The Dissertation Committee for Omar El Ayach
certifies that this is the approved version of the following dissertation:

Interference Alignment from Theory to Practice

Committee:

Robert W. Heath, Jr., Supervisor

Jeffrey G. Andrews

Angel Lozano

Sujay Sanghavi

Sanjay Shakkottai

Interference Alignment from Theory to Practice

by

Omar El Ayach, B.E., M.S.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2013

Dedicated to my family,

Marwan, Rida, Maya, and Abdallah.

Acknowledgments

I can only start by genuinely thanking my Ph.D. supervisor Prof. Robert W. Heath, Jr. for taking a chance on me back in 2008, and believing in me ever since. I don't know if I'd be writing this dissertation now in 2013 if not for his unwavering support, his unrelenting work ethic, and his excellent technical guidance. I've learned a lot from Robert, and believe I could still stand to learn a thing or two. I am very fortunate and proud to have worked with him.

I owe a great debt of gratitude to my UT professors and committee members, Prof. Jeffrey Andrews, Prof. Sanjay Shakkottai, and Prof. Sujay Sanghavi for some of the most intellectually stimulating classes at UT and for helping make WNCG the great research center it is today. I'm extremely grateful for my committee member and co-author, Prof. Angel Lozano, for hosting me in Barcelona in the spring of 2011 and for inspiring some of the work in this dissertation. It's safe to say that collaborating with Angel has truly transformed my approach to research and engineering.

It has been a pleasure to work with all my diverse and gifted, current and former, lab-mates at the WSIL: Robert Daniels, Takao Inoue, Rahul Vaze, Randa Zakhour, Namyoon Lee, Ahmad Alkhateeb, Tianyang Bai, Chao Chen, Insoo Hwang, Talha Khan, Jackson Massey, Ken Pesyna, Andrew Thornburg, Chan-Byoung Chae, and Nachiappan Valliappan. A special mention is due

to Salam Akoum, Amin Abdel Kalek, Behrang Nosrat Makouei, Ramya Bhagavatula, Steven Peters, Jonathan Starr and Kien Trung Truong. I'm very fortunate to have met my awesome WNCG friends Zrinka Puljiz, Siddhartha Banerjee, Aditya Gopalan, Harpreet Dhillon, Kumar Appaiah, Vikram Chandrasekhar and Marios Kountouris.

My time at UT would not have been the same without the rest of the notorious Lebanese gang: Marcel Nassar, Samer Chucri, Michele Saad, Bassem ElKarablieh, Shadi Abdel Khalek, and Mohammad Fakhreddine. A very special shout-out to my precious buddy, Sarah Fakhreddine, for making my time in Austin more comfortable, and that much more colorful. Back home, no words of thanks can do justice to my life-long friends, Bachar Hamade, Elie Alam, and Hisham Zebian.

Last, but most certainly not least, I am forever indebted to my family and would like thank them for believing and taking pride in me. I want to thank my mother for her unconditional love and for teaching me the value of education and perseverance. Everything I've ever done can be attributed, in no small part, to my father's endless self-sacrifice and steadfast support. I thank my dad for our first exciting encounters with technology, computers and networks that went on to shape my professional life. I'm sure he and I can perfectly recite the old nostalgic sounds of dial-up Internet and perfectly remember our first U.S. Robotics modem. Finally, I thank my sister and brother, Maya and Abdallah, with all my heart for making our family all the more wonderful.

Interference Alignment from Theory to Practice

Publication No. _____

Omar El Ayach, Ph.D.
The University of Texas at Austin, 2013

Supervisor: Robert W. Heath, Jr.

Wireless systems in which multiple users simultaneously access the propagation medium suffer from co-channel interference. Untreated interference limits the total amount of data that can be communicated reliably across the wireless links. If interfering users allocate a portion of the system's resources for information exchange and coordination, the effect of interference can be mitigated. Interference alignment (IA) is an example of a cooperative signaling strategy that alleviates the problem of co-channel interference and promises large gains in spectral efficiency. To enable alignment in practical wireless systems, channel state information (CSI) must be shared both efficiently and accurately. In this dissertation, I develop low-overhead CSI feedback strategies that help networks realize the information-theoretic performance of IA and facilitate its adoption in practical systems. The developed strategies leverage the concepts of analog, digital, and differential feedback to provide IA networks with significantly more accurate and affordable CSI when compared to existing solutions. In my first contribution, I develop an analog

feedback strategy to enable IA in multiple antenna systems; multiple antennas are one of IA's key enabling technologies and perhaps the most promising IA use case. In my second contribution, I leverage temporal correlation to improve CSI quantization in limited feedback single-antenna systems. The Grassmannian differential strategy developed provides several orders of magnitude in CSI compression and ensures almost-perfect IA performance in various fading scenarios. In my final contribution, I complete my practical treatment of IA by revisiting its performance when CSI acquisition overhead is explicitly accounted for. This last contribution settles the viability of IA, from a CSI acquisition perspective, and demonstrates the utility of the proposed feedback strategies in transitioning interference alignment from theory to practice.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Interference alignment (IA) is a revolutionary wireless transmission strategy that reduces the impact of interference. The idea of interference alignment is to coordinate multiple transmitters so that their mutual interference aligns at the receivers, facilitating simple interference cancellation techniques. Since interference alignments inception, researchers have investigated its performance and proposed several improvements. Past research usually focused on verifying interference alignments ability to achieve the maximum degrees of freedom (an approximation of sum capacity), developing algorithms for determining alignment solutions, and designing transmission strategies that relax the need for perfect alignment but yield better performance. This chapter provides an overview of the concept of interference alignment as well as an assessment of practical issues including performance in realistic propagation environments, the role of channel state information at the transmitter, and the practicality of interference alignment in large networks.

1.1 Introduction

Interference is a major impairment to successful communication in commercial and military wireless systems. In cellular systems, interference is created when different base stations share the same carrier frequency due to frequency reuse. Inter-cell interference reduces data rates throughout the cells and causes outages at the cell edges. In local area networks, interference is created when different access points share the same channel. The medium access control protocol attempts to deal with interference by avoiding packet collisions (overlapping transmissions). This conservative approach leads to an under-utilization of system bandwidth. Similarly, neighboring nodes in a dense mobile ad hoc networks interfere if they share the same time and frequency resources. The medium access control protocol again limits the number of simultaneous conversations and consequently the system performance. Interference is thus a critical impairment in most wireless systems.

Communication in the presence of interference is often analyzed using an abstraction known as the interference channel. In the example interference channel of Fig. 1.1, three different transmitters wish to communicate with three receivers. Each transmitter has a message only for its paired receiver. Assuming the transmitters share the same time and frequency resources, each transmission creates interference at the unintended receivers. There may be other sources of interference, not illustrated, such as jamming in military networks, or self-interference created from nonlinearities in the radio frequency components; those sources are not captured in the basic interference channel.

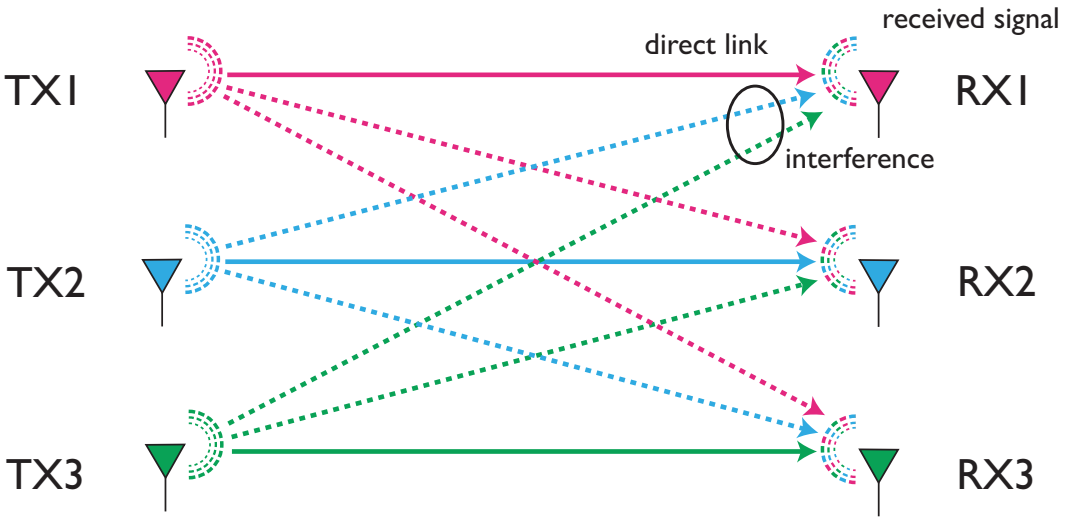

Figure 1.1: Illustration of a 3 transmit/receive pair interference channel. Each transmitter creates a signal that is interpreted as interference by its unintended receiver.

Traditional methods for dealing with interference often revolve around giving each user exclusive access to a fraction of the communication resources. In frequency division multiple access (FDMA), the system bandwidth is divided among the transmitters, e.g. in Fig. 1.1 each transmitter would be given a third of the total bandwidth. In time division multiple access (TDMA), users take turns transmitting on a periodic set of transmission intervals. In random access systems, e.g. carrier sense multiple access, transmitters listen to see if the channel is available and then transmit if it is. Random access protocols are typically much less efficient than preassigned orthogonal access like FDMA or TDMA since the spectrum may not be fully utilized and collisions may occur. Regardless of the access protocol, the unifying concept remains avoiding interference by limiting the number of overlapping transmissions. If simultaneous

transmissions are allowed, the resulting interference is often treated as noise and a loss in data rate ensues.

A new concept for communication in interference channels, called interference alignment (IA), was proposed in [8, 58]. IA is a cooperative interference management strategy that exploits the availability of multiple signaling dimensions provided by multiple time slots, frequency blocks, or antennas. The transmitters jointly design their transmitted signals in the multi-dimensional space such that the interference observed at the receivers occupies only a portion of the full signaling space. An amazing result from [8] is that alignment may allow the network's sum data rate to grow linearly, and without bound, with the network's size. This is in sharp contrast with orthogonal access strategies like FDMA or TDMA where sum rate is more or less constant since, regardless of network size, only one pair of users can communicate in a given time/frequency block. Since the development of interference alignment, there has been work on a variety of topics to understand its theory and potential applications. A good summary of key results is given in [35].

The objective of this chapter is to review interference alignment with a focus on making it practical. First I describe IA in more detail, summarizing relevant practical issues that I believe are critical for its successful deployment. Then I review different techniques for computing alignment solutions, and more general interference management solutions. Because computing these solutions relies heavily on channel state information (CSI), I describe the two competing CSI acquisition techniques, reciprocity and feedback, and focus on

their respective benefits and limitations. I highlight the fact that the dimensions needed for alignment, and the overhead of CSI acquisition, both rapidly increase with network size. This places a limit on the gains achievable via IA in large networks. For this reason I summarize work on partial connectivity and user clustering which leverages network topology information to reduce the requirements of alignment. I conclude with a discussion of the research challenges that remain in realizing interference alignment in practice.

Note that the objective of this chapter is to provide a high level introduction to the linear precoding type of interference alignment, which exploits channel state information for all users. Deeper discussions of other forms of interference alignment, including blind alignment, are found in [35].

1.2 The Interference Alignment Concept

Interference alignment in its simplest form is a precoding technique for the interference channel. It is a transmission strategy that linearly encodes signals over multiple dimensions such as time slots, frequency blocks or antennas. By coding over multiple dimensions, transmissions are designed to consolidate, i.e. align, the interfering signals observed at each receiver into a low dimensional subspace. By doing so, interference alignment maximizes the number of non-interfering symbols that can be simultaneously communicated over the interference channel, otherwise known as the multiplexing gain. Interestingly, achieving the channel's maximum multiplexing gain, also known as degrees of freedom, implies that the sum rates provided by IA can approach

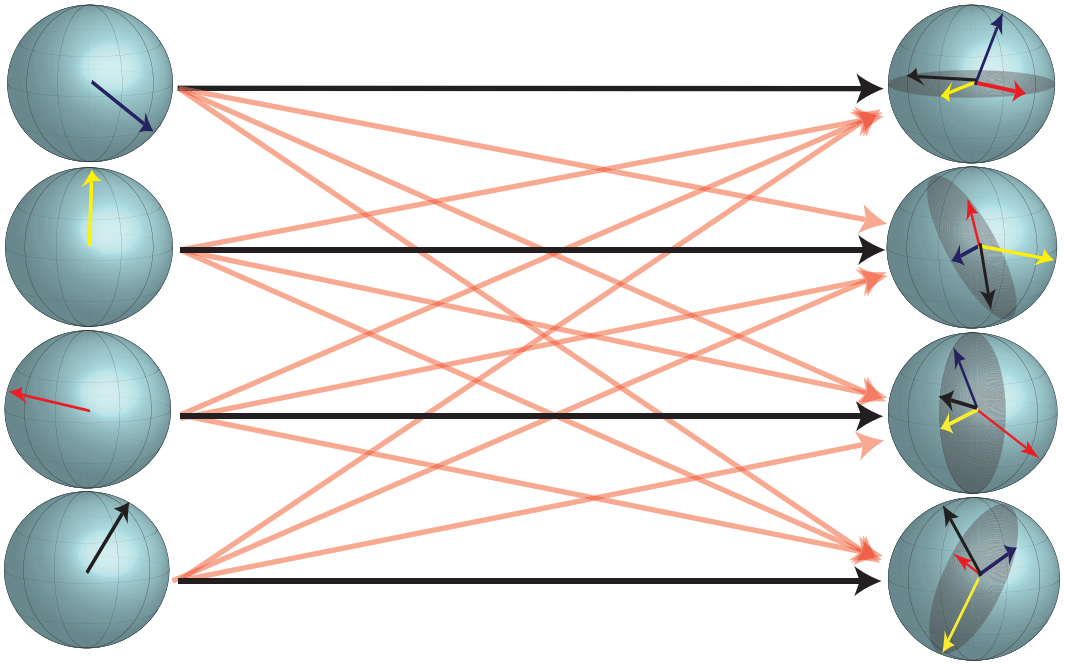

Figure 1.2: A diagram illustrating interference alignment. At each receiver, three interferers collapse to appear as two. This enables interference free decoding in a desired signal subspace.

sum capacity at high signal-to-noise ratio (SNR).

To illustrate the IA concept, consider the four user system of Fig. 1.2 where real valued signals are coded over three dimensions and communicated over real valued channels. In this four user system, a receiver observes a total of three interference signals, each of which is represented as a vector in the real 3D space. Without careful structure, the three interference signals will occupy all three signal dimensions at the receiver; the signals do not fortuitously align. IA allows users to cooperatively precode their transmissions such that each *three* interference signals are fully contained in a *two* dimensional space. Alignment

thus leaves one dimension in which users can decode their messages free from interference by projecting the received signal onto the subspace orthogonal to the interference.

The IA visualization in Fig. 1.2 translates into an equally intuitive mathematical representation. Consider a K user interference channel in which each user i transmits a set of S_i information symbols encoded in the $S_i \times 1$ vector $\mathbf{s}_i$. Symbols $\mathbf{s}_i$ are precoded using a matrix $\mathbf{F}_i$ and observed at each receiver ℓ after propagating over the matrix channel $\mathbf{H}_{\ell,i}$; the dimensions of $\mathbf{F}_i$ and $\mathbf{H}_{\ell,i}$ will be made clear shortly. For such a system, the received signal at user i is

$$\mathbf{y}_i = \mathbf{H}_{i,i}\mathbf{F}_i\mathbf{s}_i + \sum_{\ell \neq i} \mathbf{H}_{i,\ell}\mathbf{F}_\ell\mathbf{s}_\ell + \mathbf{v}_i,$$

where the vector $\mathbf{v}_i$ is the noise observed at receiver i. When only the time or frequency dimension are used for precoding, for example when coding over a bandwidth of B subcarriers, $\mathbf{H}_{i,\ell}$ is a $B \times B$ diagonal matrix; diagonal since subcarriers are orthogonal. In a MIMO system with N_T transmit antennas and N_R receive antennas, $\mathbf{H}_{i,\ell}$ is $N_\mathrm{R} \times N_\mathrm{T}$ with no special structure.

Regardless of the dimension used to align interference, IA can be summarized as calculating a set of precoders $\mathbf{F}_\ell$ such that any given user, even using a simple linear receiver $\mathbf{W}_i$, can cancel the interference it observes from all other users, i.e., $\mathbf{W}_i^*\mathbf{H}_{i,\ell}\mathbf{F}_\ell = 0 \ \forall \ell \neq i$, without nulling or destroying its desired signal $\mathbf{W}_i^*\mathbf{H}_{i,i}\mathbf{F}_i$. Early work on IA showed that a system's ability to find such IA precoders is directly related to the number of signal dimensions it can code over. Intuitively, the more time slots, frequency blocks, or antennas

available for precoding, the more flexibility a system has in aligning interference. The problem of characterizing the number of coding dimensions needed to ensure the feasibility of IA has been studied extensively [76].

1.3 The Practical Challenges of Interference Alignment

Interference alignment relies on some assumptions which must be relaxed before it is adopted in practical wireless systems. In this section, I briefly discuss some of the most pressing challenges facing the transition of IA from theory to practice.

1.3.1 Dimensionality and Scattering

IA is achieved by coding interference over multiple dimensions. Intuitively, the more interfering signals that need to be aligned, the larger the number of dimensions needed to align them. When using the frequency domain for alignment, prior work has shown that the number of signaling dimensions needed to achieve good IA performance grows *faster than exponentially* with the number of IA users. Properly aligning even as little as four users requires a potentially unreasonable number of subcarriers and a correspondingly large bandwidth [35]. This dimensionality requirement poses a major challenge for IA in practical systems. The dimensionality requirement is relatively milder in the case of MIMO IA where more users can be supported as long as the number of antennas grows linearly with network size [76]. For this reason, IA seems most likely to be implemented in MIMO systems. I thus place a special

focus on MIMO IA.

1.3.2 Signal-to-Noise Ratio

IA is often degrees of freedom optimal, meaning that the sum rates it achieves approach the channel's sum capacity at very high SNR. At *moderate SNR levels*, however, the sum rates resulting from IA may fall short of the theoretical maximum. As a result, IA may be of limited use to systems with moderate SNR unless IA algorithms are further improved. Examples of such algorithms are discussed in Section 1.4.

1.3.3 Channel Estimation and Feedback

Channel state information (CSI), be it at the transmitter or receiver, is central to calculating IA precoders. As a result, sufficient resources must be allocated to pilot transmission, and in some cases to CSI feedback, to ensure the availability of accurate CSI. Since IA precoders must be recalculated when the channel changes appreciably, the overhead of CSI acquisition in high-mobility fast-fading systems can limit the gains of IA. For this reason, low overhead signaling strategies must be devised to properly tradeoff CSI quality with CSI acquisition overhead.

1.3.4 Synchronization

IA via linear precoding is a transmission strategy for the *coherent interference channel*. Thus, IA requires tight synchronization to remove any timing

and carrier frequency offsets between cooperating nodes. In the absence of sufficient synchronization, additional interference terms are introduced to the signal model, rendering the IA solution ineffective. Synchronization strategies that leverage GPS satellite signals could help fulfill this requirement.

1.3.5 Network Organization

Nodes cooperating via IA must not only synchronize, but also negotiate physical layer parameters, share CSI, and potentially self-organize into small alignment clusters if full network alignment proves to be too costly. In the absence of centralized control, distributed network protocols must be redesigned around this more complex and cooperative physical layer.

1.3.6 Further Description of Challenges

The following three sections are devoted to surveying the research efforts invested in overcoming some of the challenges discussed in this section. Section 1.4 presents algorithmic improvements that primarily target IA's low SNR performance, Section 1.5 explores methods of obtaining the CSI required by IA, and Section 1.6 addresses the challenge of applying IA to larger networks.

1.4 Computing Interference Alignment Solutions

There are some simple formulas for calculating IA precoders in some special cases [8]. To enable IA in general network settings, however, researchers

have relied on developing iterative IA algorithms. Since then, the algorithmic focus has largely been on MIMO interference alignment.

The earliest method for finding MIMO IA precoders was the distributed solution presented in [21]. The idea for the algorithm is that at each iteration, users try to minimize the extent to which their signal leaks into the other users' desired signal subspaces. After the algorithm converges, the IA condition $\mathbf{W}_i^* \mathbf{H}_{i,\ell} \mathbf{F}_\ell = 0$ should ideally be satisfied and the desired signal spaces be free of interference. While the algorithm performs well at high SNR, it can be far from optimal in badly conditioned channels or at low SNR. The reason is that while this algorithm properly aligns interference, it is oblivious to what happens to the desired signal power during the process.

The first attempt to improve low SNR performance is the Max-SINR algorithm of [21]. Instead of minimizing interference power at each iteration, the improved algorithm accounts for the desired channel by iteratively maximizing the per-stream signal-to-interference-plus-noise ratio (SINR). By relaxing the need for perfect alignment, Max-SINR outperforms IA at low SNR and matches its performance at high SNR. Other algorithms have also relaxed the perfect alignment requirement and adopted more direct objectives such as maximizing network sum rate. For example, [92] highlights the equivalence between maximizing sum rate and minimizing the signal's sum mean square error and uses the properties of the equivalent minimum mean square error (MMSE) problem to give a simple iterative algorithm in which precoders can be updated in closed form or via traditional convex optimization.

Suboptimal low SNR performance is admittedly not the only limitation of IA and many algorithms have been developed to address various shortcomings. Most early work on IA, for example, focused on the case where all interfering users are able to cooperate. The ability to cooperate, however, is fundamentally limited by constraints such as the number of antennas and cooperation overhead. As a result, large networks will inevitably have uncoordinated interference, or colored noise, which motivates the improved IA algorithms in [75]. Another limitation of precoding for the MIMO interference channel is that it often requires sharing entire channel matrices, potentially incurring a large overhead penalty, as I will discuss in Section 1.5. The work in [91] proposes a distributed transmission strategy inspired by game theory in which both precoders and transmit powers are iteratively adjusted to maximize sum rate by sharing *scalar quantities* known as "interference prices." The strategy in [91] thus replaces channel *matrix* feedback with several iterations of *scalar* feedback. Such a strategy could potentially reduce IA's overhead.

While editorial constraints have limited the discussion to the solutions of [8, 21, 75, 91, 92], a large number of noteworthy solutions have been developed to improve upon IA performance. The interested reader is encouraged to see the extensive list of IA-inspired algorithms and results in [1].

1.5 Obtaining CSI in the Interference Channel

Calculating IA precoders requires accurate knowledge of the interference generated by each transmitter. The premise is that if a transmitter knows

the "geometry" of the interference it creates, it can conceivably shape it to mitigate its effect. Two methods of obtaining this knowledge are reciprocity and feedback.

1.5.1 Interference Alignment via Reciprocity

In time division duplexed systems where transmissions on the forward and reverse links overlap in frequency and are minimally separated in time, propagation in both directions will be identical. In such systems, channels are said to be reciprocal. Reciprocity enables IA by allowing transmitters to infer the structure of the interference they *cause* by observing the interference they *receive*.

Consider for example the IA strategy proposed in [21] in which the precoders $\mathbf{F}_i$ and the combiners $\mathbf{W}_i$ are updated iteratively to reduce the power of uncanceled interference $\mathbf{W}_i\mathbf{H}_{i,\ell}\mathbf{F}_\ell$ to zero. To do so, transmitters begin by sending pilot data using an initial set of precoders. Receivers then estimate their interference covariance matrix and construct combiners that select the receive subspaces with least interference. When the roles of transmitter and receiver switch, thanks to reciprocity, that same *receive subspace* that carried the least interference becomes the *transmit direction* which causes the least interference on the reverse link. Iterating this subspace selection on the forward and reverse links results in a set of precoders satisfying the IA conditions.

While [21] at each iteration chooses the subspace with least interference, a variety of more sophisticated objectives may be considered. For example [90]

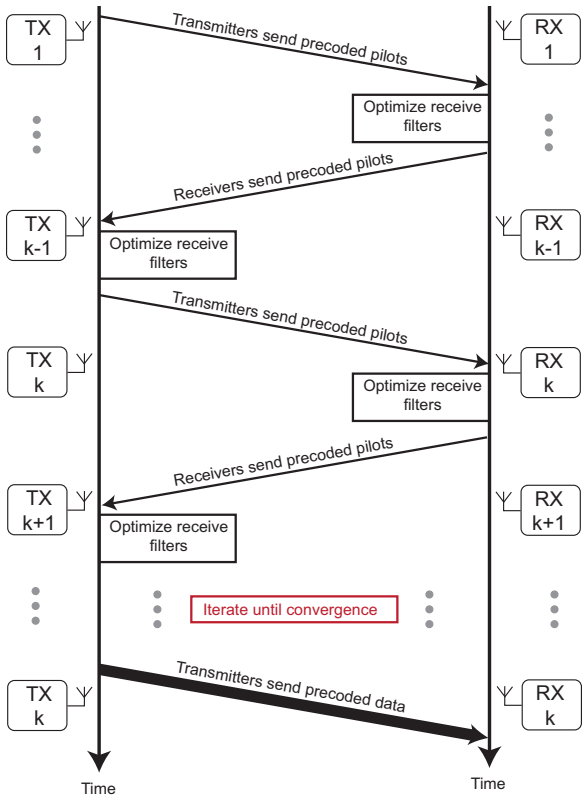

Figure 1.3: Reciprocity-based interference alignment strategies which infer the forward channel from reverse link pilots.

considers an update rule that minimizes the signal's mean square error, resulting in improved sum rate performance. Regardless of the subspace selection rule, the general framework for precoding with reciprocity is shown in Fig. 1.3 and proceeds as follows:

1. *Forward link training*: Transmitters send precoded pilot symbols using a set of initial precoders. Receivers estimate forward channel parameters and compute combiners that optimize a predefined objective.

2. *Reverse link training*: Receivers send precoded pilot symbols using the combiners from step 1 as transmit precoders. Transmitters in turn optimize their combiners/precoders and initiate a second training phase with the updated precoders.

3. *Iterations*: Communicating pairs iterate the previous steps until convergence.

4. *Data transmission*: Payload data is then communicated.

Relying on reciprocity for precoding over the interference channel has a number of potential drawbacks. First, iterating over the air incurs a non-negligible overhead due to the recurring pilot transmissions. While the results in [90] consider pilot overhead, more work is needed to settle the viability of reciprocity. Second, reciprocity may not suffice for all IA-based algorithms. For example, one of the algorithms in [75] attempts to improve IA by considering sources of uncoordinated interference. Since the uncoordinated interference observed by the transmitters and receivers is not "reciprocal", reciprocity cannot be used. Third, reciprocity does not hold in frequency duplexed systems and ensuring reciprocity with time duplexing requires tightly calibrated RF devices.

1.5.2 Interference Alignment with Feedback

Several IA results have also considered systems with CSI feedback. A general feedback framework is shown in Figure 1.4. In such a system, the

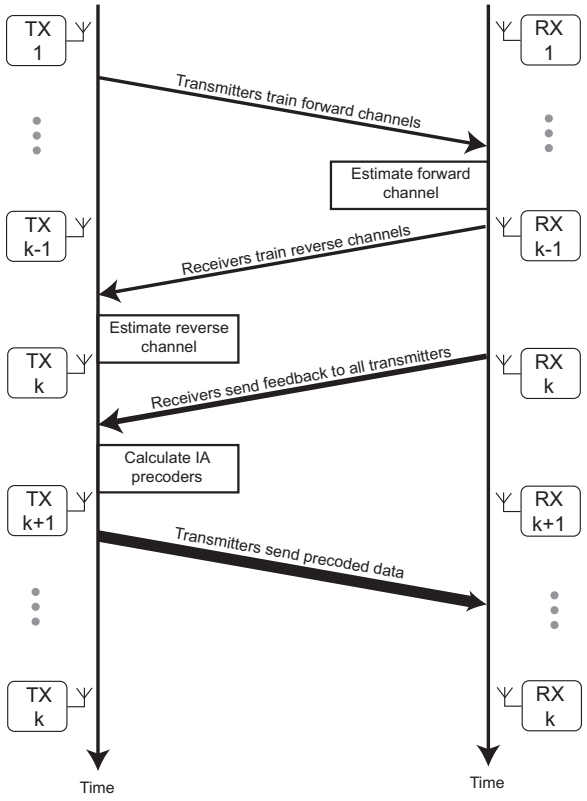

Figure 1.4: Feedback-based interference alignment strategies which rely on explicitly communicating forward channel information to the transmitters.

transmitters first send training sequences with which receivers estimate the forward channels. Receivers then feedback information about the estimated forward channels, potentially after training the reverse link. After feedback, the transmitters have the information needed to calculate IA precoders. Feedback, however, invariably introduces distortion to the CSI at the transmitters and incurs a non-negligible overhead penalty. Therefore, the difficulty lies in designing low-overhead low-distortion feedback mechanisms for IA.

An established method to provide high quality feedback with low over-

head is limited feedback, i.e., channel state quantization. Limited feedback was first considered in [7] for single antenna systems where alignment is achieved by coding over OFDM subcarriers. The feedback strategy in [7] leverages the fact that IA solutions remain unchanged if channels are scaled or rotated which allows efficient quantization via what is known as Grassmannian codebooks. Unfortunately, maintaining proper alignment requires that the accuracy of quantized CSI improve with SNR which in turn implies that the quantization codebook size must scale *exponentially* with SNR. The complexity of quantized feedback, however, increases with codebook size and large Grassmannian codebooks are difficult to design and encode. Furthermore, this strategy relies on the CSI's Grassmannian structure for efficient quantization. As a result, it cannot be applied to systems where the CSI exhibits no special structure. This alienates a main case of interest, namely IA in multiple antenna systems where the CSI to be fed back is the set of channel matrices.

While practical, feedback poses a number of challenges. Not only does the required quality of CSI scale with SNR, but the required quantity also scales super-linearly with network size. If feedback is inefficient, IA's overhead could dwarf its promised theoretical gains. Second, IA relies on sharing CSI with interfering transmitters: multicell systems that share transmitter CSI over a backhaul may suffer from queuing delays that render CSI obsolete on arrival. While work on addressing these issues is underway, most research still considers systems with a limited number of users. In large scale networks, however, practical challenges are amplified many-fold.

1.6 Interference Alignment in Large Scale Networks

Consider applying IA, or any other interference channel precoding strategy, to a large scale network, such as a regional cellular network. In this setting, the need for signaling dimensions grows, the overhead of CSI acquisition explodes, and the task of synchronization becomes daunting. This grim outlook, however, is a byproduct of a naive over-generalization of the basic interference channel. In large scale networks not all interfering links are significant and thus not all of them should be treated equally.

To examine the feasibility of IA in large networks, [26] considered a large network wherein each user receives interference from a finite subset of users, e.g., the first tier of interferers in a grid-like cellular system. Using this model, [26] showed that the number of antennas needed for network-wide alignment is a linear function of the size of the interfering subset rather than the size of the entire network. This implies that perfect alignment in an *infinitely* large network is theoretically possible with a *finite* number of antennas. This partially connected model, however, is ultimately a simplification of network interference. While some interferers may indeed be too weak to merit alignment, that threshold remains unclear. Moreover, neglecting interferers that fall below that threshold, as is implicitly done in the partially connected model, may not be optimal. Exploring the optimal transmission strategy in this gray area between full connectivity and true partial connectivity is the focus of [74].

The work in [74] considers a wireless network and assigns a finite chan-

nel coherence time and different pathloss constants to each link. This setup allows [74] to gauge the gain from aligning a set of users versus the CSI acquistion overhead involved. On the one hand, it is noticed that the large coherence time in static channels dilutes the cost of CSI and enables alignment over the whole network. On the other hand, fast fading channels do not allow enough time to acquire the CSI needed for IA. In this latter case, systems perform better with lower overhead strategies such as TDMA. In between the two extremes, it is noticed that systems benefit most from a hybrid IA/TDMA strategy wherein smaller subsets of users cooperate via alignment, and TDMA is applied across groups. The objective then becomes finding an optimal user grouping, or network partitioning, for which [74] provides various algorithms based on geographic partitioning, approximate sum rate maximization, and sum rate maximization with fairness constraints. In all grouping solutions only long-term pathloss information is used which further reduces the overhead of IA by avoiding frequent regrouping.

The IA research community has made successful initial attempts at demonstrating that the reach of IA can extend well beyond the information theoretic interference channel through work such as [26, 75] and many others. Perhaps the most crucial next step towards realizing IA gains is to rethink the upper layers above IA in the communication stack, e.g., the medium access layer. These protocols have traditionally been built around point-to-point physical layers and must be redesigned to enable more cooperative multi-user physical layers instead. Preliminary work on this front is already in progress

through the prototyping work in [51] which extends the earlier prototyping efforts by the same authors.

1.7 Summary of Contributions

In this dissertation, I develop low overhead CSI feedback strategies that help networks realize the information-theoretic performance of IA and facilitate its adoption in practical wireless systems. The proposed research leverages the concepts of analog, digital, and differential feedback to provide IA networks with significantly faster and more accurate CSI than prior work. The contributions presented in this dissertation constitute a strong set of candidate feedback solutions that result in a favorable overhead-aware assessment of IA, thus moving it from theory to practice. The main contributions of this dissertation can be summarized as follows.

1. Chapter 2: MIMO Interference Alignment with Analog CSI Feedback

   In this chapter, I develop a feedback strategy for the multiple-input multiple-output interference channel wherein receivers feedback information about the forward channel matrix entries in an analog fashion. The strategy enables the transmitters to calculate IA precoders in a centralized or distributed manner.

   (a) I consider narrow band MIMO channels and propose a two-stage CSI acquisition process wherein receivers first train the feedback

channel and then transmit the forward channel matrices in an analog fashion. The analog feedback strategy enables the transmitters to calculate IA precoders in a centralized or distributed manner. Analog feedback helps relax the reliance of CSI manifold structure and avoids the process of CSI quantization which becomes increasingly costly at high SNR.

(b) I characterize the performance of IA in systems the acquire CSI via analog feedback. I show that the degrees-of-freedom achieved with perfect channel knowledge is preserved by analog feedback and that the mean loss in sum-rate is bounded by a constant when signal-to-noise ratio is comparable in both forward and feedback channels. When signal-to-noise ratios are not quite symmetric, a fraction of the multiplexing gain is achieved.

(c) Based on the sum rate lower bounds derived in this chapter, I provide a numerical characterization of IA's degrees-of-freedom with overhead and give preliminary results on overhead scaling behavior.

2. Chapter 3: Interference Alignment with Grassmannian Differential Limited Feedback

   In this chapter, I consider single-antenna IA networks in which channel's are temporally correlated. I leverage temporal correlation to compress CSI feedback via a concept known as differential or predictive feedback.

(a) I propose a Grassmannian differential feedback framework for the single-antenna frequency-selective interference channel. The proposed framework exploits both the CSI's Grassmannian structure and the channel's temporal correlation to improve quantization accuracy and reduce the required quantization codebook size.

(b) I design quantization codebooks for the proposed differential feedback framework. The codebooks are designed to (i) efficiently fill the quantization space of interest and (ii) to adapt to the long-term dynamics of the time-varying channel thus providing better quantization accuracy.

(c) I characterize the performance of the proposed feedback strategy both analytically and numerically as a function of channel length, mobility, and the number of feedback bits. The performance characterization includes results on both average quantizer distortion and IA sum-rate.

3. Chapter 4: The Overhead of Interference Alignment

In this chapter, I analytically characterize the effective sum-rate achieved in practical MIMO interference channels in which CSI is both imperfect and accompanied with a non-negligible cost of overhead signaling.

(a) I study the performance of IA in multiple-input multiple-output that are optimized for perfect CSI operation, yet only has access to imperfect CSI through training and analog feedback. As opposed

to using the lower bounds on IA sum-rate developed in Chapter 2, I derive tractable expressions for IA sum-rate in both genie-aided systems, and systems with imperfect CSI.

(b) I analytically characterize effective sum-rate with overhead in relation to various parameters such as signal-to-noise ratio, Doppler spread, and feedback channel quality. I show that by properly designing the CSI acquisition process, IA can provide good sum-rate performance in a wide range of fading scenarios even when overhead is considered.

(c) I demonstrate how such overhead-aware analysis can help solve a number of practical network design problems. As an example on of overhead-aware network design, I solve the problem of finding the optimal number of cooperative IA users as a function of system parameters such as signal power and mobility.

1.8 Notation

Throughout this dissertation I use the following notation: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, and $(a_1, \ldots, a_k)$ is an ordered set. $\mathbf{A}^*$ and $\mathbf{a}^*$ denote the conjugate transpose of $\mathbf{A}$ and $\mathbf{a}$ respectively, $\mathbf{A} \circ \mathbf{B}$ is the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$, and $[\mathbf{a}, \mathbf{b}]$ is a horizontal concatenation of the vectors $\mathbf{a}$ and $\mathbf{b}$. $\|\mathbf{A}\|_F$ is the Frobenius norm of $\mathbf{A}$ and trace($\mathbf{A}$) is its trace, $\|\mathbf{a}\|_p$ is the p-norm of $\mathbf{a}$, and $|a|$ is the absolute value of a. $\mathbf{I}_N$ is the $N \times N$ identity matrix, 0_N is the N-dimensional zero vector, $\boldsymbol{\mathcal{F}}_N$ is the N-point discrete Fourier trans-

form (DFT) matrix, and diag($\mathbf{a}$) is the diagonal matrix obtained by putting the elements of $\mathbf{a}$ on its diagonal. $\mathbb{C}^N$ is the N-dimensional complex space, $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ is a complex Gaussian random vector with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$. The operator $\mathbb{E}[\cdot]$ denotes expectation. Other notation may be defined where needed.

Chapter 2

Interference Alignment with Analog CSI Feedback

Interference alignment (IA) is, in theory, degrees-of-freedom optimal in a variety of interference channel scenarios. While the multiplexing gain achieved with IA is much higher than previously thought possible, the improvement comes at the cost of requiring accurate and global channel state information at the transmitters. This can be achieved by explicit feedback, a flexible yet potentially costly approach that incurs large overhead. In this chapter, I propose analog feedback as an alternative to limited feedback or reciprocity based alignment. I show that the full multiplexing gain observed with perfect channel knowledge is preserved by analog feedback and that the mean loss in sum-rate is bounded by a constant when signal-to-noise ratio is comparable in both forward and feedback channels. When signal-to-noise ratios are not quite symmetric, a fraction of the multiplexing gain is achieved. I include a preliminary numerical study on the effective throughput of IA when the overhead of analog feedback is considered. I present simulation results to demonstrate the performance of IA with analog feedback, verify the theoretical analysis, and extend the conclusions on optimal training and feedback length.

2.1 Introduction

The interference channel is an information theoretic concept that models wireless networks in which several transmitters simultaneously communicate data to their paired receivers. The traditional approach for communication in such channels was orthogonalization, where resources are split among the various transceiver pairs. Unfortunately, this results in decaying per-user data rates as networks grow. Recent work in information theory, however, has shown that the interference channel can support sum-rates that, under certain assumptions, scale linearly with the number of users at high signal-to-noise ratios (SNR). This linear sum-rate scaling can be achieved by a transmission strategy known as interference alignment (IA) [8].

Interference alignment is a linear precoding technique that attempts to align interfering signals in time, frequency, or space. In multiple-input multiple-output (MIMO) networks, IA utilizes the spatial dimension offered by multiple antennas for alignment. By aligning interference at all receivers, IA reduces the dimension of interference allowing users to cancel interference via linear techniques and decode their desired signals interference free. In a single-input single-output (SISO) system with infinitely many channel extensions, IA allows users to communicate at approximately half their interference free rate [8]. IA has also been shown to achieve maximum multiplexing gain in a class of extended MIMO channels [23]. While such a general result has not been proven for the constant MIMO channel, antennas have been shown to provide a practical source of dimensionality [71, 75, 110]. While various constant

MIMO IA solutions have been proposed, all solutions assume some form of channel state information (CSI) at the transmitter obtained either implicitly, through channel reciprocity [21, 90], or explicitly through interference pricing [87] or CSI feedback [8, 12, 73, 75, 85, 99]. Reciprocity, however, does not hold in frequency duplexed systems. Moreover, with time duplexing, reciprocity requires tight calibration of RF devices. Calibration must either be accounted for in the RF hardware [38], or done dynamically using a series of forward and reverse transmissions as in [27] which again incurs overhead. Since reciprocity may not hold, methods to transfer CSI back to the transmitters are important for realizing the sum capacities promised by IA.

CSI feedback potentially incurs large overhead which reduces the effective data rates achieved. Therefore, low overhead feedback strategies that preserve IA's sum-rate performance must be used to satisfy the CSI requirement. Unfortunately, work on managing the overhead of IA, however, is rather limited. In [90], the overhead of the proposed reciprocity-based algorithm is numerically analyzed. Others have proposed network design strategies such as partitioning to reduce overhead [74]. Network partitioning, however, only reduces the number of channels that are to be fed back and makes no attempt at improving the underlying feedback strategy itself. A well established method to improve feedback and further reduce overhead is to employ channel state quantization. For example, [7] uses Grassmannian codebooks to directly quantize and feedback the wideband channel coefficients in single antenna systems which employ IA by coding over frequency. This limited feedback approach

was later extended in [45] to systems with multiple antennas at the transmitters, receivers, or both. In both [7] and [45], multiplexing gain is preserved by scaling the number of feedback bits with SNR. The need for such scaling is a characteristic of digital feedback and is indeed in line with earlier results for other multiuser channels [36]. The search and storage complexity of quantized feedback, however, increases with codebook size and large Grassmannian codebooks are hard to design.

2.2 Contributions

In this chapter, I propose using IA with analog feedback [9, 63, 84, 88]. Instead of quantizing the channel state information, analog feedback directly transmits the channel matrix elements as uncoded quadrature and amplitude modulated symbols. In the proposed strategy, assuming receiver channel knowledge, receivers train the reverse links and feedback the forward channel matrices using analog feedback. I show that, under mild assumptions on feedback quality, performing IA on the channel estimates obtained via analog feedback incurs no loss in multiplexing gain [69]. Specifically, using a cooperative analog feedback strategy, I show that when the SNR on both forward and reverse channels is comparable, the loss in sum-rate achieved by a linear zero-forcing receiver is upper bounded by a constant which implies the preservation of multiplexing gain. I extend the performance analysis to demonstrate that other analog feedback strategies, which assume no cooperation, perform similarly. I show that when such symmetry in forward and reverse link SNR is not

possible, the system still achieves a fraction of its degrees of freedom. Using the derived bounds on loss in sum-rate due to imperfect CSI, I provide a preliminary treatment of IA's effective throughput when the overhead incurred by training and feedback is accounted for. I use the constructed overhead framework to numerically optimize training and feedback lengths to find the best operating point on the trade-off curve between overhead and sum-rate.

This chapter extends the analog feedback framework that was originally proposed for the MISO broadcast channel [63] to the MIMO interference channel, and characterizes its performance when IA is used as a transmission strategy. This chapter proves that analog feedback is a viable alternative to the quantization based schemes presented in [7, 45]. As an advantage, analog feedback does not suffer from the same increasing complexity at high SNR as digital feedback, and the required feedback scaling will come naturally in most wireless ad hoc networks. Moreover, analog feedback remains optimal when the number of feedback symbols equals the number of feedback channel uses [9].

This chapter is organized as follows. Section 2.3 introduces the MIMO interference channel system model for the forward and reverse channels, and reviews the concept of interference alignment. Section 2.4 introduces the analog feedback strategy proposed and gives the main results on rate loss and multiplexing gain with imperfect CSI. Section 2.5 introduces the concept of overhead and presents results on the optimal training and feedback lengths. I then present simulation results to verify the claims made throughout the

chapter in Section 2.6 and conclude with Section 2.7.

2.3 System Model and Background

In this section, I introduce the MIMO interference channel under consideration. I review the concept of IA with perfect channel state information, with a focus on the equations and properties that will be used in the analysis of analog feedback in later sections.

2.3.1 Interference Channel Model

Consider the narrowband MIMO interference channel shown in Fig. 2.1. In this channel each of the K source nodes i communicates with its sink node i and interferes with all other sink nodes, $\ell \neq i$. I refrain from using the term transmitter-receiver since all nodes will be involved in both transmission and reception in either payload transmission or feedback intervals. For simplicity of exposition, I consider the case of the homogeneous network where each source and sink is equipped with N_{T} and N_{R} antennas respectively. Thus, source node i transmits $d_i \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$ independent spatial streams to its corresponding sink. The results can be readily generalized to a different number of antennas at each node, provided that IA remains feasible [110].

I consider a block fading channel model in which channels are drawn independently across all users and antennas and remain fixed for the interval of interest. I neglect large scale fading effects which can be accounted for at the expense of more involved exposition in Section 2.4.1. I also assume

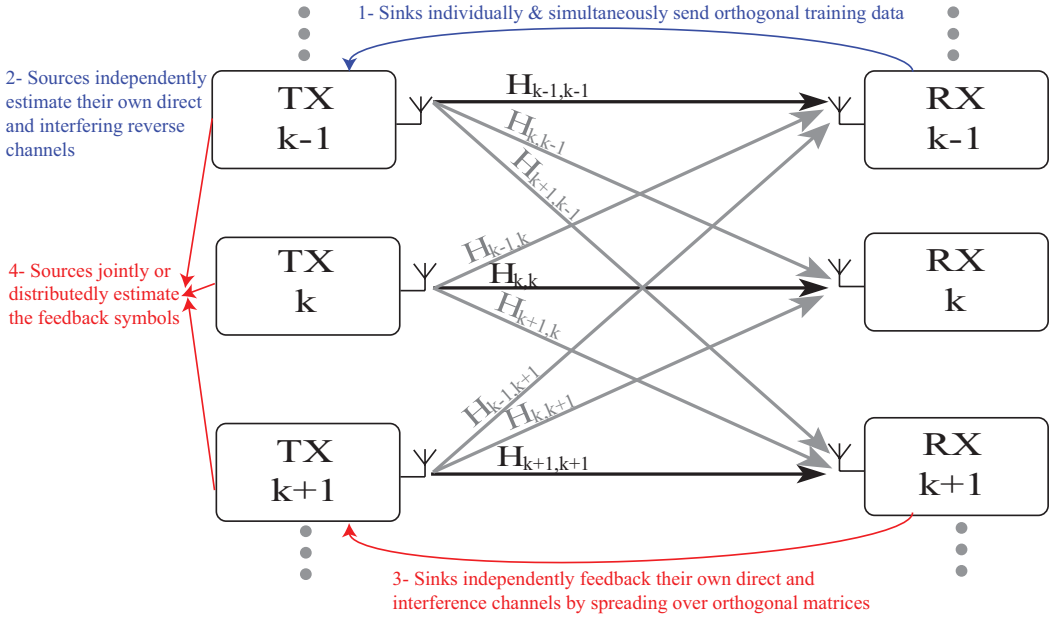

Figure 2.1: K-User MIMO interference channel with analog feedback.

perfect time and frequency synchronization when expressing received baseband signals. Under those assumptions, the received signal at sink node i can be written as

$$\mathbf{y}_i = \sqrt{\frac{P}{d_i}}\mathbf{H}_{i,i}\mathbf{F}_i\mathbf{s}_i + \sum_{\ell \neq i}\sqrt{\frac{P}{d_\ell}}\mathbf{H}_{i,\ell}\mathbf{F}_\ell\mathbf{s}_\ell + \mathbf{v}_i,$$

where $\mathbf{y}_i$ is the $N_{\mathrm{R}} \times 1$ received signal vector, $\mathbf{H}_{i,\ell}$ is the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ channel matrix from source ℓ to sink i with i.i.d $\mathcal{CN}(0,1)$ elements, $\mathbf{F}_i = [\mathbf{f}_i^1, \cdots, \mathbf{f}_i^{d_i}]$ is node i's $N_{\mathrm{T}} \times d_i$ unitary precoding matrix, $\mathbf{s}_i$ is the $d_i \times 1$ transmitted symbol vector at node i such that $\mathbb{E}\left[\|\mathbf{s}_i\|^2\right] = d_i$, and $\mathbf{v}_i$ is a complex vector of i.i.d complex Gaussian noise with covariance matrix $\sigma^2\mathbf{I}_{\mathbf{N}_{\mathrm{R}}}$. I assume equal power allocation since the gain observed from water-filling is negligible at high

SNR [61].

I place no reciprocity assumption on the forward and reverse channels, as in a frequency division duplexed system, for example. On this channel, the received signal at *source node* i is

$$\overleftarrow{\mathbf{y}}_i = \sqrt{\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\mathbf{G}_{i,i}\overleftarrow{\mathbf{x}}_i + \sum_{\ell \neq i}\sqrt{\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\mathbf{G}_{\ell,i}\overleftarrow{\mathbf{x}}_\ell + \nu_i, \tag{2.1}$$

where P_{F} is the transmit power used to transmit pilot and feedback symbols, $\mathbf{G}_{\ell,i}$ is the $N_{\mathrm{T}} \times N_{\mathrm{R}}$ reverse channel between sink node ℓ and source node i with i.i.d $\mathcal{CN}(0,1)$ elements, $\overleftarrow{\mathbf{x}}_i$ is the symbol vector with unit norm elements sent by sink i, and ν_i is a complex vector of i.i.d circularly symmetric white Gaussian noise with covariance matrix $\sigma^2\mathbf{I}_{\mathbf{N}_{\mathrm{T}}}$.

2.3.2 Interference Alignment

IA for the MIMO interference channel is a linear precoding technique that by potentially coding over infinite channel extensions achieves the channel's degrees of freedom defined as $\lim_{P\to\infty}\frac{R_{\mathrm{sum}}}{\log_2 P}$. This result originally assumed that the magnitude of continuously distributed i.i.d channel coefficients is bounded away from zero and infinity [23] to avoid the degenerate cases of equal coefficients or channels equal to zero or infinity. For constant MIMO channels with Rayleigh fading, degenerate cases happen with probability zero and IA can improve the achieved sum-rate as shown in [21, 71, 75]. In constant MIMO channels, IA computes the transmit precoders $\mathbf{F}_i$ to align interference at all receivers in a strict subspace of the received signal space, thus leaving

interference free dimensions for the desired signal. While IA is only one of the many precoding strategies for the interference channel [75, 81, 90], some of which marginally outperform it at low SNR [75], a main advantage of IA is that it is analytically tractable. Its complete interference suppression properties make it especially amenable to the study of performance with feedback and imperfect CSI.

While IA can be used with any receiver design, to simplify exposition, I consider a per-stream zero-forcing receiver in which sink node i projects its received signal on to the columns, $\mathbf{w}_i^m$, and treats residual interference as noise. Simulations in Section 2.6 indicate that the same performance can be expected from a joint signal decoder which again treats interference as noise.

Writing the per stream input-output relation at the output of $\mathbf{w}_i^m$ gives

$$\begin{aligned}(\mathbf{w}_i^m)^*\mathbf{y}_i =&(\mathbf{w}_i^m)^*\sqrt{\frac{P}{d_i}}\mathbf{H}_{i,i}\mathbf{f}_i^m s_i^m + \sum_{\ell\neq m}(\mathbf{w}_i^m)^*\sqrt{\frac{P}{d_i}}\mathbf{H}_{i,i}\mathbf{f}_i^\ell s_i^\ell \\ &+ \sum_{k\neq i}\sum_{\ell=1}^{d_k}(\mathbf{w}_i^m)^*\sqrt{\frac{P}{d_k}}\mathbf{H}_{i,k}\mathbf{f}_k^\ell s_k^\ell + (\mathbf{w}_i^m)^*\mathbf{v}_i,\end{aligned} \tag{2.2}$$

for $m \in \{1, \ldots, d_i\}$ and $i \in \{1, \ldots, K\}$, where $\|\mathbf{w}_i^m\|^2 = 1$. At the output of these linear receivers $\mathbf{w}_i^m$, the conditions for perfect interference alignment can be restated as [21]

$$(\mathbf{w}_i^m)^*\mathbf{H}_{i,k}\mathbf{f}_k^\ell = 0, \qquad \forall(k,\ell) \neq (i,m) \tag{2.3}$$

$$|(\mathbf{w}_i^m)^*\mathbf{H}_{i,i}\mathbf{f}_i^m| \geq c > 0, \qquad \forall i, m \tag{2.4}$$

where IA is guaranteed by the first condition, and the second is satisfied with high probability.

The sum-rate achieved by such a linear zero-forcing receiver, assuming Gaussian input signals and treating leakage interference as noise, is

$$R_{\text{sum}} = \sum_{i=1}^{K} \sum_{m=1}^{d_i} \log_2 \left(1 + \frac{\frac{P}{d_i} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m \right|^2}{\mathcal{I}_{i,m} + \sigma^2} \right), \tag{2.5}$$

where $\mathcal{I}_{i,m}$ is the total inter-stream and inter-user interference given by [9, 36]

$$\mathcal{I}_{i,m} = \sum_{(k,\ell) \neq (i,m)} \frac{P}{d_k} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,k} \mathbf{f}_k^\ell \right|^2.$$

In the presence of perfect channel knowledge, and for an achievable degree of freedom vector $\mathbf{d} = [d_1, d_2, \ldots, d_K]$, equations (2.3) and (2.4), are satisfied and thus $\mathcal{I}_{i,m} = 0$. This gives

$$\begin{aligned} \lim_{P \to \infty} \frac{R_{\text{sum}}}{\log_2 P} = \lim_{P \to \infty} \frac{\sum_{i,m} \log_2 \left(1 + \frac{\frac{P}{d_i} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m \right|^2}{\sigma^2} \right)}{\log_2 P} &= \sum_{i=1}^{K} d_i \\ &\leq \min(N_{\text{T}}, N_{\text{R}}) \frac{R}{R+1} K, \end{aligned} \tag{2.6}$$

in the case where $R = \frac{\max(N_{\text{T}}, N_{\text{R}})}{\min(N_{\text{T}}, N_{\text{R}})}$ is an integer such that $K > R$ [23]. The full characterization of the extended MIMO interference channel's degrees of freedom is provided in [23].

It is not immediately clear, however, if the same sum-rate scaling behavior can be expected from a network with only imperfect knowledge of the channel derived from analog feedback. Results on single user MIMO generally prove an acceptable constant rate loss due to imperfect CSI [53]. In most multi-user scenarios, however, the cost of imperfect CSI may be much higher, potentially resulting in the loss of the channel's multiplexing gain [36] which

saturates performance at high SNR [66]. In Section 2.4, I show that such performance can be expected from a realistic IA system with analog feedback, provided that the quality of channel knowledge scales sufficiently with transmit power, or effectively the forward channel's SNR. The result is based on the necessary accuracy of CSI which must scale with SNR as shown in [10]. This is similar to the results presented in [7] and [45] for IA with digital feedback where CSI quality is scaled by controlling codebook size and feedback bits.

2.4 Interference Alignment with Analog Feedback

In this section, I propose a feedback and transmission strategy based on analog feedback and naive IA, which uses the estimated channels as if they were the true propagation channels.

2.4.1 Analog Feedback

To feedback the forward channel matrices $\mathbf{H}_{i,\ell}$ reliably across the feedback channel given in (2.1), I propose dividing the feedback stage into two main phases [63]. The details of the analog feedback strategy considered are summarized in Fig. 2.1. First, each source independently learns its reverse channels. Second, the forward channels are fed back and estimated. I neglect forward channel training and estimation and, thus, assume they have been estimated perfectly. This is since imperfect CSI at the receiver adversely affects all feedback schemes and is not exclusive to analog feedback. In fact, the analysis of forward channel estimation error parallels the analysis in Section

2.4.1.1 and will simply add an extra error term that decays with power as needed in Section 2.4.2. Therefore, similar results can be readily shown.

2.4.1.1 Reverse Link Training

To learn all reverse links, the K sink nodes must transmit known pilot symbols over a period $\tau_{\mathrm{p}} \geq KN_{\mathrm{R}}$. Similar to the analysis done in [63], I let each sink independently and simultaneously transmit an $N_{\mathrm{R}} \times \tau_{\mathrm{p}}$ matrix of pilots $\mathbf{\Phi}_i$ such that $\mathbf{\Phi}_i\mathbf{\Phi}_k^* = \delta_{ik}\mathbf{I}_{N_{\mathrm{R}}}$ shown to be optimal in [62]. This training phase only requires synchronization.

Let $\overleftarrow{\mathbf{Y}}_i = \left[\overleftarrow{\mathbf{y}}_i[1] \ \ldots \ \overleftarrow{\mathbf{y}}_i[\tau_{\mathrm{p}}]\right]$ be the $N_{\mathrm{T}} \times \tau_{\mathrm{p}}$ received training matrix at source node i. Then the received training is

$$\overleftarrow{\mathbf{Y}}_i = \sqrt{\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}\sum_{k=1}^{K}\mathbf{G}_{k,i}\mathbf{\Phi}_k + \mathbf{V}_i, \quad \forall i,$$

where $\mathbf{V}_i$ is an $N_{\mathrm{T}} \times \tau_{\mathrm{p}}$ matrix of i.i.d $\mathcal{CN}(0,\sigma^2)$ noise elements. Using its received training, each source i locally computes MMSE estimates of its channels given by

$$\widehat{\mathbf{G}}_{k,i} = \frac{\sqrt{\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}}{\sigma^2 + \frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}\overleftarrow{\mathbf{Y}}_i\mathbf{\Phi}_k^*, \quad \forall k. \tag{2.7}$$

Since $\widehat{\mathbf{G}}_{k,i}$ are MMSE estimates of Gaussian variables corrupted by Gaussian noise, this results in $\widehat{\mathbf{G}}_{k,i} \sim \mathcal{CN}\left(0, \frac{\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}{\sigma^2+\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)$ and $\widetilde{\mathbf{G}}_{k,i} = \mathbf{G}_{k,i} - \widehat{\mathbf{G}}_{k,i}$ with i.i.d $\mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2+\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)$ elements.

2.4.1.2 Analog CSI Feedback

After reverse link training, each sink node i independently sends its unquantized uncoded estimates of $\mathbf{H}_{i,k}\ \forall\ k$ over a period τ_{f}. To have the sink nodes feedback their CSI simultaneously and orthogonally, each sink post multiplies its $N_{\mathrm{R}} \times KN_{\mathrm{T}}$ feedback matrix $[\mathbf{H}_{i,1} \ldots \mathbf{H}_{i,K}]$ with a $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix $\mathbf{\Psi}_i$ such that $\mathbf{\Psi}_i \mathbf{\Psi}_k^* = \delta_{i,k} \mathbf{I}_{KN_{\mathrm{T}}}$ [63]. This is a general orthogonal structure that can capture the case of orthogonality in time and requires $\tau_{\mathrm{f}} \geq K^2 N_{\mathrm{T}}$. The transmitted $N_{\mathrm{R}} \times \tau_{\mathrm{f}}$ feedback matrix $\overleftarrow{\mathbf{X}}_i$ from sink i can be written as

$$\overleftarrow{\mathbf{X}}_i = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{KN_{\mathrm{T}} N_{\mathrm{R}}}} \left[\mathbf{H}_{i,1} \ldots \mathbf{H}_{i,K}\right] \mathbf{\Psi}_i.$$

The concatenated $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix of received feedback is then given by

$$\overleftarrow{\mathbf{Y}}_{\mathrm{c}} = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{KN_{\mathrm{T}} N_{\mathrm{R}}}} \sum_{i=1}^{K} \begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix} \left[\mathbf{H}_{i,1} \ldots \mathbf{H}_{i,K}\right] \mathbf{\Psi}_i + \mathbf{V},$$

where $\mathbf{V}$ is the $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix of i.i.d Gaussian noise.

To estimate the forward channels $\mathbf{H}_{i,k}$, the source nodes first isolate the training from sink node i by post multiplying their received training by $\mathbf{\Psi}_i^*$ which gives

$$\overleftarrow{\mathbf{Y}}_{\mathrm{c}} \mathbf{\Psi}_i^* = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{KN_{\mathrm{T}} N_{\mathrm{R}}}} \underbrace{\begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix}}_{\mathbf{G}_i} \underbrace{\left[\mathbf{H}_{i,1} \ldots \mathbf{H}_{i,K}\right]}_{\mathbf{H}_i} + \mathbf{V} \mathbf{\Psi}_i^*. \tag{2.8}$$

Here we note that the columns of $\overleftarrow{\mathbf{Y}}_{\mathrm{c}} \mathbf{\Psi}_i^*$ have i.i.d noise from $\mathbf{V} \mathbf{\Psi}_i^*$ (since $\mathbf{\Psi}_i$ is orthogonal) and are functions only of the corresponding independent

columns of $\mathbf{H}_i$. Thus the columns of $\overleftarrow{\mathbf{Y}}_{\mathrm{c}}\mathbf{\Psi}_i^*$ are independent. To simplify the analysis in Section 2.4.2, I assume that the complete received feedback matrix $\overleftarrow{\mathbf{Y}}_{\mathrm{c}}$ is shared among sources. Then, assuming $KN_{\mathrm{T}} \geq N_{\mathrm{R}}$, sources compute a common least squares estimate $\widehat{\mathbf{H}}_i$ of $\mathbf{H}_i$ given by

$$\widehat{\mathbf{H}}_i \;=\; \sqrt{\frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}P_{\mathrm{F}}}}\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}\widehat{\mathbf{G}}_i^*\overleftarrow{\mathbf{Y}}_{\mathrm{c}}\mathbf{\Psi}_i^* = \underbrace{\mathbf{H}_i}_{\text{Real Channel}} + \underbrace{\widetilde{\mathbf{H}}_i}_{\text{Error}},$$

where $\widehat{\mathbf{G}}_i$ is the MMSE estimate of $\mathbf{G}_i$. Such node cooperation is only realistic in certain cases such as IA for cellular systems for example. At the end of this section, I provide alternative non-cooperative approaches that I show in Section 2.6 perform close to this special case.

Before continuing, I make the following observation. Note that each column of $\widehat{\mathbf{H}}_i$ is a function of only the corresponding column of $\overleftarrow{\mathbf{Y}}_{\mathrm{c}}\mathbf{\Psi}_i^*$ which in turn are independent and functions of the corresponding column of $\mathbf{H}_i$. This implies that $\widehat{\mathbf{H}}_{k,\ell}$ is a function of only $\mathbf{H}_{k,\ell}$. Thus, since $\mathbf{H}_{k,\ell}$ are independent, so are the estimates $\widehat{\mathbf{H}}_{k,\ell}$. This property will be important in Appendix 2.8.1.

The error in the estimates of $\mathbf{H}_i$ can then be written as

$$\widetilde{\mathbf{H}}_i = \left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}\widehat{\mathbf{G}}_i^*\left(\sqrt{\frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}P_{\mathrm{F}}}}\mathbf{V}\mathbf{\Psi}_i^* + \widetilde{\mathbf{G}}_i\mathbf{H}_i\right).$$

which makes it clear that the error in the estimate consists of two error terms: the first due to noisy feedback and the second due to a noisy estimate of the feedback channel. To quantify the effect of the error on the achieved sum-rate, I derive the variance of the error term introduced by analog feedback. Recall that the elements of $\mathbf{H}_{i,k}$ are $\mathcal{CN}(0,1)$, those of $\mathbf{V}$ are $\mathcal{CN}(0,\sigma^2)$, and

those of $\widetilde{\mathbf{G}}_i$ are $\mathcal{CN}(0, \frac{\sigma^2}{\sigma^2+\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}})$. As a result, the error term $\widetilde{\mathbf{G}}_i\mathbf{H}_i$ due to the reverse channel estimation has independent elements with a variance of $\frac{N_{\mathrm{R}}\sigma^2}{\sigma^2+\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}$. Similarly to [63], we see that the covariance of each column of $\widetilde{\mathbf{H}}_i$ denoted $\widetilde{\mathbf{H}}_i^{(\ell)}$, conditioned on $\widehat{\mathbf{G}}_i$ is

$$\mathrm{Cov}(\widetilde{\mathbf{H}}_i^{(\ell)}|\widehat{\mathbf{G}}_i) = \left(\frac{KN_{\mathrm{T}}N_{\mathrm{R}}\sigma^2}{\tau_{\mathrm{f}}P_{\mathrm{F}}} + \frac{N_{\mathrm{R}}\sigma^2}{\sigma^2+\tau_{\mathrm{p}}\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}.$$

Since the elements of the MMSE estimate $\widehat{\mathbf{G}}_i$ are Gaussian and uncorrelated, the diagonal elements of $\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}$ are reciprocals of scaled chi-squared random variables with $2(KN_{\mathrm{T}} - N_{\mathrm{R}} + 1)$ degrees of freedom. As a result, the variance of the error in the elements of $\widehat{\mathbf{H}}_{i,k}$ is

$$\sigma^2_{\widetilde{\mathbf{H}}} = \frac{\sigma^2}{(KN_{\mathrm{T}} - N_{\mathrm{R}})P_{\mathrm{F}}}\left(\frac{{N_{\mathrm{R}}}^2}{\tau_{\mathrm{p}}} + \frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}}\left(1+\frac{N_{\mathrm{R}}\sigma^2}{\tau_{\mathrm{p}}P_{\mathrm{F}}}\right)\right). \tag{2.9}$$

At high SNR this gives

$$\sigma^2_{\widetilde{\mathbf{H}}} \approx \frac{\sigma^2\left(\frac{{N_{\mathrm{R}}}^2}{\tau_{\mathrm{p}}} + \frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}}\right)}{(KN_{\mathrm{T}} - N_{\mathrm{R}})P_{\mathrm{F}}}. \tag{2.10}$$

Having computed feedback error, I return to the assumption on node cooperation. Cooperation simplifies Section 2.4.1 by making a common channel estimate known to all users. Such cooperation may not always be possible. I present two alternative practical strategies, outline the changes they incur to the feedback process, and discuss their shortcomings:

- Centralized processor: In this scheme, reverse link training, estimation and feedback transmission remain unchanged. In this strategy however,

the $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix $\overleftarrow{\mathbf{Y}}_{\mathrm{c}}$ is not shared, and instead one of the nodes calculates an estimate of the feedback information based on its locally observed $N_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix of feedback symbols (i.e. only its own rows of $\overleftarrow{\mathbf{Y}}_{\mathrm{c}}$). This can be done as long as $N_{\mathrm{T}} \geq N_{\mathrm{R}}$[1] and results in $\sigma^2_{\tilde{\mathbf{H}}} \approx \frac{\sigma^2}{(N_{\mathrm{T}}-N_{\mathrm{R}})P_{\mathrm{F}}}\left(\frac{{N_{\mathrm{R}}}^2}{\tau_{\mathrm{p}}} + \frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}}\right)$. This source then calculates the IA solution required and feeds it forward, again using "analog feedforward", to all other sources [11].

- Distributed processing: In this strategy reverse link training, estimation and feedback transmission again remain unchanged. In this scheme, however, *each* source locally calculates its own estimate of all the channels being fed back, using the $N_{\mathrm{T}} \times \tau_{\mathrm{f}}$ symbols it independently receives. Therefore, each node will obtain a different perturbed estimate of the forward channels with the same $\sigma^2_{\tilde{\mathbf{H}}}$ as in the centralized processor case. Each node then locally calculates a complete set of perturbed IA precoders based on its channel estimates.

I expand on the performance of these two alternative practical strategies in Section 2.6.

2.4.2 Multiplexing Gain with Analog Feedback

To characterize the performance of IA with analog feedback, I examine the rate loss [36] incurred by naive IA where channel estimates are used to

[1]The restriction $N_{\mathrm{T}} \geq N_{\mathrm{R}}$ can be relaxed at the expense of a longer τ_{f} by assigning each user several spreading matrices $\mathbf{\Psi}$.

calculate the columns of the precoders, $\widehat{\mathbf{f}}_i^m \ \forall i, m$ and combiners $\widehat{\mathbf{w}}_i^m \ \forall i, m$. Therefore, such transmit and receive vectors satisfy

$$(\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell = 0, \qquad \forall (k,\ell) \neq (i,m) \tag{2.11}$$

$$\left| (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m \right| \geq c > 0, \qquad \forall i, m, \tag{2.12}$$

and can be found for a feasible degree of freedom vector $\mathbf{d} = [d_1, \ldots, d_K]$. As stated earlier, sinks need not use such per-stream combiners, and can employ a number of other receiver designs based on their channel knowledge. In Section 2.6 I show the performance of a joint decoder which whitens interference using its covariance [6]. Such receiver CSI can be acquired blindly [30] or via additional training or silent phases. In the following rate loss analysis, I abstract the calculation of receiver CSI and assume that sinks have learned the combiners $\widehat{\mathbf{w}}_i^m$.

The mean loss in sum-rate is then defined as $\Delta R_{\text{sum}} \triangleq \mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \mathbb{E}_{\mathbf{H}} \widehat{R}_{\text{sum}}$, where $\mathbb{E}_{\mathbf{H}} R_{\text{sum}}$ is the average sum-rate from IA with perfect CSI, with instantaneous rate given in (2.5), and $\mathbb{E}_{\mathbf{H}} \widehat{R}_{\text{sum}}$ is the rate achieved with imperfect CSI and the vectors in (2.11), (2.12). Before proving the main result on rate loss, I will need the following lemma.

Lemma 1. *The desired signal powers,* $|(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2$ *and* $\left| (\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2$, *resulting from IA with perfect or imperfect CSI respectively are identically and exponentially distributed.*

Proof. See Appendix 2.8.1. □

Theorem 2. *The sum-rate loss experienced by IA on the K-user $N_{\mathrm{R}} \times N_{\mathrm{T}}$ interference channel with imperfect CSI obtained via the analog feedback strategy described in Section 2.4.1 is upper bounded by a constant when P_{F} scales with the transmit power P.*

Proof. Let the K-user $N_{\mathrm{R}} \times N_{\mathrm{T}}$ interference channel use the analog feedback scheme presented to achieve a vector of multiplexing gains $\mathbf{d}$. Using the imperfect vectors $\widehat{\mathbf{f}}_i^m$ and $\widehat{\mathbf{w}}_i^m$ respectively, the input-output relationship at the output of a linear zero-forcing receiver is

$$\begin{aligned}(\widehat{\mathbf{w}}_i^m)^* \mathbf{y}_i =& (\widehat{\mathbf{w}}_i^m)^* \sqrt{\frac{P}{d_i}} \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^m s_i^m + \sum_{\ell \neq m} (\widehat{\mathbf{w}}_i^m)^* \sqrt{\frac{P}{d_i}} \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^\ell s_i^\ell \\ &+ \sum_{k \neq i} \sum_{\ell=1}^{d_k} (\widehat{\mathbf{w}}_i^m)^* \sqrt{\frac{P}{d_k}} \mathbf{H}_{i,k} \widehat{\mathbf{f}}_k^\ell s_k^\ell + (\widehat{\mathbf{w}}_i^m)^* \mathbf{v}_i.\end{aligned} \tag{2.13}$$

Using this received signal and the instantaneous rate expression in (2.5) gives the following upper bound on mean loss in sum-rate:

$$\begin{aligned}\Delta R_{\mathrm{sum}} = & \mathbb{E}_{\mathbf{H}} \sum_{i,m} \log_2 \left(1 + \frac{\frac{P}{d_i} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m \right|^2}{\sigma^2} \right) \\ & - \mathbb{E}_{\mathbf{H}, \widehat{\mathbf{H}}} \sum_{i,m} \log_2 \left(1 + \frac{\frac{P}{d_i} \left| (\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2}{\mathcal{I}_{i,m} + \sigma^2} \right)\end{aligned} \tag{2.14}$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathbf{H}} \sum_{i,m} \log_2 \left(1 + \frac{\frac{P}{d_i} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m \right|^2}{\sigma^2} \right) \\
&\quad - \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \sum_{i,m} \log_2 \left(1 + \frac{\mathcal{I}_{i,m} + \frac{P}{d_i} \left| (\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2}{\sigma^2} \right) \\
&\quad + \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \sum_{i,m} \log_2 \left(1 + \frac{\mathcal{I}_{i,m}}{\sigma^2} \right) \\
&\overset{(a)}{\leq} \mathbb{E}_{\mathbf{H}\widehat{\mathbf{H}}} \sum_{i,m} \log_2 \left(1 + \frac{\mathcal{I}_{i,m}}{\sigma^2} \right) \overset{(b)}{\leq} \sum_{i,m} \log_2 \left(1 + \frac{\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \mathcal{I}_{i,m}}{\sigma^2} \right).
\end{aligned}
$$

where (a) follows from Lemma 1 which along with $\mathcal{I}_{i,m} \geq 0$ implies that $\frac{P}{d_i} \left| (\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2 + \mathcal{I}_{i,m}$ stochastically dominates $\frac{P}{d_i} \left| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m \right|^2$ and (b) follows from Jensen's inequality [36].

Since (2.11) and (2.12) are satisfied, however, the total interference term $\mathcal{I}_{i,m}$ can be simplified to include only residual interference due to the channel estimation errors $\widetilde{\mathbf{H}}_{i,\ell}$. Therefore, ΔR_{sum} can be further upper bounded by noticing that

$$
\begin{aligned}
\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left| (\widehat{\mathbf{w}}_i^m)^* (\widehat{\mathbf{H}}_{i,k} - \widetilde{\mathbf{H}}_{i,k}) \widehat{\mathbf{f}}_k^\ell \right|^2 &= \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left| (\widehat{\mathbf{w}}_i^m)^* (\widetilde{\mathbf{H}}_{i,k}) \widehat{\mathbf{f}}_k^\ell \right|^2 \\
&\leq \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \|\widetilde{\mathbf{H}}_{i,k}\|_F^2, \quad \forall k, \ \forall \ell \neq m,
\end{aligned}
\tag{2.15}
$$

which gives

$$
\Delta R_{\text{sum}} \leq \sum_{i,m} \log_2 \left(1 + \frac{1}{\sigma^2} \sum_{\ell=1}^{K} \frac{P}{d_\ell} \left(d_\ell - \delta_{i,\ell} \right) \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \|\tilde{\mathbf{H}}_{i,\ell}\|_F^2 \right). \tag{2.16}
$$

From (2.9), however, I have $\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \|\tilde{\mathbf{H}}_{i,\ell}\|_F^2 = N_{\text{T}} N_{\text{R}} \sigma_{\widetilde{\mathbf{H}}}^2 = \frac{c(\tau_{\text{p}},\tau_{\text{f}})\sigma^2}{P_{\text{F}}}$, where $c(\tau_{\text{p}}, \tau_{\text{f}})$ is a constant, independent of P_{F} at high enough SNR, given by

$$
c(\tau_{\text{p}}, \tau_{\text{f}}) = N_{\text{T}} N_{\text{R}} \frac{\left(\frac{{N_{\text{R}}}^2}{\tau_{\text{p}}} + \frac{K N_{\text{T}} N_{\text{R}} (1+\epsilon)}{\tau_{\text{f}}} \right)}{(K N_{\text{T}} - N_{\text{R}})}. \tag{2.17}
$$

Combining (2.16), (2.17) and letting $P_{\mathrm{F}} = \alpha^{-1}P$ gives the final upper bound on throughput loss

$$\begin{aligned}\Delta R_{\mathrm{sum}}(\tau_{\mathrm{p}}, \tau_{\mathrm{f}}) &\leq \sum_{i,m} \log_2 \left(1 + \frac{c(\tau_{\mathrm{p}}, \tau_{\mathrm{f}})\sigma^2 P}{\sigma^2 P_{\mathrm{F}}} \sum_{\ell=1}^{K} \frac{d_\ell - \delta_{i,\ell}}{d_\ell}\right) \\ &\leq \sum_{i} d_i \log_2 \left(1 + \alpha c(\tau_{\mathrm{p}}, \tau_{\mathrm{f}}) \left(K - \frac{1}{d_i}\right)\right).\end{aligned} \tag{2.18}$$

The bound has been presented at high SNR for simplicity of exposition only; (2.17) can be adapted for any SNR > 0 by using (2.9) instead of (2.10). □

Corollary 3. *IA with imperfect CSI obtained via the analog feedback strategy in Section 2.4.1 achieves the same average multiplexing gain as IA with perfect CSI.*

Proof. This follows immediately from the constant loss in sum-rate shown in Theorem 2 which becomes negligible as SNR is taken to infinity. □

In summary, Theorem 2 states that if the SNR on the reverse and forward link are comparable, the cost of imperfect CSI is a constant. This constant is a decreasing function of τ_{f} and τ_{p} and thus, I have written $\Delta R_{\mathrm{sum}}(\tau_{\mathrm{p}}, \tau_{\mathrm{f}})$ in (2.18) to highlight this dependence. Note that while bounding leakage interference as $\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left|(\widehat{\mathbf{w}}_i^m)^*(\widetilde{\mathbf{H}}_{i,k})\widehat{\mathbf{f}}_k^\ell\right|^2 \leq \mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}}\|\widetilde{\mathbf{H}}_{i,k}\|_F^2 = N_{\mathrm{T}}N_{\mathrm{R}}\sigma^2_{\widetilde{\mathbf{H}}}$ suffices to establish a constant rate loss, it is very conservative, and increasingly loose for larger systems. If one assumes that $\widetilde{\mathbf{H}}_{i,k}$ has a bi-unitary invariant distribution, e.g. Gaussian, then by the same reasoning as in Appendix 2.8.1, I get $\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \left|(\widehat{\mathbf{w}}_i^m)^*(\widetilde{\mathbf{H}}_{i,k})\widehat{\mathbf{f}}_k^\ell\right|^2 = \sigma^2_{\widetilde{\mathbf{H}}}$ and $c(\tau_{\mathrm{p}}, \tau_{\mathrm{f}})$ is replaced with $c_2(\tau_{\mathrm{p}}, \tau_{\mathrm{f}}) =$

$\frac{1}{KN_{\mathrm{T}}-N_{\mathrm{R}}}\left(\frac{{N_{\mathrm{R}}}^2}{\tau_{\mathrm{p}}}+\frac{KN_{\mathrm{T}}N_{\mathrm{R}}(1+\epsilon)}{\tau_{\mathrm{f}}}\right)$. Since SNRs are likely to be comparable in practice, analog feedback allows systems to overcome the problem of fast scaling complexity in digital feedback, which remains even if the actual feedback stage operates error-free close to capacity [9, 36]. In case such SNR symmetry is not possible, however, systems become interference limited and multiplexing gain is reduced.

Theorem 4. *IA on a K-user $N_{\mathrm{R}} \times N_{\mathrm{T}}$ interference channel using analog feedback with $P_{\mathrm{F}} = \alpha P^{\beta}$ such that $0 \leq \beta \leq 1$ achieves at least a β-fraction of its original multiplexing gain.*

Proof. The multiplexing gain achieved by a system using naive IA and analog feedback with $P_{\mathrm{F}} = \alpha P^{\beta}$ can be written as

$$
\begin{aligned}
M &= \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}}[\widehat{R}_{\mathrm{sum}}]}{\log_2(P)} = \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H},\widehat{\mathbf{H}}} \sum_{i,m} \log_2\left(1+\frac{\frac{P}{d_i}\left|(\widehat{\mathbf{w}}_i^m)^*\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2}{\mathcal{I}_{i,m}+\sigma^2}\right)}{\log_2(P)} \\
&= \lim_{P\to\infty} \frac{\mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}} \sum_{i,m} \log_2\left(\frac{P}{d_i}\left|(\widehat{\mathbf{w}}_i^m)^*\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2+\mathcal{I}_{i,m}+\sigma^2\right) - \mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}} \sum_{i,m} \log_2\left(\mathcal{I}_{i,m}+\sigma^2\right)}{\log_2(P)} \\
&\overset{(a)}{\geq} \frac{\mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}} \sum_{i,m} \log_2\left(\frac{P}{d_i}\left|(\widehat{\mathbf{w}}_i^m)^*\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2\right) - \sum_{i,m} \log_2\left(\mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}}\mathcal{I}_{i,m}+\sigma^2\right)}{\log_2(P)} \\
&\overset{(b)}{\geq} \sum_{k=1}^{K} d_k - \lim_{P\to\infty} \frac{\sum_{i.m} \log_2\left(\frac{(d_i-1)P}{d_i}\mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}}\|\widetilde{\mathbf{H}}_{i,i}\|_F^2+(K-1)P\mathbb{E}_{\mathbf{H},\tilde{\mathbf{H}}}\|\widetilde{\mathbf{H}}_{i,k}\|_F^2+\sigma^2\right)}{\log_2(P)}
\end{aligned}
$$

$$\begin{aligned} &\overset{(c)}{=} \sum_{k=1}^{K} d_k - \lim_{P\to\infty} \frac{\sum_{i.m} \log_2\left(\frac{(d_i-1)P^{1-\beta}c(\tau_{\mathrm{p}},\tau_{\mathrm{f}})\sigma^2}{\alpha d_i} + \frac{(K-1)P^{1-\beta}c(\tau_{\mathrm{p}},\tau_{\mathrm{f}})\sigma^2}{\alpha} + \sigma^2\right)}{\log_2(P)} \\ &= \beta\left(\sum_{k=1}^{K} d_k\right), \end{aligned} \tag{2.19}$$

where (a) follows from disregarding interference in the first term and applying Jensen's to the second; (b) follows from (2.15) and the fact that each term in the first summation has a multiplexing gain of 1. Finally, (c) is due to the scaling of residual interference with $P^{1-\beta}$. □

In the case where $\beta = 0$, feedback power is constant, and the interference limited system achieves zero multiplexing gain. In fact, the system simulations in Section 2.6 indicate that the sum-rate achieved is upper bounded by a constant. As for the case of $\beta = 1$, Theorem 4 confirms the preservation of full multiplexing gain proven in Corollary 3, but does not establish the $O(1)$ loss in sum-rate of Theorem 2. As for any $0 < \beta < 1$, Theorem 4 shows that analog feedback and interference alignment can still achieve linear sum-rate scaling even when feedback power is much smaller than transmit power. This may be the case in certain non-homogeneous networks such as cellular networks in which mobile and base station powers do not match.

2.5 Degrees of Freedom with Overhead

While the analysis done in Section 2.4.2 is a good indicator of the cost of imperfect CSI obtained through training and analog feedback, it does not

directly predict the expected throughput achieved by this strategy. Namely, the analysis done thus far neglects the cost of training and feedback overhead. In this section, I define the expected throughput with overhead and use it to optimize training and feedback.

2.5.1 Definition of Overhead

As shown in Section 2.4.2, the performance of IA is tightly related to the mean square error in the channel estimates at the transmitter. When operating in a time varying channel, training and feedback must be done periodically to ensure the validity of the channel estimate at the transmitter. Depending on the channel's coherence time, the overhead due to training and feedback may consume an arbitrarily large fraction of resources such as time or frequency slots, resulting in low net throughput. In this section, I consider the case in which training, feedback, and data transmission are all orthogonal in time, in the same coherence time or frame T [43, 74]. Using this model for training and feedback overhead, I compute the expected effective sum-rate as

$$R_{\text{eff}}(\tau_{\text{p}}, \tau_{\text{f}}) = \left(\frac{T - (\tau_{\text{f}} + \tau_{\text{p}})}{T}\right) \left(\mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \Delta R_{\text{sum}}\right). \tag{2.20}$$

Using the rate loss expression in (2.18), I write the expected sum-rate with overhead as

$$R_{\text{eff}}(\tau_{\text{p}}, \tau_{\text{f}}) \approx \left(\frac{T - (\tau_{\text{f}} + \tau_{\text{p}})}{T}\right) \left(\mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \sum_i d_i \log_2 \left(1 + \alpha c_2(\tau_{\text{p}}, \tau_{\text{f}}) \left(K - \frac{1}{d_i}\right)\right)\right), \tag{2.21}$$

where I have used $c_2(\tau_\mathrm{p}, \tau_\mathrm{f})$ rather than the looser $c(\tau_\mathrm{p}, \tau_\mathrm{f})$. As expected, and as can be seen from (2.20), insufficient training and feedback may result in poor channel estimates at the receiver, and thus a large loss in sum-rate, whereas excessive training and feedback becomes too costly as a large portion of the frame is spent on overhead.

2.5.2 Training and Feedback Optimization

Given the expression for system throughput with overhead, I propose to solve the following maximization problem

$$\begin{aligned} \max_{\tau_\mathrm{f}, \tau_\mathrm{p}} \quad & R_\mathrm{eff}(\tau_\mathrm{p}, \tau_\mathrm{f}) \\ \text{s.t.} \quad & \tau_\mathrm{p} + \tau_\mathrm{f} < T \end{aligned} \tag{2.22}$$

over the feasible set of positive integer training and feedback lengths. The sum-rate expression defined in (2.20) is not convex as it is defined on a non-convex non-continuous closed set of bounded integers, as would be the case if one were optimizing over the number of feedback bits. Nevertheless, I seek to maximize a continuous relaxation of the defined cost function by typical methods of convex optimization [9, 43].

The optimization can be simplified by realizing that for a fixed overhead length T_OHD, optimizing training and feedback lengths simplifies to minimizing $c_2(\tau_\mathrm{p}, \tau_\mathrm{f})$ and then finding the optimal overhead length. So given a fixed amount of overhead the solution of the first optimization step can be shown to be

$$\tau_\mathrm{p} = \frac{N_\mathrm{R}}{N_\mathrm{R} + \sqrt{K N_\mathrm{T} N_\mathrm{R}}} T_\mathrm{OHD}, \quad \tau_\mathrm{f} = \frac{\sqrt{K N_\mathrm{T} N_\mathrm{R}}}{N_\mathrm{R} + \sqrt{K N_\mathrm{T} N_\mathrm{R}}} T_\mathrm{OHD}, \tag{2.23}$$

which gives the optimal value $c_{opt} = \frac{(N_{\mathrm{R}}+\sqrt{KN_{\mathrm{T}}N_{\mathrm{R}}})^2}{(KN_{\mathrm{T}}-N_{\mathrm{R}})T_{\mathrm{OHD}}}$. Replacing c_{opt} into (2.22), the optimal training and feedback length T_{OHD} can now be found by solving

$$\begin{aligned}\frac{\delta}{\delta T_{\mathrm{OHD}}}\left(\frac{T-T_{\mathrm{OHD}}}{T}\right)\Bigg(\mathbb{E}_{\mathbf{H}}R_{\mathrm{sum}} \\ -\sum_i d_i \log_2\left(1+\frac{\alpha(K-1/d_i)\left(N_{\mathrm{R}}+\sqrt{KN_{\mathrm{T}}N_{\mathrm{R}}}\right)^2}{(KN_{\mathrm{T}}-N_{\mathrm{R}})T_{\mathrm{OHD}}}\right)\Bigg)=0 \\ -\frac{1}{T}\left(\mathbb{E}_{\mathbf{H}}R_{\mathrm{sum}}-\sum d_i\log\left(1+\frac{\gamma_i}{T_{\mathrm{OHD}}}\right)\right) \\ +\frac{T-T_{\mathrm{OHD}}}{T}\left(\sum\frac{d_i}{T_{\mathrm{OHD}}(T_{\mathrm{OHD}}+\gamma_i)}\right)=0,\end{aligned} \tag{2.24}$$

where $\gamma_i = \frac{\alpha(K-1/d_i)(N_{\mathrm{R}}+\sqrt{KN_{\mathrm{T}}N_{\mathrm{R}}})^2}{(KN_{\mathrm{T}}-N_{\mathrm{R}})}$. Though the exact solution to this optimization cannot be found in closed form, it can be shown that T_{OHD} increases with $\sqrt{T}$, and increases with the ratio of transmit to feedback power $\frac{P}{P_{\mathrm{F}}}$. Fig. 2.2 plots the optimal feedback vs. frame length, where I have solved the optimization in (2.24) numerically for a three 2×2 user network with $d_i = 1$. Fig. 2.2 verifies the claimed scaling and shows that for a range of reasonable frame lengths, the solution to the optimization problem is less than the minimum length required to satisfy the dimensionality constraints on the training and feedback matrices, i.e. the optimal overhead is minimal for realistic frame sizes.

While it may not be surprising that longer frames can support more training [43], it's interesting to note that small mismatches in forward and feedback SNR need not dramatically increase training and feedback lengths. Hence, a slightly noisy feedback channel does not require significant extra

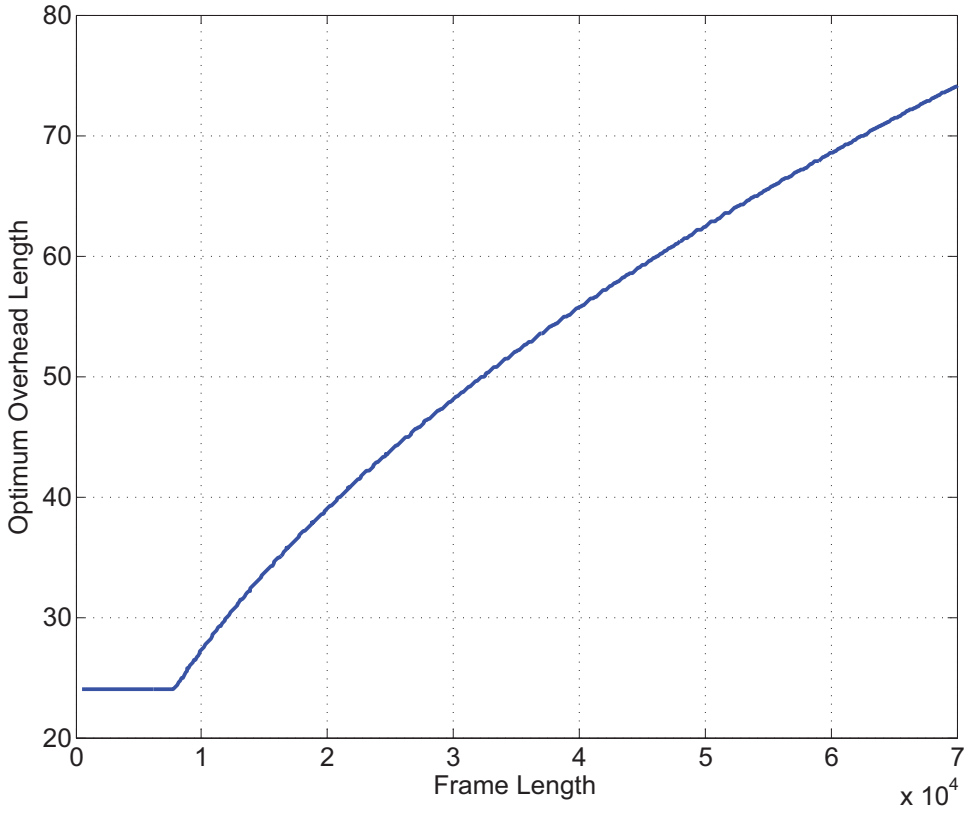

Figure 2.2: Optimal Overhead Length vs. Frame Length for a 3 user 2×2 system with $d_i = 1 \; \forall i$ and $P_{\mathrm{F}} = P$.

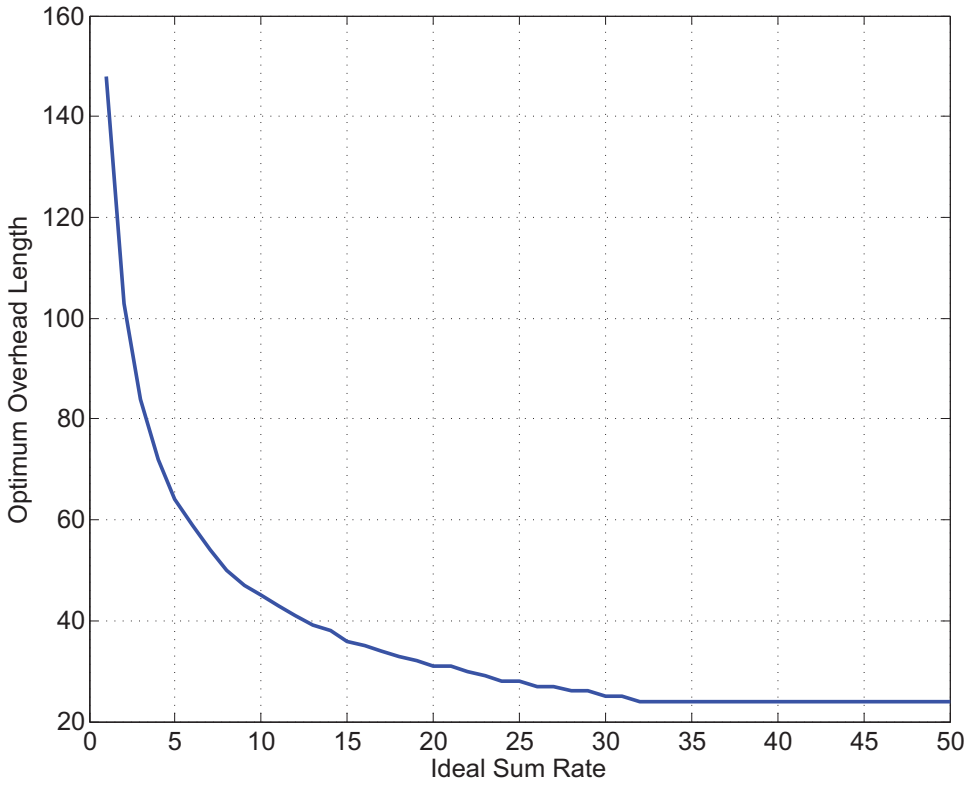

Figure 2.3: Optimal Overhead length vs. R_{sum} for a 3 user 2×2 system with $d_i = 1\ \forall i$, SNR=40dB and $P_{\text{F}} = P/10$.

training to compensate. This, however, is not true of very poor feedback channels, as the optimal overhead length does in fact increase to improve the quality of CSI. It can also be shown that, all else fixed, the optimal training length decreases with the achieved sum-rate or effectively SNR, making analog feedback especially efficient at high SNR. This is shown if Fig. 2.3.

2.6 Simulation Results

In this section, I present simulation results that validate the claims and proofs given in Sections 2.4.2. I verify the results shown in Theorems 2 and 4 which state that as long as the transmit power on the feedback channel scales sufficiently with the power on the forward channel, the multiplexing gain achieved by perfect IA is preserved. To better show the performance of both IA and IA with analog feedback, I remove the restriction of per-stream receivers and thus calculate the sum-rate of a joint decoder which also treats interference as colored Gaussian noise

$$R_{\text{sum}} = \sum_{i=1}^{K} \log_2 \left| \mathbf{I} + \frac{P}{d_i} \mathbf{H}_{i,i} \mathbf{F}_i \mathbf{F}_i^* \mathbf{H}_{i,i}^* \left(\sigma^2 \mathbf{I} + \sum_{k \neq i} \frac{P}{d_k} \mathbf{H}_{i,k} \mathbf{F}_k \mathbf{F}_k^* \mathbf{H}_{i,k}^* \right)^{-1} \right|,$$

and the precoders are calculated given ideal or estimated CSI [6].

Fig. 2.4 shows the sum-rate achieved by the IA algorithm in [75] in a 3 user 5×4 system with $d_i = 2 \; \forall i$. I show performance with perfect CSI, scaling quality CSI where $P_{\text{F}} = P/2$, slower scaling CSI with $P_{\text{F}} = P^{\beta}$ and $\beta = 0.5$ and fixed quality CSI where the SNR on the feedback channel is fixed at 5dB. In all cases, minimum training and feedback lengths are used, i.e. $\tau_{\text{p}} = 12$ and $\tau_{\text{f}} = 45$. Fig. 2.4 confirms that both perfect and scaling feedback exhibit the same sum-rate scaling or degrees of freedom. This establishes the multiplexing gain optimality of using analog feedback. Fig. 2.4 also confirms the fact that the mean loss in sum-rate at high enough SNR is indeed a constant independent of the forward channel SNR. Moreover, Fig. 2.4 shows that while the rate

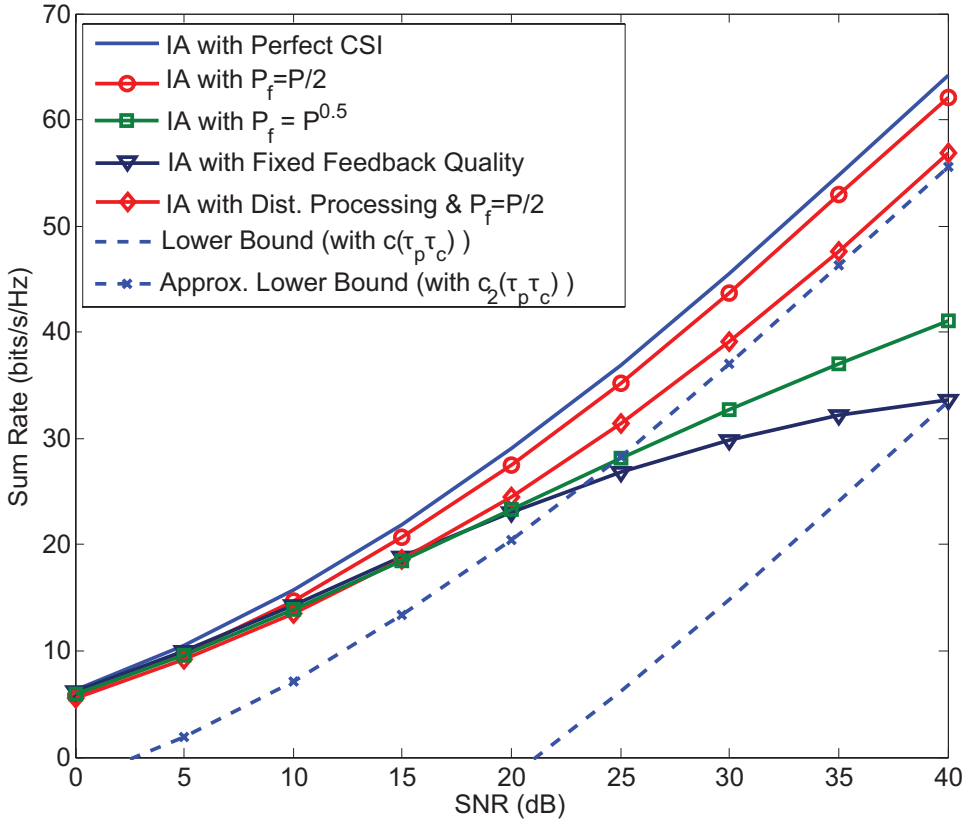

Figure 2.4: The sum-rate achieved by IA in a 3 users 5×4 system ($d_i = 2\ \forall i$) with perfect CSI, scaling feedback quality ($P_{\mathrm{F}} = P/2$), slower scaling feedback ($P_{\mathrm{F}} = \sqrt{P}$), and fixed quality feedback with $SNR_f = 5dB$.

loss bound with $c(\tau_{\mathrm{p}}, \tau_{\mathrm{f}})$ suffices to establish a constant rate loss and multiplexing gain optimality, it is loose for larger systems as mentioned in Section 2.4.2. Using $c_2(\tau_{\mathrm{p}}, \tau_{\mathrm{f}})$, by assuming a biunitary invariant error distribution, gives a better characterization of achieved sum-rate. Moreover, both rate loss characterizations, as well as simulated rate loss, naturally decrease with τ_{p} and τ_{f} as stated in Section 2.4.2; the trend can be seen in Fig. 2.5 for a 2×2 system with $P_{\mathrm{F}} = P/10$. When perfect scaling feedback is not possible, Fig. 2.4 shows

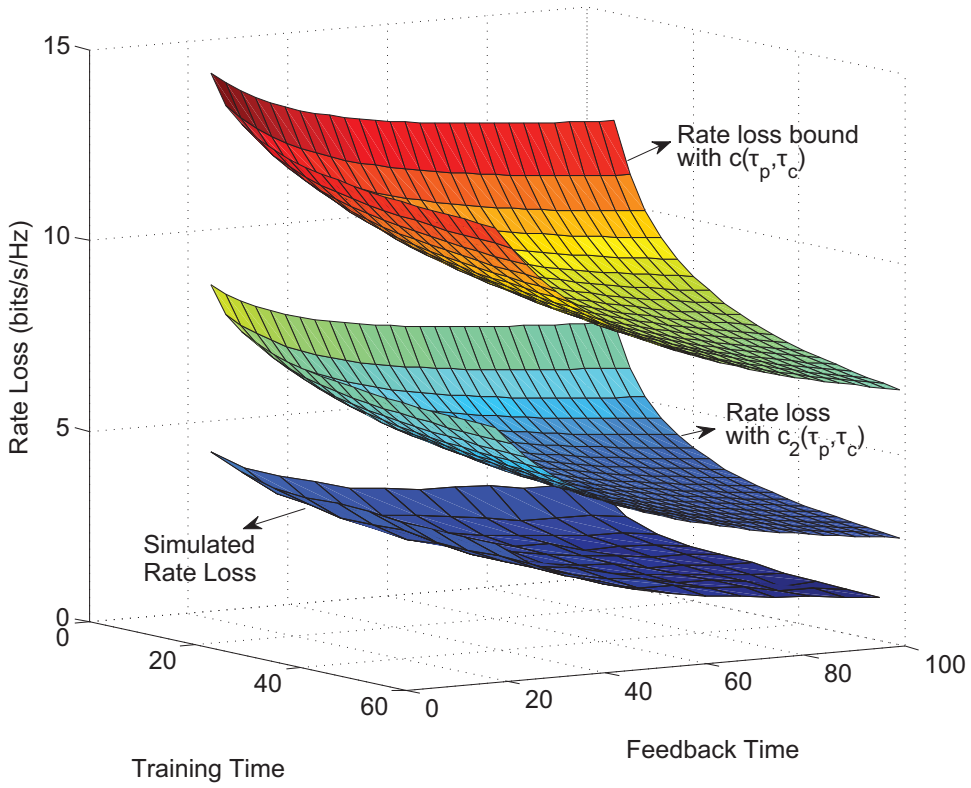

Figure 2.5: Rate loss vs. training and feedback times for a 3 user 2×2 system with $d_i = 1 \ \forall i$ and $P_{\mathrm{F}} = P/10$.

the slower yet linear scaling in the case of $P_{\mathrm{F}} = P^{\beta}$ which verifies the result shown in Theorem 4 on the preservation of a fraction of the system's original multiplexing gain. Finally, for the case of fixed feedback quality, multiplexing gain is zero and the sum-rate saturates at high SNR.

Fig. 2.4 also shows the performance of the distributed processing approach introduced in Section 2.4.1. As stated earlier, the cooperation assumed in Section 2.4.1 is only practical in certain cases such as cellular systems. If cooperation is not possible, either a central node can calculate the IA solution and feed it forward to the other sources, or nodes calculate precoders indepen-

dently. The analysis of the centralized processor strategy is straightforward. It only adds Gaussian noise to the precoders due to feed forward; i.e. all nodes will use noisy versions of the same vector, $\widehat{\mathbf{f}}_i^\ell + \widetilde{\mathbf{f}}_i^\ell$ and $\widehat{\mathbf{w}}_i^\ell + \widetilde{\mathbf{w}}_i^\ell$. So, the interference terms are now given by

$$(\widehat{\mathbf{w}}_i^\ell + \widetilde{\mathbf{w}}_i^\ell)^*(\underbrace{\widehat{\mathbf{H}}_{i,k} + \widetilde{\mathbf{H}}_{i,k}}_{\mathbf{H}_{i,k}})(\widehat{\mathbf{f}}_i^\ell + \widetilde{\mathbf{f}}_i^\ell) = (\widehat{\mathbf{w}}_i^m)^* \underbrace{\widetilde{\mathbf{H}}_{i,k}}_{{P_\mathrm{F}}^{-1}} \widehat{\mathbf{f}}_k^\ell + (\widehat{\mathbf{w}}_i^\ell)^* \mathbf{H}_{i,k} \underbrace{\widetilde{\mathbf{f}}_i^\ell}_{{P_\mathrm{F}}^{-1}}$$
$$+ \underbrace{(\widetilde{\mathbf{w}}_i^\ell)^*}_{{P_\mathrm{F}}^{-1}} \mathbf{H}_{i,k} \widehat{\mathbf{f}}_i^\ell + \underbrace{(\widetilde{\mathbf{w}}_i^\ell)^*}_{{P_\mathrm{F}}^{-1}} \mathbf{H}_{i,k} \underbrace{\widetilde{\mathbf{f}}_i^\ell}_{{P_\mathrm{F}}^{-1}}.$$

All errors in this case again decay with SNR as indicated above, and similar bounds on rate loss can be readily derived. In the case of distributed processing, the extra loss is due to the mismatch of CSI between nodes; IA solutions based on different perturbations of the same channels will be mismatched leading to extra interference leakage. Fig. 2.4, however, shows that the extra loss due to distributed processing is small, and no degrees of freedom are lost. The performance of distributed processing is not theoretically surprising. Since IA solutions and algorithms depend heavily on invariant and singular subspaces, Wedin's theorem [97, 107] and results in [96] can be used to tractably bound the angles between singular and invariant subspaces of perturbed matrices. Using the bounds in [97], I can see that even if nodes compute precoders using different perturbed channels, the angle (or error) between the different precoders calculated is bounded by a linear function of $\|\widetilde{\mathbf{H}}_{i,k}\|_F^2$. Therefore, errors still decay with feedback power. Such decay is all that is needed to prove the multiplexing gain preservation and a constant rate loss. Computing a tight constant bound on ΔR_{sum}, however, becomes significantly more involved. The

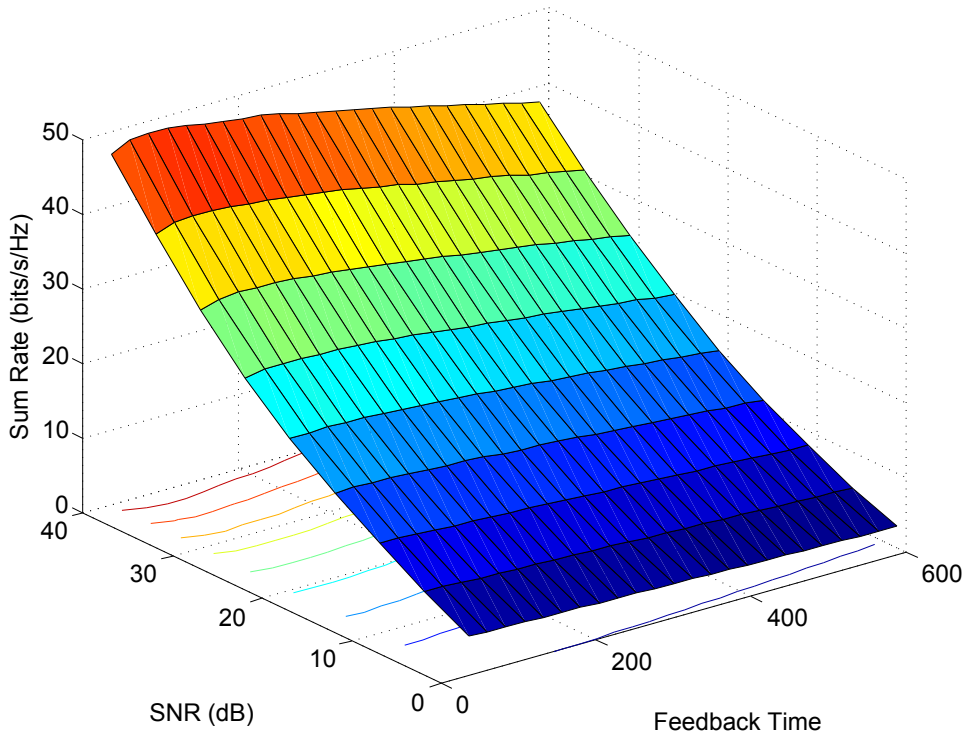

Figure 2.6: The effective throughput achieved by interference alignment in a 3 user 5×4 with $P_{\mathrm{F}} = P/100$ and a frame length $T = 2000$. This confirms that the amount of feedback needed to achieve optimal throughput is close to minimal.

small loss due to distributed processing, and the fact that it does not need extra overhead, make it a practical and viable approach.

Finally, I simulate the system's total throughput according to the overhead model presented in Section 2.5. Fig. 2.6 shows that when the forward and reverse channel SNR scale together, the optimal feedback length is close to the theoretical minimum required for a frame length of 2000 even for the

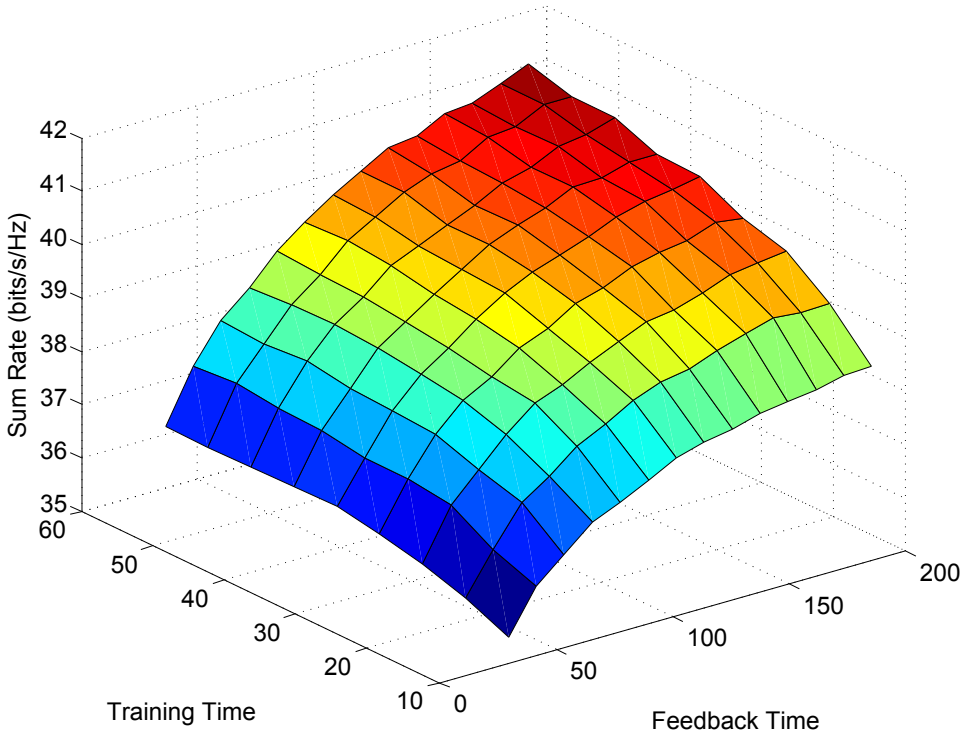

Figure 2.7: This plots the sum-rate achieved by IA in a 3 user 5×4 at 35dB with and feedback SNR of 10dB and a frame length $T = 10000$.

relatively poor feedback channel ($P_{\mathrm{F}} = P/100$) considered. Although analog feedback provides poor overhead scaling with frame length, the effect of this scaling is little since overhead is minimal for practical frames [43]. Finally, Fig 2.6 shows the optimal feedback length assuming fixed training for ease of exposition. I know, however, from Section 2.5, that this is not optimal. Fig. 2.7 shows that while generally less resources are used for training, both training and feedback must scale together for maximum throughput.

2.7 Conclusion

In this chapter, I proposed a low overhead feedback strategy for the interference channel. I showed that when combined with interference alignment, analog feedback can achieve full multiplexing gain when the forward and reverse channel SNR levels are comparable. When such symmetry is not possible, I showed a fraction of the degrees of freedom is retained. Thus a main benefit of analog feedback is that the cost of imperfect channel knowledge at the transmitter is bounded and quickly becomes negligible at high SNR. The mild requirement of comparable feedback and transmit power implies that analog feedback performs well with constant overhead, in the high SNR regime where IA is optimal. In addition to quantifying the cost of imperfect CSI, I show the scaling of required overhead with several network variables such as SNR and frame length. In simulation, I show that the throughput loss due to the overhead of training and analog feedback is often minimal.

2.8 Appendix

2.8.1 Proof of Lemma 1

Consider [8, 21, 75] or similar solutions to finding unitary IA precoders $\mathbf{F}_i = \left[\mathbf{f}_i^1, \ldots, \mathbf{f}_i^{d_i}\right]$. In all such solutions, $\mathbf{F}_i$ $\forall i$ are functions of all interfering channels $\mathbf{H}_{k,\ell}$ $\forall k, \ell,\ k \neq \ell$ only. Since $\mathbf{F}_i$ is not a function of $\mathbf{H}_{i,i}$, and channels are independent across users, this implies that $\mathbf{F}_i$ is independent of $\mathbf{H}_{i,i}$ $\forall i$. Therefore, $\mathbf{H}_{i,i}\mathbf{f}_i^m$ are Gaussian vectors with covariance $\mathbb{E}\left[\mathbf{H}_{i,i}\mathbf{f}_i^m(\mathbf{f}_i^m)^*\mathbf{H}_{i,i}^*\right] = \text{trace}\left(\mathbf{f}_i^m(\mathbf{f}_i^m)^*\right)\mathbf{I}_{N_\text{R}} = \mathbf{I}_{N_\text{R}}$ [64]. Similarly, due to the unitary property of

the precoders, $\mathbb{E}\left[\mathbf{H}_{i,i}\mathbf{f}_i^m(\mathbf{f}_i^\ell)^*\mathbf{H}_{i,i}^*\right] = \text{trace}\left(\mathbf{f}_i^m(\mathbf{f}_i^\ell)^*\right)\mathbf{I}_{N_\text{R}} = 0_{N_\text{R}}$, and therefore $\mathbf{H}_{i,i}\mathbf{f}_i^m$ and $\mathbf{H}_{i,i}\mathbf{f}_i^\ell$ are independent $\forall \ell \neq m$. Vectors $\widehat{\mathbf{f}}_i^m$ are calculated similarly based on $\widehat{\mathbf{H}}_{k,\ell}\ \forall k, \ell,\ k \neq \ell$. Since $\widehat{\mathbf{H}}_{k,\ell}\ \forall k, \ell,\ k \neq \ell$ were shown to be independent in Section 2.4.1, $\widehat{\mathbf{f}}_i^m$ are thus independent of both $\widehat{\mathbf{H}}_{i,i}$ and $\mathbf{H}_{i,i}$. Hence, $\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^\ell$ satisfy the same properties as $\mathbf{H}_{i,i}\mathbf{f}_i^\ell$.

The combiners $\mathbf{w}_i^m$ must now satisfy (2.3); for continuously distributed i.i.d channels (2.4) will be satisfied automatically [8]. This can be done by letting

$$\mathbf{w}_i^m = \mathbf{U}_{min}\left(\left[\mathbf{H}_{i,1}\mathbf{F}_1, \ldots, \mathbf{H}_{i,i-1}\mathbf{F}_{i-1}, \mathbf{H}_{i,i}\mathbf{F}_i^{(m)}, \mathbf{H}_{i,i+1}\mathbf{F}_{i+1}, \ldots, \mathbf{H}_{i,K}\mathbf{F}_K\right]\right), \tag{2.25}$$

where $\mathbf{F}_i^{(m)} = \left[\mathbf{f}_1^1, \ldots \mathbf{f}_1^{m-1}, \mathbf{f}_1^{m+1}, \ldots, \mathbf{f}_1^{d_i}\right]$, and $\mathbf{U}_{min}(\mathbf{A})$ extracts the least dominant left singular vector of $\mathbf{A}$. Given that IA is feasible, the interference matrix in (2.25) will span at most $N_\text{R} - 1$ dimensions (at most $N_\text{R} - d_i$ and $d_i - 1$ dimensions of inter-user and inter-stream interference respectively). Hence, $\mathbf{w}_i^m$ will always correspond to a 0 singular value, thus satisfying (2.3).

So $\mathbf{w}_i^m$ are a function of vectors $\mathbf{H}_{i,k}\mathbf{f}_k^\ell\ \forall k \neq i, \forall \ell$ and $\mathbf{H}_{i,i}\mathbf{f}_i^\ell,\ \forall \ell \neq m$. But $\mathbf{H}_{i,k}\mathbf{f}_k^\ell\ \forall k \neq i, \forall \ell$ are independent of $\mathbf{H}_{i,i}\mathbf{f}_i^m$ due to the independence of channels, and $\mathbf{H}_{i,i}\mathbf{f}_i^\ell,\ \forall \ell \neq m$ are independent of $\mathbf{H}_{i,i}\mathbf{f}_i^m$ as shown earlier. Thus $\mathbf{w}_i^m$ is independent of $\mathbf{H}_{i,i}\mathbf{f}_i^m$. Since $\mathbf{w}_i^m$ and $\mathbf{H}_{i,i}\mathbf{f}_i^m$ are independent, and since $\mathbf{H}_{i,i}\mathbf{f}_i^m$ has a unitary invariant Gaussian distribution as shown earlier [102], one can perform a change of basis such that $\mathbf{w}_i^m = [1, 0, \ldots, 0]^*$. After this change of basis $(\mathbf{w}_i^m)^*\mathbf{H}_{i,i}\mathbf{f}_i^m$ is simply the first element of $\mathbf{H}_{i,i}\mathbf{f}_i^m$ which I have

shown above is $\mathcal{CN}(0,1)$. Therefore, the terms $|(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2$ are exponentially distributed with parameter 1.

The vectors $\widehat{\mathbf{w}}_i^m$ are again given by (2.25) using $\widehat{\mathbf{H}}_{k,\ell}$ and $\widehat{\mathbf{f}}_k^\ell$ instead of $\mathbf{H}_{k,\ell}$ and $\mathbf{f}_k^\ell$ respectively, and are thus again functions of variables independent of both $\widehat{\mathbf{H}}_{i,i}\widehat{\mathbf{f}}_i^m$ and $\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m$. Thus $\widehat{\mathbf{w}}_i^m$ are independent of $\mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m$ and again a change of basis reveals that $(\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m$ are $\mathcal{CN}(0,1)$ and $\left|(\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2$ are exponentially distributed with parameter 1. Therefore, $|(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2$ and $\left|(\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2$ are identically and exponentially distributed with parameter 1.

Chapter 3

Grassmannian Differential Limited Feedback for Interference Alignment

As shown in Chapters 1 and 2, transmit-side channel knowledge can be used to align and cancel interference, in some cases achieving the interference channel's maximum degrees-of-freedom. When alignment is done by coding over frequency extensions in a single antenna system, the required channel state information (CSI) lies on the Grassmannian manifold and its structure can be exploited in feedback. Unfortunately, the number of channels to be shared grows with the square of the number of users, creating too much overhead with conventional limited feedback methods. This chapter proposes Grassmannian differential feedback to reduce overhead by exploiting both the channel's temporal correlation and Grassmannian structure. The performance of the proposed algorithm is characterized both analytically and numerically as a function of channel length, mobility, and the number of feedback bits. The main conclusions are that the proposed feedback strategy allows interference alignment to perform well over a wide range of Doppler spreads, and to approach perfect CSI performance in slowly varying channels. Numerical results highlight the trade-off between the frequency of feedback and the accuracy of individual feedback updates.

3.1 Introduction

Several approaches have been proposed to reduce feedback overhead in IA networks [7, 26, 45, 69, 74]. In Chapter 2, I proposed to feedback the MIMO channel matrices in an analog fashion to fulfill IA's CSI requirement. When interference alignment is considered in frequency selective single-antenna systems, however, the CSI required for alignment has additional manifold structure that can be exploited. In Section 2.1, I briefly discussed the work of [7, 45] in which memoryless quantization with Grassmannian codebooks is used in IA networks. The main limitation of [7, 45] is that good sum rate performance is only guaranteed by scaling the number of feedback bits with SNR. Such scaling leads to an exponential growth in complexity and could increase feedback overhead. The analog feedback strategy presented in Chapter 2 partially overcomes the problem of scaling complexity since the CSI distortion decreases naturally with SNR [69]. Analog feedback, however, is only optimal when the number of feedback channel uses matches the number of symbols being fed back. Further, a main limitation of both analog feedback and the quantization strategies in [7, 45] is that they neglect the channel's temporal correlation which can be exploited to further reduce overhead and improve CSI quality.

Differential and predictive feedback strategies have been proposed for compressing various forms of CSI in time-varying channels [34, 41, 48, 52, 80, 82, 83, 101]. Simplified differential strategies, such as [48], have been considered in commercial wireless standards to feedback channel quality indicators. In [83] differential feedback of multiple-input multiple-output (MIMO) chan-

nel correlation matrices is proposed by leveraging the geometry of positive definite Hermitian matrices. The strategy in [83] was shown to work well when implemented on a wireless testbed [82]. In [41] the differential feedback of unitary precoders is considered in a single-user MIMO setting. The geometry of IA's required CSI, however, neither fits the structure considered in [41, 48, 82, 83], nor the parametrization derived in [80]. Thus, the feedback strategies in [41, 48, 80, 82, 83] cannot be used. In [101], tree-structured beamforming codebooks are constructed allowing receivers to feedback differentially encoded information about a local neighborhood in the feedback tree. In [52] differential feedback is applied to beamforming vectors that are gradually rotated on the Grassmann manifold using a Householder matrix to improve quantization accuracy. In [34] Grassmannian beamforming vectors are fed back and reconstructed using predictive vector quantization tools, leveraging manifold concepts such as tangent vectors and geodesic curves. Unlike the single-user system considered in [52], however, IA systems are sensitive to leakage interference, and require more feedback than the broadcast channel [34, 101]. As a result, the application of [34, 52, 101] to IA networks is non-trivial and it is not clear whether they can achieve sufficient performance with practical codebook sizes.

3.2 Contributions

In this chapter, I propose a limited feedback strategy based on Grassmannian differential feedback to fulfill the CSI requirement in wideband single

antenna systems using IA over frequency extensions. Frequency extensions provide a practical source of dimensionality for IA precoding in SISO interference channels and play an integral role in achieving the maximum multiplexing gain in both single and multiple antenna systems [8]. Although I present results for single antenna systems, the feedback strategy can be generalized to systems with multiple antennas in a manner similar to [45]. The proposed strategy tracks the slow evolution of the normalized channel impulse responses on the Grassmannian manifold. At each feedback update, a quantized tangent vector relating consecutive channel realizations is used at the transmitter to reconstruct the channel. Since the tangent space geometry changes at each iteration, and varying Doppler spreads lead to different channel dynamics, I construct quantization codebooks that adapt to the channel's geometry and dynamics. Unlike [34], no prediction is performed and differential feedback is used in the new context of SISO interference channels. I show that the performance of the proposed differential strategy cannot be improved upon by a simple yet commonly used class of linear or geodesic predictors. Thus, the development of more sophisticated prediction algorithms that can further improve CSI quality remains a promising venue for future work.

I derive an approximation for the average distortion achieved by the proposed algorithm and show that it is accurate for a wide range of channel and feedback parameters. Unlike the analysis of the Householder strategy in [52] and the predictive strategy in [34], which neglect either the quantization process or the channel's Doppler spread, the derived performance characteri-

zation captures the effects of both fading speed and codebook size. Simulation results show that the proposed algorithm outperforms memoryless quantization and other competitive feedback strategies for temporally correlated channels [34]. Through simulation, I highlight also the trade-off between feedback refresh rate and codebook size [42]. I show that, from a CSI distortion perspective, frequent updates using smaller codebooks may often be preferred to less-frequent updates with a large codebook. Finally, while scaling codebook size is still required to preserve multiplexing gain, I show that IA with the proposed feedback strategy performs well in the range of medium to high SNR with a relatively small number of feedback bits.

This chapter is organized as follows. Section 3.3 introduces the SISO interference channel with orthogonal frequency division multiplexing (OFDM). Section 3.4 reviews the concept of interference alignment over frequency extended channels and highlights the degradation due to imperfect channel knowledge. Section 3.5 presents the proposed feedback framework while Section 3.6 elaborates on its main design parameters. Sections 3.7 and 3.8 present analytical and numerical results respectively on the performance of the Grassmannian differential feedback, as well as the performance of IA when the proposed strategy is used for CSI feedback. I conclude with Section 3.9.

3.3 System Model

Consider a frequency selective SISO interference channel with K communicating user pairs. Each user k communicates desired data to its paired

receiver k and interferes with all other receivers $\ell \neq k$. The wideband channel between transmitter ℓ and receiver k at time t is given by the L-tap channel impulse response vector $\mathbf{h}_{k,\ell}[t] = [h_{k,\ell}[t,0]^*, h_{k,\ell}[t,1]^*, \ldots, h_{k,\ell}[t,L-1]^*]^* \ \forall k,\ell \in \{1,\ldots,K\}$. Channel impulse responses are drawn independently across k and ℓ from a continuous distribution and have covariance matrix $\mathbb{E}\left[\mathbf{h}_{k,\ell}[t]\mathbf{h}_{k,\ell}[t]^*\right] = \mathbf{R}_{\mathbf{h}_{k,\ell}},\ \forall k,\ell \in \{1,\ldots,K\}$ and $\forall t$. Letting $\rho_{k,\ell}$ be the average attenuation of user ℓ's signal at user k's receiver, the channel response satisfies $\text{trace}\left(\mathbf{R}_{\mathbf{h}_{k,\ell}}\right) = \rho_{k,\ell}$. In the case of uncorrelated channel taps, $\mathbf{R}_{\mathbf{h}_{k,\ell}} = \text{diag}\left(\mathbf{p}_{k,\ell}\right)$, where $\mathbf{p}_{k,\ell}$ is the $L \times 1$ vector representing the channel's power delay profile.

Using orthogonal frequency division multiplexing (OFDM), the transmitters transform the observed frequency selective channels into a set of N_{sc} non-interfering narrowband subcarriers. Stacking each received OFDM symbol in a vector, the input-output relationship can be written in matrix form as

$$\mathbf{y}_k[t] = \mathbf{H}_{k,k}[t]\mathbf{x}_k[t] + \sum_{\ell \neq k} \mathbf{H}_{k,\ell}[t]\mathbf{x}_\ell[t] + \mathbf{v}_k[t], \tag{3.1}$$

where $\mathbf{x}_k[t]$ is the OFDM symbol sent by user k at time t with the average power constraint $\mathbb{E}\left[\|\mathbf{x}_k[t]\|^2\right] = N_{\text{sc}}P$, the $N_{\text{sc}} \times N_{\text{sc}}$ matrix $\mathbf{H}_{k,\ell}[t] = \text{diag}\left(\boldsymbol{\mathcal{F}}_{N_{\text{sc}}}\left[\mathbf{h}_{k,\ell}^*[t],\ 0_{N_{\text{sc}}-L}\right]^*\right)$ represents the channel frequency response between transmitter ℓ and receiver k at time t, and $\mathbf{v}_k[t]$ is the i.i.d. $\mathcal{CN}(0, \sigma_{\text{n}}^2\mathbf{I}_{N_{\text{sc}}})$ thermal noise observed by user k. The system model assumes perfect time and frequency synchronization, and a cyclic prefix that is long enough to accommodate the impulse response of all channels as well as the potential propagation delay between users.

Further, the relationship in (3.1) implicitly assumes that the L-tap channels $\mathbf{h}_{k,\ell}[t]$ seen by the t^{th} OFDM symbol remain constant over the OFDM symbol time and pilots can be used to estimate them at the receiver. The channels over consecutive OFDM symbols, however, are considered to be slowly varying. Let f_{D} be the Doppler frequency of the channel and T_{s} be the OFDM symbol time. I assume channels are temporally correlated according to the model proposed by Clarke [14], i.e $\mathbb{E}\left[\rho_{k,\ell}^{-1}\left|\mathbf{h}_{k,\ell}^{*}[t-m]\mathbf{h}_{k,\ell}[t]\right|\right] = J_0(2\pi f_{\text{D}} T_{\text{s}} m)$ where J_0 is the 0-th order Bessel function of the first kind. A main underlying assumption, however, is that while the channels vary slowly over time, their length L remains fixed. Since channel response length, is likely to vary slower than the channel response itself, and since it is constrained to be a small integer, the channel length can be fed back to the transmitter if necessary at a small extra overhead cost. In general, the proposed feedback algorithm can be applied to all correlated channels, however, further details on the mathematical structure of the temporally correlated channel are given whenever a result is model-dependent.

Throughout the remainder of this chapter, I make the following two system-level assumptions:

- I assume that the channels $\mathbf{h}_{k,\ell}[t]\ \forall \ell$ are known to receiver k. This implicitly assumes that the set of K transmitters coordinate (or orthogonalize) their pilot transmissions allowing receivers to estimate all incoming channels. I make a further simplification by assuming receiver CSI is perfect.

- I assume each receiver has an error-free logical feedback link to all transmitters. The feedback link from receiver k to transmitter ℓ may either be a direct link, or a link through other nodes such as transmitter k and/or a centralized processor.

While the importance of transmitter channel knowledge makes the two assumptions rather standard in the literature on precoding for the interference channel [7, 8, 12, 26, 45, 69, 73–75, 85], or interference coordination in general [17], they are by no means trivial. Coordinating training phases, feedback phases, and centralized processing constitutes significant complexity in both cellular and ad-hoc systems. I refer the interested reader to [17] and the references therein for a detailed discussion of the limitations, the benefits, and the possible solutions that enable coordination in the case of cellular systems.

3.4 Interference Alignment in Frequency

In this section I review the concept of IA over frequency extensions when perfect channel state information is available at the transmitter W then summarize the effect of imperfect transmitter channel knowledge on the performance of IA.

3.4.1 IA with Perfect CSI at the Transmitter

IA for the SISO interference channel can achieve the maximum degrees of freedom when coding over infinite channel extensions [8]. Using IA over N frequency extensions, each transmitter k at time t sends $d_k < N$ symbols,

$x_k^m[t]$, along the $N \times 1$ precoding vectors $\mathbf{f}_k^m[t]$. As a result, the transmitted symbol is

$$\mathbf{x}_k[t] = \sum_{m=1}^{d_k} \mathbf{f}_k^m[t] x_k^m[t], \tag{3.2}$$

where $\|\mathbf{f}_k^m[t]\|_2 = 1$ and $\mathbb{E}\left[|x_k^m[t]|^2\right] = NP/d_k$, such that the power in each N subcarriers is NP. The transmit directions $\mathbf{f}_k^m[t]$ are calculated such that the interference from $K-1$ users is aligned at all receivers, leaving at least d_k interference free dimensions for the desired signal. The number of transmitted symbols d_k can be chosen according to the original strategy in [8] or the improved method in [12], both of which asymptotically achieve the maximum degrees of freedom, i.e. $\lim_{N\longrightarrow\infty} \frac{1}{N} \sum d_k = K/2$.

Note that users need not code over all N_{sc} subcarriers since algorithm complexity increases significantly with the number of subcarriers whereas the achieved multiplexing gain may increase very slowly with N_{sc} as demonstrated by the IA solutions in [8]. Users could potentially treat any N-subset of subcarriers as an independent coding group and thus have $\lfloor \frac{N_{\text{sc}}}{N} \rfloor$ parallel alignment groups in an OFDM symbol. In this case (3.2) applies to each group of subcarriers, and one can write $\lfloor \frac{N_{\text{sc}}}{N} \rfloor$ relationships for each OFDM symbol. For simplicity of exposition, in the remainder of this section I assume $\lfloor \frac{N_{\text{sc}}}{N} \rfloor = 1$.

To quantify concisely the degradation due to imperfect CSI, I consider a zero-forcing receiver; other receiver designs can be used. At the output of

the linear receivers $\mathbf{w}_k^m[t]$, the received signal is

$$\begin{aligned}\mathbf{w}_k^m[t]^*\mathbf{y}_k[t] =&\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^m[t]x_k^m[t] + \sum_{\ell\neq m}\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^\ell[t]x_k^\ell[t] \\ &+ \sum_{i\neq k}\sum_{\ell=1}^{d_i}\mathbf{w}_k^m[t]^*\mathbf{H}_{k,i}[t]\mathbf{f}_i^\ell[t]x_i^\ell[t] + \mathbf{w}_k^m[t]^*\mathbf{v}_k[t],\end{aligned} \tag{3.3}$$

for $m \in \{1,\ldots,d_k\}$ and $k \in \{1,\ldots,K\}$, where $\|\mathbf{w}_k^m[t]\|^2 = 1$. Assuming a zero-forcing receiver, the conditions for perfect interference alignment are

$$\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^\ell[t] = 0 \qquad \forall k,\ \ell \neq m \tag{3.4}$$

$$\mathbf{w}_k^m[t]^*\mathbf{H}_{k,i}[t]\mathbf{f}_i^\ell[t] = 0 \qquad \forall i \neq k,\ \text{and } \forall m,\ell \tag{3.5}$$

$$|\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^m[t]| \geq c > 0 \qquad \forall k,m \tag{3.6}$$

where alignment is achieved by (3.4) and (3.5), while (3.6) ensures the decodability of the d_k desired streams.

The sum-rate achieved by the linear zero-forcing receiver [66], assuming Gaussian input signals, is

$$\bar{R}_{\text{sum}} = \sum_{k=1}^{K}\sum_{m=1}^{d_k}\frac{1}{N_{\text{sc}}}\log_2\left(1 + \frac{\frac{N_{\text{sc}}P}{d_k}\left|\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^m[t]\right|^2}{\mathcal{I}_{k,m}^1[t] + \mathcal{I}_{k,m}^2[t] + \sigma_{\text{n}}^2}\right), \tag{3.7}$$

where $\mathcal{I}_{k,m}^1[t]$ is self-interference from other transmit streams and $\mathcal{I}_{k,m}^2[t]$ is the interference from other users. The two interference terms are

$$\begin{aligned}\mathcal{I}_{k,m}^1[t] &= \sum_{\ell\neq m}\frac{N_{\text{sc}}P}{d_k}\left|\mathbf{w}_k^m[t]^*\mathbf{H}_{k,k}[t]\mathbf{f}_k^\ell[t]\right|^2, \\ \mathcal{I}_{k,m}^2[t] &= \sum_{i\neq k}\sum_{\ell=1}^{d_i}\frac{N_{\text{sc}}P}{d_i}\left|\mathbf{w}_k^m[t]^*\mathbf{H}_{k,i}[t]\mathbf{f}_i^\ell[t]\right|^2.\end{aligned} \tag{3.8}$$

If perfect channel knowledge is available, and the number of symbols d_k are feasible, equations (3.4)-(3.6) can be satisfied with probability one and thus $\mathcal{I}_{k,m}^1 = \mathcal{I}_{k,m}^2 = 0$. This gives

$$\begin{aligned}\lim_{P\to\infty} \frac{\bar{R}_{\text{sum}}}{\log_2 P} &= \lim_{P\to\infty} \frac{\sum\limits_{k,m} \frac{1}{N_{\text{sc}}} \log_2 \left(1 + \frac{\frac{N_{\text{sc}}P}{d_k} \left|\mathbf{w}_k^m[t]^* \mathbf{H}_{k,k}[t] \mathbf{f}_k^m[t]\right|^2}{\sigma_{\text{n}}^2}\right)}{\log_2 P} \\ &= \frac{1}{N_{\text{sc}}} \sum_k d_k \overset{N_{\text{sc}}\to\infty}{\longrightarrow} \frac{K}{2},\end{aligned} \tag{3.9}$$

which confirms the fact that IA achieves the $K/2$ degrees of freedom of the K user interference channel. While the achievability proof in [8] assumed independent fading across subcarriers, [7] has claimed that fading on each subcarrier need not be independent provided that the channel impulse response is long enough.

3.4.2 The Effect of Limited Feedback

In practical systems, receivers cannot feedback the channels $\mathbf{h}_{k,i}[t]$ with infinite precision, thus only a distorted version of $\mathbf{h}_{k,i}[t]$ is available when calculating IA precoders. The distortion caused by limited feedback implies that interference will not be perfectly aligned. The power of residual misaligned interference increases along with transmit power and causes the SINR in the desired signal subspace to saturate, which limits sum-rate at high SNR [66]. As a result, low overhead feedback strategies must be designed to improve CSI accuracy and allow good IA performance.

To establish the impact of limited feedback on IA sum-rate, and to

develop insights into the structure of a good quantizer, I summarize the relevant calculations done in [7] and Chapter 2. In Chapter 2, it is shown that if the transmitters use the estimated or quantized channels $\widehat{\mathbf{H}}_{k,i},\ \forall k, i \in \{1,\ 2,\ \ldots,\ k\}$ to calculate IA precoders, then the mean loss in sum-rate can be upper bounded by

$$\Delta \bar{R}_{\text{sum}} \leq \sum_{k,m} \frac{1}{N_{\text{sc}}} \log_2 \left(1 + \frac{\mathbb{E}_{\mathbf{H}} \left[\mathcal{I}_{k,m}^1 + \mathcal{I}_{k,m}^2 \right]}{\sigma_{\text{n}}^2} \right). \tag{3.10}$$

Using the results from [7, Section IV-B], the components of the sum leakage interference, $\mathcal{I}_{k,m}^1 + \mathcal{I}_{k,m}^2$, can be bounded by

$$\begin{aligned} &\frac{N_{\text{sc}} P}{d_i} \left| \widehat{\mathbf{w}}_k^m[t]^* \mathbf{H}_{k,i}[t] \widehat{\mathbf{f}}_i^\ell[t] \right|^2 \\ &\qquad \leq \frac{N_{\text{sc}} P}{d_i} \|\widehat{\mathbf{w}}_k^m[t] \circ \widehat{\mathbf{f}}_i^\ell[t]\|^2 \|\mathbf{h}_{k,i}[t]\|^2 \left(1 - \left| \frac{\mathbf{h}_{k,i}[t]^* \widehat{\mathbf{h}}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\| \|\widehat{\mathbf{h}}_{k,i}[t]\|} \right|^2 \right), \end{aligned} \tag{3.11}$$

where $\widehat{\mathbf{f}}_i^\ell[t]$ and $\widehat{\mathbf{w}}_k^m[t]$ are the precoders and combiners calculated using imperfect CSI. The bound in (3.11) relates the power of leakage interference to the quality of the quantized CSI and suggests mathematical structure that can be exploited by the quantizer. The last term in (3.11) indicates that only knowledge of the normalized channels $\frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|}$ is necessary for IA, further the CSI is phase invariant since $\left| \frac{\mathbf{h}_{k,i}[t]^* \widehat{\mathbf{h}}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\| \|\widehat{\mathbf{h}}_{k,i}[t]\|} \right|^2 = \left| \frac{e^{j\theta} \mathbf{h}_{k,i}[t]^* \widehat{\mathbf{h}}_{k,i}[t]}{\|e^{j\theta} \mathbf{h}_{k,i}[t]\| \|\widehat{\mathbf{h}}_{k,i}[t]\|} \right|^2$. This is not surprising since, if an interfering channel is scaled and rotated by $\alpha \in \mathbb{C}$, the interference subspace remains unchanged, i.e., $\text{diag}\left(\boldsymbol{\mathcal{F}}_{N_{\text{sc}}} \alpha \mathbf{h}_{k,i}[t]\right) \mathbf{f}_i^m[t] = \alpha \text{diag}\left(\boldsymbol{\mathcal{F}}_{N_{\text{sc}}} \mathbf{h}_{k,i}[t]\right) \mathbf{f}_i^m[t]$. As a result, the CSI needed for IA with frequency extensions evolves on the manifold of L-dimensional unit norm, rotationally

invariant vectors otherwise known as the Grassmannian manifold $\mathcal{G}_{L,1}$ [3, 11, 53, 65]. Define the chordal distance between two points on the Grassmannian manifold, $\mathbf{x}_1$ and $\mathbf{x}_2$, as $d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{1 - |\mathbf{x}_1^* \mathbf{x}_2|^2}$. The bound in (3.11) can be rewritten as

$$\frac{N_{\mathrm{sc}} P}{d_i} \left| \widehat{\mathbf{w}}_k^m[t]^* \mathbf{H}_{k,i}[t] \widehat{\mathbf{f}}_i^{\ell}[t] \right|^2 \leq \frac{N_{\mathrm{sc}} P}{d_i} \|\widehat{\mathbf{w}}_k^m[t] \circ \widehat{\mathbf{f}}_i^{\ell}[t]\|^2 \ \|\mathbf{h}_{k,i}[t]\|^2 \ d(\mathbf{g}_{k,i}[t], \widehat{\mathbf{g}}_{k,i}[t])^2, \tag{3.12}$$

where $\mathbf{g}_{k,i}[t] \triangleq \frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|}$ and $\widehat{\mathbf{g}}_{k,i}[t] \triangleq \frac{\widehat{\mathbf{h}}_{k,i}[t]}{\|\widehat{\mathbf{h}}_{k,i}[t]\|}$, showing that leakage interference can be bounded using the chordal distance between the true and quantized channels [2].

3.5 Grassmannian Differential Feedback

In this section, I present the proposed feedback algorithm. Under some conditions, I show that differential quantization, a special case of predictive vector quantization, cannot be improved upon using a simple yet common class of predictors.

3.5.1 Differential Feedback Framework

I propose to differentially encode the normalized channel response vectors by exploiting both their Grassmannian structure and temporal correlation. The proposed algorithm encodes CSI increments using a tangent vector which defines the geodesic path between consecutive channel realizations. The feedback framework uses tools presented in [2], which were simplified and first used for MISO broadcast channel feedback in [34]. I consider the separate quantiza-

tion of the slowly varying normalized channel vectors $\mathbf{g}_{k,i}[t] \triangleq \frac{\mathbf{h}_{k,i}[t]}{\|\mathbf{h}_{k,i}[t]\|}$ $\forall t \geq 0$; joint quantization may offer higher performance at the expense of additional complexity [5]. Therefore, I restrict my attention to one of the channels and drop the user subscripts.

The smooth structure of the manifold allows each two points to be related using the concept of a geodesic path as shown in Fig. 3.1. The geodesic curve describing the path between $\mathbf{g}[t-1]$ and $\mathbf{g}[t]$ is given by

$$G(\mathbf{g}[t-1], \mathbf{e}[t], \ell) = \mathbf{g}[t-1]\cos\left(\|\mathbf{e}[t]\|\ell\right) + \frac{\mathbf{e}[t]}{\|\mathbf{e}[t]\|}\sin\left(\|\mathbf{e}[t]\|\ell\right). \tag{3.13}$$

where the vector $\mathbf{e}[t]$, also shown in Fig. 3.1, is known as the tangent vector. The tangent vector can be viewed as a length-preserving unwrapping of the geodesic path onto the tangent space at $\mathbf{g}[t-1]$. The tangent vector is given by

$$\mathbf{e}[t] = \tan^{-1}\left(\frac{d[t]}{\rho[t]}\right)\frac{\mathbf{g}[t]/\rho[t] - \mathbf{g}[t-1]}{d[t]/\rho[t]}, \tag{3.14}$$

where $\rho[t] = \mathbf{g}[t-1]^*\mathbf{g}[t]$ and $d[t] = \sqrt{1 - |\rho[t]|^2}$ [2]. From (3.13), it can be verified that $G(\mathbf{g}[t-1], \mathbf{g}[t], 0) = \mathbf{g}[t-1]$, $G(\mathbf{g}[t-1], \mathbf{g}[t], 1) = \mathbf{g}[t]$ and that $\|G(\mathbf{g}[t-1], \mathbf{g}[t], \ell)\| = 1,\ \forall \ell$ since the tangent vector $\mathbf{e}[t]$ is orthogonal to the base vector $\mathbf{g}[t-1]$.

The tangent vector and geodesic path, which together relate two points on the manifold, can be used to build a differential feedback framework to track the evolution of CSI [34]. Given the previous channel realization and a tangent vector received by feedback, the transmitter can reconstruct the current channel by applying (3.13). The operation of the Grassmannian differential

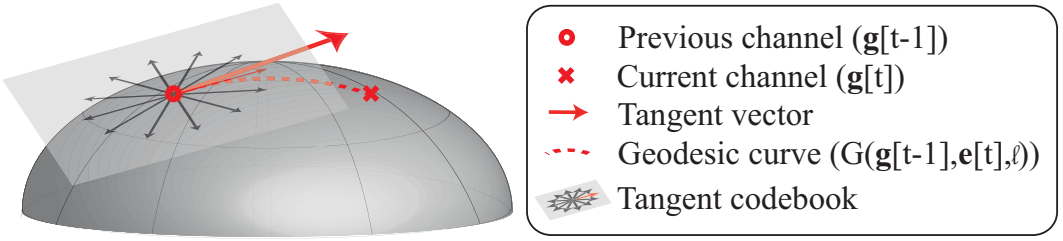

Figure 3.1: A 3D visualization of the Grassmann manifold and the tangent vector quantization process. The quantized tangent which is selected from the tangent codebook is also marked in red.

feedback algorithm at both the transmitter and receiver is given in the block diagram of Fig. 3.2. For clarity, the pseudo code used to encode the channel evolution over time is given in Algorithm 1. At each new channel realization, the receiver estimates the normalized L-tap channel $\mathbf{g}[t]$. Perfect channel estimation is assumed here to decouple the quantization error from the estimation error. Given the current observation, and the quantized channel in the previous iteration, $\widehat{\mathbf{g}}[t-1]$, the receiver calculates the tangent vector, $\mathbf{e}[t]$, from $\widehat{\mathbf{g}}[t-1]$ to $\mathbf{g}[t]$. The receiver then quantizes the tangent vector to produce a quantized tangent, $\widehat{\mathbf{e}}[t]$. The details of the quantization method are given in depth in Section 3.6. This quantized tangent is then fedback to the transmitter over a delay and error free feedback link. Equipped with the previous quantized channel, $\widehat{\mathbf{g}}[t-1]$, and the quantized error vector $\widehat{\mathbf{e}}[t]$, *both transmitter and receiver* can now calculate the new quantized channel $\widehat{\mathbf{g}}[t]$. This vector $\widehat{\mathbf{g}}[t]$ will later serve as the base point in the next iteration. The algorithm runs in parallel for all channels $\mathbf{g}_{k,i}[t],\ \forall k, i$ and the transmitters use the estimated CSI vectors $\widehat{\mathbf{g}}_{k,i}[t],\ \forall k, \ell$ to do interference alignment.

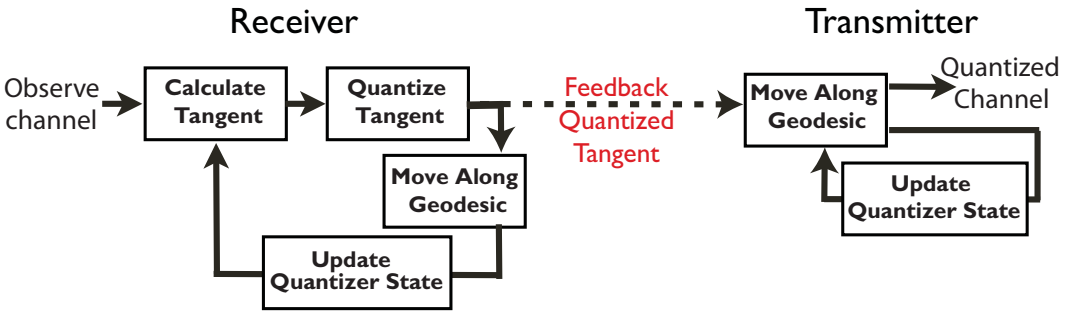

Figure 3.2: Block diagram of Grassmannian differential quantization and feedback at the transmitter and receiver.

Algorithm 1 Receiver

Input: $\widehat{\mathbf{g}}[t-1]$
for all $t = 1, 2, \ldots$ **do**
  Estimate the channel $\mathbf{g}[t]$
  Calculate tangent between $\widehat{\mathbf{g}}[t-1]$ and $\mathbf{g}[t]$ using (3.14)
  Quantize and feedback the quantized tangent vector $\widehat{\mathbf{e}}[t]$
  Reconstruct the quantized channel $\widehat{\mathbf{g}}[t] = G(\widehat{\mathbf{g}}[t-1], \widehat{\mathbf{e}}[t], 1)$
end for

Algorithm 2 Transmitter

Input: $\widehat{\mathbf{g}}[t-1]$
for all $t = 1, 2, \ldots$ **do**
  Receive the feedback tangent vector $\widehat{\mathbf{e}}[t]$
  Reconstruct the quantized channel $\widehat{\mathbf{g}}[t] = G(\widehat{\mathbf{g}}[t-1], \widehat{\mathbf{e}}[t], 1)$
end for
Output: $\widehat{\mathbf{g}}[t]$

3.5.2 Comparing Differential vs. Predictive Quantization

The feedback framework of Section 3.5.1 is a special case of predictive vector quantization, the main difference being that the proposed strategy does not attempt to further improve CSI by *predicting future channel realizations*. In this section, I begin by formalizing the main difference between predictive

and differential feedback. I then show that, for a class of channels, commonly used Grassmannian predictors such as those in [34, 52], yield no performance enhancement over the differential strategy presented. Note, however, that more sophisticated prediction functions may in fact further improve performance. Thus, the development of more intelligent Grasmannian predictors is a promising area for future work.

3.5.2.1 Predictive Quantization

The main idea behind differential feedback is to send the transmitter information that allows it to move from the *old quantized channel* to the *new channel realization*. Thus, a tangent vector $\mathbf{e}[t]$ is calculated between the base point $\widehat{\mathbf{g}}[t-1]$ and $\mathbf{g}[t]$, and is then quantized to give $\widehat{\mathbf{e}}[t]$. Instead of starting the tangent vector at the point $\widehat{\mathbf{g}}[t-1]$, general predictive quantization calculates the tangent starting at *another* point $\widetilde{\mathbf{g}}[t]$, known as the *predicted channel vector*. The premise is that if $\widetilde{\mathbf{g}}[t]$ is "closer" to $\mathbf{g}[t]$ than $\widehat{\mathbf{g}}[t-1]$ (the base point used in differential feedback), quantization accuracy can be further improved. The main difficulty now lies in constructing prediction functions that actually yield closer or better estimates of $\mathbf{g}[t]$.

Unfortunately, work on prediction on the Grassmann manifold is rather limited. In this chapter, I consider a class of linear predictors used in [34, 52]. Though the predictors in [34, 52] are presented differently, the predicted

channel vectors can be written in a unified manner as

$$\widetilde{\mathbf{g}}[t] = G(\widehat{\mathbf{g}}[t-1], \mathbf{e}_{\mathrm{p}}[t], \ell_{\mathrm{p}}) = \widehat{\mathbf{g}}[t-1]\cos\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\ell_{\mathrm{p}}\right) + \frac{\mathbf{e}_{\mathrm{p}}[t]}{\|\mathbf{e}_{\mathrm{p}}[t]\|}\sin\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\ell_{\mathrm{p}}\right), \tag{3.15}$$

where $\mathbf{e}_{\mathrm{p}}[t]$ is the *predicted tangent vector* and ℓ_{p} is a predicted step size. Since the norm of $\mathbf{e}_{\mathrm{p}}[t]$ can absorb any prediction step size ℓ_P, one can enforce $\ell_{\mathrm{p}} = 1$ without loss of generality.

The challenge now is to use commonly known information, such as $\widehat{\mathbf{g}}[t^{'}]\ \forall 0 \leq t^{'} \leq t-1$,[1] to find the best predicted tangent vector $\mathbf{e}_{\mathrm{p}}^{\mathrm{opt}}[t]$, i.e., the one that maximizes $\mathbb{E}\left[\left|\widetilde{\mathbf{g}}[t]^{*}\mathbf{g}[t]\right|^{2}\right]$ and makes $\widetilde{\mathbf{g}}[t]$ the *best estimate* of $\mathbf{g}[t]$. Thus, the prediction problem can be summarized as solving

$$\begin{aligned}\mathbf{e}_{\mathrm{p}}^{\mathrm{opt}}[t] &= \underset{\mathbf{e}_{\mathrm{p}}[t]}{\arg\max}\ \mathbb{E}\left[\left|\widetilde{\mathbf{g}}[t]^{*}\mathbf{g}[t]\right|^{2}\right] \\ &\overset{(a)}{=} \underset{\mathbf{e}_{\mathrm{p}}[t]}{\arg\max}\ \mathbb{E}\left[\left|\left(\cos\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\widehat{\mathbf{g}}[t-1] + \sin\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\frac{\mathbf{e}_{\mathrm{p}}[t]}{\|\mathbf{e}_{\mathrm{p}}[t]\|}\right)^{*}\mathbf{h}[t]\right|^{2}\right],\end{aligned} \tag{3.16}$$

where (a) is by replacing $\widetilde{\mathbf{g}}[t]$ by its definition in (3.15) and by noticing that maximizing $\mathbb{E}\left[\left|\widetilde{\mathbf{g}}[t]^{*}\mathbf{g}[t]\right|^{2}\right]$ is equivalent to maximizing $\mathbb{E}\left[\left|\widetilde{\mathbf{g}}[t]^{*}\mathbf{h}[t]\right|^{2}\right]$. In what follows, I show that for a class of temporally correlated channels, such a geodesic predictor yields no performance enhancement over the differential feedback strategy presented. Therefore, to ensure that prediction does in fact improve performance, more sophisticated predictors that better exploit the

[1] In fact [34, 52] make the simplifying assumption that $\widehat{\mathbf{g}}[t^{'}]\ \forall 0 \leq t^{'} \leq t-1$ is available during predictor design. I relax this assumption, by only allowing access to $\widehat{\mathbf{g}}[t^{'}]\ \forall 0 \leq t^{'} \leq t-1$, at the cost of adding a milder assumption on the quantization error in $\widehat{\mathbf{g}}[t]$. Results hold, and derivations are simplified, if I revert to the assumptions of [34, 52].

channel's Doppler spectrum must be derived. The development of such predictors is a promising topic for future work.

3.5.2.2 Performance Comparison & Prediction Gains

For the analysis in this section, I make the following assumptions on the channel's temporal structure.

Assumption 5. *I assume that the channels* $\mathbf{h}[t]$ *are temporally correlated according to a first order autoregressive model [4], and are thus given by* $\mathbf{h}[t] = \eta_{\mathrm{f}}\mathbf{h}[t-1] + \sqrt{1-\eta_{\mathrm{f}}{}^2}\mathbf{z}[t]$. *For simplicity, I further assume that* $\mathbf{h}[t] \sim \mathcal{CN}(0, \frac{1}{L}\mathbf{I}_L)$ *which in turn implies that* $\mathbf{z}[t] \sim \mathcal{CN}(0, \frac{1}{L}\mathbf{I}_L)$.

While I present the analysis for channels satisfying Assumption 5, similar results can be shown for other temporally correlated channel models in which changes in the channel are zero-mean and independent of the previous channel realizations.

Assumption 6. *I assume that the quantization error introduced by the quantization algorithm is isotropically distributed, i.e., the direction of* $\mathbf{g}[t] - \widehat{\mathbf{g}}[t]$ *is uniformly distributed, and that the error direction and magnitude are independent.*

For a channel following a first order autoregressive model, the predic-

tor's objective function, $\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right]$, can be written as

$$\begin{aligned}\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right] &= \mathbb{E}\left[\left|\left(\cos\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\widehat{\mathbf{g}}[t-1] + \sin\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\frac{\mathbf{e}_{\mathrm{p}}[t]}{\|\mathbf{e}_{\mathrm{p}}[t]\|}\right)^*\right.\right. \\ &\qquad \left.\left.\times\left(\eta_{\mathrm{f}}\mathbf{h}[t-1] + \sqrt{1-\eta_{\mathrm{f}}{}^2}\mathbf{z}[t]\right)\right|^2\right] \\ &= \mathbb{E}\left[\left|\eta_{\mathrm{f}}\left(\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]\boldsymbol{\beta}\right)^*\mathbf{h}[t-1] + \bar{\eta}_{\mathrm{f}}\boldsymbol{\beta}^*\mathbf{x}[t]\right|^2\right], \qquad (3.17)\end{aligned}$$

where I define $\boldsymbol{\beta} = \left[\cos\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right),\ \sin\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\right]^*$, $\bar{\eta}_{\mathrm{f}} = \sqrt{1-\eta_{\mathrm{f}}{}^2}$ and $\mathbf{x}[t] = \left[\widehat{\mathbf{g}}[t-1]^*,\ \mathbf{e}_{\mathrm{p}}[t]^*\right]^*\mathbf{z}[t]$ to simplify notation. Since $\mathbf{z}[t] \sim \mathcal{CN}(0, \frac{1}{L}\mathbf{I}_L)$ by Assumption 5, and since the matrix $\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]$ is unitary and independent of $\mathbf{z}[t]$ by definition, the vector $\mathbf{x}[t]$ has distribution $\mathcal{CN}(0, \frac{1}{L}\mathbf{I}_2)$ by the unitary invariance of the Gaussian distribution [102]. Expanding the predictor's objective function, I get

$$\begin{aligned}\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right] =& \mathbb{E}\left[\eta_{\mathrm{f}}{}^2\mathbf{h}[t-1]^*\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]\boldsymbol{\beta}\boldsymbol{\beta}^*\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]^*\mathbf{h}[t-1]\right. \\ &\left.+2\mathrm{Re}\{\eta_{\mathrm{f}}\bar{\eta}_{\mathrm{f}}\mathbf{h}[t-1]^*[\widehat{\mathbf{g}}[t-1],\mathbf{e}_{\mathrm{p}}[t]]\,\boldsymbol{\beta}\boldsymbol{\beta}^*\mathbf{x}[t]\} + \bar{\eta}_{\mathrm{f}}^2\mathbf{x}[t]^*\boldsymbol{\beta}\boldsymbol{\beta}^*\mathbf{x}[t]\right] \\ \overset{(a)}{=}& \mathbb{E}\left[\eta_{\mathrm{f}}{}^2\mathbf{h}[t-1]^*\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]\boldsymbol{\beta}\boldsymbol{\beta}^*\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]^*\mathbf{h}[t-1]\right] \\ &+ \bar{\eta}_{\mathrm{f}}^2\frac{1}{L} \\ \overset{(b)}{=}& \eta_{\mathrm{f}}{}^2\cos^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[\left|\mathbf{h}[t-1]^*\widehat{\mathbf{g}}[t-1]\right|^2\right] \\ &+ \eta_{\mathrm{f}}{}^2\sin^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[\left|\mathbf{h}[t-1]^*\mathbf{e}_{\mathrm{p}}[t]\right|^2\right] \\ &+ \mathbb{E}\left[2\mathrm{Re}\left\{\eta_{\mathrm{f}}{}^2\cos\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\sin\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\right.\right. \\ &\left.\left.\times\mathbf{h}[t-1]^*\widehat{\mathbf{g}}[t-1]\mathbf{e}_{\mathrm{p}}[t]^*\mathbf{h}[t-1]\right\}\right] + \bar{\eta}_{\mathrm{f}}^2\frac{1}{L},\end{aligned}$$

where (a) follows because $\mathbf{x}[t]$ is a zero-mean vector and independent of $\mathbf{h}[t-1]^*\left[\widehat{\mathbf{g}}[t-1],\ \mathbf{e}_{\mathrm{p}}[t]\right]\boldsymbol{\beta}\boldsymbol{\beta}^*$ and $\mathbb{E}\left[\mathbf{x}[t]^*\boldsymbol{\beta}\boldsymbol{\beta}^*\mathbf{x}[t]\right] = \frac{1}{L}$. The expectation of the third

term in (b) is zero as a result of Assumption 6 and the fact that $\mathbf{e}_{\mathrm{p}}[t]$ and $\widehat{\mathbf{g}}[t-1]$ are instantaneously orthogonal. Moreover, leveraging the independence of $\|\mathbf{h}[t-1]\|$ and $\mathbf{h}[t-1]/\|\mathbf{h}[t-1]\|$, and that $\mathbb{E}\left[\|\mathbf{h}[t-1]\|^2\right]=1$, the objective function simplifies to

$$\begin{aligned}\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right] =&{\eta_{\mathrm{f}}}^2\cos^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[|\mathbf{g}[t-1]^*\widehat{\mathbf{g}}[t-1]|^2\right]\\ &+{\eta_{\mathrm{f}}}^2\sin^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[|\mathbf{g}[t-1]^*\mathbf{e}_{\mathrm{p}}[t]|^2\right]+\bar{\eta}_{\mathrm{f}}^2\frac{1}{L}.\end{aligned}$$

Letting $\widehat{\rho}=\mathbf{g}[t-1]^*\widehat{\mathbf{g}}[t-1]$, the channel $\mathbf{g}[t-1]$ can be written as $\widehat{\rho}\widehat{\mathbf{g}}[t-1]+\sqrt{1-|\widehat{\rho}|^2}\widetilde{\mathbf{e}}[t-1]$ where $\widetilde{\mathbf{e}}[t-1]$ is a unit norm tangent vector representing the quantization error at time $t-1$. By Assumption 6, $\widetilde{\mathbf{e}}[t-1]$ is uniformly distributed in the tangent space and independent of $|\widehat{\rho}|$, which results in

$$\begin{aligned}\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right] =&{\eta_{\mathrm{f}}}^2\cos^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[|\widehat{\rho}|^2\right]\\ &+{\eta_{\mathrm{f}}}^2\sin^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\left(1-\mathbb{E}\left[|\widehat{\rho}|^2\right]\right)\mathbb{E}\left[|\widetilde{\mathbf{e}}[t-1]^*\mathbf{e}_{\mathrm{p}}[t]|^2\right]+\bar{\eta}_{\mathrm{f}}^2\frac{1}{L}\\ \overset{(a)}{=}&{\eta_{\mathrm{f}}}^2\cos^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\mathbb{E}\left[|\widehat{\rho}|^2\right]\\ &+{\eta_{\mathrm{f}}}^2\sin^2\left(\|\mathbf{e}_{\mathrm{p}}[t]\|\right)\left(1-\mathbb{E}\left[|\widehat{\rho}|^2\right]\right)\frac{1}{L-1}+\bar{\eta}_{\mathrm{f}}^2\frac{1}{L},\end{aligned}\tag{3.18}$$

where (a) is since the isotropic distribution of $\widetilde{\mathbf{e}}[t-1]$ implies that the term $|\widetilde{\mathbf{e}}[t-1]^*\mathbf{e}_{\mathrm{p}}[t]|^2$ is beta distributed with mean $\mathbb{E}\left[|\widetilde{\mathbf{e}}^*\mathbf{e}_{\mathrm{p}}[t]|^2\right]=1/(L-1)$. Therefore the predictor's objective function is independent of the predicted direction of $\mathbf{e}_{\mathrm{p}}[t]$. Further, $\mathbb{E}\left[|\widehat{\rho}|^2\right]$ is practically such that $\mathbb{E}\left[|\widehat{\rho}|^2\right]>(1-\mathbb{E}\left[|\widehat{\rho}|^2\right])$, i.e. the quantizer's output is expected to be a more accurate estimate of the channel than a random vector. Thus $\mathbb{E}\left[|\widehat{\rho}|^2\right]>(1-\mathbb{E}\left[|\widehat{\rho}|^2\right])/(L-1)$ and $\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right]$ is maximized in (3.18) by letting $\|\mathbf{e}_{\mathrm{p}}[t]\|=0$ which achieves

$$\mathbb{E}\left[|\widetilde{\mathbf{g}}[t]^*\mathbf{h}[t]|^2\right]={\eta_{\mathrm{f}}}^2\mathbb{E}\left[|\widehat{\rho}|^2\right]+\bar{\eta}_{\mathrm{f}}^2\frac{1}{L}.\tag{3.19}$$

As a result, the predictor's objective function is maximized by selecting $\|\mathbf{e}_{\mathrm{p}}[t]\| = 0$ and thus $\widetilde{\mathbf{g}}[t] = \widehat{\mathbf{g}}[t-1]$. This coincides with the differential feedback strategy presented, thus the performance of the proposed framework can not be further improved by such a predictor for the channel model considered.

The derivation in this section, however, does not preclude potential benefits from more sophisticated prediction functions, especially when the band-limited nature of the channel's Doppler spectrum is exploited. While a rich body of literature on channel prediction does exist [39, 44, 98], work on predicting processes that evolve on the Grassmann manifold is limited to [34, 52] which use predictors and channel models similar to the ones considered in this section. The design of more sophisticated Grassmannian predictors that better exploit channel structure is left for future work.

3.6 Design Considerations & Codebook Construction

In this section, I discuss the proper initialization of the proposed algorithm and construct adaptive quantization codebooks for the tangent magnitude and direction.

3.6.1 Initialization

The proper operation of the Grassmannian differential feedback algorithm is ensured by the fact that, at each iteration, both transmitter and receiver can calculate a common quantized channel based on the quantized tangent vector and the previous quantized channel. For the output of the al-

gorithm to be the same at the transmitter and receiver, however, a common initial vector, $\widehat{\mathbf{g}}[0]$, is required. Reinitialization can also be used to recover from feedback bit errors. This initial vector can, for example, be an estimate of $\widehat{\mathbf{g}}[0]$ that is fed back from receiver to transmitter in the first iteration along with $\widehat{\mathbf{e}}[0]$. Alternatively, it can be based on a memoryless quantization of the channel using methods such as [53, 65]. The additional overhead incurred is amortized when there are many subsequent feedback stages. Another option is to initialize the estimate $\widehat{\mathbf{g}}[0]$ to a common random vector from which the algorithm will be able to recover. In this chapter, I use memoryless quantization with a random vector codebook at $t = 1$.

3.6.2 Tangent Magnitude Quantization

The tangent vector calculated in (3.14) can be decomposed into a tangent magnitude and a unit norm tangent direction. I propose to quantize the tangent magnitude and direction separately, which ensures that search complexity remains manageable. The problem of quantizing the tangent magnitude is that of quantizing a non-negative scalar. I propose to quantize the tangent magnitude using a Euclidean distance distortion metric

$$\widehat{e}_{\text{mag}}[t] = \underset{e_i \in \mathcal{C}_{\text{mag}}}{\arg\min} \; |\|\mathbf{e}[t]\| - e_i| \,, \tag{3.20}$$

where $\mathcal{C}_{\text{mag}}$ is the magnitude quantization codebook. The index of the minimizer is then sent to the transmitter via a delay and error-free link that requires $N_{\text{mag}} = \log(|\mathcal{C}_{\text{mag}}|)$ bits. A locally optimal quantization codebook can

be found algorithmically using Lloyd's algorithm [25] given either the probability density function (pdf) of the tangent magnitude, or a training set of magnitudes which provide an empirical probability mass function. Finding the exact pdf of the tangent magnitude, however, has been intractable analytically thus far. Generating a good representative training set as input to the Lloyd algorithm is also difficult. The difficulty in both these approaches is that the magnitude pdf depends on the quantization process itself. That is to say that quantizing the tangent direction and magnitude itself could change their distribution in the next step. What is needed is therefore finding a pdf that is a fixed point of the quantization process, which has proven to be difficult. One solution which is adopted in [34] is to uniformly quantize a range of magnitudes, $\|\mathbf{e}[t]\| \in [0, 1]$. Unfortunately, [34] has shown that quantization error in the magnitude creates an error floor beyond which CSI can not be improved in highly correlated channels where predictive feedback is most useful. In the limit of perfectly static channels, such magnitude quantization prevents feedback from converging to perfect CSI.

I propose to adapt the quantization range, $[\alpha\widehat{e}_{\min}[t],\ \beta\widehat{e}_{\max}[t]]$, to the dynamics of the system and to uniformly quantize the magnitude in that range. To do so, the algorithm maintains a *sliding window* of the quantization range of interest. This adaptive process maintains a running average of the tangent vector magnitude and uses it to update the lower and upper limits of the quantization range. The quantization range, or window, is calculated using quantities that are commonly available at both transmitter and receiver. As

a result, both nodes can independently keep track of the current quantization range.

The moving average of quantized tangent magnitudes is given by:

$$\widehat{e}_{\text{avg}}[t+1] = \left(1 - \frac{1}{\tau}\right) \widehat{e}_{\text{avg}}[t] + \tau \widehat{e}_{\text{mag}}[t], \tag{3.21}$$

where τ is the smoothing factor that can be adjusted by the system. At each iteration either limit of the quantization range is then updated as follows:

$$\begin{aligned}
\widehat{e}_{\min}[t+1] &= \begin{cases} \left(1 - \frac{1}{\tau}\right) \widehat{e}_{\min}[t] + \frac{1}{\tau} \widehat{e}_{\text{mag}}[t] & \widehat{e}_{\text{mag}}[t] \leq \widehat{e}_{\text{avg}}[t] \\ \widehat{e}_{\min}[t] & \widehat{e}_{\text{mag}}[t] \geq \widehat{e}_{\text{avg}}[t] \end{cases} \\
\widehat{e}_{\max}[t+1] &= \begin{cases} \widehat{e}_{\max}[t] & \widehat{e}_{\text{mag}}[t] \leq \widehat{e}_{\text{avg}}[t] \\ \left(1 - \frac{1}{\tau}\right) \widehat{e}_{\max}[t] + \frac{1}{\tau} \widehat{e}_{\text{mag}}[t] & \widehat{e}_{\text{mag}}[t] \geq \widehat{e}_{\text{avg}}[t]. \end{cases}
\end{aligned}$$

This allows the feedback algorithm to accurately track the statistics of the tangent magnitude and quantize the current range of interest with higher resolution. In static channels, this allows the system to converge to perfect CSI removing the error floor caused by constant magnitude codebooks.

The fixed scalars α and β must now be such that $\alpha < 1 < \beta$ to enable the flexible adjustment of the window $[\widehat{e}_{\min}[t],\ \widehat{e}_{\max}[t]]$. To see this, note that to let $\widehat{e}_{\min}[t]$ decrease, it is necessary to be able to quantize values less than $\widehat{e}_{\min}[t]$. Similarly, allowing $\widehat{e}_{\max}[t]$ to increase requires quantizing values greater than it. In simulation I use $\tau = 5$, $\alpha = 1/2$ and $\beta = 2$. While the chosen parameter values yield good results as shown in Section 3.8, system performance can be further improved by optimizing over the values of τ, α, and β.

3.6.3 Tangent Direction Quantization

The problem of quantizing the tangent direction vector is that of quantizing a unit norm vector which lies in the tangent space orthogonal to the base vector $\widehat{\mathbf{g}}[t-1]$, that is $\mathbf{e}[t]^*\widehat{\mathbf{g}}[t-1] = 0$. Grassmannian codebooks often used for limited feedback [53, 65] are not suitable for tangent vector quantization since they do not enforce the structural constraint that requires the tangent direction codewords to be orthogonal to the base vector $\widehat{\mathbf{g}}[t-1]$. If the quantized tangent vector is not orthogonal to the base vector $\widehat{\mathbf{g}}[t-1]$, the geodesic path is undefined and the output of $G(\widehat{\mathbf{g}}[t-1], \widehat{\mathbf{e}}[t], \ell)$ does not lie on the manifold. Finally, note that the tangent space changes for each base vector $\widehat{\mathbf{g}}[t-1]$ and thus the codebook must be adapted [29], i.e., fixed codebooks should not be used.

To respect the varying tangent space geometry and orthogonality constraints, I propose a canonical generating codebook that is adapted at each iteration; only the canonical codebook is actually stored. This canonical codebook quantizes the tangent space at a special $L \times 1$ base vector. The canonical codebook is rotated at each channel realization to match the geometry of the current tangent space.

Definition 7. *The N_{dir}-bit canonical tangent codebook is a codebook of $2^{N_{\text{dir}}}$ unit norm vectors orthogonal to the special base vector $x_b \triangleq [1, 0, \ldots, 0]$. Because of the structure of x_b, the generator codebook, $\mathcal{C}_{\text{dir}}(\mathbf{x}_b)$, has elements, $\mathbf{v}_\ell \in \mathcal{C}_{\text{dir}}(\mathbf{x}_b)$, of the form $\mathbf{v}_\ell = [0,\ \tilde{\mathbf{v}}_\ell^*]^*$ where $\tilde{\mathbf{v}}_\ell$ is an $(L-1) \times 1$ unit norm vector.*

The canonical codebook can be constructed by appending a leading 0 to a codebook of $(L-1)\times 1$ vectors. The canonical codebook, which can be used to quantize the tangent space at $\mathbf{x}_b$, can now be rotated at each iteration (or channel realization) to quantize the tangent space at a new vector $\mathbf{x}$. I define this rotation as follows [29, 52].

Definition 8. *Let* $\mathbf{U}(\mathbf{x}) : \mathbb{C}^{L\times 1} \mapsto \mathcal{U}^{L\times L}$ *be the function that determines the unitary rotation matrix that rotates* $\mathbf{x}_b$ *to* $\mathbf{x}$*, i.e.* $\mathbf{U}(\mathbf{x})\mathbf{x}_b = \mathbf{x}$.

Perhaps the easiest way to find such a rotation is to consider the Householder matrix. Let $\mathbf{H}(\mathbf{x}, \mathbf{x}_b) = \mathbf{I}_L - \mathbf{u}\mathbf{u}^*/\mathbf{u}^*\mathbf{x}_b$ where $\mathbf{u} = \mathbf{x}_b - \mathbf{x}$ [13]. Note that $\mathbf{H}(\mathbf{x}, \mathbf{x}_b)$ is a unitary matrix and that if we let $\mathbf{U}(\mathbf{x}) = \mathbf{H}(\mathbf{x}, \mathbf{x}_b)$ then $\mathbf{U}(\mathbf{x})\mathbf{x}_b = \mathbf{x}$ as required. Using the rotation operation it is possible to rotate the canonical codebook to the direction of any particular base point and generate tangent direction codebooks for an arbitrary channel. For a base point, $\widehat{\mathbf{g}}[t-1]$, the quantizer chooses $\widehat{\mathbf{e}}_{\text{dir}}[t]$ from the rotated codebook

$$\widehat{\mathbf{e}}_{\text{dir}}[t] \in \mathcal{C}_{\text{dir}}(\widehat{\mathbf{g}}[t-1]) = \{\mathbf{U}(\widehat{\mathbf{g}}[t-1])\mathbf{v}_1, \mathbf{U}(\widehat{\mathbf{g}}[t-1])\mathbf{v}_2, \ldots, \mathbf{U}(\widehat{\mathbf{g}}[t-1])\mathbf{v}_{2^{N_{\text{dir}}}}\}.$$

It can be verified that all vectors $\mathbf{U}(\widehat{\mathbf{g}}[t-1])\mathbf{v}_\ell$ lie in the tangent plane at $\widehat{\mathbf{g}}[t-1]$ since $\widehat{\mathbf{g}}[t-1]^*\mathbf{U}(\widehat{\mathbf{g}}[t-1])\mathbf{v}_\ell = \mathbf{x}_b^*\mathbf{v}_\ell = 0,\ \forall \ell$.

The proposed codebook design allows us to rotate a canonical codebook to perfectly match the tangent plane at each iteration. This ensures that the output of the Grassmannian differential feedback algorithm remains on the manifold. Moreover, if changes in the channel are isotropic in the L

dimensional space, then so are the normalized tangent vectors, implying good performance with canonical codebooks that uniformly quantize the tangent space at $\mathbf{x}_b$. In Section 3.8, good performance is achieved even with random vector canonical codebooks. In general, for a given correlation matrix $\mathbf{R}_{\mathbf{h}_{k,\ell}}$, good canonical codebooks can be generated using Lloyd's algorithm, and different codebooks can be stored for environments with different correlation properties.

To formalize the tangent direction quantization, recall that the estimated channel in the next iteration is $G(\widehat{\mathbf{g}}[t-1], \widehat{\mathbf{e}}[t], 1)$. Given that the loss in sum-rate is related to the chordal distance between the channel and its estimate, the quantized tangent direction is computed as

$$\begin{aligned} \widehat{\mathbf{e}}_{\text{dir}}[t][t] &= \underset{\mathbf{x}_i \in \mathcal{C}_{\text{dir}}(\widehat{\mathbf{g}}[t-1])}{\arg\min} \; d(G(\widehat{\mathbf{g}}[t-1], \widehat{e}_{\text{mag}}[t]\mathbf{x}_i, 1), \mathbf{g}[t]) \\ &= \underset{\mathbf{x}_i \in \mathcal{C}_{\text{dir}}(\widehat{\mathbf{g}}[t-1])}{\arg\max} \; |G(\widehat{\mathbf{g}}[t-1], \widehat{e}_{\text{mag}}[t]\mathbf{x}_i, 1)^* \mathbf{g}[t]|^2, \end{aligned} \tag{3.22}$$

where the tangent magnitude, $\widehat{e}_{\text{mag}}[t]$, is given by the output of the magnitude quantization step. Note that although (3.22) intuitively includes a rotation of the full canonical codebook, computational complexity can be reduced by noticing that an equivalent rotation can be applied to the channels $\widehat{\mathbf{g}}[t-1]$ and $\mathbf{g}[t]$ instead, and the codebook need not be actually rotated. The details of this simplification are omitted but can be seen by expanding the objective function $d(G(\widehat{\mathbf{g}}[t-1], \widehat{e}_{\text{mag}}[t]\mathbf{x}_i, 1)$.

3.7 Analytical Performance Characterization

A complete performance analysis of differential or predictive quantization is difficult due to its recursive nature and, in this case, due to the CSI's Grassmannian structure [41]. Relevant performance analysis when the quantized information evolves on the Grassmannian manifold is limited to incomplete performance characterizations[34, 52]. The analysis of the Householder scheme in [52] is limited to characterizing the performance of an open loop predictor, similar to Section 3.5.2. The resulting lower bound on performance in [52] neglects the quantization process and is not a function of the number of feedback bits. The performance analysis in [34] presents partial distortion bounds that are independent of the channel's temporal correlation and Doppler spread.

In this section, I derive an approximation for the distortion achieved by the proposed quantization strategy. Unlike the analysis in [34, 52], my characterization accounts for both the number of feedback bits and the channel's temporal correlation, thus capturing most of the fundamental system parameters. Given the distortion expression, I leverage my analysis in Chapter 2 and the discussion in Section 3.4.2 to characterize IA's mean loss in sum-rate when the proposed feedback strategy is used.

I derive an approximation for the quantizer's average accuracy defined as $\mathbb{E}\left[|\mathbf{g}[t]^{*}\widehat{\mathbf{g}}[t]|^{2}\right]$, under three mild assumptions. First, I specialize the analysis to the case of channels with relatively low Doppler spread where differential feedback techniques yield the most gain [41, 52]. In the low Doppler setting, I

can make use of a small angle approximation [2, 34]. Second, I neglect the error introduced by the tangent magnitude quantizer of Section 3.6. Effectively I make the following assumption.

Assumption 9. *I assume perfect magnitude quantization, i.e.* $\widehat{e}_{\text{mag}}[t] = \|\mathbf{e}[t]\|, \forall t.$

This is a reasonable assumption for my analysis as the numerical results in Section 3.8.1 indicate that the magnitude quantizer adopted achieves very low quantization error with as little as one quantization bit and approaches perfect quantization at all Doppler. Finally, the analysis re-invokes Assumption 5 for simplicity of exposition.

I start by expanding $\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}}) = \mathbb{E}\left[|\mathbf{g}[t]^* \widehat{\mathbf{g}}[t]|^2\right]$ as

$$\begin{aligned}\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}}) = \mathbb{E}\Bigg[\Bigg|\left(\cos(\|\mathbf{e}[t]\|)\widehat{\mathbf{g}}[t-1] + \sin(\|\mathbf{e}[t]\|)\frac{\mathbf{e}[t]}{\|\mathbf{e}[t]\|}\right)^* \\ \times \left(\cos(\widehat{e}_{\text{mag}}[t])\widehat{\mathbf{g}}[t-1] + \sin(\widehat{e}_{\text{mag}}[t])\widehat{\mathbf{e}}_{\text{dir}}[t][t]\right)\Bigg|^2\Bigg] \\ = \mathbb{E}\left[\left|\cos(\|\mathbf{e}[t]\|)\cos(\widehat{e}_{\text{mag}}[t]) + \sin(\|\mathbf{e}[t]\|)\sin(\widehat{e}_{\text{mag}}[t])\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\text{dir}}[t][t]}{\|\mathbf{e}[t]\|}\right|^2\right].\end{aligned}$$

By invoking Assumption 9, $\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}})$ simplifies to

$$\begin{aligned}\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}}) = \mathbb{E}\Bigg[\cos^4(\|\mathbf{e}[t]\|) + 2\cos^2(\|\mathbf{e}[t]\|)\sin^2(\|\mathbf{e}[t]\|)\text{Re}\left\{\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\text{dir}}[t][t]}{\|\mathbf{e}[t]\|}\right\} \\ + \sin^4(\|\mathbf{e}[t]\|)\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\text{dir}}[t][t]}{\|\mathbf{e}[t]\|}\right|^2\Bigg].\end{aligned} \tag{3.23}$$

Recall that at each new channel realization the quantized tangent direction $\widehat{\mathbf{e}}_{\text{dir}}[t]$ is selected to minimize the chordal distance between $\mathbf{g}[t]$ and $\widehat{\mathbf{g}}[t]$, equivalently maximizing $|\mathbf{g}[t]^*\widehat{\mathbf{g}}[t]|^2$. By invoking Assumption 9, maximizing the objective $|\mathbf{g}[t]^*\widehat{\mathbf{g}}[t]|^2$ becomes equivalent to maximizing the instantaneous value of the right hand side of (3.23). The distortion function in (3.23), however, is new; it is neither a Euclidean quantizer's distortion function nor is it a Grassmannian quantizer's phase-invariant distortion function. As a result, neither Euclidean nor Grassmannian quantization results can be used to find an expression for $\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}})$.

By analyzing a similar alternative quantization strategy, however, it is possible to gain insight into the behavior of (3.23) and derive an approximation for $\mathcal{D}(\eta_{\text{f}}, N_{\text{dir}})$. Instead of directly quantizing the tangent direction to maximize the instantaneous value of (3.23), I propose to split $\widehat{\mathbf{e}}_{\text{dir}}[t]$ into two components, $\widehat{\mathbf{e}}_{\text{dir}}[t] = e^{-j\widehat{\theta}_e[t]}\widehat{\mathbf{e}}_{\mathcal{G}}[t]$. The variables $\widehat{\theta}_e[t]$ and $\widehat{\mathbf{e}}_{\mathcal{G}}[t]$ are then quantized separately using N_θ and $N_{\mathcal{G}}$ bits. To keep the amount of feedback constant, I enforce $N_\theta + N_{\mathcal{G}} = N_{\text{dir}}$. Such separation, naturally incurs an increase in distortion when compared to the original "joint" quantization strategy. Therefore, after deriving a distortion expression for the separate quantization strategy, I pick N_θ and $N_{\mathcal{G}}$ to optimize its performance which leads to a more accurate characterization of the original "joint" quantization strategy.

Inserting the proposed tangent decomposition into (3.23) yields

$$\begin{aligned}\mathcal{D}(\eta_{\mathrm{f}}, N_{\mathrm{dir}}) =& \mathbb{E}\left[\cos^4(\|\mathbf{e}[t]\|) + 2\cos^2(\|\mathbf{e}[t]\|)\sin^2(\|\mathbf{e}[t]\|)\,\mathrm{Re}\left\{\frac{e^{-j\widehat{\theta}_e[t]}\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right\}\right.\\ &\left. + \sin^4(\|\mathbf{e}[t]\|)\left|\frac{e^{-j\widehat{\theta}_e[t]}\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|^2\right]\\ \approx& \mathbb{E}\left[\cos^4(\|\mathbf{e}[t]\|)\right.\\ &+2\cos^2(\|\mathbf{e}[t]\|)\sin^2(\|\mathbf{e}[t]\|)\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|\cos\left(\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t]\right)\\ &\left. + \sin^4(\|\mathbf{e}[t]\|)\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|^2\right],\end{aligned} \tag{3.24}$$

where $\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]$ is the phase of the inner product $\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]$ and as a result

$$\begin{aligned}\mathrm{Re}\left\{e^{-j\widehat{\theta}_e[t]}\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]\right\} &= \mathrm{Re}\left\{e^{j(\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t])}\left|\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]\right|\right\}\\ &= \left|\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]\right|\cos\left(\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t]\right).\end{aligned}$$

Due to the decomposition of $\widehat{\mathbf{e}}_{\mathrm{dir}}[t]$ the distortion function decouples, i.e. the quantization of $\widehat{\mathbf{e}}_{\mathrm{dir}}[t]$ decouples into a scalar quantization step to minimize $|\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t]|$ and a Grassmannian quantization step to maximize $|\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]|$. Thanks to the decomposition, I can now use known results about scalar quantizers and Grassmannian quantizers to analyze the performance of the decomposed strategy.

Recalling Assumption 5 and specializing the analysis to channels with relatively low Doppler as stated earlier, I apply a small angle approximation which can be used to see that $\mathbf{e}[t] \approx \sqrt{1-\eta_{\mathrm{f}}{}^2}\mathbf{z}[t]$ [2, 34]. Therefore by Assumption 5 and the small angle approximation, one can see that the

tangent direction $\mathbf{e}[t]/\|\mathbf{e}[t]\|$ is isotropic and independent of $\|\mathbf{e}[t]\|$ and that $\sin^2(\|\mathbf{e}[t]\|) = \|\mathbf{e}[t]\|^2 = \|\sqrt{1-{\eta_{\mathrm{f}}}^2}\mathbf{z}[t]\|^2$. Using this fact, along with the approximation $\mathbb{E}[x^2] \approx \mathbb{E}[x]^2$, which is accurate when x has relatively low variance, gives

$$\mathcal{D}(\eta_{\mathrm{f}}, N_{\mathrm{dir}}) \approx {\eta_{\mathrm{f}}}^4 + 2{\eta_{\mathrm{f}}}^2(1-{\eta_{\mathrm{f}}}^2)\mathbb{E}\left[\cos\left(\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t]\right)\right]\mathbb{E}\left[\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|\right] + (1-{\eta_{\mathrm{f}}}^2)^2\mathbb{E}\left[\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|^2\right]. \tag{3.25}$$

Now, using an N_θ-bit uniform codebook over $[0, 2\pi]$ for $\widehat{\theta}_e[t]$ yields

$$\mathbb{E}\left[\cos\left(\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t] - \widehat{\theta}_e[t]\right)\right] = \frac{\sin(\pi/2^{N_\theta})}{\pi/2^{N_\theta}}, \tag{3.26}$$

since the $\mathbf{e}[t]$ is isotropic which implies that $\angle\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]$ is uniform [53]. Further, using an $L-1$ dimensional Grassmannian codebook to construct the canonical codebook for $\widehat{\mathbf{e}}_{\mathcal{G}}[t]$ yields

$$\left|\frac{\mathbf{e}[t]^*\widehat{\mathbf{e}}_{\mathcal{G}}[t]}{\|\mathbf{e}[t]\|}\right|^2 < 1 - \frac{4}{2^{N_{\mathcal{G}}/(L-2)}}. \tag{3.27}$$

The proof of (3.27) follows from the derivations in [65] and is thus omitted. Inserting (3.26) and (3.27) into (3.25) yields the final result which I summarize in the following proposition.

Proposition 10. *Under Assumptions 5 and 9, the accuracy achieved by the proposed Grassmannian differential feedback strategy with a* $2^{N_{\mathrm{dir}}}$ *direction*

codebook is

$$\mathcal{D}(\eta_{\mathrm{f}}, N_{\mathrm{dir}}) \approx \max_{N_\theta,\ N_{\mathcal{G}}} {\eta_{\mathrm{f}}}^4 + 2{\eta_{\mathrm{f}}}^2(1-{\eta_{\mathrm{f}}}^2)\frac{\sin(\pi/2^{N_\theta})}{\pi/2^{N_\theta}}\sqrt{1-\frac{4}{2^{N_{\mathcal{G}}/(L-2)}}} + (1-{\eta_{\mathrm{f}}}^2)^2\left(1-\frac{4}{2^{N_{\mathcal{G}}/(L-2)}}\right)$$
$$\text{s.t.} \quad N_\theta + N_{\mathcal{G}} = N_{\mathrm{dir}}.$$

The values of N_θ and $N_{\mathcal{G}}$ can be easily determined by a search over N_{dir} possibilities.

Finally, note that the approximation in Proposition 10 can now shed light on the performance of IA with the proposed feedback strategy. For example, it can be used in an identical manner as in Chapter 2 to evaluate the mean loss in IA sum-rate. Since deriving IA's mean loss sum-rate parallels my derivation in Chapter 2, and since the main results of [7, 69] have been summarized via (3.10) and (3.11), I present the final result on sum-rate loss and refer the reader to [69] for more detail. When the differential feedback strategy, with N_{dir} bits for the tangent direction, is used to fulfill IA's CSI requirement in a channel with $\eta_{\mathrm{f}} = J_0(2\pi f_{\mathrm{D}} T_{\mathrm{s}})$, the mean loss in sum-rate is such that

$$\Delta\bar{R}_{\mathrm{sum}} \lesssim \sum_k \frac{d_k}{N_{\mathrm{sc}}} \log_2\left(1 + \frac{N_{\mathrm{sc}}P}{\sigma_{\mathrm{n}}^2}\sum_\ell \frac{\rho_{k,\ell}\,(d_\ell - \delta_{\ell,k})}{d_\ell}\,(1-\mathcal{D}(\eta_{\mathrm{f}}, N_{\mathrm{dir}}))\right), \tag{3.28}$$

where $\delta_{\ell,k}$ is the Kronecker delta function. As a result of (3.28), IA sum-rate with differential feedback is such that $\bar{R}_{\mathrm{sum}} \gtrsim R_{\mathrm{sum}}^{\mathrm{ideal}} - \Delta\bar{R}_{\mathrm{sum}}$, where $R_{\mathrm{sum}}^{\mathrm{ideal}}$ is the sum-rate achieved with perfect CSI. Note, however, that the characterization given in (3.28) and derived using the method in [69] is known to

be rather loose especially when CSI quality is poor, mainly due to the use of Jensen's inequality. Tighter, yet more involved, sum-rate characterizations can be derived using alternative methods as in [67].

3.8 Simulation Results

In this section, I present numerical performance results. First, I characterize the performance of Grassmannian differential feedback in relation to Doppler and codebook size, and explore the trade-off between codebook size and refresh rate. Second, I demonstrate the performance of IA when CSI is obtained via the Grassmannian differential feedback strategy.

3.8.1 Quality of CSI

First, I examine the quantization error resulting from using the proposed algorithm to quantize the channel's impulse response. I consider L tap channels which are temporally correlated according to the M-order autoregressive channel model defined in [4]. Unless otherwise stated, the simulations use $M = 200$ and initialize the time series as in [4]. Although the analysis of Section 3.5 specialized $M = 1$, adopting a more complex higher order model in simulation provides more realistic performance results. I restrict my attention to an uncorrelated uniform fading profile, that is $\mathbf{R_h} = \frac{1}{L}\mathbf{I}_L$.The normalized channel vectors, $\frac{\mathbf{h}_{k,\ell}[t]}{\|\mathbf{h}_{k,\ell}[t]\|}$ constitute the correlated Grassmannian time series to be quantized.

Fig. 3.3 shows the chordal distance between the actual and quantized

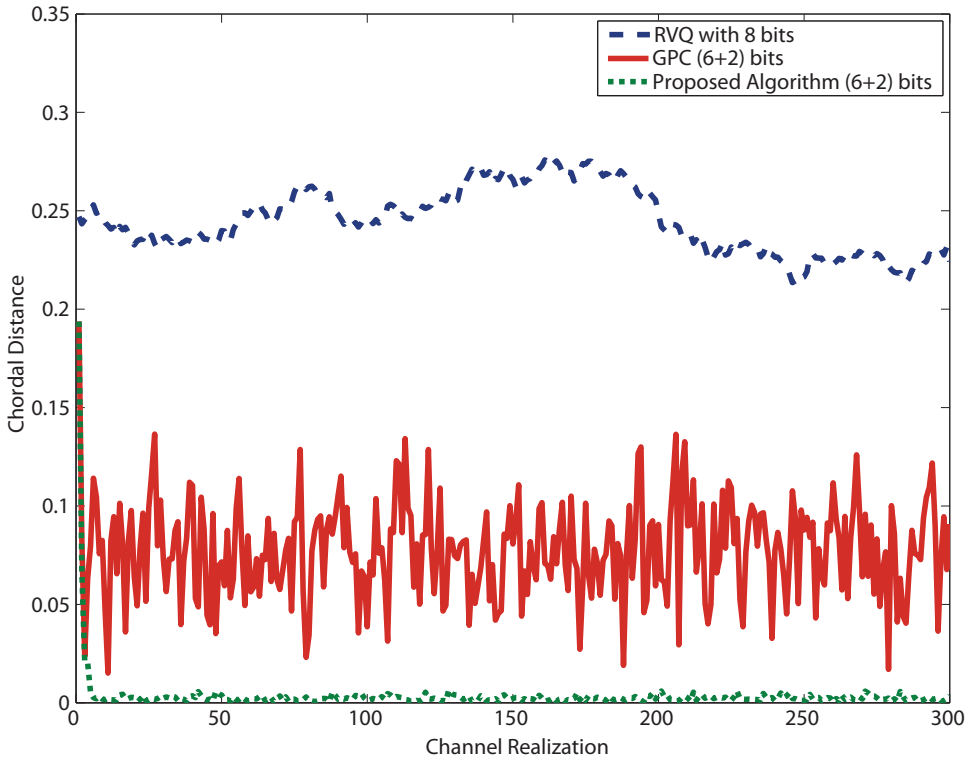

Figure 3.3: Chordal distance, $d(\widehat{\mathbf{g}}[t], \mathbf{g}[t])$, plotted over time for a slowly channel with normalized Doppler $f_{\mathrm{D}}T_{\mathrm{s}} = 0.003$. This shows the high quality CSI achieved by the proposed algorithm with 6 and 2 bits for channel direction and magnitude respectively, compared to [34] and memoryless quantization with a random codebook of 8 bits.

channel for a three tap correlated channel with $f_{\mathrm{D}}T_{\mathrm{s}} = 0.005$. Comparing Grassmannian differential feedback to memoryless random vector quantization, we see that the proposed algorithm achieves a much lower chordal distance as early as the third or fourth iteration. The initial transient phase in which the quantization error is higher than the steady state is due to initializing the algorithm with an all ones vector, i.e. $\widehat{\mathbf{g}}[0] = \frac{1}{\sqrt{3}}[1, 1, 1]^*$. The quick convergence to the steady state performance indicates that there is limited loss

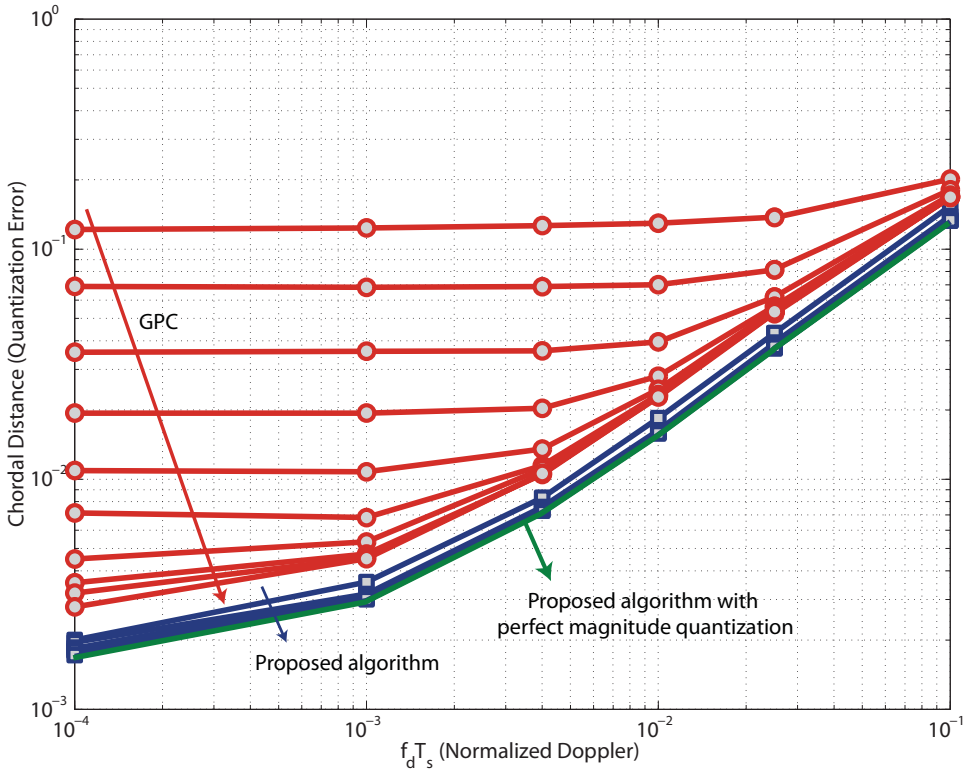

Figure 3.4: This shows the average chordal distance, or quantization error, versus the number of bits allocated for the tangent magnitude (1 to 10 bits), while keeping direction bits fixed. This shows the good performance of the proposed algorithm when compared with [34].

due to random initialization. The proposed algorithm also achieves a much lower steady state error compared to the GPC algorithm presented in [34] due to the adaptive codebooks presented. Examining the average quantization error achieved by both the proposed algorithm and the GPC algorithm of [34], we see that the proposed algorithm consistently outperforms. This can be seen in Figs. 3.4 and 3.5.

Fig. 3.4 plots the average error of both algorithms vs. Doppler spread

for a varying number of magnitude quantization bits, N_{mag}. The adaptive magnitude codebook, with even 1 bit of feedback, achieves 1.5dB lower distortion than the GPC algorithm with 10 bits of magnitude feedback. Adapting the codebook as presented in Section 3.6 allows accurate quantization and leaves very little to be gained from increasing the number of magnitude bits. Therefore, while [34] requires an increasing number of magnitude bits at low Doppler, which is the case of interest, the proposed algorithm works well with 1 bit for all Doppler spread. Moreover, if the total number of feedback bits is fixed, and the bits are allocated optimally across the direction and magnitude codebooks, it can be numerically shown that the optimal number of bits allocated to the magnitude is always 1.

Fig. 3.5 plots the average error for a varying number of direction quantization bits, N_{dir}, while keeping N_{mag} constant. Fig. 3.5 again shows that the proposed algorithm consistently outperforms existing feedback schemes. We can see that the proposed algorithm can achieve up to 13dB lower quantization error than previous schemes. This is due to the proposed magnitude and tangent codebooks which adapt to the geometry of each point on the manifold. Moreover, Fig. 3.5 shows that, for the proposed algorithm, increasing the codebook size from 3 to 4, or 4 to 5 bits offers the greatest improvements in quantization error, after which only marginal improvements are experienced. Interestingly, as opposed to further increases in codebook size, going from 3 to 5 bits does not introduce a constant improvement at all Doppler and instead changes the slope of the curve providing much larger gains in slow fading

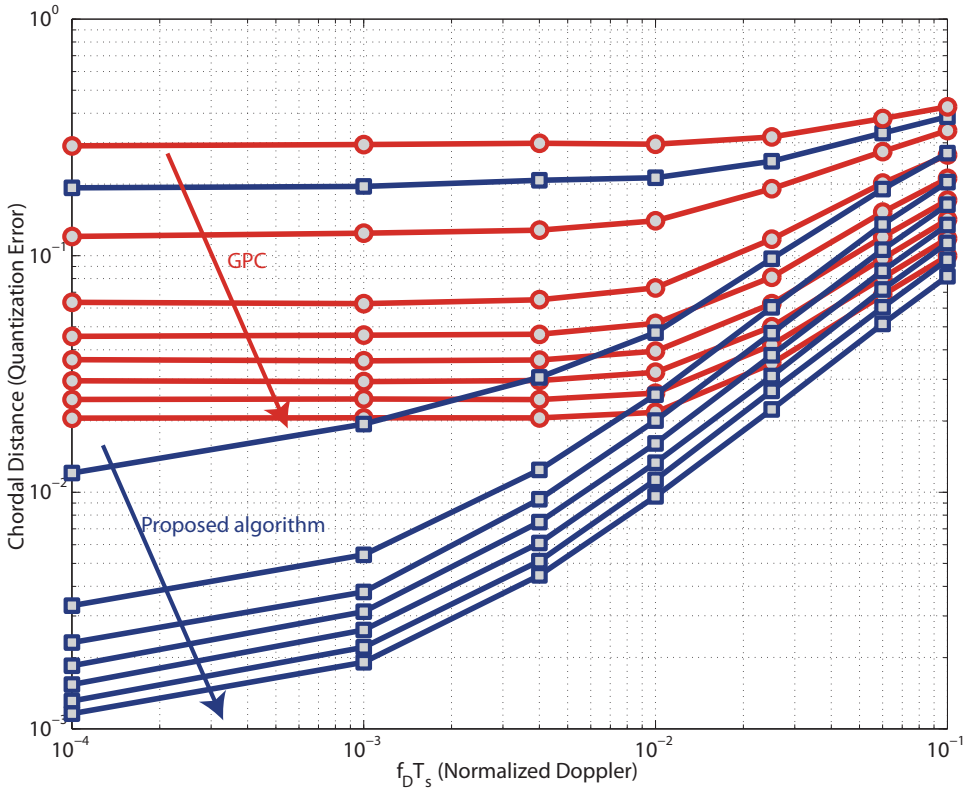

Figure 3.5: This shows the average chordal distance, or quantization error, versus the number of bits allocated for the tangent direction, while keeping magnitude bits fixed. This shows the good performance of the proposed algorithm when compared with [34].

Table 3.1: Minimum Direction Bits

Length of Channel	Suggested Direction Bits
2	3
3	5
4	7
5	9
6	11

channels.

Investigating this phase transition behavior further I notice that an

error floor exists at low Doppler for all tangent direction codebooks smaller than 5 bits. Moreover, this phase transition is independent of the number of bits allocated to the magnitude, i.e. quantizing the magnitude with infinite precision does not change this behavior. Examining channels with more taps, I notice that this behavior is recurring, and that the phase transition occurs according to Table 3.1. It is interesting to note that this phase transition behavior for an L-tap channel always occurs at $2L-1$ bits, which is the number of free variables in the tangent direction. A formal investigation of this is left for future work.

Fig. 3.6 plots $\mathbb{E}\left[|\mathbf{g}[t]^{*}\widehat{\mathbf{g}}[t]|^{2}\right]$ vs. $f_{\mathrm{D}}T_{\mathrm{s}}$ for a practical range of N_{dir} to evaluate the accuracy of the performance characterization derived in Section 3.7. Similarly to Section 3.7, I consider the performance of the differential algorithm for a first order auto regressive channel model. As expected, the performance characterization is most accurate for slow fading channels and is reasonably accurate in faster fading scenarios. The reason for this is that the small angle approximation used is naturally suited for relatively low Doppler spreads. Moreover, the performance characterization gradually becomes more accurate as N_{dir} increases. Recall that the result was derived using a "separate" quantization scheme in which the tangent direction was decomposed into a phase angle, and a phase invariant direction. The bit allocation between the 2 variables, N_{θ} and $N_{\mathcal{G}}$, was then optimized to better characterize the superior "joint quantization" approach. The penalty due to such separate quantization is large when a small number of bits is available as the "product" of the two

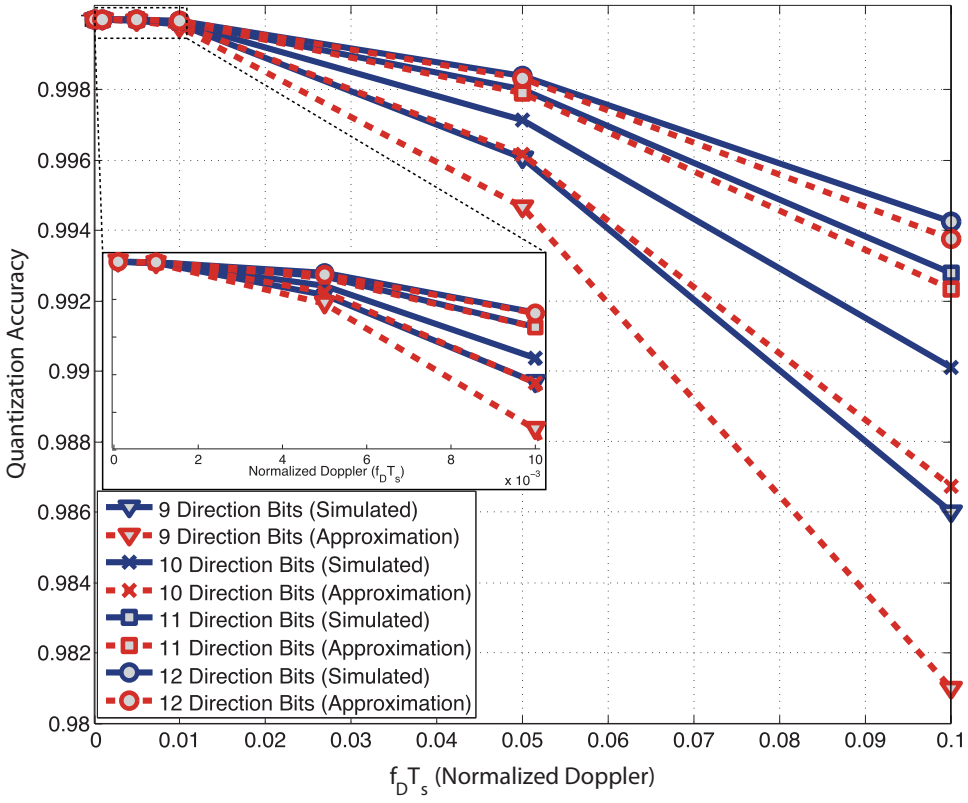

Figure 3.6: This figure compares the simulated performance of differential feedback with the performance characterization derived in Section 3.7. While the characterization is loose when feedback bits are low, due to the significant loss from separate quantization, the characterization becomes increasingly accurate as resolution increases.

codebooks does not efficiently fill the tangent space. As the number of bits increases, the penalty due to such separate quantization diminishes and the performance characterization becomes more accurate even at high Doppler. Such behavior is in line with previous results on separate vs. joint quantization [25].

Finally, I highlight an issue that has often been overlooked in CSI feed-

back, with a few exceptions such as the analysis in [32, 42] for single-user systems. Namely, the numerical analysis thus far has assumed a fixed normalized Doppler spread $f_{\mathrm{D}}T_{\mathrm{s}}$, and a fixed number of feedback bits. Both these quantities, however, can be controlled by the system while keeping the overall overhead of feedback, i.e. feedback bit rate, fixed. When such optimization is possible, feedback frequency and codebook size can be traded-off depending on channel conditions, resulting in lower distortion. Fig. 3.7 shows that increasing codebook size while keeping overall overhead constant can significantly deteriorate performance in fast fading channels. On the one hand, slow fading channels can support a large delay between high resolution feedback updates, since the channel does not change appreciably. On the other hand, decreasing the feedback time T_{s} to counter the high Doppler frequency f_{D} yields better performance even if each update carries fewer bits. A formal analysis of this effect in the context of Grassmannian differential feedback is left for future work.

3.8.2 IA Performance Analysis

In this section I present simulation results to demonstrate the performance of IA when channel knowledge at the transmitter is obtained via the Grassmannian differential feedback strategy detailed in Section 3.5. I assume that the channels are correlated according to the same autoregressive model used in Section 3.8.1, and that all the channels have equal Doppler spreads. Moreover, I assume that $\rho_{k,\ell} = \rho, \ \forall k, \ \ell$. Thus I define the signal-to-noise

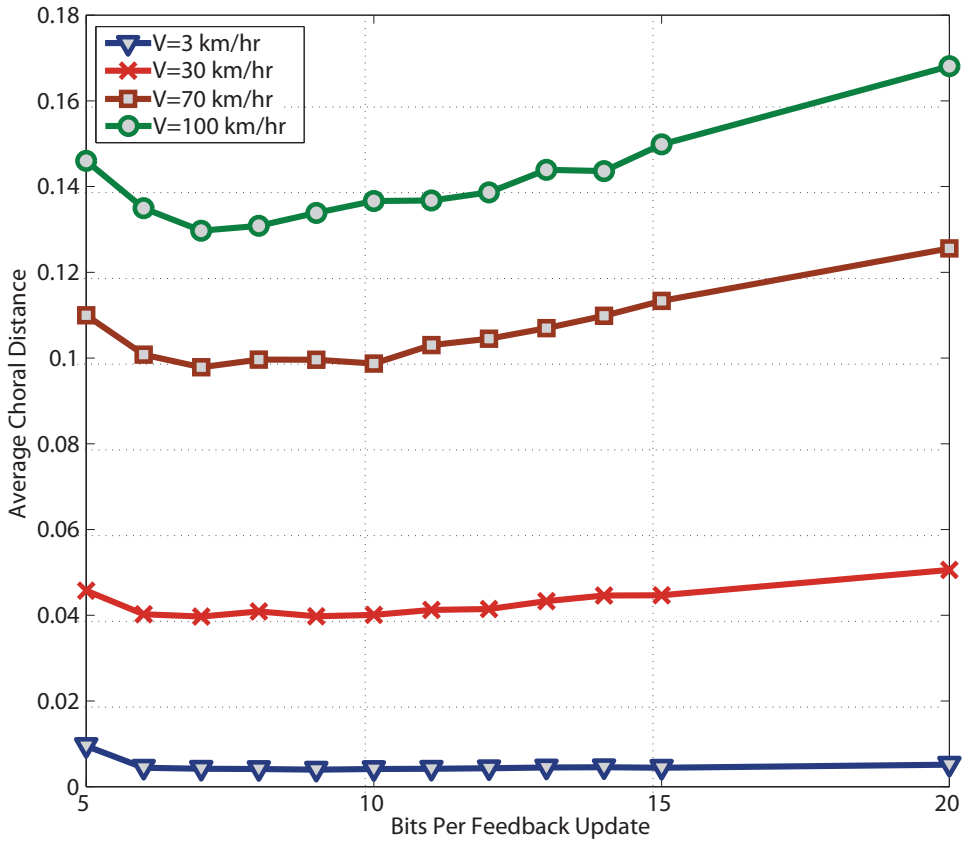

Figure 3.7: Keeping the total feedback rate fixed at 5000 bits/s, this figure shows the average chordal distance or quantization error, versus the number of feedback bits. This shows that it is most often optimal to send frequent low resolution feedback. While slow fading channels can support and benefit from large feedback update periods, the performance in faster fading channels significantly deteriorates if refresh rate is sacrificed for codebook size. In all cases $N_{\text{mag}} = 1$ and N_{dir} is varied.

ratio as $SNR = \frac{\rho P}{\sigma_{\text{n}}^2}$. Since the frequency extended system can be viewed as a

virtual MIMO system, the sum-rate achieved is calculated as [6, 57],

$$\bar{R}_{\text{sum}} = \sum_{k=1}^{K} \frac{1}{N_{\text{sc}}} \log_2 \left| \mathbf{I} + \frac{N_{\text{sc}} P}{d_k} \left(\sigma^2 \mathbf{I} + \frac{N_{\text{sc}} P}{d_k} \sum_{m \neq k} \mathbf{H}_{k,m} \widehat{\mathbf{F}}_m \widehat{\mathbf{F}}_m^* \mathbf{H}_{k,m}^* \right)^{-1} \times \left(\mathbf{H}_{k,k} \widehat{\mathbf{F}}_k \widehat{\mathbf{F}}_k^* \mathbf{H}_{kk}^* \right) \right|, \tag{3.29}$$

and the precoders, $\widehat{\mathbf{F}}_k = \left[\widehat{\mathbf{f}}_k^1, \widehat{\mathbf{f}}_k^2, \ldots, \widehat{\mathbf{f}}_k^{d_k} \right]$, are calculated given ideal or estimated CSI. Note here that the sum-rate calculated in (3.29) assumes that receivers use the perfect CSI available to them for decoding. For the results in this section, I use a modified version of the interference alignment algorithm proposed in [75], with the number of data symbols given by the conditions in [8]. Although a closed form solution for the IA precoders exists for the single antenna frequency extended interference channel in [8], it has been shown that the solution in [8] yields low sum-rates if not further improved as in [40].

Fig. 3.8 shows the sum-rate achieved by 3 users using IA over 16 channel extensions. CSI is obtained by the proposed algorithm, the strategy in [34], and memoryless vector quantization. For all the feedback strategies shown, each 3-tap channel is quantized using 10 bits. For the proposed feedback strategy and that of [34], 7 and 3 bits are used for the tangent direction and magnitude respectively. From Fig. 3.8, one can show that, in the case of perfect CSI, the rate of increase of sum-rate with SNR is 1.35, which approaches the 1.46 degrees of freedom predicted in theory over 16 subcarriers.

As for the performance of IA with Grassmannian differential feedback,

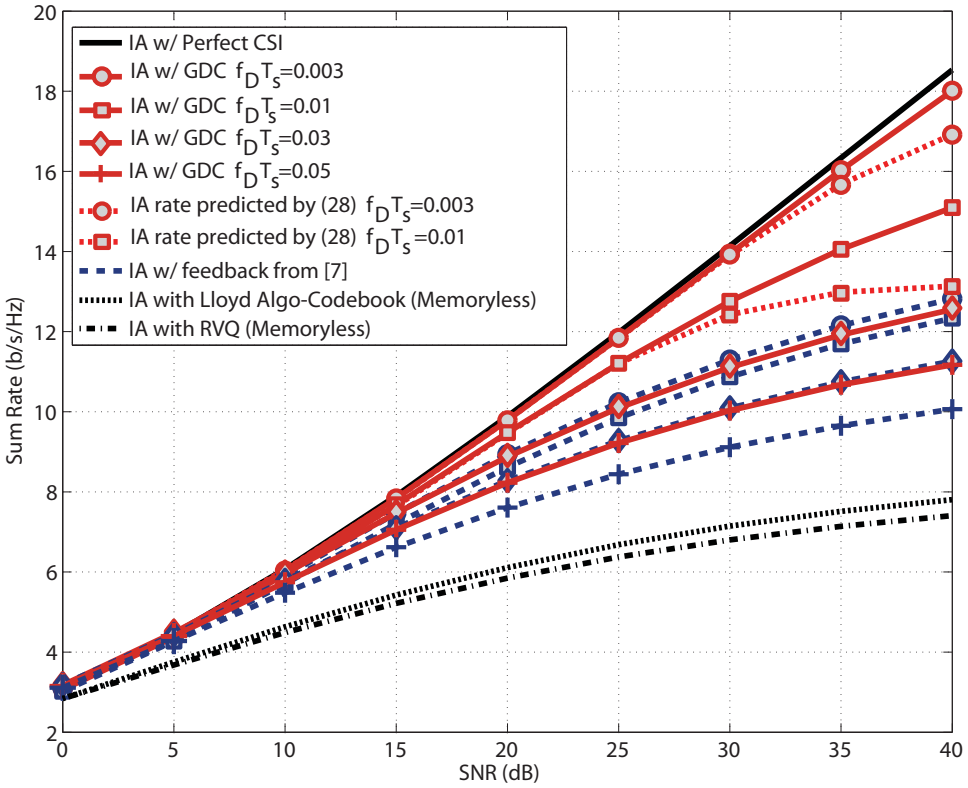

Figure 3.8: This figure shows the performance of IA with imperfect CSI obtained through the proposed algorithm, the algorithm in [34], as well as random vector quantization. This shows that for slowly varying channels, the proposed algorithm allows interference alignment networks to come very close to the perfect CSI upper bound.

Fig. 3.8 shows that the sum-rate achieved by IA is significantly improved by exploiting temporal correlation in the channel. Although multiplexing gain will likely never be preserved with a constant number of feedback bits, we see that for correlated channels, the proposed algorithm exhibits close to perfect performance over a wide range of SNR. Practically, this is in fact what is needed. Since a system will almost never function in the asymptotically high

SNR regime, the goal in practical systems is to optimize for the medium to high SNR regime where interference alignment is likely to be used. The proposed algorithm succeeds in providing negligible sum-rate loss at up to 30dB, in channels with a normalized Doppler of up to 10^{-2}. In fact this can be achieved with much less feedback bits as opposed to the 10 bit feedback performance shown in Fig. 3.8. The proposed algorithm continues to outperform memoryless quantization even at a significant Doppler of $f_{\mathrm{D}}T_{\mathrm{s}} = 0.05$ regardless of whether random codebooks (RVQ) or optimized codebooks, generated using the well-known Lloyd algorithm [25, 108], are used. The proposed algorithm also consistently outperforms the GPC algorithm in [34] as a result of the adaptive magnitude and direction codebooks used. Finally, Fig. 3.8 shows the analytical sum-rate predicted by (3.28) for the case of $f_{\mathrm{D}}T_{\mathrm{s}} = 0.003$ and 0.01. As stated in Section 3.7, the rate characterization is seen to be rather accurate for high SNRs up to 30dB, when 10 bits provide CSI of sufficient accuracy, but becomes progressively looser as SNR increases further and the CSI requirement becomes more stringent.

3.9 Conclusion

Limited feedback is a practical way to meet the CSI requirements of IA. To provide high resolution, I proposed a feedback strategy based on Grassmannian differential coding that uses the channel's temporal correlation to track channel responses by moving along geodesic paths defined via quantized tangent feedback. I showed that IA in conjunction with the proposed algorithm

can benefit from more accurate channel knowledge and provide better sum-rate performance for a large range of SNR conditions and a limited number of feedback bits. Moreover, my algorithm exhibits significant improvements over earlier methods of feeding back correlated time series on the manifold due to the optimized tangent and magnitude codebooks.

Chapter 4

The Overhead of Interference Alignment

I have thus far argued, and shown, that realizing the potential gains of interference alignment (IA) is contingent upon providing transmitters with sufficiently accurate channel state information (CSI). In addition to being accurate, however, the process of CSI acquisition must not exhaust system resources that may otherwise be used to transmit data. In this chapter, I study the performance of IA in multiple-input multiple-output systems where channel knowledge is acquired through training and the analog feedback strategy presented in Chapter 2. I improve on the multiplexing gain results of Chapter 2 by analytically designing the training and feedback system to maximize IA's *effective sum-rate*: a non-asymptotic performance metric that accounts for estimation error, training and feedback overhead, and channel selectivity. I analytically characterize effective sum-rate with overhead in relation to various parameters such as signal-to-noise ratio, Doppler spread, and feedback channel quality. This chapter shows that by properly designing the CSI acquisition process, IA can provide good sum-rate performance in a wide range of fading scenarios even when overhead is considered. Another observation from this work is that such overhead-aware analysis can help solve a number of practical network design problems. To demonstrate the concept of overhead-aware

network design, I consider the example problem of finding the optimal number of cooperative IA users based on signal power and mobility.

4.1 Introduction

The work in Chapters 2, 3, and [7, 45], have proven that limited or analog feedback algorithms can potentially provide an imperfect, yet practical and sufficiently accurate, replacement to the perfect CSI often assumed when designing IA algorithms [8, 21, 75, 100] or reporting theoretical IA gains. Namely, it was shown that no degrees-of-freedom are lost if perfect CSI was replaced by memorylessly quantized CSI in frequency selective systems, as long as codebook size scales with the system's signal-to-noise ratio. The differential feedback strategy in Chapter 3 successfully curbed the need for large codebooks in memoryless feedback systems by exploiting channel temporal correlation to improve quantization accuracy. Similarly, the work in Chapter 3 overcame the problem of scaling codebook size and relaxed the reliance on frequency selectivity for quantization by adopting a strategy known as analog feedback. Using analog feedback, a constant data rate gap from perfect CSI performance was shown, as long as the SNRs on the forward and feedback links are order-wise equal. A limitation of the analysis in [7, 45, 69], however, is that the asymptotic quantity, DoF, remains the primary performance metric considered. IA's actual sum-rate performance at *finite SNR*, especially when accounting for the time spent on overhead signaling, has yet to be considered.

Attempts to directly analyze or reduce overhead are limited to [26, 74,

109]. To analyze the effect of overhead, [109] considers the effective number of spatial DoF of an IA system with training and feedback. By considering DoF, however, [109] implicitly characterizes performance at infinitely high SNR. Alternatively, the work in [26, 74] on using network topology information to either partition users into optimally sized alignment groups [74] or apply IA to partially connected interference channels [26] does reduces the number of channels that must be shared, however, does not suggest an efficient training and feedback strategy nor characterizes the network's final performance with overhead.

4.2 Contributions

In this chapter, I characterize the performance of a MIMO IA system that is designed for perfect CSI operation yet only has access to imperfect CSI through training and analog feedback [9, 63, 71]. Thus, the performance demonstrated in this chapter constitutes a lower bound for systems that are designed to be more robust to imperfect CSI through improved precoding strategies such as [78] for example. I adopt a block-fading model wherein the channel remains constant over the block length, and varies independently across blocks. In contrast with earlier work on IA with feedback, I precisely model channel selectivity by leveraging the relationship between block-fading and continuous-fading channels shown in [37]. This relationship allows us to define the concept of Doppler spread in a block fading channel and explicitly relate the size of the coherence block to that Doppler spread. Since both CSI

acquisition and data transmission must now occur within the limits of a single coherence block, the IA system is faced with a non-trivial tradeoff: too much overhead leaves little time for payload data transmission, whereas too little overhead results in large sum-rate losses due to poor CSI quality [28, 37, 43, 55, 86]. In this chapter, I design the training and analog feedback system to maximize IA's *effective sum-rate*, a non-asymptotic performance metric that accounts for both CSI quality and CSI acquisition overhead. CSI acquisition overhead is a fundamental concept that was largely neglected in earlier work on IA with imperfect CSI.

I begin by giving a tractable expression for the IA sum-rate in genie-aided systems with perfect CSI, and extend the analysis under a general model for imperfect CSI. I then specialize the results to a system with training and analog feedback by characterizing CSI quality as a function of system parameters such as training overhead, feedback overhead and transmit power on both forward and reverse links. This results in a tractable expression for IA's effective sum-rate, which I proceed to optimize. To give a closed-form solution for the optimal effective sum-rate, I build on the method in [37] and optimize a series expansion of the objective function. In this chapter, I complete IA's performance analysis by analytically characterizing its maximum achievable effective sum-rate and the corresponding optimum overhead budget. The main insights and conclusions that can be drawn from the effective sum-rate analysis can be summarized as follows:

- Practical IA performance is not only a function of basic system param-

eters such as network size and SNR, but is tightly related to quantities such as Doppler spread, and feedback channel quality. Moreover, the dependence of both the maximum effective sum-rate, and the corresponding optimal overhead budget, on the various system parameters can be characterized accurately.

- By properly designing the training and feedback stages, IA can be made both feasible and beneficial in a wide range of fading scenarios, even when its relatively high overhead is considered.

- Overhead-aware analysis is essential to the design of IA networks. As an example of this observation, I use the overhead analysis to give simple results on the optimal number of cooperative IA users for channels with varying levels of selectivity.

The chapter is organized as follows. Section 4.3 introduces the MIMO system model for the forward and feedback channels and defines the main objective function of interest, i.e., effective sum-rate. Section 4.4 derives the sum-rate achieved by IA when perfect CSI is available, and the sum-rate achieved under a general model for imperfect CSI. Section 4.5 introduces analog feedback for the MIMO interference channel and analyzes the quality CSI it provides. Section 4.6 then presents the results on optimal training and feedback lengths. I give corroborative numerical results in Section 4.7 and conclude with Section 4.8.

4.3 System Model

Consider the K-user narrowband MIMO interference channel in which transmitter i communicates with its paired receiver i and interferes with all other receivers, $\ell \neq i$. For simplicity of exposition, consider a homogeneous network where all transmitters are equipped with N_{T} antennas and all receivers with N_{R} antennas, and each node pair communicates via $d \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$ independent spatial streams. The results can be generalized to a different number of streams or antennas at each node, provided that IA remains feasible [76].

Assuming perfect time and frequency synchronization, the sampled baseband signal at receiver i can be written as

$$\mathbf{y}_i = \sqrt{\frac{P}{d}}\mathbf{H}_{i,i}\mathbf{F}_i\mathbf{s}_i + \sum_{\ell \neq i}\sqrt{\frac{P}{d}}\mathbf{H}_{i,\ell}\mathbf{F}_\ell\mathbf{s}_\ell + \mathbf{v}_i, \tag{4.1}$$

where $\mathbf{y}_i$ is the $N_{\mathrm{R}} \times 1$ received signal vector, P is the transmit power, $\mathbf{H}_{i,\ell}$ is the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ discrete-time effective baseband channel matrix from transmitter ℓ to receiver i, $\mathbf{F}_i = \left[\mathbf{f}_i^1, \ \ldots, \ \mathbf{f}_i^d\right]$ is transmitter i's $N_{\mathrm{T}} \times d$ precoding matrix, $\mathbf{s}_i$ is the $d \times 1$ transmitted symbol vector at node i such that $\mathbb{E}\left[\mathbf{s}_i\mathbf{s}_i^*\right] = \mathbf{I}_d$, and $\mathbf{v}_i$ is a vector of i.i.d complex Gaussian noise samples with covariance matrix $\sigma^2\mathbf{I}_{N_{\mathrm{R}}}$. The channels $\mathbf{H}_{i,\ell}$ are assumed to be independent across users and each with i.i.d $\mathcal{CN}(0,1)$ entries. Large-scale fading can be included in the system model at the expense of a more involved exposition in Section 4.5.

The received signal at *transmitter* i on the feedback channel is

$$\overleftarrow{\mathbf{y}}_i = \sqrt{\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\mathbf{G}_{i,i}\overleftarrow{\mathbf{x}}_i + \sum_{\ell \neq i}\sqrt{\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\mathbf{G}_{\ell,i}\overleftarrow{\mathbf{x}}_\ell + \overleftarrow{\mathbf{v}}_i, \tag{4.2}$$

where P_{F} is the feedback power available such that $P_{\mathrm{F}}/P = \gamma$, $\mathbf{G}_{\ell,i}$ is the $N_{\mathrm{T}} \times N_{\mathrm{R}}$ discrete time feedback channel between receiver ℓ and transmitter i with i.i.d $\mathcal{CN}(0,1)$ entries, $\overleftarrow{\mathbf{x}}_i$ is the symbol vector with unit variance entries sent by receiver i, and $\overleftarrow{\mathbf{v}}_i$ is a complex vector of i.i.d circularly symmetric white Gaussian noise with covariance matrix $\sigma^2\mathbf{I}_{\mathbf{N}_{\mathrm{T}}}$. The forward and feedback channels are assumed to be independent in the error analysis of Section 4.5, i.e., a frequency division duplexed system or a general non-reciprocal system is assumed.

I adopt a block-fading channel model in which channels remain fixed for a period, T_{frame}, but vary independently from block to block. To model the effect of channel selectivity on IA performance, I set the block length to $T_{\mathrm{frame}} = \frac{1}{2f_{\mathrm{D}}}$, where f_{D} plays the role of the block fading channel's *effective Doppler spread*. The definition of f_{D} is motivated by the results in [37] showing a relationship between continuous fading and block fading systems. To enable IA over such a channel, both CSI acquisition and payload data transmission must occur within the coherence time T_{frame}, or else the CSI acquired becomes obsolete. The IA system then encounters a well-known tension between CSI acquisition and data transmission [28, 37, 43, 55, 86], and must allocate resources to each of the processes to optimize overall performance.

To account for CSI acquisition overhead, and to accurately characterize

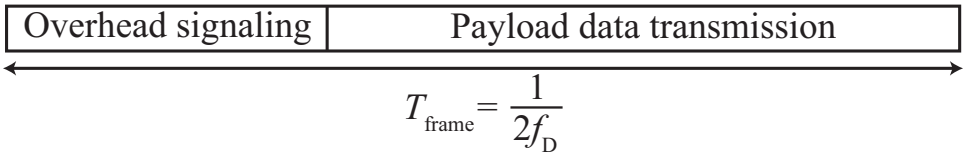

Figure 4.1: The overhead model adopted in which training and feedback consume resources that would otherwise be used for data transmission.

the *effective data rate* achieved by IA, I adopt the overhead model shown in Fig. 4.1. In this model, overhead signaling consumes time resources that could otherwise be used for data transmission, i.e., CSI acquisition penalizes effective sum-rate. For such an overhead model, the effective sum-rate (in bits/s/Hz) can be written as [28, 37, 43]

$$\bar{R}_{\text{eff}}\left(P, T_{\text{OHD}}\right) = \left(\frac{T_{\text{frame}} - T_{\text{OHD}}}{T_{\text{frame}}}\right) \bar{R}_{\text{sum}}(P, T_{\text{OHD}}) \tag{4.3}$$

where T_{OHD} is the total time spent on training and feeding back channels, and $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ is the average sum-rate in bits/s/Hz achieved by IA on the channel uses allocated for payload transmission. Using (4.3), and previous insights into IA performance, I highlight the tradeoff between overhead signaling and data transmission. Increasing overhead improves CSI quality and in turn improves $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$, but the relative period over which $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ can be achieved shrinks. A similar tension exists when lowering overhead; less overhead allows more channel uses for data transmission but the sum-rate per channel use suffers due to poor CSI quality. The objective then becomes maximizing the effective sum rate given in (4.3) by optimally trading off overhead with data transmission [28, 37, 43, 55, 86]. Throughout this chapter, I treat $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ as an information-theoretic quantity, and thus derive mutual

information-based sum-rates achievable without errors. IA performance can also be analyzed from the perspective of fixed-rate transmission where metrics such as bit error rate may be of interest [60].

4.4 Interference Alignment: An Average Sum-Rate Analysis

This section derives the average sum-rate achieved by IA in both genie-aided networks where channels are known perfectly, as well as practical systems where CSI is imperfect.

4.4.1 Interference Alignment with Perfect CSI

IA often achieves the full number of DoF supported by MIMO interference channels. In cases where the full DoF cannot be guaranteed, IA has been shown to provide significant gains in high-SNR sum-rate [21, 71, 75]. While this chapter focuses on IA, even better performance could be achieved with other precoding algorithms that seek a balance between interference minimization and signal power maximization [75, 87, 92]. The algorithms in [75, 87, 92], however, do not readily lend themselves to average sum-rate analysis.

To analyze IA sum-rates, I begin by examining the effective channels created after precoding and combining. For tractability, I focus on IA with a simple per-stream zero-forcing (ZF) receiver. Recall that in the high (but finite) SNR regime, where IA is most useful, gains from more involved receiver designs are limited. In such a system, receiver i projects its signal onto the

columns of the zero-forcing combiner $\mathbf{W}_i = \left[\mathbf{w}_i^1, \ldots, \mathbf{w}_i^m, \ldots, \mathbf{w}_i^d\right]$ which gives

$$(\mathbf{w}_i^m)^* \mathbf{y}_i = \sqrt{\frac{P}{d}} (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m s_i^m + \sqrt{\frac{P}{d}} \sum_{(k,\ell)\neq(i,m)} (\mathbf{w}_i^m)^* \mathbf{H}_{i,k} \mathbf{f}_k^\ell s_k^\ell + (\mathbf{w}_i^m)^* \mathbf{v}_i. \tag{4.4}$$

At the output of these linear receivers $\mathbf{w}_i^m$, the conditions for perfect IA can be stated as [21]

$$(\mathbf{w}_i^m)^* \mathbf{H}_{i,k} \mathbf{f}_k^\ell = 0, \qquad \forall (k,\ell) \neq (i,m) \tag{4.5}$$

$$|(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m| \geq c > 0, \qquad \forall i, m, \tag{4.6}$$

where alignment is guaranteed by (4.5), and (4.6) is satisfied almost surely [8, 21].

As a result of conditions (4.5) and (4.6), the combination of IA and ZF effectively creates Kd non-interfering scalar channels. The maximum mutual information across these channels is achieved via Gaussian signaling which yields an instantaneous sum-rate given by

$$R_{\text{sum}} = \sum_{i=1}^{K} \sum_{m=1}^{d} \log_2 \left(1 + \frac{\frac{P}{d} |(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2}{\sigma^2} \right). \tag{4.7}$$

To derive an expression for the average sum-rate, i.e., $\bar{R}_{\text{sum}} = \mathbb{E}\left[R_{\text{sum}}\right]$, I first give the following lemma.

Lemma 11 *([69, Appendix A]). The effective direct channels* $(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m$ *are independent and Gaussian distributed with unit variance if: (i) the precoders* $\mathbf{F}_i$ *are unitary and are generated by an IA solution that does not consider*

the direct channels $\mathbf{H}_{i,i}$*, and (ii) the combiners* $\mathbf{W}_i$ *are calculated to simply zero-force inter-user and inter-stream interference.*

The conditions Lemma 1 places on precoder and combiner calculation are satisfied by most IA solutions such as [8, 21, 75, 100]. Hence, as a result of Lemma 11, the scalar point-to-point channels created by the combination of IA and ZF experience Rayleigh fading. As a result, the average sum-rate can be written in exponential integral form as [72, 93]

$$\begin{aligned}\bar{R}_{\text{sum}}(\rho) &= \sum_{i=1}^{K}\sum_{m=1}^{d}\mathbb{E}\left[\log_2\left(1+\frac{\frac{P}{d}\left|(\mathbf{w}_i^m)^*\mathbf{H}_{i,i}\mathbf{f}_i^m\right|^2}{\sigma^2}\right)\right] \\ &= Kd\log_2(e)e^{1/\rho}E_1\left(\frac{1}{\rho}\right),\end{aligned} \tag{4.8}$$

which is written as a function of the per-stream SNR, $\rho = \frac{P}{d\sigma^2}$, and $E_1(\eta) = \int_1^\infty t^{-1}e^{-\eta t}dt$ is an exponential integral.

4.4.2 Interference Alignment with CSI from Training and Feedback

When the channels are not known perfectly, interference cannot be aligned perfectly. Misalignment leads to "leakage interference", which reduces the signal-to-interference-plus-noise ratio (SINR) in the desired signal space. Moreover, imperfect knowledge of the direct channel implies that receivers will perform mismatched decoding [47], again reducing effective SINR. In this section, I examine the effect of imperfect CSI on the performance of an IA system that is optimized for perfect-CSI operation, i.e., a system that does not consider CSI imperfection in its design. Thus, the performance results

demonstrated in this chapter can be improved upon by adopting precoding algorithms that are more robust to CSI errors such as [78].

Consider an IA system in which transmitters use a common set of channel estimates as input to an IA solution such as [8, 21, 75, 100], i.e., they calculate imperfect IA precoders $\widehat{\mathbf{F}}_i$ and combiners $\widehat{\mathbf{W}}_i$. Denote the channel estimates as $\widehat{\mathbf{H}}_{i,\ell}$ and the corresponding error as $\widetilde{\mathbf{H}}_{i,\ell} = \mathbf{H}_{i,\ell} - \widehat{\mathbf{H}}_{i,\ell}$. In this system, the IA solution satisfies

$$(\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell = 0, \qquad \forall (k,\ell) \neq (i,m) \tag{4.9}$$

$$\left| (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m \right| \geq c > 0, \qquad \forall i, m. \tag{4.10}$$

I assume receivers obtain perfect knowledge of the combiners $\widehat{\mathbf{W}}_i$ and the imperfect effective direct channels $\widehat{\mathbf{w}}_i^m \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m$ for detection, an assumption similar to [7, 43, 45, 69, 99][1] whose relaxation is a topic of future work. In general, receiver side information about the effective channels can be acquired blindly [30] or via additional training or silent phases [9]. For such an IA system, the received signal after projection is

$$(\widehat{\mathbf{w}}_i^m)^* \mathbf{y}_i = \sqrt{\frac{P}{d}} (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m s_i^m + \sqrt{\frac{P}{d}} \sum_{k,\ell} (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell s_k^\ell + (\widehat{\mathbf{w}}_i^m)^* \mathbf{v}_i, \tag{4.11}$$

where I have used the fact that conditions (4.9) and (4.10) are satisfied, and thus $(\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,k} \widehat{\mathbf{f}}_k^\ell = (\widehat{\mathbf{w}}_i^m)^* (\widehat{\mathbf{H}}_{i,k} + \widetilde{\mathbf{H}}_{i,k}) \widehat{\mathbf{f}}_k^\ell = (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell$.

[1] In fact [7, 45, 99] place a stronger assumption summarized by the receivers' knowledge of the exact imperfect CSI known to the transmitters. The two assumptions are functionally equivalent from the perspective of the sum-rate analysis, i.e., all that is needed is the receivers' knowledge of $\widehat{\mathbf{w}}_i^m$ and of the scalars $\widehat{\mathbf{w}}_i^m \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m$.

Analyzing the maximum sum-rates achievable on the channel in (4.11) is in general difficult, as it requires optimizing the distribution of the input symbols $\mathbf{s}_i$ for the interference channel in (4.11). Recall, however, that the objective is to analyze a system optimized for *perfect CSI operation*, i.e. one that does not account for CSI imperfection. This enables making the following assumptions that would be expected from a system optimized for perfect CSI operation.

Assumption 12. *Transmitters use a typical Gaussian codebook made up of i.i.d. symbols to form the symbol vectors* $\mathbf{s}_i$*. Such a signaling codebook, which was optimal for the interference free channels created by IA with perfect CSI, may no longer be optimal now that CSI is imperfect.*

Assumption 13. *Receivers perform nearest neighbor decoding using the estimates* $\widehat{\mathbf{H}}_{i,i}$*. Nearest neighbor decoding would again be optimal with perfect CSI. The nearest neighbor decoder, the channel estimates and the signaling codebook together satisfy the conditions outlined in [47] for Corollary 3.0.1 of [47] to hold with equality, meaning that the estimation error plays the role of an additional source of Gaussian noise irrespective of its actual distribution.*

Under Assumptions 12-13, and combining the results of [47] and [46], the average sum-rate achieved can be written as

$$\bar{R}_{\text{sum}}(\rho) = \sum_{i=1}^{K} \sum_{m=1}^{d} \mathbb{E}\left[\log_2 \left(1 + \frac{\frac{P}{d} \left| (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2}{\sum\limits_{k,\ell} \mathbb{E}\left[\frac{P}{d} \left| (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell \right|^2 \right] + \sigma^2} \right) \right] \tag{4.12}$$

where it should be noted that the outer expectation is now only over the fading on the direct channel and not the interference. Therefore, the leakage interference terms $(\widehat{\mathbf{w}}_i^m)^*\widetilde{\mathbf{H}}_{i,k}\widehat{\mathbf{f}}_k^\ell$ indeed play the role of independent sources of additive Gaussian noise, regardless of their distribution.

When the entries of $\widetilde{\mathbf{H}}_{i,k}$ are zero-mean and uncorrelated with a variance of $\sigma^2_{\widetilde{\mathbf{H}}}$, it follows that $\mathbb{E}\left[\frac{P}{d}\left|(\widehat{\mathbf{w}}_i^m)^*\widetilde{\mathbf{H}}_{i,k}\widehat{\mathbf{f}}_k^\ell\right|^2\right] = \frac{P}{d}\sigma^2_{\widetilde{\mathbf{H}}}$, thus the denominator in (4.12) is simply $KP\sigma^2_{\widetilde{\mathbf{H}}} + \sigma^2$. Moreover, if the estimates $\widehat{\mathbf{H}}_{i,k}$ are MMSE estimates of $\mathbf{H}_{i,k}$, the entries of $\widehat{\mathbf{H}}_{i,k}$ have a variance of $1 - \sigma^2_{\widetilde{\mathbf{H}}}$. This results in an effective average SINR that can be written as

$$\rho_{\text{eff}} = \frac{\rho(1-\sigma^2_{\widetilde{\mathbf{H}}})}{\rho K d\sigma^2_{\widetilde{\mathbf{H}}} + 1}, \tag{4.13}$$

where ρ is the per-stream SNR defined in (4.8). If the estimated direct channels $\widehat{\mathbf{H}}_{i,i}$ is Gaussian, the average sum-rate achieved by IA with imperfect CSI is again given in exponential integral form as

$$\bar{R}_{\text{sum}}(\rho_{\text{eff}}) = Kd\log_2(e)e^{1/\rho_{\text{eff}}}E_1\left(\frac{1}{\rho_{\text{eff}}}\right). \tag{4.14}$$

To evaluate sum-rate achieved by IA, one must now characterize ρ_{eff} or equivalently $\sigma^2_{\widetilde{\mathbf{H}}}$. In Section 4.5, I specialize my result for a system with training and analog CSI feedback and later optimize IA's *effective data rate with overhead* in Section 4.6.

4.5 Training and Analog feedback

I propose to split the acquisition of CSI at the transmitter into three main phases. First, the transmitters train the forward channels via pilots.

Second, the receivers train the feedback channels via pilots, setting the stage for the forward transmitters to estimate the feedback information in the next stage. Finally, the receivers feedback information about the forward channels in an analog fashion, i.e., as unquantized complex symbols. I can then characterize the CSI error introduced in the CSI acquisition phase by examining the three stages.

4.5.1 Forward and Feedback Channel Training

In the first training phase, each transmitter k sends an orthogonal pilot sequence matrix $\mathbf{\Phi}_k$, i.e., $\mathbf{\Phi}_i\mathbf{\Phi}_k^* = \delta_{ik}\mathbf{I}_{N_\mathrm{T}}$, over a training period τ_t [62]. Pilot orthogonality imposes the constraint $\tau_\mathrm{t} \geq KN_\mathrm{T}$. Each receiver i then observes the $N_\mathrm{R} \times \tau_\mathrm{t}$ matrix

$$\mathbf{Y}_i = \sqrt{\frac{\tau_\mathrm{t}P}{N_\mathrm{T}}} \sum_{k=1}^{K} \mathbf{H}_{i,k}\mathbf{\Phi}_k + \mathbf{V}_i, \quad \forall i, \tag{4.15}$$

where $\mathbf{V}_i$ is an $N_\mathrm{R} \times \tau_\mathrm{t}$ matrix of noise terms. Using $\mathbf{Y}_i$, receiver i calculates an MMSE estimate of its incoming channels $\mathbf{H}_{i,k}$ $\forall k$ given by

$$\widehat{\mathbf{H}}_{i,k}^r = \frac{\sqrt{\frac{\tau_\mathrm{t}P}{N_\mathrm{T}}}}{\sigma^2 + \frac{\tau_\mathrm{t}P}{N_\mathrm{T}}} \mathbf{Y}_i\mathbf{\Phi}_k^*, \quad \forall k. \tag{4.16}$$

where the superscript $(\cdot)^r$ emphasizes that $\widehat{\mathbf{H}}_{i,k}^r$ are the channel estimates gathered at the receiver before they are relayed back to the transmitters and further corrupted. At the output of this first training stage, the channel estimates $\widehat{\mathbf{H}}_{i,k}^r$ have i.i.d. $\mathcal{CN}(0, \frac{\tau_\mathrm{t}P/N_\mathrm{T}}{\sigma^2+\tau_\mathrm{t}P/N_\mathrm{T}})$ entries with corresponding errors $\widetilde{\mathbf{H}}_{i,k}^r \sim \mathcal{CN}(0, \frac{\sigma^2}{\sigma^2+\tau_\mathrm{t}P/N_\mathrm{T}})$.

The feedback channel training phase proceeds similarly. Namely, the receivers transmit orthogonal pilot sequences over a training period $\tau_{\mathrm{p}} \geq KN_{\mathrm{R}}$. The transmitters independently compute MMSE estimates of their incoming channels, resulting in estimates $\widehat{\mathbf{G}}_{k,i} \sim \mathcal{CN}\left(0, \frac{\frac{\tau_{\mathrm{p}} P_{\mathrm{F}}}{N_{\mathrm{R}}}}{\sigma^2 + \frac{\tau_{\mathrm{p}} P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)$ with corresponding error terms $\widetilde{\mathbf{G}}_{k,i} \sim \mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_{\mathrm{p}} P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)$.

4.5.2 Analog Feedback

After forward and feedback channel training, the receivers feedback their channel estimates $\widehat{\mathbf{H}}^r_{i,k}$ in an analog fashion during a feedback period τ_{f}. This is accomplished by first post-multiplying each $N_{\mathrm{R}} \times KN_{\mathrm{T}}$ feedback matrix $\left[\widehat{\mathbf{H}}^r_{i,1} \ldots \widehat{\mathbf{H}}^r_{i,K}\right]$ with a $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix $\mathbf{\Psi}_i$ such that $\mathbf{\Psi}_i \mathbf{\Psi}_k^* = \delta_{i,k} \mathbf{I}_{KN_{\mathrm{T}}}$ [63, 69]. The spreading matrices $\mathbf{\Psi}_i$ orthogonalize the feedback from different users and facilitate estimation. This orthogonality constraint requires that $\tau_{\mathrm{f}} \geq K^2 N_{\mathrm{T}}$. The transmitted $N_{\mathrm{R}} \times \tau_{\mathrm{f}}$ feedback matrix $\overleftarrow{\mathbf{X}}_i$ from receiver i can be written as [63, 69]

$$\overleftarrow{\mathbf{X}}_i = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{KN_{\mathrm{T}} N_{\mathrm{R}}} \left(\frac{\tau_{\mathrm{t}} P / N_{\mathrm{T}}}{\sigma^2 + \tau_{\mathrm{t}} P / N_{\mathrm{T}}}\right)^{-1}} \left[\widehat{\mathbf{H}}^r_{i,1} \ldots \widehat{\mathbf{H}}^r_{i,K}\right] \mathbf{\Psi}_i, \tag{4.17}$$

where the leading scalar is to ensure that the average transmit power constraints are satisfied with equality, i.e., one can verify that $\mathbb{E}\left[\mathrm{trace}\left(\overleftarrow{\mathbf{X}}_i \overleftarrow{\mathbf{X}}_i^*\right)\right] = \tau_{\mathrm{f}} P_{\mathrm{F}}$. I write the concatenated $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix of feedback symbols observed

by all transmitters as

$$\overleftarrow{\mathbf{Y}}_{\mathrm{f}} = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{K N_{\mathrm{T}} N_{\mathrm{R}}} \left(\frac{\tau_{\mathrm{t}} P / N_{\mathrm{T}}}{\sigma^2 + \tau_{\mathrm{t}} P / N_{\mathrm{T}}} \right)^{-1}} \sum_{i=1}^{K} \begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix} \left[\widehat{\mathbf{H}}^{r}_{i,1} \dots \widehat{\mathbf{H}}^{r}_{i,K} \right] \mathbf{\Psi}_i + \mathbf{V}, \tag{4.18}$$

where $\mathbf{V}$ is the $KN_{\mathrm{T}} \times \tau_{\mathrm{f}}$ matrix of i.i.d Gaussian noise.

To simplify the performance analysis, I make the same assumption as in [69]: at the end of the feedback phase, the transmitters cooperate by sharing their rows of the received feedback matrix $\overleftarrow{\mathbf{Y}}_{\mathrm{f}}$ which enables them to form a unified estimate of the forward channels $\mathbf{H}_{i,k}$. I refer the reader to [69] for a discussion of this cooperative assumption and for alternative non-cooperative approaches that are shown to perform close to this special case.

Under this cooperative assumption, the transmitters estimate $\mathbf{H}_{i,k}$ $\forall k$ by first isolating the feedback sent by receiver i. They post-multiply their received symbols by $\mathbf{\Psi}_i^*$ to compute

$$\overleftarrow{\mathbf{Y}}_{\mathrm{f}} \mathbf{\Psi}_i^* = \sqrt{\frac{\tau_{\mathrm{f}} P_{\mathrm{F}}}{K N_{\mathrm{T}} N_{\mathrm{R}}} \left(\frac{\tau_{\mathrm{t}} P / N_{\mathrm{T}}}{\sigma^2 + \tau_{\mathrm{t}} P / N_{\mathrm{T}}} \right)^{-1}} \underbrace{\begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix}}_{\mathbf{G}_i} \left[\widehat{\mathbf{H}}^{r}_{i,1} \dots \widehat{\mathbf{H}}^{r}_{i,K} \right] + \mathbf{V} \mathbf{\Psi}_i^*. \tag{4.19}$$

The transmitters then compute a common linear MMSE estimate of the forward channels $\mathbf{H}_{i,k}$ $\forall i,k$ using their feedback channel estimates $\widehat{\mathbf{G}}_{i,k}$ $\forall i,k$, and assuming that $KN_{\mathrm{T}} \geq N_{\mathrm{R}}$ so that the estimation problem is well posed. After a lengthy yet standard application of the orthogonality principle and the

matrix inversion lemma, the MMSE estimate is given by

$$\widehat{\mathbf{H}}_i = \sqrt{\frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}P_f}\left(\frac{\tau_{\mathrm{t}}P/N_{\mathrm{T}}}{\sigma^2+\tau_{\mathrm{t}}P/N_{\mathrm{T}}}\right)^{-1}}\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i+\gamma_1\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i+\gamma_2\mathbf{I}_{N_{\mathrm{R}}}\right)^{-1}\widehat{\mathbf{G}}_i^*\overleftarrow{\mathbf{Y}}_{\mathrm{f}}\mathbf{\Psi}_i^*, \tag{4.20}$$

where I have written (4.20) in terms of $\widehat{\mathbf{H}}_i = \left[\widehat{\mathbf{H}}_{i,1},\ \ldots,\ \widehat{\mathbf{H}}_{i,K}\right]\ \forall i$, the concatenated estimate of the channels $\mathbf{H}_i = [\mathbf{H}_{i,1},\ \ldots,\ \mathbf{H}_{i,K}]\ \forall i$, for the sake of notational brevity. The constants γ_1 and γ_2 are the MMSE regularization factors. For completeness, γ_1 and γ_2 are given by

$$\gamma_1 = \frac{N_{\mathrm{T}}\sigma^2}{P\tau_{\mathrm{t}}}, \qquad \gamma_2 = \left(1+\frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P}\right)\left(\frac{\sigma^2KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}P_{\mathrm{F}}}+\frac{N_{\mathrm{R}}\sigma^2}{\sigma^2+\tau_{\mathrm{p}}P_{\mathrm{F}}/N_{\mathrm{R}}}\right). \tag{4.21}$$

In essence, γ_1 captures the effect of the noise in the transmitted estimates $\widehat{\mathbf{H}}_{i,k}^r$, while γ_2 captures the effect of the noise in the estimates $\widehat{\mathbf{G}}_{i,k}$ as well as the noise observed during feedback.

Having formalized the three training and analog feedback stages, I now analyze the variance,[2] $\sigma^2_{\widetilde{\mathbf{H}}}$, of the CSI error $\mathbf{H}_{i,k}-\widehat{\mathbf{H}}_{i,k}$, which automatically yields an estimated CSI variance of $1-\sigma^2_{\widetilde{\mathbf{H}}}$. Unfortunately, writing $\sigma^2_{\widetilde{\mathbf{H}}}$ exactly yields rather cumbersome expressions. For this reason, I replace the variance of the MMSE estimation error by that of a zero-forcing estimator in a manner similar to [9, 111]. This ZF simplification intuitively amounts to deriving a high SNR result [9] and mathematically amounts to neglecting the constants γ_1 and γ_2; recall that moderately high SNR is after all the main operating

[2] I in fact derive the entire covariance matrix for the columns of $\mathbf{H}_{i,k}-\widehat{\mathbf{H}}_{i,k}$. I show that the covariance matrices are scaled identities and thus the second order statistics of the error are entirely described by a scalar variance.

region of interest for IA. Numerical results in Section 4.7 will demonstrate that the effect of this simplification is negligible.

By neglecting γ_1 and γ_2, and after some algebraic manipulation, I find that the error $\widetilde{\mathbf{H}}_i = \mathbf{H}_i - \widehat{\mathbf{H}}_i$ at the end of the three training and feedback phases can be written as

$$\widetilde{\mathbf{H}}_i = \sqrt{1 + \frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P}} \times \left[\widetilde{\mathbf{H}}_i^r + \left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}\widehat{\mathbf{G}}_i^*\left(\sqrt{1 + \frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P}}\widetilde{\mathbf{G}}_i\widehat{\mathbf{H}}_i^r + \sqrt{\frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{p}}P_f}}\mathbf{V}\boldsymbol{\Psi}_i^*\right)\right]. \tag{4.22}$$

As can be seen from (4.22), the resulting CSI error is a combination of three terms: the first due to forward channel estimation error $\widetilde{\mathbf{H}}_i^r$, the second due to feedback channel estimation error $\widetilde{\mathbf{G}}_i$, and the third due to feedback noise.

To derive the statistics of $\widetilde{\mathbf{H}}_i$, I note the following three facts about the three terms in (4.22):

1. The entries of $\widetilde{\mathbf{H}}_i^r$ are uncorrelated $\mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_{\mathrm{t}}P}{N_{\mathrm{T}}}}\right)$ variables as shown in Section 4.5.1.

2. Similarly, the entries of $\widetilde{\mathbf{G}}_i$ are $\mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)$ implying that $\widetilde{\mathbf{G}}_i\widehat{\mathbf{H}}_i^r$ has independent entries with variance equal to $\frac{N_{\mathrm{R}}\sigma^2}{\sigma^2 + \frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}} \frac{\tau_{\mathrm{t}}PN_{\mathrm{T}}}{\sigma^2 + \tau_{\mathrm{t}}P/N_{\mathrm{T}}}$.

3. The entries of $\mathbf{V}$ are uncorrelated $\mathcal{CN}(0, \sigma^2)$ variables and so are the elements of $\mathbf{V}\boldsymbol{\Psi}_i^*$ since the matrix $\boldsymbol{\Psi}_i$ is unitary.

Combining the properties stated, the conditional covariance of each column of $\widetilde{\mathbf{H}}_i$ denoted $\widetilde{\mathbf{H}}_i^{(\ell)}$, conditioned of $\widehat{\mathbf{G}}_i$ is [63, 69]

$$\mathbb{E}\left(\widetilde{\mathbf{H}}_i^{(\ell)}\widetilde{\mathbf{H}}_i^{(\ell)*}|\widehat{\mathbf{G}}_i\right) = \frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P}\mathbf{I}_{N_{\mathrm{R}}} + \left(\frac{KN_{\mathrm{T}}N_{\mathrm{R}}\sigma^2}{\tau_{\mathrm{f}}P_{\mathrm{F}}} + \frac{N_{\mathrm{R}}\sigma^2}{\sigma^2+\tau_{\mathrm{p}}\frac{P_{\mathrm{F}}}{N_{\mathrm{R}}}}\right)\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}. \tag{4.23}$$

Since the entries of the MMSE estimate $\widehat{\mathbf{G}}_i$ are Gaussian, the matrix $\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}$ has an inverse-Wishart distribution [102]. Moreover, since $\widehat{\mathbf{G}}_i$ has uncorrelated entries with a variance of $\frac{\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}{\sigma^2+\frac{\tau_{\mathrm{p}}P_{\mathrm{F}}}{N_{\mathrm{R}}}}$, $\left(\widehat{\mathbf{G}}_i^*\widehat{\mathbf{G}}_i\right)^{-1}$ has a covariance matrix equal to a properly scaled identity [63, 69, 102]. Thus marginalizing (4.23) over $\widehat{\mathbf{G}}_i$, I find that $\widetilde{\mathbf{H}}_i$ has independent columns with scaled identity covariance matrices with diagonal entries given by

$$\sigma^2_{\widetilde{\mathbf{H}}} = \frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P} + \frac{\sigma^2}{(KN_{\mathrm{T}}-N_{\mathrm{R}})P_{\mathrm{F}}}\left(\frac{{N_{\mathrm{R}}}^2}{\tau_{\mathrm{p}}} + \frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\tau_{\mathrm{f}}}\left(1+\frac{N_{\mathrm{R}}\sigma^2}{\tau_{\mathrm{p}}P_{\mathrm{F}}}\right)\right). \tag{4.24}$$

The same high SNR simplification adopted earlier to replace MMSE estimation error by ZF estimation error, however, allows us to further simplify (4.24) by writing

$$\sigma^2_{\widetilde{\mathbf{H}}} = \frac{N_{\mathrm{T}}\sigma^2}{\tau_{\mathrm{t}}P} + \frac{\sigma^2}{P(KN_{\mathrm{T}}-N_{\mathrm{R}})}\left(\frac{{N_{\mathrm{R}}}^2}{\gamma\tau_{\mathrm{p}}} + \frac{KN_{\mathrm{T}}N_{\mathrm{R}}}{\gamma\tau_{\mathrm{f}}}\right), \tag{4.25}$$

which completes the characterization of the distortion introduced by training and analog feedback.

Note: Finally, a word on applying the results of Section 4.4 to the analog feedback system described. First, note that the analog feedback system satisfies Assumption 13, and the estimates yield $\mathbb{E}\left[\frac{P}{d}\left|(\widehat{\mathbf{w}}_i^m)^*\widetilde{\mathbf{H}}_{i,k}\widehat{\mathbf{f}}_k^\ell\right|^2\right] = \frac{P}{d}\sigma^2_{\widetilde{\mathbf{H}}}$ and $\mathbb{E}\left[\frac{P}{d}\left|(\widehat{\mathbf{w}}_i^m)^*\widehat{\mathbf{H}}_{i,i}\widehat{\mathbf{f}}_i^m\right|^2\right] = \frac{P}{d}(1-\sigma^2_{\widetilde{\mathbf{H}}})$ as needed. One subtlety though is that

the fading on the feedback channel introduces non-Gaussian terms into the estimates $\widehat{\mathbf{H}}_{i,i}$, yet (4.14) is only exact when the estimates are truly Gaussian. For fairly accurate estimation, however, $\widehat{\mathbf{H}}_{i,i}$ can be well approximated by a Gaussian. Moreover, it will be clear from the results of Section 4.7 that the effect of this is negligible.

4.6 Optimizing Overhead and Effective Sum-Rate

Having formally quantified IA sum-rate as a function of SNR and CSI quality, and characterized CSI quality in terms of training and feedback resources, I redefine both the optimization problem and objective function as

$$\bar{R}^{\star}_{\text{eff}}(P) = \max_{\tau_{\text{t}},\tau_{\text{p}},\tau_{\text{f}}} \left(\frac{T_{\text{frame}} - (\tau_{\text{t}} + \tau_{\text{p}} + \tau_{\text{f}})}{T_{\text{frame}}} \right) \bar{R}_{\text{sum}}(\rho_{\text{eff}}), \tag{4.26}$$

where I have used $(\cdot)^{\star}$ to denote optimality. Note from (4.26) that ρ_{eff} depends on $\sigma^2_{\widetilde{\mathbf{H}}}$ and thus on τ_{t}, τ_{p}, and τ_{f}. The problem in (4.26) can be rewritten in a more tractable form as [43]

$$\bar{R}^{\star}_{\text{eff}}(P) = \max_{\substack{\alpha \\ \alpha_{\min} \leq \alpha \leq 1}} \left[(1-\alpha) \max_{\substack{\tau_{\text{t}},\ \tau_{\text{p}},\ \tau_{\text{f}} \\ \tau_{\text{t}}+\tau_{\text{p}}+\tau_{\text{f}} = \alpha T_{\text{frame}}}} \bar{R}_{\text{sum}}(\rho_{\text{eff}}) \right], \tag{4.27}$$

where $\alpha_{\min} = K(N_{\text{T}} + N_{\text{R}} + KN_{\text{T}})/T_{\text{frame}}$ and is dictated by the minimum number of training and feedback symbols needed to render the estimation problems in Section 4.5 well defined. The inner maximization in (4.27) optimizes sum-rate for a fixed overhead length of $T_{\text{OHD}} = \alpha T_{\text{frame}}$ and the outer maximization finds the optimal α thereby completing the solution.

Since $\bar{R}_{\text{sum}}(\rho_{\text{eff}})$ is decreasing in $\sigma^2_{\widetilde{\mathbf{H}}}$, the inner maximization step simplifies to

$$\sigma^{2\star}_{\widetilde{\mathbf{H}}} = \min_{\tau_{\text{t}},\ \tau_{\text{p}},\ \tau_{\text{f}}} \quad \frac{N_{\text{T}}\sigma^2}{\tau_{\text{t}}P} + \frac{\sigma^2}{P(KN_{\text{T}} - N_{\text{R}})}\left(\frac{KN_{\text{T}}N_{\text{R}}}{\gamma\tau_{\text{f}}} + \frac{{N_{\text{R}}}^2}{\gamma\tau_{\text{p}}}\right) \qquad (4.28)$$
$$s.t. \qquad \tau_{\text{t}} + \tau_{\text{p}} + \tau_{\text{f}} = \alpha T_{\text{frame}}.$$

Although (4.28) is an integer problem, its continuous relaxation is convex. Applying standard convex optimization techniques, the Lagrangian for the inner maximization is

$$\Lambda(\tau_{\text{t}}, \tau_{\text{p}}, \tau_{\text{f}}, \lambda) = \frac{N_{\text{T}}\sigma^2}{\tau_{\text{t}}P} + \frac{\sigma^2}{P(KN_{\text{T}} - N_{\text{R}})}\left(\frac{KN_{\text{T}}N_{\text{R}}}{\gamma\tau_{\text{f}}} + \frac{{N_{\text{R}}}^2}{\gamma\tau_{\text{p}}}\right) + \lambda\left(\tau_{\text{t}} + \tau_{\text{p}} + \tau_{\text{f}} - \alpha T_{\text{frame}}\right). \qquad (4.29)$$

Solving for the first order KKT conditions, I obtain the optimal training and feedback times as a function of the total overhead budget αT_{frame} as

$$\begin{aligned}
{\tau_{\text{t}}}^{\star} &= \frac{\sqrt{\gamma N_{\text{T}}(KN_{\text{T}} - N_{\text{R}})}}{\mu}\alpha T_{\text{frame}}, \\
{\tau_{\text{p}}}^{\star} &= \frac{N_{\text{R}}}{\mu}\alpha T_{\text{frame}}, \\
{\tau_{\text{f}}}^{\star} &= \frac{\sqrt{KN_{\text{T}}N_{\text{R}}}}{\mu}\alpha T_{\text{frame}},
\end{aligned}$$

where $\mu = \sqrt{\gamma N_{\text{T}}(KN_{\text{T}} - N_{\text{R}})} + N_{\text{R}} + \sqrt{KN_{\text{T}}N_{\text{R}}}$. After solving the problem's continuous relaxation, convexity implies that for any given feasible overhead budget αT_{frame} simply examining the few integer neighbors of the points ${\tau_{\text{t}}}^{\star}$, ${\tau_{\text{p}}}^{\star}$, ${\tau_{\text{f}}}^{\star}$ yields the integer training and feedback times that minimize CSI distortion, i.e., optimal integer training and feedback times can be found by a simple search over the grid neighbors of the non-integer solution. Proceeding

with the continuous relaxation, the minimum CSI distortion for an overhead budget αT_{frame} is

$$\sigma^{2\star}_{\tilde{\mathbf{H}}} = \frac{\sigma^2 \left(\sqrt{K N_{\text{T}} N_{\text{R}}} + N_{\text{R}} + \sqrt{\gamma N_{\text{T}} (K N_{\text{T}} - N_{\text{R}})} \right)^2}{\gamma P (K N_{\text{T}} - N_{\text{R}}) \alpha T_{\text{frame}}}. \tag{4.30}$$

Having found the optimal allocation of τ_{t}, τ_{p}, and τ_{f} for a fixed overhead budget, what remains is to optimize the budget itself. The outer optimization in (4.26), however, does not admit a closed form solution. To circumvent this problem, prior work on single user and broadcast channels has specialized their results to the limiting high or low SNR regimes [28, 55], relied on numerical optimization [69], or resorted to characterizing the scaling of overhead with various system parameters based on sum-rate lower bounds [43]. To give accurate results on finite-SNR sum-rate, I propose to optimize a series expansion of (4.3) with respect to the channel's Doppler spread around the point $f_{\text{D}} = 0$ [37]. Recall that T_{frame} which I have been using thus far is related to f_{D} by the relationship $T_{\text{frame}} = \frac{1}{2 f_{\text{D}}}$. To that end, I give the following result on the series expansion of $\bar{R}_{\text{eff}}(P, T_{\text{OHD}})$.

Proposition 14. *The effective sum-rate achieved by IA with training and feedback expands as*

$$\begin{aligned} \bar{R}_{\text{eff}}(P, T_{\text{OHD}}) =& (1-\alpha)(1+\rho K d) \left[\frac{\bar{R}_{\text{sum}}(\rho)}{1+\rho K d} - \frac{2\beta}{d\alpha} \dot{R}_{\text{sum}}(\rho) f_{\text{D}} \right. \\ & \left. + \left(\frac{2\beta}{d\alpha} \right)^2 \left(\ddot{R}_{\text{sum}}(\rho)(1+\rho K d) + 2 K d \dot{R}_{\text{sum}}(\rho) \right) \frac{{f_{\text{D}}}^2}{2} \right] + O({f_{\text{D}}}^3), \end{aligned} \tag{4.31}$$

where

$$\beta = \frac{\left(\sqrt{KN_{\mathrm{T}}N_{\mathrm{R}}} + N_{\mathrm{R}} + \sqrt{\gamma N_{\mathrm{T}}(KN_{\mathrm{T}} - N_{\mathrm{R}})}\right)^2}{\gamma(KN_{\mathrm{T}} - N_{\mathrm{R}})}, \tag{4.32}$$

whereas $\dot{R}_{\mathrm{sum}}(\rho)$ *and* $\ddot{R}_{\mathrm{sum}}(\rho)$ *are the first and second derivatives of* perfect CSI sum-rate, $\bar{R}_{\mathrm{sum}}(\rho)$*, which can be conveniently expressed as*

$$\dot{R}_{\mathrm{sum}}(\rho) = \frac{1}{\rho}\left(Kd\log_2(e) - \frac{\bar{R}_{\mathrm{sum}}(\rho)}{\rho}\right), \tag{4.33}$$

$$\ddot{R}_{\mathrm{sum}}(\rho) = -\frac{1}{\rho^2}\left(Kd\log_2(e) + \dot{R}_{\mathrm{sum}}(\rho) - 2\frac{\bar{R}_{\mathrm{sum}}(\rho)}{\rho}\right). \tag{4.34}$$

Proof. Given in Appendix 4.9.1. □

Thus, by expanding effective sum-rate w.r.t. f_{D}, I have transformed the complicated non-linear dependence of effective sum-rate on system parameters such as P, T_{frame}, f_{D}, and T_{OHD} to a simpler polynomial dependence. The expansion in Proposition 14 can now be used to derive the expansion of the optimal overhead budget, $\alpha^\star$, along with the performance it achieves. Relaxing the constraint that the overhead fraction α must be rational, simply differentiating the series expansion in Proposition 14 and equating it to zero yields the optimal overhead budget $\alpha^\star$.

Proposition 15. *The optimum overhead fraction* $\alpha^\star$ *for an IA system with*

training and analog feedback expands as

$$\alpha^\star = \sqrt{\frac{2\beta(1+\rho K d)}{d}\frac{\dot{\bar{R}}_{\text{sum}}(\rho)}{\bar{R}_{\text{sum}}(\rho)}f_{\text{D}}} - \frac{\beta}{d}\left(\frac{\ddot{\bar{R}}_{\text{sum}}(\rho)}{\dot{\bar{R}}_{\text{sum}}(\rho)}(1+\rho K d)+2Kd\right)f_{\text{D}} + O({f_{\text{D}}}^{3/2}), \tag{4.35}$$

which results in the optimal effective sum-rate

$$\bar{R}^\star_{\text{eff}}(P) = \bar{R}_{\text{sum}}(\rho) - 2\sqrt{\frac{2\beta}{d}(1+\rho K d)\dot{\bar{R}}_{\text{sum}}(\rho)\bar{R}_{\text{sum}}(\rho)f_{\text{D}}} + O(f_{\text{D}}). \tag{4.36}$$

Note that if f_{D} is large enough that $\alpha^\star < \alpha_{\min}$ the optimal overhead budget must be adjusted to $\alpha_{\min}$ and the expression for $\bar{R}^\star_{\text{eff}}(P)$ correspondingly updated.

Proof. The proof follows directly from differentiating the expansion in Proposition 14 w.r.t. α and solving the resulting cubic polynomial for its relevant root. □

Therefore, Proposition 15 along with the solution to (4.28) gives the effective sum-rate-maximizing amount of forward training, feedback channel training, and analog feedback as simple functions of fundamental system parameters such as SNR, Doppler spread (equivalently T_{frame}), and perfect CSI sum-rate. Numerical results in Section 4.7 will show that the overhead expansion in Proposition 15 is accurate for a wide range of system parameters and can thus obviate the need for numerical overhead optimization. Furthermore, the derived results allow us to draw several interesting insights into IA system design and performance:

1. The optimal overhead budget α scales with $\sqrt{f_{\mathrm{D}}}$. As stated, for high enough Doppler $\alpha^{\star}$, must be adjusted to $\alpha_{\min}$ meaning that overhead subsequently increases with f_{D}. This scaling behavior is in line with previous results on other single and multiuser channels.

2. The sum-rate penalty due to overhead and imperfect CSI behaves similarly, i.e., increases with $\sqrt{f_{\mathrm{D}}}$ initially and with f_{D} at high Doppler.

3. Examining the leading term in $\alpha^{\star}$ we note that, similarly to [37], the term $(1+\rho K d)\frac{\dot{R}_{\mathrm{sum}}(\rho)}{R_{\mathrm{sum}}(\rho)}$ behaves like $Kd/\log_e(1+\rho)$ and thus the optimal overhead budget decreases with SNR roughly as $\sqrt{Kd/\log_e(1+\rho)}$.

4. Since overhead decreases with SNR, a minimum overhead interval of $KN_{\mathrm{T}}+KN_{\mathrm{R}}+K^2N_{\mathrm{T}}$ is always optimal at sufficiently high SNR. Thus, the effective number of spatial DoF achieved by IA with the analog feedback strategy described is $\left(1-\frac{KN_{\mathrm{T}}+KN_{\mathrm{R}}+K^2N_{\mathrm{T}}N_{\mathrm{R}}}{T_{\mathrm{frame}}}\right)Kd$, i.e., the DoF penalty increases linearly with f_{D}. Further, we note that this minimal training budget increases at least with K^2, and more realistically with K^3 if the linear dependence of N_{T} and N_{R} on K is accounted for.

5. Again examining the leading term in $\alpha^{\star}$ we note that it increases with $\sqrt{\beta}$. Recalling the definition of β in (4.32), I conclude that the optimal overhead budget increases with $\sqrt{P/P_{\mathrm{F}}}$. This formalizes the relationship between overhead and feedback link quality.

In addition to highlighting the dependence of overhead and effective sum-rate on various system parameters, the derived results can provide simple answers

to various network design questions. For example, by simply comparing IA's effective sum-rate expression to those achieved by other transmission strategies, one can choose the optimal transmission strategy for a given fading environment. Moreover, since overhead and channel selectivity have been shown to place fundamental limits on the gains of cooperation in wireless network [56], the overhead-aware analysis presented in this chapter can help in determining the optimal number of cooperative IA users at a given level of selectivity.

Consider, as a simple example, a K-user single-stream cooperation cluster with a variable number of antennas in which extra users are allowed to cooperate via IA if they do not incur a loss in effective sum-rate, else the extra users are not allowed access to the propagation medium and presumably left to transmit on a separate channel. In this model, additional cooperating users can be incorporated into the cluster as long as ${\mathcal{I}^{\star}}_{K+1}(P) - {\mathcal{I}^{\star}}_{K}(P) > 0$ where I have made cluster size explicit in the effective sum-rate subscript. Consequently, the effective sum-rate-maximizing cluster size becomes the smallest K such that ${\mathcal{I}^{\star}}_{K+1}(P) - {\mathcal{I}^{\star}}_{K}(P) < 0$. Moreover, note that minimizing overhead and maintaining IA feasibility imposes the constraint $N_{\mathrm{T}} + N_{\mathrm{R}} = K + 1$ [76]. Thus, writing N_{T} and N_{R} in terms of K, e.g. $N_{\mathrm{T}} = \lceil (K+1)/2 \rceil$, the user admission rule can be simplified to a function of only K, f_{D}, SNR, and γ. To simplify the user admission rule even further, I make the following approximations: (i) I consider the leading term in $\bar{R}_{\mathrm{eff}}(P, T_{\mathrm{OHD}})$ thus focusing on IA's effective DoF given in the fourth observation after Proposition 15, (ii) I assume that $N_{\mathrm{T}} = (K+1)/2$ and thus relax its integer constraint. Using these

simplifications, the user admission rule ${\mathcal{I}^{\star}}_{K+1}(P) - {\mathcal{I}^{\star}}_{K}(P) > 0$ simplifies to

$$4K^3 + 15K^2 + 17K + 6 < \frac{1}{f_{\mathrm{D}}}, \tag{4.37}$$

i.e., a K-user cluster can be extended to $K+1$ as long as (4.37) is satisfied. Interestingly, this implies that in such single-stream IA scenarios the effective sum-rate-maximizing cluster size grows with ${f_{\mathrm{D}}}^{-1/3}$. While the approximate admission rule is a rather simplified version of ${\mathcal{I}^{\star}}_{K+1}(P) - {\mathcal{I}^{\star}}_{K}(P) > 0$, I show in Section 4.7 that it is very accurate at predicting optimal cluster size. Finally, note that while I provide this example to illustrate problems that can be solved using my analysis, the rule in (4.37) is by no means universal. When parameters such as large-scale fading or uncoordinated interference are considered, both the analysis and the admission rule must be adjusted.

4.7 Simulation Results

Consider a three-user IA cluster with two transmit antennas, two receive antennas, and one spatial stream per user and let $\gamma = \frac{P_{\mathrm{F}}}{P} = 1$. Fig. 4.2 shows the effective sum-rate achieved by IA in systems with various levels of mobility or normalized Doppler spreads, f_{D}. To quantify the degradation in effective sum-rate caused by overhead and imperfect CSI, I include the performance of a baseline genie-aided system in which CSI is both perfect and free. Fig. 4.2 indicates that IA achieves good performance in a system with vehicular-levels of mobility. In fact, if typical wireless parameters are adopted, such as a wavelength of $\lambda = 0.15\ m$ (corresponding to a carrier frequency of

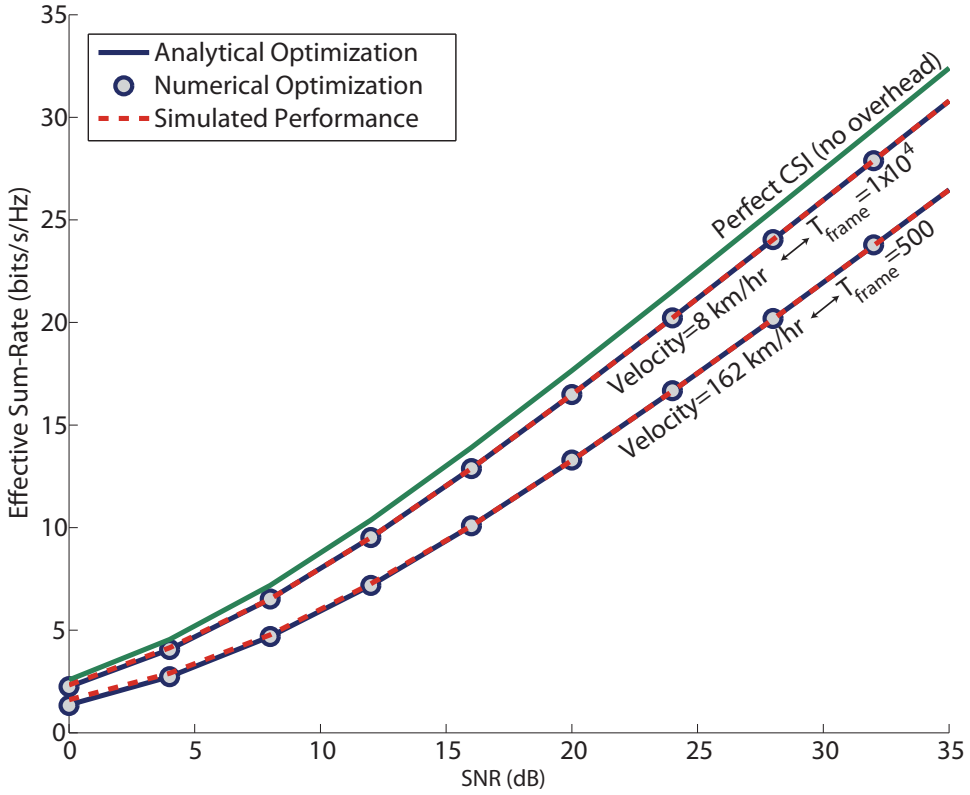

Figure 4.2: Effective Sum-Rate vs. SNR for systems with different normalized Doppler spreads. This quantifies the loss in sum-rate due to both imperfect CSI and overhead and shows that the performance predicted by the analytical results presented is an accurate representation of optimal performance.

2 GHz), a coherence bandwidth of $W_{\mathrm{C}} = 300\ kHz$, and a normalized Doppler given by $f_{\mathrm{D}} = \frac{v}{\lambda W_{\mathrm{C}}}$ where v is the user's velocity, Fig. 4.2 indicates that IA could theoretically perform well even at a speed of more than 160 km/hr. The rate of performance degradation over a wider range of Doppler spread can be seen in Fig. 4.3. Both Figs. 4.2 and 4.3 indicate that the analytical results of Section 4.6 are very effective in optimizing the effective sum-rate of IA systems as the resulting performance closely matches that of a numerically optimized

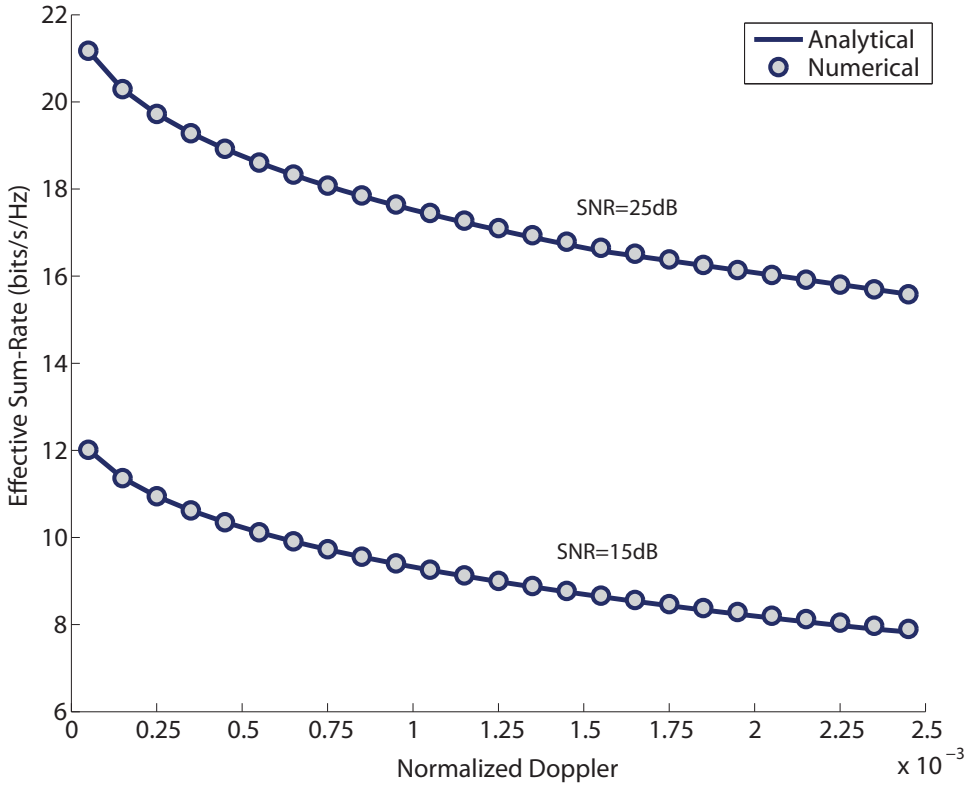

Figure 4.3: Effective Sum-Rate vs. Normalized Doppler for IA systems at different SNR levels. This quantifies the degradation in sum-rate as mobility increases resulting in an increased overhead penalty.

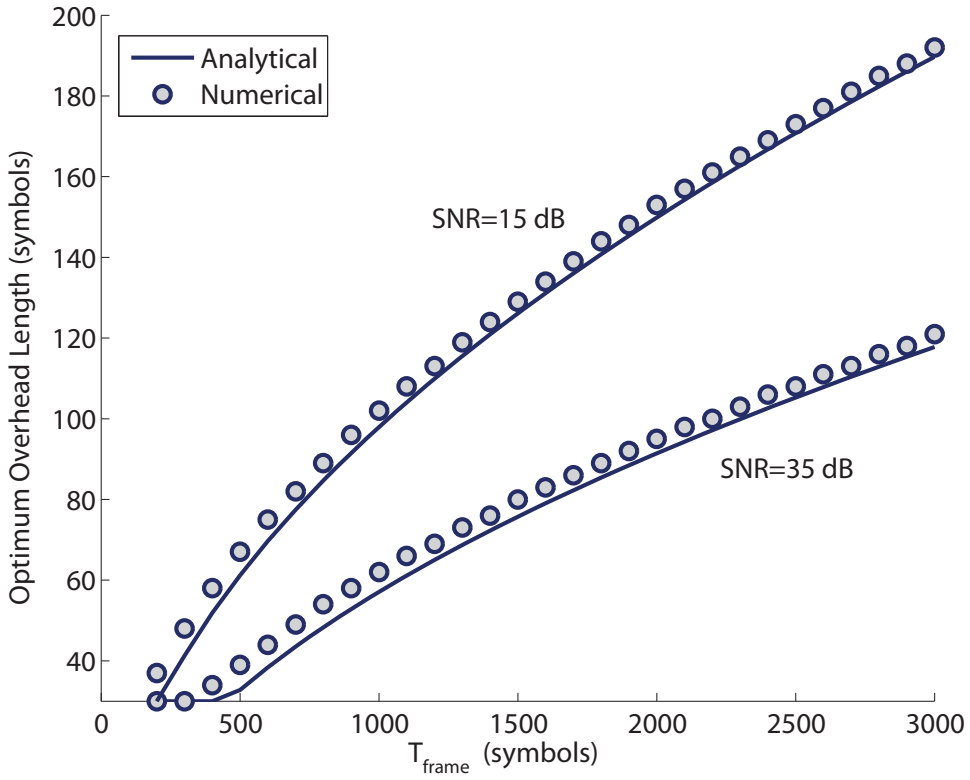

Figure 4.4: T_{OHD} vs. T_{frame}. This confirms that the optimal value of T_{OHD} scales with $\sqrt{T_{\mathrm{frame}}}$ as predicted, and shows that optimizing a series expansion of the objective yields remarkably accurate results.

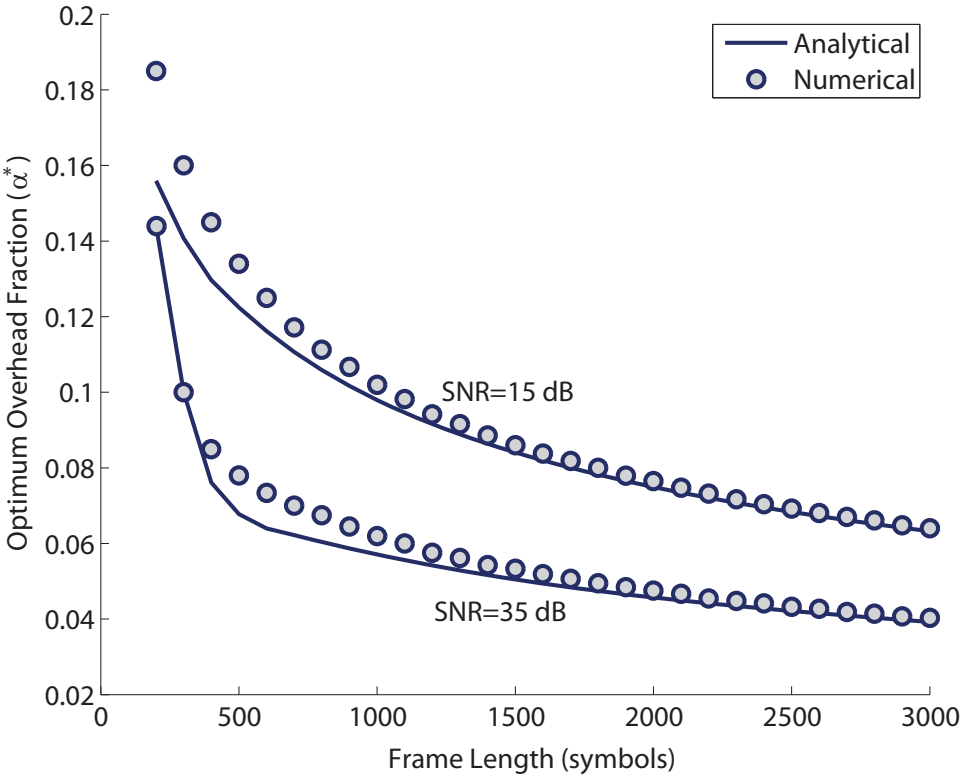

Figure 4.5: $\alpha^\star$ vs. T_{frame}. This confirms that the optimal value of α scales with $\frac{1}{\sqrt{T_{\text{frame}}}}$ and thus scales with $\sqrt{f_{\text{D}}}$ as predicted.

system. Finally, Fig. 4.2 indicates that the effect of the simplifying assumptions made in Section 4.5 is negligible since the effective sum-rate predicted by the derived rate expressions closely matches simulated IA performance. A very slight deviation is noticed at very low SNR where the ZF simplification in Section 4.5 is a less accurate approximation of MMSE performance.

Fig. 4.4 shows the optimal overhead budget for systems with varying frame lengths and again includes both the analytical overhead budget from Section 4.6 as well as the result of numerically optimizing the same system.

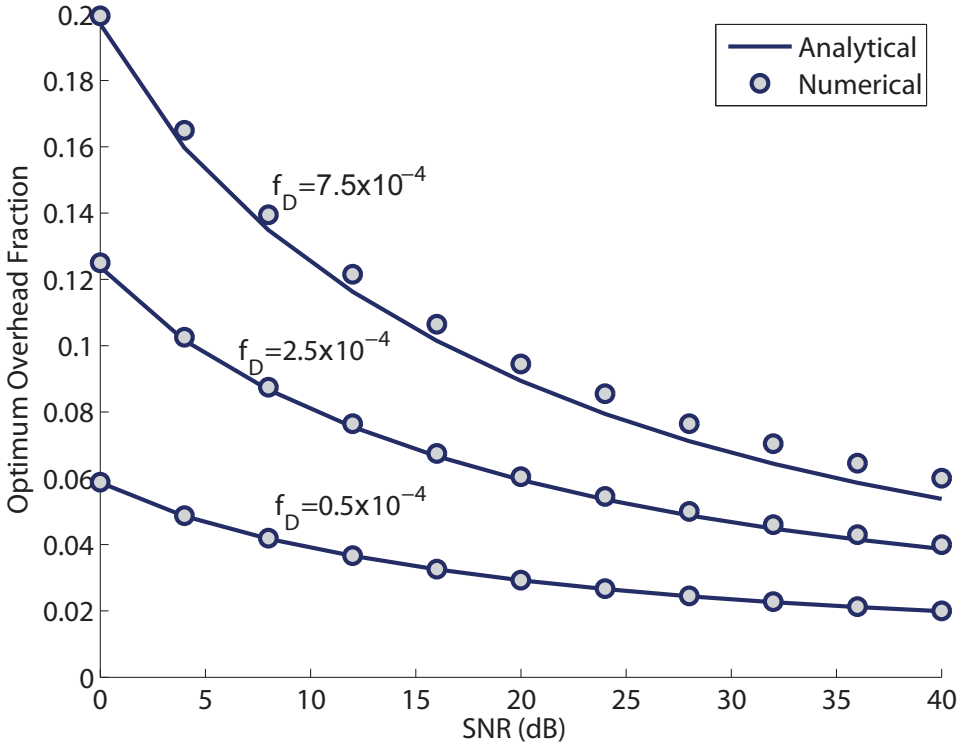

Figure 4.6: $\alpha^\star$ vs. SNR. This shows the decrease of the optimal overhead budget with SNR. As stated in Section 4.6, it can be shown that the decrease is logarithmic with SNR. The figure also demonstrates that my expansion-based results are very accurate, deviating only slightly in high-SNR high-mobility scenarios.

Fig. 4.4 confirms that T_{OHD} increases with frame size T_{frame} at a rate proportional to $\sqrt{T_{\text{frame}}}$. Thus $\alpha^\star$ indeed decreases with $\frac{1}{\sqrt{T_{\text{frame}}}}$, as shown in Fig. 4.5, or equivalently increases with $\sqrt{f_{\text{D}}}$ (and with f_{D} for sufficiently high Doppler). Fig. 4.4 also shows that the expansion in Proposition 15 provides an accurate characterization of IA's effective sum-rate-maximizing overhead budget over a wide range of SNRs and frame sizes. Fig. 4.6 in turn verifies the decrease of $\alpha^\star$ with SNR, which as stated in Section 4.6 follows the relationship

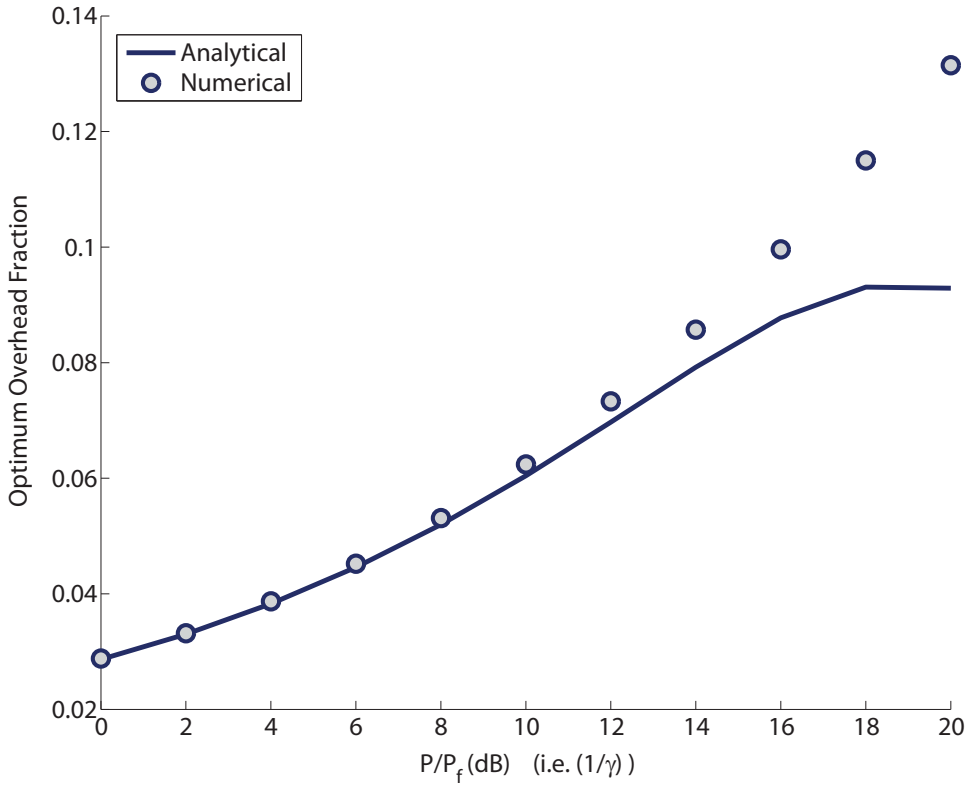

Figure 4.7: This figure shows the relationship between $\alpha^\star$ and feedback channel quality for a system with $T_{\text{frame}} = 10^4$. The figure verifies the increase of overhead with $1/\gamma$, and when plot in linear scale the square root rate of increase can be verified.

$\alpha^\star \sim \sqrt{\log_e(1+\rho)^{-1}}$. To complete the characterization of overhead and effective sum-rate, Fig. 4.7 quantifies the deleterious effect of a weak feedback channel on overhead and effective sum-rate. Fig. 4.7 also indicates that the expansion results of Section 4.6 could significantly underestimate $\alpha^\star$ in very weak feedback channels, though the final effect on throughput in Fig. 4.8 remains limited.

Finally I examine the efficiency of the overhead analysis in further net-

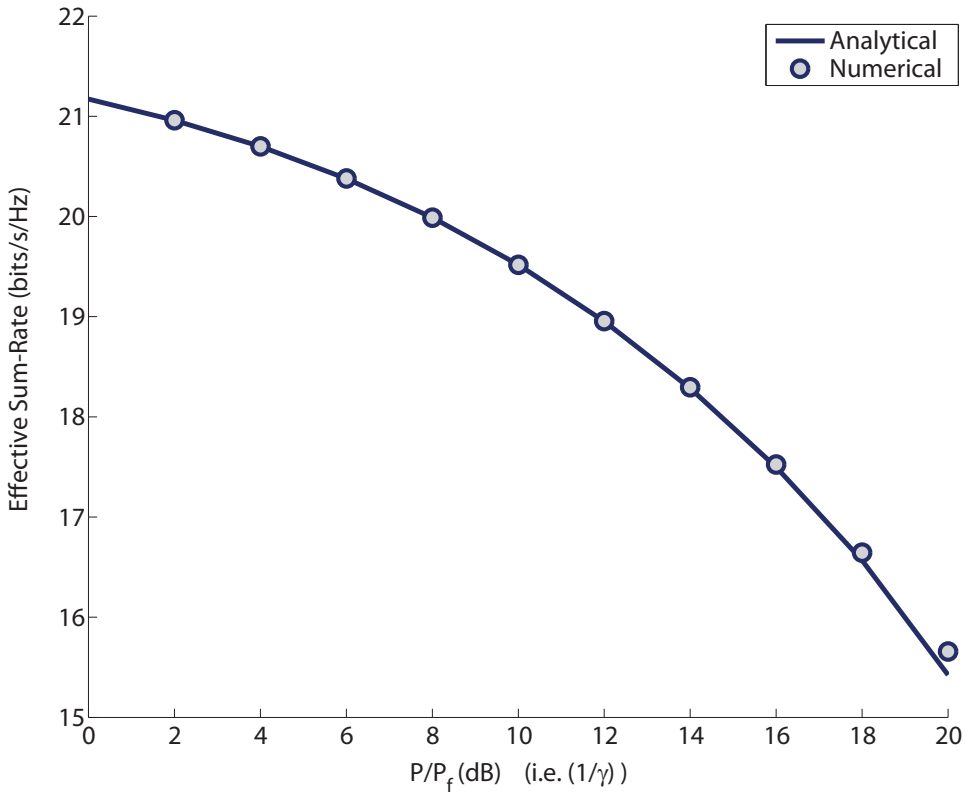

Figure 4.8: This figure shows the relationship between $\bar{R}^{\star}_{\text{eff}}(P)$ and feedback channel quality for a system with $T_{\text{frame}} = 10^4$. The figure verifies the rate of decrease of optimal effective sum-rate with feedback link quality.

work design. I consider the motivating example given in Section 4.6 of a K-user system for which I seek to optimize the cooperation cluster size as a function of mobility. Fig. 4.9 shows the optimal cluster size as a function of T_{frame} for an IA system at 35 dB SNR. Fig 4.10 shows the corresponding effective sum-rate achieved. I plot the cluster size and effective sum-rate resulting from (i) an exhaustive search over all possible K, and (ii) the simple overhead-based user admission rule in (4.37). Note that the cluster size predicted by the two

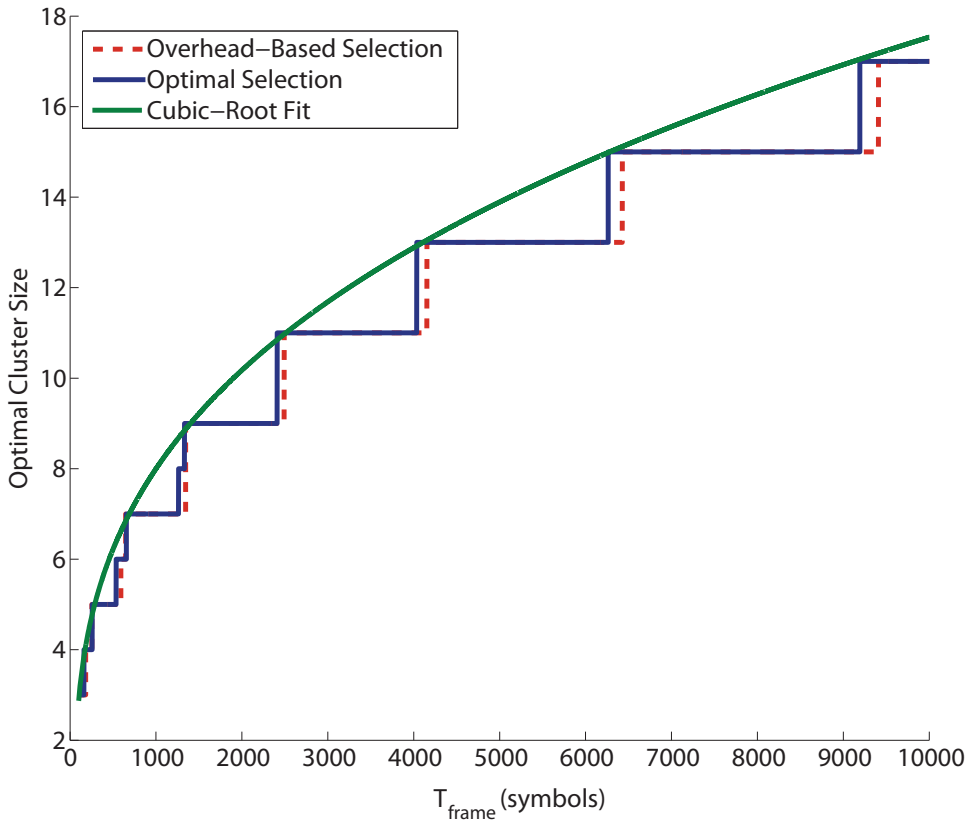

Figure 4.9: Optimal Cluster Size vs. T_{frame}. This shows the optimal number of users to coordinate via IA which increases channels coherence time. This also shows that comparing overhead, i.e., overhead based selection, provides accurate decisions on optimal cluster size.

methods are in close agreement, and that the asymptotic cube-root relationship predicted in Section 4.6 between optimal cluster size and T_{frame} is quite accurate even for small values of K. While the overhead-based rule tends to underestimate cluster size for small intervals of T_{frame}, Fig. 4.10 indicates that the resulting rate gap from optimal sizing is negligible. The same can be said about the rate loss when applying the same overhead based rule to a system

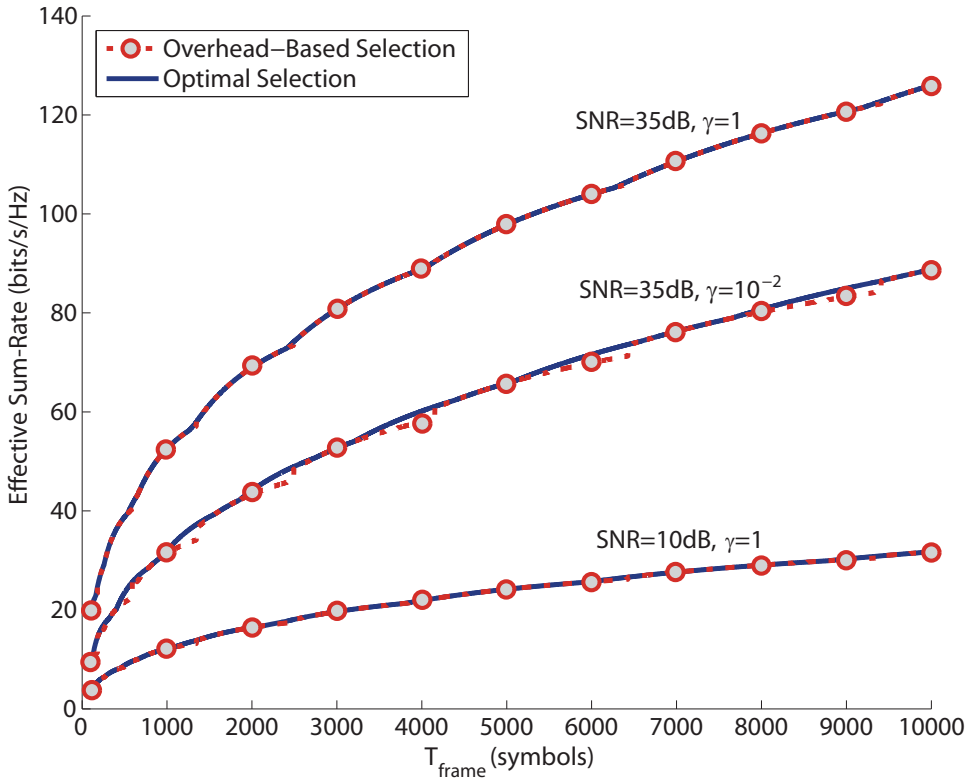

Figure 4.10: Effective Sum-Rate with Cluster Size Optimization vs. T_{frame}. This shows the increase in effective sum-rate as a function of T_{frame} when the cluster size is chosen to maximize rate. This also quantifies the minimal sum-rate loss due to sub-optimal overhead-only based cluster sizing.

at an SNR of 10 dB and a system with $\gamma = 10^{-2}$. Further, examining Fig. 4.11 which plots the effective data rate achieved per user, we notice that per-user rates are not affected significantly by varying the size of the IA cluster. As expected, per-user data rates increase with frame length when network size is fixed (due to better CSI), drop when networks increase in size (due to additional overhead), and then recover as frame length continues to grow. Further, we note that per-user rates do not exhibit a decaying trend with in-

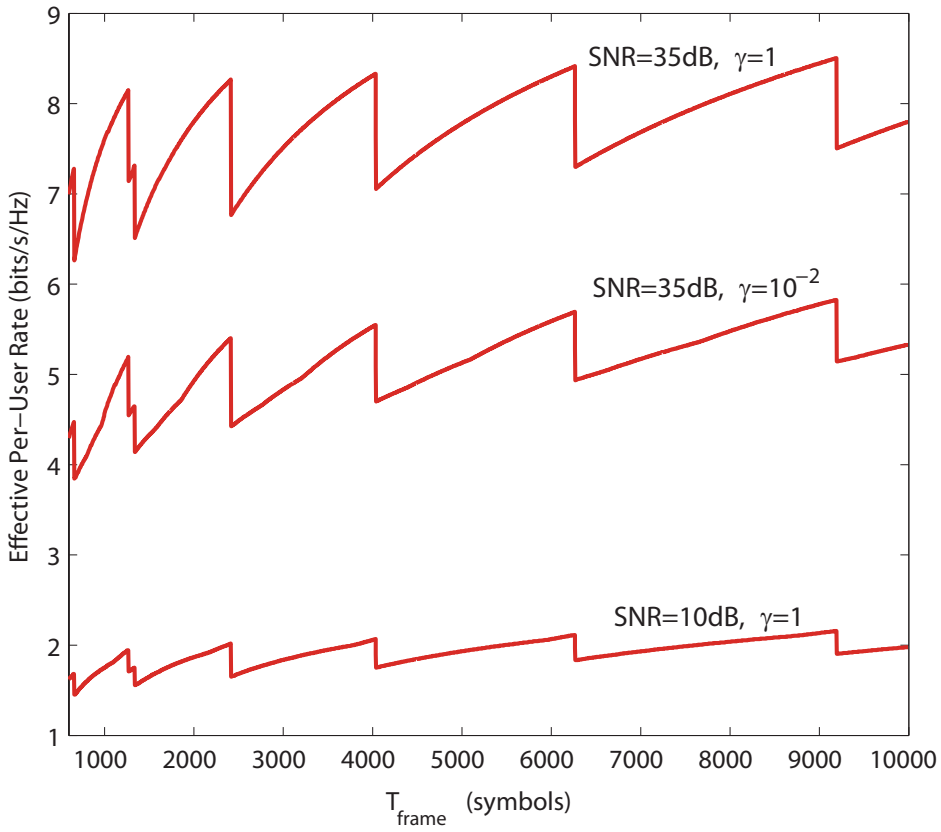

Figure 4.11: Effective Per-User Rate with Cluster Size Optimization vs. T_{frame}. This shows how per-user rate is affected by overhead-aware network sizing. Interestingly, per-user rate does not decay to zero as the size of the IA cluster increases (as frame size increases), provided that training and feedback overhead is properly factored into the choice of IA cluster size.

creasing network size, provided that CSI acquisition overhead is factored into the network-sizing decision. As a result, IA networks and their corresponding effective sum-rate can be allowed to grow without sacrificing the data rates achieved by individual users.

4.8 Conclusion

I considered IA's effective sum-rate in practical systems where CSI is imperfect and comes with an associated overhead cost. I showed that training and feedback overhead can be optimized to ensure good IA performance over a wide range of SNR and Doppler spread. I quantified the dependence between overhead and various system parameters such as feedback link quality. More sophisticated precoding algorithms, designed to be robust to imperfect CSI, could further improve the demonstrated performance and thus remain a promising area for future work. The derived results provide a formal method to gauge true IA performance vs. other transmission strategies, and can thus highlight settings under which IA provides tangible gains. The derived analysis can also be used for further network design as demonstrated by the motivating example given at the end of Section 4.6 on overhead-aware user admission and optimal network sizing.

4.9 Appendix

4.9.1 Proof of Proposition 14

To expand effective sum-rate around $f_{\mathrm{D}} = 0$, I start computing its first order derivative

$$\frac{\partial \bar{R}_{\mathrm{eff}}(P, T_{\mathrm{OHD}})}{\partial f_{\mathrm{D}}} = (1-\alpha)\dot{R}_{\mathrm{sum}}(\rho)\frac{\partial \rho_{\mathrm{eff}}}{\partial f_{\mathrm{D}}}|_{f_{\mathrm{D}}=0} = -(1-\alpha)\dot{R}_{\mathrm{sum}}(\rho)(1+Kd\rho)\frac{2\beta}{d\alpha} \tag{4.38}$$

where $\frac{\partial \rho_{\mathrm{eff}}}{\partial f_{\mathrm{D}}}|_{f_{\mathrm{D}}=0}$ is evaluated by noticing that after solving the inner maximization in (4.28) and obtaining $\sigma^{2\,\star}_{\tilde{\mathbf{H}}}$ in (4.30) we have $\frac{\partial \sigma^2_{\tilde{\mathbf{H}}}}{\partial f_{\mathrm{D}}}|_{f_{\mathrm{D}}=0} = \frac{2\beta}{d\alpha}$. The

term $\dot{R}_{\text{sum}}(\rho)$ can be obtained by a standard derivation of the exponential integral rate expression in (4.8) w.r.t ρ and is given directly in the statement of Proposition 14; $\ddot{R}_{\text{sum}}(\rho)$ is obtained similarly. As for the second order term, we have

$$\begin{aligned}
\frac{\partial^2 \bar{R}_{\text{eff}}(P, T_{\text{OHD}})}{\partial {f_{\text{D}}}^2} &= (1-\alpha)\left[\ddot{R}_{\text{sum}}(\rho)\left(\frac{\partial \rho_{\text{eff}}}{\partial f_{\text{D}}}\right)^2 + \dot{R}_{\text{sum}}(\rho)\frac{\partial^2 \rho_{\text{eff}}}{\partial {f_{\text{D}}}^2}\right]|_{f_{\text{D}}=0} \\
&\overset{(a)}{=} (1-\alpha)\left[\ddot{R}_{\text{sum}}(\rho)\left(\frac{\partial \rho_{\text{eff}}}{\partial f_{\text{D}}}\right)^2 \right. \qquad (4.39) \\
&\qquad \left. + \dot{R}_{\text{sum}}(\rho)\left(\frac{\partial^2 \rho_{\text{eff}}}{\partial {\sigma^2_{\widetilde{\mathbf{H}}}}^{\star 2}}\left(\frac{\partial {\sigma^2_{\widetilde{\mathbf{H}}}}^{\star}}{\partial f_{\text{D}}}\right)^2 + \frac{\partial \rho_{\text{eff}}}{\partial {\sigma^2_{\widetilde{\mathbf{H}}}}^{\star}}\frac{\partial^2 {\sigma^2_{\widetilde{\mathbf{H}}}}^{\star}}{\partial {f_{\text{D}}}^2}\right)\right]|_{f_{\text{D}}=0} \\
&\overset{(b)}{=} (1-\alpha)\left(\frac{2\beta}{d\alpha}\right)^2 (1+\rho K d) \\
&\qquad \times \left(\ddot{R}_{\text{sum}}(\rho)(1+\rho K d) + 2Kd\dot{R}_{\text{sum}}(\rho)\right) \qquad (4.40)
\end{aligned}$$

where (a) expands $\frac{\partial^2 \rho_{\text{eff}}}{\partial {f_{\text{D}}}^2}$ for clarity and (b) is by noticing that $\frac{\partial^2 {\sigma^2_{\widetilde{\mathbf{H}}}}^{\star}}{\partial {f_{\text{D}}}^2} = 0$ since ${\sigma^2_{\widetilde{\mathbf{H}}}}^{\star}$ is linear in f_{D} and otherwise replacing the values of the different variables. Combining (4.38) and (4.40) I get the resulting second order expansion. Higher order expansions can be found if additional accuracy is needed, however, the second order expansion is in general sufficient.

Chapter 5

Conclusion

5.1 Summary

In this dissertation, I developed methods for providing interference alignment systems with accurate channel knowledge. The proposed solutions enable practical systems to translate alignment's theoretical degrees-of-freedom optimality to tangible gains in spectral efficiency. In my first contribution, I leveraged the concept of analog feedback to develop a low-complexity CSI acquisition strategy for the MIMO interference channel. In my second contribution, I leveraged temporal correlation and manifold structure to compress CSI in single antenna IA systems. In my last contribution, I showed that the proposed feedback strategies not only provide accurate enough channel knowledge, but do so at a very reasonable overhead cost. The contributions presented in this dissertation constitute a strong set of candidate feedback solutions for future wireless systems that seek to reduce the effect of inter-user interference.

In my initial work, I developed a feedback strategy for the multiple-input multiple-output frequency-flat interference channel. The proposed feedback strategy is a two step process, (i) pilot transmission over the reverse

channel to enable coherent decoding of feedback data, and (ii) the feedback of forward channel matrices in an analog fashion. As opposed to the concept of limited feedback in which channel states are quantized before transmission, analog feedback directly transmits the channel matrix elements as uncoded quadrature and amplitude modulated symbols. Bypassing quantization allows systems to avoid the large codebooks needed in limited feedback solutions and thus avoid the exponential search complexity that comes with those codebooks. I characterized the performance of IA with analog feedback and showed that the degrees-of-freedom achieved with perfect channel knowledge are preserved by analog feedback. In fact, I showed that the spectral efficiency gap between genie-aided systems with perfect CSI and practical systems with analog feedback is bounded by a constant when signal-to-noise ratio is comparable in both forward and feedback channels, i.e., when forward and reverse-link SNRs are equal up to a constant factor (or constant additive gap in the dB-scale). When signal-to-noise ratios are not quite symmetric, a fraction of the degrees-of-freedom are still achieved.

To improve the efficiency and lower the complexity of limited feedback, which in theory could outperform the analog feedback concept, I proposed a Grassmannian differential feedback framework for the single-antenna frequency-selective interference channel. The proposed framework exploits both the CSI's Grassmannian structure and the channel's temporal correlation to improve quantization accuracy and dramatically reduce the required quantization codebook size. I designed quantization codebooks to be used with the

proposed differential feedback solution. The constructed codebooks have the ability to adapt to the long-term dynamics of time-varying channels; a capability which ultimately translates to better quantization accuracy. I analytically characterized the performance of differential feedback and investigated its dependence on the length of the channel's impulse response, user mobility, quantization codebook structure, and the number of feedback bits. The analytical performance characterization includes results on average quantizer distortion, and more importantly, on the sum-rate achieved by interference alignment systems with differential feedback.

To complete the treatment of interference alignment systems with practical methods of CSI acquisition, I revisit the analog feedback strategy developed and more formally investigate network performance when CSI acquisition overhead is explicitly accounted for. First, I augment the analog feedback strategy developed earlier in the dissertation to include forward channel training. Further, instead of using the lower bounds on sum-rate developed earlier, I derive more accurate and tractable expressions for IA sum-rate in both genie-aided systems and systems with using analog feedback. I then analytically characterize *effective sum-rate with overhead* in relation to various parameters such as signal-to-noise ratio, Doppler spread, and feedback channel quality. I show that by using the analog feedback framework developed in this dissertation, and by optimizing its various parameters, IA's information theoretic optimality can be translated into practical system gains in a wide range of fading scenarios even when overhead is considered. Further, I demonstrate the

utility of my proposed overhead analysis in solving practical system/network design problems. As an example of overhead-aware network design, I solve the problem of finding the optimal number of cooperative IA users as a function of system parameters such as signal power and mobility.

5.2 Future Research Directions

In this section I present some promising areas for future research on interference alignment and practical communication strategies for interference channels in general.

1. **Improved Feedback Strategies**: The analog feedback strategy developed in this dissertation can elegantly handle MIMO interference alignment and do so with sufficient accuracy and limited complexity. Rate-distortion results from information theory, however, indicate that (in some cases of interest) a properly designed digital scheme will ultimately outperform analog feedback. Thus, practical limited feedback strategies remain of interest. Existing strategies such as [7] can be generalized to better support MIMO IA and then improved upon to decrease complexity. Further, the development of limited feedback strategies can benefit from the lessons learned in the MIMO single user and broadcast channels where feedback can be significantly compressed by exploiting the mathematical properties of the quantized CSI. While manifold structure can be simply highlighted in point-to-point and broadcast channels, no equivalent results exist for the interference channel. The work in [79] succeeds

in highlighting some mathematical invariance properties in the alignment problem, however, [79] does not claim to have identified the *minimal set of channel parameters* needed for interference alignment. Another interesting, and somewhat unexplored, avenue for further feedback reduction in MIMO interference channels has recently been highlighted in [15, 16]. The authors in this paper challenge the common assumption that cooperating users should share global (network-wide) CSI and revisit this CSI dissemination process in MIMO interference channels using either IA or joint-transmission (also known as network MIMO). They find that in many fully-connected interference channels, global channel knowledge is in fact not needed at the transmitter and thus degrees-of-freedom may be preserved not only with imperfect CSI, but also with *incomplete CSI*.

2. **Interference Alignment with Delayed CSI**: The results of this dissertation prove that replacing perfect CSI with imperfect CSI leads to limited performance losses. However, the results in this dissertation have neglected delays in CSI reporting. Delayed CSI is traditionally considered an obstacle that prohibits cooperative transmission. Recent results on alignment-based coding strategies, however, indicate that our perspective of delayed CSI is overly pessimistic [19, 22, 49, 59, 104–106]. Namely, [59] showed that completely outdated CSI can still be very useful (in a degrees-of-freedom sense) in the wireless broadcast channel. More recent results indicate that, if the CSI delay is only a fraction of the channel's coherence time, delay causes no loss in degrees-of-freedom

for both broadcast and interference channels [49]. Interestingly, while delayed CSI coding strategies are not identical to IA in channels with instantaneous CSI, the concepts of interference alignment and neutralization play a central role in all delayed-CSI solutions. However, since these coding strategies utilize CSI in a different method than the IA strategies considered in this dissertation, they present new problems in feedback design. For example, in strategies such as [22], which use a combination of instantaneous and delayed CSI to cancel interference, one can ask the following questions: What is the performance of delayed CSI strategies when said CSI is also imperfect? Which should be quantized with higher fidelity, current or past CSI? Can coding strategies for channels with delayed CSI be extended to systems with heterogeneous CSI, i.e. where information about different channels is available at different times for different users? The heterogeneous CSI setup is practically motivated by systems that acquire knowledge about direct channels via training and reciprocity, yet can only acquire delayed information about the other user's channels via explicit CSI sharing over a delay-limited backhaul. The problem of examining networks with such heterogeneously delayed CSI has been explored in the context of queuing systems [77] but, to the best of my knowledge, has not been considered from the perspective of physical layer transmission strategies.

3. **IA in Multihop Interference Channels**: Another promising direction for future work on interference alignment is its applications to the

multihop interference channels, i.e., interference channels with relays. The importance of practical precoding algorithms and CSI acquisition strategies for the relay aided interference channel is further amplified by the standards community's growing interest in deploying relays in future wireless systems. Preliminary work on multihop IA suggests that relays can greatly reduce the coding dimensions needed to achieve a network's maximum degrees-of-freedom. In some cases, an exponentially increasing need for coding dimensions is reduced to a linear or constant requirement. As a result, in many cases, relays can simplify the degrees-of-freedom optimal transmission strategies [35]. The key ingredient that enables such efficient interference cancellation in multihop systems is the existence of multiple propagation paths between each transmitter and receiver. This multiplicity of paths can be leveraged, along with relay processing, to realize over-the-air interference neutralization [24, 50, 103]. While the existing work on such systems has succeeded in highlighting the potential of relays, there are a number of assumptions that must be relaxed before the proposed solutions can be used in practice. Some of the main assumptions include full duplex operation, global CSI at the relay, and perhaps the most pressing assumption is the reliance on the cascaded interference channel model in which interference between the different full-duplex relays is neglected.

4. **IA in Modern Cellular Networks**: To make IA a viable multi-cell cooperation technique, IA research must characterize the effect of modern

cellular system complexities such as scheduling, resource allocation, data traffic modeling, and the possibility of device-to-device communication. IA must also be re-evaluated using models that more accurately resemble modern cellular systems which could include heterogeneous infrastructure such as macrocells, picocells, small cells, relays, and distributed antennas. Heterogeneous architectures may hold both great promise and challenges for interference alignment. For example, early work on IA in cellular systems shows lackluster gains for cell edge users due to low cell-edge SNR. Recall that IA is, by design, a high-SNR cooperation strategy. Heterogeneity, however, typically seeks to boost SNR at the cell-edge thus potentially making IA more applicable. My preliminary work on IA in distributed antenna systems (DAS), for example, indicates that DAS may significantly increase the utility of IA in cellular systems [94, 95]. Applying IA to DAS, however, requires improved precoding and CSI acquisition strategies that account for the heterogeneous pathloss in DAS propagation channels. The work in [94, 95] establishes feasibility conditions that demonstrate the possibility of IA in DAS, as well as develops simple algorithms for DAS-IA precoding. The algorithms proposed in [94, 95], however, can be improved upon by using more sophisticated constrained optimization techniques.

Bibliography

[1] Interference alignment bibliography. *www.profheath.org/research/interference-alignment/interference-alignment-bibliography/*, 2012.

[2] TA Arias, A. Edelman, and ST Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl*, 20:303–353, 1998.

[3] Alexei Ashikhmin and Ravikiran Gopalan. Grassmannian packings for efficient quantization in MIMO broadcast systems. *in Proc. of IEEE International Symposium on Information Theory*, pages 1811 –1815, Jun. 2007.

[4] K.E. Baddour and N.C. Beaulieu. Autoregressive modeling for fading channel simulation. *IEEE Transactions on Wireless Communications*, 4(4):1650–1662, Jul. 2005.

[5] R. Bhagavatula, Robert W. Heath, Jr., and B. Rao. Limited feedback with joint CSI quantization for multicell cooperative generalized eigenvector beamforming. *in Proc. of IEEE International Conference on Acoustics Speech and Signal Processing*, pages 2838–2841, Mar. 2010.

[6] R.S. Blum. MIMO capacity with interference. *IEEE Journal on Selected Areas in Communications*, 21(5):793–801, Jun. 2003.

[7] H. Bölcskei and I.J. Thukral. Interference alignment with limited feedback. *in Proc. of IEEE International Symposium on Information Theory*, pages 1759 –1763, Jul. 2009.

[8] V. Cadambe and S. Jafar. Interference Alignment and Degrees of Freedom of the K-User Interference Channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, Aug. 2008.

[9] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Transactions on Information Theory*, 56(6):2845–2866, Jun. 2010.

[10] G. Caire, N. Jindal, and S. Shamai. On the required accuracy of transmitter channel state information in multiple antenna broadcast channels. *Proc. of the 41st Asilomar Conference on Signals, Systems and Computers*, pages 287–291, Nov. 2007.

[11] Chan-Byoung Chae, D. Mazzarese, T. Inoue, and Robert W. Heath, Jr. Coordinated Beamforming for the Multiuser MIMO Broadcast Channel With Limited Feedforward. *IEEE Transactions on Signal Processing*, 56(12):6044–6056, Dec. 2008.

[12] Sang Won Choi, S.A. Jafar, and Sae-Young Chung. On the beamforming design for efficient interference alignment. *IEEE Communications Letters*, 13(11):847–849, Nov. 2009.

[13] Kuo-Liang Chung and Wen-Ming Yan. The complex Householder transform. *IEEE Transactions on Signal Processing*, 45(9):2374–2376, Sep. 1997.

[14] R. H. Clarke. A statistical theory of mobile-radio reception. *Bell System Technical Journal*, 47(6):957–1000, 1968.

[15] Paul de Kerret and David Gesbert. Interference alignment with incomplete CSIT sharing. *arXiv preprint arXiv:1211.5380*, 2012.

[16] Paul de Kerret and David Gesbert. Csi sharing strategies for transmitter cooperation in wireless networks. *arXiv preprint arXiv:1302.3166*, 2013.

[17] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu. Multi-cell mimo cooperative networks: A new look at interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, 2010.

[18] D. Gesbert and M. Kountouris. Rate scaling laws in multicell networks under distributed power control and user scheduling. *IEEE Transactions on Information Theory*, 57(1):234–244, Jan. 2011.

[19] A. Ghasemi, A.S. Motahari, and A.K. Khandani. Interference alignment for the MIMO interference channel with delayed local CSIT. *arXiv preprint arXiv:1102.5673*, 2011.

[20] A. Goldsmith. *Wireless communications*. Cambridge Univ Pr, 2005.

[21] K. Gomadam, V.R. Cadambe, and S.A. Jafar. A distributed numerical approach to interference alignment and applications to wireless interference networks. *IEEE Transactions on Information Theory*, 57(6):3309–3322, 2011.

[22] T. Gou and S.A. Jafar. Optimal use of current and outdated channel state information: Degrees of freedom of the miso bc with mixed CSIT. *IEEE Communications Letters*, 16(7):1084–1087, 2012.

[23] Tiangao Gou and S.A. Jafar. Degrees of freedom of the K user MxN MIMO interference channel. *IEEE Transactions on Information Theory*, 56(12):6040–6057, Dec. 2010.

[24] Tiangao Gou, Syed A Jafar, Sang-Woon Jeon, and Sae-Young Chung. Aligned interference neutralization and the degrees of freedom of the 2× 2× 2 interference channel. *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2751–2755, 2011.

[25] R. Gray. Vector quantization. *IEEE ASSP Magazine*, 1(2):4–29, April 1984.

[26] M. Guillaud and D. Gesbert. Interference alignment in the partially connected K-user MIMO interference channel. *Proc. of European Signal Processing Conference (EUSIPCO), Barcelona, Spain*, pages 1–5, Sept. 2011.

[27] M. Guillaud, D.T.M. Slock, and R. Knopp. A practical method for wireless channel reciprocity exploitation through relative calibration. *Proc. of the 8th Int. Symp. on Signal Processing and Its Applications*, pages 403 – 406, Aug. 2005.

[28] B. Hassibi and B.M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, Apr. 2003.

[29] Robert W. Heath, Jr., Tao Wu, and A.C.K. Soong. Progressive refinement for high resolution limited feedback multiuser MIMO beamforming. *in Proc. of 42nd Asilomar Conference on Signals, Systems and Computers*, pages 743–747, Oct. 2008.

[30] M. Honig, U. Madhow, and S. Verdu. Blind adaptive multiuser detection. *IEEE Transactions on Information Theory*, 41(4):944–960, Jul. 1995.

[31] C Huang, S A Jafar, S Shamai, and S Vishwanath. On degrees of freedom region of MIMO networks without CSIT. *2009. [Online]. Available*, page http//wwwcitebaseorg/abstract?id=oaiarXivorg090940.

[32] K. Huang, R.W. Heath, and J.G. Andrews. Limited feedback beamforming over temporally-correlated channels. *IEEE Transactions on Signal Processing*, 57(5):1959–1975, 2009.

[33] Kaibin Huang, Robert W. Heath, Jr., and J.G. Andrews. Limited feedback beamforming over temporally-correlated channels. *IEEE Transactions on Signal Processing*, 57(5):1959–1975, May 2009.

[34] Takao Inoue and Robert W. Heath, Jr. Grassmannian predictive coding for limited feedback in multiple antenna wireless systems. *submitted to IEEE Transactions on Signal Processing*, 2011.

[35] S.A. Jafar. *Interference Alignment-A New Look at Signal Dimensions in a Communication Network*. Now Publications, 2011.

[36] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Transactions on Information Theory*, 52(11):5045 –5060, Nov. 2006.

[37] N. Jindal and A. Lozano. A unified treatment of optimum pilot overhead in multipath fading channels. *IEEE Transactions on Communications*, 58(10):2939–2948, 2010.

[38] V. Jungnickel, U. Kruger, G. Istoc, T. Haustein, and C. von Helmolt. A MIMO system with reciprocal transceivers for the time-division duplex mode. *Proc. of IEEE Antennas and Propagation Society International Symposium*, 2:1267–1270, Jun. 2004.

[39] Y.H. Kho and D.P. Taylor. MIMO channel estimation and tracking based on polynomial prediction with application to equalization. *IEEE Transactions on Vehicular Technology*, 57(3):1585–1595, May 2008.

[40] D. Kim and M. Torlak. Optimization of interference alignment beamforming vectors. *IEEE Journal on Selected Areas in Communications*, 28(9):1425–1434, Dec. 2010.

[41] Kyeongyeon Kim, Il Han Kim, and D.J. Love. Utilizing temporal correlation in multiuser MIMO feedback. *in Proc. of 42nd Asilomar Conference on Signals, Systems and Computers*, pages 121–125, Oct. 2008.

[42] Taejoon Kim, D.J. Love, and B. Clerckx. Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems? *IEEE Transactions on Signal Processing*, 59(4):1654–1669, Apr. 2011.

[43] M. Kobayashi, N. Jindal, and G. Caire. Training and Feedback Optimization for Multiuser MIMO Downlink. *IEEE Transactions on Communications*, 59(8):2228–2240, Aug. 2011.

[44] C. Komninakis, C. Fragouli, A.H. Sayed, and R.D. Wesel. Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *IEEE Transactions on Signal Processing*, 50(5):1065–1076, May 2002.

[45] R.T. Krishnamachari and M.K. Varanasi. Interference alignment under limited feedback for MIMO interference channels. *in Proc. of IEEE International Symposium on Information Theory*, pages 619–623, Jun. 2010.

[46] A. Lapidoth. Nearest neighbor decoding for additive non-Gaussian noise channels. *IEEE Transactions on Information Theory*, 42(5):1520–1529, 1996.

[47] A. Lapidoth and S. Shamai. Fading channels: How perfect need perfect side information be? *IEEE Transactions on Information Theory*, 48(5):1118–1134, 2002.

[48] J. Lee, J.K. Han, and J. Zhang. MIMO technologies in 3GPP LTE and LTE-advanced. *EURASIP Journal on Wireless Communications and Networking*, pages 1–10, March 2009.

[49] N. Lee and Robert W. Heath, Jr. Not too delayed csit achieves the optimal degrees of freedom. *arXiv preprint arXiv:1207.2211*, 2012.

[50] Namyoon Lee and Syed A Jafar. Aligned interference neutralization and the degrees of freedom of the 2 user interference channel with instantaneous relay. *arXiv preprint arXiv:1102.3833*, 2011.

[51] K.C.J. Lin, S. Gollakota, and D. Katabi. Random access heterogeneous MIMO networks. *in Proc. of the ACM SIGCOMM Conference*, pages 146–157, 2011.

[52] Li Liu and H. Jafarkhani. Novel transmit beamforming schemes for time-selective fading multiantenna systems. *IEEE Transactions on Signal Processing*, 54(12):4767–4781, Dec. 2006.

[53] D.J. Love and Robert W. Heath, Jr. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Transactions on Information Theory*, 51(8):2967–2976, Aug. 2005.

[54] D.J. Love, Robert W. Heath, Jr., V.K.N. Lau, D. Gesbert, B.D. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE Journal on Selected Areas in Communications*, 26(8):1341–1365, Oct. 2008.

[55] A. Lozano. Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication. *IEEE Transactions on Wireless Communications*, 7(12):5020–5029, 2008.

[56] A. Lozano, Robert W. Heath, Jr., and J. G. Andrews. Fundamental limits of cooperation. *Arxiv preprint arXiv:1204.0011*, Mar. 2012.

[57] A. Lozano and A.M. Tulino. Capacity of multiple-transmit multiple-receive antenna architectures. *IEEE Transactions on Information Theory*, 48(12):3117–3128, 2002.

[58] M.A. Maddah-Ali, A.S. Motahari, and A.K. Khandani. Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, 54(8):3457–3470, 2008.

[59] M.A. Maddah-Ali and D. Tse. Completely stale transmitter channel state information is still very useful. *IEEE Transactions on Information*

Theory, 58(7):4418–4431, 2012.

[60] AbdulRahman Mahmoud, Mostafa El-Khamy, and Khaled Elsayed. Interference alignment performance on MIMO X channels with imperfect channel knowledge. *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications, Cesme, Turkey*, Jun. 2012.

[61] E. Martinian. Waterfilling gains O (1/SNR) at high SNR.

[62] T.L. Marzetta. Blast training: Estimating channel characteristics for high capacity space-time wireless. *Proceedings of the Annual Allerton Conference on Communication Control and Computing*, 37:958–966, 1999.

[63] T.L. Marzetta and B.M. Hochwald. Fast transfer of channel state information in wireless systems. *IEEE Transactions on Signal Processing*, 54(4):1268–1278, Apr. 2006.

[64] A.M. Mathai and S.B. Provost. *Quadratic forms in random variables: theory and applications*. M. Dekker, New York, 1992.

[65] K.K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Transactions on Information Theory*, 49(10):2562–2579, Oct. 2003.

[66] B. Nosrat-Makouei, J.G. Andrews, and Robert W. Heath, Jr. A simple SINR characterization for linear interference alignment over uncertain

MIMO channels. *in Proc. of IEEE Int. Symposium on Inf. Theory*, pages 2288–2292, Jun. 2010.

[67] O. El Ayach, A. Lozano, and Robert W. Heath, Jr. On the overhead of interference alignment: Training, feedback, and cooperation. *to appear in IEEE Transactions on Wireless Communication*, 2012.

[68] O. El Ayach and Robert W. Heath, Jr. Grassmannian differential limited feedback for interference alignment. *submitted to IEEE Transactions on Signal Processing*, Nov. 2011.

[69] O. El Ayach and Robert W. Heath, Jr. Interference alignment with analog channel state feedback. *IEEE Transactions on Wireless Communication*, 11(2):626–636, Feb. 2012.

[70] O. El Ayach, A. Lozano, and Robert W. Heath, Jr. Optimizing Training and Feedback for MIMO Interference Alignment. *in Proc. of Asilomar Conference on Signals, Systems, and Computers*, Nov. 6-9 2011.

[71] O. El Ayach, S. Peters, and Robert W. Heath, Jr. The feasibility of interference alignment over measured MIMO-OFDM channels. *IEEE Transactions on Vehicular Technology*, 59(9):4309–4321, Nov. 2010.

[72] L.H. Ozarow, S. Shamai, and A.D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43(2):359–378, 1994.

[73] D.S. Papailiopoulos and A.G. Dimakis. Interference alignment as a rank constrained rank minimization. *in Proc. of IEEE Global Telecommunications Conference*, pages 1–6, Dec. 2010.

[74] SW Peters and Robert W. Heath, Jr. Orthogonalization to reduce overhead in MIMO interference channels. *in Proc. of International Zurich Seminar*, pages 126–129, March 2010.

[75] S.W. Peters and Robert W. Heath, Jr. Cooperative algorithms for MIMO interference channels. *IEEE Transactions on Vehicular Technology*, 60(1):206–218, Jan. 2011.

[76] M. Razaviyayn, G. Lyubeznik, and Zhi-Quan Luo. On the degrees of freedom achievable through interference alignment in a MIMO interference channel. *IEEE Transactions on Signal Processing*, 60(2):812–821, Feb. 2012.

[77] Akula Aneesh Reddy, Siddhartha Banerjee, Aditya Gopalan, Sanjay Shakkottai, and Lei Ying. On distributed scheduling with heterogeneously delayed network-state information. *Queueing Systems*, pages 1–26, 2012.

[78] Mohsen Rezaee, Romain Couillet, Maxime Guillaud, and Gerald Matz. Sum-Rate Optimization for the MIMO IC under Imperfect CSI: a Deterministic Equivalent Approach. *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications, Cesme, Turkey*, Jun. 2012.

[79] Mohsen Rezaee and Maxime Guillaud. Limited feedback for interference alignment in the K-user MIMO Interference Channel. In *IEEE Information Theory Workshop (ITW)*, pages 667–671. IEEE, 2012.

[80] J.C. Roh and B.D. Rao. Efficient feedback methods for MIMO channels based on parameterization. *IEEE Transactions on Wireless Communications*, 6(1):282–292, Jan. 2007.

[81] C. Rose, S. Ulukus, and R.D. Yates. Wireless systems and interference avoidance. *IEEE Transactions on Wireless Communications*, 1(3):415 –428, Jul. 2002.

[82] D. Sacristan-Murga, F. Kaltenberger, A. Pascual-Iserte, and A.I. Perez-Neira. Differential feedback in MIMO communications: Performance with delay and real channel measurements. *in Proc. of Workshop on Smart Antennas*, pages 1–8, 2009.

[83] D. Sacristan-Murga and A. Pascual-Iserte. Differential feedback of MIMO channel gram matrices based on geodesic curves. *IEEE Transactions on Wireless Communications*, 9(12):3714–3727, Dec. 2010.

[84] D. Samardzija and N. Mandayam. Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems. *IEEE Transactions on Communications*, 54(7):1335–1345, Jul. 2006.

[85] I. Santamaría, O. González, Robert W. Heath, Jr., and S. W. Peters. Maximum sum-rate interference alignment algorithms for MIMO chan-

nels. *in Proc. of IEEE Global Telecommunications Conference*, pages 1–6, Dec. 2010.

[86] W. Santipach and M.L. Honig. Optimization of training and feedback overhead for beamforming over block fading channels. *IEEE Transactions on Information Theory*, 56(12):6103–6115, 2010.

[87] D.A. Schmidt, Changxin Shi, R.A. Berry, M.L. Honig, and W. Utschick. Minimum mean squared error interference alignment. *Conference Record of the 43rd Asilomar Conference on Signals, Systems and Computers*, pages 1106–1110, Nov. 2009.

[88] Maryam Modir Shanechi, Ron Porat, and Uri Erez. Comparison of practical feedback algorithms for multiuser MIMO. *IEEE Transactions on Communications*, 58(8):2436 –2446, Aug. 2010.

[89] Minqi Shen, A. Host-Madsen, and J. Vidal. An improved interference alignment scheme for frequency selective channels. *in Proc. of IEEE International Symposium on Information Theory*, pages 559–563, Jul. 2008.

[90] C. Shi, R. A. Berry, and M. L. Honig. Adaptive Beamforming in Interference Networks via Bi-Directional Training. *ArXiv preprint arXiv:1003.4764*, March 2010.

[91] Changxin Shi, D.A. Schmidt, R.A. Berry, and M.L. Honig. Local interference pricing for distributed beamforming in MIMO networks. *in*

Proc. of IEEE Military Communications Conference, Oct. 2009.

[92] Qingjiang Shi, M. Razaviyayn, Zhi-Quan Luo, and Chen He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, Sept. 2011.

[93] H. Shin and J.H. Lee. Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole. *IEEE Transactions on Information Theory*, 49(10):2636–2647, 2003.

[94] J. Starr, O. El Ayach, and Robert W. Heath, Jr. Interference alignment with per-antenna power constraints. *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 2746–2750, 2011.

[95] J. Starr, O. El Ayach, and Robert W. Heath, Jr. Mimo interference alignment with distributed antenna systems. *in preparation, to be submitted to IEEE Transactions on Wireless Communications*, 2013.

[96] GW Stewart. Stochastic perturbation theory. *SIAM review*, 32(4):579–610, 1990.

[97] G.W. Stewart. Perturbation theory for the singular value deomposition. *SVD and Signal Processing II, Algorithms, Analysis and Applications (Elsevier Science)*, pages 99 –109, 1991.

[98] T. Svantesson and A.L. Swindlehurst. A performance bound for prediction of MIMO channels. *IEEE Transactions on Signal Processing*, 54(2):520–529, Feb. 2006.

[99] R. Tresch and M. Guillaud. Cellular interference alignment with imperfect channel knowledge. *in Proc. of IEEE International Conference on Communications*, pages 1–5, Jun. 2009.

[100] R. Tresch, M. Guillaud, and E. Riegler. On the achievability of interference alignment in the K-user constant MIMO interference channel. *in Proc. of IEEE/SP 15th Workshop on Statistical Signal Processing*, pages 277–280, Sep. 2009.

[101] M. Trivellato, F. Boccardi, and H. Huang. On transceiver design and channel quantization for downlink multiuser mimo systems with limited feedback. *IEEE Journal on Selected Areas in Communications*, 26(8):1494–1504, 2008.

[102] A.M. Tulino and S. Verdú. *Random matrix theory and wireless communications*. Now Publishers Inc, 2004.

[103] Chinmay S Vaze and Mahesh K Varanasi. The degrees of freedom of the $2\times 2\times 2$ interference network with delayed csit and with limited shannon feedback. *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 824–831, 2011.

[104] C.S. Vaze and M.K. Varanasi. The degrees of freedom region of the two-user MIMO broadcast channel with delayed CSIT. *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 199–203, 2011.

[105] C.S. Vaze and M.K. Varanasi. The degree-of-freedom regions of mimo broadcast, interference, and cognitive radio channels with no csit. *IEEE Transactions on Information Theory*, 58(8):5354–5374, 2012.

[106] C.S. Vaze and M.K. Varanasi. The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT. *IEEE Transactions on Information Theory*, 58(7):4396–4417, 2012.

[107] P.Å. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.

[108] P. Xia, S. Zhou, and G.B. Giannakis. Achieving the welch bound with difference sets. *IEEE Transactions on Information Theory*, 51(5):1900–1907, 2005.

[109] Baile Xie, Yang Li, Hlaing Minn, and A. Nosratinia. Interference alignment under training and feedback constraints. *In Proc. of IEEE Global Telecommunications Conference*, Dec. 5–9 2011.

[110] C.M. Yetis, Tiangao Gou, S.A. Jafar, and A.H. Kayran. On feasibility of interference alignment in mimo interference networks. *IEEE Transactions on Signal Processing*, 58(9):4771–4782, 2010.

[111] T. Yoo and A. Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Transactions on Information Theory*, 52(5):2203–2214, 2006.

Vita

Omar El Ayach received the Bachelor of Engineering in Computer and Communications Engineering (with High Distinction) from the American University of Beirut (AUB), Lebanon in 2008. He joined the Wireless Networking and Communication Group (WNCG) at The University of Texas at Austin in Fall 2008. He obtained his Masters of Science in Engineering from The University of Texas in 2010. Since then, he has been pursuing his PhD in Electrical Engineering at the same university. He was an engineering research intern at the Dallas Technology Lab at Samsung Telecommunications America in the summers of 2011 and 2012. There he worked on developing novel precoding strategies for the millimeter wave single-user and broadcast channels. His research interests include signal processing, information theory, and the design of transmission strategies for multiple-input multiple-output systems.

Email address: oelayach@utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.