\}|\{S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} \sum_s \lambda_s\ I(\overline{X}_\Omega; Y_{\Omega^c}|S_{[0:N]}{=}s, \overline{X}_{\Omega^c}) \\
&\qquad + \log(N+1) + N\log(\kappa) \\
&\overset{(c)}{\leq} \max_{\mathbb{P}_{\{S_i\}}(\cdot)} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}} \max_{\mathbb{P}_{\{\overline{X}_i\}|\{S_i\}}(\cdot)} \sum_s \lambda_s\ I(\overline{X}_\Omega; Y_{\Omega^c}|S_{[0:N]}{=}s, \overline{X}_{\Omega^c}) \\
&\qquad + \log(N+1) + N\log(\kappa), \qquad (5.26)
\end{aligned}
$$

where: (i) $\Omega^c = [0:N]\backslash\Omega$; (ii) $\mathbb{P}_{\{\overline{X}_i,S_i\}}(\cdot)$ is the probability distribution of the channel input $\{(S_i, \overline{X}_i)\}_{i=0}^N$; (iii) $S_\Omega = \{S_i | i \in \Omega\}$; (iv) the inequality in (a) is due to the fact that the state variable at the source takes $N+1$ values (the source can only be transmitting to at most one node), while at each relay/destination, the state variable can take κ values, where κ depends on the mode of operation at the relays, namely

$$
\kappa = \begin{cases} N(N+1) & \text{if relays operate in FD} \\ 2N & \text{if relays operate in HD} \end{cases};
$$

(v) we use s in the equality (b) to enumerate the possible network states $S_{[0:N]}$ and we denote with $\lambda_s = \mathbb{P}(S_{[0:N]} = s)$ the joint distribution of the states; (vi) the inequality in (c) follows from the max-min inequality.

For a network state s, we define the channel matrix $\widehat{H}_s$, where the element $[\widehat{H}_s]_{i,j}$ is defined as

$$[\widehat{H}_s]_{i,j} = \begin{cases} h_{ij} & \text{if } i \in s_{j,t} \text{ and } j \in s_{i,r} \\ 0 & \text{otherwise,} \end{cases} \tag{5.27}$$

where h_{ij} is the channel coefficient of the link from node j to node i. It is not difficult to see that every row (and column) of $\widehat{H}_s$ has at most one non-zero element and thus there exists a permutation matrix Π such that $\Pi\widehat{H}_s$ is a diagonal matrix. Also, let $s^+ = \{i | s_{i,t} \neq \emptyset, \forall i \in [0:N]\}$ and $s^- = \{i | s_{i,r} \neq \emptyset, \forall i \in [1:N]\}$.

With this, we can further simplify the mutual information expression in (5.26) as follows

$$\begin{aligned} &\max_{\mathbb{P}_{\{\overline{X}_i\}|\{S_i\}}(\cdot)} \sum_s \lambda_s \; I(\overline{X}_\Omega; Y_{\Omega^c} | S_{[0:N]}{=}s, \overline{X}_{\Omega^c}) \\ &\overset{(a)}{=} \max_{\mathbb{P}_{\{\overline{X}_i\}|\{S_i\}}(\cdot)} \sum_s \lambda_s \; I(\overline{X}_{s^+,\Omega}; Y_{s^-,\Omega^c} | S_{[0:N]}{=}s, \overline{X}_{\Omega^c}) \\ &\overset{(b)}{=} \sum_s \lambda_s \log\det\left(I + \widehat{H}_{s,\Omega} \; K_{s,\Omega} \; \widehat{H}_{s,\Omega}^H\right) \\ &= \sum_s \lambda_s \log\det\left(I + \widehat{H}_{s,\Omega}^H \widehat{H}_{s,\Omega} \; K_{s,\Omega} \; \right), \end{aligned} \tag{5.28}$$

where: (i) we define $\overline{X}_{s^+,\Omega}$ as $\overline{X}_{s^+,\Omega} = \left\{ \overline{X}_i(s_{i,t}) \middle| \, i \in \Omega \cap s^+ \right\}$ and $\overline{Y}_{s^-,\Omega^c}$ as $\overline{Y}_{s^-,\Omega^c} = \left\{ \overline{Y}_i \middle| \, i \in \Omega^c \cap s^- \right\}$; (ii) the equality in (a) follows since, given the state s, all variables $\overline{X}_i(j)$, with $j \neq s_{i,t}$, as well as all Y_i with $s_{i,r} = \emptyset$ are deterministic; (iii) the equality in (b) follows due to the maximization of the mutual information by the Gaussian distribution; (iv) $\widehat{H}_{s,\Omega}$ is a submatrix of $\widehat{H}_s$ (defined in (5.27)) and is defined as $\widehat{H}_{s,\Omega} = [\widehat{H}_s]_{\Omega^c,\Omega}$ and $K_{s,\Omega}$ is the submatrix of the covariance matrix of the random vector $\left[\overline{X}_0(s_{0,t}) \; \overline{X}_1(s_{1,t}) \ldots \overline{X}_N(s_{N,t})\right]^T$, where the rows and columns are indexed by Ω.

We now further upper bound the Right-Hand Side (RHS) of (5.28) using [LKE11, Lemma 1]. In particular, using [LKE11, Lemma 1], we have that for any $\gamma \geq e-1$, the following holds

$$\begin{aligned}\log\det\left(I+\widehat{H}_{s,\Omega}^{H}\widehat{H}_{s,\Omega}\ K_{s,\Omega}\right) &\leq \log\det\left(I+\gamma^{-1}P\widehat{H}_{s,\Omega}^{H}\widehat{H}_{s,\Omega}\right)+|\Omega|\log\alpha(\Omega,s,\gamma)\\ &\overset{(a)}{\leq} \log\det\left(I+P\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{H}\right)+|\Omega|\log\alpha(\Omega,s,\gamma), \end{aligned} \tag{5.29}$$

where the inequality (a) follows since $\gamma > 1$ and by applying Sylvester's determinant identity and $\alpha(\Omega,s,\gamma)$ is defined based on [LKE11, Lemma 1] as

$$\alpha(\Omega,s,\gamma)=\begin{cases} e^{\gamma/e} & \text{if } \gamma \leq e\frac{\operatorname{rank}(H_{s,\Omega})}{\operatorname{trace}(K_{s,\Omega}/P)} = e\frac{\operatorname{rank}(H_{s,\Omega})}{|s^{+}\cap\Omega|}\\ \left(\gamma\frac{|s^{+}\cap\Omega|}{\operatorname{rank}(H_{s,\Omega})}\right)^{\frac{\operatorname{rank}(H_{s,\Omega})}{|s^{+}\cap\Omega|}} & \text{otherwise.}\end{cases} \tag{5.30}$$

If we select $\gamma = e$, then we have that

$$\alpha(\Omega,s,e)=\left(e\frac{|s^{+}\cap\Omega|}{\operatorname{rank}(H_{s,\Omega})}\right)^{\frac{\operatorname{rank}(H_{s,\Omega})}{|s^{+}\cap\Omega|}} \leq \max_{x\geq 0}\ (ex)^{\frac{1}{x}} = e. \tag{5.31}$$

Now, if we substitute (5.28), (5.29) and (5.31) in (5.26), we get that

$$\begin{aligned}\mathsf{C}_{\mathrm{cs}}^{\mathrm{multi}} &\leq \max_{\mathbb{P}_{\{S_i\}}(\cdot)}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^{c}}}\left[\sum_{s}\lambda_s\log\det\left(I+P\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{H}\right)+|\Omega|\log e\right]+\log(N+1)+N\log(\kappa)\\ &\leq \underbrace{\max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s\lambda_s=1}}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^{c}}}\sum_{s}\lambda_s\log\det\left(I+P\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{H}\right)}_{\mathsf{C}_{\mathrm{cs,iid}}^{\mathrm{multi}}}\\ &\quad+\underbrace{(N+1)\log e+\log(N+1)+N\log(\kappa)}_{\mathsf{GAP}(N)}. \end{aligned} \tag{5.32}$$

The main observation in (5.32) is that an i.i.d Gaussian distribution on the inputs and a fixed schedule are within a constant additive gap from the information-theoretic cut-set upper bound on the capacity of the Gaussian 1-2-1 network. With this, we can argue that $\mathsf{C}_{\mathrm{cs,iid}}^{\mathrm{multi}}$ is within a constant additive gap of the capacity. This is due to the fact that $\mathsf{C}_{\mathrm{cs,iid}}^{\mathrm{multi}}$ can be achieved using quantize-map-and-forward as in [OD13] or noisy network coding as in [CTK14].

Due to the special structure of the Gaussian 1-2-1 network we can further simplify $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ by making use of the structure of $\widehat{H}_{s,\Omega}$ in (5.32). In particular, recall that, since every row (and column) in $\widehat{H}_{s,\Omega}$ has at most one non-zero element, then there exists a permutation matrix $\Pi_{s,\Omega}$ such that $\Pi_{s,\Omega}\widehat{H}_{s,\Omega}$ is a diagonal matrix (not necessarily square). Thus, we have

$$\begin{aligned}\log\det\left(I+P\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{H}\right) &\overset{(a)}{=} \log\det\left(I+P\Pi_{s,\Omega}\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{H}\Pi_{s,\Omega}^{T}\right)\\ &\overset{(b)}{=} \sum_{i\in[1:\min\{|\Omega|,|\Omega^c|\}]}\log\left(1+P\left|[\Pi_{s,\Omega}\widehat{H}_{s,\Omega}]_{i,i}\right|^2\right), \end{aligned} \tag{5.33}$$

where: (i) the equality in (a) follows since permutation matrices are orthogonal matrices and thus multiplying by them only permutes the singular values of a matrix; (ii) the equality in (b) follows since the permuted channel matrix $\Pi_{s,\Omega}\widehat{H}_{s,\Omega}$ can be represented as a parallel MIMO channel with $\min\{|\Omega|,|\Omega^c|\}$ active links. We can rewrite the expression in (5.33) as

$$\begin{aligned}\log\det\left(I+P\widehat{H}_{s,\Omega}\widehat{H}_{s,\Omega}^{T}\right) &= \sum_{\substack{(i,j):\\ i\in s^+\cap\Omega,\ j\in s^-\cap\Omega^c,\\ j\in s_{i,t},\ i\in s_{j,r}}}\log\left(1+P\left|[\widehat{H}]_{j,i}\right|^2\right)\\ &= \sum_{\substack{(i,j):\\ i\in s^+\cap\Omega,\ j\in s^-\cap\Omega^c,\\ j\in s_{i,t},\ i\in s_{j,r}}}\log\left(1+P\left|h_{ji}\right|^2\right). \end{aligned} \tag{5.34}$$

Thus, by letting $\ell_{j,i}=\log\left(1+P\left|h_{ji}\right|^2\right)$, we arrive at the following expression for $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$

$$\begin{aligned}\mathsf{C}_{\text{cs,iid}}^{\text{multi}} &= \max_{\substack{\lambda:\|\lambda\|_1=1\\ \lambda\geq 0}}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}}\sum_s\lambda_s\sum_{\substack{(i,j):\\ i\in s^+\cap\Omega,\ j\in s^-\cap\Omega^c,\\ j\in s_{i,t},\ i\in s_{j,r}}}\ell_{j,i}\\ &= \max_{\substack{\lambda:\|\lambda\|_1=1\\ \lambda\geq 0}}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}}\sum_s\lambda_s\sum_{(i,j)\in[0:N]^2}\mathbb{1}_{\{j\in s_{i,t},\ i\in s_{j,r}\}}\mathbb{1}_{\{i\in\Omega,\ j\in\Omega^c\}}\ell_{j,i}\\ &= \max_{\substack{\lambda:\|\lambda\|_1=1\\ \lambda\geq 0}}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}}\sum_{(i,j)\in[0:N]^2}\mathbb{1}_{\{i\in\Omega,\ j\in\Omega^c\}}\sum_s\lambda_s\mathbb{1}_{\{j\in s_{i,t},\ i\in s_{j,r}\}}\ell_{j,i}\\ &= \max_{\substack{\lambda:\|\lambda\|_1=1\\ \lambda\geq 0}}\ \min_{d\in\mathcal{D}}\ \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega,\\ d\in\Omega^c}}\sum_{\substack{(i,j):i\in\Omega,\\ j\in\Omega^c}}\left(\sum_{\substack{s:\\ j\in s_{i,t},\\ i\in s_{j,r}}}\lambda_s\right)\ell_{j,i}\end{aligned}$$

$$= \max_{\substack{\lambda:\|\lambda\|_1=1 \\ \lambda\geq 0}} \min_{d\in\mathcal{D}} \min_{\substack{\Omega\subseteq[0:N]:0\in\Omega, \\ d\in\Omega^c}} \sum_{\substack{(i,j):i\in\Omega, \\ j\in\Omega^c}} \ell_{j,i}^{(s)}, \tag{5.35}$$

where $\ell_{j,i}^{(s)}$ is defined as

$$\ell_{j,i}^{(s)} = \left(\sum_{\substack{s: \\ j\in s_{i,t}, \\ i\in s_{j,r}}} \lambda_s \right) \ell_{j,i}.$$

This concludes the proof that the multicast capacity $\mathsf{C}^{\text{multi}}$ of the Gaussian 1-2-1 network described in (5.4) and in (5.5) can be characterized to within a constant additive gap as expressed in (5.6).

CHAPTER 6

Efficient Scheduling in Gaussian Half-Duplex 1-2-1 Networks

In this chapter, we continue our study of Gaussian 1-2-1 networks. The main result of this chapter is the design of two polynomial time algorithms that: (i) compute the approximate multicast capacity of the 1-2-1 HD network and, (ii) find the network schedule optimal for the approximate multicast capacity. The chapter starts by expressing the approximate multicast capacity as a linear program with an exponential number of constraints. A core technical component consists of building a polynomial time separation oracle for this linear program, by using algorithmic tools such as perfect matching polytopes and Gomory-Hu trees.

6.1 Introduction

In Chapter 5, we introduced the 1-2-1 network model and proved an max-min expression for its approximate multicast capacity (both in FD and HD) modes of operation. The efficient polynomial time scheduling of the network in the case of the FD mode of operation turns out to be a direct consequence of casting the approximate capacity as an equivalent linear program. In the HD case, however, showing a similar result requires some additional machinery. In what follows we state our key result of this chapter that shows that polynomial time scheduling can be performed even in the HD mode of operation before delving into proving the result and exploring the proof machinery.

Theorem 6.1.1. *For the N-relay Gaussian HD 1-2-1 network, we have:*

(a) The approximate multicast capacity $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ *can be found in polynomial time in* N*;*

(b) An optimal schedule for the approximate capacity $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ *can be found in polynomial time in* N*.*

To the best of our knowledge, Gaussian HD 1-2-1 networks represent the first class of HD relay networks for which the approximate capacity and schedule can be computed efficiently independently of the network topology.

6.2 Scheduling in Gaussian HD 1-2-1 Networks

Theorem 6.1.1 Part (a). The proof of the first part of Theorem 6.1.1 is a direct consequence of two results that we present and discuss in what follows. Our first result shows that calculating $\mathsf{C}_{\text{cs,iid}}^{\text{multi}}$ for Gaussian HD 1-2-1 networks is equivalent to solving an LP, where the state activation times are replaced by the link activation times. In particular, we have the following theorem for which the proof is delegated to Section 6.3.

Theorem 6.2.1. *For any* N*-relay Gaussian HD 1-2-1 network, we have that* P1 *in* (5.9) *is equivalent to the LP below*

$$\mathrm{P3}_{\text{HD}}: \mathsf{C}_{\text{cs,iid}}^{\text{multi}} = \max_{\alpha,\mathbf{F}} \min_{d\in\mathcal{D}} \left\{ \sum_{j\in[0:N]\setminus d} F_{d,j}^{(d)} \right\}$$

$$\text{(P3}a\text{)} \quad 0 \le F_{j,i}^{(d)} \le \alpha_{ji}\ell_{j,i} \qquad \forall (i,j)\in[0:N]\times[1:N], d\in\mathcal{D},$$

$$\text{(P3}b\text{)} \quad \sum_{j\in[1:N]\setminus\{i\}} F_{j,i}^{(d)} = \sum_{k\in[0:N]\setminus\{i\}} F_{i,k}^{(d)} \quad \forall i\in[1{:}N]\setminus\{d\}, d\in\mathcal{D}$$

$$\text{(P3}c\text{)} \quad \hat{\alpha}_{ij} = \alpha_{ij} + \alpha_{ji}, \quad i\in[0{:}N],\ j\in[i{+}1{:}N],$$

$$\text{(P3}d\text{)} \quad \alpha_{ij} \ge 0 \quad (i,j)\in[0{:}N]\times[1{:}N],$$

$$\text{(P3}e\text{)} \quad \sum_{(i,j):i=v \text{ or } i<j,\ j=v,} \hat{\alpha}_{ij} \le 1, \quad \forall v\in[0:N],$$

$$(\text{P3}f)\ \sum_{\substack{i\in S, j\in S,\\ i<j}} \hat{\alpha}_{ij} \leq \frac{|S|-1}{2},\ \forall S \subseteq [0:N],\ |S| \text{ odd},$$

$$(\text{P3}g)\ \hat{\alpha}_{ij} \geq 0 \quad i \in [0{:}N],\ j \in [i{+}1{:}N],$$

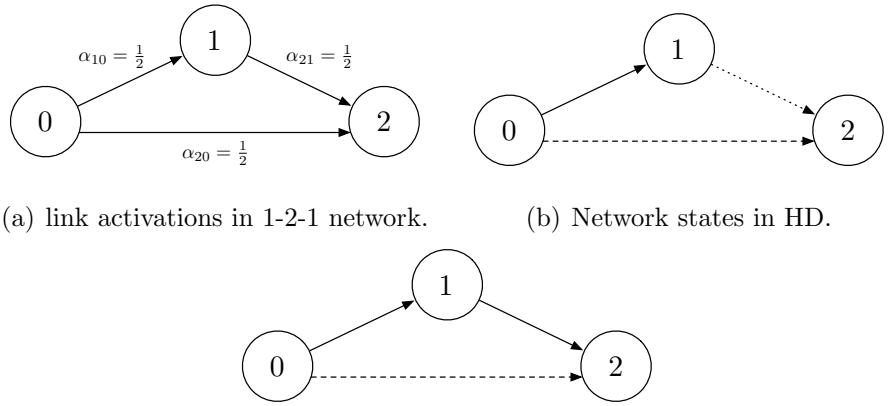
where $F_{j,i}^{(d)}$ *represents the data flow through the link of capacity* $\ell_{j,i}$ *intended for destination* d *and* α_{ji} *represents the fraction of time for which the link is active and* $\alpha = \{\alpha_{ji}\},\ \forall i \in [0:N], j \in [1:N]$ *and* $i \neq j$.

Similar to the development used in Section 5.4 for the FD case, the LP P3_{HD} is very similar to the LP representation of the max-flow problem where the edge capacities are given by $\alpha_{ji}\ell_{j,i}$. The key difference with FD is that α_{ji} is now a variable that is subject to the feasibility constraints in $(\text{P3}c) - (\text{P3}f)$, which stem from the nature of the scheduling in HD 1-2-1 networks.

Remark 6.2.1. There exists a fundamental difference in HD with respect to FD, which is captured by the constraints in $(\text{P3}f)$ that are not needed in FD. To illustrate the need of the constraints in $(\text{P3}f)$ in P3_{HD}, consider the network in Fig. 6.1(a) with a single destination node. Assume that each of the three links in the network is active for a fraction of time equal to 1/2 (as shown in Fig. 6.1(a)). Clearly, these link activation times satisfy the constraints in $(\text{P3}c)$–$(\text{P3}e)$ with $\hat{\alpha}_{01}{=}\hat{\alpha}_{02}{=}\hat{\alpha}_{12} = \frac{1}{2}$. However, we note that these link activation times do not satisfy the constraints in $(\text{P3}f)$ since, by considering $S = \{0,1,2\}$, we have

$$\sum_{i\in S, j\in S,\ i<j} \hat{\alpha}_{ij} = \hat{\alpha}_{01} + \hat{\alpha}_{02} + \hat{\alpha}_{12} = \frac{3}{2} > \frac{|S|-1}{2} = 1.$$

Thus, if the constraints in $(\text{P3}f)$ were not there, then one would conclude that the link activation times illustrated in Fig. 6.1(a) are feasible. However, we now show that this is not the case, which highlights the need of the constraints in $(\text{P3}f)$. For the Gaussian 1-2-1 network in Fig. 6.1(a), when the relay operates in HD, there are three possible useful states of the network, in each of which exactly one link is active. These three states are depicted with different line styles (i.e., solid, dashed, dotted) in Fig. 6.1(b). Note that the links



(a) link activations in 1-2-1 network.

(b) Network states in HD.

(c) Network states in FD.

Figure 6.1: Gaussian 1-2-1 network examples and network states in HD and FD.

$0\to1$ and $1\to2$ cannot be active simultaneously because of the HD constraint at the relay. Additionally, the links $0\to1$ and $0\to2$ cannot be activated simultaneously since the source has only a single transmit beam. A similar argument also holds for the links $1\to2$ and $0\to2$. Thus, for this network we have a one-to-one mapping between λ_s in (5.9) and α_{ji} in P3_{HD}, with $\lambda_s = \alpha_{ji}$ if state s activates the link of capacity $\ell_{j,i}$. Hence, if we use the values from Fig. 6.1(a), we would obtain $\sum_s \lambda_s = 3/2 > 1$ which clearly does not satisfy the constraint in P1 in (5.9). We therefore conclude that the link activation times illustrated in Fig. 6.1(a), which are feasible if the relay operates in FD (see Fig 6.1(c)), are not feasible when the relay operates in HD. This simple example shows why the constraints in (P3f) in P3_{HD} are needed for Gaussian HD 1-2-1 networks.

We note that the LP P3_{HD} has a number of variables that is polynomial in N (two per each edge in the network) compared to the number of variables in the LP P1, which instead is exponential in N (one per each state in the network). However, we also note that P3_{HD} now has an exponential number of constraints of the type (P3f). Thus, it follows that algorithms such as the simplex method and the standard interior point method cannot solve P3_{HD} in polynomial time in N. However, as we show next, the ellipsoid method [GLS81] can indeed be adapted to solve P3_{HD} in polynomial time in N. The key step of the ellipsoid method,

that incorporates the constraints of an LP, relies on the existence of an oracle which, given the problem and a point in space, can decide whether the point is feasible or not and, if not, it returns one constraint of the linear program that is violated by that point. If such a decision can be taken in polynomial time, then this is referred to as a *polynomial time separation oracle* and the LP can be solved in polynomial time. Our next result focuses on showing that a polynomial time separation oracle for $\mathrm{P3_{HD}}$ exists such that, given the graph representing the N-relay Gaussian HD 1-2-1 network and an assignment of $\{F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij}\}$, it can verify in polynomial time in N if $\{F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij}\}$ is feasible in $\mathrm{P3_{HD}}$ and, if not feasible, it returns one of the constraints that is violated. In other words, if one of the constraints is violated, then the oracle returns a hyperplane that separates the given point (assignment of $\{F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij}\}$) from the feasible polytope in $\mathrm{P3_{HD}}$. This result is formalized in the theorem below for which the proof is given in Section 6.4.

Theorem 6.2.2. *A polynomial time separation oracle exists that, provided with a network of $N+1$ nodes and an assignment $\{F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij}\}$ for its link activations, can verify in polynomial time in N if $\{F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij}\}$ is feasible with respect to the constraints in $\mathrm{P3_{HD}}$, and if not feasible it returns one of the constraints in $\mathrm{P3_{HD}}$ that is violated.*

Theorem 6.2.1, Theorem 6.2.2 and the existence of the ellipsoid method [GLS81] directly imply the result in Part (a) of Theorem 6.1.1.

Theorem 6.1.1 Part (b). The proof of the second part of Theorem 6.1.1 makes use of Theorem 6.2.2 and an algorithmic version of Caratheodory's theorem to find a feasible schedule $\{\lambda_s\}$ for the approximate capacity in (5.9), such that each link is activated for the amount given by the solution of $\mathrm{P3_{HD}}$. Theorem 6.1.1 Part (b) is proved in Appendix 6.5.

Remark 6.2.2. Through Theorem 6.2.1, the 1-2-1 network model in HD is equivalent to link scheduling in CDMA networks as it was considered in [HS88], where a similar algorithm was independently developed to perform scheduling in polynomial time in the number of nodes.

6.3 Proof of Theorem 6.2.1

Note that all variables, except λ_s and α_{ij}, are the same in the two LPs P1 and P3_{HD}. Thus, to make P1 and P3_{HD} equivalent, we want to find a bijective mapping between λ_s's and α_{ij}'s.

If we observe a state s in a Gaussian HD 1-2-1 network and hence in P1, we see that a state does not activate two adjacent links, i.e., a state represents a matching in the graph representing the network topology. Furthermore, the constraints in (P1a), (P1c) and (P1d) in P1 suggest that the link activation times should be in the convex hull of the 0-1 activations representing a matching. To state this more formally, let $F^{(d)}$ be a matrix that is populated by the flow variables $F_{j,i}^{(d)}$, $\forall d \in \mathcal{D}$ in P1, and let L be a matrix populated by the link capacities $\ell_{j,i}$ in P1. For each state s, let M_s be a binary matrix such that its (j,i)-th component is 1 if the link from node i to node j is activated by state s, and 0 otherwise. Then, we can write the constraint in $(1a)$ in (P1a) as

$$\text{(P1}a\text{)} \quad F^{(d)} \leq \left(\sum_s \lambda_s M_s \right) \odot L, \quad \forall d \in \mathcal{D}, \tag{6.1}$$

where the operator $\odot$ denotes the element-wise multiplication (Hadamard product). Now, if we consider P3_{HD}, then the constraint in (P3a) can be written as

$$\text{(P3}a\text{)} \quad F^{(d)} \leq \Lambda \odot L, \quad \forall d \in \mathcal{D}, \tag{6.2}$$

where Λ is populated by the variables α_{ji}.

From (6.1) and (6.2), we can see that the constraint needed in P3_{HD} is to have Λ in the convex hull of $\{M_s\}$, i.e., we want the link activations (or weights) that are in the convex hull of points representing matchings in a graphs. A result [Edm65] by Edmonds for undirected graphs characterizes the constraints that represent the Matching polytope (M-polytope). The M-polytope of an undirected graph G is the polytope that has all matchings as its extreme points. By massaging this result to apply to the directed graph in our problem, we define $\hat{\alpha}_{ij} = \alpha_{ij} + \alpha_{ji}$ for all $i < j$ and we apply Edmonds's constraints on $\hat{\alpha}_{ij}$. With this, we precisely get the linear program in P1. In particular, the M-polytope characterized by

Edmonds defines the constraints in (P3e) and (P3f) on $\hat{\alpha}_{ij}$. The mapping from $\hat{\alpha}$ to α_{ij} is defined by the constraint in (P3c) in P3_{HD}. This concludes the proof of Theorem 6.2.1.

6.4 Proof of Theorem 6.2.2

In this section, we prove Theorem 6.2.2, namely we show the existence of a polynomial time separation oracle that, provided with a weighted graph with $N+1$ nodes – representing our Gaussian HD 1-2-1 network – and a point $y=(F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij})$ in the space of P3_{HD}, can verify in polynomial time in N if y is feasible in P3_{HD} and if not, it returns a hyperplane that separates y from the feasible region (i.e., an inequality satisfied for the feasible region but not for y).

Our oracle can be divided into two parts: (i) a simple oracle that checks the constraints in (P3a)-(P3e), and (ii) a more involved oracle for checking the constraint in (P3f). Note that since the number of variables and constraints in (P3a)-(P3e) is polynomial in N, then we can directly check these constraints for y in polynomial time in N. If one constraint is violated, then we return that constraint as the hyperplane that separates y from the feasible set. In what follows, we prove that the constraint in (P3f) can also be checked in polynomial time in N. Towards this end, in Section 6.4.1 we first overview some results from [Edm65], and define the M-polytope and the Perfect Matching polytope (PM-polytope) of an undirected graph, and show a useful relationship between them. Note that (similar to the definition of the M-polytope), the PM-polytope of an undirected graph G is the polytope that has all perfect-matchings as its extreme points. Then, in Section 6.4.2 we show how a set S that violates the constraint in (P3f) can be found by first constructing a Gomory-Hu tree [GH61] of the weighted graph representing our network, and then checking cuts with a particular structure in it. Finally, in Section 6.4.3, we show how these results can be leveraged to build our polynomial time separation oracle.

Table 6.1: Quantities of interest used throughout Section 6.4.

Quantity	**Definition**
$x_G(e)$	Weight corresponding to edge $e = (i, j)$ between nodes $i \in V_G$ and $j \in V_G$
$\delta_G(S)$	Set of edges in G that have only one endpoint in the set of vertices $S \subseteq V_G$
$E_G(A, B)$	Set of edges with one endpoint in A and the other endpoint in B with $A, B \subseteq V_G$
$E_G(A)$	Set of edges with both endpoints in the set of vertices A
$x_G(F)$	Sum of the weights of the edges that belong to F, i.e., $x_G(F) = \sum_{e \in F} x_G(e)$

6.4.1 M-polytope and PM-polytope

We here overview some results from [Edm65], and show a useful relationship between the M-polytope and the PM-polytope of an undirected graph, which we next define. In particular, we will use the following graph theory notation. For an undirected graph $G = (V_G, E_G, x_G)$, with set of vertices V_G, set of edges E_G and edge weight function $x_G : E_G \to \mathbb{R}_+$, we use the convention $e = (i, j)$ with $i < j$. Furthermore, in the remainder of this section, we use the definitions in Table 6.1.

As an example of the used notation, with reference to the weighted graph G in Fig. 6.2, we have

$$\delta_G(\{1,3,4,5\}) = \{(0,1),(1,2),(2,3),(2,4),(2,5)\},$$
$$E_G(\{1,3,4,5\},\{2\}) = \{(1,2),(2,3),(2,4),(2,5)\},$$
$$E_G(\{1,3,4,5\}) = \{(1,3),(1,4),(3,4),(3,5),(4,5)\},$$
$$x_G(\{(0,2),(1,3),(2,5)\}) = 8+4+2 = 14.$$

We let $G = (V_G, E_G, \hat{\alpha})$ define the weighted undirected graph that describes our Gaussian HD 1-2-1 network, where the weight of edge e equals $\hat{\alpha}(e) = \hat{\alpha}_e$ as in the constraint in (P3c) in P3$_{\text{HD}}$. Then, we note that the constraints in (P3d) $-$ (P3f) in P3$_{\text{HD}}$ are those introduced by Edmonds [Edm65] to define the M-polytope for the graph G. By using the notation

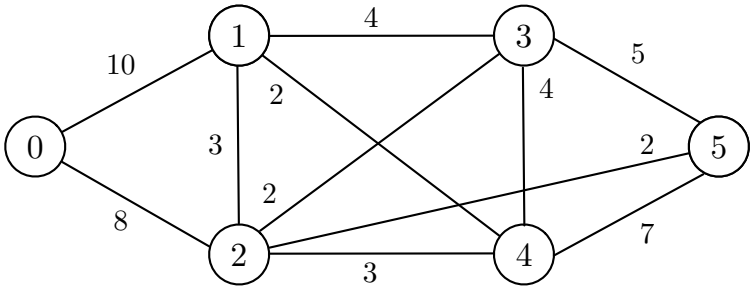


Figure 6.2: Example of a weighted graph G.

introduced above, we can rewrite the constraints in (P3d) $-$ (P3f) in P3$_{\text{HD}}$, as follows

$$\begin{aligned}
&\text{M} - \text{polytope}: \\
&\quad \hat{\alpha}(e) \geq 0, && \forall e \in E_G, \\
&\quad \hat{\alpha}(\delta_G(v)) \leq 1, && \forall v \in V_G, \\
&\quad \hat{\alpha}(E_G(S)) \leq \frac{|S|-1}{2}, && \forall S \subseteq V_G, |S| \text{ odd}.
\end{aligned} \tag{6.3}$$

The PM-polytope is represented similarly to the M-polytope, where now the second constraint in (6.3) is forced to be satisfied with equality, namely

$$\begin{aligned}
&\text{PM} - \text{polytope}: \\
&\quad \hat{\alpha}(e) \geq 0, && \forall e \in E_G, \\
&\quad \hat{\alpha}(\delta_G(v)) = 1, && \forall v \in V_G, \\
&\quad \hat{\alpha}(E_G(S)) \leq \frac{|S|-1}{2}, && \forall S \subseteq V_G, |S| \text{ odd}.
\end{aligned} \tag{6.4}$$

By using the second constraint in (6.4), we can manipulate the third constraint, and rewrite the PM-polytope as

$$\begin{aligned}
&\text{PM} - \text{polytope}: \\
&\quad \hat{\alpha}(e) \geq 0, && \forall e \in E_G, \\
&\quad \hat{\alpha}(\delta_G(v)) = 1, && \forall v \in V_G, \\
&\quad \hat{\alpha}(\delta_G(S)) \geq 1, && \forall S \subseteq V_G, |S| \text{ odd}.
\end{aligned} \tag{6.5}$$

We now show that, for any graph $G = (V_G, E_G, \hat{\alpha})$ with $|V_G|$ vertices, there is an injection from the M-polytope of G to the PM-polytope of a constructed graph $\widetilde{G}$ with double the number of vertices. Towards this end, we first create a copy $G' = (V_{G'}, E_{G'}, \hat{\alpha}')$ of the original graph G; note that with this construction we have $\hat{\alpha}' = \hat{\alpha}$. We then let $\widetilde{G} = (V_{\widetilde{G}}, E_{\widetilde{G}}, \widetilde{\alpha})$ be a graph with: (i) $V_{\widetilde{G}} = V_G \cup V_{G'}$, (ii) $E_{\widetilde{G}} = E_G \cup E_{G'} \cup \{(v, v') | v \in V_G,\ v' \text{ is the copy of } v \text{ in } G'\}$, and (iii) $\widetilde{\alpha}(e)$ defined as

$$\widetilde{\alpha}(e) = \begin{cases} \hat{\alpha}(e) & \text{if } e \in E_G, \\ \hat{\alpha}'(e) & \text{if } e \in E_{G'}, \\ 1 - \hat{\alpha}(\delta_G(v)) & \text{if } e = (v, v'). \end{cases} \tag{6.6}$$

It turns out that if $\hat{\alpha}$ is in the M-polytope of G, then $\widetilde{\alpha}$ is in the PM-polytope of $\widetilde{G}$ (check [ECF19a] for a detailed proof of this standard claim). This result implies that we can check whether $\hat{\alpha}(e)$ is in the M-polytope of G by checking whether $\widetilde{\alpha}(e)$, with the construction in (6.6), is in the PM-polytope of $\widetilde{G}$.

6.4.2 PM-polytope and Gomory-Hu Tree

In the previous subsection, we have defined the PM-polytope for the weighted undirected graph $\widetilde{G}$ as in (6.5). In particular, in (6.5) we can check in polynomial time if the first two sets of constraints are satisfied. Therefore, our main concern lies in the third group of constraints, since there is an exponential number of them. We here show that a set S that violates the third constraint in (6.5) can be found by first constructing a Gomori-Hu tree [GH61] of $\widetilde{G}$ (defined below), and then checking cuts with a particular structure in it.

We start by noting that the third group of constraints in (6.5) can be written in a compact way as

$$\min_{\substack{S \subseteq V_{\widetilde{G}}, \\ |S| \text{ odd}}} \hat{\alpha}(\delta_{\widetilde{G}}(S)) \geq 1. \tag{6.7}$$

In words, what this says is that the *minimum odd* cut in $\widetilde{G}$ has a value greater than or equal to 1. An *odd* cut is a vertex partition of $V_{\widetilde{G}}$ into S and S^c such that either S and/or S^c has an odd cardinality (S^c is the complement of S).

In [PR82], Padberg and Rao provided an efficient algorithm to find the minimum odd cut and its value for any graph G. An appealing feature of the algorithm designed in [PR82] is that it runs in polynomial time in $|V_G|$. In particular, the method introduced, which we summarize below, consists of using *Gomory-Hu* trees.

Definition 6.4.1 (Gomory-Hu Tree)**.** Let $G = (V_G, E_G, x_G)$ be a capacitated (weighted) undirected graph with capacity function $x_G : E_G \to \mathbb{R}_+$. A *Gomory-Hu* tree (for G and x_G) is a capacitated tree $T = (V_G, F, \beta_T)$ with capacity β_T, such that for each edge $e = (s, t) \in F$, the two components of $T \backslash e$ give a a minimum capacity $s - t$ cut in G. The capacity of the cut in G is equal to $\beta_T(s, t)$.

Note that, for any capacitated undirected graph with $|V_G|$ vertices, a Gomory-Hu tree always exists and can be constructed by the algorithm in [GH61] using $|V_G| - 1$ runs of the max-flow problem. Given the triangle inequality of min-cuts in a graph, Definition 6.4.1 implies the following property of Gomory-Hu trees.

Property 6.4.1. Let $G = (V_G, E_G, x_G)$ be a capacitated undirected graph and $T = (V_G, F, \beta_T)$ be a Gomory-Hu tree of G. Consider any two vertices $u, v \in V_G$, let $\mathcal{P}_{uv}$ be the path connecting u and v in T and let (s, t) be the edge with the minimum capacity $\beta_T(s, t)$ along the path $\mathcal{P}_{uv}$. Then, we have the two following properties:

1. The two components $T \backslash (s, t)$ give a minimum capacity $u - v$ cut in G;
2. The value of the minimum capacity $u - v$ cut is given by $\beta_T(s, t)$.

As an example of a Gomory-Hu tree, consider Fig. 6.3, which represents a Gomory-Hu tree T of the graph G in Fig. 6.2. From Property 6.4.1, it follows that if for G in Fig. 6.2,

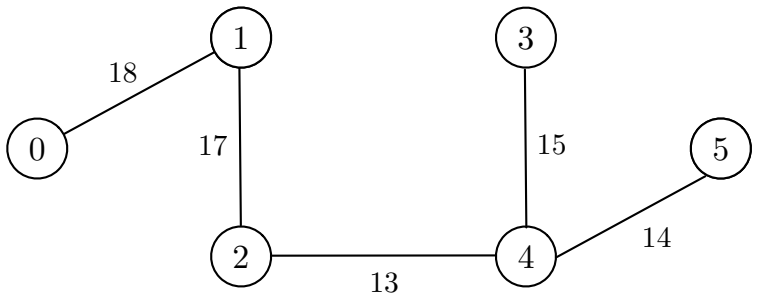


Figure 6.3: A Gomory-Hu tree T of the graph G in Fig. 6.2.

we would like to know the minimum cut between vertices 1 and 5, then all we need to do is to look at the unique path connecting 1 and 5 in T in Fig. 6.3. The edge with the minimum capacity is $(2, 4)$ and hence the min-cut between 1 and 5 equals 13. A cut that has this value in G is the partition $U = \{0, 1, 2\}$ and $U^c = \{3, 4, 5\}$.

We now state a simplified version of the result proved in [PR82] that shows how a Gomory-Hu tree of a graph can be leveraged to verify if the minimum odd cut of a graph satisfies (6.7).

Theorem 6.4.1 ([PR82])**.** Let $\widetilde{G} = (V_{\widetilde{G}}, E_{\widetilde{G}}, \widetilde{\alpha})$ be an undirected capacitated graph with $|V_{\widetilde{G}}|$ even, and let $T = (V_{\widetilde{G}}, F, \beta_T)$ be a Gomory-Hu tree for $\widetilde{G}$. Then, one of the cuts determined by $T \backslash e$, $\forall e \in F$, is a minimum capacity odd cut in $\widetilde{G}$.

We can use the result in Theorem 6.4.1 to verify whether the minimum odd cut has a weight greater than or equal to one, by following the procedure illustrated in Algorithm 1. We note that, even though we need only one among $|W_f|$ and $|W_f^c|$ to be odd, in a graph with even number of vertices, if one is odd, then also the other is odd.

In summary, we have here shown that we can find a set S (if any) that violates the third constraint of the PM-polytope of $\widetilde{G} = (V_{\widetilde{G}}, E_{\widetilde{G}}, \widetilde{\alpha})$ defined in (6.5) by first constructing a Gomory-Hu tree $T = (V_{\widetilde{G}}, F, \beta_T)$ and then checking odd cuts in it. Since the number of cuts in T is $|V_{\widetilde{G}}| - 1 = 2N + 1$, then by using this procedure we only need to perform $O(N)$ checks (and not an exponential number of them).

Algorithm 1 Check if minimum odd cut in $\widetilde{G}$ satisfies (6.7)

function CHECKMINIMUMODDCUT($\widetilde{G}$)
    Build a Gomory-Hu tree $T = (V_{\widetilde{G}}, F, \beta_T)$ of $\widetilde{G}$
    for each $f \in F$ **do**
        Let W_f and W_f^c be the two components of $T \backslash f$
        if $|W_f|$ is odd **then**
            if $\beta_T(f) < 1$ **then**
                return W_f as the set that violates (6.7)
    return $W_f = \emptyset$

6.4.3 Polynomial time Separation Oracle

We here show how the results that we discussed and stated in the previous two subsections can be combined and leveraged to build our polynomial time separation oracle. This oracle either returns that the point $y = (F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij})$ is feasible or it returns one of the constraints in (P3a)-(P3f) that is violated. In this latter case, if the returned violated inequality constraint is evaluated with equality, then it defines the hyperplane that separates y from the feasible polytope. In particular, Algorithm 2 provides the pseudocode of our separation oracle. It is worth noting that each step in Algorithm 2 can be performed in polynomial time in N and hence our oracle runs in polynomial time in N. In particular, the complexity is dominated by the step where a Gomory-Hu tree is constructed for $\widetilde{G}$. This construction algorithm, in fact, involves performing the max-flow problem $2N + 1$ times and hence the complexity of this step – and consequentially of our oracle – is $O(N^4)$ in the worst-case. This concludes the proof of Theorem 6.2.2.

Algorithm 2 Polynomial time separation oracle for P3_{HD}

Input: Network NET, point $y = (F_{i,j}, \alpha_{ij}, \hat{\alpha}_{ij})$ to test for feasibility

Output: Feasible flag, violated constraint in P3_{HD}

Check if all constraints in (P3a)-(P3e) in P3_{HD} are satisfied
if one constraint in (P3a)-(P3e) is violated **then**
    return Feasible = False, constraint violated
Construct undirected simple graph $G = (V_G, E_G, \hat{\alpha})$ with the same set of nodes in NET, edges representing links in NET (but only one direction) and $\hat{\alpha}$ given by y
Construct undirected graph $\widetilde{G}$ from G that has double number of vertices as described in Section 6.4.1
$W_f = \textsc{CheckMinimumOddCut}(\widetilde{G})$
if $W_f = \emptyset$ **then**
    return Feasible = True
else
    Feasible = False
    Let $W_{f,a} = W_f \cap V_G$ and $W'_{f,b} = W_f \cap V_{G'}$
    Let $W'_{f,a}$ be a copy of $W_{f,a}$ in $V_{G'}$
    Let $W_{f,b}$ be a copy of $W'_{f,b}$ in V_G
    if $|W_{f,a} \backslash W_{f,b}|$ odd **then**
        $Z = W_{f,a} \backslash W_{f,b}$
    else
        $Z = W'_{f,b} \backslash W'_{f,a}$
    return Feasible = False, constraint Z in (P3f) violated

6.5 Appendix: Proof of Theorem 6.1.1(b): Constructing an Optimal Schedule in Polynomial Time

To prove part (b) of Theorem 6.1.1, recall that for a Gaussian HD 1-2-1 network, a state s in (5.9) does not activate two adjacent links. Thus, a state is a matching of directed edges in a directed graph representing the network topology. Now, assume that we are given a feasible point in the LP P3_{HD} (obtained by solving P3_{HD}). The main objective of this section is to efficiently (i.e., in polynomial time in the number of nodes) construct a set of matchings (representing states in the network) and find their corresponding activation times (representing λ_s in (5.9)), such that the fraction of time a link $i \to j$ is active is equal to α_{ji} in P3_{HD}. For any pair of nodes $i < j$, we refer to $\hat{\alpha}_{ji}$ in P3_{HD} as the *connection* activation time, i.e., $\hat{\alpha}_{ji}$ represents the duration of time nodes i and j are connected, without considering the direction of communication between them. Thus, from a connection activation time perspective, the network is represented by an undirected graph where an edge (i, j) is active for a fraction $\hat{\alpha}_{ji}$ of time. We first discuss how we can decompose the connection activation times into undirected matchings, and then show how these can be leveraged to construct our set of directed matchings (states). The goal is to show that both these tasks can be performed in polynomial time in the number of nodes.

6.5.1 Decomposition into Undirected Matchings

We define the undirected graph $G = (V_G, E_G, \hat{\alpha})$, where: (i) the graph vertices in V_G represent the nodes in our Gaussian HD 1-2-1 network, (ii) $E_G = \{(i, j) | i > j,\ \hat{\alpha}_{ij} > 0\}$ is the set of edges, and (iii) the edge weights are equal to the values of $\hat{\alpha}_{ij}$ from the feasible point in P3_{HD}. Note that, without loss of generality, in the definition of E_G we do not include any edge e for which $\hat{\alpha}_e = 0$.

Let $\hat{\alpha} \in \mathbb{R}_+^{|E_G|}$ be the vector comprised of $\hat{\alpha}_e,\ \forall e \in E_G$. As highlighted in Section 6.4, the constraints on $\{\hat{\alpha}_e\}$ in $(\text{P3}e) - (\text{P3}g)$ describe the M-polytope of the undirected graph

G [Edm65]. Our goal here is to efficiently find a set of K matchings $M_k \in \{0,1\}^{|E_G|}, k \in [1:K]$ (vertices of the M-polytope) such that

$$\hat{\alpha} = \sum_{k\in[1:K]} \varphi_k M_k, \quad \sum_{k\in[1:K]} \varphi_k = 1, \quad \varphi_k \geq 0, \ \forall k \in [1{:}K]. \tag{6.8}$$

By Caratheodory's theorem [Bar82], we know that for some $K \leq |E_G|+1$, such a decomposition of $\hat{\alpha}$ exists. However, the key challenge is to discover this decomposition in polynomial time in N. Towards this end, we appeal to a result in combinatorial optimization [GLS12, Theorem 6.5.11]. This theorem states that, if we can optimize an objective function over the M-polytope using a separation oracle that runs in polynomial time, then an algorithmic implementation of Caratheodory's theorem can be performed in polynomial time in the number of variables. Our result in Theorem 6.2.2 proves that such a polynomial time separation oracle exists, and hence [GLS12, Theorem 6.5.11] ensures that the decomposition in (6.8) can be performed in polynomial time.

The remainder of this subsection is devoted to describing the decomposition algorithm [GLS12, Theorem 6.5.11] and explaining why we can apply it to our M-polytope. For this, we need to explicitly mention some properties of polyhedra.

Definition 6.5.1. *The dimension of a polyhedron $P \subseteq \mathbb{R}^n$, $dim(P)$ is the maximum number of affinely independent points in P minus 1. If $dim(P) = n$, we say that P is* **fully-dimensional**. *A polyhedron P is said to be* **bounded** *if there exists a ball B in $\mathbb{R}^n$ centered around the origin with radius $r < \infty$ such that $P \subseteq B$.*

Definition 6.5.2. *A polyhedron $P \subseteq \mathbb{R}^n$ is called* **rational** *if all its vertices and at least one point in its interior belong to $\mathbb{Q}^n$. A polyhedron P is called* **well-described**, *if a finite number of bits is needed to encode a single constraint of the polyhedron.*

Definition 6.5.3. *The subset F is called a* **face** *of polyhedron $P \subseteq \mathbb{R}^n$, if there exists an inequality such that $a^T x \leq a_0, \ \forall x \in P$ and $F = \{x \in P | a^T x = a_0\}$. We say that the inequality $a^T x \leq a_0$ defines the face F.*

Definition 6.5.4. *A face F of the polyhedron $P \subseteq \mathbb{R}^n$, is called a* **facet** *of P if $dim(F) = dim(P) - 1$. If the polyhedron P is fully-dimensional, then each non-redundant inequality constraint of P defines a facet of P.*

It is not difficult to see that the M-polytope is rational, well-described and fully-dimensional. To observe full-dimensionality, note that the set of matchings (vertices) such that each selects only one edge in the graph, together with the all-zero matching, form a set of affinely independent points with $|E_G| + 1$ elements. As a result (by Definition 6.5.4) each constraint in (P3e) – (P3g) defines a facet of the M-polytope.

To describe the decomposition algorithm for our polytope, we need to shorthand two abstract oracles that we will use with their shorthand (SEP and OPT), extensively.

1. $\underline{SEP(P, y)}$ denotes the separation oracle that takes a polyhedron P and a point y. This oracle determines if $y \in P$ or else returns a constraint of P that is violated by y. If P is fully-dimensional, then this returned constraint defines a facet of P as aforementioned above. Note that P might not be explicitly defined to the oracle with a set of constraints, but rather as an object (e.g., a graph) and a condition on the object (e.g., minimum odd cut is greater than some value). An example of such an oracle is the result in Theorem 6.2.2.

2. $\underline{OPT(P, c)}$ denotes the optimization oracle that given a polyhedron P and an affine objective function parameterized with c, maximizes $c^T x$ over $x \in P$. If P is bounded then for any c, the oracle returns a vertex of P at which $c^T x$ is maximized.

We are now ready to describe the algorithm used for the decomposition in (6.8) over the M-polytope. The exact details are described in Algorithm 3. The skeleton of Algorithm 3 is the following:

1. We start with our desired point to decompose $y_0 = \hat{\lambda}$ and find any vertex M_1 of the M-polytope (using OPT).

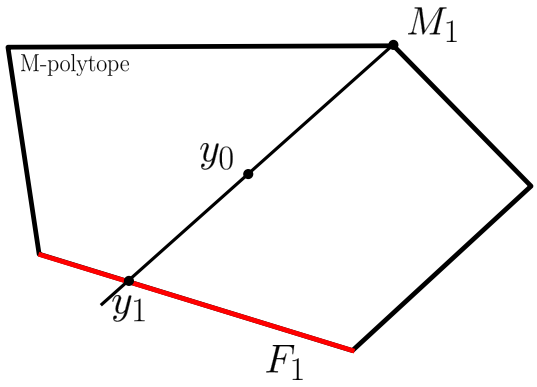


Figure 6.4: An example of the decomposition of y_0 into M_1 and y_1. The red color represents the facet F_1 that y_1 belongs to.

2. We maximize along the line connecting y_0 and M_1 in the M-polytope in the direction $y_0 - M_1$. The maximizer, denoted as y_1, is the intersection of the line and some boundary of the M-polytope. A simple illustration of y_0, y_1 and M_1 is shown in Fig. 6.4.

3. Since y_0 is in between M_1 and y_1 on a line, then we can write $y_0 = \theta_1 M_1 + (1 - \theta_1)y_1$ for $0 \leq \theta \leq 1$.

4. Next, we find a facet F_1 (defined as $a_1^T x = b_1$) of the M-polytope containing y_1 by finding the constraint separating $y_1 + \varepsilon(y_0 - M_1)$ from the M-polytope for some $\varepsilon > 0$ (using SEP).

5. We now want to find a vertex of this facet F_1, so we maximize the objective function $c_1 = a_1$ (using OPT) to get a vertex M_2 in the F_1.

6. Since both y_1 and M_2 belong to F_1 then the line projected along them to get y_2 (similar to how in Step 2, we got y_1 from y_0 and M_1) also belongs to F_1.

7. We repeat Steps 2) to 6) for $i > 1$ until we hit a vertex in the end. Note that in

Step 5) to get M_{i+1} that belongs to the intersection of all facets $F_1, F_2, \cdots, F_i$, we use $c_i = \sum_{j=1}^{i} a_i$ to ensure that the vertex satisfies $a_j^T x = b_j,\ \forall j \in [1:i]$.

8. At the end, we have that for each $i \in [1 : |E_G| + 1]$,

$$y_{i-1} = \theta_i y_i + (1 - \theta_i) M_i, \qquad 0 \leq \theta_i \leq 1.$$

Thus, by applying recursion, we can express our desired point y_0 as

$$y_0 = \sum_{i \in [1:|E_G|+1]} \varphi_i M_i,$$

with

$$\varphi_i = \theta_i \prod_{j \in [1:i-1]} (1 - \theta_j).$$

Note that Algorithm 3 iterates over the oracle for a polynomial number of times equal to $|E_G| + 1$, since at each time, we are restricting our search with a new equation ($a_j^T x = b_j$), i.e., at every iteration we decrease the dimension of the currently considered polytope by 1. Therefore, if the oracles SEP and OPT can run in polynomial time, then the algorithm is polynomial in time. Fortunately, the M-polytope has a polynomial time separation oracle (by Theorem 6.2.2) and a polynomial time optimization oracle (by Theorem 6.1.1). Thus, by consequence Algorithm 3 runs in polynomial time in the number of nodes.

6.5.2 Post-processing for Directional Matchings

We now need to utilize the matchings $\{M_k\}, k \in [1 : K]$ and their activation times $\{\varphi_k\}$ output by the algorithm discussed in the previous subsection to construct network states and find their activation times such that each link $i \to j$ is activated for a duration α_{ji} (output by $\mathrm{P3_{HD}}$). We can perform this decomposition in polynomial time by iterating over the edges of the undirected graph G constructed in the previous subsection. For notational ease, we can rewrite each of the matchings M_k discovered in the previous subsection as a lower triangular

matrix $\widetilde{M_k} \in \{0,1\}^{(N+1)\times(N+1)}$, where

$$\widetilde{M_k}(j,i) = \begin{cases} M_k(j,i) & (i-1,j-1) \in E_G,\ i<j \\ 0 & \text{otherwise,} \end{cases}$$

where $i-1$ and $j-1$ are used since the matrix entries are matched with positive num-

Algorithm 3 Decomposition into Caratheodory points

Input: Graph $G = (V_G, E_G, \hat{\alpha})$

Output: $\{\varphi_i\}$, $\{M_i\}$

$y_0 = \hat{\alpha}$

P = M-polytope of G

$c_0 = 1_{|E_G|\times 1}$ // Set first objective function to any value (all ones here)

for each $i \in [1 : |E_G|+1]$ **do**

  $M_i = OPT(P, c_{i-1})$ // Get vertex maximizing c_{i-1}

  // Maximize in the direction of $y_{i-1} - M_i$ to get the point y_i on the boundary of P

  $y_i = OPT(P \cap \text{line}(y_{i-1}, M_i), y_{i-1} - M_i)$

  // Convex combination of M_i and y_i gets y_{i-1}

  $\theta_i = \frac{M_i - y_i}{y_{i-1} - y_i}$ // Solve for $0 \le \theta_i \le 1 : \theta_i M_i + (1-\theta_i) y_i = y_{i-1}$

  // Get the facet containing y_i. Perturb y_i by a small amount in the direction of $y_{i-1} - M_i$ to get a point outside of the polyhedron, and then apply the separation oracle

  $a_i^T, b_i = SEP(P, y_i + \varepsilon \frac{y_{i-1} - M_i}{|y_{i-1} - M_i|})$ // $a_i^T x = b_i$ defines a facet containing y_i

  //We want the new vertex in the next iteration to be in the intersection of all facets F_j visited before. This is done by adding all the inequalities defining these facets

  $c_i = \sum_{j=1}^{i} a_j$

for each $i \in [1 : |E_G|+1]$ **do**

  $\varphi_i = \theta_i \prod_{j=1}^{i-1} (1-\theta_j)$

bers while our nodes are indexed from 0. The rows of $\widetilde{M_k}$ represent the receiving modules of the nodes, while the column indexes represent the transmitting modules of the nodes. As an illustrative example, consider a Gaussian HD 1-2-1 network with $N = 3$ (2 intermediate relays and 1 destination node), where the undirected graph G has edge set $E_G = \{(0,1),(0,2),(1,2),(1,3),(2,3)\}$. For the matching M_k shown below, we have the

corresponding $\widetilde{M}_k$

$$M_k = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \Longrightarrow \quad \widetilde{M}_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

With this notation, we can design an algorithm that generates the optimal $\{\lambda_s\}$ for (5.9) such that they collectively activate each link $i \to j$ for the duration α_{ji} in the LP P3_{HD}. The formal procedure of the algorithm is described in Algorithm 4. The main idea behind Algorithm 4 is the following. From (6.8), the definition of $\widetilde{M}_k$ and the LP P3_{HD}, we know that $\forall (i,j) \in [0:N]$, such that $i < j$, we have

$$\sum_{k \in [1:|E_G|+1]} \varphi_k \ \widetilde{M}_k(j,i) = \hat{\alpha}_{ij} \overset{(\text{P3}_{\text{HD}})}{=} \alpha_{ji} + \alpha_{ij}.$$

Thus, for each connection (i,j) in the network, we just need to break the matchings $\{\widetilde{M}_k\}$ into two sets contributing to the activation of the links $i \to j$ and $j \to i$. Before processing a connection (i,j), the default direction is $i \to j$. The algorithm iterates over each connection (i,j), and adds up the activation times for the matchings one by one until the sum exceeds α_{ji}. For the remaining matchings, we change the assigned direction to $j \to i$. The matching $\widetilde{M}_k(j,i)$ that caused the sum to exceed α_{ji} is split into two copies, one where the direction is $i \to j$ and the other with $j \to i$.

Note that, in each iteration over the elements in E_G, we split at most one matching (state). Thus, starting with $|E_G|+1$ matchings, we end up with at most $2|E_G|+1$ matchings. Moreover, the inner loop iterates over at most $2|E_G|+1$ matchings. Thus, the algorithm runs in $O(|E_G|^2)$ time which in the worst case is $O(N^4)$ for a network with $N+1$ nodes.

Algorithm 4 Constructing digraph matchings from undirected graph matchings

Input: Graph $G = (V_G, E_G, \hat{\alpha})$, $\{\widetilde{M}_k\}_{k=1}^{|E_G|+1}$, $\{\varphi\}_{k=1}^{|E_G|+1}$

Output: $\{\varphi'_k\}$, $\{\widetilde{M}'_k\}$

1: $K = |E_G| + 1$

2: $\widetilde{M}'_k = \widetilde{M}_k, \quad \varphi'_k = \varphi_k, \qquad \forall k \in [1 : |E_G| + 1]$

3: **for each** $(i, j) \in E_G$ **do**

4: $\quad \omega = 0$

5: $\quad$ **for each** $k \in [1 : K]$ **do**

6: $\quad\quad$ **if** $\widetilde{M}'_k(j, i) == 1$ **then**

7: $\quad\quad\quad$ **if** $\omega + \varphi'_k < \alpha_{ji}$ **then** // All matchings that contain (i, j) connection so far are less than α_{ji}

8: $\quad\quad\quad\quad \omega = \omega + \varphi'_k$

9: $\quad\quad\quad$ **else if** $\omega + \varphi'_k = \alpha_{ji}$ **then** // All matchings that contain (i, j) connection so far are exactly α_{ji}

10: $\quad\quad\quad\quad$ // For all remaining matchings containing (i, j) assign to $j \to i$

11: $\quad\quad\quad\quad$ **for each** $\ell \in [k + 1 : K]$ *s.t.* $\widetilde{M}'_\ell(j, i) == 1$ **do**

12: $\quad\quad\quad\quad\quad \widetilde{M}'_\ell(j, i) = 0, \quad \widetilde{M}'_\ell(i, j) = 1$

13: $\quad\quad\quad\quad$ **Break**

14: $\quad\quad\quad$ **else**

15: $\quad\quad\quad\quad$ // Make another copy $(K + 1)$ of the current state, and assign $j \to i$ instead of $i \to j$

16: $\quad\quad\quad\quad \widetilde{M}'_{K+1} = \widetilde{M}'_k$ // Make a new copy at the end of the list

17: $\quad\quad\quad\quad \varphi'_{K+1} = \omega + \varphi'_k - \alpha_{ji}, \qquad \varphi'_k = \alpha_{ji} - \omega$

18: $\quad\quad\quad\quad \widetilde{M}'_{K+1}(j, i) = 0, \quad \widetilde{M}'_{K+1}(i, j) = 1$

19: $\quad\quad\quad\quad$ // For all remaining matchings in $[k + 1 : K]$containing (i, j) assign to $j \to i$

20: $\quad\quad\quad\quad$ **for each** $\ell \in [k + 1 : K]$ *s.t.* $\widetilde{M}'_\ell(j, i) == 1$ **do**

21: $\quad\quad\quad\quad\quad \widetilde{M}'_\ell(j, i) = 0, \quad \widetilde{M}'_\ell(i, j) = 1$

22: $\quad\quad\quad\quad K = K + 1$ // Increase the number of states by one due to the copy created in line 16

23: $\quad\quad\quad\quad$ **Break**

CHAPTER 7

Network Simplification Results in Gaussian 1-2-1 Networks

In this chapter, we build on the 1-2-1 model introduced in Chapter 5 and Chapter 6 to prove interesting operational properties of 1-2-1 models in terms of the number of routes that need to be used in the network to achieve its approximate unicast capacity as well as provide guarantees on routing in mmWave networks through their 1-2-1 network model.

7.1 Introduction

After deriving the approximate multicast capacity for Gaussian 1-2-1 networks and proving that they can be efficiently scheduled (in polynomial time), we now focus on network simplification results that can be derived from its approximate unicast capacity in 1-2-1 networks. Recall that unicast is a special case of multicast where the destination set $\mathcal{D}$ contains a single element. As a result, without loss of generality, we consider a network with $N+2$ nodes where the source is represented by node 0, the destination is represented by node $N+1$ and we have N intermediate relay nodes. For brevity, the approximate unicast capacity in this case is denoted by $\mathsf{C}_{\text{cs,iid}}$ by dropping the superscript in (5.6b). This gives us the unicast approximate capacity expression below

$$\mathsf{C}_{\text{cs,iid}} = \max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s \lambda_s=1}} \min_{\substack{\Omega\subseteq[0:N+1]:0\in\Omega,\\ N+1\in\Omega^c}} \sum_{\substack{(i,j):i\in\Omega,\\ j\in\Omega^c}} \left(\sum_{\substack{s:\\ j\in s_{i,t},\\ i\in s_{j,r}}} \lambda_s \right) \ell_{j,i} \tag{7.1}$$

We first explore how the approximate unicast capacity of Gaussian FD 1-2-1 networks can be computed as the sum of fractions of the FD Shannon capacity of the paths in the network from the source to the destination. We then study the implications of such a formulation from a network simplification perspective, i.e., how the performance of the network changes by operating only a subset of its nodes.

7.2 A Path Formulation for the Approximate Unicast Capacity

The key result of this section is summarized by the following theorem which casts the unicast approximate capacity of a Gaussian FD 1-2-1 network as an LP in terms of path activations. The proof of the theorem is delegated to Appendix 7.5.1.

Theorem 7.2.1. *For any N-relay Gaussian FD 1-2-1 network, the approximate unicast capacity is equivalent to*

$$\begin{aligned} \text{P5}: \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{p\in\mathcal{P}} x_p \mathsf{C}_p \\ & (\text{P5}a)\ x_p \geq 0 && \forall p \in \mathcal{P}, \\ & (\text{P5}b)\ \sum_{p\in\mathcal{P}_i} x_p f^p_{p.nx(i),i} \leq 1 && \forall i \in [0:N], \\ & (\text{P5}c)\ \sum_{p\in\mathcal{P}_i} x_p f^p_{i,p.pr(i)} \leq 1 && \forall i \in [1:N+1], \end{aligned} \tag{7.2}$$

where: (i) $\mathcal{P}$ is the collection of all paths from the source node 0 to the destination node $N+1$; (ii) $\mathcal{P}_i \subseteq \mathcal{P}$ is the collection of paths that pass through node $i \in [0:N+1]$ (clearly, $\mathcal{P}_0 = \mathcal{P}_{N+1} = \mathcal{P}$ since all paths pass through the source and the destination); (iii) C_p is the FD capacity of the path $p \in \mathcal{P}$, i.e., $\mathsf{C}_p = \min_{(i,j)\in p} \ell_{j,i}$; (iv) $p.nx(i)$ (respectively, $p.pr(i)$) with $i \in [0:N+1]$ is the node following (respectively, preceding) node $i \in [0:N+1]$ in path $p \in \mathcal{P}$ (clearly, $p.pr(0) = p.nx(N+1) = \emptyset$); (v) $f^p_{j,i}$ is the optimal activation time for the link of capacity $\ell_{j,i}$ when the path $p \in \mathcal{P}$, such that $(i,j) \in p$, is operated, i.e.,

$$f^p_{j,i} = \frac{\mathsf{C}_p}{\ell_{j,i}}. \tag{7.3}$$

Remark 7.2.1. In the LP P5, the variable x_p represents the fraction[1] of time the path $p \in \mathcal{P}$ is utilized in the network. Moreover, each of the constraints in (P5b) (respectively, (P5c)) ensures that a node $i \in [0 : N+1]$ - even though it can appear in multiple paths in the network - does not transmit (respectively, receive) for more than 100% of the time.

7.3 Simplification Results for General Topology FD 1-2-1 Networks

Building on the LP P5 in Theorem 7.2.1 presented in the previous section, we can now present the key result for network simplification in this chapter that applies to networks with arbitrary topology. This is given by Lemma 7.3.1 below, which states that although the number of paths $|\mathcal{P}|$ in general is exponential in the number of relays N, we need to use at most a linear number of paths to characterize the approximate capacity $\mathsf{C}_{\text{cs,iid}}$; this can also be translated to a guarantee on the rate that can be achieved using only the best path.

Lemma 7.3.1. *For any N-relay Gaussian FD 1-2-1 relay network with source node 0 and destination node $N+1$, we have the following guarantees:*

(L1) For a network with arbitrary topology, the approximate capacity $\mathsf{C}_{\text{cs,iid}}$ can always be achieved by activating at most $2N+2$ paths in the network.

(L2) For a network with arbitrary topology, the best path has an FD capacity C_1 such that $\mathsf{C}_1 \geq \frac{1}{2N+2}\mathsf{C}_{\text{cs,iid}}$.

(L3) For an N-relay layered network with two layers of relays[2] and $M = N/2$ relays per layer, the approximate capacity $\mathsf{C}_{\text{cs,iid}}$ can be achieved by activating at most $2M+1$

[1]Note that x_p in P5 implicitly satisfies that $x_p \leq 1,\ \forall p \in \mathcal{P}$. This is due to the fact that for any path $p \in \mathcal{P}$, the definition of $f_{j,i}^p$ in (7.3) implies that at least one constraint in (P5b) and (P5c) has $f_{j,i}^p = 1$.

[2]*A layered network with T layers of relays consists of an initial layer with the source, a final layer with the destination and T layers of relays in between. A node in a layer can only receive from nodes in the preceding layer and transmit only to nodes in the following layer.*

paths in the network.

(L4) For an N-relay layered network with two layers of relays and $M = N/2$ relays per layer, the best path has an FD capacity C_1 such that $\mathsf{C}_1 \geq \frac{1}{2M+1}\mathsf{C}_{\text{cs,iid}}$. .

Proof. **Proof of (L1):** The LP P5 in (7.2) is bounded and hence there always exists an optimal corner point. In particular, at any corner point in P5, we have at least $|\mathcal{P}|$ constraints satisfied with equality among (P5a), (P5b) and (P5c). Therefore, we have at least $|\mathcal{P}| - 2N - 2$ in (P5a) satisfied with equality (since (P5b) and (P5c) combined represent $2N + 2$ constraints). Thus, at least $|\mathcal{P}| - 2N - 2$ paths are not operated, which proves the statement (L1).

Proof of (L3): Similar to the proof above for (L1), a corner point in the LP P5 has at most $2N + 2$ constraints among (P5b) and (P5c) satisfied with equality. To prove (L3), we need to show that in the case of a 2-layered network when we have $2N+2$ equality satisfying constraints, then at least one of these equations is redundant. Note that for a 2-layer relay network, any path p in the network from the source to the destination has four nodes and is written as $0 - p(1) - p(2) - N + 1$ where $p(1)$ and $p(2)$ represent the node in the path from layer 1 and layer 2, respectively. We assume that the relays in the first layer are indexed with $[1 : N/2]$ and the second layer relays are indexed with $[N/2 + 1 : N]$. Thus, for any path, $p(1) \in [1 : N/2]$ and $p(2) \in [N/2 + 1 : N]$. Now assume that all the constraints (P5b) and (P5c) are satisfied with equality. Then, by adding all (P5b) constraints for $i \in [1 : N/2]$ and subtracting from them all the constraints from (P5c) for $j \in [N/2 + 1 : N]$, we get

$$\begin{aligned}\text{LHS}: \sum_{i\in[1:N/2]} (\text{P5}b)_i - \sum_{j\in[N/2+1:N]} (\text{P5}c)_j &= \sum_{i\in[1:N/2]} \sum_{p\in\mathcal{P}_i} x_p f^p_{p.\text{nx}(i),i} - \sum_{j\in[N/2+1:N]} \sum_{p\in\mathcal{P}_j} x_p f^p_{j,p.\text{pr}(j)} \\ &= \sum_{p\in\mathcal{P}} x_p f^p_{p(2),p(1)} - \sum_{p\in\mathcal{P}} x_p f^p_{p(2),p(1)} = 0\end{aligned}$$

$$\text{RHS}: \sum_{i\in[1:N/2]} (\text{P5}b)_i - \sum_{j\in[N/2+1:N]} (\text{P5}c)_j = \sum_{i\in[1:N/2]} 1 - \sum_{j\in[N/2+1:N]} 1 = 0.$$

Thus, when all constraints (P5b) and (P5c) are satisfied with equality, at least one of them is redundant, which proves the statement (L3).

Proof of (L2) and (L4): The proof of (L2) follows directly from (L1) by considering only the $2N+2$ paths needed to achieve $\mathsf{C}_{\text{cs,iid}}$ and picking the path that has the largest FD capacity among them. In particular, the guarantee in (L2) is true for the selected path due to the fact that for any feasible point in P5, $x_p \leq 1,\ \forall p \in \mathcal{P}$. The proof of (L4) from (L3) follows by the same argument used to prove (L2) from (L1). □

7.4 Simplification Results for Diamond Topology 1-2-1 Networks

The result in Lemma 7.3.1 (L3) proved in the previous section suggests that for Gaussian 1-2-1 networks with particular structures, we can further reduce the number of active paths needed to achieve the approximate unicast capacity $\mathsf{C}_{\text{cs,iid}}$. In what follows, we explore this observation in the context of Gaussian 1-2-1 diamond networks operating in FD and HD, through the following two lemmas proved in Appendix 7.5.2, Appendix 7.5.3 and Appendix 7.5.4.

Lemma 7.4.1. *For the N-relay Gaussian 1-2-1 diamond network (a layered network with 1 layer of relays) with source node 0 and destination node $N+1$, we can calculate the approximate unicast capacity $\mathsf{C}_{\text{cs,iid}}$ in FD and HD as*

$$\begin{aligned}
\text{P6}^{(\text{dia})}: \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{p\in[1:N]} x_p \mathsf{C}_p \\
& (\text{P6}a)\ 0 \leq x_p \leq 1 \qquad\qquad \forall p \in [1{:}N], \\
& (\text{P6}b)\ \textstyle\sum_{p\in[1:N]} x_p \frac{\mathsf{C}_p}{\ell_{p,0}} \leq 1, \\
& (\text{P6}c)\ \textstyle\sum_{p\in[1:N]} x_p \frac{\mathsf{C}_p}{\ell_{N+1,p}} \leq 1,
\end{aligned} \tag{7.4}$$

where C_p is the capacity (approximate capacity in HD) of the path $0 \to p \to N+1$ and its

value depends on whether the network is operating in FD or HD, namely

$$\mathsf{C}_p = \begin{cases} \min\{\ell_{p,0}, \ell_{N+1,p}\} & \text{if relays operate in FD} \\ \frac{\ell_{p,0}\ \ell_{N+1,p}}{\ell_{p,0}+\ell_{N+1,p}} & \text{if relays operate in HD} \end{cases}.$$

Lemma 7.4.2. *For an N-relay Gaussian 1-2-1 diamond relay network, we have the following guarantees:*

(L1) If the network is operating in FD, then the approximate unicast capacity $\mathsf{C}_{\text{cs,iid}}$ can always be achieved by activating at most 2 relays in the network, independently of N.

(L2) If the network is operating in HD, then the approximate unicast capacity $\mathsf{C}_{\text{cs,iid}}$ can always be achieved by activating at most 3 relays in the network, independently of N.

(L3) In both FD and HD networks, the best path has an approximate unicast capacity C_1 such that $\mathsf{C}_1 \geq \frac{1}{2}\mathsf{C}_{\text{cs,iid}}$; furthermore, this guarantee is tight for both the FD and HD cases, i.e., there exists a class of Gaussian 1-2-1 diamond networks such that $\mathsf{C}_1 \leq \frac{1}{2}\mathsf{C}_{\text{cs,iid}}$ both for the FD and HD cases (see Remark 7.4.1).

Remark 7.4.1. The results in Lemma 7.4.2 (L1) and (L2) are surprising as they state that, independently of the total number of relays in the network, there always exists a subnetwork of 2 (in FD) and 3 (in HD) relays that achieves the full network approximate capacity. Moreover, the guarantee provided by (L3) is tight. To see this, consider $N = 2$ and

$$\begin{aligned} \ell_{1,0} &= \ell_{3,2} = 1, \\ \ell_{3,1} &= \ell_{2,0} = X \to \infty. \end{aligned} \tag{7.5}$$

For this network, it is not difficult to see by applying the LP $\text{P6}^{(\text{dia})}$ that the approximate unicast capacity is $\mathsf{C}_{\text{cs,iid}} = \ell_{1,0} + \ell_{3,2} = 2$, while the capacity of each path (both in FD and HD) is $\mathsf{C}_1 = \min\{1, X\} = 1$, hence $\mathsf{C}_1/\mathsf{C}_{\text{cs,iid}} = 1/2$.

Remark 7.4.2. For FD mode of operation, the result in Lemma 7.4.1 (L3) mirrors the first result in network simplification for classical Gaussian wireless networks [NOF14]. In

this case, it was also shown that by selecting $k = 1$ relay that has the largest capacity path in a diamond network, at least $1/2$ of the approximate capacity is retained. However, different from the classical FD network case, where selecting $k > 2$ relays only guarantees $\frac{k}{k+1}$ of the approximate capacity, in 1-2-1 networks by selecting $k = 2$, the full value of the approximate capacity is retained. When the relays operate in HD, Lemma 7.4.1 (L3) highlights a fundamental difference with respect to the classical wireless network [CEF19]. In particular, in [CEF19] the authors proved that when $k = 1$, then $1/2$ of the full network approximate capacity can always be retained when $N = 2$; however, this ratio decreases as N grows, e.g., when $N \gg 1$ the ratio is equal to $1/4$. Differently, Lemma 7.4.1 (L3) shows that with 1-2-1 constraints selecting $k = 1$ relay always ensures that $1/2$ of the full network approximate capacity can be retained, independently of N.

7.5 Appendices

7.5.1 Proof of Theorem 7.2.1

The proof of Theorem 7.2.1 follows a sequence of LP equivalence relations starting with the LP P2_{FD} in Theorem 5.4.2 and ending with P5 in Theorem 7.2.1. Recall that for a network with source node 0 and single destination node $N + 1$, the LP P2_{FD} in Theorem 5.4.2 to compute the approximate unicast capacity can be simplified to be

$$
\begin{aligned}
\text{P7}: \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{j\in[1:N+1]} F_{j,0} \\
& (\text{P7}a) \quad 0 \le F_{j,i} \le \alpha_{ji}\ell_{j,i} \qquad \forall (i,j)\in[0:N]\times[1:N+1], \\
& (\text{P7}b) \quad \sum_{j\in[1:N+1]\setminus\{i\}} \alpha_{ji} \le 1 \qquad \forall i \in [0:N], \\
& (\text{P7}c) \quad \sum_{i\in[0:N]\setminus\{j\}} \alpha_{ji} \le 1 \qquad \forall j \in [1:N+1], \\
& (\text{P7}d) \quad \alpha_{ji} \ge 0 \qquad\qquad \forall (i,j)\in[0:N]\times[1:N+1].
\end{aligned}
\tag{7.6}
$$

Note that for fixed link activation times α_{ji}, the remaining constraints define a max-flow problem with link capacities $\alpha_{ji}\ell_{j,i}$. The max-flow problem can be equivalently written as an LP with path flows instead of link flows, and hence (7.6) can be written as P7′ described next by using the path flows representation of the max-flow problem. A variable F_p is used for the flow through the path $p \in \mathcal{P}$, where $\mathcal{P}$ is the set of all paths from the source to the destination. With this, we obtain

$$
\begin{aligned}
\text{P7}': \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{p\in\mathcal{P}} F_p \\
& (\text{P7}'a)\ F_p \geq 0 && \forall p \in \mathcal{P}, \\
& (\text{P7}'b) \sum_{\substack{p\in\mathcal{P},\\ (i,j)\in p,\\ j=p.\text{nx}(i)}} F_p \leq \alpha_{ji}\ell_{j,i} && \forall (i,j)\in[0\!:\!N]\times[1\!:\!N\!+\!1], \\
& (\text{P7}'c) \sum_{j\in[1:N+1]\setminus\{i\}} \alpha_{ji} \leq 1 && \forall i \in [0:N], \\
& (\text{P7}'d) \sum_{i\in[0:N]\setminus\{j\}} \alpha_{ji} \leq 1 && \forall j \in [1:N+1], \\
& (\text{P7}'e)\ \alpha_{ji} \geq 0 && \forall (i,j)\in[0\!:\!N]\times[1\!:\!N\!+\!1].
\end{aligned}
\tag{7.7}
$$

The next equivalence relation is to show that P7′ is equivalent to the LP P8 below

$$
\begin{aligned}
\text{P8}: \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{p\in\mathcal{P}} F_p \\
& (\text{P8}a)\ F_p \geq 0 && \forall p \in \mathcal{P}, \\
& (\text{P8}b)\ F_p = \alpha^p_{p.\text{nx}(i),i}\ell_{p.\text{nx}(i),i} && \forall i \in p\setminus\{N+1\}, \forall p \in \mathcal{P}, \\
& (\text{P8}c)\ F_p = \alpha^p_{i,p.\text{pr}(i)}\ell_{i,p.\text{pr}(i)} && \forall i \in p\setminus\{0\}, \forall p \in \mathcal{P}, \\
& (\text{P8}d)\ \textstyle\sum_{p\in\mathcal{P}_i} \alpha^p_{p.\text{nx}(i),i} \leq 1 && \forall i \in [0:N], \\
& (\text{P8}e)\ \textstyle\sum_{p\in\mathcal{P}_i} \alpha^p_{i,p.\text{pr}(i)} \leq 1 && \forall i \in [1:N+1],
\end{aligned}
\tag{7.8}
$$

where remember that, for any path $p \in \mathcal{P}$, we have that $\alpha^p_{i,j}$ is the activation time of the link of capacity $\ell_{i,j}$. Moreover, $p.\text{nx}(i)$ (respectively, $p.\text{pr}(i)$) with $i \in [0 : N+1]$ is the node following (respectively, preceding) node $i \in [0 : N+1]$ in path $p \in \mathcal{P}$ (clearly, $p.\text{pr}(0) = p.\text{nx}(N+1) = \emptyset$).

Finally, we show that P8 is equivalent to P5 in Theorem 7.2.1.

P7′ → P8 . For $(i,j) \in p$ such that $j = p.\mathrm{nx}(i)$, define the variable $\alpha^p_{j,i}$ to be

$$\alpha^p_{j,i} = \frac{F_p}{\ell_{j,i}}. \tag{7.9}$$

Note that, the definition above automatically satisfies the constraints (P8a), (P8b) and (P8c) in P8. Then, by always using the definition in (7.9), we can equivalently rewrite the constraint (P7′b) as

$$(\mathrm{P7}'b): \quad \sum_{\substack{p\in\mathcal{P},\\ (i,j)\in p,\\ j=p.\mathrm{nx}(i)}} \alpha^p_{j,i} \le \alpha_{ji}, \qquad \forall (j,i) \in [1:N+1]\times[0:N].$$

Now, if we fix $\hat{i} \in [0:N]$ and add the left-hand side and right-hand side of (P7′b) for $(j,i) \in [1:N+1]\times\{\hat{i}\}$, then we get

$$\forall \hat{i} \in [0:N], \qquad \sum_{j\in[1:N+1]} \sum_{\substack{p\in\mathcal{P},\\ (\hat{i},j)\in p,\\ j=p.\mathrm{nx}(\hat{i})}} \alpha^p_{j,\hat{i}} \le \sum_{j\in[1:N+1]\setminus\{\hat{i}\}} \alpha_{j,\hat{i}}$$

$$\implies \sum_{p\in\mathcal{P}_{\hat{i}}} \alpha^p_{p.\mathrm{nx}(\hat{i}),\hat{i}} \le \sum_{j\in[1:N+1]\setminus\{\hat{i}\}} \alpha_{j,\hat{i}} \overset{(\mathrm{P7}'c)}{\le} 1 \implies (\mathrm{P8}d).$$

Similarly, by adding the constraints in (P7′b) for a fixed $\hat{j} \in [1:N+1]$, one can show that, under the transformation in (7.9), the constraint in (P8e) is satisfied. Thus, for any feasible point in P7′, we can get a feasible point in P8 using the transformation in (7.9). Regarding the objective function, note that we did not perform any transformation on the variables F_p from P7′ to P8. Therefore, it follows that the objective function value achieved in P7′ is the same as the one achieved in P8.

P8 → P7′. Given a feasible point in P8, we define the following variables for each link in the network

$$\alpha_{ji} = \sum_{\substack{p\in\mathcal{P},\\ (i,j)\in p,\\ j=p.\mathrm{nx}(i)}} \alpha^p_{j,i}.$$

Based on this transformation, we automatically have that (P7′e) is satisfied. Moreover, we have that

$$\begin{aligned}
&(\text{P8}a) \;:\quad \forall p \in \mathcal{P}, \quad 0 \le F_p && \implies (\text{P7}'a)\\
&(\text{P8}d) \;:\quad \forall i, \quad 1 \ge \sum_{p\in\mathcal{P}_i} \alpha^p_{p.\text{nx}(i),i} = \sum_{\substack{p\in\mathcal{P},\\(i,j)\in p}} \alpha_{ji} = \sum_{j\in[1:N+1]\setminus\{i\}} \alpha_{ji} && \implies (\text{P7}'c)\\
&(\text{P8}e) \;:\quad \forall i, \quad 1 \ge \sum_{p\in\mathcal{P}_i} \alpha^p_{i,p.\text{pr}(i)} = \sum_{\substack{p\in\mathcal{P},\\(j,i)\in p}} \alpha_{ij} = \sum_{j\in[0:N]\setminus\{i\}} \alpha_{ij} && \implies (\text{P7}'d)\\
&(\text{P8}b)\ \&\ (\text{P8}c) \;:\quad \sum_{\substack{p\in\mathcal{P},\\(i,j)\in p,\\ j=p.\text{nx}(i)}} \frac{F_p}{\ell_{j,i}} = \sum_{\substack{p\in\mathcal{P},\\(i,j)\in p,\\ j=p.\text{nx}(i)}} \alpha^p_{j,i} = \alpha_{ji} && \implies (\text{P7}'b).
\end{aligned}$$

Thus, the problems P7, P7′ and P8 are equivalent. We now show that P8 is equivalent to P5 in Theorem 7.2.1.

<u>P8 $\to$ P5.</u> Define x_p to be

$$x_p = \frac{F_p}{\mathsf{C}_p}, \quad \forall p \in \mathcal{P}. \tag{7.10}$$

Using this transformation, we get that the constraints in P8 imply the following

$$\begin{aligned}
&(\text{P8}a) \;:\quad \forall p \in \mathcal{P}, \quad 0 \le F_p = x_p\mathsf{C}_p && \implies (\text{P5}a)\\
&(\text{P8}d) \;:\quad \forall i \in [0:N], \quad 1 \ge \sum_{p\in\mathcal{P}_i} \alpha^p_{p.\text{nx}(i),i} \overset{(\text{P8}b)}{=} \sum_{p\in\mathcal{P}_i} \frac{F_p}{\ell_{p.\text{nx}(i),i}}\\
&\qquad\qquad = \sum_{p\in\mathcal{P}_i} \frac{x_p\mathsf{C}_p}{\ell_{p.\text{nx}(i),i}} \overset{(7.3)}{=} \sum_{p\in\mathcal{P}_i} x_p f^p_{p.\text{nx}(i),i} && \implies (\text{P5}b)\\
&(\text{P8}e) \;:\quad \forall i \in [1:N+1], \quad 1 \ge \sum_{p\in\mathcal{P}_i} \alpha^p_{i,p.\text{pr}(i)} \overset{(\text{P8}c)}{=} \sum_{p\in\mathcal{P}_i} \frac{F_p}{\ell_{i,p.\text{pr}(i)}}\\
&\qquad\qquad = \sum_{p\in\mathcal{P}_i} \frac{x_p\mathsf{C}_p}{\ell_{i,p.\text{pr}(i)}} \overset{(7.3)}{=} \sum_{p\in\mathcal{P}_i} x_p f^p_{i,p.\text{pr}(i)} && \implies (\text{P5}c).
\end{aligned}$$

Moreover, we have that

$$(\text{P8 objective function}) \;: \sum_{p\in\mathcal{P}} F_p = \sum_{p\in\mathcal{P}} x_p\mathsf{C}_p = (\text{P5 objective function}).$$

Thus, for any feasible point in P8, we get a feasible point in P5 using the transformation in (7.10) that has the objective function with the same value as the original point in P8.

P5 $\rightarrow$ P8. Define F_p, $\alpha^p_{p.\mathrm{nx}(i),i}$ and $\alpha^p_{i,p.\mathrm{pr}(i)}$ as

$$F_p = x_p \mathsf{C}_p, \quad \alpha^p_{p.\mathrm{nx}(i),i} = \frac{x_p \mathsf{C}_p}{\ell_{p.\mathrm{nx}(i),i}} \forall i \in p \backslash \{N+1\}, \quad \alpha^p_{i,p.\mathrm{pr}(i)} = \frac{x_p \mathsf{C}_p}{\ell_{i,p.\mathrm{pr}(i)}} \forall i \in p \backslash \{0\} \tag{7.11}$$

for all $p \in \mathcal{P}$. Note that the transformation above directly implies conditions (P8b) and (P8c) in P8. Now, we are going to show that the constraints in P5 when applied to (7.11) imply the rest of the constraints in P8 as follows

$$\begin{aligned}
&(\text{P5}a) \ : \forall p \in \mathcal{P}, \quad 0 \leq x_p = \frac{F_p}{\mathsf{C}_p} && \implies (\text{P8}a)\\
&(\text{P5}b) \ : \forall i \in [0:N] \quad 1 \geq \sum_{p \in \mathcal{P}_i} x_p f^p_{p.\mathrm{nx}(i),i} \overset{(7.3)}{=} \sum_{p \in \mathcal{P}_i} \frac{x_p \mathsf{C}_p}{\ell_{p.\mathrm{nx}(i),i}} = \sum_{p \in \mathcal{P}_i} \alpha^p_{p.\mathrm{nx}(i),i} && \implies (\text{P8}d)\\
&(\text{P5}c) \ : \forall i \in [1:N+1] \quad 1 \geq \sum_{p \in \mathcal{P}_i} x_p f^p_{i,p.\mathrm{pr}(i)} \overset{(7.3)}{=} \sum_{p \in \mathcal{P}_i} \frac{x_p \mathsf{C}_p}{\ell_{i,p.\mathrm{pr}(i)}} = \sum_{p \in \mathcal{P}_i} \alpha^p_{i,p.\mathrm{pr}(i)} && \implies (\text{P8}e)
\end{aligned}$$

Moreover, we have that

$$(\text{P5 objective function}) \ : \sum_{p \in \mathcal{P}} x_p \mathsf{C}_p = \sum_{p \in \mathcal{P}} F_p = (\text{P8 objective function}).$$

Thus, for any feasible point in P5, we get a feasible point in P8 using the transformation in (7.11) that has the objective function with the same value as the original point in P5. Thus, the two problems P5 and P8 are equivalent. In conclusion, the problems P5, P7, P7$'$ and P8 are equivalent. This concludes the proof of Theorem 7.2.1.

7.5.2 Proof of Lemma 7.4.1 for a Gaussian FD 1-2-1 Diamond Network

In this section, we prove Lemma 7.4.1 for the Gaussian FD 1-2-1 network with a diamond topology. In this network the source communicates with the destination by hopping through one layer of N non-interfering relays. For this network the LP P5 in (7.2) can be further simplified by leveraging the four following implications of the sparse diamond topology:

1. In a Gaussian 1-2-1 diamond network, we have N disjoint paths from the source to the destination, each passing through a different relay. We enumerate these paths with the index $i \in [1:N]$ depending on which relay is in the path. Moreover, each path $i \in [1:N]$ has a FD capacity equal to $\mathsf{C}_i = \min\{\ell_{i,0}, \ell_{N+1,i}\}$.

2. In the Gaussian 1-2-1 diamond network, each relay $i \in [1:N]$ appears in only one path from the source to the destination. Thus, when considering constraints (P5b) and (P5c) in (7.2) for $i \in [1:N]$ gives us that
$$x_i \frac{\mathsf{C}_i}{\ell_{i,0}} \leq 1 \quad \& \quad x_i \frac{\mathsf{C}_i}{\ell_{N+1,i}} \leq 1. \tag{7.12}$$

3. Note that for the Gaussian FD 1-2-1 diamond network, $\mathsf{C}_i = \min\{\ell_{i,0}, \ell_{N+1,i}\}$, $\forall i \in [1:N]$. Therefore, one of the coefficients $\mathsf{C}_i/\ell_{i,0}$ or $\mathsf{C}_i/\ell_{N+1,i}$ in (7.12) is equal to 1. This implies that for the Gaussian FD 1-2-1 diamond network, a feasible point of P5 has $x_i \leq 1$. Therefore, the constraints $x_i \leq 1$, $\forall i \in [1:N]$, albeit redundant, can be added to the LP without reducing the feasible region.

4. In the Gaussian FD 1-2-1 diamond network, the constraints due to the source and destination nodes, namely (P5b) for $i = 0$ and (P5c) for $i = N+1$ in (7.2) give us that
$$\sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{i,0}} \leq 1, \qquad \sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{N+1,i}} \leq 1. \tag{7.13}$$
Note that the constraints in (7.13) make the constraints in (7.12) redundant.

By considering the two implications above, we can readily simplify P5 in (7.2) for Gaussian FD 1-2-1 networks with a diamond topology as follows
$$\begin{aligned} \text{P6}^{(\text{dia})}: \quad & \mathsf{C}_{\text{cs,iid}} = \max \sum_{i\in[1:N]} x_i \mathsf{C}_i \\ & (\text{P6}a)\ 0 \leq x_i \leq 1 \qquad\qquad \forall i \in [1:N], \\ & (\text{P6}b)\ \sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{i,0}} \leq 1, \\ & (\text{P6}c)\ \sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{N+1,i}} \leq 1, \end{aligned} \tag{7.14}$$

which is the LP we have in Lemma 7.4.1.

7.5.3 Proof of Lemma 7.4.1 for a Gaussian HD 1-2-1 Diamond Network

Throughout this section, we slightly abuse notation by defining $\ell_i = \ell_{i,0}$ and $r_i = \ell_{N+1,i}$. Similarly, we also use $\alpha_{\ell_i} = \alpha_{i,0}$ and $\alpha_{r_i} = \alpha_{N+1,i}$. Based on this definition, we can rewrite the approximate unicast capacity expression by leveraging the two following observations:

1. For the diamond topology network with a single destination node $N+1$, the constraint (P3f) is redundant. This is due to the fact that any subset S of nodes with odd cardinality can only fall in one of the following three cases:

   - S contains the source but not the destination (or vice versa): In this case all links with endpoint in S have one of their endpoints as the source. Recall that the relays do not communicate with one another in a diamond topology. Thus, the constraint (P3e) for the source node 0 makes the constraint for S in (P3f) redundant.

   - S contains neither the source nor the destination: In this case the constraint (P3f) is trivial as there are no links connecting nodes in S since the relay nodes (comprising S) do not communicate among themselves.

   - S contains both the source and destination nodes: In this case, all the links with endpoint in S have either the source node 0 or the destination node $N+1$ as one of their endpoints. Thus, it follows from the constraints in (P3e) for nodes 0 and $N+1$, that the constraint (P3f) is redundant for all $|S| \geq 5$ with $|S|$ odd. In the remaining situation where $|S| = 3$, the constraint in (P3e) for the only relay node in S makes (P3f) redundant.

2. All the links in the diamond network are unidirectional (either going from the source to the different relays or from a relay node to the destination). Thus, the use of the

notation $\hat{\alpha}_{ji}$ in $\mathrm{P3_{HD}}$ is redundant.

Based on the above observations, we can write a simplified version of $\mathrm{P3_{HD}}$ in Theorem 6.2.1 for the approximate unicast capacity of the Gaussian HD 1-2-1 diamond network as

$$\begin{aligned}
\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})} : \mathsf{C}_{\mathrm{cs,iid}} = \max & \sum_{j\in[1:N]} F_{N+1,j} \\
(\mathrm{P3}a) \quad & 0 \le F_{N+1,i} \le \alpha_{r_i} r_i && \forall i \in [1:N], \\
(\mathrm{P3}a') \quad & 0 \le F_{i,0} \le \alpha_{\ell_i} \ell_i && \forall i \in [1:N], \\
(\mathrm{P3}b) \quad & F_{N+1,i} = F_{i,0} && \forall i \in [1:N], \\
(\mathrm{P3}d) \ & \alpha_{\ell_i}, \alpha_{r_i} \ge 0 && \forall i \in [1:N], \\
(\mathrm{P3}e) \ & \alpha_{\ell_i} + \alpha_{r_i} \le 1 && \forall i \in [1:N], \\
(\mathrm{P3}e') \ & \sum_{i\in[1:N]} \alpha_{\ell_i} \le 1 && , \\
(\mathrm{P3}e'') \ & \sum_{i\in[1:N]} \alpha_{r_i} \le 1 && .
\end{aligned} \tag{7.15}$$

Our first step is to show that solving the LP $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$ in (7.15) is equivalent to solving the LP P9 below

$$\begin{aligned}
\mathrm{P9}: \quad \text{maximize} \quad & \sum_{i=1}^{N} f_i \\
\text{subject to} \quad & (\mathrm{P9}a) \ f_i = \alpha'_{\ell_i} \ell_i = \alpha'_{r_i} r_i && \forall i \in [1:N], \\
& (\mathrm{P9}b) \ \sum_{i=1}^{N} \alpha'_{\ell_i} \le 1, \quad \sum_{i=1}^{N} \alpha'_{r_i} \le 1, \\
& (\mathrm{P9}c) \ \alpha'_{\ell_i} + \alpha'_{r_i} \le 1 && \forall i \in [1:N], \\
& (\mathrm{P9}d) \ f_i \ge 0 && \forall i \in [1:N],
\end{aligned} \tag{7.16}$$

where: (i) $f_i = \alpha'_{\ell_i} \ell_i = \alpha'_{r_i} r_i$ represents the data flow through the i-th relay; (ii) α_{ℓ_i} (respectively, α_{r_i}) represents the fraction of time in which the link from the source to relay i (respectively, from relay i to the destination) is active. Note that, since the network is

operating in HD, the fact that a beam can be used for either reception or transmission but not both simultaneously is captured by the constraint (P9c) above.

It is not difficult to see that the feasible set in P9 is a subset of the feasible set in $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$ with appropriate renaming of the variables and setting $F_{N+1,i} = F_{i,0} = f_i$. Thus, a feasible point in P9 gives a feasible point in $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$ and the following relation is true

$$(\text{P9 optimal value}) \leq (\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}\text{optimal value}). \tag{7.17}$$

We next show the opposite direction, i.e., a feasible point in the LP $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$ gives a feasible point in the LP P9 with a higher objective value in P9 than $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$. We define the following transformation

$$\begin{aligned} \forall i \in [1:N] \quad : \quad & f_i = \min\left\{\alpha_{\ell_i}\ell_i,\ \alpha_{r_i}r_i\right\}, \\ & \alpha'_{\ell_i} = \frac{f_i}{\ell_i}, \quad \alpha'_{r_i} = \frac{f_i}{r_i}. \end{aligned} \tag{7.18}$$

Using this transformation, we have that

- $\alpha'_{\ell_i}\ell_i = f_i = \alpha'_{r_i}r_i, \quad \forall i \in [1:N] \Longrightarrow (\mathrm{P9}a)$
- $$\begin{aligned} \sum_{i\in[1:N]} \alpha'_{\ell_i} = \sum_{i\in[1:N]} \frac{f_i}{\ell_i} = \sum_{i\in[1:N]} \frac{1}{\ell_i}\min\left\{\alpha_{\ell_i}\ell_i,\ \alpha_{r_i}r_i\right\} &\leq \frac{1}{\ell_i}\min\left\{\ell_i \sum_{i\in[1:N]} \alpha_{\ell_i},\ r_i \sum_{i\in[1:N]} \alpha_{r_i}\right\} \\ &\leq \frac{\min\{\ell_i, r_i\}}{\ell_i} \leq 1 \Longrightarrow (\mathrm{P9}b) \end{aligned}$$
- $$\begin{aligned} \sum_{i\in[1:N]} \lambda'_{r_i} = \sum_{i\in[1:N]} \frac{f_i}{r_i} = \sum_{i\in[1:N]} \frac{1}{r_i}\min\left\{\alpha_{\ell_i}\ell_i,\ \alpha_{r_i}r_i\right\} &\leq \frac{1}{r_i}\min\left\{\ell_i \sum_{i\in[1:N]} \alpha_{\ell_i},\ r_i \sum_{i\in[1:N]} \alpha_{r_i}\right\} \\ &\leq \frac{\min\{\ell_i, r_i\}}{r_i} \leq 1 \Longrightarrow (\mathrm{P9}b) \end{aligned}$$
- $$\begin{aligned} \alpha'_{\ell_i} + \alpha'_{r_i} = \left[\frac{1}{\ell_i} + \frac{1}{r_i}\right]\min\left\{\alpha_{\ell_i}\ell_i,\ \alpha_{r_i}r_i\right\} &= \min\left\{\alpha_{\ell_i}, \alpha_{r_i}\frac{r_i}{\ell_i}\right\} + \min\left\{\alpha_{\ell_i}\frac{\ell_i}{r_i}, \alpha_{r_i}\right\} \\ &\leq \alpha_{\ell_i} + \alpha_{r_i} \leq 1 \Longrightarrow (\mathrm{P9}c). \end{aligned}$$

Additionally, the constraints (P3d) directly imply the constraints (P9d). Thus, a feasible point in $\mathrm{P3}_{\mathrm{HD}}^{(\mathrm{dia})}$ implies a feasible point in the LP P9 in (7.16). Furthermore, by substitut-

ing (7.18) in $\text{P3}_{\text{HD}}^{(\text{dia})}$, we get that

$$\begin{aligned}(\text{P9 optimal value}) \ : \ \sum_{i\in[1:N]} f_i &= \sum_{i\in[1:N]} \min\{\alpha_{\ell_i}\ell_i,\ \alpha_{r_i}r_i\} \\ &\geq \max \sum_{i\in[1:N]} F_{N+1,i} = (\text{P3}_{\text{HD}}^{(\text{dia})}\text{optimal value}) \end{aligned} \tag{7.19}$$

Thus from (7.17) and (7.19), solving the LP $\text{P3}_{\text{HD}}^{(\text{dia})}$ is equivalent to solving the LP P9.

We are now going to relate the LP P9 discussed above to the LP $\text{P6}^{(\text{dia})}$ in Lemma 7.4.1. Recall that for a two-hop path with link capacities ℓ_i and r_i, the approximate HD capacity is given by

$$\mathsf{C}_i = \frac{\ell_i r_i}{\ell_i + r_i}.$$

Thus, we can write the LP $\text{P6}^{(\text{dia})}$ as

$$\begin{aligned} \text{P6}^{(\text{dia})}: \quad \text{maximize} \quad & \sum_{i\in[1:N]} x_i \frac{\ell_i r_i}{\ell_i + r_i} \\ \text{subject to} \quad & (\text{P6}a)\ 0 \leq x_i \leq 1 \qquad\qquad \forall i \in [1:N], \\ & (\text{P6}b)\ \sum_{i\in[1:N]} x_i \frac{\ell_i}{\ell_i + r_i} \leq 1, \\ & (\text{P6}c)\ \sum_{i\in[1:N]} x_i \frac{r_i}{\ell_i + r_i} \leq 1. \end{aligned} \tag{7.20}$$

We are now going to show that the LP $\text{P6}^{(\text{dia})}$ is equivalent to the LP P9 and, as a consequence, it is equivalent to $\text{P3}_{\text{HD}}^{(\text{dia})}$ in (7.15). To do this, we are going to show how a feasible point in P9 can be transformed into a feasible point in $\text{P6}^{(\text{dia})}$ and vice versa.

1. <u>P9 $\to$ $\text{P6}^{(\text{dia})}$.</u> Define x_i to be

$$x_i = \alpha'_{\ell_i} \frac{\ell_i + r_i}{r_i}, \quad \forall i \in [1:N]. \tag{7.21}$$

Using this transformation, we get that the constraints in P4 imply the following

$$(\text{P9}b) \ : \quad 1 \geq \sum_{i\in[1:N]} \alpha'_{\ell_i} = \sum_{i\in[1:N]} x_i \frac{r_i}{\ell_i + r_i} \qquad \implies (\text{P6}c)$$

$$(\text{P9}b)\ :\quad 1 \geq \sum_{i\in[1:N]} \alpha'_{r_i} \overset{(\text{P9}a)}{=} \sum_{i\in[1:N]} \alpha'_{\ell_i} \frac{\ell_i}{r_i} = \sum_{i\in[1:N]} x_i \frac{\ell_i}{\ell_i + r_i} \qquad \Longrightarrow (\text{P6}b)$$

$$(\text{P9}c)\ :\quad \forall i \in [1:N],\quad 1 \geq \alpha'_{\ell_i} + \alpha'_{r_i} \overset{(\text{P9}a)}{=} \alpha'_{\ell_i} \left(1 + \frac{\ell_i}{r_i}\right)$$
$$= x_i \frac{r_i}{\ell_i + r_i} \left(1 + \frac{\ell_i}{r_i}\right) = x_i \qquad \Longrightarrow (\text{P6}a)$$

$$(\text{P9}d)\ :\quad \forall i \in [1:N],\quad 0 \leq \frac{f_i}{\ell_i} \frac{\ell_i + r_i}{r_i} = \alpha'_{\ell_i} \frac{\ell_i + r_i}{r_i} = x_i \qquad \Longrightarrow (\text{P6}a)$$

and

$$(\text{P9 objective function})\ : \sum_{i\in[1:N]} \alpha'_{\ell_i} \ell_i = \sum_{i\in[1:N]} x_i \frac{r_i}{\ell_i + r_i} \ell_i = (\text{P6}^{(\text{dia})} \text{ objective function}).$$

Thus, for any feasible point in P9, we get a feasible point in $\text{P6}^{(\text{dia})}$ using the transformation in (7.21) that has the same objective function with the same value as the original point in P9.

2. $\underline{\text{P6}^{(\text{dia})} \rightarrow \text{P9.}}$ Define α_{ℓ_i} and α_{r_i} to be

$$\alpha'_{\ell_i} = x_i \frac{r_i}{\ell_i + r_i}, \quad \alpha'_{r_i} = x_i \frac{\ell_i}{\ell_i + r_i}, \quad \forall i \in [1:N]. \tag{7.22}$$

Note that the transformation above directly implies the constraint (P9a) in P9. Now, we are going to show that the constraints in $\text{P6}^{(\text{dia})}$ when applied to (7.22) imply the rest of the constraints in P9 as follows

$$(\text{P6}a)\ :\quad 1 \geq x_i = x_i \left(\frac{r_i}{\ell_i + r_i} + \frac{\ell_i}{\ell_i + r_i}\right) = \alpha'_{\ell_i} + \alpha'_{r_i} \qquad \Longrightarrow (\text{P9}c)$$

$$(\text{P6}a)\ :\quad 0 \leq x_i \frac{r_i}{r_i + \ell_i} = \alpha'_{\ell_i} = \frac{f_i}{\ell_i} \qquad \Longrightarrow (\text{P9}d)$$

$$(\text{P6}b)\ :\quad 1 \geq \sum_{i\in[1:N]} x_i \frac{\ell_i}{\ell_i + r_i} = \sum_{i\in[1:N]} \alpha'_{r_i} \qquad \Longrightarrow (\text{P9}b)$$

$$(\text{P6}c)\ :\quad 1 \geq \sum_{i\in[1:N]} x_i \frac{r_i}{\ell_i + r_i} = \sum_{i\in[1:N]} \alpha'_{\ell_i} \qquad \Longrightarrow (\text{P9}b)$$

$$(\text{P6}^{(\text{dia})} \text{ objective function})\ : \sum_{i\in[1:N]} x_i \frac{r_i}{\ell_i + r_i} \ell_i = \sum_{i\in[1:N]} \alpha'_{\ell_i} \ell_i = (\text{P9 objective function}).$$

Thus, for any feasible point in P6$^{(\text{dia})}$, we get a feasible point in P9 using the transformation in (7.22) that has the same objective function with the same value as the original point in P6$^{(\text{dia})}$.

Thus, the two problems P6$^{(\text{dia})}$ and P9 are equivalent. This concludes the proof of Lemma 7.4.1 for the HD case.

7.5.4 Proof of Lemma 7.4.2

7.5.4.1 Proof of Lemma 7.4.2 (L1) for a Gaussian FD 1-2-1 Diamond Network

To prove Lemma 7.4.2 (L1), we observe that for a bounded LP, there always exists an optimal corner point. Furthermore, at any corner point in the LP P6$^{(\text{dia})}$, we have at least N constraints satisfied with equality among (P6a)), (P6b) and (P6c). Therefore, we have at least $N-2$ constraints in (P6a) satisfied with equality that form linearly independent equations (since (P6b) and (P6c) combined represent only two constraints). Moreover, recall that, as mentioned earlier, all constraints $x_i \leq 1$ are redundant. Thus, at least $N-2$ relays are turned off (i.e., $x_i = 0$), i.e., at most two relays are sufficient to characterize the approximate capacity of any N-relay Gaussian FD 1-2-1 network with a diamond topology, as stated in Lemma 7.4.2 (L1).

7.5.4.2 Proof of Lemma 7.4.2 (L2) for a Gaussian HD 1-2-1 Diamond Network

We first prove the following property of the optimal corner points in the LP P6$^{(\text{dia})}$ in (7.20) for the HD case.

Property 7.5.1. *If we have a Gaussian HD 1-2-1 diamond network, then for any optimal corner point solution of* P6$^{(\text{dia})}$*, at least one of the constraints in* (P6b) *and* (P6c) *is satisfied with equality.*

Proof. We are going to prove Property 2 by contradiction. Note that, since the LP P6$^{(\text{dia})}$ has

N variables, then any corner point in P6$^{(\text{dia})}$ has at least N constraints satisfied with equality. Now, assume that we have an optimal point $(x_1^\star, x_2^\star, \ldots, x_N^\star)$ such that neither (P6b) nor (P6c) is satisfied with equality. This implies that the constraints satisfied with equality are only of the type (P6a). Thus, from the constraints in (P6a), we have that $x_i^\star \in \{0,1\}, \forall i \in [1:N]$. Additionally, (P6$b$) and (P6$c$) being strict inequalities implies that $\sum_{i=1}^N x_i^\star < 2$. Thus, there exists at most one i', such that $x_{i'}^\star = 1$, while $x_j^\star = 0, \forall j \in [1:N] \backslash \{i'\}$.

Now, if we pick some $k \neq i'$ and set $x_k^\star = \varepsilon > 0$ such that both (P6b) and (P6c) are still satisfied, then we increase the objective function by $\varepsilon \frac{\ell_k r_k}{\ell_k + r_k}$, which contradicts the fact that $(x_1^\star, x_2^\star, \ldots, x_N^\star)$ is an optimal solution. □

Now using Property 7.5.1, we are going to prove Lemma 7.4.2 (L2) by considering the following two cases: (i) there exists an optimal corner point for which only one of the constraints in (P6b) and (P6c) is satisfied with equality, and (ii) all optimal corner points have both (P6b) and (P6c) satisfied with equality.

1. **An optimal corner point exists with only one among** (P6b) **and** (P6c) **satisfied with equality.** We denote this optimal corner point as $(x_1^\star, x_2^\star, \ldots, x_N^\star)$. Since only one among (P6b) and (P6c) is satisfied with equality, then at least $N-1$ constraints of the type (P6a) are satisfied with equality. Also note that, since only one among (P6b) and (P6c) is satisfied with equality, then this implies that $\sum_{i=1}^N x_i < 2$. This implies that, although we have at least $N-1$ constraints in (P6a) satisfied with equality, we have at most one i' such that $x_{i'}^\star = 1$. As a result, at least $N-2$ of the constraints satisfied with equality from (P6a) are of the form $x_i = 0$. This proves that at least $N-2$ relays are not utilized at this optimal corner point, which proves Lemma 7.4.2(L2) in this case.

2. **All optimal corner points have** (P6b) **and** (P6c) **satisfied with equality.** Pick an optimal corner point and denote it as $(x_1^\star, x_2^\star, \ldots, x_N^\star)$. Define $\mathcal{F}_x^\star = \{i | 0 < x_i^\star < 1\}$ and $\mathcal{I}_x^\star = \{i | x_i^\star = 1\}$, i.e., the sets of indices of the variables with non-integer and

unitary values, respectively. The fact that both (P6b) and (P6c) are satisfied with equality implies that $\sum_{i=1}^{N} x_i^\star = 2$, which implies that $|\mathcal{I}_x^\star| \leq 2$. Additionally, since we are considering a corner point, then we have that at least $N-2$ constraints of the type (P6a) are satisfied with equality. This implies that $|\mathcal{F}_x^\star| \leq 2$. Note that, if $|\mathcal{F}_x^\star| + |\mathcal{I}_x^\star| \leq 3$ for all optimal corner points, then we have proved Lemma 7.4.2 (L2) for this case. Thus, we now show that the events $\{|\mathcal{F}_x^\star| = 2\}$ and $\{|\mathcal{I}_x^\star| = 2\}$ are mutually exclusive (i.e., disprove the possibility that $|\mathcal{F}_x^\star| + |\mathcal{I}_x^\star| = 4$). This follows by observing the following relation

$$2 = \sum_{i \in [1:N]} x_i^\star = \sum_{i \in [1:N] \setminus \mathcal{I}_x^\star} x_i^\star + \sum_{i \in \mathcal{I}_x^\star} x_i^\star = \sum_{i \in [1:N] \setminus \mathcal{I}_x^\star} x_i^\star + |\mathcal{I}_x^\star|.$$

Thus

$$|\mathcal{I}_x^\star| = 2 \implies \sum_{i \in [1:N] \setminus \mathcal{I}_x^\star} x_i^\star = 0 \implies |\mathcal{F}_x^\star| = 0,$$

which proves that the two events are mutually exclusive. This concludes the proof of Lemma 7.4.2 (L2).

7.5.4.3 Proof of Lemma 7.4.2 (L3)

The proof of Lemma 7.4.2 (L3) for the FD case follows directly from Lemma 7.4.2 (L1) by considering only the two paths (relays) needed to achieve $\mathsf{C}_{\text{cs,iid}}$. Without loss generality, we assume that relays 1 and 2 are the relays in question. Then, using the optimal fractions $x_1^\star$ and $x_2^\star$ we have that

$$\mathsf{C}_{\text{cs,iid}} = x_1^\star \mathsf{C}_1 + x_2^\star \mathsf{C}_2 \overset{\text{(P6}a\text{)}}{\leq} \mathsf{C}_1 + \mathsf{C}_2,$$

which proves that at least one among C_1 or C_2 is always greater than or equal half $\mathsf{C}_{\text{cs,iid}}$.

To prove Lemma 7.4.2 (L3) for the HD case, note that for an HD network C_i in P6$^{(\text{dia})}$ is given by

$$\mathsf{C}_i = \frac{\ell_{i,0}\ \ell_{N+1,i}}{\ell_{i,0}\ +\ \ell_{N+1,i}}.$$

Thus, by adding the constraints (P6b) and (P6c), we have the following implication for any feasible point in P6$^{(\text{dia})}$

$$\begin{aligned} 2 &\geq \sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{i,0}} + \sum_{i\in[1:N]} x_i \frac{\mathsf{C}_i}{\ell_{N+1,i}} \\ &= \sum_{i\in[1:N]} x_i \frac{\ell_{N+1,i}}{\ell_{i,0}+\ell_{N+1,i}} + \sum_{i\in[1:N]} x_i \frac{\ell_{i,0}}{\ell_{i,0}+\ell_{N+1,i}} = \sum_{i\in[1:N]} x_i. \end{aligned} \tag{7.23}$$

Now, assume without loss of generality that the path through relay 1 has the largest HD approximate capacity. Then, for any optimal point $x_i^\star$ that solves P6$^{(\text{dia})}$ in the HD case, we have

$$\mathsf{C}_{\text{cs,iid}} = \sum_{i\in[1:N]} x_i^\star \mathsf{C}_i \leq \left(\sum_{i\in[1:N]} x_i^\star \right) \mathsf{C}_1 \leq 2\mathsf{C}_1.$$

This proves that the approximate capacity of the best path in the network is at least half the of $\mathsf{C}_{\text{cs,iid}}$ and concludes the proof of Lemma 7.4.2 (L3).

CHAPTER 8

Gaussian 1-2-1 Networks with Imperfect Beamforming

In this chapter, we study bounds on the capacity of Full-Duplex (FD) Gaussian 1-2-1 networks with *imperfect* beamforming. In particular, different from the *ideal* 1-2-1 network model introduced in Chapter 5, in this model beamforming patterns result in side-lobe leakage that cannot be perfectly suppressed. We characterize the gap between the approximate capacities of the imperfect and ideal 1-2-1 models for the same channel coefficients and transmit power. We show that, under some conditions, this gap only depends on the number of nodes. Moreover, we evaluate the achievable rate of schemes that treat the resulting side-lobe leakage as noise, and show that they offer suitable solutions for implementation.

8.1 Introduction

Our study of networks with mmWave nodes equipped with directional beams in Chapters 5, 6 and 7 focused on modeling the high-directional nature of communication. Towards this end, our model in the previous chapters considered mmWave nodes equipped with perfect directional beams. In this chapter, we expand our study in order to understand how the capacity approximation is affected when the beamforming is not ideal.

In particular our focus is on Gaussian 1-2-1 FD networks where instead of perfect beams, the nodes are equipped with *imperfect* beams that have side-lobe leakage (a scenario that is closer to practice). In this new imperfect 1-2-1 network model, it is still possible to approximate the capacity using collaborative schemes such as Quantize-Map-Forward

(QMF) [ADT11, OD13] and Noisy Network Coding (NNC) [LKE11] to make use of the multiple access and broadcast channels present in the network. However, our previous study for the *ideal* 1-2-1 model with perfect main-lobe beamforming beams naturally suggests the following two questions: *(i) When is the ideal 1-2-1 model a good approximation for the imperfect 1-2-1 model? (ii) Under what conditions can simple schemes involving point-to-point decoding approximate the performance of QMF and NNC in imperfect 1-2-1 networks?*

Our first main result in this chapter is to characterize the gap between the approximate capacity of imperfect and ideal 1-2-1 networks (introduced in Chapter 5) for the same channel coefficients and transmit power. We give sufficient conditions on the parameters of the beamforming pattern for the gap to be constant. Surprisingly, these conditions are independent of the transmit power used by the nodes in the network and only depend on the channel coefficients through a ratio between their values. Under such conditions, the ideal 1-2-1 network model offers a good approximation for the imperfect model; we can thus utilize tools developed for the ideal model, such as high-efficiency scheduling algorithms, without incurring significant losses over the imperfect model.

Our second result explores the gap between the approximate capacity of the ideal 1-2-1 model and the rate achieved by a simple scheme that consists of decoding point-to-point transmissions while treating side-lobe leakage as noise. We show that we can characterize this gap which, different from our first result, depends on the transmitted power in the network and on the individual channel coefficient values.

8.2 System Model and Capacity Formulation

We consider an N-relay Gaussian 1-2-1 network as studied in Chapter 7 where N relays assist the communication between a source node (node 0) and a destination node (node $N+1$).

We assume FD mode of operation for the relays, where each relay $i \in [1 : N]$ can be

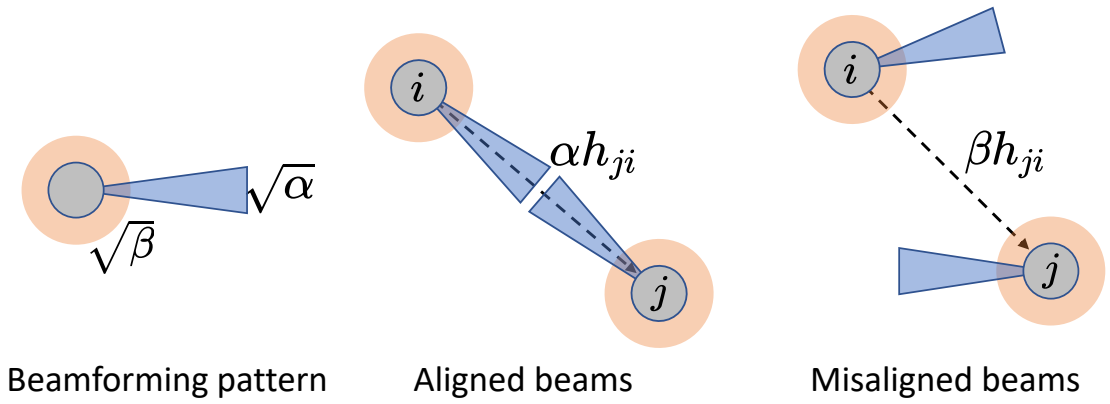


Figure 8.1: Imperfect Beamforming model.

simultaneously receiving and transmitting. Similar to the vanilla model studied in Chapter 5, each node $i \in [0 : N+1]$ in the network is characterized by two states, namely $s_{i,t}$ and $s_{i,r}$, that represent the node towards which node i is beamforming its transmissions and the node towards which node i is pointing its receiving beam, respectively. In particular, $\forall i \in [0 : N+1]$, we have that

$$\begin{aligned} s_{i,t} &\subseteq [1 : N+1] \backslash \{i\}, \quad |s_{i,t}| \leq 1, \\ s_{i,r} &\subseteq [0 : N] \backslash \{i\}, \qquad |s_{i,r}| \leq 1, \end{aligned} \tag{8.1}$$

where $s_{0,r} = s_{N+1,t} = \emptyset$ since the source node always transmits and the destination node always receives. Next, we highlight the difference between the two models that we compare in this chapter

Vanilla 1-2-1 network (Chapter 5): At any particular time, a node can only direct (beamform) its transmission towards at most one other node through a *perfect* main-lobe beamforming beam with no side-lobes. Similarly, a node can only receive transmissions from at most another node (to which its receiving main-lobe beam points towards). Node j receives transmission from node i only if node i points its transmitting beam towards node j, and simultaneously, node j points its receiving beam towards node i. The channel coefficient between nodes i and j is enhanced by a gain $\alpha > 0$.

Imperfect 1-2-1 network: We here introduce the *imperfect* 1-2-1 network model,

where transmissions are still achieved by aligning main-lobes, but in addition, transmissions/receptions through side-lobes also occur and are not suppressed as in the ideal model. In particular, we assume that at any point in time, the channel coefficient h_{ji} from node i to node j is enhanced by a gain $\alpha \geq 1$ when the main-lobes are aligned, and is attenuated by a factor $\beta \leq 1$ otherwise, as shown in Fig. 8.1. Thus, we have the following memoryless channel model

$$Y_j = Z_j + \sum_{i \in [0:N] \setminus \{j\}} \widehat{h}_{ji} X_i, \quad \forall j \in [1 : N+1] \tag{8.2a}$$

$$\widehat{h}_{ji} = \begin{cases} \alpha h_{ji} & \text{if } s_{i,t} = \{j\},\ s_{j,r} = \{i\}, \\ \beta h_{ji} & \text{otherwise}, \end{cases} \tag{8.2b}$$

where: (i) $s_{i,t}$ and $s_{i,r}$ are defined in (8.1); (ii) X_i (respectively, Y_i) denotes the channel input (respectively, output) at node i; (iii) $h_{ji} \in \mathbb{C}$ is the complex channel coefficient from node i to node j without beamforming (i.e., using omnidirectional antennas); the channel coefficients are assumed to be time-invariant; (iv) the channel inputs are subject to an individual power constraint, i.e., $\mathbb{E}[|X_k|^2] \leq P,\ k \in [0 : N]$; (v) $Z_j,\ j \in [1 : N+1]$ indicates the additive white Gaussian noise at the j-th node; noises across the network are assumed to be i.i.d. as $\mathcal{CN}(0, 1)$. We use a matrix H to record all the channel coefficients h_{ji} between any two network nodes, where the rows are indexed by $[1\!:\!N\!+\!1]$ and columns by $[0\!:\!N]$.

Let s be a variable that enumerates all the possible network states of the 1-2-1 network in FD, where each network state corresponds to specific values for the variables in (8.1) for each network node. For an active state s, the effective channel is given by $H_s = B_s \odot H$, where B_s is the beamforming matrix that defines which links are multiplied by the main-lobes gain α and which links are attenuated by β in state s.

Remark 8.2.1. Note that the ideal 1-2-1 network studied in Chapter 5 can be recovered from the imperfect model by setting $\beta = 0$. For $\beta > 0$, the beamforming model, see Fig. 8.1, abstracts the sectored antenna model used to approximate antenna patterns in mmWave ad-hoc networks [TBH16, BAH14].

The Shannon capacity C of the imperfect 1-2-1 network in (8.2) is not known. However, using similar arguments as we used in Chapter 5 to approximate the capacity of the ideal model, we can show that $\mathsf{C}_{\text{cs,iid}}$, namely the cut-set upper bound evaluated with i.i.d. channel inputs and a fixed schedule (i.e., independent of the transmitted message) offers a constant gap approximation for C. In particular, C can be bounded as

$$\mathsf{C}_{\text{cs,iid}} \leq \mathsf{C} \leq \mathsf{C}_{\text{cs,iid}} + O(N \log N), \tag{8.3a}$$

$$\mathsf{C}_{\text{cs,iid}} = \max_{\substack{\lambda_s:\lambda_s\geq 0 \\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega\subseteq[0:N], \\ 0\in\Omega}} \sum_s \lambda_s \text{logdet}\Big(I + P H_{s,\Omega} H_{s,\Omega}^{\dagger}\Big), \tag{8.3b}$$

where: (i) Ω enumerates all possible cuts in the graph representing the network, such that the source belongs to a set of vertices Ω and the destination belongs to Ω^c; (ii) $\Omega^c = [0 : N+1] \backslash \Omega$; (iii) s enumerates all possible network states of the 1-2-1 network in FD, where each network state corresponds to specific values for the variables in (8.1) for each network node; (iv) $H_{s,\Omega}$ is the submatrix of the matrix $H_s = B_s \odot H$, which is obtained by retaining only the rows indexed by Ω^c and the columns indexed by Ω and reorganizing the rows and columns such that the links that are multiplied by α are along the diagonal; (v) λ_s, i.e., the optimization variable, is the fraction of time for which state s is active; a collection of λ_s's for all feasible states, such that they sum up to at most one, is referred to as network *schedule*.

8.3 From Imperfect to Ideal 1-2-1 networks

In this section, we derive an upper bound on the difference between two approximate capacities: (i) the approximate capacity of the ideal Gaussian 1-2-1 networks (i.e., with $\beta = 0$) which we denote by $\mathsf{C}_{\text{ideal}}$ in this chapter to avoid conflict of notation; (ii) the approximate capacity $\mathsf{C}_{\text{cs,iid}}$ of imperfect Gaussian 1-2-1 networks given by (8.3b). Note that, from (7.1), $\mathsf{C}_{\text{ideal}}$ is given by

$$\mathsf{C}_{\text{ideal}} = \max_{\substack{\lambda_s:\lambda_s\geq 0 \\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega\subseteq[0:N+1]:0\in\Omega, \\ N+1\in\Omega^c}} \sum_{\substack{(i,j):i\in\Omega, \\ j\in\Omega^c}} \left(\sum_{\substack{s: \\ j\in s_{i,t},\ i\in s_{j,r}}} \lambda_s \right) \ell_{ji}, \tag{8.4a}$$

$$\ell_{ji} = \log\left(1 + P\alpha^2 \left|h_{ji}\right|^2\right), \tag{8.4b}$$

where $s_{i,t}$ and $s_{i,r}$ denote the transmitting and receiving states for node i in the network state s as defined in (8.1).

The expression $\mathsf{C}_{\text{ideal}}$ in (8.4a) is appealing as it evaluates the approximate capacity in terms of the point-to-point link capacities ℓ_{ji}. This, in turn, leads to interesting properties on how the network should be optimally operated and how to efficiently find an optimal schedule $\{\lambda_s\}$ in polynomial time in N as discussed in the previous chapters. Thus, we would like to understand when the ideal 1-2-1 network model is a good proxy for the imperfect 1-2-1 network. We explore this by characterizing the gap between the approximate capacities of the two models.

We characterize the gap between the approximate capacities of the two models under the following two assumptions that are reasonable for beamforming applications:

• **(Main-lobe is always stronger):** We assume that $\forall i \neq j \neq k$ such that $h_{ji}, h_{jk} \neq 0$, we have that $\alpha|h_{ji}| \geq \beta|h_{jk}|$.

• **(Each cut is diagonally dominant):** We assume that $\forall s, \Omega$, the matrix $A_{s,\Omega} = I + PH_{s,\Omega}H_{s,\Omega}^{\dagger}$ is *diagonally dominant*, i.e.,

$$\rho_{s,\Omega}(H) = \max_{i \in [1:|\Omega|^c]} \left\{ \sum_{j \in [1:|\Omega|] \setminus \{i\}} \frac{|[A_{s,\Omega}]_{ij}|}{|[A_{s,\Omega}]_{ii}|} \right\} \leq 1. \tag{8.5}$$

Remark 8.3.1. Note that in the ideal 1-2-1 network model, the matrix $H_{s,\Omega}$ (and hence, by design also $A_{s,\Omega}$) is diagonal with $\rho_{s,\Omega}(H) = 0$, for all cuts Ω. The condition in (8.5) is a relaxation of the diagonal requirement on $H_{s,\Omega}$ to cases where side-lobes contribution is allowed but not overwhelming.

Our main results are provided by the following two theorems, which are proved in Appendices 8.5.1 and 8.5.2, respectively.

Theorem 8.3.1. *Consider an N-relay Gaussian 1-2-1 network with channel matrix H. Assuming that (α, β) are selected such that the two assumptions above are satisfied, then the*

gap between $\mathsf{C}_{\text{ideal}}$ *and* $\mathsf{C}_{\text{cs,iid}}$ *is upper bounded by*

$$|\mathsf{C}_{\text{cs,iid}} - \mathsf{C}_{\text{ideal}}| \leq N \max\left\{\log N,\ f(H,\alpha,\beta)\right\}, \tag{8.6}$$

where $f(H,\alpha,\beta) = \max_{s,\Omega} |\log\left(1-\rho_{s,\Omega}(H)\right)|$*, with* $\rho_{s,\Omega}(H)$ *defined in* (8.5)*.*

The gap expressed in Theorem 8.3.1 depends on the beamforming parameters (α,β) and the channel coefficients through the expression of $\rho_{s,\Omega}(H)$. In order for the ideal 1-2-1 network model to be a valid approximation for the imperfect 1-2-1 network model (i.e., to ensure that the two approximate capacities are a constant gap away), we would like to operate in the range of parameters (α,β) such that the gap in Theorem 8.3.1 is bounded by $N\log(N)$. Our second theorem provides sufficient conditions on (α,β), as a function of the channel coefficients, such that $|\mathsf{C}_{\text{cs,iid}} - \mathsf{C}_{\text{ideal}}| \leq N\log(N)$.

Theorem 8.3.2. *Consider an* N*-relay Gaussian 1-2-1 network with channel matrix* H*, and let* Δ *be the maximum degree of the graph representing the network topology. If the beamforming parameters* (α,β) *satisfy that*

$$\frac{\alpha}{\beta} \geq \Delta^2 \frac{N}{N-1} \max_{\substack{(i,j,m,n):|h_{ji}|>0,\\ i\neq j\neq m\neq n}} \frac{|h_{mn}|^2}{|h_{ji}|^2}, \tag{8.7}$$

then we have that $|\mathsf{C}_{\text{cs,iid}} - \mathsf{C}_{\text{ideal}}| \leq N\log(N)$*.*

Remark 8.3.2. Note that the condition in (8.7) above is independent of the operating power P, i.e., it is valid for any operational transmit power used in the network. Furthermore, the condition does not depend on the single channel coefficients, but rather is related to the maximum ratio between the magnitudes of any two non-zero channel coefficients in the network. Thus, for any given network with finite channel coefficient magnitudes, there exists an $(\alpha^\star,\beta^\star)$ pair such that the approximate capacity of the imperfect 1-2-1 network model is at most a constant gap away from the approximate capacity of the ideal 1-2-1 network model that uses $(\alpha^\star,0)$ as described in Remark 8.2.1.

Remark 8.3.3. The coupling of the approximate capacity of the imperfect 1-2-1 network to the approximate capacity of its ideal 1-2-1 counterpart, allows to translate results already proven for the ideal 1-2-1 network model to the imperfect 1-2-1 network. In particular, in Chapter 5, we proved that we can find an optimal schedule for the ideal 1-2-1 network model (i.e., a schedule that achieves $\mathsf{C}_{\text{ideal}}$) in polynomial time in the number of nodes. The proof of Theorem 8.3.1 and the result in Theorem 8.3.2 imply that, by applying a schedule developed for the ideal model to an imperfect model that satisfies the condition in Theorem 8.3.2, we can achieve a rate that is at most a constant gap away from the rate achieved in the ideal model.

8.4 Treating Side-lobe Transmissions as Noise

The approximate capacity $\mathsf{C}_{\text{cs,iid}}$ for imperfect 1-2-1 networks in (8.3) can be achieved using schemes such as QMF [ADT11] and NNC [LKE11]. However, given the relation that we have established between the imperfect and ideal approximate capacities in Theorem 8.3.2, it is worth exploring how far the rate achieved by simple schemes that rely on point-to-point decoding is from the approximate capacity.

In this section, we focus on characterizing the gap between the approximate capacity in (8.4) of the ideal 1-2-1 network model and the rate achieved when side-lobe transmissions are treated as noise and only the aligned main-lobes transmissions are decoded at their intended receiver node. Our main result of this section is summarized by the following theorem.

Theorem 8.4.1. *Consider an N-relay Gaussian 1-2-1 network with channel matrix H, and let Δ be the maximum degree of the graph representing the network topology. Let R_{TSN} be the rate achieved by Treating Side-lobes as Noise. Then, we have*

$$|\mathsf{C}_{\text{ideal}} - \mathsf{R}_{\text{TSN}}| \leq N \log(\Delta) + N \max_{i,j} \widetilde{\ell}_{ji}, \tag{8.8}$$

where $\widetilde{\ell}_{ji} = \log\left(1 + \max_{m:m\neq i} \beta^2 P|h_{jm}|^2\right)$.

Before delving into the proof of Theorem 8.4.1, we would like to highlight the following remarks on how the values in Theorem 8.4.1 reflect in practice.

Remark 8.4.1. Note that, for R_{TSN} to be a constant gap away from $\mathsf{C}_{\text{ideal}}$, β should be selected such that for all channel coefficients in the network we have that $\widetilde{\ell}_{ji} = O(1)$.

Remark 8.4.2. The conditions imposed by (8.8) and Theorem 8.3.2 on the beamforming parameters are realistic for several envisioned applications of mmWave communications. For example, in a typical vehicle platooning scenario [ZSS19], the inter-platoon distance is around 10 meters. Thus, with an operating frequency of 60 GHz and bandwidth of 1 GHz, the largest channel coefficient magnitude $\max_{ij} |h_{ij}|^2 = O(10^{-4})$. Assuming the line-of-sight path loss model and a transmit signal-to-noise ratio of 100, we would have $P|h_{ij}|^2 = O(10^{-2})$. As a result, even with $\beta = 1$, the gap in (8.13) would be upper bounded by $N[\log(\Delta) + 1]$.

8.4.1 Proof of Theorem 8.4.1

For a given network state s, node i can communicate to node j, treating side-lobes as noise, at a rate $\widehat{\ell}_{ji}$ given by

$$\widehat{\ell}_{ji} = \begin{cases} \log\left(1 + \frac{\alpha^2 P|h_{ji}|^2}{1+\sum_{m:m\neq i} \beta^2 P|h_{jm}|^2}\right) & \text{if } j\in s_{i,t}, i\in s_{j,r} \\ 0 & \text{otherwise} \end{cases}. \tag{8.9}$$

In other words, the rate $\widehat{\ell}_{ji}$ is either a positive value if the beams are aligned and zero otherwise. Thus, it is not difficult to see that the maximum rate achieved by this scheme can be computed by considering a ideal 1-2-1 network where the point to point link capacities ℓ_{ji} are replaced by $\widehat{\ell}_{ji}$ in (8.9). It therefore follows that the achievable rate R_{TSN} is given by

$$\mathsf{R}_{\text{TSN}} = \max_{\substack{\lambda_s:\lambda_s\geq 0 \\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega:\Omega\subseteq[0:N], \\ 0\in\Omega}} \sum_{\substack{(i,j):i\in\Omega, \\ j\in\Omega^c}} \left(\sum_{\substack{s: \\ j\in s_{i,t},\ i\in s_{j,r}}} \lambda_s\right) \widehat{\ell}_{ji}. \tag{8.10}$$

We now focus on characterizing the gap $|\mathsf{C}_{\text{ideal}} - \mathsf{R}_{\text{TSN}}|$. If the aforementioned gap is upper bounded by a quantity that only depends on N for some conditions on the beamforming parameters (α, β), then by invoking the result in Theorem 8.3.2 together with the triangle inequality, we can show that $|\mathsf{C}_{\text{cs,iid}} - \mathsf{R}_{\text{TSN}}|$ is also upper bounded by a constant gap.

It is not difficult to see that $\mathsf{R}_{\text{TSN}} \leq \mathsf{C}_{\text{ideal}}$, since for all i, j, we have $\widehat{\ell}_{ji} \leq \ell_{ji}$. For a lower bound on R_{TSN}, we can use the following lower bound on $\widehat{\ell}_{ji}$

$$\begin{aligned}
\widehat{\ell}_{ji} &= \log\left(1 + \alpha^2 P|h_{ji}|^2 + \sum_{m:m\neq i} \beta^2 P|h_{jm}|^2\right) \\
&\quad - \log\left(1 + \sum_{m:m\neq i} \beta^2 P|h_{jm}|^2\right) \\
&\geq \log\left(1 + \alpha^2 P|h_{ji}|^2\right) \\
&\quad - \underbrace{\log\left(1 + \max_{m:m\neq i} \beta^2 P|h_{jm}|^2\right)}_{\widetilde{\ell}_{ji}} - \log(\Delta) \\
&\geq \ell_{ji} - \max_{i,j} \widetilde{\ell}_{ji} - \log(\Delta), \qquad (8.11)
\end{aligned}$$

where Δ is the maximum degree of the graph representing the network topology. We can then lower bound R_{TSN} as follows

$$\begin{aligned}
\mathsf{R}_{\text{TSN}} &= \max_{\substack{\lambda_s:\lambda_s\geq 0 \\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega:\Omega\subseteq[0:N], \\ 0\in\Omega}} \sum_{\substack{(i,j):i\in\Omega, \\ j\in\Omega^c}} \left(\sum_{\substack{s: \\ j\in s_{i,t},\ i\in s_{j,r}}} \lambda_s\right) \widehat{\ell}_{ji} \\
&\geq \max_{\substack{\lambda_s:\lambda_s\geq 0 \\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega:\Omega\subseteq[0:N], \\ 0\in\Omega}} \sum_{\substack{(i,j):i\in\Omega, \\ j\in\Omega^c}} \left(\sum_{\substack{s: \\ j\in s_{i,t},\ i\in s_{j,r}}} \lambda_s\right) \ell_{ji} \\
&\quad - N\log(\Delta) - N\max_{i,j} \widetilde{\ell}_{ji}. \qquad (8.12)
\end{aligned}$$

As a result, we have

$$|\mathsf{C}_{\text{ideal}} - \mathsf{R}_{\text{TSN}}| \leq N\log(\Delta) + N\max_{i,j} \widetilde{\ell}_{ji}. \qquad (8.13)$$

8.5 Appendices

8.5.1 Proof of Theorem 8.3.1

To prove Theorem 8.3.1, we focus on $\mathsf{C}(s,\Omega)$ below for all valid states s and cuts Ω

$$\mathsf{C}(s,\Omega) = \log\det\left(I + PH_{s,\Omega}H_{s,\Omega}^{\dagger}\right). \tag{8.14}$$

We seek to understand how this term relates to its counterpart in the expression of $\mathsf{C}_{\text{ideal}}$ in (8.4).

Without loss of generality, we assume that for the given s and Ω, the considered $H_{s,\Omega}$ matrix is wide, otherwise, we can consider the expression in (8.14) with the conjugate transpose matrix. We can get an upper bound on $\mathsf{C}(s,\Omega)$ in (8.14) as follows

$$\begin{aligned}\mathsf{C}(s,\Omega) &= \log\det\left(I + PH_{s,\Omega}H_{s,\Omega}^{\dagger}\right)\\ &\overset{\text{(a)}}{\leq} \sum_{j=1}^{|\Omega^c|}\log\left(1+P\max_{i\in[1:|\Omega|]}\left|\left[H_{s,\Omega}\right]_{ji}\right|^2\right) + |\Omega^c|\log(|\Omega|),\end{aligned} \tag{8.15}$$

where (a) follows from the Hadamard-Fischer inequality [RWH17] that upper bounds the determinant of a positive semidefinite matrix with the product of its diagonal elements.

By using our "main-lobe is always stronger" assumption, i.e., α and β are such that the side-lobe transmissions are weaker than those on the main-lobe, we can simplify (8.15) as

$$\mathsf{C}(s,\Omega) \leq \sum_{j=1}^{|\Omega^c|}\log\left(1+P\left|\left[H_{s,\Omega}\right]_{jj}\right|^2\right) + |\Omega^c|\log(|\Omega|). \tag{8.16}$$

Recall that each entry $\left[H_{s,\Omega}\right]_{jj}$ in the matrix $H_{s,\Omega}$ corresponds to the enhanced channel coefficient between two nodes that have their main-lobe beams aligned. Thus, we have the following upper bound on the approximate capacity $\mathsf{C}_{\text{cs,iid}}$ in (8.3b)

$$\mathsf{C}_{\text{cs,iid}} = \max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s\lambda_s=1}}\ \min_{\substack{\Omega\subseteq[0:N],\\ 0\in\Omega}}\mathsf{C}(s,\Omega)$$

$$\leq N\log(N) + \max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s \lambda_s = 1}} \min_{\substack{\Omega\subseteq[0:N],\\ 0\in\Omega}} \sum_s \lambda_s \sum_{\substack{(i,j)\in\Omega\times\Omega^c:\\ j\in s_{i,t},\ i\in s_{j,r}}} \ell_{ji}$$

$$= N\log(N) + \mathsf{C}_{\text{ideal}}. \tag{8.17}$$

We now want to find a lower bound for $\mathsf{C}_{\text{cs,iid}}$ in terms of $\mathsf{C}_{\text{ideal}}$. To do this, we will again focus on each of the terms $\mathsf{C}(s,\Omega)$ in (8.14). Recall that, by our "each cut is diagonally dominant" assumption, we have that for all (s,Ω) pairs, the matrix $I + PH_{s,\Omega}H_{s,\Omega}^\dagger$ is diagonally dominant.

For a diagonally dominant $n\times n$ matrix A, we can use the result in [Ost52] to derive the following lower bound on the determinant of A

$$\begin{aligned}\det(A) &\overset{\text{[Ost52]}}{\geq} \prod_{i=1}^n \left([A]_{ii} - \sum_{j\in[1:n]\setminus\{i\}} |[A]_{ij}| \right) \\ &\geq \prod_{i=1}^n \left([A]_{ii} - \rho_A [A]_{ii} \right) = (1-\rho_A)^n \prod_{i=1}^n A_{ii}, \end{aligned} \tag{8.18}$$

where ρ_A is given by

$$\rho_A = \max_{i\in[1:n]} \left\{ \sum_{j\in[1:n]\setminus\{i\}} \frac{|[A]_{ij}|}{|[A]_{ii}|} \right\}. \tag{8.19}$$

Now, by employing (8.18) on the matrix $I + PH_{s,\Omega}H_{s,\Omega}^\dagger$ in (8.14), we have that

$$\begin{aligned}\mathsf{C}(s,\Omega) &= \log\det\left(I + PH_{s,\Omega}H_{s,\Omega}^\dagger\right) \\ &\geq \sum_{j=1}^{|\Omega^c|} \log\left(1 + P\left\|[H_{s,\Omega}]_j\right\|^2\right) + \log\left((1-\rho_{s,\Omega}(H))^{|\Omega^c|}\right) \\ &\geq \sum_{j=1}^{|\Omega^c|} \log\left(1 + P\left|[H_{s,\Omega}]_{jj}\right|^2\right) - |\Omega^c|\,|\log(1-\rho_{s,\Omega}(H))| \\ &\geq \sum_{j=1}^{|\Omega^c|} \log\left(1 + P\left|[H_{s,\Omega}]_{jj}\right|^2\right) - N\underbrace{|\log(1-\rho_{s,\Omega}(H))|}_{f(\rho,s,\Omega)}. \end{aligned} \tag{8.20}$$

Thus, we have a lower bound on the approximate capacity as follows

$$\begin{aligned}
\mathsf{C}_{\text{cs,iid}} &= \max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s \lambda_s=1}} \min_{\substack{\Omega\subseteq[0:N],\\ 0\in\Omega}} \mathsf{C}(s,\Omega) \\
&\geq \max_{\substack{\lambda_s:\lambda_s\geq 0\\ \sum_s \lambda_s=1}} \min_{\substack{\Omega\subseteq[0:N],\\ 0\in\Omega}} \sum_s \lambda_s \left(\sum_{\substack{(i,j)\in\Omega\times\Omega^c:\\ j\in s_{i,t},\ i\in s_{j,r}}} \ell_{ji} - N f(\rho,s,\Omega) \right) \\
&\geq \mathsf{C}_{\text{ideal}} - N \max_{s,\Omega} f(\rho,s,\Omega) \\
&= \mathsf{C}_{\text{ideal}} - N \max_{s,\Omega} \left|\log(1-\rho_{s,\Omega}(H))\right|. \qquad (8.21)
\end{aligned}$$

By taking the maximum among the gaps in the bounds in (8.15) and (8.21), we get the result in Theorem 8.3.1.

8.5.2 Proof of Theorem 8.3.2

To prove Theorem 8.3.2, we would like to derive bounds on the pair (α,β) such that $f(H,\alpha,\beta)$ in (8.6) is upper bounded by $\log(N)$. By simple arithmetic manipulation, it is not difficult to see the following equivalence

$$\begin{aligned}
f(H,\alpha,\beta) = \max_{s,\Omega} \left|\log\left(1-\rho_{s,\Omega}(H)\right)\right| &\leq \log(N) \\
\iff \max_{s,\Omega} \rho_{s,\Omega}(H) &\leq \frac{N-1}{N}. \qquad (8.22)
\end{aligned}$$

We now find an upper bound on $\rho_{s,\Omega}(H)$ and then derive the sufficient condition in Theorem 8.3.2 by enforcing that the upper bound on $\rho_{s,\Omega}(H)$ is less than or equal to $(N-1)/N$, $\forall s,\Omega$.

From the definition of $\rho_{s,\Omega}(H)$ in (8.5), we can show that

$$\begin{aligned}
\rho_{s,\Omega}(H) &= \max_{i\in[1:|\Omega^c|]} \left\{ \sum_{j\in[1:|\Omega|]\setminus\{i\}} \frac{\left|P\left[H_{s,\Omega}\right]_i \left[H_{s,\Omega}\right]_j^\dagger\right|}{1+P\left[H_{s,\Omega}\right]_i \left[H_{s,\Omega}\right]_i^\dagger} \right\} \\
&\leq \max_{i\in[1:|\Omega^c|]} \left\{ \sum_{j\in[1:|\Omega|]\setminus\{i\}} \frac{\left|P\left[H_{s,\Omega}\right]_i \left[H_{s,\Omega}\right]_j^\dagger\right|}{P\alpha^2\left|\left[H_{s,\Omega}\right]_{ii}\right|^2} \right\}
\end{aligned}$$

$$= \max_{i\in[1:|\Omega^c|]} \left\{ \sum_{j\in[1:|\Omega|]\setminus\{i\}} \frac{\left|[H_{s,\Omega}]_i [H_{s,\Omega}]_j^\dagger\right|}{\alpha^2 \left|[H_{s,\Omega}]_{ii}\right|^2} \right\}. \tag{8.23}$$

Now, note that

$$\begin{aligned}\left|[H_{s,\Omega}]_i [H_{s,\Omega}]_j^\dagger\right| &\overset{(a)}{=} \left|\alpha\beta(\widehat{h}_{ii}\widehat{h}_{ji}^*+\widehat{h}_{ij}\widehat{h}_{jj}^*)+\beta^2 \sum_{k\neq i,j} \widehat{h}_{ik}\widehat{h}_{jk}^*\right| \\ &\overset{(b)}{\leq} \left[2\alpha\beta + (\Delta-2)\beta^2\right] \max_{i,j} |h_{ij}|^2, \end{aligned} \tag{8.24}$$

where: (a) uses $\widehat{h}_{ij} = [H_{s,\Omega}]_{ij}$ as in (8.2) for space limitation; (b) follows from the triangle inequality and the fact that, in the dot product $[H_{s,\Omega}]_i [H_{s,\Omega}]_j^\dagger$, there are at most Δ non-zero terms since Δ is the maximum degree of the graph representing the network topology.

By substituting (8.24) in (8.23), we have that

$$\begin{aligned}\rho_{s,\Omega}(H) &\leq \max_{i\in[1:|\Omega^c|]} \left\{ \sum_{j\in[1:|\Omega|]\setminus\{i\}} \frac{\left|[H_{s,\Omega}]_i [H_{s,\Omega}]_j^\dagger\right|}{\alpha^2 \left|[H_{s,\Omega}]_{ii}\right|^2} \right\} \\ &\leq \max_{i\in[1:|\Omega^c|]} \left\{ \sum_{j\in[1:|\Omega|]\setminus\{i\}} \frac{\left[2\alpha\beta + (\Delta-2)\beta^2\right] \max_{m,n} |h_{mn}|^2}{\alpha^2 \left|[H_{s,\Omega}]_{ii}\right|^2} \right\} \\ &\leq \frac{\left[2\alpha\beta+(\Delta-2)\beta^2\right](\Delta-1)}{\alpha^2} \max_{\substack{(i,j,m,n):|h_{ji}|>0,\\ i\neq j\neq m\neq n}} \frac{|h_{mn}|^2}{|h_{ij}|^2}. \end{aligned} \tag{8.25}$$

Given the upper bound on $\rho_{s,\Omega}(H)$ in (8.25) that is independent of (Ω, s), we can now use (8.22) and get the sufficient condition

$$\begin{aligned}&\frac{\left[2\alpha\beta+(\Delta-2)\beta^2\right](\Delta-1)}{\alpha^2} \max_{\substack{(i,j,m,n):|h_{ji}|>0,\\ i\neq j\neq m\neq n}} \frac{|h_{mn}|^2}{|h_{ij}|^2} \leq \frac{N-1}{N} \\ &\implies \frac{\alpha}{\beta} \geq \Delta^2 \frac{N}{N-1} \max_{\substack{(i,j,m,n):|h_{ji}|>0,\\ i\neq j\neq m\neq n}} \frac{|h_{mn}|^2}{|h_{ij}|^2}. \end{aligned}$$

This concludes the proof of Theorem 8.3.2.

CHAPTER 9

Conclusion and Open Questions

In this dissertation, we discussed two enabling aspects for communication next-generation wireless networks: multi-hopping and mmWave transmission. Our key focus throughout the dissertation was on providing fundamental guarantees for operating a subset of these multi-hop networks (through the network simplification framework) and developing efficient approaches for selecting and scheduling these networks/subnetworks.

9.1 Network Simplification in Full-Duplex Networks

For the problem of network simplification in full-duplex networks, we considered arbitrary topology relay networks: N relays arranged arbitrarily to assist communication between a source and a destination node. We prove fundamental worst-case guarantees on selecting the best route in the network and show that the fraction guaranteed is inversely proportional with the number of nodes in the network. This presents the first such result on network simplification that extends beyond the diamond network topology. Towards proving these results, we proved an auxiliary fundamental guarantee for selecting a $k_t \times k_r$ MIMO sub-channel from an existing $n_t \times n_r$ MIMO channel, which can be of independent interest given the increasing deployment of multiple-antenna systems in next-generation networks.

In this line of work, there is a number of interesting open questions. First, the extension of the network simplification result to arbitrary topology was possible for the single route case but choosing larger subnetworks (for instance choosing k relays per layer in a

layered network topology) is still an open question. Selecting larger subnetworks would allow the information-flow to make use of physical layer cooperation schemes developed in the information theory community with a reduced level complexity that is associated with their implementation across nodes in then network. Another open question is how do these fundamental bounds behave on average. Given a predefined distribution for independently sampled network channel coefficients, what guarantees can be provided in expectation on the retained approximate capacity from selecting a subnetwork from an N-relay network. These expectation bounds would retain dependence on the underlying network topology, but may not be limited by corner low-probability instances.

9.2 Network Simplification in Half-Duplex Networks

Next, we discussed the network simplification problem in half-duplex diamond networks. We provided fundamental worst-case guarantees on selecting the best subnetwork of $N-1$ out of N relays. The proof of this fundamental bound involved exploring and deriving equivalent definition of submodular functions in terms of N sets, which can be independent interest beyond the scope of this dissertation. Next, we presented efficient algorithms for selecting a subnetwork of $1 < k < N$ relays with theoretical guarantees on the retained approximate capacity.

Several open questions remain in this line of work. First, our fundamental worst-case bounds for the diamond topology are proved to be tight only for the case where we are selecting $k = N - 1$ relays out of N. In fact, we believe that the derived guarantees can be tightened for almost all $k < N - 1$. In fact, most recently the authors in [JMC20] derived a tight fundamental worst-case bound for the case of selecting $k = 1$ relays in the diamond network, which include sinusoidal values in terms of the number of relays N. Bounds for the case of $2 \leq k \leq N - 2$ remain an open problem in this line of work. Beyond the diamond network, it is also interesting to understand how these bounds extend even for simple values

of k relays to arbitrary topology half-duplex networks. For all aforementioned cases, it still remains an open direction to understand whether efficient selection algorithms can be coupled with these fundamental bounds. Such selection algorithms for the case of $k = 1$ and $k = N - 1$ are straightforward, however for different values of k and in different topologies, such algorithms may not be as direct. In fact, based on our work on half-duplex routing, selecting the best single route in an arbitrary topology network is NP-hard and selecting different types of subnetworks may prove to be as difficult. This is an open problem for future investigation.

9.3 Information-theoretic Modeling of mmWave Networks

The second part of this dissertation focused on the modeling of information flow in mmWave networks and how this network model can be operated efficiently. A key characteristic of mmWave networks used in 5G mobile networks is the highly-directive short-range transmission used to counter path loss. This feature is not explicitly captured in current information-theoretic models for multi-hop wireless networks. Instead, broadcast communication is assumed from a node to all neighboring devices in range as well as interference due to signals superposition at a receiving node. To inherently model the directivity property of mmWave networks and the need to schedule beam orientations, we proposed a new information-theoretic model for multi-hop wireless networks referred to as "1-2-1 network". We first used this model to characterize the Shannon unicast and multicast capacities, up to a constant gap, for mmWave networks with arbitrary topology operating with full-duplex or half-duplex mmWave nodes. Next, we developed a provably optimal polynomial time algorithm to compute the approximate capacity and an optimal beam-orientation schedule in full-duplex and half-duplex mmWave network. Finally, we studied the impact of relaxing the perfect beamforming modeled in the full-duplex 1-2-1 network model to the case where signal leakage can occur from side-lobes and showed that within some operational range of the beamforming

parameters, the ideal 1-2-1 network is still a viable constant gap approximation of imperfect case.

Several open research questions remain. For the ideal 1-2-1 network model, the scheduling solutions are centralized in nature: a linear program is solved using knowledge of the different link capacities in the network. This inherently imposes an overhead of collecting these link capacities and can potentially limit the application of these solutions to fixed backhaul mmWave deployments. An ambitious goal is to develop distributed solutions that are guaranteed to achieve (or approximately achieve) the same rates by the centralized scheduling approaches. Another open direction to explore in the context of scheduling 1-2-1 networks is the resilience of the scheduling solution to time-varying link capacities in the network. A more attractive scheduling approach will require fewer edits or updates to achieve the maximum communication rate of the network after a few link capacities in the network have changed. In imperfect 1-2-1 network, the same set of questions naturally follow. In addition, although we show that the ideal model can approximate the imperfect model under some sufficient conditions, this only extends as far as projecting the scheduling solution and operational properties of the network from the ideal model onto the imperfect 1-2-1 model. To operate the network, to the best of our knowledge, we still need to employ complicated schemes from network information theory such as [ADT11, LKE11]. It remains an open problem to understand whether simpler solutions can be used to achieve the approximate capacity under the regime of beamforming parameters that couple the imperfect and ideal models.

9.4 Other Explored Research Problems

The focus of our discussion in this dissertation was on operating next-generation wireless networks in terms of scheduling and simplification of multi-hop communication. Another intriguing aspect of next-generation networks is the evolving nature of applications that are

envisioned on top of their deployments to include computationally heavy machine learning applications that are distributed over the network or in general distributed computation solutions that makes use of the wireless medium and its flexible network deployment cost. In this section, we discuss three such applications over next-generation networks that were studied during the course of this doctoral degree but that deviate from the central theme of this dissertation.

Distributed MapReduce over broadcast medium. Distributed computation across a set of wireless networked servers is well motivated for several practical constraints: we may want to speed up computation time so as to finish a computation faster; we may have partial view of the files needed for computation across servers; we may have limited memory in each server; or we may be motivated by energy constraints. In this work, we considered the distributed computing framework over a broadcast medium studied that follows the architecture of MapReduce [DG08]. We introduced a communication-computation-storage tradeoff in this system that expands the considered system parameters studied in [LMA16] (communication-computation). We proposed a scheme for minimizing computation given communication and storage constraints. This was later proved to be optimal by [YYW18].

Distributed quantization for classification tasks. In this line of work, we considered the problem of distributed feature quantization, where the goal is to enable a pretrained classifier at a central node to carry out its classification on features that are gathered from distributed nodes through communication constrained channels. This is motivated by applications in wireless cyberphysical systems, immersive environments and supported health. We proposed the design of distributed quantization schemes specifically tailored to the classification task: unlike quantization schemes that help the central node reconstruct the original signal as accurately as possible, our focus is not reconstruction accuracy, but instead correct classification. Our designs leveraged discrete neural representations and training data, and could be designed in polynomial time for any number of features, any number of classes,

and arbitrary division of features across the distributed nodes. We found that tailoring the quantizers to the classification task can offer significant savings: as compared to alternatives, we could achieve more than a factor of two reduction in terms of the number of bits communicated, for the same classification accuracy.

Security in mmWave Networks. In next-generation networks, a large portion of the enormous data to be exchanged is sensitive in nature, such as banking, health, personal, and IoT control packets. Therefore, we need to securely exchange this sensitive information against eavesdropping adversaries who have interests in gaining access to this information. We consider the problem of information-theoretic security in arbitrary 1-2-1 networks that model mmWave networks, and derive lower and upper bounds on the secure capacity. Since we need to use beamforming and align beams to activate links, we cannot use all the underlying links of the network simultaneously. However, the degree of freedom in choosing the links to activate can be leveraged for secure communication against an eavesdropper. We show that we can achieve a secure capacity that in some cases, can be very close to the unsecure capacity.

REFERENCES

[AA16] Tulika Agrawal and Samar Agnihotri. "Network Simplification for Secure AF Relaying." *arXiv preprint arXiv:1607.00185*, 2016.

[ADT11] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse. "Wireless Network Information Flow: A Deterministic Approach." *IEEE Transactions on Information Theory*, **57**(4):1872–1905, April 2011.

[AHR04] Baruch Awerbuch, David Holmer, and Herbert Rubens. "High throughput route selection in multi-rate ad hoc wireless networks." *Wireless On-Demand Network Systems*, pp. 201–205, 2004.

[AJC12] S. Agnihotri, S. Jaggi, and M. Chen. "Analog Network Coding in General SNR Regime: Performance of Network Simplification." In *IEEE Information Theory Workshop (ITW)*, pp. 632–636, September 2012.

[AL09] Mustaq Ahmed and Anna Lubiw. "Shortest Paths Avoiding Forbidden Subpaths." In *26th International Symposium on Theoretical Aspects of Computer Science STACS 2009*, pp. 63–74. IBFI Schloss Dagstuhl, 2009.

[Are81] Mohammad R Aref. "Information flow in relay networks." *Ph.D. thesis, Stanford University*, 1981.

[Ari84] Erdal Arikan. "Some complexity results about packet radio networks (Corresp.)." *IEEE Transactions on Information Theory*, **30**(4):681–685, 1984.

[BAH14] T. Bai, A. Alkhateeb, and R. W. Heath. "Coverage and capacity of millimeter-wave cellular networks." *IEEE Communications Magazine*, **52**(9):70–77, Sep. 2014.

[Bar82] Imre Bárány. "A generalization of Carathéodory's theorem." *Discrete Mathematics*, **40**(2-3):141–152, 1982.

[BF14a] S. Brahma and C. Fragouli. "A simple relaying strategy for diamond networks." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 1922–1926, June 2014.

[BF14b] S. Brahma and C. Fragouli. "Structure of optimal schedules in diamond networks." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 641–645, June 2014.

[BFO16] Siddhartha Brahma, Christina Fragouli, and Ayfer Özgür. "On the complexity of scheduling in half-duplex diamond networks." *IEEE Transactions on Information Theory*, **62**(5):2557–2572, 2016.

[Bir46] Garrett Birkhoff. "Tres observaciones sobre el algebra lineal." *Univ. Nac. Tucuman, Ser. A*, **5**:147–154, 1946.

[BM11] A. Bondy and U.S.R. Murty. *Graph Theory*. Graduate Texts in Mathematics. Springer London, 2011.

[BMJ98] Josh Broch, David A Maltz, David B Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. "A performance comparison of multi-hop wireless ad hoc network routing protocols." In *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, pp. 85–97. ACM, 1998.

[BMK14] H. Bagheri, A.S. Motahari, and A.K. Khandani. "On the Capacity of the Half-Duplex Diamond Channel Under Fixed Scheduling." *IEEE Transactions on Information Theory*, **60**(6):3544–3558, June 2014.

[BOF12] S. Brahma, A. Özgür, and C. Fragouli. "Simple schedules for half-duplex networks." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 1112–1116, July 2012.

[Bro] "What Role Will Millimeter Waves Play in 5G Wireless Systems?" `https://www.mwrf.com/systems/what-role-will-millimeter-waves-play-5g-wireless-systems`.

[BSF14] S. Brahma, A. Sengupta, and C. Fragouli. "Switched local schedules for diamond networks." In *IEEE Information Theory Workshop (ITW)*, pp. 656–660, November 2014.

[CCH99] Cheng-Shang Chang, Wen-Jyh Chen, and Hsiang-Yi Huang. "On service guarantees for input-buffered crossbar switches: a capacity decomposition approach by Birkhoff and von Neumann." In *IEEE Seventh International Workshop on Quality of Service (IWQoS)*, pp. 79–86. IEEE, 1999.

[CCH07] Tao Cui, Lijun Chen, and Tracey Ho. "Distributed optimization in wireless networks using broadcast advantage." In *46th IEEE Conference on Decision and Control*, pp. 5839–5844, 2007.

[CE79] T. Cover and A. El Gamal. "Capacity theorems for the relay channel." *IEEE Transactions on Information Theory*, **25**(5):572 – 584, September 1979.

[CEF19] Martina Cardone, Yahya H Ezzeldin, Christina Fragouli, and Daniela Tuninetti. "Network simplification in half-duplex: Building on submodularity." *IEEE Transactions on Information Theory*, 2019.

[CG79] Thomas Cover and A EL Gamal. "Capacity theorems for the relay channel." *IEEE Transactions on information theory*, **25**(5):572–584, 1979.

[CJ03] Thomas Clausen and Philippe Jacquet. "Optimized Link State Routing Protocol (OLSR)." *RFC 3626, DOI 10.17487/RFC3626*, 2003.

[CO12] B. Chern and A. Özgür. "Achieving the capacity of the N-relay Gaussian diamond network within logn bits." In *IEEE Information Theory Workshop (ITW)*, pp. 377–380, September 2012.

[CO15] T. A. Courtade and A. Özgür. "Approximate capacity of Gaussian relay networks: Is a sublinear gap to the cutset bound plausible?" In *IEEE International Symposium on Information Theory (ISIT)*, pp. 2251–2255, June 2015.

[CT12] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[CTK14] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim. "Gaussian Half-Duplex Relay Networks: improved constant gap and connections with the assignment problem." *IEEE Transactions on Information Theory*, **60**(6):3559 – 3575, June 2014.

[CTK15] M. Cardone, D. Tuninetti, and R. Knopp. "The approximate optimality of simple schedules for half-duplex multi-relay networks." In *IEEE Information Theory Workshop (ITW)*, pp. 1–5, April 2015.

[CTK16] M. Cardone, D. Tuninetti, and R. Knopp. "On the Optimality of Simple Schedules for Networks With Multiple Half-Duplex Relays." *IEEE Transactions on Information Theory*, **62**(7):4120–4134, July 2016.

[CVG16] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath. "Millimeter-Wave Vehicular Communication to Support Massive Automotive Sensing." *IEEE Communications Magazine*, **54**(12):160–167, December 2016.

[DAB05] Douglas SJ De Couto, Daniel Aguayo, John Bicket, and Robert Morris. "A high-throughput path metric for multi-hop wireless routing." *Wireless networks*, **11**(4):419–434, 2005.

[DFP17] Z. Ding, P. Fan, and H. V. Poor. "Random Beamforming in Millimeter-Wave NOMA Networks." *IEEE Access*, **5**:7667–7681, 2017.

[DG08] Jeffrey Dean and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." *Communications of the ACM*, **51**(1):107–113, 2008.

[Dij59] Edsger W Dijkstra. "A note on two problems in connexion with graphs." *Numerische mathematik*, **1**(1):269–271, 1959.

[DSA14] Melissa Duarte, Ashutosh Sabharwal, Vaneet Aggarwal, Rittwik Jana, KK Ramakrishnan, Christopher W Rice, and NK Shankaranarayanan. "Design and characterization of a full-duplex multiantenna system for WiFi networks." *IEEE Transactions on Vehicular Technology*, **63**(3):1160–1177, 2014.

[DTF17] Junquan Deng, Olav Tirkkonen, Ragnar Freij-Hollanti, Tao Chen, and Navid Nikaein. "Resource allocation and interference management for opportunistic relaying in integrated mmwave/sub-6 GHz 5G networks." *IEEE Communications Magazine*, **55**(6):94–101, 2017.

[Dul55] Lloyd Dulmage. "On a theorem of Frobeniuskonig and J. von Neumann's game of hide and seek." *Transactions of the Royal Society of Canada, Section III*, **49**:23, 1955.

[ECF17] Y. H. Ezzeldin, M. Cardone, C. Fragouli, and D. Tuninetti. "Finding Simple Half-Duplex Schedules in Gaussian Relay Line Networks." *to appear in IEEE International Symposium on Information Theory (ISIT)*, June 2017.

[ECF18] Yahya H Ezzeldin, Martina Cardone, Christina Fragouli, and Giuseppe Caire. "Gaussian 1-2-1 networks: Capacity results for mmwave communications." In *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2569–2573. IEEE, 2018.

[ECF19a] Yahya H Ezzeldin, Martina Cardone, Christina Fragouli, and Giuseppe Caire. "Polynomial-time capacity calculation and scheduling for half-duplex 1-2-1 networks." In *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 460–464. IEEE, 2019.

[ECF19b] Yahya H Ezzeldin, Martina Cardone, Christina Fragouli, and Daniela Tuninetti. "Network simplification in half-duplex: Building on submodularity." *IEEE Transactions on Information Theory*, **65**(10):6801–6818, 2019.

[ECF20a] Yahya H Ezzeldin, Martina Cardone, Christina Fragouli, and Giuseppe Caire. "Gaussian 1-2-1 Networks with Imperfect Beamforming." *arXiv preprint arXiv:2005.06655*, 2020.

[ECF20b] Yahya H Ezzeldin, Martina Cardone, Christina Fragouli, and Daniela Tuninetti. "The Approximate Capacity of Half-Duplex Line Networks." *IEEE Transactions on Information Theory*, 2020.

[Edm65] Jack Edmonds. "Maximum matching and a polyhedron with 0, 1-vertices." *Journal of Research of the National Bureau of Standards B*, **69**(125-130):55–56, 1965.

[EK11] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge university press, 2011.

[EPS14] R.H. Etkin, F. Parvaresh, I. Shomorony, and A.S. Avestimehr. "Computing Half-Duplex Schedules in Gaussian Relay Networks via Min-Cut Approximations." *IEEE Transactions on Information Theory*, **60**(11):7204–7220, November 2014.

[ESF16] Yahya H Ezzeldin, Ayan Sengupta, and Christina Fragouli. "Wireless network simplification: Beyond diamond networks." In *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2594–2598. IEEE, 2016.

[ESF20] Yahya H Ezzeldin, Ayan Sengupta, and Christina Fragouli. "Wireless Network Simplification: The Performance of Routing." *IEEE Transactions on Information Theory*, 2020.

[ESZ16] Evan Everett, Clayton Shepard, Lin Zhong, and Ashutosh Sabharwal. "SoftNull: Many-Antenna Full-Duplex Wireless via Digital Beamforming." *IEEE Transactions on Wireless Communications*, **15**(12):8077–8092, 2016.

[GGP03] Alexei Gorokhov, Dhananjay A Gore, and Arogyaswami J Paulraj. "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms." *IEEE Transactions on signal processing*, **51**(11):2796–2807, 2003.

[GH61] Ralph E Gomory and Tien Chung Hu. "Multi-terminal network flows." *Journal of the Society for Industrial and Applied Mathematics*, **9**(4):551–570, 1961.

[GK00] P. Gupta and P. R. Kumar. "The capacity of wireless networks." *IEEE Transactions on Information Theory*, **46**(2):388–404, 2000.

[GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver. "The ellipsoid method and its consequences in combinatorial optimization." *Combinatorica*, **1**(2):169–197, 1981.

[GLS12] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[GTM16] Xiaohu Ge, Song Tu, Guoqiang Mao, Cheng-Xiang Wang, and Tao Han. "5G ultra-dense cellular networks." *IEEE Wireless Communications*, **23**(1):72–79, 2016.

[Hea16] R. Heath. "Vehicle-to-X communication using millimeter waves." 2016.

[HEJ10] T. Ho, M. Effros, and S. Jalali. "On equivalence between network topologies." In *48th Annual Allerton Conference on Communication, Control, and Computing*, pp. 391–398, September 2010.

[HS88] Bruce Hajek and Galen Sasaki. "Link scheduling in polynomial time." *IEEE Transactions on Information Theory*, **34**(5):910–917, 1988.

[JEC19] S. Jain, M. Elyasi, M. Cardone, and S. Mohajer. "On Simple Scheduling in Half-Duplex Relay Diamond Networks." In *IEEE International Symposium on Information Theory (ISIT)*, July 2019.

[JEH11] S. Jalali, M. Effros, and T. Ho. "On the impact of a single edge on the network coding capacity." In *Information Theory and Applications Workshop (ITA)*, pp. 1–5, February 2011.

[JM96] David B Johnson and David A Maltz. "Dynamic Source Routing in Ad Hoc Wireless Networks." In *Mobile computing*, pp. 153–181. Springer, 1996.

[JMC20] Sarthak Jain, Soheil Mohajer, and Martina Cardone. "Best Relay Selection in Gaussian Half-Duplex Diamond Networks." *arXiv preprint arXiv:2001.02851*, 2020.

[JV09] Yi Jiang and M. K. Varanasi. "The RF-Chain Limited MIMO System- Part I: Optimum Diversity-Multiplexing Tradeoff." *IEEE Transactions on Wireless Communications*, **8**(10):5238–5247, October 2009.

[JX06] Mikael Johansson and Lin Xiao. "Cross-layer optimization of wireless networks using nonlinear column generation." *IEEE Transactions on Wireless Communications*, **5**(2):435–445, 2006.

[KAH17] Hlabishi I Kobo, Adnan M Abu-Mahfouz, and Gerhard P Hancke. "A survey on software-defined wireless sensor networks: Challenges and design requirements." *IEEE Access*, **5**:1872–1899, 2017.

[Kar72] Richard M Karp. "Reducibility among combinatorial problems." In *Complexity of computer computations*, pp. 85–103. Springer, 1972.

[KM17] Alper Köse and Muriel Médard. "Scheduling wireless ad hoc networks in polynomial time using claw-free conflict graphs." In *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–7. IEEE, 2017.

[Kra04] G. Kramer. "Models and theory for relay channels with receive constraints." In *42nd Annual Allerton Conference on Communication, Control, and Computing*, pp. 1312–1321, September 2004.

[LA15] X. Lin and J. G. Andrews. "Connectivity of Millimeter Wave Networks With Multi-Hop Relaying." *IEEE Wireless Communications Letters*, **4**(2):209–212, April 2015.

[LHK12] Tobias Lutz, Christoph Hausl, and Ralf Kotter. "Bits through deterministic relay cascades with half-duplex constraint." *IEEE transactions on information theory*, **58**(1):369–381, 2012.

[LKE11] S.H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung. "Noisy Network Coding." *IEEE Transactions on Information Theory*, **57**(5):3132 –3152, May 2011.

[LKK14] S. H. Lim, K. T. Kim, and Y. H. Kim. "Distributed decode-forward for multicast." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 636–640, June 2014.

[LMA16] Songze Li, Mohammad Ali Maddah-Ali, and A Salman Avestimehr. "Fundamental tradeoff between computation and communication in distributed computing." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 1814–1818, 2016.

[Mos07] Thomas Moscibroda. "The worst-case capacity of wireless sensor networks." In *2007 6th International Symposium on Information Processing in Sensor Networks*, pp. 1–10. IEEE, 2007.

[MSK16] M. Mueck, E. C. Strinati, I. Kim, A. Clemente, J. Dore, A. De Domenico, T. Kim, T. Choi, H. K. Chung, G. Destino, A. Parssinen, A. Pouttu, M. Latva-aho, N. Chuberre, M. Gineste, B. Vautherin, M. Monnerat, V. Frascolla, M. Fresia, W. Keusgen, T. Haustein, A. Korvala, M. Pettissalo, and O. Liinamaa. "5G CHAMPION - Rolling out 5G in 2018." In *IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, December 2016.

[MWC05] Andreas F Molisch, Moe Z Win, Yang-Seok Choi, and Jack H Winters. "Capacity of MIMO systems with antenna selection." *IEEE Transactions on Wireless Communications*, **4**(4):1759–1772, 2005.

[NEO11] C. Nazaroglu, J. B. Ebrahimi, A. Özgür, and C. Fragouli. "Network Simplification: The Gaussian Diamond Network with Multiple Antennas." In *IEEE International Symposium on Information Theory (ISIT)*, pp. 79–83, July 2011.

[NOF14] C. Nazaroglu, A. Özgür, and C. Fragouli. "Wireless Network Simplification: The Gaussian N-Relay Diamond Network." *IEEE Transactions on Information Theory*, **60**(10):6329–6341, October 2014.

[Nor19] Amy Nordrum. "Facebook pushes networking tech: The company's Terragraph technology will soon be available in commercial gear-[News]." *IEEE Spectrum*, **56**(4):8–9, 2019.

[OD13] A. Özgür and S. N. Diggavi. "Approximately Achieving Gaussian Relay Network Capacity With Lattice-Based QMF Codes." *IEEE Transactions on Information Theory*, **59**(12):8275–8294, December 2013.

[OMJ12] Lawrence Ong, Mehul Motani, and Sarah J Johnson. "On capacity and optimal scheduling for the half-duplex multiple-relay channel." *IEEE Transactions on Information Theory*, **58**(9):5770–5784, June 2012.

[Ost52] Alexander M Ostrowski. "Note on bounds for determinants with dominant principal diagonal." *Proceedings of the American Mathematical Society*, **3**(1):26–30, 1952.

[Pap16] 5GMF White Paper. "5G Mobile Communications systems for 2020 and beyond." July 2016.

[PB94] Charles E Perkins and Pravin Bhagwat. "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers." In *ACM SIGCOMM Computer Communication Review*, volume 24, pp. 234–244, October 1994.

[PE14] F. Parvaresh and R. Etkin. "Efficient Capacity Computation and Power Optimization for Relay Networks." *IEEE Transactions on Information Theory*, **60**(3):1782–1792, March 2014.

[Pol60] Maurice Pollack. "Letter to the Editor – The Maximum Capacity Through a Network." *Operations Research*, **8**(5):733–736, 1960.

[PR82] Manfred W Padberg and M Ram Rao. "Odd minimum cut-sets and b-matchings." *Mathematics of Operations Research*, **7**(1):67–80, 1982.

[PR99] Charles E Perkins and Elizabeth M Royer. "Ad-hoc On-Demand Distance Vector Routing." In *Proceedings of the Second IEEE Workshop on Mobile Computer Systems and Applications*, p. 90, 1999.

[RRE14] S. Rangan, T.S. Rappaport, and E. Erkip. "Millimeter-wave Cellular Wireless Networks: Potentials and Challenges." In *Proceedings of IEEE*, volume 102, pp. 366–385, March 2014.

[RWH17] Michał Różański, Roman Wituła, and Edyta Hetmaniok. "More subtle versions of the Hadamard inequality." *Linear Algebra and its Applications*, **532**:500–511, 2017.

[SC20] Xiaoshen Song and Giuseppe Caire. "Queue-Aware Beam Scheduling for Half-Duplex mmWave Relay Networks." *arXiv preprint arXiv:2001.05586*, 2020.

[SMK07] Jay Kumar Sundararajan, Muriel Médard, MinJi Kim, Atilla Eryilmaz, Devavrat Shah, and Ralf Koetter. "Network coding in a multicast switch." In *IEEE 26th International Conference on Computer Communications (INFOCOM)*, pp. 1145–1153. IEEE, 2007.

[SN04] Shahab Sanayei and Aria Nosratinia. "Antenna selection in MIMO systems." *IEEE Communications Magazine*, **42**(10):68–73, 2004.

[SWF12] A. Sengupta, I-Hsiang Wang, and C. Fragouli. "Optimizing Quantize-Map-and-Forward relaying for Gaussian diamond networks." In *IEEE Information Theory Workshop (ITW)*, pp. 381–385, September 2012.

[TBH16] A. Thornburg, T. Bai, and R. W. Heath. "Performance Analysis of Outdoor mmWave Ad Hoc Networks." *IEEE Transactions on Signal Processing*, **64**(15):4065–4079, August 2016.

[THM12] Danail Traskov, Michael Heindlmaier, Muriel Médard, and Ralf Koetter. "Scheduling for network-coded multicast." *IEEE/ACM Transactions on Networking*, **20**(5):1479–1488, 2012.

[Tho68] R. C. Thompson. "Principal submatrices V: Some results concerning principal submatrices of arbitrary matrices." *Journal of Research of the Natural Bureau of Standards, Section B - Mathematical Sciences*, **72**:115–125, April 1968.

[WLA17] Y-P Eric Wang, Xingqin Lin, Ansuman Adhikary, Asbjorn Grovlen, Yutao Sui, Yufei Blankenship, Johan Bergman, and Hazhir S Razaghi. "A primer on 3gpp narrowband internet of things." *IEEE Communications Magazine*, **55**(3):117–123, 2017.

[WO15] X. Wu and A. Özgür. "Cut-set bound is loose for Gaussian Relay Networks." In *53rd Annual Allerton Conference on Communication, Control, and Computing*, pp. 1135–1142, October 2015.

[YPK03] Su Yi, Yong Pei, and Shivkumar Kalyanaraman. "On the capacity improvement of ad hoc wireless networks using directional antennas." In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pp. 108–116. ACM, 2003.

[YYW18] Qifa Yan, Sheng Yang, and Michèle Wigger. "Storage, computation, and communication: A fundamental tradeoff in distributed computing." In *2018 IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 2018.

[ZJS15] N. Zlatanov, V. Jamali, and R. Schober. "On the Capacity of the Two-Hop Half-Duplex Relay Channel." In *2015 IEEE Global Communications Conference (GLOBECOM), arXiv:1411.5299*, pp. 1–7, December 2015.

[ZLZ14] Degan Zhang, Guang Li, Ke Zheng, Xuechao Ming, and Zhao-Hua Pan. "An energy-balanced routing method based on forward-aware factor for wireless sensor networks." *IEEE transactions on industrial informatics*, **10**(1):766–773, 2014.

[ZSS19] Tengchan Zeng, Omid Semiari, Walid Saad, and Mehdi Bennis. "Joint communication and control for wireless autonomous vehicular platoon systems." *IEEE Transactions on Communications*, **67**(11):7907–7922, 2019.