Decisions in Distributed Wireless Networks with Imprecise Information

Vaneet Aggarwal

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Electrical Engineering
Adviser: A. Robert Calderbank

September 2010

Abstract

The use of wireless technology is rapidly growing. The demand is so huge that the limited supply of resources is becoming the bottleneck. Thus, network designs need to be rethought. Most of the analyses to date consider complete network information, perfect knowledge of channel state at the receivers, perfect knowledge of location of destination or perfect feedback link. This is an idealization and new design strategies accounting for the imperfect or incomplete information are needed. In this thesis, we will consider the effect of various forms of incomplete and imperfect knowledge motivated by practical protocol designs. The basic theme of the results is an old adage "If we know more, we can achieve more." This thesis applies this adage to networks, where more information about the network translates into higher throughput or diversity.

We will first study a diversity multiplexing tradeoff for both frequency division duplex (FDD) and time division duplex (TDD) systems, when both receiver and transmitter knowledge about the channel is noisy and potentially mismatched. We break the mold of all current channel state based protocols by using multiple rounds of conferencing to extract more bits about the actual channel. Multiple rounds of conferencing provide more refined information about the network at the nodes, leading to improved diversity order with every round of communication. The protocols are on-demand in nature, using high powers for training and feedback only when the channel is in poor states. The key result in FDD systems is that the diversity multiplexing tradeoff with perfect training and K levels of perfect feedback can be achieved, even when there are errors in training the receiver and errors in the feedback link, with a multi-round protocol which has K rounds of training and $K-1$ rounds of binary feedback. For TDD systems, we also develop new achievable strategies with multiple rounds of communication between the transmitter and the receiver, which use the reciprocity of the forward and the feedback channel. The multi-round TDD protocol achieves a diversity-multiplexing tradeoff which uniformly dominates its FDD counterparts, where no channel reciprocity is available.

We will then focus on the case when the destination is mobile and the placement of base station and the relay station in the network has to be decided. To make progress, we introduce an alternative perspective where the objective is maximizing *coverage* for a given rate. The new objective captures the problem of how to deploy relays to provide a given level of service to a particular geographic area, where the relay locations become a design parameter that can be optimized. We evaluate the decode and forward (DF) and compress and forward (CF) strategies for the relay channel with respect to the new objective of maximizing coverage. When the objective is maximizing rate, different locations of the destination favor different strategies. When the objective is coverage for a given rate, and the relay is able to decode, DF is uniformly superior in that it provides coverage at any point served by CF. While the coverage provided by DF is sensitive to changes in the location of the relay and the path loss exponent, CF exhibits a more graceful degradation with respect to such changes.

Finally, we formalize the increase of sum-rate with increased knowledge of the network state in an interference network. The knowledge of network state is measured in terms of the number of hops of information available to each node and is labeled each node's local view. To understand how much capacity is lost due to limited information, we propose to use the metric of normalized sum-capacity, which is the h-hop local view sum-capacity divided by global-view sum-capacity. For the cases of one and two-local view, we characterize the normalized sum-capacity for many classes of deterministic and Gaussian interference networks. In many cases, a scheduling scheme called maximal independent graph scheduling is shown to achieve normalized sum-capacity. We also show that its generalization for one-hop local view, labeled coded maximal independent set scheduling, achieves capacity whenever its uncoded counterpart fails to do so.

Acknowledgements

First and foremost, I would like to express deep and sincere gratitude to my adviser, Professor Robert Calderbank for his generous encouragement, support, and guidance during the past five years. He has habituated me how to choose research problems, and how to manage different research problems. One of the great pieces of wisdom he gave me is that it is always better to think of more problems in parallel so that if you get stuck on one, there is always other to think about rather than wasting time in which no new idea emerges. Further, he taught me that there should always be a long term problem which you think will be significant but is not very well formalized and a short term problem which is more formalized. This has led me to work on different problems in the last five years even though I was not able to solve completely many of the problems that I had started working on. I am very fortunate to have such an adviser. He has always encouraged me to find my own problems and collaborators to work on them which has led me to enjoy research discussions with many great researchers from academia and research labs and increased my motivation in research. It is his unique way of approaching the problems that has led me to do research and finally I am getting a feeling of great accomplishment.

I also wish to thank profusely Professor Ashustosh Sabharwal for his encouragement and expertise advise in the last three years. There are three stages to a research problem - problem formulation, problem solving and presenting the problems and solutions to a wider audience. He has given me guidance at each step of the research. At different stages of my graduate student life, his suggestions were appropriate for that level of maturity in research. His enthusiasm for research and his exceptional way of presenting and teaching always provides me an encouragement and motivation. I would also like to thank Professor Ashutosh Sabharwal and Professor Suhas Diggavi for my research internship visits to Rice University and EPFL, Switzerland respectively. These visits helped me to learn more about the research going over at different places and broaden my horizon.

I wish to express my sincere thanks to Professor Sharad Malik for his help in getting

me admitted to Princeton University and Professor Sun-Yuan Kung for advising me related to academia and research in my first year at Princeton University. I would also like to thank Professor Rakesh Bansal, my undergraduate thesis adviser at the Indian Institute of Technology Kanpur and other faculty at I.I.T. Kanpur for their role in me being able to boost today that I am graduating with a doctoral degree from a premier institution in the world. I also acknowledge the support of Professor Daniel Mlynek from EPFL, Professor Mukul Goyal from University of Winsconsin Milwaukee for guiding me in research in the undergraduate and for writing my references for the graduate school.

I am very grateful to Professor Salman Avestimehr (Cornell), Professor Youjian Liu (Colorado Boulder), Professor Vincent Poor (Princeton), Professor Aylin Yener (Penn State) for many thoughtful discussions. I also acknowledge the researchers at AT&T, MITRE Research, Bell Labs and Siemens, that I got a chance to work with in my doctoral studies. In specific, thanks to Rittwik Jana, K. K. Ramakrishnan, Vijay Gopalakrishnan, Christopher Rice, and Jack Murray from AT&T, Gerald Gilbert, Michael Hamrick, and Yaakov Weinstein from MITRE Research, Alexei Ashikhmin from Bell Labs and Aik Chindapol from Siemens.

I am thankful to Professor Vincent Poor and Professor Sergio Verdu for serving on my oral generals committee and guiding me in the research. In addition, many thanks to Professor Robert Calderbank, Professor Vincent Poor, and Professor Ashutosh Sabharwal for reading this thesis carefully and providing constructive comments.

Professor Stuart Schwartz has been a great teacher, and I am fortunate to work as an Assistant in Instruction for the course “Signal Analysis and Communication Systems” taught by him twice in a row. I am grateful to him for training me as an effective teacher and teaching me the subtleties and nuances of being an excellent instructor. Working with Dr. Paul Henry on this course while I was the Assistant in Instructor for the course the third time was also a great pleasure, and I thoroughly enjoyed working with him.

I am very grateful to the administrative staffs at Princeton University including Sarah

Braude, Beth Jarvie, Lisa Lewis, and Dorothy Coakley, who made many things very easy. They are always free for the students and are great to talk to. I am also grateful to Andrea Zorbas, the administrative secretary of the CMC at Rice University for helping me with many vouchers and travel.

I would also like to thank all my current and former colleagues (students, post-docs and visiting researchers) at Princeton University, Rice University and EPFL for their invaluable companionship, and with whom I have shared the joys and successes on one hand and the graduate school frustrations and anxieties on the other. In specific, I would like to acknowledge Lorne Applebaum (Princeton), Amir Bennatan (Princeton), Satashu Goel (Penn State), Gajanana Krishna (Stanford), Lifeng Lai (Princeton), Achaleshwar Sahai (Rice), Lalitha Sankar (Princeton), Yiyue Wu (Princeton), Jun Xiao (Colorado Boulder), Melda Yuksel (TOBB, Ankara) for working with me on the research problems and for their companionship in addition to Waheed Bajwa, Bryan Conroy, Debasis Dash, Melissa Duarte, Ankit Gupta, Deniz Gunduz, Sina Jafarpour, Aman Jain, Tung Kim, Satish Babu Korada, Tian Lan, Yury Polyanskiy, Pedero Santacruz, Dinkar Vasudevan, Guanchun Wang, Yihong Wu and many others with whom I had many helpful discussions.

Of course, the acknowledgments would not be complete without thanking my family. Without their love, support and sacrifices none of this would have been possible.

To my parents for their love, endless support and encouragement.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Use of wireless networks is everywhere these days - from disaster areas to less-troubled climes. Wireless technology has literally changed our lives. Just 20 years ago, contacting people within a country was difficult, forget about across continents. A letter would take a few days to a few weeks, a phone call was not accessible in every home, there were no cell-phones to communicate while outside the residence or office. Wireless technology now has made our lives easier but the quest for improvement has not stopped. These days we need wifi everywhere, even in air-planes. We want to access movies over wireless, make wireless video calls, read books, play games over wireless, surf the web, etc. all in a small device like a cell phone. This quest has led to the development of several hand-held wireless devices like smartphones, iPad, Amazon Kindle etc. in the last few years. The use of wireless technology is not only limited to wifi and handheld wireless devices, but can also be used in large sensor networks for instance to measure weather conditions, body area networks to send the messages from parts of the body to the hospital if needed. The use of wireless technology has increased so much that the network service providers are having a hard time to meet the demands of the customers.

Since the nodes in a wireless network are mobile, the information about the network cannot be fully obtained throughout the network. As a result, each of the nodes in the network has incomplete or imperfect information about the network. This knowledge would lead to different sets of decisions at the nodes - when, what and how to transmit? Less information or noisier information about the network leads to poor decisions in terms of the achievable throughput. Thus, understanding what can be done with a given amount of local information and how the increase in local information yields improved decisions at the nodes is a fundamental challenge.

The problem of finding decisions with less or imperfect information has been studied from various aspects. Some of these formulations include compound channels, knowledge or non-knowledge of channel state information at the nodes, quantized perfect channel state information at the nodes, etc. However, the study of local information is still very idealized in the current literature which keeps the theory distant from the actual implementations. Even in the case of point-to-point channels, neither the forward nor the feedback channel is perfect. The more information that can be propagated over the channels, the better is the performance but there is more overhead in passing the information. The problems increase with more nodes in the network. The users in a network are mobile, so the system cannot be designed considering a fixed user location. The network design has to support the largest region in which a user can be present since there is uncertainty of the location of the users. Further, the mobility of nodes causes incomplete channel knowledge at different points in the network. The main motivation is to understand the networks with limited information at the nodes which would lead to better system designs and bridge the gap between theory and implementation.

In this thesis, we will focus on three main issues with imperfect or incomplete information in wireless networks. These issues are explained in the next three sections.

1.2 Imperfect Channel Estimates and Imperfect Channel State Feedback

Consider the feedback point-to-point system shown in Figure 1.1, where the receiver first measures the channel and sends a feedback signal to the transmitter about the channel. The transmitter uses the channel information about the channel conditions to adapt its transmission strategy. A common feedback model involves encoding the channel knowledge at the receiver into a finite number of bits or a real number. The majority of the work to-date assumes that either the channel knowledge at the receiver is perfect and/or the feedback channel is noise-free [14, 20, 21, 25, 27, 28, 30, 33–36, 40, 45, 46, 50, 52]. The above body of work thus serves as an upper bound to the performance of a system where channel information at the transmitter and receiver is approximate and potentially mismatched.

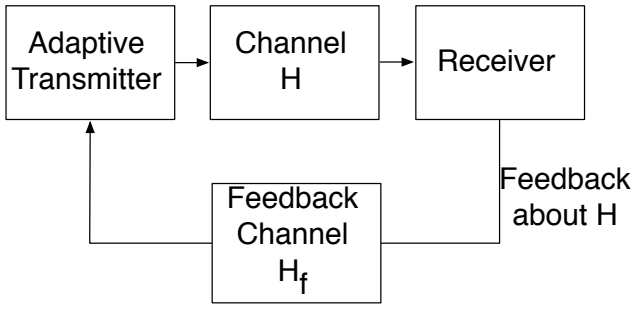

Figure 1.1: A feedback based adaptive transmission system.

Imperfect channel knowledge at the transmitter and receiver is unavoidable in wireless communications since the channel H (see Figure 1.1) is time-varying and hence has to be measured periodically by the receiver to assist in transmitter adaptation. To measure the channel, the transmitter sends a periodic finite length training signal, which is used at the receiver to form an estimate of the channel. The channel estimate is then encoded by the receiver and sent over a feedback channel which often is also a wireless link, and is thus prone to errors. In this scheme, a fundamental question is how much information *about* the channel can be extracted, which *both the transmitter and receiver can agree upon*. In short,

the transmitter and receiver have to agree upon the state of the channel by conversing over a noisy fading channel. In this thesis, we make progress towards the above question for the case of power control, with the objective to minimize the outage probability.

If the channel information is known perfectly at the transmitter and the receiver, then the transmitter can perform power control to invert the effect of a multiplicative fading channel. The inversion conserves power in good channel states by using less transmit power, while using more transmit power in poor channel conditions. If the transmitter knowledge is limited to a finite-bit approximation of the perfect receiver knowledge, then the quantized power control [33, 34, 36] is an approximation of the ideal power inversion. Much like the ideal channel inversion, the quantized power control delivers an average received signal-to-noise ratio (SNR) by exploiting the fact that poor channel states are very rare and hence large peak powers are also used rarely.

In this thesis, we consider the achievable diversity multiplexing tradeoff when the transmitter and receiver "conference" about the channel, which allows the transmitter to perform a more precise power control and achieve better performance. The conferencing assumes no genie-knowledge at any point, and can be viewed as a consensus problem over noisy channels. The use of multiple rounds is similar in spirit to the improvement in error exponents by using the Schalkwijk-Kailath-like feedback-based coding scheme [37, 43, 49] for feedback in Gaussian channels. The multi-round strategies involve zooming into the region where there is some uncertainty. In the Schalkwijk-Kailath coding scheme, the transmitter starts with a coarse version of the message and then refines it with more rounds of feedback. In our scheme, the channel is unknown and the feedback is used to resolve the channel at the transmitter and the receiver. We model the forward and the backward channels as fading channels, but the nodes are not interested in both these channel gains. Thus only partial information is needed to decide upon the feedback and the power levels. In our scheme, the receiver learns the channel state information which is a continuous random variable unlike learning a discrete message in [49]. Moreover in our scheme, the feedback channel

has noise and is quantized in the case of FDD protocols unlike [49] where the feedback is a real number received noiselessly at the transmitter. Hence the basic idea of successive refinement in the two schemes is similar, but the schemes differ significantly.

1.3 Accounting Mobility of Users to Decide Placement of Fixed Nodes

The users in a network are mobile, and hence the base station and relay station cannot be placed in a network for a fixed destination location. In these cases, the design problem at hand is to maximize *coverage* for a fixed desired transmission rate. To make progress at the design problem, we will consider relay channels, which have recently attracted significant attention as a model for *ad-hoc* networks [42]. These channels model problems where one or more *relays* help a pair of terminals communicate. The general channel model was first considered by van der Meulen [55–57] and further studied in a ground breaking work by Cover and El Gamal [26]. Although the capacity region for the channel is still unknown, the results of [26] include two achievable coding strategies which were subsequently named decode-and-forward (DF) and compress-and-forward (CF). The Gaussian relay channel was examined by Kramer *et al.* [38] and Høst-Madsen and Zhang [31].

While a destination at some locations may benefit from relocating the relay, a destination at other locations will suffer. However, in many cases of practical interest, the location of the relay is determined at a time when the destination's location is unknown. The destination is typically a mobile station, while the relay is often a fixed terminal, whose location is determined once, at the time that the network is designed.

Nevertheless, while the location of the destination may be unknown at the time of network design, the target transmission rate is typically known. Thus, the effect of changes in the relay (change of location or communication strategy) on the *coverage region* can be evaluated.

1.4 Accounting Mobility of Users to Decide Strategies with Local Information

Node mobility in wireless networks leads to constant changes in network connectivity at long time-scales and per link channel gains at short time-scales. The optimal rate allocation and associated encoding and decoding rules depend on both the network connectivity and the current channel gains of all links (commonly referred as network state). However, in large wireless networks, acquiring full network connectivity and state information for making optimal decisions is typically infeasible. Thus, in the absence of centralization of network state information, nodes have limited local view of the whole network. As a result, the local view of the nodes are mismatched and different from local views of other nodes. Thus, each node has potentially a different snapshot of the whole network. Due to mismatched local views, nodes' decisions about their transmission (like rate, power, codebook) and reception (method of decoding) parameters are inherently distributed. The key question then is how do optimal distributed decisions perform when compared to the optimal decisions which have full network state information.

We immediately acknowledge the difficulty in answering the above question. Even with full global information, where each node knows the full network connectivity and current state perfectly, the capacity of general networks is an open problem. In light of that fact, our driving question adds additional complexity to the analysis by asking nodes to rely only on their local views.

In this thesis, we limit our attention to K-user single-hop interference networks with K transmitters and K receivers. Each transmitter communicates with its receiver in a single-hop fashion but in the process can interfere with an arbitrary number of receivers. The special cases include the classic two-user interference network, Z-network, one-to-many, many-to-one and fully-connected interference networks. In this thesis, we will consider both the deterministic [18, 24] and the Gaussian models for the network.

To model the local view, we will borrow the concept of hop distance from the networking literature and consider the case where each transmitter has a perfect knowledge of all channels within h hops from it and has no knowledge of links beyond h hops. As a result, if h is less than the network diameter, a subset of transmitters will end with mismatched knowledge about the state of the channels. Since each channel gain can range from zero to a maximum value, our formulation is similar to compound channels [23, 47] with one major difference. In the multi-terminal compound network formulations, *all* nodes are missing identical information about the channels in the network. In our formulation, the hop-based model of local view leads to nodes with *asymmetric* information about the channels in the network. Thus to emphasize that the lack of knowledge is asymmetric, we have labeled the resulting compound channel capacity formulation as *local view capacity*. Finally, we assume that the nodes know the connectivity, i.e., which pairs of the links can exist but may or may not know the actual value of the channel gains on those links. In graph-theoretic parlance, the nodes are assumed to know the edges of the graph (i.e. the shape of the network) but not their weights which represent channel gains. This is partially motivated by the fact that the network connectivity often changes at a much slower time-scales than the channel gains.

Finally, realizing the difficulty of directly characterizing capacity (sum or the whole region), we propose to study the best guaranteed ratio of the sum rate with local view to the sum-capacity with full global view at each node. We label this as normalized sum-capacity, $\alpha^* \in [0, 1]$. Our goal then is to characterize the normalized sum-capacity as a function of the hops of information about the network that is available at the nodes. In many cases, it turns out that the normalized capacity is easier to characterize than the actual capacity since this involves finding sum-capacity for a smaller range of the values of channel gains.

1.5 Contributions of the Thesis

Our first objective is to understand the effect of imperfect channel estimation at the receiver and imperfect feedback, which is covered in Chapter 2. We give new conferencing protocols, which assume no genie-knowledge at any point, and can be viewed as a consensus problem over noisy channels. These protocols use novel concepts of power-controlled feedback and power-controlled training. It can be shown that the feedback sent using symbols of the same power provides very limited gains, and hence the power levels have to be optimized. Further, the training power levels are a function of the feedback symbol since all the error events have to be driven down using conferencing. The results for FDD and TDD systems can be described as follows, and will be provided in Chapter 2.

For FDD systems, suppose that the receiver quantizes the channel state and sends a feedback index consisting of K levels. It was shown in [33, 36] that the diversity increases exponentially in K for $mn > 1$ and linearly in K for $mn = 1$ (where m and n are the number of transmit and receive antennas respectively). However, we show in this thesis that if the receiver obtains the channel state information by training and the feedback from the receiver is sent over a noisy channel, feedback bits help in resolving the channel state till the noise floor becomes dominant. If the channel estimate is below the noise floor, the channel estimate does not have sufficient resolution to extract additional bits of information about the channel. This restricts the increase in diversity with feedback bits. We further show that the diversity equivalent of 1-bit perfect feedback and perfect channel state at the receiver can be achieved with 1 bit of noisy feedback with two rounds of training at the receiver. These results extend to a quantized iterative protocol where we use K rounds of training with $K-1$ rounds of feedback to achieve the same diversity as K levels of perfect feedback when the receiver knows perfect channel state information. The protocol uses K rounds of training and $K-1$ rounds of feedback to extract $\log_2(K)$ bits about the channel. The main idea in the proposed multi-round protocol is that the transmitter opportunistically sends higher training power when the previous round of training indicated that the channel

is below the noise floor. The power used by the nodes to send a training/feedback/data symbol in a channel event is governed by the probability of the channel event itself, so that a weighted average of the transmit power at the nodes meet the average power constraint. The above *on-demand use* of powers over many rounds helps increase the diversity beyond single shot estimation methods considered in all prior work. This is because the channel estimates at the nodes can be refined with increasing rounds of communication leading to larger diversities.

For TDD systems, the reciprocity in the forward and the backward channel can be utilized to achieve the gains in diversity. Due to reciprocity, if any of the nodes have perfect channel state information, the diversity gain is unbounded. However, if none of the nodes know the channel state, reciprocity allows the receiver to detect if the transmitter has made an error in understanding the previous feedback signal. Thus, the receiver can correct the transmitter's actions more rapidly compared to the case of FDD protocols, where such immediate error detection at the receiver is not possible. The proposed protocol is able to achieve better diversity than the FDD results with 1.5 rounds. More precisely, this scheme achieves a diversity of $mn(mn+1)-(m+n-1)r$ for multiplexing $0<r<\min(m,n)$. For a $m=1$ or $n=1$, this strategy achieves the same diversity-multiplexing tradeoff as in the FDD system as the number of feedback levels go to ∞. The knowledge at the nodes can be further refined by more rounds of training and feedback. In general, a $(K-1).5$ round strategy is able to achieve a diversity of $mn(1+mn+\cdots+(mn)^{K-1})-(mn)^{K-2}(m+n-1)r$ for multiplexing $0<r<\min(m,n)$. The FDD and TDD iterative schemes have the same diversity in the limit as $r\rightarrow 0$, but the TDD scheme leads to higher gains at all non-zero multiplexing gains. All the results can be easily extended to Multiple-Access channels and have not been covered in this thesis.

In Chapter 3, we will consider our second objective, which is to consider placement of nodes to allow maximum coverage. The following simple example illustrates the difference between the performance measure considered in this thesis (maximizing coverage), and the

classic measure (maximizing rate). We assume the relay is at a normalized distance of 1 from the source, the average power constraints at the source and relay are 1, and the channel fade coefficient is $\alpha = 2$. If we fix the desired communication rate at $R = 1$ bits/channel use, our analysis will assert that DF achieves ubiquitously superior performance in terms of coverage as compared to CF. Specifically, the region of potential destination locations where communication at a rate of R is possible using CF, is contained within the equivalent DF region.

In the above discussion, we fixed the target rate R and simultaneously considered all potential destinations. We now switch to the classic problem formulation, where the attention is confined to one fixed destination, and the focus is on the maximum achievable rate at this destination. Specifically, we place the destination at an equal distance of 1 from the source and the relay and leave the rest of the parameters unchanged. In this case, DF can achieve a maximum rate of 1 bit/channel use at the destination, yet CF can do better: 1.17. Thus, CF achieves superior performance in terms of rate at this particular destination. This would appear to contradict our previous discussion, which asserted the ubiquitous superiority of DF. However, in the context of that discussion, we were not concerned with the precise value of the maximum achievable rate, only whether it was greater than the desired R. Using this concept of coverage, we show that DF is better than CF as long has relay has some minimum average power.

In Chapter 4, we will consider our third objective, which is to maximize global sum-rate in networks with mismatched local views. Nodes have to base their decision only on their local asymmetric views which in turn implies that their decisions are naturally distributed. One intuitive solution is for nodes to coordinate their transmissions such that the nodes beyond h hops transmit only if they can cause no interference with h-hop size sub-network and thus each connected sub-network operates as if it is a network with full global information. This is formalized through the notion of an independent graph, which is defined as a sub-graph which admits a distributed encoding and decoding scheme which

achieves same sum-capacity as a scheme with full global information. We use this intuition to propose *maximal independent graph scheduling*, where the network is divided into sub-graphs (equivalently sub-networks) and the sub-graphs are scheduled orthogonally over time. The sub-graphs are chosen such that they are maximal independent graphs which ensure highest spatial reuse of the users.

For one hop information at the transmitters, maximal independent graphs are equivalent to maximal independent sets (MIS), which are largest subsets with non-interfering transmitter-receiver pairs. Note that maximal independent set scheduling or maximal weighted independent sets are often the optimal schedules under traditional SINR (signal to interference noise ratio) based protocol models for networks [53]. Our results show that MIS schedule is information-theoretically optimal in several cases. Hence, we provide an information-theoretic notion of optimality for MIS scheduling algorithms in those cases.

We show that in several cases, a maximal independent graph (MIG) scheduling algorithm achieves the maximum normalized sum-rate among all distributed encoding and decoding schemes, when the transmitters have no more than two hops of channel information. The MIG schedule is shown to be optimal for most three-user bipartite interference topologies, K-user cyclic chain, K-user d-to-many interference channel etc.

However, we show that the MIG schedule is not optimal in general for all network topologies and higher rates can be achieved by using a more sophisticated coding schemes. For example, in the case of one hop information at the transmitters in 3-user cyclic chain network, we show that a coded MIS (CMIS) schedule, where the coding is performed over two scheduling time-slots achieves a higher rate than pure scheduling based network. In CMIS scheduling, receivers of inactive transmitters continue listening and train themselves on the interference caused by other nodes. Then, they use this interference in a later slot to aid reliable decoding of their own codeword.

Parts of this thesis are published in [2–8, 10–16].

Chapter 2

Communication with Errors in Training and Feedback

In this chapter, we will provide multi-round iterative protocols that account for imperfection in the channel estimates at the receiver and imperfect feedback in point-to-point systems. We consider two systems- FDD where the forward and the feedback channel gains are independent, and TDD where the forward and the feedback channel gains are identical in Sections 2.2 and 2.3 respectively. We will provide on-demand power control algorithms with multiple rounds of conferencing between the terminals, assuming no genie-aided information of noiseless feedback or perfect training.

2.1 Problem Formulation and Preliminaries

2.1.1 Two-way Channel Model

Consider a multiple input multiple output (MIMO) channel with the transmitting node denoted by T and receiving node denoted by R. We will assume that there are m transmit antennas at the source node and n receive antennas at the destination node, such that the

input-output relation is given by

$$\mathsf{T} \rightarrow \mathsf{R} : Y = HX + W, \tag{2.1}$$

where the elements of H and W are assumed to be independent and identically distributed (i.i.d.) with complex normal distribution of zero mean and unit variance, $CN(0,1)$. The matrices Y, H, X and W are of dimension $n \times T_{coh}, n \times m, m \times T_{coh}$ and $n \times T_{coh}$, respectively. We assume that T_{coh} is the coherence interval so that the channel H is fixed during a fading block of T_{coh} consecutive channel uses, and statistically independent from one block to another. We further assume that T_{coh} is finite and does not scale with SNR. The transmitter is assumed to be power-limited, such that the long-term power is upper bounded, i.e, $\frac{1}{T_{coh}}\text{trace}(\mathbb{E}\left[XX^{\dagger}\right]) \leq \mathsf{SNR}$.

We consider both frequency-division and time-division duplex models for the feedback channel. In both cases, we assume that the same multiple antennas at the transmitter and receiver are available to send feedback, in a half-duplex manner. For the feedback path, the receiver will act as a transmitter and the transmitter as a receiver. As a result, the feedback source (which is the destination for data bits) will have n transmit antennas and the feedback destination (which is the source of data bits) will be assumed to have m receive antennas. Furthermore, a block fading channel model is assumed for the feedback channel,

$$\mathsf{R} \rightarrow \mathsf{T} : Y_f = H_f X_f + W_f, \tag{2.2}$$

where H_f is the MIMO fading channel for the feedback link, and W_f is the additive noise at the receiver of the feedback; both are assumed to have i.i.d. $CN(0,1)$ elements. The feedback transmissions are also assumed to be power-limited with a long-term power constraint given by $\frac{1}{T_{coh}}\text{trace}(\mathbb{E}\left[X_f X_f^{\dagger}\right]) \leq \mathsf{SNR}_f$. Without loss of generality, we will assume the case where the transmitter and receiver have *symmetric resources*, such that $\mathsf{SNR} = \mathsf{SNR}_f$. All our results can be easily generalized to the case of asymmetric resource usage.

For the case of FDD systems, H and H_f are statistically independent. On the other hand, for the case of TDD, we assume that the H and H_f are perfectly correlated within one coherence interval, and adopt a phase-symmetric two-way channel model with $H_f = H^T$ [40, 52].

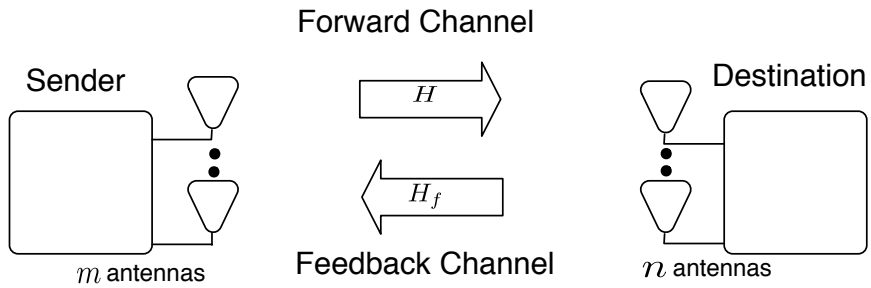

Figure 2.1: Two-way fading channel, where the forward and feedback channels use the same antennas.

2.1.2 Multi-round Protocols to Acquire Channel State

The two-way channel model allows the transmitter and receiver to conduct multiple rounds of information exchange. In this chapter, we will consider coding strategies where the transmitter-receiver perform a multi-round exchange to estimate the channel H, before sending the data in order to maximize system diversity order. A *round* is defined to consist of two transactions: a transmission sent from node T to node R and a return transmission from node R to node T. We also define *half-round*, where only one node sends a transmission without receiving a transmission in exchange.

Figure 2.2 depicts the input-output signals and the temporal dependence between them. In round i, transmission from node T is denoted as $X_i\left(X^{i-1}, Y_f^{i-1}\right)$ where $X^{i-1} = \{X_1, X_2, \ldots, X_{i-1}\}$ are all the previous inputs and $Y_f^{i-1} = \{Y_{f,1}, Y_{f,2}, \ldots, Y_{f,i-1}\}$ are all the previous outputs of the feedback channel. Analogously, we define feedback channel input $X_{f,i}\left(X_f^{i-1}, Y^i\right)$. Since both T and R are assumed to be half-duplex and the multi-round protocol is always initiated by the transmitter T, the signal $X_{f,i}$ from node R can depend on the last round of received signal Y_i.

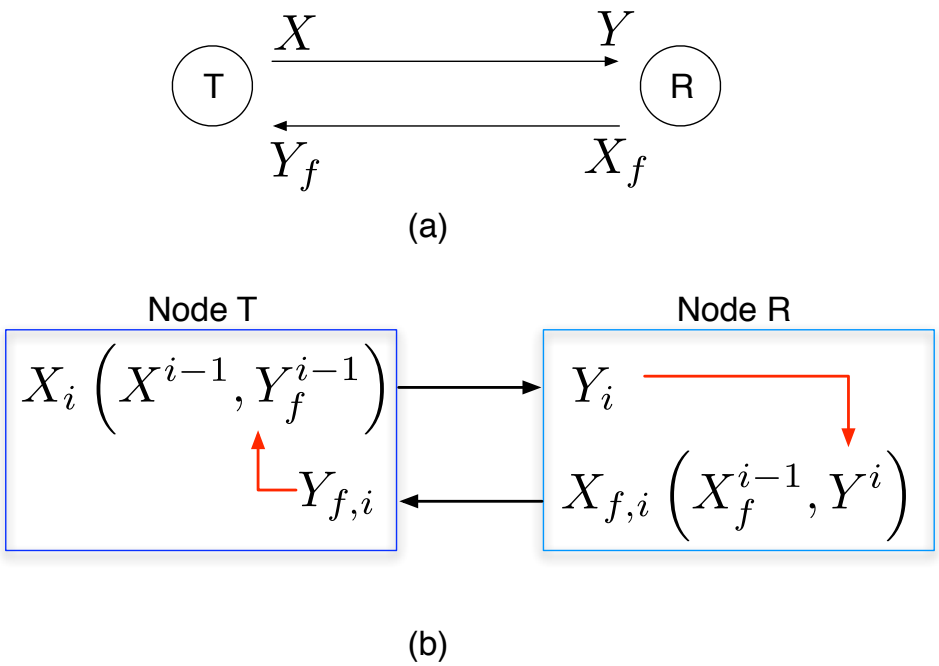

Figure 2.2: Multi-round protocols: (a) input-output relations and (b) temporal dependence of signals.

2.1.3 Diversity-Multiplexing Tradeoff

In this chapter, we will focus on the asymptotic regime of large SNR. Thus, we adopt the notation of [60] in which $\doteq$ represents exponential equality as $a \doteq b$ if $\lim_{\mathsf{SNR}\to\infty} (\log(a) - \log(b)) / \log(\mathsf{SNR}) = 0$. However, $a \doteq 0$ would mean that a decays faster than any polynomial in SNR. In this chapter, we are only concerned with probability decaying polynomially with SNR and hence the probability of events that decay faster than polynomial decay (example, exponential decay) will not enter the calculations. We similarly use $\dot{<}$, $\dot{>}$, $\dot{\leq}$, $\dot{\geq}$ to denote exponential inequalities.

We will only consider the case when the codeword X spans a single fading block. Based on the transmitter channel knowledge G, the transmitted codeword is chosen from the codebook $C_G = \{X_G(1), X_G(2), \cdots, X_G(2^{RT_{coh}})\}$, where R is the rate of the codebook. All $X_G(k)$'s are matrices of size $m \times T_{coh}$. In this chapter, we will only consider single rate transmission where the rate of the codebooks does not depend on the transmitter knowledge, and the codebooks are derived from the same base codebook C by power scal-

ing of the codewords. In other words $C_G = \sqrt{P_G}C$, where the product implies that each element of every codeword is multiplied by $\sqrt{P_G}$ where each codeword $X(k) \in C$ has unit power. Thus, P_G is the power of the transmitted codewords. Recall that there is an average transmit power constraint, such that $\mathbb{E}(P_G) \leq \mathsf{SNR}$. All coding strategies in this chapter assume Gaussian input distribution.

Since our focus is on the delay-limited regime of single codewords, we will use outage as our metric. Outage is defined as the event that the mutual information of the channel for a channel knowledge $H \in \xi$ with some distribution $P_{H|\xi}$ (for ξ a subset of all $n \times m$ matrices) at the receiver, $I_\xi(X;Y)$ is less than the desired rate R [59]. Let $\Pi(\mathcal{O})$ denote the probability of outage, where $\mathcal{O}$ is the set of all the channels where the transmitted rate R is less than the maximum supportable rate $I_\xi(X;Y)$. The system is said to have diversity order of d if $\Pi(\mathcal{O}) \doteq \mathsf{SNR}^{-d}$. Note that all the index mappings, codebooks, rates, and powers are dependent on the average signal to noise ratio, SNR. Specifically, the dependence of rate R on SNR is explicitly given by $R = r \log \mathsf{SNR}$, where r is labeled as the multiplexing gain. The diversity-multiplexing tradeoff is then described as the maximum diversity order $d(r)$ that can be achieved for a given multiplexing gain r.

If ξ is a singleton set containing H, the channel H is known at the receiver in which case $I_\xi(X;Y) = \log\det\left(I + \frac{P_G}{m}HQH^\dagger\right)$ is the mutual information of a point-to-point link with m transmit and n receive antennas, transmit signal to noise ratio P_G and input distribution Gaussian with covariance matrix Q [60]. The dependence of the index at the transmitter is made explicit by writing the transmit SNR as a function of transmitter channel knowledge G. If the receiver knows the channel estimate which is the actual channel with some additive noise, we will consider treating channel estimation error as noise to provide an achievability in this chapter as in [60].

For the case of perfect receiver information and no transmitter information, let the codebooks of rate $R \doteq r \log \mathsf{SNR}$ and power $P \doteq \mathsf{SNR}^p$ be used. The outage event is defined as $\mathcal{O}(R,P) = \{H : \left(I + \frac{P}{m}HQH^\dagger\right) < R\}$, the probability of which is $\Pi(\mathcal{O}(R,P)) \doteq$

$\mathsf{SNR}^{-G(r,p)}$ thus giving the diversity-multiplexing tradeoff of $d = G(r,p)$ [12], where

$$G(r,p) \triangleq \inf_{\alpha_1^{\min(m,n)} \in A(r,p)} \sum_{i=1}^{\min(m,n)} (2i - 1 + \max(m,n) - \min(m,n))\, \alpha_i,$$

and

$$A(r,p) \triangleq \{\alpha_1^{\min(m,n)} | \alpha_1 \geq \ldots \alpha_{\min(m,n)} \geq 0, \sum_{i=0}^{\min(m,n)} (p - \alpha_i)^+ < r\}.$$

The function $G(r,p)$ defines a piecewise linear curve connecting the points $(r, G(r,p))$= $(kp, p(m-k)(n-k))$, $k = 0, 1, \ldots, \min(m,n)$ for fixed m, n and $p > 0$. We define $G_u(r)$ recursively as follows. Let $G_0(r) = 0$, and $G_u(r) = G(r, 1 + G_{u-1}(r))$ for $u \geq 1$.

Remark 1. Consider a MIMO system with m transmit and n receive antennas with perfect channel state information at the receiver (CSIR). The diversity multiplexing tradeoff with no channel state information at the transmitter (CSIT) is $d_{\text{CSIR}} = G(r,1) = G_1(r)$ [60]. Further, the diversity multiplexing tradeoff of $d_{\text{CSIRT}_\text{q}} = G_K(r)$ can be achieved with K levels (or equivalently $b = \log_2(K)$ bits) of quantized feedback from the receiver about the channel [36].

2.1.4 Training the receiver

In this section, we will consider minimum mean square error (MMSE) channel estimation for a single user MIMO channel. The channel is estimated using a training signal that is known at the receiver. From the received signal, MMSE estimation is done as in [22] to get an estimate $\widehat{H}$ of the original channel H. Let X_T be the training signal of size $m \times N$ for some $m \leq N < T_{\text{coh}}$ that is known at the receiver and transmitter. The transmitter sends X_T and the destination receives $Y_T = HX_T + W_T$ where Y_T is a $n \times N$ received signal and W_T is the additive Gaussian noise with each entry from $CN(0,1)$. Following [22], the

optimal training signal is

$$X_T = \left[\begin{array}{cc} \sqrt{(n\mu_0 - 1)^+} I_m & 0_{m\times(N-m)} \end{array} \right],$$

where μ_0 is tuned to satisfy the power constraint, I_m denotes the $m \times m$ identity matrix and $0_{m\times(N-m)}$ represents the $m \times (N-m)$ matrix having all entries 0. Hence, $n\mu_0 - 1 = \frac{N\mathsf{SNR}}{m}$. Further, the channel estimate is given by

$$\widehat{H} = Y_T(X_T^\dagger X_T + I_N)^{-1} X_T^\dagger = Y_T \left[\begin{array}{c} \frac{\sqrt{\frac{N\mathsf{SNR}}{m}}}{1+\frac{N\mathsf{SNR}}{m}} I_m \\ 0_{N-m\times m} \end{array} \right],$$

which can be rewritten as

$$\widehat{H} = H \frac{1}{1 + \frac{m}{N\mathsf{SNR}}} + W_2 \frac{\sqrt{\frac{N\mathsf{SNR}}{m}}}{1 + \frac{N\mathsf{SNR}}{m}} \tag{2.3}$$

where W_2 is the left $n \times m$ submatrix of W_T. We now note some properties of the MMSE estimate. First, it is easy to see that the expected value of $(H - \widehat{H})\widehat{H}$ is zero confirming the orthogonality of the error with the unbiased estimate. Further, the variance of any entry in $(H - \widehat{H})$ is $\frac{1}{1+\frac{N\mathsf{SNR}}{m}}$. Thus, H and $\widehat{H}$ are matrices in which each corresponding element is highly correlated, with the correlation coefficient between corresponding elements of H and $\widehat{H}$ being $\rho = \frac{1}{\sqrt{1+\frac{m}{N\mathsf{SNR}}}}$. Further note that any $N \geq m$ channel uses are equivalent for analyzing the asymptotic performance.

In general, if G and H are correlated with correlation coefficient ρ, the joint probability distribution function of the eigenvalues of $HH^\dagger$ and $GG^\dagger$ is given by the following result:

Lemma 1. *[58] Consider two $n \times m$ random matrices $H = (h_{ij})$ and $G = (g_{ij})$, $i \in [1, n]$, $j \in [1, m]$, each with i.i.d complex zero-mean unit-variance Gaussian entries, i.e., $E[h_{ij}] = E[g_{ij}] = 0, \forall i, j, E[h_{ij}h_{pq}^\dagger] = E[g_{ij}g_{pq}^\dagger] = \delta_{ip}\delta_{jq}$, where the Kronecker symbol δ_{ij} is 1 or 0 when $i = j$ or $i \neq j$ respectively. Moreover, the correlation among the two*

random matrices is given by $E[h_{ij}g_{pq}^{\dagger}] = \rho\delta_{ip}\delta_{jq}, \forall i,j,p,q$, *where* $\rho = |\rho|e^{j\theta}$ *is a complex number with* $|\rho| < 1$. *Let* $n \leq m$ *and* $\nu = m-n$. *The joint probability distribution function of the unordered eigenvalues of* $HH^{\dagger}$ *and* $GG^{\dagger}$ *is*

$$\begin{aligned} p(\lambda, \widehat{\lambda}) &= \frac{\exp\left(-\frac{\sum_{k=1}^{n}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)\triangle(\lambda)\triangle(\widehat{\lambda})}{n!n!\Pi_{j=0}^{n-1}j!(j+\nu)!|\rho|^{mn-n}(1-|\rho|^2)^n}\Pi_{k=1}^{n}\left(\sqrt{\lambda_k\widehat{\lambda}_k}\right)^{\nu} \\ &\quad \det\left|I_{\nu}\left(\frac{2|\rho|\sqrt{\lambda_k\widehat{\lambda}_l}}{1-|\rho|^2}\right)\right|, \end{aligned} \tag{2.4}$$

where $\triangle(.)$ *represents the* $n-$*dimensional Vandermonde determinant,* $I_k(.)$ *denotes the* k^{th} *order modified Bessel function of the first kind, and the eigenvalues of* $HH^{\dagger}$ *and* $GG^{\dagger}$ *are given by* $\lambda = (\lambda_1, \cdots, \lambda_n)$ *and* $\widehat{\lambda} = (\widehat{\lambda}_1, \cdots, \widehat{\lambda}_n)$ *respectively.*

Note that although Lemma 1 assumed $n \leq m$, it can be extended to the other case of $m > n$ since the nonzero eigenvalues of $HH^{\dagger}$ and $H^{\dagger}H$ are the same. Hence for all n and m, let $m_N = \min(m,n)$ and $\nu = |m-n|$. Then, the joint probability density function of the unordered eigenvalues of $HH^{\dagger}$ and $GG^{\dagger}$ is

$$\begin{aligned} p(\lambda, \widehat{\lambda}) &= \frac{\exp\left(-\frac{\sum_{k=1}^{m_N}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)\triangle(\lambda)\triangle(\widehat{\lambda})}{m_N!m_N!\Pi_{j=0}^{n-1}j!(j+\nu)!|\rho|^{mn-m_N}} \\ &\quad \frac{\Pi_{k=1}^{m_N}\left(\sqrt{\lambda_k\widehat{\lambda}_k}\right)^{\nu}\det\left|I_{\nu}\left(\frac{2|\rho|\sqrt{\lambda_k\widehat{\lambda}_l}}{1-|\rho|^2}\right)\right|}{(1-|\rho|^2)^{m_N}}. \end{aligned} \tag{2.5}$$

Let the eigenvalues of $HH^{\dagger}$ be $(\lambda_1, \cdots, \lambda_{m_N})$, $\lambda_i \doteq \mathsf{SNR}^{-\alpha_i}$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{m_N})$. Similarly, let the eigenvalues of $\widehat{H}\widehat{H}^{\dagger}$ be $(\widehat{\lambda}_1, \cdots, \widehat{\lambda}_{m_N})$, $\widehat{\lambda}_i \doteq \mathsf{SNR}^{-\widehat{\alpha}_i}$ and $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \cdots, \widehat{\alpha}_{m_N})$. We will now find the joint distribution of α_i's and $\widehat{\alpha}_i$'s. Let $\alpha_1 \geq \alpha_2 \geq \cdots \alpha_{m_N}$ and $\widehat{\alpha}_1 \geq \widehat{\alpha}_2 \geq \cdots \widehat{\alpha}_{m_N}$. Further, we define

$$\begin{aligned} E_k &= \{(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) : \min(\alpha_i, \widehat{\alpha}_i) \geq p \; \forall i = 1, \cdots, k, \text{ and} \\ &\quad 0 \leq \alpha_i = \widehat{\alpha}_i < p \; \forall i > k\} \text{ for all } 0 \leq k \leq m_N. \end{aligned} \tag{2.6}$$

Theorem 1. *Let H be the channel and $\widehat{H}$ be the estimated channel when the training is done with power SNR^p for some $p > 0$. In the limit of high SNR, the probability density function of the SNR exponents of the eigenvalues of $HH^\dagger$ and $\widehat{H}\widehat{H}^\dagger$ is given by*

$$p(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \doteq \sum_{k=0}^{m_N} e_k \mathbf{1}_{E_k} \tag{2.7}$$

where $\alpha_1 \geq \alpha_2 \geq \cdots \alpha_{m_N}$, $\widehat{\alpha}_1 \geq \widehat{\alpha}_2 \geq \cdots \widehat{\alpha}_{m_N}$, and

$$e_k = \mathsf{SNR}^{kp(|n-m|+k)} \Pi_{i=1}^{k} \mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_i} \Pi_{i=1}^{m_N} \mathsf{SNR}^{-(2i-1+|n-m|)\alpha_i}. \tag{2.8}$$

Proof. Since all the results were symmetric about interchanging m and n, without loss of generality, we take $m \leq n$ for the purpose of this proof and the rest of the chapter. The proof for $p = 1$ is provided in Appendix 2.A. The proof for general $p > 0$ is a simple extension. □

Remark 2. *Since the receiver is trained at SNR^p, we note that for all $\alpha_i < p$, $\alpha_i = \widehat{\alpha}_i$ with probability 1, which means that the channel estimate is a reliable proxy for the actual channel. On the other hand, if $\alpha_i \geq p$, all we can state is that $\widehat{\alpha}_i \geq p$ with probability 1. For example, if $\widehat{\alpha}_i \geq 100p$, all we can reliably say about α_i is that it is $\geq p$. In the single input single output (SISO) case, the above property of $\widehat{\alpha}_i$ implies that the channel cannot be resolved below the noise floor since the noise dominates the training signal. The interesting implication in MIMO is that this result of noise dominance holds for all the eigen-values. None of the eigen-values of the channel can be resolved beyond $\alpha_i \geq p$ if $\widehat{\alpha}_i \geq p$.*

A novel scheme called power-controlled training was suggested in [52] which takes care of the estimation error in training by estimating the power-controlled channel. If the actual channel is H and the transmit power is SNR^{p_j}. Suppose that the receiver do not know p_j. The receiver makes an estimate of $\sqrt{\mathsf{SNR}^{p_j}} H$. With this estimate, the estimation error given by trace $\left(\mathsf{SNR}^{p_j} \tilde{H}\tilde{H}^\dagger\right)$ will be at-most of the order of constant. This is because the channel

estimation error in H, $\tilde{H}$ have complex normal entries with 0 mean and variance SNR^{p_j}. Thus, let eigen-values of $\mathsf{SNR}^{p_j}\tilde{H}\tilde{H}^{\dagger}$ be $\tilde{\lambda}_i \doteq \mathsf{SNR}^{-\tilde{\alpha}_i}$ where $\tilde{\alpha}_i$ have $\doteq 0$ probability if $\alpha_i < 0$ as given before. Then, trace$(\mathsf{SNR}^{p_j}\tilde{H}\tilde{H}^{\dagger}) \doteq \mathsf{SNR}^{-\tilde{\alpha}_m} \dot{\leq} \mathsf{SNR}^0$, Thus, the estimation error is at the noise floor. Theorem 1 gives the relation between eigen-values of trained estimate and the actual channel which will guide us in giving new protocols and to analyze the performance of these protocols.

2.1.5 Power Controlled Feedback

We will use the concept of power-controlled feedback throughout the chapter which is illustrated for a case of distinguishing between two feedback levels in this section. Suppose feedback levels a and b are transmitted from the receiver using power levels of SNR^{p_1} and SNR^{p_2} respectively. Without loss of generality, assume that $p_1 < p_2$. The transmitter will observe the received power and determine whether the receiver sent a or b. If the receiver transmitted a symbol at SNR^{p_i}, the received power is $\doteq \mathsf{SNR}^{(p_i-\alpha_m)^+}$ where α_m is the smallest eigen-value exponent of the channel as defined in the last subsection. We will use the threshold detection at the transmitter. We can prove that a threshold of $\mathsf{SNR}^{p_1+\delta}$ is optimal for arbitrarily small $\delta > 0$. To see this, we consider the error events at the transmitter:

1. $\Pi(\text{Transmitter received } a|\text{Receiver transmitted } b) \doteq \Pi(p_2 - \alpha_m < p_1 + \delta)$
   $\doteq \mathsf{SNR}^{-mn(p_2-p_1-\delta)}$.

2. $\Pi(\text{Transmitter received } b|\text{Receiver transmitted } a) \doteq \Pi(p_1 - \alpha_m > p_1 + \delta) \doteq 0$.

Since the second event happens with probability $\doteq 0$, δ is chosen arbitrarily close to 0 to get the optimal detection at the transmitter.

2.2 FDD Iterative Protocol

In this section, we will show how multiple rounds can be used by the transmitter and receiver to iteratively refine the knowledge about the channel H for *all* multiplexing gains. In each round, the receiver sends a binary signal back to the transmitter, providing an additional level of channel information. To perform the iterative quantization, both transmitter and receiver occasionally and opportunistically use large transmit power in some blocks. However, the use of large power is performed very rarely, allowing both nodes to stay within their prescribed average power constraints over the long-term.

2.2.1 Iterative Quantization: $\log_2(K)$ bits in $(K-1).5$ Rounds

In this subsection, we will describe the protocol and give its diversity-multiplexing tradeoff performance.

The protocol can be described as follows.

1. **Round 1 (forward)**: T sends training signal at power SNR, $X_1 = \sqrt{\mathsf{SNR}}\,\beta$, where β is a known fixed sequence. The signal received at R is $Y_1 = HX_1 + W$, which is used to form an MMSE estimate $\widehat{H}_1$. The feedback from R is a binary quantization, $q_1(\widehat{H}_1)$ such that

$$q_1(\widehat{H}_1) = \begin{cases} 0, & \text{if } \widehat{H}_1 \in \widehat{\mathcal{O}}_1^c \\ 1, & \text{otherwise} \end{cases}, \tag{2.9}$$

where $\widehat{\mathcal{O}}_1 = \{\widehat{H}_1 : \det(I + \widehat{H}_1\widehat{H}_1^\dagger \mathsf{SNR}) < \mathsf{SNR}^{r+\epsilon}\}$.

Round 1 (reverse): The quantized channel $q_1(\widehat{H}_1)$ is modulated as follows:

$$X_{f,1}\left(Y_1\right) = \begin{cases} \sqrt{\mathsf{SNR}}\,\beta_f, & q_1(\widehat{H}_1) = 0 \\ 0\,.\beta_f, & q_1(\widehat{H}_1) = 1 \end{cases}, \tag{2.10}$$

where β_f is a known sequence. The signal received at node T is $Y_{f,1} = H_f X_{f,1} + W_f$. The transmitter T estimates q_1 as $\widehat{q}_1$ using a MAP detector for power level with a threshold of $\mathsf{SNR}^{\epsilon/mn}$.

At the end of round $i-1$, the transmitter forms an estimate $\widehat{q}_{i-1}$ of the quantization performed at the receiver, which is denoted by the function $q_{i-1}(\cdot)$. To state the following recursion, we assume $\widehat{q}_0 = 1$. The estimate $\widehat{q}_{i-1}$ is used in the ith round by the transmitter as described below.

2. **Round** $i \in \{2, \ldots, K-1\}$ **(forward)**: The transmitter trains the receiver with $X_i = \sqrt{P_i}\beta$ where the power level P_i is chosen as follows

$$P_i = \begin{cases} 0, & \text{if } \widehat{q}_{i-1} = 0 \\ \mathsf{SNR}^{1+G_{i-1}(r+\epsilon)}, & \text{otherwise} \end{cases}. \tag{2.11}$$

The receiver's actions can be described as follows.

- If any of $q_u(\widehat{H}_u, q^{u-1}) = 0$ for $u \leq i-1$, then the receiver performs a maximum a-posteriori (MAP) power estimation to check if the transmitter sent a training at power level of 0 or $\mathsf{SNR}^{1+G_{i-1}(r+\epsilon)}$. The associated MAP threshold is SNR^{ϵ}. If power estimate $\geq \mathsf{SNR}^{\epsilon}$, $q_i(\widehat{H}_i, q^{i-1}) = 0$, else $q_i(\widehat{H}_i, q^{i-1}) = 1$.
- If $q_u(\widehat{H}_u, q^{u-1}) = 1$ for all $u \leq i-1$, the receiver estimates the channel as $\widehat{H}_i$ assuming that the training power is $\mathsf{SNR}^{1+G_{i-1}(r+\epsilon)}$ and bases the feedback signal based on the estimate as follows

$$q_i(\widehat{H}_i, q^{i-1}) = \begin{cases} 0, & \text{if } \widehat{H}_i \in \widehat{\mathcal{O}}_i^c \\ 1, & \text{otherwise} \end{cases}, \tag{2.12}$$

where $\widehat{\mathcal{O}}_i = \{\widehat{H}_i : \det(I + \widehat{H}_i \widehat{H}_i^{\dagger} \mathsf{SNR}^{1+G_u(r+\epsilon)}) < \mathsf{SNR}^{r+\epsilon}\}$.

3. **Round** $i \in \{2, \ldots, K-1\}$ **(reverse)**: The receiver sends the signal

$$X_{f,i} = \begin{cases} \sqrt{\mathsf{SNR}^{1+G_{i-1}(r+\epsilon)}}\,\beta_f, & q_i(\widehat{H}_i, q^{i-1}) = 0 \\ 0\,.\beta_f, & \text{otherwise} \end{cases}, \tag{2.13}$$

where β_f is a known sequence. The signal received at node T is $Y_{f,i} = H_f X_{f,i} + W_f$. If $\widehat{q}_{i-1} = 0$, then $\widehat{q}_i = 0$. That is, if the transmitter had concluded that it was a "good" channel after round $i-1$, then it does not attempt to re-estimate the feedback signal. Otherwise, the transmitter decodes the power level by a MAP detection with a threshold of $\mathsf{SNR}^{\epsilon/mn}$ to get an estimate of q_i to get $\widehat{q}_i$.

4. **Round** K **(forward only)**: The transmitter then trains the receiver and then sends the data both at a power level of $\mathsf{SNR}^{1+G_u(r+\epsilon)}$ where $u = \min\{\{v \in [1, K-1] : \widehat{q}_v = 0\} \bigcup K\} - 1$.

Theorem 2. *The iterative quantization protocol, described by the above Steps 1-5, achieves a diversity order of* $d_{(K-1).5,1} = G_K(r)$ *as* $\epsilon \to 0$.

Proof. The proof is provided in Appendix 2.B. □

Note that this diversity is the same as that with K levels of perfect feedback when the receiver knows perfect channel state information as given in [36]. Thus $(K-1).5$ rounds allow the transmitter and receiver to *agree* upon $\log_2(K)$ bits describing the channel. Note that the bits describing the channel H relate to the outage regions, and hence the Voronoi regions are concentric spheres, all of which are centered at the origin as explained in the next sub-section.

2.2.2 Discussion of Iterative Quantization

Although, the diversity of $G_K(r)$ can be obtained with perfect channel state at the receiver and perfect quantized feedback of K levels, achieving this gain when both the channel es-

timates and the feedback are noisy is a challenge. Since the receiver has to use training symbols to estimate the channel, it cannot resolve the channel gains that are smaller than the estimation error. Recall that power control gain results from using large instantaneous power in poor channel conditions. Since such poor channel states are rare, the transmitter has to use large power rarely and hence can still meet the average power constraint. However, it is precisely the identification of the poor channel states which is not possible due to error in channel estimation at the receiver.

Ideally, one should use a large power, say $\mathsf{SNR}^p, p > 1$, for the training signal, which is the sequence β in the protocol. The large training power means that the error floor is reduced to SNR^{-p} and we can identify channels gains of the order of SNR^{-p} with high reliability. However, the use of $\mathsf{SNR}^p, p > 1$ in every frame will violate the average power constraint of SNR.

The solution is to use large training power *only* when the channel gain is too small to be resolved using lesser training power. Thus, the probability of using large training power should be governed by the probability of a channel event, and thus the weighted average of transmitter power will meet the average power constraint of SNR. The above *on-demand use* of training power is precisely the main idea behind iterative quantization described in Section 2.2.1.

The other key ingredient is power-controlled feedback, which skews the feedback errors in one direction. The receiver in each iteration communicates to the transmitter whether the current power level will be enough to avoid outage. However some feedback errors are worse than others. Consider the case of binary feedback. If the channel is Good ($\log\det(I + HH^\dagger\mathsf{SNR}^p) \geq \tau$ for appropriate p and τ) but the transmitter decodes it as a Bad channel and uses larger power than needed, no outage occurs during data transmission. So confusing Good channels for Bad channels is a non-issue, especially due to the additional round of power-controlled training. However, if a Bad channel is confused as a Good channel, then the transmitter will use lower power than needed and hence as a result, there will be a

definite outage. Thus, confusing Bad channels with Good channels is the main contributor of outage due to feedback errors.

Thus the desired situation will be that $\mathsf{Prob}(\mathsf{Bad} \rightarrow \mathsf{Good})$ is as small as possible while ensuring that the probabilities of the events $\{\mathsf{Good}, \mathsf{Bad}\}$ as seen by the transmitter is the same as those seen by the receiver; this last constraint on event probabilities at the transmitter ensures that the average power consumed is SNR. All of the power-controlled feedback designs proposed in previous sections skew the feedback power allocation to lower $\mathsf{Prob}(\mathsf{Bad} \rightarrow \mathsf{Good})$ at the expense of limiting the maximum power. This is because the probability of the Good event is asymptotically much higher than the probability of the Bad event and thus if higher power is used for the Good event, the maximum power would be limited by $\mathsf{SNR}/\Pi(\mathsf{Good})$. The Bad event is encoded with a codeword of zero power by the receiver on the feedback link, which is decoded by the transmitter using an energy-based estimator. The probability of confusing a zero power codeword with any codeword of power greater than $\mathsf{SNR}^{\delta}, \delta > 0$, i.e. $\mathsf{Prob}(\mathsf{Bad} \rightarrow \mathsf{Good})$, decays exponentially fast. In comparison, the probability of confusing a Good channel with a Bad due to decoding an SNR codeword as a zero power codeword decays polynomially fast.

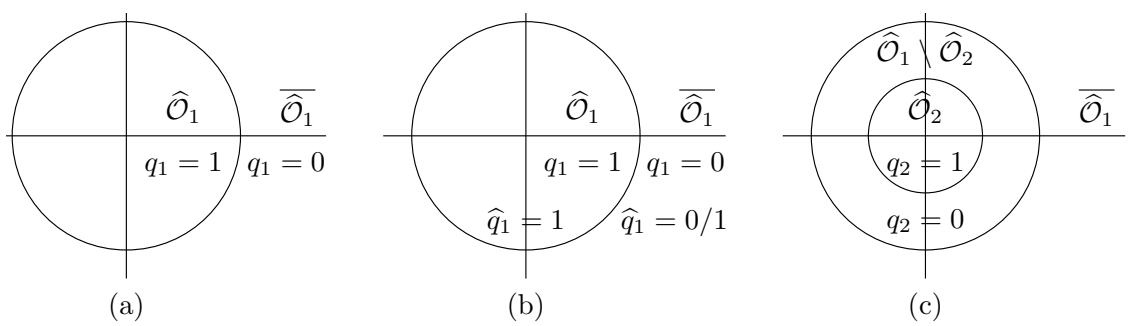

Figure 2.3: Channel events (a) in the first round (forward), (b) in the first round (reverse) and (c) in the second round(forward).

The iterative quantization can be more easily understood by a pictorial view of the decision process for $m = n = 1$. In 2.5 rounds with 1-bit per feedback round, we can resolve channels into one of 3 regions depicted in Figure 2.3(c). In the first round (forward), the receiver can distinguish between the events $\widehat{\mathcal{O}}_1$ and $\overline{\widehat{\mathcal{O}}_1}$. This is communicated to the

transmitter using a power-controlled feedback, where $q_1 = 1$ representing event $\overline{\widehat{\mathcal{O}}_1}$ is encoded with zero power codeword. As a result, if $q_1 = 1$, then $\widehat{q}_1$ will be 1. Now, consider the case when the actual event is $\widehat{\mathcal{O}}_1$. In this case, in the second round (forward), the transmitter sends a training at power SNR^2 which allows the receiver to distinguish between $\widehat{\mathcal{O}}_1 \setminus \widehat{\mathcal{O}}_2$ and $\widehat{\mathcal{O}}_2$. Once again, if $q_2 = 1$, a zero-power codeword is used by the receiver, which implies that the transmitter concludes $\widehat{q}_2 = 1$ with exponentially small error probability. Hence, the transmitter knows $\widehat{q}_1 = 1$ inside $\widehat{\mathcal{O}}_1$ and $\widehat{q}_2 = 1$ inside $\widehat{\mathcal{O}}_2$ (with probability 1). Thus, we find that the transmitter has received an index that is the same or bigger than that transmitted by the receiver in all the scenarios. This indicates that there will be no error because of the transmitter receiving the index by feedback incorrectly since the transmitter will send at at-least the power that the receiver suggested (with probability 1). Since the feedback errors do not dominate and in each feedback round the receiver adds one level precision, the diversity corresponding to K levels of perfect feedback is attained with $(K-1).5$ rounds of iterative quantization protocol.

The key to iterative quantization is that the channel is resolved to a finer level only when needed. As and when the channel is below the channel estimation error floor, the transmitter sends a higher power training compared to previous rounds. The on-demand nature of the protocol ensures neither the transmitter, nor the receiver exceed their power budget and use high powers only in rare cases.

2.3 TDD Protocols

In the FDD systems, the forward and feedback channels are independent. Thus, the feedback channel fading and errors had to be dealt with independently of the forward channel, or in other words, information about one channel cannot be exploited for better encoding on the other channel. In contrast, the symmetry of TDD channels can be effectively exploited by using the knowledge of one channel for transmission on the other. We will show

that this symmetry allows us to achieve higher diversity-multiplexing tradeoffs.

The study of training based TDD systems from a diversity multiplexing perspective was initiated in [52] for the case when $m = 1$ or $n = 1$, where a receiver-initiated strategy was proposed and can be cast as a 1.5 round scheme. This was further extended to a general MIMO system in [39], where it was shown that a diversity order up to $2mn$ (for zero multiplexing gain) was achievable.

In the case of FDD systems, if the receiver knows perfect channel state information and there is noisy feedback from the receiver, the diversity is limited by the noise in the feedback. Further, if the transmitter learns the channel state through a quantized noiseless link and the receiver does not know the channel state information, the diversity is limited by the noise in the forward channel and the quantization. Even if the feedback link is perfect and has infinite capacity, the training cannot be sent on the feedback channel since the training will indicate the quality of the feedback channel rather than the forward channel. Hence the best estimate that the transmitter can receive is the noisy estimate from the receiver. By contrast in TDD systems, infinite diversity can be achieved if the receiver or the transmitter knows the channel state perfectly. This is shown in Appendix 2.C.

In this section, we will propose a multi-round TDD protocol which iteratively refines the channel information at both ends, and also keeps the two nodes aligned in their knowledge about the channel estimate. Unlike the FDD iterative protocol, this protocol uses the fact that the errors in the forward and the feedback path will be correlated since both paths have same eigen-values due to the reciprocity of the channel. With this protocol, the diversity-multiplexing tradeoff is a straight line and dominates the diversity-multiplexing tradeoff for the FDD iterative quantization for all $0 < r < m$ with $(K-1).5$ rounds for any $K \geq 2$. For the special case of $K = 2$, the diversity multiplexing tradeoff of $mn(mn+1)-(m+n-1)r$ can be achieved using a power controlled strategy at the transmitter based on its own channel estimate. This power control is similar to that in [35] where the transmitter has a channel estimate with some error variance. While the model in [35]

has no feedback channel, we make use of reciprocity of forward and feedback channels.

2.3.1 Iterative Protocol with (K-1).5 Rounds

In this subsection, we will provide an iterative protocol for a TDD system considering the noise in the training and the feedback channels. This protocol achieves better diversity multiplexing tradeoff than the known results even for the special case of $K = 2$. We will now prove a Lemma that will be used to analyze our proposed iterative protocol.

Lemma 2. *Suppose that the transmitter is trained with a power of* SNR^p *for* $p \geq 1$*, so that the transmitter estimates the channel* $\widehat{H}_1$ *with the eigenvalues of* $\widehat{H}_1\widehat{H}_1^\dagger$ *being* $\doteq \mathsf{SNR}^{-\widehat{\alpha}_i}$*. Further, suppose that the transmit power for the training signal from the transmitter is* $P(\widehat{H}_1) \doteq \frac{\mathsf{SNR}^{1-\epsilon/mn}}{\Pi_{i=1}^{mN}\mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_i}}$*. Then, the probability that the channel cannot support the rate of* R *on this power controlled channel is asymptotically equivalent to*

$$\Pi(\log\det(I+P(\widehat{H}_1)HH^\dagger) < R) \doteq \mathsf{SNR}^{-mnp-G(r,1+(mn-1)p)} = \mathsf{SNR}^{-mn(1+pmn)+(m+n-1)r},$$

as $\epsilon \to 0$.

Proof. This case results in the transmitter knowing noisy CSIT, and hence the result is similar in nature to that in [35]. Although the result is similar in nature, we provide the proof in Appendix 2.D for completion. □

Remark 3. We note from the proof of Lemma 2 that the outage event (the event that $\log\det(I + P(\widehat{H}_1)HH^\dagger) < R$) has exponentially small probability given $\alpha_m < p$. So, the outage happens when all the $\alpha_i \geq p$, or in other words when all the eigen-values of $HH^\dagger$ are in the bad state. Further, in this bad state the channel estimation does not work well in the sense that given that $\widehat{\alpha}_i \geq p$, all one can state about α_i is that $\alpha_i \geq p$ with probability one. Raising the power exponent p for the feedback reduces the probability of the bad state ($\widehat{\alpha}_m \geq p$) thus increasing diversity.

We now describe the iterative, multi-round TDD protocol. The basic idea of the protocol lies in Lemma 2. If a higher power is allowed on the feedback or training or both, higher diversity can be obtained. However, if higher power is used for all transmissions, then the nodes will violate their power constraint. Hence, we iteratively refine the information about the channel and use higher power for the feedback on-demand. Since the large powers are used rarely, both the nodes stay within their prescribed long-term power constraints.

Let $W_1(r) = 0$, $W_2(r) = mn + G(r, mn)$, $W_k(r) = mn(1 + W_{k-1}(r)) - \epsilon$ for $k > 2$.

1. **Round 1(forward):** The transmitter remains silent.

   Round 1(reverse): The receiver sends the training signal using power SNR, $X_{f,1} = \sqrt{\mathsf{SNR}}\,\beta$, where β is a known fixed sequence. The signal received at T is $Y_{f,1} = H_f X_{f,1} + W$, which is used to form an MMSE estimate $\widehat{H}_1$. Let the eigenvalues of $\widehat{H}_1\widehat{H}_1^\dagger$ be $\widehat{\lambda}_{1,i} \doteq \mathsf{SNR}^{-\widehat{\alpha}_{1,i}}$. Further, the transmitter will save a quantization index of the channel which is an estimate based on the received power and is labeled $\widehat{q}_{u-1}(Y_{f,i})$ at the end of round $u-1$. To state the recursion, we assume $\widehat{q}_1 = 1$.

2. **Round** $u \in \{2, K-1\}$**(forward):** The transmitter trains the receiver with $X_u = \sqrt{P_{u-1}(\widehat{H}_{u-1})}\beta$ where the power level $P(\widehat{H}_{u-1})$ is chosen as follows

$$P_{u-1}(\widehat{H}_{u-1}) = \begin{cases} 0, & \text{if } \widehat{q}_{u-1} = 0 \\ \frac{\mathsf{SNR}}{\Pi_{i=1}^{mN}\mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_{u-1,i}}}, & \text{otherwise.} \end{cases} \tag{2.14}$$

   Note that for power constraint, SNR in the numerator can be replaced by $\mathsf{SNR}^{1-\delta/mn}$ for δ very small and that will not affect the analysis as was seen in the proof of Lemma 2.

   If $u = 2$, the receiver estimates the power controlled channel $G_2 = \widehat{\sqrt{P_1(\widehat{H}_1)}H}$ and

decides an index $q_2(G_2)$ as

$$q_2(G_2) = \begin{cases} 0, & \text{if } G_2 \in \mathcal{O}_2^c \\ 1, & \text{otherwise,} \end{cases} \tag{2.15}$$

where $\mathcal{O}_2 = \{G_2 : \det(I + G_2 G_2^\dagger) < \mathsf{SNR}^{r+\epsilon}\}$.

For $u > 2$, the receiver's actions can be described as follows.

- If $q_{u-1} = 0$, $q_u(q^{u-1}, G_u) = 0$.
- If $q_{u-1} = 1$, a MAP power estimation is done by the receiver. If the power $\dot{<}\mathsf{SNR}^{\epsilon/2mn}$, $q_u = 1$. Otherwise, the receiver estimates the power controlled channel $G_u = \widehat{\sqrt{P(\widehat{H}_{u-1})}}H$, and

$$q_u(q^{u-1}, G_u) = \begin{cases} 0, & \text{if } G_u \in \mathcal{O}_u^c \\ 1, & \text{otherwise,} \end{cases} \tag{2.16}$$

  where $\mathcal{O}_u = \{G_u : \det(I + G_u G_u^\dagger) < \mathsf{SNR}^{r+\epsilon}\}$.

Round $u \in \{2, K-1\}$ **(reverse):** The receiver sends the signal

$$X_{f,u} = \begin{cases} 0\ .\beta_f, & q_u = 0 \\ \sqrt{\mathsf{SNR}^{1+W_u(r+\epsilon)}}\ \beta_f, & q_u = 1 \end{cases}, \tag{2.17}$$

where β_f is a known sequence. The signal received at node T is $Y_{f,u} = H_f X_{f,u} + W_f$.

The transmitter estimates the index sent by the receiver using estimate of the received power with a threshold of $\mathsf{SNR}^{\epsilon/mn}$. If the received power $\dot{>}\mathsf{SNR}^{\epsilon/mn}$, then $\widehat{q}_u = 1$ else $\widehat{q}_u = 0$. If $\widehat{q}_u = 1$, the transmitter estimates the channel as $\widehat{H}_{u-1}$. Let the eigenvalues of $\widehat{H}_{u-1}\widehat{H}_{u-1}^\dagger$ be $\widehat{\lambda}_{u-1,i} \doteq \mathsf{SNR}^{-\widehat{\alpha}_{u-1,i}}$.

3. **Round K (forward only):** The transmitter trains the receiver and then sends the data both at a power level of

$$P_K(\widehat{H}_1, \cdots, \widehat{H}_{K-1}) = \mathsf{SNR}^{1+\max\left(\sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_{1,i},\Gamma\right)}, \tag{2.18}$$

where $\Gamma = \max_{u\in[2,K-1]:\widehat{q}_u=1} \sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_{u,i}$.

Theorem 3. *If the iterative TDD protocol is used for $K \geq 2$, following diversity-multiplexing tradeoff can be achieved as $\epsilon \to 0$*

$$d_{(K-1).5} = \begin{cases} mn\frac{(mn)^K-1}{mn-1} - (mn)^{K-2}(m+n-1)r & mn > 1 \\ K - r & mn = 1 \end{cases}.$$

Proof. We will show in Appendix 2.E that diversity of $mn(1 + W_{K-1}(r + \epsilon))$ can be obtained, which for $\epsilon \to 0$ converges to the statement of the Theorem. □

Remark 4. As a special case of the theorem, for $0 < r < \min(m, n)$, diversity of $mn + G(r, mn) = mn(mn + 1) - (m + n - 1)r$ can be obtained in the TDD system with the 1.5 rounds of training and feedback. In the FDD system, the transmitter learns a constant number of bits about the channel estimate of the forward channel (channel from the transmitter to the receiver) in the first round. However in the TDD case, the transmitter knows the forward channel estimate (due to reciprocity in the two channels) with larger resolution. Hence, even though the transmitter remains silent in the first round, it receives more information in the first round. From the point of view of the receiver, only the power-controlled channel estimate in the second round is used for decoding the data since it has at least the same information about the channel (asymptotically) as the first round of training and hence the first training is neglected for decoding. Hence, the TDD protocol performs better than its FDD counterpart. However, the TDD strategy results in the same diversity multiplexing tradeoff performance for a SIMO/MISO system as the FDD strategy when the

number of feedback levels $K \to \infty$.

In the iterative TDD protocol, we used a training symbol from the transmitter in each round. However, we could have used one bit of data from the transmitter in each round (since the protocol is receiver initiated, this is like the quantized feedback from the transmitter.). One bit of quantized feedback can be used to achieve a diversity of $G_K(r)$. Although $G_K(r)$ is smaller than what we achieve with this iterative TDD protocol, it takes less time to send one bit of data (one channel use) than to train (m channel uses). Thus when accounting for the resources used in training and feedback, an optimization over the use of a quantized feedback symbol or the training symbol in each round of communication would also be required.

2.3.2 Discussion of TDD Protocols

As in the iterative protocol for FDD systems, we may extend the TDD protocol to an iterative protocol where higher powers are used rarely, leading to arbitrarily large gains in diversity with increase in communication rounds. The TDD protocols use reciprocity in addition to the asymmetric properties of the power control. Due to the channel symmetry of the forward and the backward channels, the receiver will be able to find whether the transmitter made an error in understanding the power level. Thus, the channel symmetry can be used to increase the diversity order beyond that achievable using FDD protocols.

Channel reciprocity enables the receiver to detect whether the transmitter has made an error in understanding the feedback signal sent by the receiver. Thus, if the receiver indicated that the channel was Bad in the previous feedback phase but the transmitter understood it as Good, then the transmitter's actions will allow the receiver to detect this error with high probability. In this case, the receiver can re-send its feedback signal with higher power.

To understand the specific operation of the protocol, we will use the pictorial view depicted in Figure 2.4 for a SISO system with 3.5 rounds of the TDD protocol. After the

forward phase of second round, the receiver knows if the channel is Good (outside the circle in Figure 2.4(a)) or Bad (inside the circle), and indicates its first feedback as q_2. Since the feedback channel is the same as the forward channel, the transmitter can make an error in understanding the feedback about only the Bad channel ($q_2 = 1$). That is because when the forward channel is Good, so is the reverse channel. Thus, the feedback in this case is received with exponentially small error probability. On the other hand, when the forward channel is Bad, indicated by $q_2 = 0$, so is the feedback channel. In this case, the transmitter will make errors ($\widehat{q}_2 = 0$ when $q_2 = 1$) if the actual channel was inside the smaller circle of Figure 2.4(b), while not make errors in the shell between the two circles since the channel is not that bad ($\widehat{q}_2 = 1$ when $q_2 = 1$). In the inner most circle, since $q_2 = 1$, the receiver knows (with exponentially small error probability) that the channel is Bad but the transmitter send no more training assuming it is a Good channel since it thinks that no more channel resolution is required. The receiver can detect the erroneous training power and realize that the transmitter made an error. Detecting an error, the receiver will re-send the same information as $q_3 = 1$ but with higher power to ensure safer delivery. For clarity, we step through each phase of the protocol.

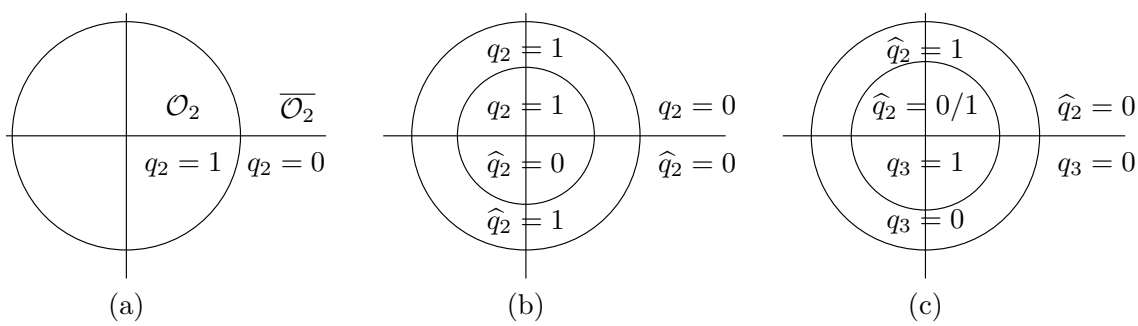

Figure 2.4: Channel events (a) in the second round (forward), (b) in the second round (reverse) and (c) in the third round (forward).

In the first round, the receiver sends training signal to the transmitter based on which the transmitter finds the estimated channel. Using this estimated channel, the transmitter trains the receiver in the forward phase of the second round. With this training, the receiver finds the regions where this power level will be sufficient to avoid outage. It sends $q_2 = 0$

when the channel is Good (the event that the channel is not in outage based on the estimate of power-controlled channel) while $q_2 = 1$ when it is Bad (not Good). Thus, the channel space is divided in two parts at the receiver. This is depicted in Figure 2.4(a). The receiver sends this index to the transmitter and due to the asymmetry in the feedback errors, the Good channel state is received as the Good state ($\widehat{q_2} = 0$) with probability 1. However, the Bad channel state ($q_2 = 1$) may be mistakenly understood as the Good channel state ($\widehat{q_2} = 0$) by the transmitter. Hence, as shown in Figure 2.4(b), the transmitter receives the index in error inside the inner circle.

In the third round, the transmitter sends a training symbol if it understood that the channel is in the Bad state ($\widehat{q_2} = 1$) which is in the shell between the two circles. Note that no training is sent in the inner circle where the transmitter received incorrect feedback. Since outside the outer circle, the receiver has resolved the channel as being Good, it continues to inform the Good channel state ($q_3 = 0$). In the interior of the inner circle where the transmitter was mistaken, the receiver will be able to know that the transmitter made an error (This is because the no training is sent and the power-controlled forward channel will be in outage, i.e. $\log(1 + GG^\dagger) < \mathsf{SNR}^{r+\epsilon}$ where G is the power-controlled channel estimate, with probability 1). This is due to the symmetry which allowed receiver to know that the transmitter has made an error and the channel is Bad. Hence, the receiver informs the transmitter that the channel is Bad ($q_3 = 1$) in this region. However, when the receiver sent the Bad channel state indicator in the previous round which is correctly decoded as Bad at the transmitter ($q_2 = \widehat{q_2} = 1$), the receiver will make an estimate and decide if this power level is enough to avoid outage. Based on this, it classifies the region into Good ($q_3 = 0$) or Bad ($q_3 = 1$) as can be seen in Figure 2.4(c). Outside the outermost circle, the transmitter receives Good ($\widehat{q_3} = 0$). In the shell between the two circles also, the transmitter receives Good ($\widehat{q_3} = 0$) due to the asymmetry in the feedback errors. However, the inside circle would be divided in two parts since the transmitter may mistake the Bad channel state ($q_3 = 1$) as the Good channel state. The inner most would be received as

Good ($\widehat{q}_3 = 0$) while the outer part will be received as Bad ($\widehat{q}_3 = 1$). Thus, we see that the region corresponding to the uncertain part at the transmitter is always a circle centered at the origin and whose radius keeps on shrinking with more rounds of the iterative protocol.

There are two main outage events in this protocol. The first outage event is the channel state being in the inner most circle where the transmitter received Good ($\widehat{q}_3 = 0$) since lower power would be used for transmission than is needed to avoid outage. The second outage event is the area that the receiver classified as Bad ($q_3 = 1$) and the transmitter correctly decoded it as Bad ($\widehat{q}_3 = 1$) and the higher power level used by the transmitter is insufficient to avoid outage. Since both these outage events can be represented as one inner-circle which shrinks with increasing rounds, the resulting diversity order increases in accordance.

Note that both FDD and TDD protocols have the element of using information collected in previous rounds to form the next feedback signal. This memory in encoding channel state information is critical to *adaptive zooming* into the actual channel state. As apparent from the above discussion, the protocol zooms into the poor channel states on-demand, and at the same time tries to keep both the transmitter and receiver aligned in their information about the channel.

2.A Proof of Theorem 1

We first note some properties of $I_\nu(x)$, the modified Bessel function of first kind, that will be used in the proof. The series expansion of $I_\nu(x)$ is given as [1, Equation 9.6.10][19],

$$I_\nu(x) = \sum_{i=0}^{\infty} \frac{1}{i!(i+\nu)!} \left(\frac{x}{2}\right)^{2i+\nu}. \tag{2.19}$$

When $|x|$ is large and $|\arg(x)| < \frac{\pi}{2}$, the asymptotic expansion of $I_\nu(x)$ is given by [1, Equation 9.7.1]

$$I_\nu(x) \doteq \frac{e^x}{\sqrt{2\pi x}} \left\{ 1 - \frac{\mu-1}{x} + \frac{(\mu-1)(\mu-9)}{2!(8x)^2} - \frac{(\mu-1)(\mu-9)(\mu-25)}{3!(8x)^3} + \cdots \right\}, \tag{2.20}$$

where $\mu = 4\nu^2$.

Now, for the proof of Theorem 1, we will use Lemma 1. We will further suppose that $m_N = m$ without loss of generality. Let $A = \{i : \alpha_i + \widehat{\alpha}_i \leq 2\}$. We will evaluate $p(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}})$ in the following five disjoint cases which comprise the whole space of possibilities.

1. $\min(\alpha_m, \widehat{\alpha}_m) \geq 1$ (or $(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \in E_m$).
2. $\min(\alpha_m, \widehat{\alpha}_m) < 1$, $\alpha_m + \widehat{\alpha}_m \geq 2$.
3. $\min(\alpha_m, \widehat{\alpha}_m) < 1$, $\alpha_m + \widehat{\alpha}_m < 2$, $\alpha_i \neq \widehat{\alpha}_i$ for some $i \in A$.
4. $\min(\alpha_m, \widehat{\alpha}_m) < 1$, $\alpha_m + \widehat{\alpha}_m < 2$, $\alpha_i = \widehat{\alpha}_i$ for all $i \in A$, $(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \notin \bigcup_{k=0}^{m-1}(E_k)$.
5. $(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \in \bigcup_{k=0}^{m-1}(E_k)$.

Now, we consider all of the cases one by one as follows.

1. $\min(\alpha_m, \widehat{\alpha}_m) \geq 1$: Using Equation (2.5),

$$p(\lambda, \widehat{\lambda}) = \frac{\exp\left(-\frac{\sum_{k=1}^m \lambda_k + \widehat{\lambda}_k}{1-|\rho|^2}\right) \triangle(\lambda)\triangle(\widehat{\lambda}) \Pi_{k=1}^m (\sqrt{\lambda_k \widehat{\lambda}_k})^\nu \det \left| I_\nu \left(\frac{2|\rho|\sqrt{\lambda_k \widehat{\lambda}_l}}{1-|\rho|^2} \right) \right|}{m! m! \Pi_{j=0}^{m-1} j!(j+\nu)! |\rho|^{mn-m} (1-|\rho|^2)^m}. \tag{2.21}$$

Denote $L = \exp\left(-\frac{\sum_{i=1}^m \mathsf{SNR}^{-\alpha_i} + \mathsf{SNR}^{-\widehat{\alpha}_i}}{1/\mathsf{SNR}}\right)$. Since $\rho = \frac{1}{\sqrt{1+\frac{m}{\mathsf{SNR}}}}$, we substitute $\rho \doteq 1$ and $1 - \rho^2 \doteq 1/\mathsf{SNR}$ to obtain

$$\begin{aligned} p(\alpha, \widehat{\alpha}) \;\; \doteq \;\; & L \triangle(\mathsf{SNR}^{-\alpha}) \triangle(\mathsf{SNR}^{-\widehat{\alpha}}) \mathsf{SNR}^{-\sum_{i=1}^m \alpha_i} \mathsf{SNR}^{-\sum_{i=1}^m \widehat{\alpha}_i} \\ & \frac{\mathsf{SNR}^{-\sum_{i=1}^m \frac{n-m}{2}(\alpha_i + \widehat{\alpha}_i)} \det \left| I_{n-m} \left(\frac{2\sqrt{\mathsf{SNR}^{-\alpha_k} \mathsf{SNR}^{-\widehat{\alpha}_l}}}{1/\mathsf{SNR}} \right) \right|}{m! m! \Pi_{j=0}^{m-1} j!(j+n-m)!(1/\mathsf{SNR}^m)}, \end{aligned} \tag{2.22}$$

which leads to

$$p(\alpha,\widehat{\alpha}) \doteq \frac{L\triangle(\mathsf{SNR}^{-\alpha})\triangle(\mathsf{SNR}^{-\widehat{\alpha}})\mathsf{SNR}^{-\sum_{i=1}^{m}\alpha_i}\mathsf{SNR}^{-\sum_{i=1}^{m}\widehat{\alpha}_i}\mathsf{SNR}^{-\sum_{i=1}^{m}\frac{n-m}{2}(\alpha_i+\widehat{\alpha}_i)}\det\left|I_{n-m}\left(\frac{2\sqrt{\mathsf{SNR}^{-\alpha_k}\mathsf{SNR}^{-\widehat{\alpha}_l}}}{1/\mathsf{SNR}}\right)\right|}{m!m!\Pi_{j=0}^{m-1}j!(j+n-m)!(1/\mathsf{SNR}^m)}. \quad (2.23)$$

As $\triangle(\mathsf{SNR}^{-\alpha}) \doteq \mathsf{SNR}^{-\sum_{i=1}^{m}(i-1)\alpha_i}$, we get

$$p(\alpha,\widehat{\alpha}) \doteq \mathsf{SNR}^{m}\mathsf{SNR}^{-\sum_{i=1}^{m}(i+\frac{n-m}{2})(\alpha_i+\widehat{\alpha}_i)}\det\left|I_{n-m}\left(\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}\right)\right|. \quad (2.24)$$

We will now find $\det\left|I_\nu\left(\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}\right)\right|$. Using (2.19), we get

$$\begin{aligned} I_\nu\left(\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}\right) &= \sum_{i=0}^{\infty}\frac{1}{i!(i+\nu)!}\left(\frac{\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}}{2}\right)^{2i+\nu} \\ &= \left(\frac{\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}}{2}\right)^{\nu}\sum_{i=0}^{\infty}\frac{1}{i!(i+\nu)!}\left(\frac{\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}}{2}\right)^{2i} \quad (2.25) \end{aligned}$$

Let $\mathsf{per}(k_1,k_2,\cdots,k_m)$ for $(k_1,k_2,\cdots,k_m)$ a permutation of $(1,\cdots,m)$ be defined as follows. $\mathsf{per}(k_1,k_2,\cdots,k_m) \triangleq$

$$\begin{cases} 0 & \text{if } (k_1,k_2,\cdots,k_m) \text{ is an even permutation of } (1,\cdots,m). \\ 1 & \text{if } (k_1,k_2,\cdots,k_m) \text{ is an odd permutation of } (1,\cdots,m) \end{cases}. \quad (2.26)$$

Thus,

$$\begin{aligned}
&\det\left|I_\nu\left(\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}\right)\right| \\
= &\sum_{\mathbf{k}}(-1)^{\mathsf{per}(k_1,k_2,\cdots,k_m)}\Pi_{l=1}^m\left(\frac{\mathsf{SNR}^{1-\frac{\alpha_{k_l}+\widehat{\alpha}_l}{2}}}{2}\right)^\nu\sum_{i=0}^\infty\frac{1}{i!(i+\nu)!}\left(\frac{\mathsf{SNR}^{1-\frac{\alpha_{k_l}+\widehat{\alpha}_l}{2}}}{2}\right)^{2i} \\
\doteq &\ \mathsf{SNR}^{\nu\sum_{i=1}^m(1-\frac{\alpha_i+\widehat{\alpha}_i}{2})}\sum_{\mathbf{k}}(-1)^{per(k_1,k_2,\cdots,k_m)}\Pi_{l=1}^m\sum_{i=0}^\infty\frac{1}{i!(i+\nu)!}\left(\frac{\mathsf{SNR}^{2-\alpha_{k_l}+\widehat{\alpha}_l}}{4}\right)^i.
\end{aligned}$$

The above equation is same as Equation (56) in [51] with $K=\frac{\sqrt{2}-1}{2}$, $\phi_i=\mathsf{SNR}^{1-\alpha_i}$ and $\lambda_i=\mathsf{SNR}^{1-\widehat{\alpha}_i}$, and hence

$$\begin{aligned}
\det\left|I_\nu\left(\mathsf{SNR}^{1-\frac{\alpha_k+\widehat{\alpha}_l}{2}}\right)\right| &\doteq \mathsf{SNR}^{\nu\sum_{i=1}^m(1-\frac{\alpha_i+\widehat{\alpha}_i}{2})}\triangle(\mathsf{SNR}^{1-\alpha_i})\triangle(\mathsf{SNR}^{1-\widehat{\alpha}_i}) \\
&\doteq \mathsf{SNR}^{\nu\sum_{i=1}^m(1-\frac{\alpha_i+\widehat{\alpha}_i}{2})}\mathsf{SNR}^{2\sum_{i=1}^m(i-1)(1-\frac{\alpha_i+\widehat{\alpha}_i}{2})} \\
&\doteq \mathsf{SNR}^{\sum_{i=1}^m(2i-2+\nu)\left(1-\frac{\alpha_i+\widehat{\alpha}_i}{2}\right)}. \qquad (2.27)
\end{aligned}$$

Substituting in Equation (2.24), we get

$$\begin{aligned}
p(\alpha,\widehat{\alpha}) &\doteq \mathsf{SNR}^m\mathsf{SNR}^{-\sum_{i=1}^m(i+\frac{n-m}{2})(\alpha_i+\widehat{\alpha}_i)}\mathsf{SNR}^{\sum_{i=1}^m(2i-2+\nu)\left(1-\frac{\alpha_i+\widehat{\alpha}_i}{2}\right)} \\
&\doteq \mathsf{SNR}^{mn}\mathsf{SNR}^{-\sum_{i=1}^m(i+\frac{n-m}{2})(\alpha_i+\widehat{\alpha}_i)}\mathsf{SNR}^{-\sum_{i=1}^m(i-1+\frac{n-m}{2})(\alpha_i+\widehat{\alpha}_i)} \\
&\doteq \mathsf{SNR}^{mn}\mathsf{SNR}^{-\sum_{i=1}^m(2i+n-m-1)(\alpha_i+\widehat{\alpha}_i)}=e_m. \qquad (2.28)
\end{aligned}$$

2. $\min(\alpha_m,\widehat{\alpha}_m)<1$, $\alpha_m+\widehat{\alpha}_m\geq 2$: Using Equation (2.23), we see that all terms except $\exp\left(-\frac{\sum_{i=1}^m\mathsf{SNR}^{-\alpha_i}+\mathsf{SNR}^{-\widehat{\alpha}_i}}{1/\mathsf{SNR}}\right)$ remain the same and hence are polynomial in SNR while

   $\exp\left(-\frac{\sum_{i=1}^m\mathsf{SNR}^{-\alpha_i}+\mathsf{SNR}^{-\widehat{\alpha}_i}}{1/\mathsf{SNR}}\right)$ decreases exponentially in SNR and hence $p(\alpha,\widehat{\alpha})\doteq 0$.

3. $\min(\alpha_m,\widehat{\alpha}_m)<1$, $\alpha_m+\widehat{\alpha}_m<2$, $\alpha_i\neq\widehat{\alpha}_i$ for some $i\in A$: We will prove $p(\alpha,\widehat{\alpha})\doteq 0$

in this case. For this we consider $\exp\left(-\frac{\sum_{k=1}^{m}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)\det\left|I_\nu\left(\frac{2|\rho|\sqrt{\lambda_k\widehat{\lambda}_l}}{1-|\rho|^2}\right)\right|$ and prove this part to decrease exponentially and we would be done since rest of the terms are polynomial in SNR. Using (2.20),

$$\exp\left(-\frac{\sum_{k=1}^{m}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)\det\left|I_\nu\left(\frac{2|\rho|\sqrt{\lambda_k\widehat{\lambda}_l}}{1-|\rho|^2}\right)\right| \doteq \exp\left(-\frac{\sum_{k=1}^{m}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)$$
$$\left(\sum_{\mathbf{k}}(-1)^{\mathsf{per}(k_1,k_2,\cdots,k_m)}\Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l\leq 2}^{m}\frac{\exp\left(\frac{2|\rho|\sqrt{\lambda_{k_l}\widehat{\lambda}_l}}{1-\rho^2}\right)}{\sqrt{\frac{2\sqrt{\lambda_{k_l}\widehat{\lambda}_l}}{1-|\rho|^2}}}(1-\frac{\mu-1}{\frac{16\sqrt{\lambda_{k_l}\widehat{\lambda}_l}}{1-\rho^2}}+\cdots)\right.$$
$$\left.\Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l>2}^{m}\mathsf{Poly}(\mathsf{SNR})\right), \tag{2.29}$$

where $\mathsf{Poly}(\mathsf{SNR})$ represents the term that are polynomial in SNR. First observe that

$$\exp\left(-\frac{\lambda_{k_l}+\widehat{\lambda}_l}{1-\rho^2}\right)\exp\left(\frac{2|\rho|\sqrt{\lambda_{k_l}\widehat{\lambda}_l}}{1-\rho^2}\right)\leq 1, \tag{2.30}$$

and hence the product of such terms cannot increase exponentially with SNR.

$$\begin{aligned}
&\exp\left(-\frac{\sum_{k=1}^{m}\lambda_k+\widehat{\lambda}_k}{1-|\rho|^2}\right)\det\left|I_\nu\left(\frac{2|\rho|\sqrt{\lambda_k\widehat{\lambda}_l}}{1-|\rho|^2}\right)\right| \\
&\doteq \sum_{\mathbf{k}}(-1)^{\mathsf{per}(k_1,k_2,\cdots,k_m)}\Pi_{l=1}^{m}\exp(-\mathsf{SNR}^{1-\min(\alpha_{k_l},\widehat{\alpha}_l)}) \\
&\quad \Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l\leq 2}^{m}\exp(\mathsf{SNR}^{1-\frac{\alpha_{k_l}+\widehat{\alpha}_l}{2}})\mathsf{Poly}(\mathsf{SNR}) \\
&\doteq \sum_{\mathbf{k}}(-1)^{\mathsf{per}(k_1,k_2,\cdots,k_m)}\Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l>2}^{m}\exp(-\mathsf{SNR}^{1-\min(\alpha_{k_l},\widehat{\alpha}_l)}) \\
&\quad \Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l\leq 2}^{m}\exp(-\mathsf{SNR}^{1-\min(\alpha_{k_l},\widehat{\alpha}_l)})\exp(\mathsf{SNR}^{1-\frac{\alpha_{k_l}+\widehat{\alpha}_l}{2}})\mathsf{Poly}(\mathsf{SNR}) \\
&\dot{\leq} \sum_{\mathbf{k}}(-1)^{\mathsf{per}(k_1,k_2,\cdots,k_m)}\Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l>2}^{m}\mathbf{1}_{\min(\alpha_{k_l},\widehat{\alpha}_l)\geq 1}\Pi_{l=1,\alpha_{k_l}+\widehat{\alpha}_l\leq 2}^{m}\mathbf{1}_{\alpha_{k_l}=\widehat{\alpha}_l}\mathsf{Poly}(\mathsf{SNR}).
\end{aligned}$$

Now, we will show that each term under the sum decays exponentially with SNR.

For this, the product of the above indicators should be 0. Let us now consider all the scenarios when the above product of indicators is 1. If some $\widehat{\alpha}_l < 1$, and pairs with α_{k_l} to give $\widehat{\alpha}_l + \alpha_{k_l} \leq 2$, then the two must be equal and if it pairs to give $\widehat{\alpha}_l + \alpha_{k_l} > 2$, the product of indicators will always be 0. If $\widehat{\alpha}_l \geq 1$ and pairs with α_{k_l} to give $\widehat{\alpha}_l + \alpha_{k_l} \leq 2$, it can only happen when the two are equal and if it pairs to give $\widehat{\alpha}_l + \alpha_{k_l} > 2$, then $\alpha_{k_l} \geq 1$. Hence the only pairing that will work is that all elements of $\alpha_i < 1$ are matched to $\widehat{\alpha}_j < 1$ and also $\alpha_i \geq 1$ is not mapped to $\widehat{\alpha}_i < 1$. This can happen only when $\alpha_i = \widehat{\alpha}_i$ whenever $\alpha_i < 1$ or $\widehat{\alpha}_i < 1$ and all the rest are ≥ 1. If $\alpha_i \neq \widehat{\alpha}_i$ for some $i \in A$, the above condition do not hold. This proves Case 3.

4. $\min(\alpha_m, \widehat{\alpha}_m) < 1$, $\alpha_m + \widehat{\alpha}_m < 2$, $\alpha_i = \widehat{\alpha}_i$ for all $i \in A$, $(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) \notin \bigcup_{k=0}^{m-1}(E_k)$: We see that all the analysis of Case 3 holds for this case and hence if $\alpha_i = \widehat{\alpha}_i$ for all $i \in A$ and $(\alpha, \widehat{\alpha}) \notin \bigcup_{k=0}^{m-1}(E_k)$, an element ≥ 1 is mapped to an element < 1 which makes the product of indicators zero and hence the probability decreases exponentially with SNR.

5. $\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}} \in \bigcup_{k=0}^{m-1}(E_k)$. These are the cases we sum over in the statement of the Lemma. In each of these cases, $p(\alpha, \widehat{\alpha})$ exist. When we integrate over $\widehat{\alpha}$, we find that integral of $\sum_k e_k \mathbf{1}_{E_k}$ w.r.t. $\widehat{\alpha}$ is $\Pi_{i=1}^{m} \mathsf{SNR}^{-(2i-1+n-m)\alpha_i} \mathbf{1}_{\alpha_1, \cdots, \alpha_k \geq 1} \mathbf{1}_{0 \leq \alpha_{k+1}, \cdots, \alpha_{m_N} < 1}$. We next note that in order to find the polynomial expressions associated, the two exponentials multiplication cannot decrease with SNR for which we would need $k_l = l$ in (2.29) (since all the λ's and $\widehat{\lambda}$'s are ordered) and hence from all the expressions, we find that $p(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) = p_1(\boldsymbol{\alpha}) p_2(\widehat{\boldsymbol{\alpha}})$ for any E_k. Using the separability, we find that

$$e_k = \mathsf{SNR}^{k(n-m+k)} \Pi_{i=1}^{k} \mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_i} \Pi_{i=1}^{m_N} \mathsf{SNR}^{-(2i-1+|n-m|)\alpha_i}. \tag{2.31}$$

2.B Proof of Theorem 2

We will first verify that the power levels used in the protocol satisfy the average power constraints.

1. **Round 1**: Transmitter trains receiver with power SNR, and hence the average power constraint is satisfied. The receiver uses a maximal power level of SNR and hence the average power constraint is satisfied.

2. **Round** $2 \leq u+1 \leq K-1$:

$$\begin{aligned}
\Pi(\widehat{q}_1 = 1) &= \Pi(\widehat{q}_1 = 1, q_1(\widehat{H}_1) = 0) + \Pi(J_{\mathsf{T}1} = 1, q_1(\widehat{H}_1) = 1) \\
&\leq \Pi(\widehat{q}_1 = 1 | q_1(\widehat{H}_1) = 0) + \Pi(q_1(\widehat{H}_1) = 1) \\
&\dot{\leq} \mathsf{SNR}^{-mn+\epsilon} + \mathsf{SNR}^{-G_1(r+\epsilon)} \\
&\doteq \mathsf{SNR}^{-G_1(r+\epsilon)}, \qquad (2.32)
\end{aligned}$$

where we used $\Pi(\widehat{q}_1 = 1 | q_1(\widehat{H}_1) = 0) \doteq \mathsf{SNR}^{-mn+\epsilon}$ since the threshold is at $\mathsf{snr}^{\epsilon/mn}$ and by [16]. Further, $\Pi(q_1(\widehat{H}_1) = 1) = \det(I + \widehat{H}_1 \widehat{H}_1^{\dagger} \mathsf{SNR}) < \mathsf{SNR}^{r+\epsilon} \doteq \mathsf{SNR}^{-G_1(r+\epsilon)}$. The last step follows since $G_1(r+\epsilon) \leq G_1(\epsilon) = mn - \epsilon(m+n-1) \leq mn - \epsilon$.

Hence, the power constraint is satisfied at the transmitter for $u = 1$. For other rounds consider

$$
\begin{aligned}
\Pi(q_{u+1}=0) &= \Pi(q_{u+1}=0, q_u=0) + \Pi(q_{u+1}=0, q_u=1) \\
&\doteq \Pi(q_{u+1}=0, q_u=0, \widehat{q}_u=0) + \Pi(q_{u+1}=0, q_u=0, \widehat{q}_u=1) \\
&\quad + \Pi(q_{u+1}=0, \widehat{q}_u=1, q_u=1) \\
&\doteq \Pi(q_{u+1}=0, q_u=0, \widehat{q}_u=1) + \Pi(q_{u+1}=0, \widehat{q}_u=1, q_u=1) \\
&\dot{\leq} \Pi(\widehat{q}_u=1) \\
&\dot{\leq} \mathsf{SNR}^{-G_u(r+\epsilon)}, \qquad (2.33)
\end{aligned}
$$

where we used three things:

(a) $\Pi(q_{u+1}=0, q_u=1, \widehat{q}_u=0) \doteq 0$. If $u=1$, $\Pi(q_u=1, \widehat{q}_u=0) \doteq 0$. For this consider two cases. If $q_i=0$ for some $i<u$, then as $\widehat{q}_u=0$, $q_{u+1}=1$ since the power estimate $\geq \mathsf{SNR}^{\epsilon}$ happens with probability $\doteq 0$. Further, if $q_i=1$ for all $i \leq u$, $\widehat{q}_u=0$ happens with probability $\doteq 0$ since any transmitted 1 from the receiver gets received as 0 with exponentially small probability.

(b) $\Pi(q_{u+1}=0, q_u=0, \widehat{q}_u=0) \doteq 0$. This is because when transmitter sends at 0 power, the receiver receives at power $\geq \mathsf{SNR}^{\epsilon}$ with exponentially small probability.

(c) $\Pi(\widehat{q}_u=1) \dot{\leq} \mathsf{SNR}^{-G_u(r+\epsilon)}$ which has been earlier shown to be true for $u=1$, and we will show in general by induction.

Hence, the average power constraint at the receiver is satisfied for all u. To check the induction argument for $\widehat{q}_u$ and transmit power constraint,

$$
\Pi(\widehat{q}_{u+1}=1) = \Pi(\widehat{q}_{u+1}=1, q_{u+1}=0) + \Pi(\widehat{q}_{u+1}=1, q_{u+1}=1). \quad (2.34)
$$

For the first term $\Pi(\widehat{q}_{u+1} = 1, q_{u+1} = 0)$. This happens when

$$\text{trace}(HH^\dagger)\mathsf{SNR}^{1+G_u(r+\epsilon)} + 1 \dot{\leq} \mathsf{SNR}^{\epsilon/mn},$$

which happens with probability $\mathsf{SNR}^{-mn(1+G_u(r+\epsilon))+\epsilon}$.

The second term can occur only when all the q_1 to $q_{u+1} = 1$, and all $\widehat{q}_1$ to $\widehat{q}_{u+1} = 1$. The reason for all q_1 to $q_{u+1} = 1$ is because if any of $q_i = 0$ all $q_{i+1} = q_{u+1} = 0$ since the forward channel was good and cannot be bad in the next rounds. This happens when $\det(I + \widehat{H}_{u+1}\widehat{H}_{u+1}^\dagger \mathsf{SNR}^{1+G_u(r+\epsilon)}) < \mathsf{SNR}^{r+\epsilon}$, which happens with probability $\mathsf{SNR}^{-G_{u+1}(r+\epsilon)}$.

Note that

$$\begin{aligned} G_{u+1}(r+\epsilon) &= G(r+\epsilon, 1+G_u(r+\epsilon)) \\ &\leq G(\epsilon, 1+G_u(r+\epsilon)) = mn(1+G_u(r+\epsilon)) - (m+n-1)\epsilon \\ &\leq mn(1+G_u(r+\epsilon)) - \epsilon. \end{aligned} \tag{2.35}$$

Hence, $\Pi(\widehat{q}_{u+1} = 1) \dot{\leq} \mathsf{SNR}^{-G_{u+1}(r+\epsilon)}$. Thus, the induction steps hold, and the power constraint is satisfied.

3. **Round K (forward only)**:

$$\begin{aligned} & \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1) \\ = \; & \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_2 = 0) + \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_1 = 0, q_2 = 1) \\ & + \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_1 = 1, q_2 = 1) \\ \dot{\leq} \; & \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_2 = 0, q_1 = 0) + \Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_2 = 0, q_1 = 1) \\ & + \Pi(q_1 = 0, \widehat{q}_1 = 1, q_2 = 1) + \Pi(q_2 = 1 | q_1 = 1, \widehat{q}_1 = 1). \end{aligned} \tag{2.36}$$

The first term $\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_2 = 0, q_1 = 0$ happens asymptotically with probability that there is error in the reverse channel both times. This happens with probability $\mathsf{SNR}^{-mn(1+G_1(r+\epsilon))+\epsilon}$.

The second term $\widehat{q}_1 = 1, \widehat{q}_2 = 1, q_2 = 0, q_1 = 1$ happens when $q_1 = 1$ and when there is error in the feedback channel in the second round. This happens with probability less than $\mathsf{SNR}^{-mn(1+G_1(r+\epsilon))+\epsilon}$.

The third term $q_1 = 0, \widehat{q}_1 = 1, q_2 = 1$ happens with exponentially small probability since the probability $\det(I+H_1H_1^\dagger\mathsf{SNR}) \geq \mathsf{SNR}^{r+\epsilon}$ and trace$(H_2H_2^\dagger\mathsf{SNR}^{1+G_1(r+\epsilon)})+1 < \mathsf{SNR}^\epsilon$ is $\doteq 0$. Note that i, j^{th} term of H_1 and H_2 are correlated with correlation $\rho_{i,j}$ s.t. $1-\rho_{i,j}^2 \doteq \mathsf{SNR}^{-1}$. Let the eigenvalues of $H_1H_1^\dagger$ and $H_2H_2^\dagger$ be $\lambda_1, \cdots, \lambda_{\min(m,n)}$ and $\mu_1, \cdots, \mu_{\min(m,n)}$ respectively. Let $\lambda_i = \mathsf{SNR}^{\alpha_i}$ and $\mu_i = \mathsf{SNR}^{-\widehat{\alpha_i}}$.

$$\begin{aligned}
& \Pi\left(\det(I+H_1H_1^\dagger\mathsf{SNR}) \geq \mathsf{SNR}^{r+\epsilon}, \text{trace}(H_2H_2^\dagger\mathsf{SNR}^{1+G_1(r+\epsilon)})+1 < \mathsf{SNR}^\epsilon\right) \\
\doteq\ & \Pi\left(\sum(1-\alpha_i)^+ \geq r+\epsilon, (1+G_1(r+\epsilon)-\min_i \widehat{\alpha}_i)^+ < \epsilon\right) \\
\doteq\ & \Pi\left(\sum(1-\widehat{\alpha}_i)^+ \geq r+\epsilon, (1+G_1(r+\epsilon)-\min_i \widehat{\alpha}_i)^+ < \epsilon\right) \\
\doteq\ & 0 \qquad (2.37)
\end{aligned}$$

The fourth term in right hand side of (2.36),

$$\Pi(q_2 = 1 | q_1 = 1, \widehat{q}_1 = 1) \doteq \mathsf{SNR}^{-G_2(r+\epsilon)}. \qquad (2.38)$$

Hence, (2.36) reduces to

$$\begin{aligned}
\Pi(\widehat{q}_1 = 1, \widehat{q}_2 = 1) &\doteq \mathsf{SNR}^{-mn(1+G_1(r+\epsilon))+\epsilon} + \mathsf{SNR}^{-G_2(r+\epsilon)} \\
&\doteq \mathsf{SNR}^{-G_2(r+\epsilon)}. \qquad (2.39)
\end{aligned}$$

Similarly, we see that for $u \geq 2$, if $\widehat{q}_1 = \cdots = \widehat{q}_u = 1$. Then, if any of $q_u = i$ for some $i \leq u$, then $q_u = 0$. Also, the feedback channel transforms $q_u = 0$ to $\widehat{q}_u = 1$, which happens with probability $\mathsf{SNR}^{-mn(1+G_{u-1}(r+\epsilon))+\epsilon}$. However, if $q_i = 1$ for all $i \leq u$, this happens when $\log\det(I + H_u H_u^\dagger \mathsf{SNR}^{1+G_{u-1}(r+\epsilon)}) \geq \mathsf{SNR}^{r+\epsilon}$ which happens with probability $\mathsf{SNR}^{-G_u(r+\epsilon)}$. Hence, the event $\widehat{q}_1 = \cdots = \widehat{q}_u = 1$ happens with probability $\doteq \mathsf{SNR}^{-G_u(r+\epsilon)}$. Hence, the power constraint is satisfied.

We will now show that the outage probability is $\doteq \mathsf{SNR}^{-G(r,1+G_{K-1}(r+\epsilon))}$. For this, we see the region of H when there will be outage.

1. $\det(I + HH^\dagger \mathsf{SNR}) \geq \mathsf{SNR}^r$. In this case, a power level of SNR is sufficient, while we will be using a higher power level. Hence, there can be no outage.

2. $\det(I + HH^\dagger \mathsf{SNR}) < \mathsf{SNR}^r$ and $\det(I + HH^\dagger \mathsf{SNR}^{1+G_1(r+\epsilon)}) \geq \mathsf{SNR}^r$. In this case, $q_1 = 1$ with probability 1. The reason is that

$$\Pi\left(\sum_{i=1}^{\min(m,n)} (1-\alpha_i)^+ \leq r, \sum_{i=1}^{\min(m,n)} (1-\widehat{\alpha}_i)^+ \geq r+\epsilon\right) \doteq 0,$$

where α_i and $\widehat{\alpha}_i$ are the negative exponents of the eigenvalues of $HH^\dagger$ and $\widehat{H}_1\widehat{H}_1^\dagger$. Since $q_1 = 1$, $\widehat{q}_1 = 1$ with probability 1. Since $\widehat{q}_1 = 1$, power of atleast $\mathsf{SNR}^{2-r-\epsilon}$ is used and hence no outage in this range.

3. For $u \in [2, K-1]$, $\det(I + HH^\dagger \mathsf{SNR}^{1+G_{u-1}(r+\epsilon)}) < \mathsf{SNR}^r$ and $\det(I + HH^\dagger \mathsf{SNR}^{1+G_u(r+\epsilon)}) \geq \mathsf{SNR}^r$. As before, $q_1 = \widehat{q}_1 = 1$. Now, since training is at $\mathsf{SNR}^{1+G_1(r+\epsilon)}$,

$$\Pi\left(\sum_{i=1}^{\min(m,n)} (1+G_1(r+\epsilon)-\alpha_i)^+ \leq r, \sum_{i=1}^{\min(m,n)} (1+G_1(r+\epsilon)-\widehat{\alpha}_i)^+ \geq r+\epsilon\right) \doteq 0$$

(where α_i and $\widehat{\alpha}_i$ are the negative exponents of the eigenvalues of $HH^\dagger$ and $\widehat{H}_2\widehat{H}_2^\dagger$). Hence, $q_2 = 1$ which implies $\widehat{q}_2 = 1$ with probability 1. Similarly extending all till

$q_u = \widehat{q}_u = 1$ with probability 1. Hence there is no outage in this range

Hence, the outage happens only when $\det(I + HH^{\dagger}\mathsf{SNR}^{1+G_u(r+\epsilon)}) < \mathsf{SNR}^r$ which happens with probability $\mathsf{SNR}^{-G(r,1+G_{K-1}(r+\epsilon))}$. For ϵ very small, diversity of $G_K(r)$ can be achieved.

2.C TDD: Perfect CSIR or Perfect CSIT

In this Appendix, we consider two cases. The first case assumes perfect channel state information at the transmitter, where only the receiver is unaware of channel conditions. Second, we consider the case of perfect information at the receiver (CSIR), with no channel information at the transmitter. In both the above cases, infinite diversity order can be achieved with 1.5 round protocols which is unlike the case of FDD systems [16].

For the case of perfect channel knowledge at the transmitter and receiver, it is well known [21] for the case of $\min(m,n) > 1$, zero outage probability is possible at a finite SNR value and thus the diversity order is unbounded. For the case of single antenna system, i.e $\min(m,n) = 1$, channel inversion leads to an exponential decay in probability and hence also has an infinite diversity order. We show in the next two results that infinite diversity order can be achieved even if only one of the nodes, transmitter or receiver has perfect channel information.

We begin the discussion with the case when the channel information is perfect at the transmitter and receiver relies on estimated channel knowledge.

Lemma 3 (No CSIR, Perfect CSIT)**.** *For all* $0 < r < \min(m,n)$*, an infinite diversity order is achievable with a 0.5 round protocol.*

Proof. The protocol proceeds as follows.

Round 1 (forward only): The transmitter trains the receiver with a power level

$$P(H) \doteq \frac{\mathsf{SNR}^{1-\epsilon/mn}}{\Pi_{i=1}^{m_N}\mathsf{SNR}^{-(2i-1+|n-m|)\alpha_i}},$$

where $\lambda_i \doteq \mathsf{SNR}^{-\alpha_i}$ are the eigen-values of the channel realization H at the transmitter, and sends data at the same power. It is straightforward to conclude that the power constraint will be satisfied, and the outage probability $\doteq 0$ for $r < \min(m,n) - \epsilon$. (By choosing ϵ small enough, the above holds for $r < \min(m,n)$.) Note that the power control from the transmitter is only asymptotically satisfied. $\square$

Since the transmitter knows the channel perfectly, it can "invert" the channel exactly for the case of $\min(m,n) > 1$ and in this case, the receiver sees a channel whose SNR does not fluctuate and is always above the threshold needed to avoid the outage. Similarly, for the case of no information at the transmitter but full information at the receiver, we can achieve an error probability decay which is faster than any polynomial decay in SNR.

Lemma 4 (Perfect CSIR, No CSIT)**.** *For* $0 < r < \min(m,n)$*, an infinite diversity order is achievable with a 1.5 round protocol.*

Proof. The protocol proceeds as follows.

1. **Round 1 (forward):** The transmitter remains silent.

   Round 1 (reverse): Since the receiver knows the channel H, it finds the index of the power level from $i \in [1, K]$ that would be sufficient to avoid outage as in [36]. Let $\lambda_i \doteq \mathsf{SNR}^{-\alpha_i}$ be the eigen-values of H with $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m$. Further, the receiver communicates this power level to the transmitter using a power level of $\mathsf{SNR}^{\alpha_m + \frac{i+1}{2K}}$ to communicate index i. The transmitter estimates the received power and hence gets the index i with error probability $\doteq 0$.

2. **Round 2 (forward only):** Since the transmitter now knows the power level to use, it sends data at the required power level.

First note that the probability with which the power level of $\mathsf{SNR}^{\alpha_m + \frac{i+1}{2K}}$ is used is $\dot{\leq}\mathsf{SNR}^{-\alpha_m}$ and hence the power constraint is satisfied. The only step to prove is to show that the probability with which the transmitter will make a mistake in the index transmitted by

the receiver is $\doteq 0$. Suppose that the transmitter uses a threshold for level i at $\mathsf{SNR}^{\frac{i+1}{2K}+\frac{1}{4K}}$. Since the transmitted power is $\mathsf{SNR}^{\alpha_m+\frac{i+1}{2K}}$, the received power is $\doteq \mathsf{SNR}^{\alpha_m+\frac{i+1}{2K}}\mathsf{SNR}^{-\alpha_m} = \mathsf{SNR}^{\frac{i+1}{2K}}$. Here, we exploited the reciprocity in the channels since α_m for the forward and the backward channels are the same. Since there is nothing random in the above exponent of SNR in this received power (asymptotically), using the above threshold, the index will be mistaken with probability $\doteq 0$.

The above strategy attains the diversity of $G_K(r)$ for any finite K. Since $G_K(r)$ monotonically increases with K for $0 < r < \min(m,n)$, for any given $x \geq 0$ diversity $\geq x$ can be achieved by choosing K large enough. Thus, the above protocol yields unbounded diversity. $\square$

2.D Proof of Lemma 2

The probability that the channel cannot support rate R is given by

$$\begin{aligned}
& \Pi\left(\log\det(I + P(\widehat{H}_1)HH^\dagger) < R\right) \\
\doteq\ & \Pi\left(\sum_{i=1}^{\min(m,n)}(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_i - \alpha_i)^+ < r\right) \\
\doteq\ & \int_{\mathcal{A}} p(\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}) d\boldsymbol{\alpha} d\widehat{\boldsymbol{\alpha}} \\
\doteq\ & \sum_{k=0}^{\min(m,n)} \int_{\mathcal{A}\cap E_k} \mathsf{SNR}^{k(|n-m|+k)} \Pi_{i=1}^{k} \mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_i} \Pi_{i=1}^{m_N} \mathsf{SNR}^{-(2i-1+|n-m|)\alpha_i} d\boldsymbol{\alpha} d\widehat{\boldsymbol{\alpha}},
\end{aligned}$$

where $\mathcal{A} = \{\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}} : \sum_{i=1}^{\min(m,n)}\left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_i - \alpha_i\right)^+ < r$.

We first note that $\mathcal{A} \cap E_k = \Phi$ for $k < \min(m,n)$. Let $m \leq n$, as the other case is similar. For any E_k where $k < \min(m,n)$, $0 \leq \alpha_i = \widehat{\alpha}_i < p$ for all$i > k$ and $\min(\alpha_i, \widehat{\alpha}_i) \geq p$ for all $i = 1, \cdots, k$. Thus,

$$
\begin{aligned}
& \sum_{i=1}^{\min(m,n)} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{\min(m,n)} (2i-1+|n-m|)\widehat{\alpha}_i - \alpha_i\right)^+ \\
= & \sum_{i=1}^{k} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{\min(m,n)} (2i-1+|n-m|)\widehat{\alpha}_i - \alpha_i\right)^+ \\
& + \sum_{i=k+1}^{m} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{k}(2i-1+|n-m|)\widehat{\alpha}_i + \sum_{i=k+1}^{m} (2i-1+|n-m|)\alpha_i - \alpha_i\right)^+ \\
\geq & \sum_{i=k+1}^{m} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{k}(2i-1+|n-m|)\widehat{\alpha}_i + \sum_{i=k+1}^{m} (2i-1+|n-m|)\alpha_i - \alpha_i\right)^+ \\
\geq & \sum_{i=k+1}^{m} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{k}(2i-1+|n-m|)\widehat{\alpha}_i\right)^+ \\
\geq & \sum_{i=k+1}^{m} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{k}(2i-1+|n-m|)\right)^+ \\
= & (m-k)\left(1 - \frac{\epsilon}{mn} + k(k+|n-m|)\right) \\
\geq & (m-k)\left(1 + k^2 - \frac{\epsilon}{mn}\right). \qquad (2.40)
\end{aligned}
$$

For $k = 0$, above is $\geq m(1 - \epsilon/mn)$ and for any $r < m - \epsilon$, the above cannot be less than r. Similarly for $k = 1, m > 1$, the above cannot be $< r$. For $k > 1$ we can lower bound $k^2 - k - \frac{\epsilon}{mn}$ by 0 (by choosing $\epsilon < 1$). Thus,

$$
\begin{aligned}
& \sum_{i=1}^{\min(m,n)} \left(1 - \frac{\epsilon}{mn} + \sum_{i=1}^{\min(m,n)} (2i-1+|n-m|)\widehat{\alpha}_i - \alpha_i\right)^+ \\
\geq & (m-k)\left(1 + k^2 - \frac{\epsilon}{mn}\right) \\
\geq & (m-k)(1+k) \\
= & m - k + mk - k^2 \\
= & m + (m-1-k)k \\
\geq & m. \qquad (2.41)
\end{aligned}
$$

Hence,

$$\begin{aligned}
&\Pi\left(\log\det(I+P(\widehat{H}_1)HH^\dagger)<R\right)\\
\doteq\ &\int_{\mathcal{A}\cap E_m}\mathsf{SNR}^{mp(n-m+m)}\Pi_{i=1}^m\mathsf{SNR}^{-(2i-1+n-m)\widehat{\alpha}_i}\Pi_{i=1}^m\mathsf{SNR}^{-(2i-1+n-m)\alpha_i}d\boldsymbol{\alpha}d\widehat{\boldsymbol{\alpha}}\\
\doteq\ &\int_{\mathcal{A}\cap E_m}\mathsf{SNR}^{mnp}\Pi_{i=1}^m\mathsf{SNR}^{-(2i-1+n-m)\widehat{\alpha}_i}\Pi_{i=1}^m\mathsf{SNR}^{-(2i-1+n-m)\alpha_i}d\boldsymbol{\alpha}d\widehat{\boldsymbol{\alpha}}. \quad (2.42)
\end{aligned}$$

Hence, the negative of the SNR exponent of the above probability is

$$\min_{(\boldsymbol{\alpha},\widehat{\boldsymbol{\alpha}})\in\mathcal{A}\cap E_m}\sum_{i=1}^m(2i-1+n-m)(\widehat{\alpha}_i+\alpha_i)-mnp. \quad (2.43)$$

The optimal $\widehat{\alpha}_i=p$. Now, let $\alpha_i=p+\beta_i$. The above negative SNR exponent is given by

$$mnp+\min_{(\beta_1,\cdots,\beta_m)\in\mathcal{B}}\sum_{i=1}^m(2i-1+n-m)\beta_i, \quad (2.44)$$

where $\mathcal{B}=\{(\beta_1,\cdots,\beta_m):\beta_i\geq 0,\sum_{i=1}^m(1+(mn-1)p-\frac{\epsilon}{mn}-\beta_i)^+<r$. By the definition of the function G, the above reduces to

$$mnp+G(r,1+(mn-1)p-\frac{\epsilon}{mn}). \quad (2.45)$$

For $\epsilon\to 0$, the probability that the channel cannot support rate R is thus $\mathsf{SNR}^{-mnp-G(r,1+(mn-1)p-\frac{\epsilon}{mn})}$.

2.E Proof of Theorem 3

To verify that the power constraint is satisfied, observe the following.

1. **Round 1:** The receiver trains the transmitter with power SNR, and hence the power constraint is satisfied.

2. **Round 2:** The transmitter trains the receiver with a power level $P_1\left(\widehat{H}_1\right) \doteq \frac{\mathsf{SNR}}{\Pi_{i=1}^{m_N} \mathsf{SNR}^{-(2i-1+|n-m|)\widehat{\alpha}_{1,i}}}$, and the average power constraint is satisfied.

$$\begin{aligned} \Pi\left(q_2 = 1\right) &= \Pi\left(\det\left(I + G_2 G_2^{\dagger}\right) \geq \mathsf{SNR}^{r+\epsilon}\right) && (2.46) \\ &\dot{\leq} \mathsf{SNR}^{-W_2(r+\epsilon)}, && (2.47) \end{aligned}$$

where the second steps follows from Lemma 2 with $p = 1$.

3. **Round** $u = 3 : K-1$:

$$\begin{aligned} \Pi\left(\widehat{q}_2 = 1\right) &= \Pi\left(\widehat{q}_2 = 1, q_2 = 1\right) + \Pi\left(\widehat{q}_2 = 1, q_2 = 0\right) && (2.48) \\ &\dot{\leq} \Pi\left(q_2 = 1\right) + \Pi\left(\widehat{q}_2 = 1 | q_2 = 0\right) && (2.49) \\ &\dot{\leq} \mathsf{SNR}^{-W_2(r+\epsilon)} + e^{-\mathsf{SNR}^{\epsilon}} && (2.50) \\ &\doteq \mathsf{SNR}^{-W_2(r+\epsilon)} && (2.51) \end{aligned}$$

Further, the training power from the transmitter satisfies the power constraint. For the receiver power constraint, consider

$$\Pi\left(q_u = 1\right) = \Pi\left(q_u = 1, \widehat{q}_{u-1} = 1, q_{u-1} = 1\right) + \Pi\left(q_u = 1, \widehat{q}_{u-1} = 0, q_{u-1} = 1\right).$$

Let us first consider the first term. As $q_{u-1} = 1$ is received as $\widehat{q}_{u-1} = 1$, $\alpha_m \leq 1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$. Since α_m is less than the SNR exponent of the power at which the training symbol was sent, $G_u \in \mathcal{O}_u$ with probability $\doteq 0$ by Lemma 2. Also, the power estimation cannot fail since $\alpha_m = \widehat{\alpha}_m$. Thus, the first term is $\doteq 0$.

The second term is upper bounded by $\Pi\left(\widehat{q}_{u-1}=0|q_{u-1}=1\right)$ which happens with probability

$\mathsf{SNR}^{-mn(1+W_{u-1}(r+\epsilon))+\epsilon} \doteq \mathsf{SNR}^{-W_u(r+\epsilon)}$. Hence,

$$\begin{aligned}\Pi\left(q_u=1\right) &= \Pi\left(q_u=1,\widehat{q}_{u-1}=1,q_{u-1}=1\right)+\Pi\left(q_u=1,\widehat{q}_{u-1}=0,q_{u-1}=1\right)\\ &\dot{\leq} \mathsf{SNR}^{-W_u(r+\epsilon)}. \end{aligned} \tag{2.52}$$

The power constraint at the receiver is satisfied since the average power is bounded by SNR.

4. **Round K (forward only)**: The transmitter trains the receiver and then sends the data both at a power level of

$$\left\{\mathsf{SNR}^{\max\left(1+\sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_{1,i},\max_{u\in[2,K-1]:\widehat{q}_u=1}\left(1+\sum_{i=1}^{\min(m,n)}(2i-1+|n-m|)\widehat{\alpha}_{u,i}\right)\right)}\right. \tag{2.53}$$

Note that the power constraint is satisfied since for any maximizing term in the power, the probability that that particular $\widehat{q}=1$ would balance the exponent.

We will now show that the outage probability is $\doteq mn\left(1+W_{k-1}\left(r+\epsilon\right)\right)$. For $K=2$, it follows directly by Lemma 2 for p=1. We now assume that $K>2$. For this, we see the region of H when there will be outage.

1. $\det\left(I+HH^{\dagger}P_1\left(\widehat{H}_1\right)\right)\geq \mathsf{SNR}^r$. In this case, a power level of $P\left(\widehat{H}_1\right)$ is sufficient, while we will be using a higher power level. Hence, there can be no outage.

2. $\det\left(I+HH^{\dagger}P_1\left(\widehat{H}_1\right)\right)<\mathsf{SNR}^r$ and $\alpha_m<1+W_2\left(r+\epsilon\right)-\frac{\epsilon}{mn}$. In this case, $q_2=1$ with probability 1. The reason is that

$$\Pi\left(\sum_{i=1}^{\min(m,n)}\left(p-\alpha_i\right)^+\leq r,\ \sum_{i=1}^{\min(m,n)}\left(p-\widehat{\widehat{\alpha}}_i\right)^+\geq r+\epsilon\right)\doteq 0,$$

where α_i and $\widehat{\widehat{\alpha}}_i$ are the negative exponents of the eigenvalues of $HH^\dagger$ and $H_2H_2^\dagger$, and where H_2 is the non-power controlled channel estimate at the receiver and p is the SNR exponent in $P\left(\widehat{H}_1\right)$. Further, $\widehat{q}_2 = 1$ because $\alpha_m < 1 + W_2\left(r+\epsilon\right) - \frac{\epsilon}{mn}$ makes the received power $\dot{>}\mathsf{SNR}^{\epsilon/mn}$. As $\widehat{q}_2 = 1$, using power $P\left(\widehat{H}_2\right)$ does not produce any outage when $\alpha_m < 1 + W_2\left(r+\epsilon\right) - \frac{\epsilon}{mn}$.

3. For $u \in [3, K-1]$, $\det\left(I + HH^\dagger P_1\left(\widehat{H}_1\right)\right) < \mathsf{SNR}^r$, $1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn} \leq \alpha_m < 1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$. As before, $q_2 = 1$. Now, since $\alpha_m \geq 1 + W_2\left(r+\epsilon\right) - \frac{\epsilon}{mn}$, the transmitter will receive this index as 0 which the receiver would know. Hence, $q_3 = 1$. This continues till $q_u = 1$. However, $\widehat{q}_u = 1$ since $\alpha_m < 1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$. Moreover, sending at power $P_u\left(\widehat{H}_u\right)$ when $\alpha_m < 1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$ does not produce any outage.

Hence, the outage happens only when $\det\left(I + HH^\dagger P_1\left(\widehat{H}_1\right)\right) < \mathsf{SNR}^r$ and $\alpha_m \geq 1 + W_{u-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$, whose probability is less than the probability of $\alpha_m \geq 1 + W_{K-1}\left(r+\epsilon\right) - \frac{\epsilon}{mn}$ which happens with probability $\mathsf{SNR}^{-W_K(r+\epsilon)}$. For ϵ very small, the diversity multiplexing tradeoff as given in the statement of the Theorem can be achieved.

Chapter 3

Placement of a Relay with Uncertainty in Destination Location

In this chapter, we consider the design problem of maximizing *coverage* for a fixed desired transmission rate. This would lead to maximize the possibility that the destination is covered. In Sec. 3.2 we compare the coverage regions of the CF and DF achievable strategies, for different locations of the relay. In this comparison, we assume that the relay's location is the same with both strategies. We extend the comparison in Sec 3.3, and allow each strategy its own preferred relay location and find that when the power of relay is large enough, DF is superior. In Sec. 3.4 we provide bounds on the area (measured in normalized units of area) of the coverage region of DF, as a function of the relay's location which leads us to study optimal relay location in some cases.

3.1 Background and Definitions

3.1.1 The Channel Model

Our model is depicted in Fig. 3.1. The channel consists of three nodes: a source (node 1), a relay (node 2) and a destination (node 3). We consider a two-dimensional domain for our

three-node network. This means that the source, relay and destination are associated with two-dimensional location vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, respectively. For simplicity, and without loss of generality, we may assume that $\mathbf{a}_1 = (0, 0)$, and $\mathbf{a}_2 = (d, 0)$ where $d > 0$ is the distance between the source and the relay. The relations between the channel outputs and

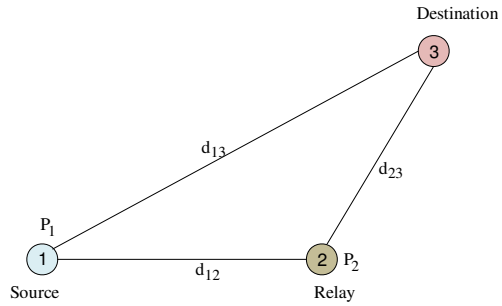

Figure 3.1: Single relay network

inputs are a function of the distances between the various nodes. We let $d_{kl}, k, l = 1, 2, 3$ denote the distances between nodes k and l. With this notation, $d_{12} = \|\mathbf{a}_2\| = d, d_{13} = \|\mathbf{a}_3\|$ and $d_{23} = \|\mathbf{a}_3 - \mathbf{a}_2\|$ where $\|\cdot\|$ denotes the Euclidean norm. See Fig. 3.1. The channel equations are now given by,

$$\begin{aligned} y_2[i] &= \tfrac{1}{d_{12}^{\alpha/2}} x_1[i] + z_1[i] \\ y_3[i] &= \tfrac{1}{d_{13}^{\alpha/2}} x_1[i] + \tfrac{1}{d_{23}^{\alpha/2}} x_2[i] + z_2[i] \end{aligned} , \tag{3.1}$$

where $x_1[i]$ and $x_2[i]$ are the signals transmitted from the source and relay, respectively, at time i. These signals are subject to average power constraints P_1 and P_2, respectively. The quantities $y_2[i]$ and $y_3[i]$ denote the observed signals at the relay and destination, respectively. The quantities $z_1[i]$ and $z_2[i]$ are mutually independent i.i.d circularly-symmetric complex Gaussian noise processes with variance 1. $\alpha \geq 2$ is the path loss exponent.

3.1.2 Codes and Achievable Strategies

A code for the relay channel of rate R and block-length n, consists of a pair $(C, \{f_i\}_{i=1}^n)$. C is a set of 2^{nR} codewords of length n. The source encoder constructs its signal by selecting a codeword $\mathbf{x}_1 \in C$. At time i, the source sends index $x_1[i]$ and the relay sends $x_2[i]$ using f_i based on its past observation. That is, $x_2[i] = f_i(y_2[i-1], ..., y_2[1])$.

A relay transmission *strategy* $\mathcal{S}$ is formally a collection of relay codes.

Definition 1. *Given locations* $\mathbf{a}_2 = (d, 0)$ *and* $\mathbf{a}_3$ *of the relay and the destination respectively, a rate* R *is defined to be achievable by a strategy* $\mathcal{S}$ *(equivalently:* $\mathcal{S}$ *supports* R*) if for any* $\epsilon > 0$*, there exist* $(C, \{f_i\}_{i=1}^n) \in \mathcal{S}$ *such that the rate of* C *is at least* R*, and the probability of error, under maximum-likelihood decoding is at most* ϵ*.*

We define the capacity of $\mathcal{S}$ at relay location $\mathbf{a}_2 = (d, 0)$ and destination location $\mathbf{a}_3$ as,

$$C_{\mathcal{S}}(d, \mathbf{a}_3) = \sup\{R \,:\, \mathcal{S} \text{ supports } R\}.$$

Cover and El Gamal [26] introduced two achievable coding strategies which were subsequently named decode-and-forward (DF) and compress-and-forward (CF). With DF, the relay decodes the message transmitted by the source. It then cooperates with the source to transmit the message to the destination. With CF, the relay considers the observed signal from the source as a raw signal, compresses it, and transmits it to the destination. The destination then combines this observation with its own observation, and uses both to decode the source's message.

An important distinction between the two strategies, is that with DF, the relay attempts to decode the source's message, while with CF it does not. A comprehensive description of the strategies is available e.g. in [26, 31, 38].

The achievable rates with both strategies, for our channel model, were computed in [31, 38], and are provided in Appendix 3.A. Following their example, we confine our attention to CF when the random variables used in the generation of the codebooks are Gaussian.

3.1.3 Coverage

We are now ready to formally define the concept of coverage.

Definition 2. *Let $R > 0$ be a desired transmission rate. For a fixed distance d between the source and the relay, and a fixed transmission strategy $\mathcal{S}$, we define the coverage region as,*

$$\mathcal{G}_{\mathcal{S}}(d) \triangleq \{\mathbf{a}_3 \ : \ C_{\mathcal{S}}(d, \mathbf{a}_3) \geq R\}.$$

The concept of coverage is closely related to *outage* - we fix a target rate, and seek to maximize the geographic region, outside which an outage occurs.

3.2 Comparison of CF and DF

In this section, we consider the coverage region when using the DF and CF approaches, for a fixed, given location of the relay $\mathbf{a}_2$. For reference, we also consider the *no-relay* (NR) coverage region, i.e, the coverage region when the relay is not used.

We would expect different locations of the destination $\mathbf{a}_3$ to favor different strategies. That is, for a fixed rate R, some locations would be covered only by CF, others only by DF and yet others by both. Surprisingly, however, the following result indicates a monotonic ordering of the coverage regions.

Theorem 4. *Assume the channel model of Sec. 3.1 and let $R > 0$. Let d_c be defined as follows,*

$$d_c = \left(\frac{P_1}{2^R - 1}\right)^{1/\alpha}. \tag{3.2}$$

1. *If $d \leq d_c$, then*

$$\mathcal{G}_{\mathrm{DF}}(d) \supseteq \mathcal{G}_{\mathrm{CF}}(d) \supseteq \mathcal{G}_{\mathrm{NR}}(d).$$

2. *If $d > d_c$, then*

$$\mathcal{G}_{\mathrm{CF}}(d) \supseteq \mathcal{G}_{\mathrm{NR}}(d) \supseteq \mathcal{G}_{\mathrm{DF}}(d) = \varnothing.$$

The proof is provided in Appendix 3.B.

We now interpret the results of Theorem 4. The condition $d \leq d_c$ determines whether the relay is still able to decode the information that is transmitted to it by the source[1]. Whenever the relay is able to decode, we see that DF is the method of choice. DF is uniformly superior in transmission to any point, in the sense that it provides coverage at any point served by CF. However, if the relay is not able to decode the data, then DF cannot be applied, and $\mathcal{G}_{\mathrm{DF}}(d) = \varnothing$. For such values of d, the best approach is CF, where the relay and destination combine their channel observations and perform collaborative decoding.

From a design perspective, while for $d \leq d_c$ DF enjoys a larger coverage region than CF, it suffers from a sharp drop in performance when d crosses d_c. In practical settings, when the path loss exponent α is not known (and consequently d_c, as defined by (3.2), is not known), this may become an important disadvantage. CF, in comparison, enjoys a more graceful degradation with respect to changes in d.

The above results can equivalently be stated as follows: Consider the combined transmission strategy $\mathrm{CF} \vee \mathrm{DF}$, under which the various terminals are free to select the best of CF and DF. Theorem 4 implies that,

$$\mathcal{G}_{\mathrm{CF}\vee\mathrm{DF}}(d) = \begin{cases} \mathcal{G}_{\mathrm{DF}}(d), & d \leq d_c \\ \mathcal{G}_{\mathrm{CF}}(d), & d > d_c \end{cases}.$$

Figs. 3.2 and 3.3 present numerical examples of the regions considered in Theorem 4. In Fig. 3.2, $d \leq d_c$ and in Fig. 3.3, $d > d_c$. In both figures we also compare the coverage regions with a region computed according to the upper-bound (UB) as provided by [38].

[1]More precisely, if the transmitter dedicated its power to communication with the relay, then d_c is the maximal distance at which communication at rate of R would still be possible.

Note that in neither of the figures $\mathcal{G}_{\mathrm{CF}\vee\mathrm{DF}}(d)$ equals $\mathcal{G}_{\mathrm{UB}}(d)$.

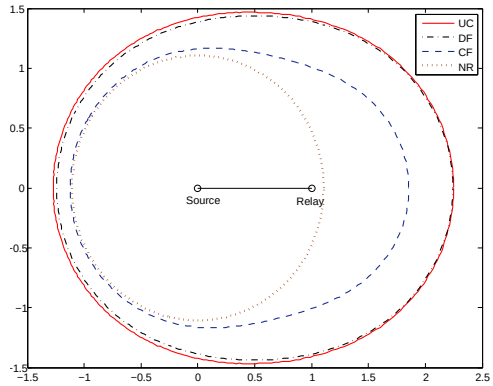

Figure 3.2: Coverage regions when $d \leq d_c, P_1 = 10, P_2 = 10, \alpha = 3.52, R = 3$

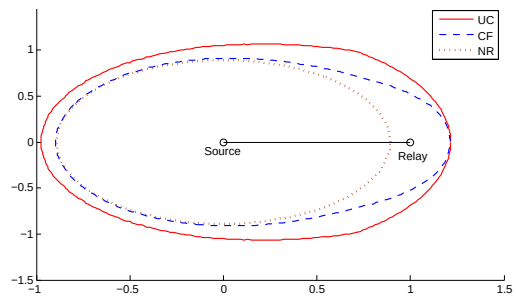

Figure 3.3: Coverage regions when $d > d_c, P_1 = 10, P_2 = 10, \alpha = 3.52, R = 4$

3.3 Comparison with Unequal Relay Placement

The analysis of Sec. 3.2 focused on coverage regions for a given distance d between the source and the relay. However, different strategies may favor different locations of the relay. Thus, in this section, we provide some interesting results that consider CF and DF when they are allowed *different* locations of the relay.

Theorem 5. *Let $R > 0$. Let d_c be defined as in Theorem 4. Then the following assertions holds:*

1. *Assume $\alpha = 2$. Then there exists a non-negative positive fraction $0 < \gamma < 1/9$, independent of R, such that if $P_2 > \gamma \cdot P_1$, then for all $d > d_c$,*

$$\mathcal{G}_{\mathrm{CF}}(d) \subseteq \mathcal{G}_{\mathrm{DF}}(d_c).$$

2. *There exists $\beta(R) > 0$ such that if $P_2 < \beta(R) \cdot P_1$, then there exists d_0 such that,*

$$\mathcal{G}_{\mathrm{CF}}(d_0) \nsubseteq \bigcup_{d>0} \mathcal{G}_{\mathrm{DF}}(d).$$

The proof of this theorem is provided in Appendix 3.C. We now provide some intuition for these results. We begin with Part 2 of the theorem. Its proof follows from the observation that regardless of how low the power at the relay may be, if the destination is close enough to the relay, then arbitrarily high reliability may be achieved in the link between them. With CF, the relay and the destination cooperate and form a virtual antenna array. With DF, the relay must decode alone, and is thus confined to the no-relay coverage region (i.e., d must not exceed d_c). If its power is sufficiently low, it cannot service a destination that is beyond the no-relay region. Turning to Part 1, we observe that with both CF and DF, as P_2 increases, the coverage regions naturally stretch around the relay. However, with DF, the relay transmits a higher-quality signal (having stripped off the noise) than with CF. Thus, the region stretches more rapidly. Part 1 now follows by extensive arithmetic.

Theorem 5, combined with Theorem 4, provide us with insight into the choice between strategies DF and CF when the relay location may be optimized independently for each strategy. In the case of $\alpha = 2$, when the relay power is sufficiently large, DF is the method of choice regardless of the desired R. If we choose to place the relay at location $d \leq d_c$, then Theorem 4 tells us that $\mathcal{G}_{\mathrm{CF}}(d) \subseteq \mathcal{G}_{\mathrm{DF}}(d)$, and thus DF is superior. If $d > d_c$, then Theorem 5 tells us that $\mathcal{G}_{\mathrm{CF}}(d) \subseteq \mathcal{G}_{\mathrm{DF}}(d_c)$, and thus switching from CF and DF and repositioning the relay at $d = d_c$ would render superior performance.

Regardless of α, if the power at the relay is sufficiently *low*, Theorem 5 tells us that there are locations and rates that may only be supported by CF. These cannot be supported by DF, regardless of where we place the relay.

3.4 Bounds on the DF Coverage Area

So far, our discussion has focused on a comparison of coverage regions with DF and CF. However, a natural question that arises is what the *area* (in normalized units of area) of the coverage region is, and the effect of the distance d on it. In this section we partially answer

this question for the DF coverage region in a few specific cases. We let $|\mathcal{G}_{\mathrm{DF}}(d)|$ denote this area, and begin with the following theorem.

Theorem 6. *Assume $P_1 = P_2$ and $\alpha = 2$. Then for all $0 < d \leq d_c$,*

$$\pi\sqrt{\lambda\gamma} \cdot d^2 \leq |\mathcal{G}_{\mathrm{DF}}(d)| \leq \pi\sqrt{\lambda\gamma} \cdot \frac{1 - a/2}{\sqrt{1-a}} \cdot d^2, \tag{3.3}$$

where

$$\rho = \sqrt{1 - \left(\frac{d}{d_c}\right)^2},, \tag{3.4}$$

$$\lambda = \frac{1}{4} + \frac{1}{1-\rho} + \frac{\sqrt{2(1+\rho)}}{1-\rho^2}, \tag{3.5}$$

$$\gamma = \frac{2}{1-\rho} - \frac{1}{4}, \textit{and} \tag{3.6}$$

$$a = 1 - \frac{\gamma}{\lambda}. \tag{3.7}$$

Remark 5.

1. *Recall, from Theorem 6, that when $d > d_c$, $\mathcal{G}_{\mathrm{DF}}(d) = \varnothing$ and thus $|\mathcal{G}_{\mathrm{DF}}(d)| = 0$.*

2. *Observe that ρ is is a function of d, and implicitly a function of of R and P_1 through its dependence on d_c, which was defined by* (3.2)*. Consequently, the other parameters are functions of these values too.*

The proof of this theorem is achieved by bounding the coverage region from within by an ellipse whose boundary is given by the points $(\frac{d}{2} + \sqrt{\lambda}d\cos(\theta), \sqrt{\gamma}d\sin(\theta))$, and from outside by a conic whose boundary is given by the points $(\frac{d}{2} + \sqrt{\lambda}d\cos(\theta)\sqrt{1 - a\sin^2(\theta)}$, $\sqrt{\gamma}d\sin(\theta)\frac{\sqrt{1-a\sin^2(\theta)}}{\sqrt{1-a}})$ (the detailed proof is omitted in the thesis). These two bounding shapes, for the case of $P_1 = P_2 = 10, R = 1$, are plotted in Fig. 3.4, along with the true region (computed numerically).

Fig. 3.5 presents the bounds in (3.3), along with the true area (computed numerically) corresponding to the same parameters as Fig. 3.4. Examining this figure, we see that the

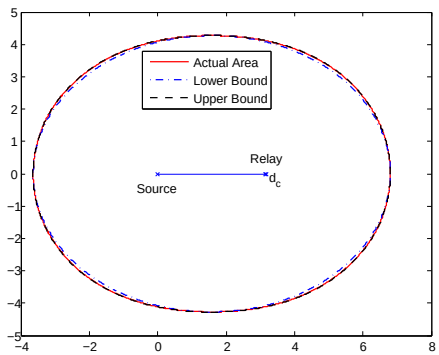

Figure 3.4: Bounds for $\mathcal{G}_{\mathrm{DF}}$, $P_1 = P_2 = 10$, $R = 1$, $\alpha = 2$

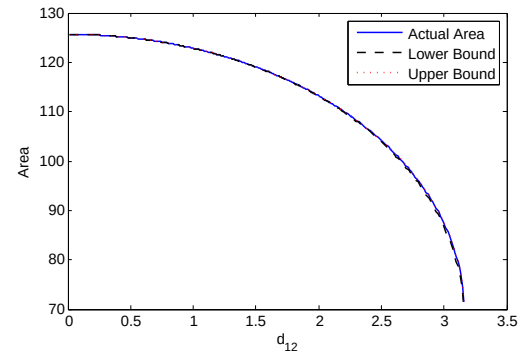

Figure 3.5: Bounds for $\mathcal{G}_{\mathrm{DF}}$, $P_1 = P_2 = 10$, $R = 1$, $\alpha = 2$

bounds are very tight. By (3.3), the ratio between the two bounds is $(1 - a/2)/(\sqrt{1-a})$, which increases with d from 1 to $\frac{6+2\sqrt{2}}{\sqrt{7(5+4\sqrt{2})}} \approx 1.02216$. Thus, the gap between the two bounds never exceeds 2.22%.

Examining Fig. 3.5, we may observe that $|\mathcal{G}_{\mathrm{DF}}(d)|$ approaches its maximum as $d \to 0$. To see this, observe that both bounds coincide as $d \to 0$, and the upper bound decreases with d. Thus, from the point of view of maximizing coverage, the relay should optimally be placed as close to the base station as possible. While this results holds for $\alpha = 2$, it is not generally true, as the discussion below will show.

The following theorem extends the *lower* bound of (3.3) to the case of $\alpha = 4$.

Theorem 7. *Assume $P_1 = P_2$ and $\alpha = 4$. Then for all $0 < d \leq d_c$,*

$$|\mathcal{G}_{\mathrm{DF}}(d)| \geq \pi\sqrt{\lambda\gamma} \cdot d^2, \tag{3.8}$$

where

$$\rho = \sqrt{1 - \left(\frac{d}{d_c}\right)^4}, \tag{3.9}$$

$$\gamma = \sqrt{\frac{2}{1-\rho}} - 1/4, \tag{3.10}$$

and λ is the largest real-valued solution of the equation,

$$\left(x-\frac{1}{4}\right)^4-\frac{2}{1-\rho}\left(x-\frac{1}{4}\right)^2-\frac{4}{1-\rho^2}\left(x-\frac{1}{4}\right)-\frac{1}{1-\rho^2}=0. \qquad (3.11)$$

This solution can be found analytically by applying the Ferrari method, see e.g. [54, page 32].

The proof of this theorem follows by bounding the coverage area from within by an ellipse like in Theorem 6 and is omitted.

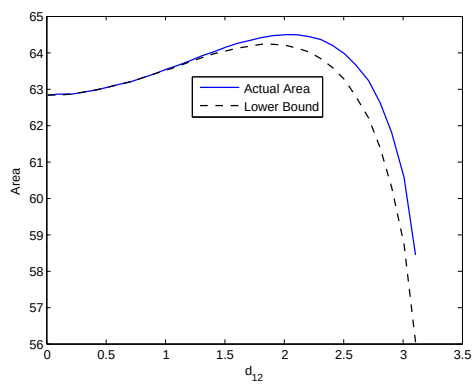

Figure 3.6: Lower bound for $\mathcal{G}_{\mathrm{DF}}$, $P_1 = P_2 = 100$, $R = 1$, $\alpha = 4$

Fig. 3.6 compares the lower bound of Theorem 7 with the true area (computed numerically), when $P_1 = P_2 = 100$ and $R = 1$. It can be shown that the lower bound (3.8) becomes tight as $d \to 0$.

Recall that with $\alpha = 2$, we proved that the maximum coverage area was achieved when $d \to 0$. However, this is *not* the case with $\alpha = 4$. To prove this, all we need to do is find a nonzero value of d at which the lower bound is greater than its value at $d = 0$. Changing variables from d to ρ, we obtain that at $\rho = .93$,

$$\begin{aligned} |\mathcal{G}_{\mathrm{DF}}|_{\rho=.93} &\geq \pi\sqrt{\lambda\gamma}d^2_{\rho=.93} = 2.0441..\pi d_c^2 \\ &> 2\pi d_c^2 = |\mathcal{G}_{\mathrm{DF}}(d=0)|. \end{aligned}$$

3.A General Expressions and Notation

3.A.1 DF and CF Achievable Rates

The achievable rates, with DF and CF, that were computed by [31, 38] are

$$\begin{aligned} C_{\mathrm{DF}} &= \max_{0\le\rho\le 1} \min\left\{\log\left(1+\frac{P_1}{d_{12}^{\alpha}}(1-\rho^2)\right),\right. \\ &\qquad \left.\log\left(1+\frac{P_1}{d_{13}^{\alpha}}+\frac{P_2}{d_{23}^{\alpha}}+\frac{2\rho\sqrt{P_1P_2}}{d_{13}^{\alpha/2}d_{23}^{\alpha/2}}\right)\right\}, \qquad (3.12) \\ C_{\mathrm{CF}} &= \log\left(1+\frac{P_1}{d_{12}^{\alpha}(1+\hat{N}_2)}+\frac{P_1}{d_{13}^{\alpha}}\right), \qquad (3.13) \end{aligned}$$

where $\hat{N}_2$ is given by,

$$\hat{N}_2 = \frac{P_1(1/d_{12}^{\alpha}+1/d_{13}^{\alpha})+1}{P_2/d_{23}^{\alpha}}. \qquad (3.14)$$

3.A.2 Switch to Polar Coordinates

We frequently use the polar coordinates x and θ to parameterize the destination's location, i.e. we assume the destination is placed at,

$$\mathbf{a}_3 = (x\cos\theta, x\sin\theta). \qquad (3.15)$$

Letting d denote the distance between the source and the relay, the distances d_{12}, d_{13} and d_{23} (see Sec. 3.1.1) now satisfy, $d_{12} = d$, $d_{13} = x$, and $d_{23}^2 = d^2 + x^2 - 2dx\cos\theta$.

3.B Proof of Theorem 4

We begin by noting that for all $d > 0$, the result $\mathcal{G}_{\mathrm{CF}}(d) \supseteq \mathcal{G}_{\mathrm{NR}}(d)$ is straightforward from (3.13) and from the observation that,

$$C_{\mathrm{NR}} = \log\left(1 + \frac{P_1}{d_{13}^{\alpha}}\right). \tag{3.16}$$

We now prove Part 2 of the theorem, which is the easier of the two parts. Whenever $d > d_c$, by (3.2),

$$R > \log\left(1 + \frac{P_1}{d^{\alpha}}\right) = \max_{\rho \in [0,1]} \log\left(1 + \frac{P_1}{d^{\alpha}}(1-\rho^2)\right) \geqslant C_{\mathrm{DF}},$$

where the last inequality was obtained by (3.12), recalling that $d_{12} = d$. This is true regardless of the destination location $\mathbf{a}_3$, and thus DF cannot support a rate of R anywhere and $\mathcal{G}_{\mathrm{DF}}(d) = \phi$.

Before proceeding to the proof of Part 1 of the theorem, we introduce the following notation. For a given strategy $\mathcal{S}$ ($\mathcal{S}$ being DF or CF), and a given θ, we define $x_{\mathcal{S}}(\theta)$ as the maximum x such that a destination with polar coordinates (x, θ) (see Appendix 3.A) is contained in $\mathcal{G}_{\mathcal{S}}(d)$. The proof now focuses on showing that $x_{\mathrm{DF}}(\theta) \geq x_{\mathrm{CF}}(\theta)$ for all θ.

By (3.12), we may obtain a lower bound on C_{DF} by restricting the maximization to $\rho = 0$. Thus,

$$C_{\mathrm{DF}} \geq \min\left\{\log\left(1 + \frac{P_1}{d_{12}^{\alpha}}\right), \log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}}\right)\right\}. \tag{3.17}$$

The following lemma examines this expression at $x = x_{\mathrm{DF}}(\theta)$.

Lemma 5. *For all θ, consider the point $\mathbf{a}_3$ whose polar coordinates are given by $(\theta, x_{\mathrm{DF}}(\theta))$.*

Then the following holds at $\mathbf{a}_3$*,*

$$\frac{P_1}{d_{12}^{\alpha}} \geqslant \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}}. \tag{3.18}$$

Proof. Assume, by contradiction, that (3.18) does not hold. Letting $x = x_{\mathrm{DF}}(\theta)$, we will now show that we may increase x and preserve $C_{\mathrm{DF}} \geq R$, contradicting the maximality of $x_{\mathrm{DF}}(\theta)$.

Consider (3.12). Whenever (3.18) does not hold, we have, for all ρ,

$$\begin{aligned}\log\left(1 + \frac{P_1}{d_{12}^{\alpha}}(1-\rho^2)\right) &\leq \log\left(1 + \frac{P_1}{d_{12}^{\alpha}}\right) \\ &< \log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}} + \frac{2\rho\sqrt{P_1 P_2}}{d_{13}^{\alpha/2} d_{23}^{\alpha/2}}\right).\end{aligned}$$

Thus, the minimization in (3.12) is achieved by the first term. Changing x does not affect this term. Increasing x affects d_{13} and d_{23} and thus reduces the second term. However, by a continuity argument, we may increase x slightly without altering the invalidity of (3.12). Thus, the minimization in (3.12) would still be achieved by the first term, and C_{DF} will not change. Therefore, if $C_{\mathrm{DF}} \geq R$ at $x = x_{\mathrm{DF}}(\theta)$, it will remain so after we increase x a little, producing the desired contradiction. □

By Lemma 5, (3.17) implies that at $x = x_{\mathrm{DF}}(\theta)$,

$$C_{\mathrm{DF}} \;\geq\; \log\left(1 + \frac{P_1}{d_{13}^{\alpha}} + \frac{P_2}{d_{23}^{\alpha}}\right) > C_{\mathrm{CF}}. \tag{3.19}$$

The last inequality may be obtained by simple arithmetic using (3.13) and (3.14). By a continuity argument, $C_{\mathrm{DF}} = R$ at $(\theta, x_{\mathrm{DF}}(\theta))$, and thus the above inequality implies that $C_{\mathrm{CF}} < R$ at this point.

We now argue that this result implies that $\mathcal{G}_{\mathrm{DF}}(d) \supseteq \mathcal{G}_{\mathrm{CF}}(d)$. This would follow if we could show that C_{CF} decreases in the range $x > x_{\mathrm{DF}}(\theta)$, for all θ.

In the range $x > d$, d_{13}, d_{23} can easily be shown to be increasing functions. This implies that $\hat{N}_2$ increases (see (3.14)), and consequently C_{CF} decreases (see (3.13)). Thus, our desired result would now follow if we could prove that $x_{\mathrm{DF}}(\theta) > d$ for all θ. To see this, observe that by (3.16) and (3.2), $\mathcal{G}_{\mathrm{NR}}$ is a circle with radius d_c. We have shown that $\mathcal{G}_{\mathrm{NR}} \subset \mathcal{G}_{\mathrm{CF}}$. Thus, if $C_{\mathrm{CF}} < R$ at $(\theta, x_{\mathrm{DF}}(\theta))$ (as we have shown above), then $x_{\mathrm{DF}}(\theta) \geq d_c \geq d$ (the last inequality being one of the conditions of this part of the theorem). This is precisely the result we sought, thus completing the proof of Part 1 of the theorem.

3.C Proof of Theorem 5

Before beginning with the proof, we argue that we may assume, without loss of generality, that P_1 (the power of the source) is 1. To see this, observe that if $P_1 \neq 1$, we may replace P_1 and P_2 by $\hat{P}_1 = 1$ and $\hat{P}_2 = P_2/P_1$. This substitution preserves the ratio $\hat{P}_1/\hat{P}_2 = P_1/P_2$, and has the effect of scaling the DF and CF coverage regions. That is, if we relocate the relay and the destination from any locations $\mathbf{a}_2$ and $\mathbf{a}_3$ (respectively) to $1/P_1^{1/\alpha} \cdot \mathbf{a}_2$ and $1/P_1^{1/\alpha} \cdot \mathbf{a}_3$, the achievable rates C_{CF} and C_{DF} would remain unchanged.

Under these assumptions, given that the relay is placed at $(d, 0)$, and using the notation of Sec. 3.A, the expressions for the DF and CF achievable rates (3.12) and (3.13) may be rewritten as,

$$\begin{aligned} C_{\mathrm{DF}} &= \max_{0 \leq \rho \leq 1} \min \left\{ \log\left(1 + \frac{1}{d^\alpha}(1-\rho^2)\right), \right. \\ &\qquad \left. \log\left(1 + \frac{1}{x^\alpha} + \frac{P_2}{d_{23}^\alpha} + \frac{2\rho\sqrt{P_2}}{d^{\alpha/2} d_{23}^{\alpha/2}}\right) \right\}, \qquad (3.20) \\ C_{\mathrm{CF}} &= \log\left(1 + \frac{1}{d^\alpha(1+\hat{N}_2)} + \frac{1}{x^\alpha}\right), \qquad (3.21) \end{aligned}$$

and (3.14) by,

$$\hat{N}_2 = \frac{1/d^\alpha + 1/x^\alpha + 1}{P_2/d_{23}^\alpha}. \tag{3.22}$$

3.C.1 Proof of Part 1 of Theorem 5

In this section, we assume that $\alpha = 2$. Rather than focus on C_{DF} and C_{CF} as in (3.20) and (3.21), the following lemma allows us to revert to two alternative functions.

Lemma 6. *Let $C_{\mathrm{DF}}(x, \theta, d)$ and $C_{\mathrm{CF}}(x, \theta, d)$ denote the achievable rates when the relay is placed at $(d, 0)$ and the destination location is derived from x and θ. Let*

$$\tilde{C}_{\mathrm{DF}}(x, \theta) \triangleq \log\left(1 + \frac{1}{x^2} + \frac{P_2}{d_{23,c}^2}\right), \tag{3.23}$$

$$C_{\mathrm{CF}}^{+}(x, \theta) \triangleq \log\left(1 + \frac{1}{x^2} + \frac{P_2}{\inf\limits_{d \geqslant d_c}(P_2 d^2 + d_{23}^2)}\right), \tag{3.24}$$

where $d_{23,c}$ is the distance from a relay placed at $(d_c, 0)$ to the destination, and d_{23} is the same for a relay placed at $(d, 0)$. Then the following holds:

1. *$\tilde{C}_{\mathrm{DF}}(x, \theta) \geq R$ if and only if $C_{\mathrm{DF}}(x, \theta, d_c) \geq R$. Equivalently,*

$$\mathcal{G}_{\mathrm{DF}}(d_c) = \left\{(x\cos\theta, x\sin\theta) \;:\; \tilde{C}_{\mathrm{DF}}(x, \theta) \geq R\right\}.$$

2. *$C_{\mathrm{CF}}^{+}(x, \theta)$ upper bounds $C_{\mathrm{CF}}(x, \theta, d)$ for all $d \geq d_c$, and therefore,*

$$\bigcup_{d \geq d_c} \mathcal{G}_{\mathrm{CF}}(d) \subset \left\{(x\cos\theta, x\sin\theta) \;:\; C_{\mathrm{CF}}^{+}(x, \theta) \geq R\right\}.$$

Proof. Part 1 follows from the observation that when $d = d_c$, by (3.2), any choice of ρ in (3.20) other than $\rho = 0$ would render $C_{\mathrm{DF}} < R$.

Part 2 follows by bounding the term $d^2(1+\hat{N}_2)$, which appears in (3.21), as

$$d^2(1+\hat{N}_2) = \frac{d^2(1/d^2 + 1/x^\alpha + 1 + P_2/d_{23}^2)}{P_2/d_{23}^2} \geq \frac{d^2 d_{23}^2(1/d^2 + P_2/d_{23}^2)}{P_2} = \frac{d_{23}^2 + P_2 d^2}{P_2}.$$

□

Our aim now is to show that wherever $C_{\mathrm{CF}}^{+}(x,\theta) \geq R$, also $\tilde{C}_{\mathrm{DF}}(x,\theta) \geq R$. In the sequel, we use the shorthand notation $s \triangleq \cos\theta$.

Lemma 7. *Let θ be fixed,*

1. *If $s \leq 0$, then for all $x > 0$, $\tilde{C}_{\mathrm{DF}}(x,\theta) \geq C_{\mathrm{CF}}^{+}(x,\theta)$.*

2. *Assume $s > 0$ and let x_1 be given by,*

$$x_1 = \frac{1}{s}\frac{\sqrt{P_2+1}}{\sqrt{P_2+1}-\sqrt{P_2}} \cdot d_c. \tag{3.25}$$

Then, $C_{\mathrm{CF}}^{+}(x,\theta) \leq \tilde{C}_{\mathrm{DF}}\left(\min(x,x_1),\theta\right)$.

Proof. Observe, by (3.23) and (3.24) that $\tilde{C}_{\mathrm{DF}}(x,\theta) \geq C_{\mathrm{CF}}^{+}(x,\theta)$ if and only if $f_{\min} \geq d_{23,c^2}$ where $f_{\min}$ is given by, $f_{\min} = \inf_{d \geq d_c}(P_2 d^2 + d_{23}^2)$. Let $f(d)$ be defined by,

$$f(d) \quad \triangleq \quad P_2 d^2 + d_{23}^2 = (P_2+1)d^2 + x^2 - 2dxs, \tag{3.26}$$

where the last equality was obtained by the cosine theorem. Taking the derivative of $f(d)$, we obtain that the unconstrained minimum (i.e., the minimum in the range $d \in (-\infty,\infty)$) is obtained at,

$$d_{\min} = \frac{xs}{P_2+1}. \tag{3.27}$$

If $s \leq 0$ then $d_{\min} \leq 0 < d_c$, and thus $f_{\min}$, which is the minimum in the range $d \geq d_c$, is

obtained at d_c. Thus,

$$f_{\min} = f(d_c) = P_2 d_c^2 + d_{23,c}^2 > d_{23,c}^2,$$

and our desired result of Part 1 of the lemma follows.

To prove Part 2, we begin by considering the range $x \in [0, x_1]$. We further divide the interval $x \in [0, x_1]$ into two overlapping intervals, $[0, x']$ and $[x'', x_1]$ and prove the desired result separately in each interval. We define,

$$\begin{aligned} x' &= \frac{1}{s} \cdot d_c, \\ x'' &= \frac{1}{s} \frac{\sqrt{P_2+1}}{\sqrt{P_2+1}+\sqrt{P_2}} \cdot d_c. \end{aligned}$$

Clearly, $x'' < x' < x_1$, and thus the two intervals overlap.

If $x \leq x'$ then $d_{\min} < d_c$ (by (3.27)), and thus $f_{\min} \geq d_{23,c}^2$ as in the proof of Part 1 of the lemma. Thus, the desired result is obtained for $x \in [0, x']$.

We proceed to consider the interval $x \in [x'', x_1]$. The minimum $f_{\min}$ (constrained to $d \geq d_c$) clearly satisfies $f_{\min} \geq f(d_{\min})$. By (3.26) and (3.27), we have, after some algebraic manipulations,

$$f(d_{\min}) = x^2 \left(1 - \frac{s^2}{P_2+1}\right). \tag{3.28}$$

Thus, a sufficient condition for $f_{\min} \geq d_{23,c}^2$ is,

$$x^2 \left(1 - \frac{s^2}{P_2+1}\right) \geq d_{23,c}^2. \tag{3.29}$$

Applying the cosine theorem and rearranging the terms, we obtain the inequality,

$$\frac{s^2}{P_2+1} \cdot x^2 - (2d_c s)x + d_c^2 \leq 0. \tag{3.30}$$

The left hand side of this inequality is a parabola in x, whose roots are x'' and x_1. Thus, the inequality is satisfied for all $x \in [x'', x_1]$. Consequently, by the above discussion, $f_{\min} \geq d_{23,c}^2$ in this interval, as desired. This completes the proof on the lemma in the interval $x \in [0, x_1]$.

Before proceeding to the interval $[x_1, \infty)$, observe that x_1 satisfies (3.30) and consequently (3.29), with equality. Equivalently,

$$x_1^2 \left(1 - \frac{s^2}{P_2 + 1}\right) = d_{23,c}^2. \tag{3.31}$$

We now consider $x \geq x_1$. Observe that the discussion leading to (3.28) remains valid in this range. By (3.24),

$$\begin{aligned} C_{\mathrm{CF}}^{+}(x,\theta) = \; & \log\left(1 + \tfrac{1}{x^2} + \tfrac{P_2}{\inf\limits_{d \geqslant d_c}(f(d,x,\theta)}\right) & \leq \log\left(1 + \frac{1}{x^2} + \frac{P_2}{\inf\limits_{d \geqslant 0} f(d,x,\theta)}\right) \\ \overset{(a)}{=} \; & \log\left(1 + \tfrac{1}{x^2} + \tfrac{P_2}{x^2\left(1 - \frac{s^2}{P_2+1}\right)}\right) & \overset{(b)}{\leq} \log\left(1 + \frac{1}{x_1^2} + \frac{P_2}{x_1^2\left(1 - \frac{s^2}{P_2+1}\right)}\right) \\ \overset{(c)}{=} \; & \log\left(1 + \tfrac{1}{x_1^2} + \tfrac{P_2}{d_{23,c}^2}\right) & = \tilde{C}_{\mathrm{DF}}(x_1,\theta), \end{aligned}$$

where (a) is obtained by (3.28), (b) follows from $x \geq x_1$, and (c) is obtained by (3.31). □

Lemmas 6 and 7 imply that a sufficient condition for Part 1 of Theorem 5 is $\tilde{C}_{\mathrm{DF}}(x_1, \theta) \leq R$ for all θ such that $s = \cos\theta > 0$ (recall that x_1 is a function of θ) and all $P_2 > 1/9$ (recall that we have assumed that $P_1 = 1$).

Using (3.2) and (3.23), $\tilde{C}_{\mathrm{DF}}(x_1, \theta) \leq R$ can be rewritten as,

$$\log\left(1 + \frac{P_2}{d_{23,c}^2} + \frac{1}{x_1^2}\right) \leq \log\left(1 + \frac{1}{d_c^2}\right). \tag{3.32}$$

Plugging (3.25) and (3.31) in (3.32), we get

$$P_2 \leq \left(1 - \frac{s^2}{P_2+1}\right)((\frac{1}{s}\frac{\sqrt{P_2+1}}{\sqrt{P_2+1}-\sqrt{P_2}})^2 - 1). \tag{3.33}$$

The right hand side of the above inequality is a descending function of s in the range $s \in [0,1]$. Thus, a necessary and sufficient condition for the inequality to be valid for all $s \in [0,1]$ is that the inequalities hold for $s = 1$. We thus confine our attention to this case, and rewrite (3.33) as,

$$P_2 \leq \left(1 - \frac{1}{P_2+1}\right)\left[\left(\frac{\sqrt{P_2+1}}{\sqrt{P_2+1}-\sqrt{P_2}}\right)^2 - 1\right],$$

which can be verified to be satisfied for $P_2 \geq 1/9$.

3.C.2 Proof of Part 2 of Theorem 5

An outline of the following proof was provided in Sec. 3.3. We confine our attention in this section to destinations $\mathbf{a}_2$ of the form $(0,x)$ where $x > d_c$. We begin by examining C_{DF}. By (3.20) (assuming, as discussed above, that $P_1 = 1$),

$$\begin{aligned}\limsup_{P_2\to 0} C_{\text{DF}} &\leq \lim_{P_2\to 0}\max_{0\leq\rho\leq 1}\log\left(1+\frac{1}{x^\alpha}+\frac{P_2}{d_{23}^\alpha}+\frac{2\rho\sqrt{P_2}}{x^{\alpha/2}d_{23}^{\alpha/2}}\right)\\ &= \log\left(1+\frac{1}{x^\alpha}\right) < \log\left(1+\frac{1}{d_c^\alpha}\right) = R.\end{aligned}$$

The last equality follows by (3.2). Thus, at any $x > d_c$, for P_2 that is too small, $C_{\text{DF}} < R$, and thus we may not support the desired rate of R.

We now turn to examine C_{CF}. We wish to find a destination location $\mathbf{a}_3 = (0,x)$, where $x > d_c$, such that regardless of how low the relay power P_2 may be, by positioning the relay close enough to the destination, we will be able to support the desired rate of R. Combining this with our above result for DF will conclude the proof.

Let $x > d_c$ be arbitrary. We will examine relay locations given by $(0, d)$, where $d = x - P_2^{1/\alpha}$. With this choice, $P_2/d_{23}^{\alpha} = 1$, and

$$\lim_{P_2 \to 0} C_{\text{CF}} = \lim_{P_2 \to 0} \log \left(1 + \frac{1}{d^{\alpha}(1 + \hat{N}_2)} + \frac{1}{x^{\alpha}} \right),$$

where we have used $d_{12} = d$, and $d_{13} = x$. Now, it is straightforward to verify that $\lim_{P_2 \to 0} d^{\alpha}(1 + \hat{N}_2) = 2(1 + x^{\alpha})$. Thus,

$$\lim_{x \to d_c} \left(\lim_{P_2 \to 0} C_{\text{CF}} \right) = \log \left(1 + \frac{1}{2(1 + d_c^{\alpha})} + \frac{1}{d_c^{\alpha}} \right) > \log \left(1 + \frac{1}{d_c^{\alpha}} \right) = R,$$

where the last equality follows by (3.2). Thus, there exists $x_0 > d_c$, such that for arbitrarily small P_2, the achievable rate with CF (when the relay is appropriately placed) is greater than R. This is exactly what we set out to prove, thus concluding the proof of the theorem.

Chapter 4

Uncertain Channel Gains in Interference Networks

In this chapter, we consider the case when the network connectivity is known to all the nodes in the network. However, the channel gains of the links are only known locally. We measure the local knowledge of the link gains via a metric called the number of hops and measure the performance of a strategy in terms of the worst case ratio of sum throughput achieved to the maximum sum throughput possible with global information. The best ratio achieved by any strategy is called normalized sum-capacity. We propose a strategy called Maximal Independent Graph (MIG) Scheduling, which achieves normalized sum-capacity in many scenarios. In section 4.1, we give the system and network model, and provide some definitions that will be used throughout the chapter. We will also consider the example of a Multiple Access Network to gain understanding. In Section 4.2, we define MIG scheduling and derive the independent graphs in the cases when the transmitters have 1 or 2 hops of channel gain information. In Section 4.3, we characterize the cases where maximal independent graph scheduling is optimal. In Section 4.4, we give an example where maximal independent graph scheduling is not optimal, and extend the achievable scheme with 1-hop knowledge at transmitters to coded maximal independent set scheduling.

4.1 Problem Formulation

In this section, we will first describe the system and network models. We will then define normalized sum-rate and normalized sum-capacity which will be used to evaluate the performance with asymmetric network information at the nodes. Finally, we will also formalize the specific notion of local view used in this chapter to model asymmetric network information.

4.1.1 System Model

As shown in Figure 4.1, consider a wireless network with K transmitters and K receivers. Each node in the network is either a transmitter or a receiver. For each transmitter k, let the message index m_k be encoded as X_k^n using the encoding functions $e_k(m_k|N_k, \mathsf{SI})$, which depend on the local view, N_k, and side information about the network, SI. Only receiver k is interested in message m_k. The message is decoded at the receiver k using the decoding function $d_k(Y_k^n|N_k', \mathsf{SI})$, where N_k' is the receiver local view and SI is the side information. A strategy is defined as the set of all encoding and decoding functions in the network, $\{e_k(m_k|N_k, \mathsf{SI}), d_k(Y_k^n|N_k', \mathsf{SI})\}$. We note that the local view at transmitter k and receiver k can be different, as will be the case in our subsequent development. The relationship between the transmit signals and the received signals is specified by the network model that is described in the next section.

4.1.2 Network Model

We will consider two models for interference networks. We use a deterministic model, which was proposed as an approximation to the Gaussian model in [18] to get insights and then proceed to Gaussian network model both of which are described as follows.

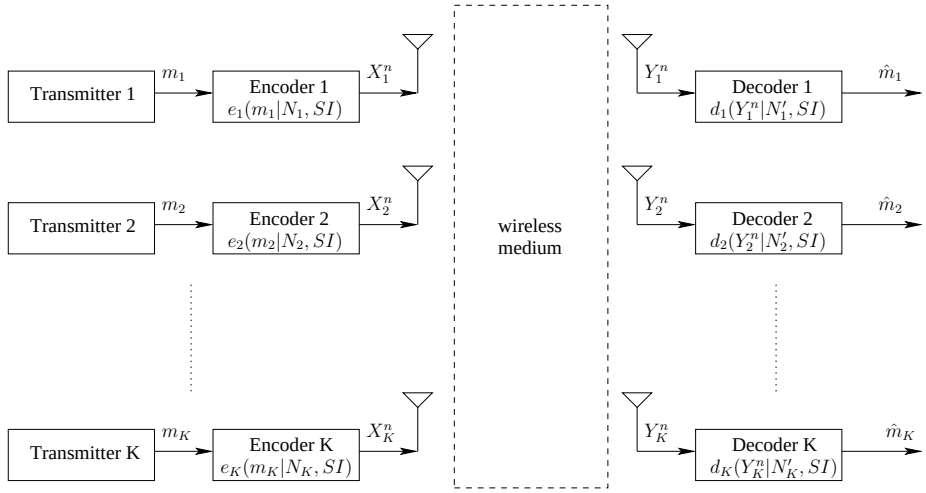

Figure 4.1: System-level depiction of the problem.

Deterministic Model

In a deterministic interference network, the input of the kth transmitter at time i can be written as $X_k[i] = \left[X_{k_1}[i]\ X_{k_2}[i]\ \ldots X_{k_q}[i]\right]^T$, $k = 1, 2, \cdots, K$, such that $X_{k_1}[i]$ and $X_{k_q}[i]$ are the most and the least significant bits, respectively. The received signal of user j, $j = 1, 2, \cdots, K$, at time i is denoted by the vector $Y_j[i] = \left[Y_{j_1}[i]\ Y_{j_2}[i]\ \ldots\ Y_{j_q}[i]\right]^T$. Associated with each transmitter k and receiver j is a non-negative integer n_{kj} that represents the gain of the channel between them. The maximum number of bits supported by any link is $q = \max_{k,j}(n_{kj})$. The received signal $Y_j[i]$ is given by

$$Y_j[i] \quad = \quad \sum_{k=1}^{K} \mathbf{S}^{q-n_{kj}} X_k[i], \tag{4.1}$$

where q is the maximum of the channel gains (*i.e.* $q = \max_{j,k}(n_{jk})$), the summation is in $\mathbb{F}_2^q$, and $\mathbf{S}^{q-n_{jk}}$ is a $q \times q$ shift matrix with entries $\mathbf{S}_{m,n}$ that are non-zero only for $(m, n) = (q - n_{jk} + n, n), n = 1, 2, \ldots, n_{jk}$. We will also use X_k^n, Y_k^n to denote $(X_{k1}, \cdots, X_{kn})$, $(Y_{k1}, \cdots, Y_{kn})$. The network can be represented by a square matrix H whose $(i, j)^{th}$ entry is $H_{ij} = n_{ij}$. We note that H need not be symmetric.

Gaussian Model

In a Gaussian interference network, the inputs of the kth transmitter at time i are denoted by $X_k[i] \in \mathbb{C}$, $k = 1, 2, \cdots, K$, and the outputs at j^{th} receiver in time i can be written as $Y_j[i] \in \mathbb{C}$, $j = 1, 2, \cdots, K$. The received signal $Y_j[i]$, $j = 1, 2, \cdots, K$ is given by

$$Y_j[i] \quad = \quad \sum_{k=1}^{K} h_{kj} X_k[i] + Z_j[i], \tag{4.2}$$

where $h_{kj} \in \mathbb{C}$ is the channel gain associated with each transmitter k and receiver j, and $Z_j[i]$ are additive white complex Gaussian random variables of unit variance. Much like the deterministic case, we will use X_k^n, Y_k^n to denote $(X_k[1], \cdots, X_k[n])$, $(Y_k[1], \cdots, Y_k[n])$. Further, the input $X_k[i]$ has an average power constraint of unity, i.e. $\mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} |X_k[i]|^2) \leq 1$, where $\mathbb{E}$ denotes the expectation of the random variable.

Like the deterministic case, we represent the network by a square matrix H whose $(i,j)^{th}$ entry is $H_{ij} = |h_{ij}|^2$ and can similarly define the set of network states. Thus we will use the matrix H for both the deterministic and the Gaussian model, where the usage will be clear from the context.

4.1.3 Normalized Sum-Capacity

As we discussed earlier, at each receiver k, the desired message m_k is decoded using the decoding function $d_k(Y_k^n | N_k', \mathsf{SI})$, where N_k' is the receiver local view of the network and SI is the side information. The corresponding probability of decoding error $\lambda_j(n)$ is defined as $\Pr[m_k \neq d_k(Y_k^n | N_k', \mathsf{SI})]$. A rate tuple $(R_1, R_2, \cdots, R_K)$ is said to be achievable if there exists a sequence of codes such that the error probabilities $\lambda_1(n), \cdots \lambda_K(n)$ go to zero as n goes to infinity for all network states consistent with the side information. The sum-capacity is the supremum of $\sum_i R_i$ over all possible encoding and decoding functions.

We will now define normalized sum-rate and normalized sum-capacity that will be used throughout the chapter. These notions represent the percentage of the global-view

sum-capacity that can be achieved with partial information about the network.

Definition 3. *Normalized sum-rate of α is said to be achievable for a set of network states with partial information if there exists a strategy such that following holds. The strategy yields a sequence of codes having rates R_i at the transmitter i such that the error probabilities at the receiver, $\lambda_1(n), \cdots \lambda_K(n)$, go to zero as n goes to infinity, satisfying*

$$\sum_i R_i \geq \alpha C_{sum} - \tau$$

for all the sets of network states consistent with the side information, and for a constant τ that is independent of the channel gains but may depend on the side information SI. *Here C_{sum} is the sum-capacity of the whole network with the full information.*

Definition 4. *Normalized sum-capacity, α^*, is defined as the supremum over all achievable normalized sum rates α.*

Note that $\alpha^* \in [0, 1]$. In [11], we defined the concept of *universal optimality* of a strategy. A universally optimal strategy is the one which achieves $\alpha^*(h) = 1$ for a given network. Thus, universal optimality is the special case where the distributed scheme achieves global-view sum-capacity in *all* network states and hence is universally optimal for all network states.

4.1.4 Local View Based on Hop Distance

We assume that that there is a direct link between each transmitter T_i and its intended receiver D_i. On the other hand, if a cross-link between transmitter i and receiver j does not exist, then $H_{ij} \equiv 0$. For large part, we will treat the network as a weighted undirected graph, $\mathsf{G} = (\mathsf{V}, \mathsf{E}, \mathsf{W})$, where transmitters and receivers are the vertices of the graph, $\mathsf{V} = \{\mathsf{T}_i, \mathsf{D}_i\}$, and an edge $e \in \mathsf{E}$ exists between any two nodes if they have a possibility of non-zero channel gain. In other words, if the channel gain between two nodes is identically

zero, there is no edge between them[1]. Finally, the actual channel gain n_{ij} (for deterministic model) or h_{ij} (for Gaussian model) is the edge weight $w(e) \in \mathsf{W}$. The resulting bipartite graph thus has $2K$ vertices and no more than K^2 edges.

We realize that the current formulation of distributed encoding is very general and encompasses a large class of $\{N_k, N'_k\}_k$ and SI. To make progress we will focus on a special structure of local view and side information at the nodes, which is largely inspired by common characteristics of existing network protocols. We will assume that the side information at all the nodes is the network connectivity characterized by (E, V). We identify (E, V) with the long time-scale characteristics of the network, which changes slowly. However, the network state captured by edge weights W, which gives the weights of edges is not part in the side information.

The local view at the nodes is defined using the metric of hop count (h). For any node, the links that are incident on the node have a distance of 1-hop. In general, hop-distance of a link from a node is one plus the minimum amount of links traversed starting from the node and terminating at the link. An example of the minimum distance of the links from a node is shown in Figure 4.2. We say that there is h-local view when all the transmitters know the weights (equivalently the channel gains) of those links which are at a distance of h-hops from them while the receivers know the weight of only those links which are at most distance of $h + 1$ hops from them. This definition of h-local information is based on our prior work in [11] where we proposed a multi-round protocol abstraction to show how different nodes have different amounts of network information. In the message-passing abstraction, it was convenient to have receivers know one more hop than their corresponding transmitters, which allowed coherent decoding.

Thus, we will consider the side information SI to be the network connectivity while the local information at each node is the h-local information. As h increases, the normalized

[1]The model is inspired by fading channels, where the existence of a link is based on its average channel gain. On the average the link gain may be above the noise floor but its instantaneous value can be below the noise floor.

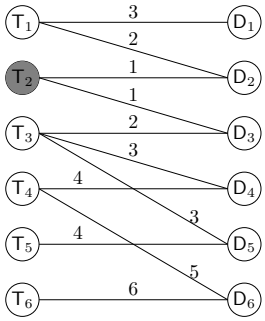

Figure 4.2: The hop-distances of each link from transmitter, T_2 (the dark circle), are labeled above each link.

sum-capacity increases. When h is the network diameter, which is the maximum hop distance between any link and any node, all the nodes have full network information. This is called the global view, since every node knows the complete network state, $\mathsf{G} = (\mathsf{V}, \mathsf{E}, \mathsf{W})$. In this setting, normalized sum-capacity $\alpha^* = 1$. When $h = 0$, none of the nodes know any weights and thus following compound channel arguments [23], $\alpha^* = 0$ since none of the nodes know any link weight and have to assume that all channel gains are zero.

4.1.5 A Warmup Example: Multiple Access Network

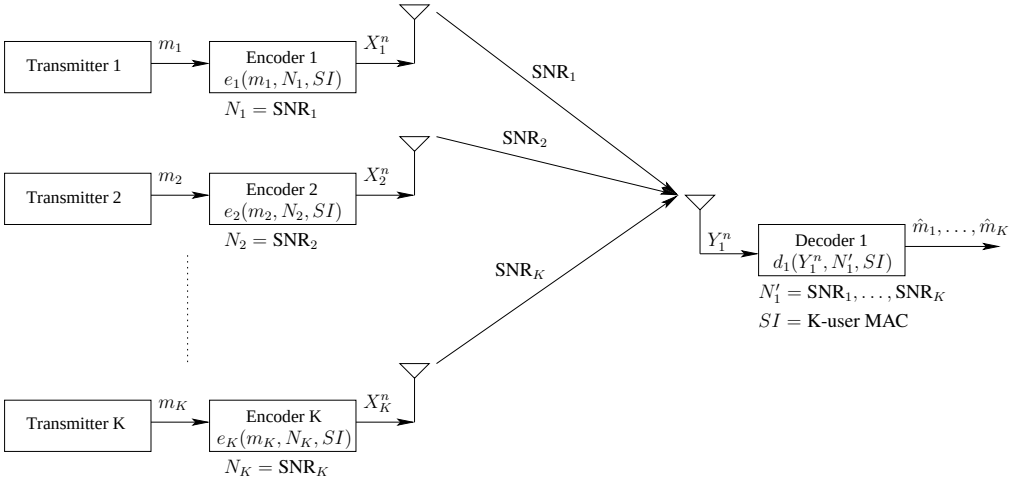

Figure 4.3: Example: multiple-access network with 1-hop local information.

We start with a simple example to illustrate these concepts. As shown in Figure 4.3,

we consider the K-user Gaussian multiple access network with the channel gain from i^{th} transmitter to the receiver being h_i such that $|h_i|^2 = \sqrt{\mathsf{SNR}_i}$ and the power constraint at each transmitter being unity. Note that the network diameter is two, which implies 2-local is equivalent to global view implying $\alpha^*(2) = 1$. Thus the interesting case is that of 1-local view.

We show that when there is 1-local view, the normalized sum-capacity is $1/K$ which can be achieved by simply scheduling one user at a time in a total of K time-slots. It can also be achieved by letting each user simultaneously send at $1/K$ fraction of its direct link capacity.

The main challenge is to show the converse. Let $K > 1$, as otherwise the result holds trivially. Assume that normalized sum-rate of $\alpha = (1/K + \epsilon)$ is achievable. Then, we should be able to achieve a rate tuple satisfying

$$R_i \geq \left(\frac{1}{K} + \epsilon\right) \log(1 + \mathsf{SNR}_i) - \tau, \ \forall\, 1 \leq i \leq K. \tag{4.3}$$

This is because each node is unaware of the other channel gains. To achieve a normalized sum-rate larger than α, each user should send at a rate larger than a fraction α of its channel capacity up to a difference τ (otherwise in the case when all other channel gains are zero, achievable normalized sum-rate is smaller than α). Now, we will show that this rate-tuple cannot be achieved. With the capacity bound of full information,

$$\begin{aligned} R_K &\leq \log\left(1 + \sum_{i=1}^{K} \mathsf{SNR}_i\right) - \sum_{i=1}^{K-1} R_i \\ &\overset{(4.3)}{\leq} \log\left(1 + \sum_{i=1}^{K} \mathsf{SNR}_i\right) - \left(\frac{1}{K} + \epsilon\right) \sum_{i=1}^{K-1} \log(1 + \mathsf{SNR}_i) + (K-1)\tau. \end{aligned}$$

Since the Kth transmitter does not know SNR_i for $1 \leq i \leq K-1$,

$$\begin{aligned} R_K &< \min_{\mathsf{SNR}_i, 1 \leq i \leq K-1} \left[\log\left(1 + \sum_{i=1}^{K} \mathsf{SNR}_i\right) \right. \\ &\qquad \left. - \left(\frac{1}{K} + \epsilon\right) \sum_{i=1}^{K-1} \log(1 + \mathsf{SNR}_i) + (K-1)\tau \right] \\ &\leq \frac{1}{K} \log(1 + \mathsf{SNR}_K) - (K-1)\epsilon \log(1 + \mathsf{SNR}_K) + \log(K) + (K-1)\tau \end{aligned} \tag{4.4}$$

For the above to hold, $(K-1)\epsilon \log(1 + \mathsf{SNR}_K) \leq \log(K) + (K-1)\tau$ which cannot hold for all SNR_K with τ and K independent of SNR_K. Thus, $\alpha^* \leq \frac{1}{K}$.

Since all the links are at-most two hops from each transmitter, the normalized sum-capacity in the case when each transmitter knows all the links that are at-most two hop distant from it is 1.

For the rest of the chapter, we will focus on interference networks, some examples of which will be defined in the next section.

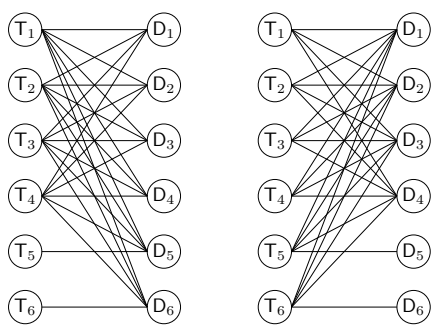

Figure 4.4: (a) 4-to-many interference network, and (b) many-to-4 interference network with 6 users.

4.1.6 Examples of Interference Networks

In this chapter, some special interference networks will be used as examples. They are defined as follows.

Definition 5. *A <u>d-to-many</u> interference network with K users is an interference network*

specified by $\mathsf{E} = \bigcup_{i=1}^{K}\{(\mathsf{T}_i, \mathsf{D}_i)\} \bigcup \bigcup_{i=1}^{d} \bigcup_{j=1}^{K}\{(\mathsf{T}_i, \mathsf{D}_j)\}$. *This network has links from the first* d *transmitters to all the receivers.*

Definition 6. *A* <u>*many-to-*d</u> *interference network of* K *users is an interference network specified by* $\mathsf{E} = \bigcup_{i=1}^{K}\{(\mathsf{T}_i, \mathsf{D}_i)\} \bigcup \bigcup_{i=1}^{K} \bigcup_{j=1}^{d}\{(\mathsf{T}_i, \mathsf{D}_j)\}$. *This network has links from all transmitters to the first* d *receivers.*

Example of 4-to-many interference network and many-to-4 interference networks with 6 users are depicted in Figure 4.4.

Definition 7. *A* <u>*fully-connected*</u> *interference network with* K *users is many-to-*K *interference network with* K *users which is also the same as a* K*-to-many interference network with* K *users.*

Definition 8. *A* <u>*chain*</u> *of* K *users is an interference network defined by* $\mathsf{E} = \bigcup_{i=1}^{K}\{(\mathsf{T}_i, \mathsf{D}_i)\} \bigcup \bigcup_{i=1}^{K-1}\{(\mathsf{T}_i, \mathsf{D}_{i+1})\}$. *This network has links from each transmitter to its next receiver. A Z-network is a chain of* 2 *users.*

Definition 9. *A* <u>*cyclic-chain*</u> *of* K *users is an interference network defined by* $\mathsf{E} = \bigcup_{i=1}^{K}\{(\mathsf{T}_i, \mathsf{D}_i)\} \bigcup \bigcup_{i=1}^{K-1}\{(\mathsf{T}_i, \mathsf{D}_{i+1})\} \bigcup\{(\mathsf{T}_K, \mathsf{D}_1)\}$. *This network is similar to a* $K - user$ *chain of Definition 8 except that the last transmitter interferes with the first receiver, thereby making the network a circular chain.*

4.2 Subgraph Scheduling

In this section, we will present a scheduling-based scheme which uses partial information at every node. The main idea is to divide the network into smaller disjoint sub-networks, each of which can operate optimally such that the normalized sum-rate of $\alpha^*(h) = 1$ for each sub-network. The choice of sub-networks thus becomes important and will be addressed in the form of independent sub-graphs as discussed below.

We will use the graph-theoretic terminology introduced in Section 4.1.4 to describe the scheduling algorithm. The graph theoretic formulation will allow us to compare our results to existing results in the literature for the special case of single-hop local view, as discussed in Section 4.3. Further, the graph-theoretic formulation will facilitate parallels between our proposed scheduling method and graph-concepts of chromatic number, again discussed in Section 4.3.

In Section 4.2.1, we will first describe the scheduling algorithm and derive its achievable normalized sum-rate performance for arbitrary hop-view, assuming independent graphs are known. In Section 4.2.2, we will derive the form of independent sub-graphs for 1- and 2-local view. An example is provided in Section 4.2.3.

4.2.1 Maximal Independent Graph Scheduling

Following standard graph theory terminology, a subgraph $A \subseteq \mathsf{G}$, is a subset of vertices and edges in G. The complement of A is A^c such that $(\mathsf{V}, \mathsf{E}) = A \cup A^c$. In this section, we will only consider subgraphs where both transmitter T_i and its corresponding receiver D_i are either in the subgraph together or in its complement. We will remove this restriction on subgraphs in Section 4.4 to propose a generalization which can achieve strictly higher rates for some networks compared to the following sub-graph schedule. Note that while the graph edges are weighted with the channel gains, the edge weights will not play a role in the description of the scheduling algorithm. Hence in our definition of subgraphs, we do not include edge weights. Since the network connectivity is known as side information to all the nodes and the schedules only depend on the connectivity, each user knows the schedule and hence when to transmit or when not to transmit.

With the above (restricted) definition of subgraph, any strict subgraph $A \subseteq \mathsf{G}$ represents a valid interference network with a reduced number of transmitter-receiver pairs. For that subgraph A, the normalized sum-rate $\alpha^*_A(h)$ can be defined, which is the ratio of sum-capacity with h-local view to the sum-capacity with global view ($h = \text{diameter}(A)$) for

network A.

Armed with the above framework, we can now define Independent Graph Scheduling as follows. Let $A_1, A_2, \ldots, A_t$ be t sub-graphs (not necessarily distinct) of the network G such that for each sub-graph A_i, $\alpha^*_{A_i}(h) = 1$. Since transmitter-receiver pairs are either part of A_i or A_i^c, each pair either appears in a subgraph A_i or it does not appear in A_i.

Definition 10 (Independent Graph Scheduling)**.** *Independent Graph Scheduling parametrized by t independent sub-graphs $A_1, A_2, \ldots, A_t$ uses t time-slots and schedules the sub-graph A_i in time-slot i.*

Define the indicator function

$$\mathbf{1}_{j \in A_i} = \begin{cases} 1 & \mathsf{T}_j \in A_i \\ 0 & \mathsf{T}_j \notin A_i \end{cases}. \tag{4.5}$$

For *any* given tuple of independent subgraphs, $\{A_i\}_{i=1}^t$, which satisfy $\alpha^*_{A_i}(h) = 1$, the next theorem gives the normalized sum-rate that can be achieved by sub-graph scheduling.

Theorem 8 (Achievable Normalized Sum-rate of Independent Graph Scheduling)**.** *Independent Graph Scheduling parametrized by t independent sub-graphs $A_1, A_2, \ldots, A_t$ achieves a normalized sum-rate of d/t, where*

$$d = \min_{j \in \{1,2,\ldots,K\}} \sum_{i=1}^{t} \mathbf{1}_{j \in A_i}. \tag{4.6}$$

Proof. Let $(C_1, \cdots, C_K)$ be any point in the full knowledge capacity region. The achievable rate in time-slot i is $R^{(i)} \geq \sum_{\{j\} \subseteq A_i} C_i - \tau_i$ by the choice of subgraphs A_i which satisfy $\alpha^*_{A_i}(h) = 1$. Note that τ is dependent on i since it can change in each time-slot due to selection of different subgraphs. Hence, the overall rate is $\frac{1}{t}\sum_{i=1}^t R_i \geq \frac{1}{t}\sum_{i=1}^t \sum_{\{j\} \subseteq A_i} C_i - \frac{1}{t}\sum_{i=1}^t \tau_i \geq \frac{d}{t}(C_1 + \cdots + C_K) - \frac{1}{t}\sum_{i=1}^t \tau_i$. By the definition of normalized sum-rate, $\alpha = d/t$. $\square$

First note that the sub-graphs A_i need not be distinct, which allows allocating more than one time-slot to a particular subgraph if needed. Second, the subgraph set $\{A_i\}_{i=1}^t$ and the number of subgraphs t are both design variables and should be chosen to maximize d/t, such that the overall network rate is maximized. The d/t-maximizing choice of subgraphs is labeled as a *maximal independent graph* (MIG) schedule.

The main idea behind MIG scheduling is to decouple transmissions of nodes from the unknown part of the network. This is done by switching off some of the flows such that the network gets partitioned into disconnected subgraphs. However, switching off flows means potentially lost rate compared to global-view optimal sum-capacity, so the sub-graphs have to be selected to maximize *spatial reuse*. That is, this involves operating as many flows as possible in parallel while still satisfying $\alpha^*_{A_i}(h) = 1$. Such subgraphs are labeled *maximal independent graphs* and form the core of MIGS. We characterize independent graphs next.

4.2.2 Identifying Independent Graphs

Since MIG scheduling schedules a subgraph A_i satisfying $\alpha^*_{A_i}(h) = 1$ in time-slot i, we need a characterization of independent sub-graphs. The problem turns out to be very challenging for a general h. We provide complete characterization for two important cases of $h = 1$ and $h = 2$, for both deterministic and Gaussian networks, in the next two theorems.

We note that the sufficient and necessary conditions in following two theorems are stated in terms of the graph properties of G. Theorem 9 uses the node degree, which is the number of edges incident on the node. Theorem 10 uses the definitions in Section 4.1.6.

Theorem 9 (1-Local View Independent Subgraphs)**.** *The normalized sum-capacity of a K-user interference network (deterministic or Gaussian) with 1-local view is equal to one,* i.e. $\alpha^*(1) = 1$, **if and only if** *all the receivers have degree* 1.

Proof. First, let us assume a Z-network. We will first show that in this network, $\alpha^* \leq 1/2$.

For a deterministic network model, assume that a normalized sum rate of α is achiev-

able; then

$$R_i \geq \alpha n_{ii} - \tau, \ \forall \ 1 \leq i \leq 2. \tag{4.7}$$

When all the channel gains are n, the condition that data can be decoded at the intended destinations gives

$$R_1 + R_2 \quad \leq \quad n.$$

Thus,

$$\alpha(2n) - 2\tau \leq R_1 + R_2 \leq n,$$

which is equivalent to $(2\alpha - 1)n \leq 2\tau$. Since this has to hold for all values of n where α and τ are independent of n, $\alpha \leq 1/2$.

For a Gaussian network model, assume that a normalized sum rate of α is achievable; then

$$R_i \geq \alpha \log(1 + |h_{ii}|^2) - \tau, \ \forall \ 1 \leq i \leq 2. \tag{4.8}$$

Further, when all $h_{11} = h_{12} = h_{22}$,

$$R_1 + R_2 \quad \leq \quad \log(1 + 2|h_{11}|^2).$$

This gives $(2\alpha - 1)\log(1 + |h_{11}|^2) \leq 1 + 2\tau$. Since this has to hold for all values of $|h_{11}|$ where α and τ are independent of h_{11}, $\alpha \leq 1/2$.

This shows that for a Z-network, $\alpha^*(1) \leq 1/2$. If there is a network containing a link from T_i to D_j for $i \neq j$, then as a genie consider a system of two users i and j where all other links are 0 and known to all. In this two user system, the Z-network will be an outer bound and thus $\alpha^*(1) \leq 1/2$. This proves that if there is a link from T_i to D_j for $i \neq j$, $\alpha^*(1) \leq 1/2$, thus proving the theorem. $\square$

Thus, with 1-local view, the only network that can support $\alpha^*(1) = 1$ is the one where

no transmitter interferes with other receivers, i.e, a network with K completely isolated flows. As a result, for a two-user interference network where transmitters *can* cause interference at other receivers, MIG scheduling will require the two flows to operate in a TDMA fashion. This is because the transmitters do not know any of the interfering link gains and thus have to optimize for the worst case in our formulation. The worst case network conditions are when the interfering network gains are the same as the direct link ($h_{12} = h_{11} = h_{22}$), where the network has only one degree of freedom and each node can thus transmit only half the time [48]. Thus, for the two-user case, the above conclusion can be derived from the results in [48]. Theorem 9 is a generalization to arbitrary K-user interference network.

We next provide the characterization of independent sub-graphs for two-local view, $h = 2$.

Theorem 10 (2-local View Independent Subgraphs)**.** *The normalized sum-capacity of a K-user interference network (deterministic or Gaussian) with 2-local view is equal to one (*i.e. $\alpha^*(2) = 1$*)* **if and only if** *all the connected components are of one of the following forms:*

1. *a one-to-many interference network*
2. *a fully-connected interference network*

Proof. A fully-connected network implies all nodes are within two-hops from each other. Thus, in this case, the diameter of such a network is two and thus $h = 2$ constitutes global knowledge. By the definition of normalized sum-rate, $\alpha^*(2) = 1$ for a fully-connected subnetwork.

The proof for condition (a) is provided in Appendix 4.A. Further, a converse to the theorem is also provided in Appendix 4.A. □

Contrasting Theorems 9 and 10, we see that increasing the local horizon from $h = 1$ to $h = 2$ increases the number of networks under which universally optimal performance can

be obtained. While for $h = 1$, universal optimality required no simultaneous transmissions, the independent subgraphs for $h = 2$ constitute a richer class. Not only are the fully connected interference networks possible (since their diameter is 1 for $K = 1$ and 2 for $K \geq 2$), one-to-many subgraphs are also possible even though their diameter is 4 for $K \geq 3$. For one-to-many subgraphs, the interfering transmitter is two hops away from all nodes and thus has full network knowledge. As a result, the optimal strategy is to allow $K - 1$ links to operate at their near-maximum link capacity and for the interfering flow to adjust its rate to cause no harmful interference (either the interference is below the noise floor or completely decodable and thus can be cancelled out). This is proven in Appendix 4.A.

4.2.3 An Example

Figure 4.4(a) gives a case of a six-user 4-to-many interference network. With 1-local view, the MIG Scheduling algorithm can be described as follows. Let $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{3\}$, $A_4 = \{4\}$, and $A_5 = \{5, 6\}$. Note that we have used a shorthand notation in describing these sets; $A_1 = \{a, b\}$ represents that A_1 is subgraph containing T_j, D_j for all $j \in A_1$ and all edges between the members of A_1 are also implied by this shorthand notation. We use a five time-slot strategy. In the ith time-slot, users in A_i transmit. MIG Scheduling achieves $\alpha(1) = 1/5$. We will show that this scheduling is optimal in Theorem 11.

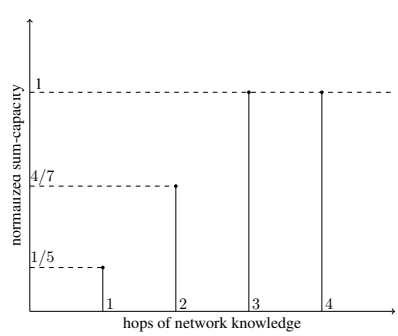

Figure 4.5: Normalized sum-capacity vs. h-local information for six-user 4-to-many interference network.

With 2-local view, the MIG Scheduling algorithm can be described as follows. Let $A_1 = A_2 = A_3 = \{1, 2, 3, 4\}$, $A_4 = \{1, 5, 6\}$, $A_5 = \{2, 5, 6\}$, $A_6 = \{3, 5, 6\}$, and $A_7 = \{4, 5, 6\}$. We use a seven time-slot strategy. In the ith time-slot, users in A_i transmit. MIG Scheduling achieves $\alpha(2) = 4/7$. We will show that this scheduling is optimal in Theorem 13. The normalized sum-capacity for increasing local information is depicted in Figure 4.5.

4.3 Optimality of MIG Scheduling

Now a natural question is: How good is the MIG scheduling? In this section, we address the question and show that MIG scheduling is optimal for several K-user networks with 1-local and 2-local view. Our results are limited to 1- and 2-local view only because independent graphs are known only for these two cases.

The reader will immediately note that much like capacity analyses of different multi-terminal networks (multiple access, interference network, Z-network etc), our proofs are largely taken on a case by case basis. At the current moment, there appears to be no general algorithmic procedure to derive general capacity region and as a result, we do not have an algorithmic procedure to derive normalized sum-capacity. However, we do note that we can derive normalized sum-capacity in our formulation for many cases while the global-view sum-capacity is still unknown.

4.3.1 1-Local View

Our main result in this section is determining the networks for which MIGS with one-local view is optimal. Recall that one-local view MIGS is equivalent to scheduling of non-interfering links in the network.

The key step in the proof is a derivation of an upper bound. The proof for the upper bound follows the following recipe in all cases for the deterministic model (the Gaussian

model is similar).

1. When any transmitter sees the direct channel capacity as n, it has to send at a rate $R_i \geq \alpha n - \tau$. This is because if the rate is $< \alpha n - \tau$, then when all other channel gains are 0, the worst-case guarantee of α is not achievable.

2. Find an upper bound on global-view sum capacity when all the channel gains in the network are n. Let the global sum capacity be bounded from above by $cn + d$ for some constants c and d which are independent of n. For example, one trivial outer bound is Kn for all K-user networks. To yield a useful bound, it is important to find the smallest constant c.

3. Combining Steps 1 and 2, an outer bound on α as $\alpha \leq c/K$ can be obtained where K is the number of users.

The proof follows the above three steps for each subset of users, and chooses the tightest outer bound thereafter.

Let $A \subseteq \mathsf{G}$ represents a valid interference network with $|A| \leq K$ transmitter-receiver pairs. Suppose the global view sum capacity of A when all the link capacities in A are n is upper bounded by $c_A n + d_A$ for some constants c_A and d_A which may depend on A but remain constant with changing A. Then,

$$\alpha \leq \min_A \frac{c_A}{|A|}. \tag{4.9}$$

The following theorem characterizes the cases where we can prove that MIG Scheduling is optimal.

Theorem 11 (1-Local View Optimality of MIG Scheduling)**.** *MIG scheduling is optimal with* 1*-local view when the network is of one of the following forms, and we also derive* α^* *for each case.*

1. *All the three user interference networks, except the 3-user cyclic-chain, (In Figure 4.6, $\alpha^*(1) = 1$ in (a), $\alpha^*(1) = 1/2$ in (b), (c), (d), (e), (f), (g), (j), and (k), and $\alpha^*(1) = 1/3$ in (h), (l), (m), (n), (o), and (p))*

2. *chain interference metwork, ($\alpha^*(1) = 1/2$ for $K \geq 2$),*

3. *d-to-many interference network, ($\alpha^*(1) = \frac{1}{d+1}$ for $K \geq 2$ and $1 \leq d < K$),*

4. *many-to-d interference network, ($\alpha^*(1) = \frac{1}{d+1}$ for $K \geq 2$ and $1 \leq d < K$),*

5. *fully-connected interference network, ($\alpha^*(1) = \frac{1}{K}$),*

Further, the achievability holds with $\tau = 0$ for both the deterministic and the Gaussian models.

Proof. 1. For a three user interference network, we will consider all the possible networks as shown in Figure 4.6 up to relabeling of the users. In networks (b), (c), (d), (e), (f), (g), (j), and (k), the same upperbound as that for the Z-network ($\alpha^* \leq 1/2$) holds since the channel gains except those that forms a Z-network can be made 0 and are known to all as a genie (Since there is only 1-local view, existence of zero capacity links do not help get more information about the network). Further, this can be achieved with MIG scheduling with two time-slots. For (h), (l), (m), (n), (o), and (p), consider the topology equivalent to (h) by setting all other network gains to 0 and make this global information. With this, the outer bound for the case (h) holds for all these cases. In the case (h), suppose all the network gains are the same. Then, D_1 decodes the message of T_1. Thus, D_2 will be able to decode the message of T_1 as well as T_2 since after decoding message of T_2 and subtracting, the equivalent signal is the same as that at D_1. Similarly, D_3 will be able to decode the message of T_1, T_2 as well as T_3 since after decoding the message of T_3 and subtracting, the equivalent signal is same as that at D_2. Thus, the normalized sum capacity is upper bounded by that of the Multiple Access Network to D_3 thus giving $1/3$ as an upper bound.

Further, $1/3$ can be achieved using MIG scheduling, scheduling the three users in three different time-slots.

2. The achievability of $1/2$ follows by using two time-slots, scheduling odd numbered users in the first time-slot and even numbered in the second time-slot, while the outer bound for the Z-network also holds here by the same arguments as in the previous part. Thus, $\alpha^*(1) = 1/2$.

3. As an outer bound, consider $d+1$ users containing the first d users that are interfering at all receivers and the $d+1^{th}$ user as one other user. Consider the rest of the direct channel gains as 0 and known to all. In this case, it is easy to see that when all the channel gains are equal, D_{d+1} has to decode all the messages, thus upper bounding the normalized sum capacity by that of the Multiple Access Network to this receiver. For achievability, consider MIG scheduling using $d+1$ time-slots with $A_1 = \{1\}, \cdots, A_d = \{d\}$ and $A_{d+1} = \{d+1, \cdots, K\}$. Note that this extends the example of a 4-to-many interference network with 6 users with 1-local view provided in Section 4.2.3.

4. As an outer bound, consider $d+1$ users containing the first d users that are receiving interference from all transmitters and the $d+1^{th}$ user as one other user. Assume the rest of the direct channel gains are 0 and known to all. In this case, it is easy to see that when all the channel gains are equal, D_1 has to decode all the messages, thus upper bounding the normalized sum capacity by that of the Multiple Access Network to this receiver. For achievability, consider MIG scheduling using $d+1$ time-slots with $A_1 = \{1\}, \cdots, A_d = \{d\}$ and $A_{d+1} = \{d+1, \cdots, K\}$.

5. When all the channel gains are equal, each destination has to decode all the messages and is thus upper-bounded by that of the Multiple Access Network, giving $1/K$ as the upper bound. This is achievable using MIG scheduling, scheduling each user in a separate time-slot.

□

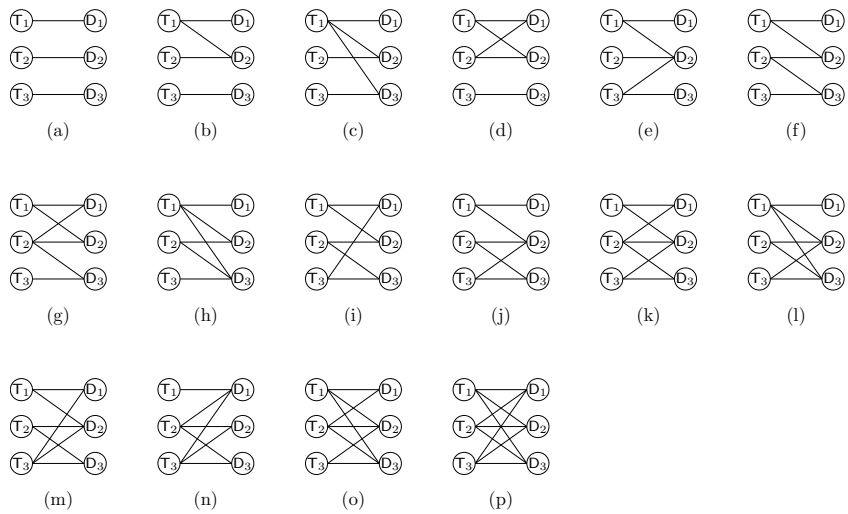

Figure 4.6: All possible canonical network topologies in a three-user interference network.

Thus, maximal scheduling of non-interfering links can be information-theoretically optimal for many networks. The theorem only gives sufficient conditions and thus not a sharp characterization of all networks which can be operated optimally with scheduling. However, observing the class of networks given in the theorem, it appears that MIG scheduling might be optimal for a large class of networks. We, thus, explore the connection further in the next section and also discuss the relationship with graph coloring.

4.3.2 1-Local View: Relation to Maximal Independent Set Scheduling

For one-local view, the MIG scheduling strategy reduces to maximal independent set scheduling (MIS scheduling) that can be described as follows. An independent set $A_i \subseteq \{1, \cdots, K\}$ is a set that contains mutually non-interfering nodes. A maximal independent set (MIS) is an independent set A_i such that $A_i \cup \{x\}$ is not an independent set for any $x \in \{1, \cdots, K\} \setminus A_i$. Using t time-slots, a maximal independent set A_i is scheduled in each time-slot such that

$$\min_i \frac{1}{t} \sum_{j=1}^{t} \mathbf{1}_{i \in A_j}$$

is maximized over the choice of t and $A_1 \cdots, A_t$. When a user is scheduled, it sends at the direct channel rate (and uses power of 1 in a Gaussian network). The resulting strategy achieves a normalized sub-rate of $\alpha = \min_i \frac{1}{t} \sum_{j=1}^{t} \mathbf{1}_{i \in A_j}$.

This is similar to the following vertex coloring algorithm. To relate to vertex coloring, we will need the concept of conflict graph [44, Chapter 2.2] derived from G as follows. Consider a graph C with K vertices (half as many as present in G), where two vertices i and j are connected if there is an edge between T_i and D_j or between T_j and D_i in G. Suppose that there are t colors, labeled $1, 2, \cdots, t$. We assign $k \leq t$ of these colors to each vertex in C such that the sets of colors associated with two vertices connected by an edge are disjoint. In conventional graph coloring [41], each vertex has only one color and the objective is to assign a color to each vertex such that adjoining vertices have different colors. In contrast, the generalized *set* coloring algorithm can assign multiple colors to each vertex as long as the color sets for adjoining vertices are disjoint. This is similar to fractional coloring considered in [32]. The best set coloring corresponds to MIS schedule and maximizes k/t with k and t as variables. The scheduling algorithm uses t time-slots and schedules the vertices with color i in the ith time-slot.

This algorithm is similar to Maximal Weight Independent Set Scheduling in [53] except that the weights are decided not by the queue lengths, but by the weights that maximize the minimum proportion each link is used.

A k-fold coloring of a graph is an assignment of sets of size k to vertices of a graph such that adjacent vertices receive disjoint sets. A $t : k$-coloring is a k-fold coloring out of t available colors. The k-fold chromatic number ξ_k is the least t such that a $t : k$-coloring exists. Note that MIS Scheduling achieves $\alpha = \max_{k \in \mathbb{N}} \frac{k}{\xi_k}$, where ξ_k is the k-fold chromatic number of the conflict graph. The following theorem gives an optimality condition of MIS Scheduling algorithm in terms of the k-fold chromatic number of the conflict graph.

Theorem 12 (1-Local View Optimality of MIS Scheduling)**.** *If the conflict graph of an*

interference network has k-fold chromatic number of at most $2k$ for some $k \in \mathbb{N}$, then the MIS scheduling algorithm is optimal, i.e. achieves normalized sum-capacity with 1-local view.

Corollary 1. *Chain interference networks, different configurations of two-user interference networks, 1-to-many and many-to-1 interference networks are some special cases that have chromatic number ≤ 2 in the conflict graph. Moreover, the normalized sum capacity is the inverse of the 1-fold chromatic number of the conflict graph in these cases.*

Proof. If the chromatic number in the conflict graph is 1, then there is no link between any T_i and D_j for $j \neq i$. Since this connectivity satisfies the condition of $\alpha^*(1) = 1$, the theorem holds.

If the chromatic number in the conflict graph is 2, then there is at-least one link between T_i and D_j for some $j \neq i$. In this case, $\alpha^*(1) \leq 1/2$ by the same arguments as in Theorem 9. Further, this can be achieved by MIS scheduling; scheduling the vertices of two different colors in two time-slots. □

4.3.3 2-Local View

We start with a theorem which provides sufficient conditions under with MIG scheduling is two-local view optimal.

Theorem 13 (2-Local View Optimality of MIG Scheduling)**.** *MIG Scheduling achieves normalized sum-capacity with 2-local view when the network is of one of the following forms. We also derive their normalized sum-capacity.*

1. *Two user interference network, ($\alpha^*(2) = 1$),*
2. *Chain interference network, ($\alpha^*(2) = 2/3$ for $K > 2$),*
3. *d-to-many interference network, ($\alpha^*(2) = \frac{d}{2d-1}$ for $1 \leq d < K$ and $K > 2$),*

4. *many-to-one interference network, ($\alpha^*(2) = \frac{K-1}{2K-3}$ for $K > 2$),*

5. *fully-connected interference network, ($\alpha^*(2) = 1$).*

Proof. 1. In this case, the result follows from Theorem 10.

2. The outer bound of topology (f) in Appendix 4.A holds in this case by assuming all other channel gains to be 0 and known to all. For achievability, the MIG scheduling algorithm can be described as follows. Let $A_j = \{3i + j : i \in \mathbb{Z}, 3i + 1 \leq K\}$ for $j = 1, 2, 3$. According to the MIG scheduling algorithm, three time-slots are used and users in A_i use a strategy that achieves $\alpha(2) = 1$ in the ith time-slot. The MIG scheduling strategy achieves $\alpha(2) = 2/3$.

3. Let $d > 1$ because for $d = 1$ the statement holds by Theorem 10. For the outer bound, consider a $d + 1$ user d-to-many interference network. Suppose that there exists a scheme achieving normalized sum capacity of α. We first prove the result for the deterministic model. For user $d + 1$, since it does not know any other direct channel gain, it has to use $R_{d+1} \geq \alpha n - \tau$ when it sees that all the channel gains within 2 hops have capacity equal to n. Suppose that all other direct links have capacity of n while all other cross links have zero capacity. Then, all $R_i \leq (1 - \alpha)n + \tau$ for $i \in [1, d]$ yielding that sum rate $\leq (d - (d - 1)\alpha)n + (K - 2)\tau$. This sum rate has to be at-least $\alpha(dn) - \tau$. Since this holds for all n, $\alpha \leq \frac{d}{2d-1}$. Similar proof holds in the Gaussian model as follows. For user $d + 1$, since it does not know any other direct channel gain, it has to use $R_{d+1} \geq \alpha \log(1 + \mathsf{SNR}) - \tau$ when it sees that all the channel gains within 2 hops have channel gain equal to $\sqrt{\mathsf{SNR}}$. Suppose that all other direct links have capacity of $\sqrt{\mathsf{SNR}}$ while all other cross links have zero capacity. Then, all $R_i \leq (1 - \alpha) \log(1 + \mathsf{SNR}) + \tau + 1$ for $i \in [1, d]$ yielding that sum rate $\leq (d - (d - 1)\alpha) \log(1 + \mathsf{SNR}) + (K - 2)\tau + d$. This sum rate has to be at-least $\alpha(d \log(1 + \mathsf{SNR})) - \tau$. Since this holds for all SNR, $\alpha \leq \frac{d}{2d-1}$.

For achievability, consider $2d-1$ time-slots in which the first $d-1$ time-slots only users 1 to d transmit. In the remaining d time-slots one user among the first d and all the users $> d$ transmit making it an equivalent one-to-many configuration (or, $A_1 = \cdots = A_{d-1} = \{1, \cdots, d\}$ and $A_{d-1+j} = \{j, d+1, \cdots, K\}$ for $j = 1, \cdots, d$). Thus, this is MIG scheduling with each user scheduled in d time-slots achieving $\alpha^*(2) = \frac{d}{2d-1}$. Note that this extends the example of a 4-to-many interference network with 6 users with 2-local view provided in Section 4.2.3.

4. Suppose that a normalized sum rate of α can be achieved. We first consider a deterministic model. $R_K > \alpha n_{KK} - \tau$ since the Kth user has to send at this rate when all other direct channel gains are 0 and are not known to user K. Now, suppose all the channel gains are n. In this case, $R_i < (1-\alpha)n+\tau$ for $1 \leq i \leq K-1$. Thus, the sum rate achieved is less than $(K-2)(1-\alpha)n+(K-2)\tau+n$. This sum rate has to be at-least $\alpha(K-1)n-\tau$. Since this has to hold for all n, $\alpha \leq \frac{K-1}{2K-3}$. For a Gaussian model, $R_K \geq \alpha \log(1+|h_{KK}|^2) - \tau$ since the Kth user has to send at this rate when all other direct channel gains are 0 and are not known to user K. Now, suppose all the channel gains are $\sqrt{\mathsf{SNR}}$. In this case, $R_i \leq (1-\alpha)\log(1+\mathsf{SNR})+\tau+1$ for $1 \leq i \leq K-1$. Thus, the sum rate achieved is $\leq (K-2)(1-\alpha)\log(1+\mathsf{SNR})+(K-2)\tau+n+K-1$. This sum rate has to be at-least $\alpha(K-1)\log(1+\mathsf{SNR})-\tau$. Since this has to hold for all SNR, $\alpha \leq \frac{K-1}{2K-3}$.

   For achievability, consider the data transfer over $2K-3$ time slots. In the timeslot i satisfying $1 \leq i \leq K-1$ users i and K transmit. They form a Z-network and use the optimal strategy for this channel with partial information. In the remaining $K-2$ timeslots, users $1, \cdots, K-1$ transmit at full rate. Let $(R_1, R_2, \cdots, R_K)$ be any point in the global information capacity region. In the ith time-slot where $1 \leq i \leq K-1$, sum rate of at least $R_i + R_K$ can be achieved while in the remaining $K-2$ timeslots, sum rate of at least $\sum_{1 \leq i \leq K-1} R_i$ can be achieved. Thus, the sum-capacity with a factor of $\frac{K-1}{2K-3}$ can be achieved.

5. In this case, the condition of $\alpha^* = 1$ is satisfied by Theorem 10 thus proving the statement.

□

Note that for all the cases in the statement of Theorem 13, we have characterized normalized sum-capacity in the case of 1-local and 2-local view. For a fully-connected interference network, larger subgraphs increased $\alpha^*(2) = 1$ from $\alpha^*(1) = 1/K$. For a d-to-many interference network, one-to-many configurations that satisfy $\alpha^*(2) = 1$ could be exploited to get $\alpha^*(2) = \frac{d}{2d-1}$ from $\alpha^*(1) = 1/2$. With one-local view, only single user encoding and decoding operations are performed while with 2-local view, optimal encoding and decoding operations for one-to-many interference network and fully-connected network need to be performed.

We consider the sixteen network configurations shown in Figure 4.6 for the two-local view separately in the following theorem. The next theorem shows that MIG Scheduling is normalized sum-capacity achieving for 12 out of 16 canonical cases.

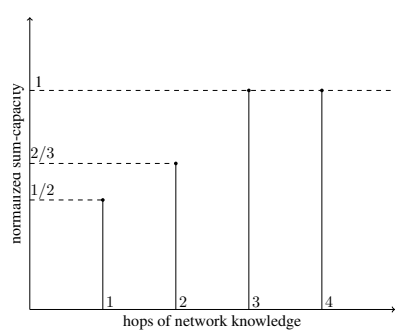

Figure 4.7: Normalized sum-capacity vs. h-local information for cases (e), (f), (h), (i), (j), (m), (n) in Figure 4.6.

Theorem 14 (2-Local View Optimality of MIG Scheduling in Three-user Interference Network)**.** *The MIG Scheduling is optimal with 2-local view when the three-user interference network is one of the following types in Figure 4.6: (a), (b), (c), (d), (e), (f), (h), (i), (j), (m), (n), (p).*

Proof. For cases (a), (b), (c), (d), and (p), $\alpha^* = 1$ by Theorem 10. For the remaining cases, the outer bounds of $2/3$ hold as shown in Appendix 4.A. The achievability follows by choosing $A_1 = \{1, 2\}$, $A_2 = \{2, 3\}$, and $A_3 = \{1, 3\}$. The normalized sum-capacity with h-local view for varying h in these remaining cases is depicted in Figure 4.7 □

Here, we do not prove the optimality of MIG scheduling for the remaining four cases. We conjecture that the outer bound is tight in cases (g) and (k). The achievability would require the capacity region in these cases to give better schemes, and is left as future work.

4.4 Optimality of MIG Scheduling: Extension of MIG Scheduling with 1-Local View

Is MIG scheduling always optimal? In this section, we will illustrate an example where MIG scheduling is not optimal. This example will use 1-local view and achieve a normalized sum capacity better than MIS scheduling (MIG scheduling with 1-local information). This gives a way to extend the MIS scheduling with 1-local information to involve coding across the time-slots and hence we define a new strategy called Coded Maximal Independent Set (CMIS) scheduling. This will be followed by some cases when this algorithm is optimal.

4.4.1 An Example Where MIS Scheduling is Not Optimal

We will now illustrate the only case when MIS Scheduling is not optimal in a 3-user interference network, which is a cyclic-chain interference network. The MIS Scheduling algorithm uses three time-slots, scheduling user i in time-slot i. Thus, MIG Scheduling achieves $\alpha(1) = 1/3$ (Note that there are only 3 independent sets consisting of individual users and thus optimality of $1/3$ using MIS scheduling is straightforward). We will now describe another strategy for this example, which uses two time-slots as follows (and de-

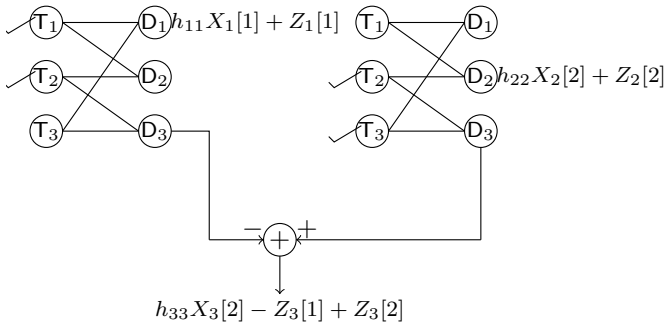

Figure 4.8: Two time-slots for CMIS scheduling. The transmitters with a tick sign transmit, the second user repeats X_2 ($X_2[1] = X_2[2]$) in the two time-slots.

picted in Figure 4.8). The main idea is to perform coding across time. In the first time-slot, we schedule $A_1 = \{1, 2\}$ and in second time-slot, we schedule $A_2 = \{2, 3\}$ such that the codeword of the second user is repeated in the two time-slots. All the users send at the rate equal to the direct link capacity to the intended receiver (n_{ii} in the deterministic and $\log(1 + |h_{ii}|^2)$ in the Gaussian model). In the Gaussian model, power of 1 is used at the first two transmitters while power 2 is used for the third transmitter. Note that this does not effect average power since this transmitter will be used half the time. We will now show that the data can be decoded at the intended receivers. The first receiver can decode its data in the first time-slot since it receives no interference. The second receiver can similarly decode the data in the second time-slot. The third receiver on the other hand subtracts the data received in the first time-slot from that in the second time-slot which gives a interference-free direct signal which can be decoded; double power level at the third transmitter is used since the noise power will also be double the single slot noise power. Thus, all the receivers can decode the data and this strategy achieves $\alpha(1) = 1/2$.

This example motivates an extension of the MIS Scheduling algorithm to involve coding. This new scheduling algorithm is called Coded Maximal Independent Set Scheduling (CMIS Scheduling) which will be described in the next subsection.

4.4.2 Definition of CMIS Scheduling

In this subsection, we will define the CMIS algorithm for the deterministic model and the Gaussian model separately. In this section, we will only consider subgraphs $A \subseteq G$ with a set of transmitters T_i and all the receivers $\{\mathsf{D}_1, \cdots, \mathsf{D}_K\}$ in the subgraph because we do not want to throw away any received signal. Let the in-degree at D_i be denoted by d_i. Suppose that each transmitter generates k independent codewords (The rate of these codewords will be n_{ii} for the deterministic model, and $\log(1 + |h_{ii}|^2/b_i)$ for the Gaussian model where b_i will be defined in the Gaussian subsection below). Let $S_{i,j}$ be a vector of time-slots in which transmitter T_i is transmitting the jth codeword. Note that each time-slot should be used at a transmitter T_i for only one codeword, thus giving $S_{i,u}$ and $S_{i,v}$ disjoint for $u \neq v$. Thus, in time-slot u, the subgraph A_u used has transmitters T_i where i satisfies $S_{i,j} \supseteq \{u\}$ for some $1 \leq j \leq k$. The sets S_{ij} and thus A_u, t and k are all design variables for the CMIS scheduling algorithm that satisfy some conditions on the constraint matrix, which is defined next.

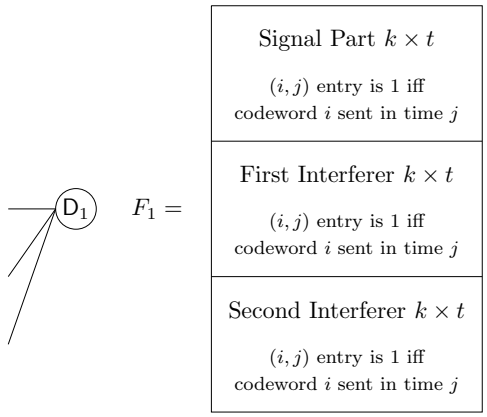

Figure 4.9: Constraint matrix with t columns $\mathsf{M}_{i,1}, \cdots \mathsf{M}_{i,t}$ where each column represents the different codewords being sent for the direct signal and the interferers.

We form a binary constraint matrix F_i of size $kd_i \times t$ at each receiver i which is defined as follows. The constraint matrix has d_i blocks of size $k \times t$ where the top block corresponds to the transmitted signal from T_i while the rest belong to the different transmitters causing interference at D_i as depicted in Figure 4.9. In each $k \times t$ subpart of this matrix, only the

entries $(j, S_{i,j})$ are 1 for all $1 \leq j \leq k$. Suppose that the t columns of the constraint matrix are denoted as $\mathsf{M}_{i,1}, \cdots \mathsf{M}_{i,t}$ respectively. The constraints that the constraint matrix has to satisfy are different for deterministic and Gaussian network model, and is explained below separately for the two cases.

Deterministic Model

For a deterministic network model, CMIS Scheduling can be described as follows. Suppose that a $kd_i \times t$ matrix with the top $k \times k$ part an identity and rest of the elements 0 can be formed by choosing each column j as $\sum_{l=1}^{t} a_{jl}\mathsf{M}_{i,l}$ where a_l's are binary and addition is binary addition. If such a transformation exist at destination i, this configuration is feasible at vertex i. If the assignment of C_i and S_{ij} is feasible at each vertex, this strategy achieves α of k/t. The strategy that achieves the maximum k/t is called Coded Maximal Independent Set (CMIS) Scheduling.

The scheduling algorithm uses t time-slots. Each user forms k independent codewords at rate n_{ii}. User i transmits codeword j in time-slots corresponding to $S_{i,j}$. It is easy to see that the data can be decoded at the receivers. The constraint matrix reduction represents that all the k independent codewords can be decoded in the presence of the interference from other transmitters.

Gaussian Model

For a Gaussian network model, CMIS Scheduling can be described as follows. Suppose that a $kd_i \times t$ matrix with the top $k \times k$ part an identity and rest of the elements 0 can be formed by choosing each column j as $\sum_{l=1}^{t} a_{jl}\mathsf{M}_{i,l}$ where $a_l \in \mathbb{R}$ and addition is real addition. If such a transformation exist at destination i, this configuration is feasible at vertex i. If the assignment of C_i and S_{ij} is feasible at each vertex, this strategy achieves α of k/t. The strategy that achieves the maximum k/t is called Coded Maximal Independent Set (CMIS) Scheduling.

Note that a_{jl} can be chosen to be 0 for $j > k$. The matrix formed by a_{jl} satisfies

$$\begin{aligned} &\begin{bmatrix} \mathsf{M}_{i,1} & \cdots & \mathsf{M}_{i,t} \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{k1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ a_{1t} & \cdots & a_{kt} & 0 & \cdots & 0 \end{bmatrix} \\ = &\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{4.10}$$

Since this is an under-determined system, the following is a solution.

$$\begin{aligned} &\begin{bmatrix} a_{11} & \cdots & a_{k1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ a_{1t} & \cdots & a_{kt} & 0 & \cdots & 0 \end{bmatrix} \\ = &\begin{bmatrix} \mathsf{M}_{i,1} & \cdots & \mathsf{M}_{i,t} \end{bmatrix}^{\dagger} \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{4.11}$$

where $A^{\dagger}$ represents pseudo-inverse of matrix A. Let $b_i = \max_{l=1}^{k} \sum_{m=1}^{t} a_{lm}^2$, where i represents that the constraint matrix is formed for receiver i.

User i forms k independent codewords at rate $\log(1 + \mathsf{SNR}_i/b_i)$. User i transmits codeword j in the time-slots in $S_{i,j}$ with power SNR_i. It is easy to see that the data can be decoded at the receivers and the strategy would achieve $\alpha = k/t$ with $\tau \leq \alpha \sum_{i=1}^{K} \log(b_i)$ which is independent of channel gains. This can be further optimized in certain cases by changing the corresponding rates and powers. The constraint matrix specifies the interference and the data at each user; existence of a_{ij}'s represents that the data can be decoded in the presence of interference. The value of b_i represents that while decoding the data, the noise gets added up which has to be compensated by the decrease in rate. Since this rate gap is not a function of channel gains, we get a constant τ that is independent of the channel gains.

4.4.3 Optimality of CMIS Scheduling

CMIS scheduling achieves normalized sum rate which is at-least that achieved by MIS scheduling. In this section, we prove that CMIS Scheduling is optimal in a K-user cyclic-chain, while MIS scheduling is not in an odd user cyclic-chain.

Theorem 15 (1-Local View Optimality of CMIS Scheduling)**.** *CMIS scheduling is optimal with 1-local view for a K-user cyclic-chain, while MIS is not for odd $K \geq 3$. Further, an achievable strategy with $\tau = 0$ is possible in this case for both the deterministic and the Gaussian model.*

Proof. For a K-user cyclic-chain, an outer bound of $1/2$ holds from the Z-network by the same arguments as in Theorem 9. If the number of users is even, then using two time-slots and scheduling even and odd users as in MIS scheduling will be optimal. If the number of users is odd, consider two time-slots. In the first time-slot, all odd users transmit while in the even time-slot all even users transmit except that the last user also transmits but it repeats the data in the previous time-slot.

In a deterministic model, all the users send at full rate (n_{ii} for T_i) and the data can be proved to be decoded.

For a Gaussian network, user i sends at a rate of $\log(1 + |h_{ii}|^2)$. All the users except the first use power of 1 to send the data while the first transmitter uses power of 2.

The data can be decoded in the same way as explained for a 3-user cyclic-chain thus proving the result. □

4.A Universally Optimal Strategies with 2 Hops in Three User Topologies

We first note that for $K < 3$, all the topologies have connected components that satisfy the property in the statement of the theorem and thus the result holds trivially.

In a three-user interference network, there are at-most six cross links, existence or non-existence of which gives rise to $2^6 = 64$ cases. Some of the cases are topologically equivalent (up to relabeling of users) and hence that will reduce the total number of possibilities considered in this chapter to the sixteen that are shown in Figure 4.6.

(e): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the second user sees the channel gains as $n_{22} = n_{12} = n_{32} = n$. In this case, the rate allocated by the second user has to be at-least $\alpha n - \tau$ for some τ independent of n. This is because the achievable sum rate has to be at-least $\alpha C_{sum} - \tau$ even if $n_{11} = n_{33} = 0$. Now suppose all the channel gains are equal to n. In this case since $R_1 + R_2 \leq n$, we have $R_1 \leq (1-\alpha)n + \tau$. Further since $R_2 + R_3 \leq n$, we have $R_1 + R_2 + R_3 \leq (2-\alpha)n + \tau$. This sum rate has to be at-least $\alpha(2n) - \tau$ since the sum capacity is $2n$. Thus, $2\alpha n - \tau \leq (2-\alpha)n + \tau$ or $(3\alpha - 2)n \leq 2\tau$. If $3\alpha - 2 > 0$, n can be chosen large enough to not satisfy the inequality. So, the inequality can be satisfied for all n only when $3\alpha - 2 \leq 0$ which gives $\alpha \leq 2/3$.

(f): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the first user sees the channel gains as $n_{11} = n_{12} = n_{22} = n$. In this case, the second user will send at rate $\geq \alpha n - \tau$ if $n_{33} = 0$ by the same arguments as in part (e) which implies $R_1 \leq (1-\alpha)n + \tau$. Now suppose all the channel gains are equal to n. In this case since $R_2 + R_3 \leq n$, we have $R_1 + R_2 + R_3 \leq (2-\alpha)n + \tau$. This sum rate has to be at-least $\alpha(2n) - \tau$ since the sum capacity is $2n$. Thus, $2\alpha n - \tau \leq (2-\alpha)n + \tau$ or $(3\alpha - 2)n \leq 2\tau$. Similar arguments as in (e) yields $\alpha \leq 2/3$.

(g): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the first user sees the channel gains as $n_{11} = n_{12} = n_{21} = n$, $n_{22} = 2n$. In this case, $2R_1 + R_2 \leq 2n$ gives that if $R_1 = x$, $R_1 + R_2 \leq 2n - x$. If $n_{33} = 0$, the first user should give a strategy such that (R_1, R_2) satisfy $R_1 + R_2 \geq 2n\alpha - \tau$ giving $R_1 \leq 2n(1-\alpha) + \tau$. Now suppose that $n_{23} = n_{33} = 2n$ giving $R_2 + R_3 \leq 2n$. We thus have $R_1 + R_2 + R_3 \leq 2n(2-\alpha) + \tau$. Since this has to be at-least $3n\alpha - \tau$, using similar arguments as in (e) yields $\alpha \leq 4/5$.

(h): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the third user sees the channel gains as $n_{33} = n_{13} = n_{23} = n$. In this case, the third user will send at rate $\geq \alpha n - \tau$ if $n_{11} = n_{22} = 0$ by the same arguments as in part (e). Further suppose $n_{11} = n_{22} = n$, $n_{12} = 0$ which implies $R_1 \leq (1-\alpha)n + \tau$. Since $R_2 + R_3 \leq n$, we have $R_1 + R_2 + R_3 \leq (2-\alpha)n + \tau$. This sum rate has to be at-least $\alpha(2n) - \tau$ since the sum capacity is $2n$. Thus, $2\alpha n - \tau \leq (2-\alpha)n + \tau$ or $(3\alpha - 2)n \leq 2\tau$. Similar arguments as in (e) yields $\alpha \leq 2/3$.

(i): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the first user sees all the channel gains in two hops equal to n. In this case, if $n_{33} = 0$, the second user will have to send at rate $\geq \alpha n - \tau$ and thus $R_1 \leq (1-\alpha)n + \tau$. Further suppose all the channel gains are equal to n which implies $R_2 + R_3 \leq n$ thus giving $R_1 + R_2 + R_3 \leq (2-\alpha)n + \tau$. This sum rate has to be at-least $\alpha(2n) - \tau$ since the sum capacity is $2n$. Thus, $2\alpha n - \tau \leq (2-\alpha)n + \tau$ or $(3\alpha - 2)n \leq 2\tau$. Similar arguments as in (e) yields $\alpha \leq 2/3$.

(j): The same steps as in (i) yield $\alpha \leq 2/3$.

(k): Let $n_{32} = 0$ be the global information. Applying the same steps as in (g) for the other channel gains gives $\alpha \leq 4/5$.

(l): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the second user sees the channel gains as $n_{22} = n_{23} = n_{32} = n$, $n_{33} = 2n$. In this case, we get $R_2 \leq 2n(1-\alpha) + \tau$ as in (g). Now suppose that $n_{13} = n_{33} = 2n$, $n_{12} = 0$ giving $R_1 + R_3 \leq 2n$. Using similar arguments as in (g) yields $\alpha \leq 4/5$.

(m): Let $n_{31} = 0$ be the global information. Applying same steps for the remaining channel gains as in (i) yields $\alpha \leq 2/3$.

(n): Suppose that the normalized sum rate of α can be achieved. Then, if the first user sees all the channel gains as n in two hops, $R_1 \geq \alpha n - \tau$. Suppose that $n_{22} = n_{33} = n$, $n_{23} = n_{32} = 0$. In this case, $R_1 + R_2 + R_3 \leq (2-\alpha)n + \tau$ and since it has to be at-least $2\alpha n - \tau$ for all n, $\alpha \leq 2/3$.

(o): Suppose that the normalized sum rate of α can be achieved. Further, suppose that the third user sees the channel gains as $n_{33} = n_{23} = n_{32} = n$, $n_{22} = 2n$, $n_{13} = 0$. In this case, we get $R_3 \leq 2n(1-\alpha) + \tau$ as in (g). Now suppose that $n_{12} = n_{11} = n_{21} = 2n$ giving $R_1 + R_2 \leq 2n$. Using similar arguments as in (g) yields $\alpha \leq 4/5$.

We will now consider $K > 3$. Consider that there exist a connected component with $K > 3$ users which is not in the one-to-many configuration or in the fully-connected configuration. Then, two cases arise:

1. There exists a transmitter (say T_1) which has degree d satisfying $1 < d < K$.

2. All the transmitter nodes have degrees 1 or K, such that the number of nodes n having degree K satisfy $1 < n < K$.

For the first case, take the nodes $1, \cdots, d$ as the nodes whose receivers are connected to T_1. Now, there exist a transmitter-receiver pair among $d+1, \cdots, K$ whose transmitter or receiver is connected to any of the nodes $1, \cdots, d$. Choose any such pair and call it pair $d+1$. The receiver of $d+1$ is not connected to transmitter 1. Now if the receiver of the first node is connected to the transmitter of $d+1$, then choose the nodes $1, 2, d+1$ and assume that the direct link of all other users is zero and this information is given as a genie to all the nodes. This creates a genie-aided system in which the nodes 1, 2 and $d+1$ have the uncertainties about all the links connecting them and know 2-local view among these links only. In this genie-aided system, there does not exist any universally optimal strategy, thus proving the claim (since it makes a connected three-user component which is not in the one-to-many configuration or in the fully-connected configuration). If pair $d+1$ is not connected to pair 1, let us say it is connected to pair $2 \leq j \leq d$. Then, choosing nodes 1, j, $d+1$ and repeating the same argument as above proves the statement.

For the second case, choose the three nodes as any two nodes in which the transmitter has degree K and one in which the transmitter has degree 1. Repeating the above genie-aided proof for these three nodes proves the theorem.

This completes the proof that there does not exist a universally optimal strategy for a topology that does contain a connected component which is not in the one-to-many configuration or in the fully-connected configuration.

It is easy to see that there exists a strategy with $\alpha = 1$ if all the connected components of the topology are in one-to-many configuration or fully-connected configuration. For fully-connected components, all the nodes know their connected components and thus each node in the component can use the optimal strategy for its component. For the one-to-many components, each of the users whose transmitters have degree 1 send at rate equal to the rate that the direct channel can support and the remaining user knows all the channel gain and adjust its rate correspondingly. Assume that it is a one-to-many component of L users with the first transmitter having degree L. The above strategy achieves a sum rate of $R_{sum} = \sum_{i=2}^{L} n_{ii} + \sum_{i=1}^{n_{11}} \mathbf{1}_{|U_k|=0}$, where $|U_k|$ is the number of users potentially experiencing interference from the kth signal level of first transmitter which is the same as the sum capacity with global channel information in [24].

We now see the steps extended to a Gaussian network model. For this, we only consider case (e). Extension of the remaining steps is similar and is thus omitted.

(e): Suppose that a normalized sum rate of α can be achieved. Further, suppose that the second user sees the channel gains as $h_{22} = h_{12} = h_{32} = \sqrt{\mathsf{SNR}}$. In this case, the rate allocated by the second user has to be at-least $\alpha \log(1 + \mathsf{SNR}) - \tau$ for some τ independent of n. This is because the achievable sum rate has to be at-least $\alpha C_{sum} - \tau$ even if $h_{11} = h_{33} = 0$. Now suppose all the channel gains are equal to $\sqrt{\mathsf{SNR}}$. In this case since $R_1 + R_2 \leq \log(1 + \mathsf{SNR}) + 1$, we have $R_1 \leq (1 - \alpha) \log(1 + \mathsf{SNR}) + \tau + 1$. Further since $R_2 + R_3 \leq \log(1 + \mathsf{SNR}) + 1$, we have $R_1 + R_2 + R_3 \leq (2 - \alpha)n + \tau + 2$. This sum rate has to be at-least $\alpha(2 \log(1 + \mathsf{SNR})) - \tau$ since the sum capacity is at-least $2 \log(1 + \mathsf{SNR})$. Thus, $2\alpha \log(1 + \mathsf{SNR}) - \tau \leq (2 - \alpha) \log(1 + \mathsf{SNR}) + \tau + 2$ or $(3\alpha - 2) \log(1 + \mathsf{SNR}) \leq 2\tau + 2$. If $3\alpha - 2 > 0$, SNR can be chosen large enough to not satisfy the inequality. So, the inequality can be satisfied for all SNR only when $3\alpha - 2 \leq 0$ which gives $\alpha \leq 2/3$.

For the achievability, sum capacity can be achieved for a fully-connected interference network since every user knows the global network state. For one-to-many network, the result in [24] gives that all users except the first using rate $(\log(\mathsf{SNR}_i))^+$ and the first user backing off will achieve a sum rate within $3K - 2$ bits of the sum capacity.

Chapter 5

Conclusions

As the use of wireless technology is expanding, major assumptions in the design of the network need to be rethought. The knowledge about the network flows by some message passing protocols in actual systems, and the rounds of messaging between nodes cannot be very large. In this study, we provide new protocols which are based on the implemented protocols and analyze the performance with partial messages. This leads to a new study of system designs with incomplete or imperfect information at the terminals.

In this thesis, we formulated two models for channel state feedback, and derived the diversity tradeoff in which the errors in channel estimation and the feedback channel are fully accounted. The two models considered in this thesis cover all the cases of the correlations between the forward and the backward channel such that correlation coefficient is independent of signal-to-noise ratio (This is because any constant correlation, $|\rho| < 1$, will make the two channels look independent in high SNR regime). This thesis finds the diversity multiplexing tradeoffs when the receiver and the transmitter are allowed to exchange multiple rounds of messages. The diversity multiplexing tradeoff increases with each round of message passing between the nodes. This thesis gives a way of resolving errors with more rounds of message passing in a system where the errors in training and feedback are considered.

We then model the notion of incomplete information about the network. From a network design perspective, the location of the destination is important to find the location of a fixed base station and the relay station. Since the users are mobile, practical designs rely on the placement of nodes which maximize the region in which coverage can be provided. Thus, we introduced a new perspective on the relay channel, switching from maximizing rate at a fixed given destination location, to maximizing coverage for a given rate. This perspective opens up an array of new possibilities for research. We have obtained the surprising result that for any given placement of the relay, one of the two common strategies (CF and DF) is uniformly optimal, and thus a relay that switches between the two is not required. The comparison is extended for the case of $\alpha = 2$ by allowing each of the strategies its own preferred relay location. In this case, it is interesting that the results depend on the power constraint on the relay. A natural question that arises, when considering coverage regions, is the area of the region, as a function of the distance between the source and the relay. In this thesis, we have provided bounds in two special cases.

Finally, we extend the concept of incomplete knowledge to large networks, in specific interference networks. We have given a framework for optimality of distributed decisions. The optimality is measured in terms of normalized sum-capacity which is the best worst case guarantee of the distributed decisions. We gave an achievable algorithm called maximal independent graph scheduling, and characterized its performance in several examples. This algorithm reduces to a fractional coloring algorithm for the case when there is very limited information at the nodes, which has been studied in the scheduling literature. In this thesis, we give the first information-theoretic optimality of such mechanisms. The thesis extend these concepts to novel strategies that can be used with more information at the nodes. Even though this algorithm achieves normalized sum-capacity in several cases, we also show that this algorithm is not always optimal.

In separate work, we have also studied the effect of incomplete information on security [9, 17, 29]. In [29], we studied the effect on a graph when the location of eavesdroppers

are not precisely known and the existence of nodes are known only up to a limited distance. Knowledge of more precise information leads to better decisions and thus better performance. This holds not only in wireless networks, but also has been studied in many different fields. For example, finite rate of innovation is used to compress ECG signals because there is more information about the structure of the signals. Recently, compressed sensing has come up as a field which uses a particular knowledge in the structure of signals for better compression. Asymmetric information is also studied in game theory and economics; this topic won a Nobel Prize in 2001. This tells that the concept is widely applicable and this thesis applies the theme to networks and tells some first order effects that incomplete or imperfect information can have on system design and the design of strategies. Better knowledge leads to better performance, but more overhead is needed to learn more. A system designer can then use the tradeoff to choose the amount of learning needed and the protocols that need to be used with the obtained knowledge. I believe that understanding the effect of imperfect and incomplete information along the lines suggested in this thesis will have a great impact on next generation wireless system designs.

Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. Courier Dover Publications, 1995.

[2] V. Aggarwal, S. Avestimehr, and A. Sabharwal, "Distributed universally optimal strategies for interference channels with partial message passing," *in Proc. Allerton Conference on Communication, Control, and Computing*, Sept-Oct 2009.

[3] ——, "How (information theoretically) optimal are distributed decisions?" *in Proc. IEEE Conference on Information Sciences and Systems*, Mar 2010.

[4] ——, "Normalized sum-capacity of interference networks with partial information," *in Proc. IEEE International Symposium on Information Theory*, June 2010.

[5] ——, "On achieving local view capacity via maximal independent graph scheduling," *submitted to Special Issue of the IEEE Transactions on Information Theory on Interference Networks*, Apr 2010.

[6] V. Aggarwal, A. Bennatan, and A. R. Calderbank, "On maximizing coverage in Gaussian relay networks," pp. 1–5, Jul 2007.

[7] ——, "On maximizing coverage in Gaussian relay channels," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2518–2536, June 2009.

[8] V. Aggarwal, G. Krishna, S. Bhashyam, and A. Sabharwal, "Two models for noisy feedback in MIMO channels," *in Proc. Asilomar Conference on Signals, Systems and Computers*, Oct 2008.

[9] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor, "Wiretap channel type II with an active eavesdropper," *in Proc. IEEE International Symposium on Information Theory*, pp. 1944–1948, Jun-Jul 2009.

[10] V. Aggarwal, Y. Liu, and A. Sabharwal, "Message passing in distributed wireless networks," *in Proc. IEEE International Symposium on Information Theory*, pp. 1090–1094, Jun-Jul 2009.

[11] ——, "Sum-capacity of interference channels with a local view: Impact of distributed decisions," *submitted to IEEE Transactions on Information Theory*, Oct 2009, available at arXiv:0910.3494v1.

[12] V. Aggarwal and A. Sabharwal, "Diversity order gain with noisy feedback in multiple access channels," *in Proc. IEEE International Symposium on Information Theory*, pp. 404–408, Jul 2008.

[13] ——, "On multiple access channels with asymmetric feedback," *in Proc. IEEE International Symposium on Information Theory*, pp. 399–403, Jul 2008.

[14] ——, "Performance of multiple access channels with asymmetric feedback," *IEEE Journal on Selected Areas in Communication*, vol. 26, no. 8, pp. 1516–1525, Oct 2008.

[15] ——, "Bits about the channel: Multi-round protocols for two-way fading channels," *submited to IEEE Transactions on Information Theory*, Sept 2009.

[16] ——, "Power-controlled training and feedback for two-way MIMO channels," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3310–3331, July 2010.

[17] V. Aggarwal, L. Sankar, A. R. Calderbank, and H. V. Poor, "Secrecy capacity of a class of orthogonal relay eavesdropper channels," *EURASIP Journal on Wireless Communications and Networking*, 2009.

[18] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *submitted to IEEE Transactions on Information Theory*, Aug 2009, available at arXiv:0906.5394v2.

[19] C. M. Bender, D. C. Brody, and B. K. Meister, "On powers of bessel functions," *Journal of Mathematical Physics*, vol. 44, no. 1, pp. 309–314, Jan 2003.

[20] S. Bhashyam, A. Sabharwal, and B. Aazhang, "Feedback gain in multiple antenna systems," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 785–798, May 2002.

[21] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance for block-fading channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[22] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, Mar 2006.

[23] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Annals of Mathematical Statistics*, vol. 30, pp. 1229–1241, Dec 1959.

[24] G. Bresler, A. Parekh, and D. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *submitted to IEEE Transactions on Information Theory*, Sept 2008, available at arXiv:0809.3554v1.

[25] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2007–2019, Sept 1999.

[26] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, Sept 1979.

[27] S. Ekbatani, F. Etemadi, and H. Jafarkhani, "Outage behavior of slow fading channels with power control using noisy quantized CSIT," *arXiv:0804.0790v1*, Apr 2008.

[28] H. E. Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Transactions on Information Theory*, vol. 52, pp. 3601–3621, Aug 2006.

[29] S. Goel, V. Aggarwal, A. Yener, and A. R. Calderbank, "Modeling location uncertainty for eavesdroppers: A secrecy graph approach," *in Preparation*.

[30] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov 1997.

[31] A. H.-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2020–2040, Jun 2005.

[32] B. Hajek and G. Sasaki, "Link scheduling in polynomial time," *IEEE Transactions on Information Theory*, vol. 34, pp. 910–917, May 1988.

[33] A. Khoshnevis and A. Sabharwal, "Achievable diversity and multiplexing in multiple antenna systems with quantized power control," *in Proc. IEEE International Conference on Communications*, May 2005.

[34] ——, "On the asymptotic performance of multiple antenna channels with quantized feedback," *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, pp. 3869–3877, Oct 2008.

[35] T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1618–1626, Apr 2009.

[36] T. Kim and M. Skoglund, "Diversity-multiplexing tradeoff in MIMO channels with partial CSIT," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2743–2759, Aug 2007.

[37] A. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Transactions on Information Theory*, vol. 15, no. 1, pp. 52–60, Jan 1969.

[38] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, sept. 2005.

[39] G. G. Krishna, "Feedback with resource accounting in MIMO systems," *B. Tech. Project Report - Indian Institute of Technology Madras*, May 2008.

[40] G. Krishna, S. Bhashyam, and A. Sabharwal, "Decentralized power control with two-way training for multiple access," *in Proc. IEEE International Symposium on Information Theory*, pp. 609–613, 2008.

[41] M. Kubale, *Graph Colorings*. American Mathematical Society, 2004.

[42] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec 2004.

[43] S. Lavenberg, "Feedback communication using orthogonal signals," *IEEE Transactions on Information Theory*, vol. 15, no. 4, pp. 478–483, Jul 1969.

[44] L. Meng, A. Zipf, and S. Winter, *Map-based mobile services: Design, interaction, and usability*. Springer, 2008.

[45] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, Oct 2003.

[46] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.

[47] A. Raja, V. Prabhakaran, and P. Viswanath, "The two user Gaussian compound interference channel," *in Proc. IEEE International Symposium on Information Theory*, pp. 569–573, Jun 2008.

[48] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 786–788, Nov 1981.

[49] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, Apr 1966.

[50] V. Sharma, K. Premkumar, and R. Swamy, "Exponential diversity achieving spatio-temporal power allocation scheme for fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 188–208, Jan 2008.

[51] W.-Y. Shin, S.-Y. Chung, and Y. Lee, "Diversity-multiplexing tradeoff and outage performance for rician MIMO channels," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1186–1196, Mar 2008.

[52] C. Steger and A. Sabharwal, "Single-input two-way simo channel: diversity-multiplexing tradeoff with two-way training," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4877–4885, Dec 2008.

[53] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[54] J.-P. Tignol, *Galois theory of algebraic equations*. Longman Scientific and Technical Press, 1998.

[55] E. van der Meulen, "Ttransmission of information in a t-terminal discrete memoryless channel," *Ph.D. dissertation - UC Berkeley*, Jun 1968.

[56] ——, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, pp. 120–154, 1971.

[57] ——, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 1–37, Jan 1977.

[58] S. Wang and A. Abdi, "Joint singular value distribution of two correlated rectangular Gaussian matrices and its application," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 3, pp. 972–981, Oct 2007.

[59] L. Zheng and D. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.

[60] ——, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.