Linear Precoding in Wireless Networks with Channel State Information Feedback

LINEAR PRECODING IN WIRELESS NETWORKS WITH CHANNEL STATE INFORMATION FEEDBACK

BY

AHMED MEDRA, M.Sc., B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2014)	McMaster University
(Electrical & Computer Engineering)	Hamilton, Ontario, Canada

TITLE:	Linear Precoding in Wireless Networks with Channel State Information Feedback
AUTHOR:	Ahmed Medra M.Sc., (Electrical and Electronic Engineering) Alexandria University, Egypt.
SUPERVISOR:	Timothy N. Davidson Professor, Department of Electrical and Computer Engineering.
NUMBER OF PAGES:	xv, 192

To my father, my mother, my siblings,

Safaa, Malak and Abdullah

Abstract

This thesis focuses on the design of linear precoding schemes for downlink multiple-input multiple-output (MIMO) networks. These schemes are designed to be amenable to implementation in wireless networks that allow rate-limited feedback of channel state information (CSI). In the first half of this thesis, memoryless quantization codebooks are designed and incremental vector quantization techniques are developed for the representation of CSI in MIMO point-to-point links and isolated (single-cell) downlink networks. The second half of the thesis seeks to design linear precoding schemes for the multi-cell downlink networks that can achieve improved performance without requiring significantly more communication resources for CSI feedback than those required in the case of an isolated single-cell.

For the quantization problem, smooth optimization algorithms are developed for the design of codebooks that possess attractive features that facilitate their implementation in practice in the addition to having good quantization properties. As one example, the proposed approach is used to design rank-2 codebooks that have a nested structure and elements from a phase-shift keying (PSK) alphabet. The designed codebooks have larger minimum distances than some existing codebooks, and provide tangible performance gains.

To take advantage of temporal correlation that may exist in the wireless channel,

an incremental approach to the Grassmannian quantization problem is proposed. This approach leverages existing codebooks for memoryless quantization schemes and employs a quantized form of geodesic interpolation. Two schemes that implement the principles of the proposed approach are presented. A distinguishing feature of the proposed approach is that the direction of the geodesic interpolation is specified implicitly using a point in a conventional codebook. As a result, the approach has an inherent ability to recover autonomously from errors in the feedback path.

In addition to the development of the Grassmannian quantization techniques and codebooks, this thesis studies linear precoder design for the downlink MIMO networks in the cases of small networks of arbitrary topology and unbounded networks that have typical architectures. In particular, a linear precoding scheme for the isolated 2-cell network that achieves the optimal spatial degrees of freedom of the network is proposed. The implementation of a limited feedback model for the proposed linear precoding scheme is developed as well. Based on insight from that model, other linear precoding schemes that can be implemented in larger networks, but with finite size, are developed. For unbounded networks of typical architecture, such as the hexagonal arrangement of cells, linear precoding schemes that exploit the partial connectivity of the network are presented under a class of precoding schemes that is referred to as spatial reuse precoding. These precoding schemes provide substantial gains in the achievable rates of users in the network, and require only local feedback.

Acknowledgements

As my Ph.D program comes to an end, it is the time to express my gratitude towards everyone who helped me accomplish this achievement. I am very sorry that I will not be able to mention every one who influenced my career and participated in my progress. Further, for those who I will mention, I believe that it is impossible to express my deep thanks for them.

I would like to thank my father, mother and siblings for their support during my entire life. Indeed, no word can express my feelings towards the miracle actions that my parents did in order to be in my current position. I would like to thank every member of my family for the happy moments we have shared together.

As to my wife, Safaa my beloved, I am very grateful to her. We shared a lot of memories and happy times with our kids. During these moments where I had some difficulties in my research, she did her best to change my mood and to make me more confident. Really, I would not be able to achieve anything without her support and patience. Thanks my angel !

I would like to thank my friends and colleagues at McMaster University, Mohamed Elshimy, Osman Ahmed, Mohamed Negm and all the other friends that I did not mention for their valuable support. Special thanks to my friend and brother Ahmed Hamza.

I would like to thank my professors in the electrical and electronic engineering department at Alexandria University. In particular, I would like to express my graduate towards Prof. Hossam Shalaby for his sincere help and advices. I would like also to thank all the professors who I worked with as a teaching assistant. I am also indebted to Prof. Hassan El-Kamchouchi, my Msc. supervisor for his support. Furthermore, I would like to thank the electrical and computer engineering (ECE) department at McMaster University for the friendly and warm atmosphere it provides for the research students. In particular, I would like to thank Mrs. Cheryl Gies for all the support she provided to us. Also, I like to thank the technical staff for keeping our machine and labs running.

I would like to gratefully thank my supervisory committee members, Prof. Steve Hranilovic and Prof. Jun Chen for their insightful remarks and discussions during our committee meetings. Their comments were very useful in the development of my research.

Finally, I would like to express my deepest gratitude to my supervisor Prof. Timothy N. Davidson. During the past four years I enjoyed being one of this students. His discussions were interesting and fruitful. I really liked his way of thinking and his approaches when dealing with a hard problem. He has been a source of inspiration and guidance for me. Regretfully, no word can express my feelings towards him.

Contents

List of Figures

Chapter 1

Introduction

The focus of this thesis is on the design and implementation of linear precoding schemes for downlink multiple-input multiple-output (MIMO) wireless networks. These networks enable communication from a base station, or base stations, to wireless devices. In these systems, the base station is assumed to have more than one antenna, and the mobile device may have more than one antenna, too. The rate at which a base station can communicate, reliably, to the wireless devices that are assigned to it is fundamentally dependent on the accuracy of the information regarding the current channel state of the channel (the channel state information, CSI) that is available to the transmitter (Caire and Shamai, 2003; Jindal, 2006). For instance, for communication at high signal-to-noise ratios (SNRs) from a base station with M antennas to K single antenna receivers in a richly-scattered environment, the sum of the (ergodic) achievable rates in a system with perfect CSI at the base station grows as $\min\{M, K\} \log(\mathrm{SNR})$, whereas in the absence of any CSI at the base station it grows as $\log(\mathrm{SNR})$ (Caire and Shamai, 2003).

In time division duplex (TDD) systems, if the channel changes sufficiently slowly, and if appropriate compensation for analog front end components is performed (e.g., Kaltenberger *et al.*, 2010), then the base station can estimate the CSI during the training phase of uplink communication. However, in frequency domain duplex (FDD) systems, the channel is not reciprocal, and the channel state must be estimated at the receiver and then fed back to the base station. This feedback requires the allocation of resources (e.g., time, bandwidth and energy) that would otherwise allocated to direct communication. To ensure that the benefits of making CSI available to the base station outweigh the cost of getting it there, the feedback mechanism must be designed so that the feedback is limited in an appropriate sense (Love *et al.*, 2008). Fundamentally, this is a source-channel coding problem in which the “source” is the CSI that is to be communicated to the base station. Typically, the source and channel coding schemes for the CSI are separated. That separation will be made in this thesis, and the focus will be on the source coding problem. The first half of this thesis considers MIMO point-to-point links and isolated (single-cell) MIMO downlink networks and designs memoryless and incremental vector quantization schemes for representing the CSI. In the second half of the thesis, multi-cell downlink networks are considered. In these more complex networks, the question of limited feedback addresses not only how information is quantized or compressed, but also which information is provided to which node in the network (de Kerret and Gesbert, 2013).

In this chapter, the basic principles of CSI feedback are discussed, first in the context of single-user (SU) and single-cell schemes, and then in the context of a multi-cell MIMO downlink network. An overview of the underlying concepts for the proposed linear precoding schemes is also provided. Based on insight from these models, the main challenges that arise in the design and implementation of limited

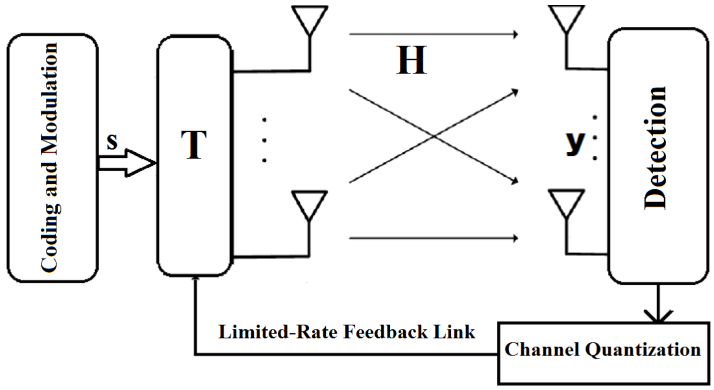

Figure 1.1: Limited feedback SU-MIMO system

feedback schemes are outlined. The resolution of some of these challenges is the main goal of this thesis.

1.1 Principles of Limited Feedback in Single-User MIMO Systems

To set the stage for the developments in this thesis, let us first consider a narrow-band single-user (SU)-MIMO system, as shown in Figure 1.1. In this system, the transmitter has M_t transmit antennas and communicates to a receiver with M_r receive antennas using a linear precoder. The received signal can be modelled as:

$$\mathbf{y} = \sqrt{\frac{E_s}{Q}}\mathbf{HTs} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the matrix of channel gains from the transmit antennas to the receive antennas, and is assumed to be constant within the communication interval. The matrix $\mathbf{T} \in \mathbb{C}^{M_t \times Q}$ is the transmit precoding matrix, where $Q \leq \min\{M_t, M_r\}$ is the number of transmitted data streams, and $\mathbf{T}$ is normalized so that trace$(\mathbf{T}^H\mathbf{T}) = Q$. The data symbols form a vector $\mathbf{s}$ that is scaled so that $\mathrm{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_Q$. This vector is then multiplied by $\sqrt{\frac{E_s}{Q}}\mathbf{T}$ to generate the transmitted signal $\mathbf{x} = \sqrt{\frac{E_s}{Q}}\mathbf{T}\mathbf{s}$ with E_s being the total transmit energy. The additive white Gaussian noise (AWGN) at the receiver, denoted by $\mathbf{n}$, is modelled as being a zero-mean circular complex Gaussian with $\mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \mathbf{I}$.

Under the assumption that the receiver obtains perfect knowledge of the channel $\mathbf{H}$, the achievable rate of this scheme is given by (Telatar, 1999)

$$R = \log\det\Big(\mathbf{I}_Q + \frac{E_s}{Q}\mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T}\Big). \tag{1.2}$$

A precoder structure that maximizes the achievable rate under an average transmitted power constraint is $\sqrt{\frac{E_s}{Q}}\mathbf{T} = \mathbf{V}\mathbf{\Phi}$, where $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ is the singular value decomposition of the channel matrix $\mathbf{H}$ and $\mathbf{\Phi}$ is a positive semidefinite diagonal matrix whose elements are obtained by applying the "water-filling" precodure to the singular values of $\mathbf{H}$ (Telatar, 1999). At higher SNRs, this scheme can be approximated by performing uniform power allocation over the subchannels with significant gains (e.g., Tse and Viswanath, 2005). That is $\mathbf{T} = \mathbf{V}_Q$, where $\mathbf{V}_Q$ is the first Q columns of $\mathbf{V}$. The resulting rate gap of such a scheme is small (Yu and Cioffi, 2006). Furthermore, in a limited feedback setting, this scheme has the advantage that the receiver need only inform the transmitter of $\mathbf{V}_Q$ and does not need to send $\mathbf{\Phi}$ as well. This reduces the amount of feedback that is required. (Actually, as will be discussed in Chapter 2, the transmitter only needs to know the subspace spanned by $\mathbf{V}_Q$, and this further reduces the feedback requirements.)

A natural approach to quantizing $\mathbf{V}_Q$ so that it can be fed back to the transmitter is to employ memoryless vector quantization (e.g., Gersho and Gray, 1991). In that scheme, a codebook $\mathcal{F} = \{\mathbf{F}_i\}$ of "tall" candidate matrices with orthonormal columns is designed offline and made available to both the transmitter and the receiver (Love and Heath, Jr., 2005b). Having identified the channel matrix $\mathbf{H}$, the receiver solves the problem

$$i^* = \arg\max_{\mathbf{F}_i \in \mathcal{F}} \log\det\Big(\mathbf{I}_Q + \frac{E_s}{Q}\mathbf{F}_i^H\mathbf{H}^H\mathbf{H}\mathbf{F}_i\Big) \tag{1.3}$$

and sends back the index i^* via a feedback channel to the transmitter, as depicted in Figure 1.1. At the transmitter side, the index i^* is used to select $\mathbf{F}_{i^*}$ from the transmitter's copy of the codebook and the transmit precoder is chosen to be $\mathbf{T} = \mathbf{F}_{i^*}$.

1.2 Isolated Single-Cell MU-MIMO Downlink Network

The basic principles of the above CSI feedback scheme extend naturally to the case of an isolated single-cell multi-user (MU) MIMO downlink network, as depicted in Figure 1.2. In this network, a base station of M_t transmit antennas communicates to K users, each of which has M_r receive antennas. As in the SU-MIMO case, the base station employs linear precoding, in which the transmitted signal takes the form of

$$\mathbf{x} = \sqrt{\frac{E_s}{Q}}\mathbf{T}\mathbf{s} = \sqrt{\frac{E_s}{Q}}\sum_{k=1}^{K}\mathbf{T}^k\mathbf{s}^k, \tag{1.4}$$

where $\mathbf{T} = [\mathbf{T}^1, \mathbf{T}^2, \ldots, \mathbf{T}^K]$ is the transmit precoder used at the base station, $\mathbf{s}^k \in \mathbb{C}^{d^k}$ is the vector of data symbols for user k and $\mathbf{s}^T = [\mathbf{s}^{1^T}, \mathbf{s}^{2^T}, \ldots, \mathbf{s}^{K^T}]$. The received

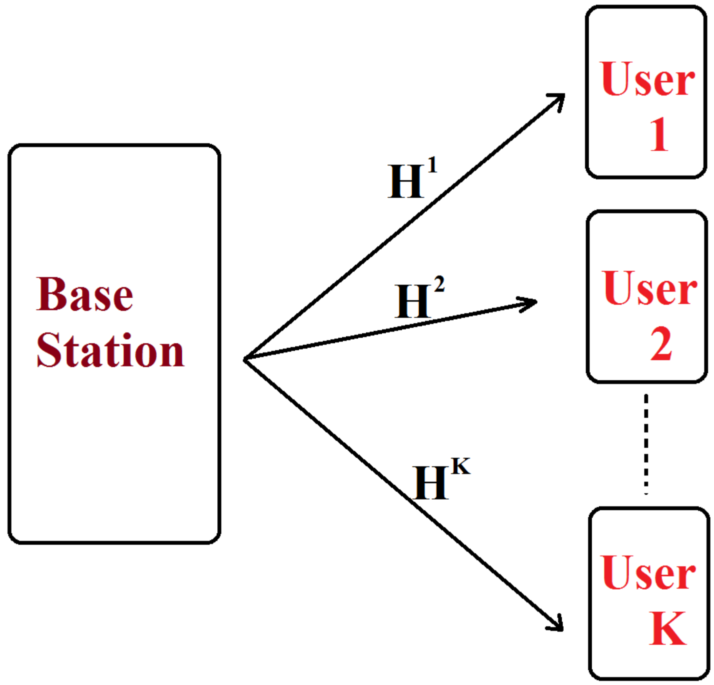

Figure 1.2: An isolated single-cell downlink network

signal at the kth user $\mathbf{y}^k \in \mathbb{C}^{M_r}$, can be modelled as

$$\mathbf{y}^k = \mathbf{H}^k\mathbf{x} + \mathbf{n}^k = \sqrt{\frac{E_s}{Q}}\mathbf{H}^k\mathbf{T}^k\mathbf{s}^k + \sqrt{\frac{E_s}{Q}}\mathbf{H}^k\sum_{\ell \neq k}^{K}\mathbf{T}^\ell\mathbf{s}^\ell + \mathbf{n}^k, \tag{1.5}$$

where $\mathbf{H}^k \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix between the base station and the kth user, $\mathbf{x}$ is the transmitted signal from the base station, and $\mathbf{n}^k$ is the zero-mean additive white circular Gaussian noise with unit variance. Comparing (1.1) and (1.5), it can be seen that the received signal contains an additional component besides the desired signal, namely the "intra-cell interference" due to the transmissions of the base station to other users in the same cell. This appears as the second term on the right hand side of (1.5). For the base station to be able to design the precoder $\mathbf{T}$ to suppress the intra-cell interference at non-intended receivers, the presence of sufficiently accurate

CSI is required (Jindal, 2006). In particular, each user feeds back a quantized version of a measure of its channel matrix $\mathbf{H}^k$ to the base station. The base station collects this information and designs the linear precoder $\mathbf{T}$ using the quantized CSI according to one of a variety of design strategies, such as transmit matched filtering (Hanly, 1996), zero-forcing beamforming (Spencer *et al.*, 2004b), regularized channel inversion (Spencer *et al.*, 2004a), or robust downlink precoding (e.g., Shenouda and Davidson, 2007).

1.3 Multi-cell MIMO Downlink Network

The extension of the basic principles of CSI feedback to the case of a multi-cell downlink network is now considered. An example of a multi-cell network is the network presented in Figure 1.3. In this example, the network consists of 3 cells, and each base station (BS) communicates to 2 users in its cell. The signals transmitted by any base station cause inter-cell interference to users in the other two cells.

Let us consider the general case of a multi-cell MIMO downlink network that consists of G-cells, where each base station communicates to K users in its cell. The base stations are equipped with M_t transmit antennas and each receiver has M_r receive antennas. As in the MU-MIMO case, attention is focused on linear precoding schemes, in which the signal transmitted from base station j takes the form

$$\mathbf{x}_j = \mathbf{T}_j\mathbf{s}_j = \sum_{k=1}^{K} \mathbf{T}_j^k \mathbf{s}_j^k, \tag{1.6}$$

where $\mathbf{T}_j = [\mathbf{T}_j^1, \mathbf{T}_j^2, \ldots, \mathbf{T}_j^K]$ is the transmit precoder used at base station j, $\mathbf{s}_j^k \in \mathbb{C}^{d^{j,k}}$ is the vector of data symbols for user (j,k), and $\mathbf{s}_j^T = [\mathbf{s}_j^{1^T}, \mathbf{s}_j^{2^T}, \ldots, \mathbf{s}_j^{K^T}]$. (Here, the

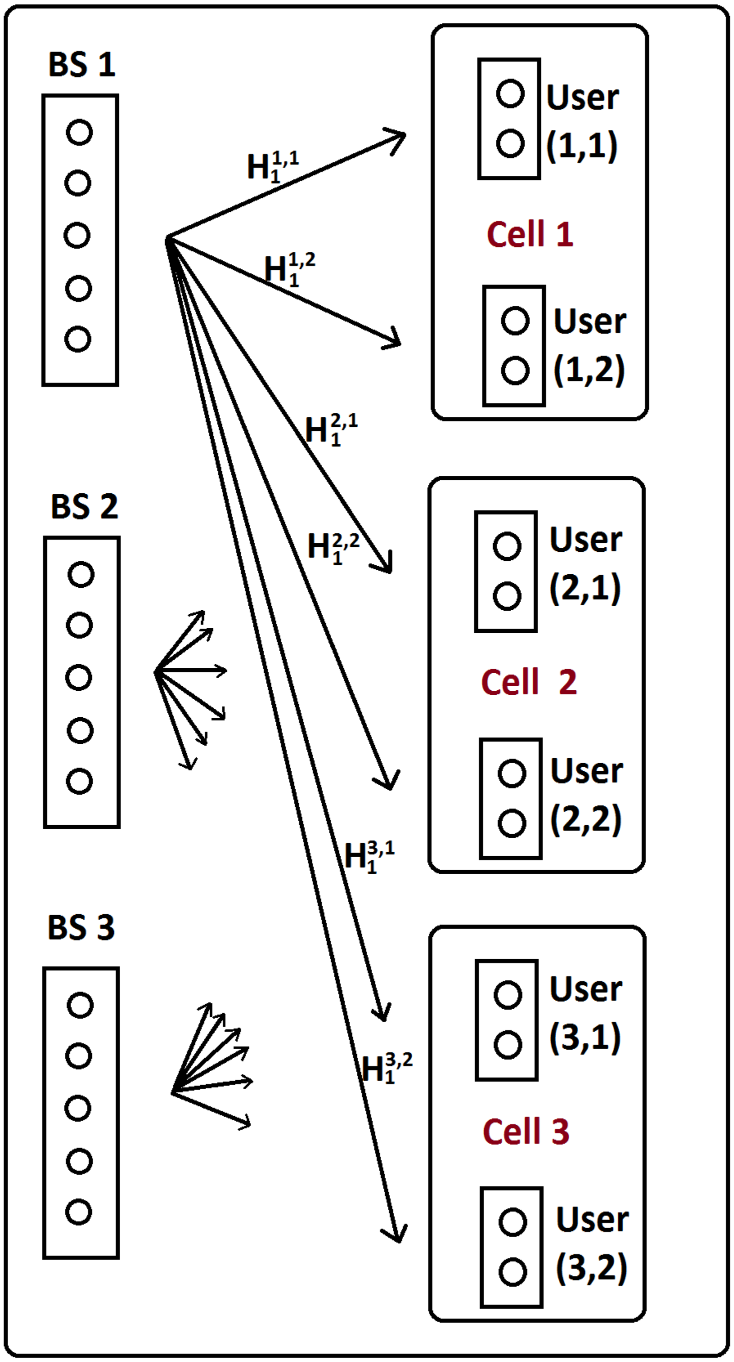

Figure 1.3: A 3-cell 2-user-per-cell MIMO downlink network

power allocation has been absorbed into the columns of $\mathbf{T}$.) Accordingly, the received signal at user k in cell i, i.e., user (i, k), is

$$\mathbf{y}^{i,k} = \sum_{j=1}^{G} \mathbf{H}_j^{i,k} \mathbf{x}_j + \mathbf{n}^{i,k} \tag{1.7a}$$

$$= \mathbf{H}_i^{i,k} \mathbf{T}_i^k \mathbf{s}_i^k + \mathbf{H}_i^{i,k} \sum_{\ell \neq k}^{K} \mathbf{T}_i^\ell \mathbf{s}_i^\ell + \sum_{j \neq i}^{G} \mathbf{H}_j^{i,k} \mathbf{T}_j \mathbf{s}_j + \mathbf{n}^{i,k}, \tag{1.7b}$$

where $\mathbf{y}^{i,k} \in \mathbb{C}^{M_r}$, $\mathbf{H}_j^{i,k} \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix between BS j and user (i, k), $\mathbf{x}_j$ is the transmitted signal from BS j, and $\mathbf{n}^{i,k}$ is the zero-mean additive white circular Gaussian noise with unit variance. The first term on the right hand side of (1.7b) is the desired signal term, the second term is the “intra-cell interference” from transmissions that the ith base station makes to other users in its cell, and the third term is the “inter-cell interference” from transmissions made by other base stations. The expression in (1.7b) implicitly outlines the significant expansion in the channel state information that is to be communicated amongst the nodes of the network. Not only should base station i be informed of the channels to each user in its own cell, as was the case in the isolated single-cell downlink, but it appears that it should be also be informed of the channels through which interference will “leak” to users in the other cells, $\{\mathbf{H}_j^{i,k}\}_{j=1, j \neq i}^{G,K}$. This is a substantial increase in the number of channels about which each base station has to acquire information. For this reason, in multi-cell networks, the CSI quantization and feedback scheme should first identify which CSI is really required by each base station (de Kerret and Gesbert, 2013). Otherwise the resources dedicated for feedback may not suffice. In particular, it would be preferable if one could design precoding schemes that can achieve improved performance without requiring a significant increase in the resources dedicated for feedback over the resources required in the single cell case. Another challenge in the design of the linear precoding scheme is whether it requires a central processing unit or not. It may be desirable if each base station is able to design its linear transmit precoder $\mathbf{T}_j$ without cooperation with the other base stations.

1.4 Key Challenges of Limited Feedback in Downlink Networks

The system models for the SU-MIMO case, the single-cell downlink, and the general G-cell downlink network described in previous sections generate key insights into the main components of limited feedback schemes and the main challenges that need to be resolved. These challenges can be summarized as:

1. **The quantization process:**

   Following the estimation of the channel by the receiver, typically through the use of a training sequence of known symbols sent by the transmitter, the receivers quantize the information regarding the physical channel. This process involves the following:

   - **Which information to be quantized?**

     Since the resources dedicated for the feedback process are limited, the design of the quantization process and the precoding scheme that is subsequently implemented at the transmitter should address the problem of identifying which information needs to be compressed in order to achieve improved performance for the network using the limited available resources.

- **The quantization scheme:**

  The quantization scheme should exploit the structure of the evolution of the channel. In the case of temporally uncorrelated channels, memoryless vector quantization schemes have shown to be effective techniques for channel quantization. On the other hand, in temporally correlated channels, the implementation of quantization schemes that exploit the memory in the wireless channel have the potential to provide improved performance when compared to the single shot memoryless quantization schemes assuming the same feedback budget.

- **The quantization codebook:**

  A key component in the design of any vector quantization system is the design of the quantization codebook that will be used at the receiver to convey useful information about the channel to the transmitter. This codebook should reflect the properties of the channel, and, at the same time, should possess other features that facilitate its implementation and storage.

2. **The linear precoding scheme:**

   In the SU-MIMO system and the isolated single-cell MIMO downlink, the core issues that arise in the design of the linear precoder are reasonably well established. This is due, in part, to the fact that the architecture of the CSI feedback system is well developed, and to the fact that analysis of the rates that can be achieved within that architecture has generated considerable design insights. The design of linear precoding schemes for the multi-cell downlink networks is rather more intricate task. This intricacy arises from several facts, some of which are as follows:

- The system architecture is still being developed, and is not well established, even for the perfect CSI case.
- It would be desirable if the designed precoding schemes could achieve improved performance without requiring a significant increase in the communication resources dedicated for the feedback process.
- It would be desirable if the design of a linear precoding scheme were generic, in the sense that it could be implemented in an arbitrary network under different configurations for the number of cells, the number of users, and other network parameters. This complicates the design problem substantially.
- Even though the capacities of most wireless network architectures are still unknown, significant steps have been taken in the characterization of the high signal-to-noise (SNR) regime for several networks, or what is known as "Degrees of Freedom" (DoF). The study of the DoF of networks has generated key insights for the development of linear precoding schemes that can provide improved performance against conventional interference techniques, (e.g., the interference alignment linear precoding scheme, Maddah-Ali *et al.*, 2008; Jafar, 2011). This analysis suggests that one approach to the design of linear precoding schemes for these networks would be to enable the scheme to achieve a large fraction of the DoF of the network, while being amenable to implementation with limited feedback.

The above challenges have provided much of the motivation for the topics that will be investigated in this thesis. These topics are outlined in the next section.

1.5 Thesis Outline

The focus of this thesis is on developing the key components of limited feedback schemes for a variety of downlink MIMO networks. Networks in which a single base station communicates to a single node, as in point-to-point MIMO links, or, communicates to multiple-nodes, as in the isolated single-cell MIMO network, have established architectures that provide insights on how limited feedback models can be designed and implemented (Love *et al.*, 2008). For these networks, the focus will be on the design of the quantization scheme, and the underlying codebook. For multicell downlink networks, the problem becomes more involved as it involves the design of the system architecture, in addition to the quantization scheme and the underlying codebook.

Accordingly, this thesis begins with the development of quantization schemes for the SU-MIMO and MU-MIMO systems. In those systems, linear precoding schemes designed for the case of perfect CSI can be adapted to be implemented in a limited feedback version of the system. In Chapter 2, design techniques are developed for Grassmannian quantization codebooks that are flexible enough to be able to design codebooks with particular features that simplify their practical implementation. These features include codebooks of constant modulus elements, codebooks with elements chosen from defined alphabets such as 4-QAM, and codebooks with nested structure. In order to design these codebooks, the use of optimization techniques based on smooth objective functions of matrix variables that lie on the Grassmanian manifold is proposed. In some cases, reduced dimensional exhaustive search is used as well. The codebooks produced by these optimization methods are compared with existing codebooks, and it is shown that they generate codebooks with significantly improved distance properties. In certain scenarios, the designed codebooks are optimal in the sense that they achieve a known bound on the minimum distances between codewords.

While memoryless quantization schemes have been shown to be effective for uncorrelated channels, in the case of temporally-correlated channels under the same feedback budget, incremental/differential quantization schemes provide the possibility of improved performance. Accordingly, in Chapter 3 an incremental feedback scheme on the Grassmannian manifold is developed that exploits the temporal correlation that exists in the wireless channel. The scheme has many attractive features as it needs only one codebook for initialization and updates, and that codebook can be designed using the techniques in Chapter 2. Another feature for the proposed scheme is that it can recover autonomously from feedback errors. Using statistical analysis, the key elements of the proposed incremental schemes are characterized and through extensive simulations, their superiority when compared to existing schemes under different channel environments and conditions is illustrated.

In Chapter 4, the case of multi-cell downlink MIMO networks is considered, with a focus on small scale networks. A review of the concept of interference alignment (IA) precoding along with the definition of the degrees of freedom (DoF) is presented. Then linear IA precoding schemes for the isolated 2-cell MIMO downlink network are proposed in case of perfect CSI and their corresponding implementations for limited feedback systems. The main conclusion is that designing precoding schemes that require only local "in-cell" feedback leads to a significant reduction in the feedback budget without a dramatic reduction in performance. With appropriate scaling of the feedback budget with the operating SNR, these schemes remain optimal in the DoF sense. Hence, these schemes are very attractive for limited feedback systems. For the remainder of the thesis, the focus is on designing precoding scheme that can be implemented without cooperation between transmitters.

The precoding schemes presented in Chapter 4 provide good insights and guidelines for designing linear precoding schemes that can be implemented in a larger network of $G > 2$ cells. In Chapter 5 precoding schemes are designed that achieve more DoF than conventional interference avoidance schemes, such as time division multiple access (TDMA), without requiring a significant increase in the resources required for CSI feedback. These schemes exploit the inherent partial connectivity in a cellular network due to path loss. An application for such precoders in heterogeneous networks is also presented.

Finally, Chapter 6 concludes this thesis and presents directions for future work.

Chapter 2

Flexible Codebook Design for Limited Feedback Systems via Sequential Smooth Optimization on the Grassmannian Manifold

As discussed in Chapter 1 and explained in more detail in this chapter, the presence of the CSI at the transmitter plays an important role in increasing the achievable rate of SU-MIMO and MU-MIMO systems with slowly fading channels. For such schemes, the architecture of limited feedback has been well established, and in this chapter one seeks to optimize aspects of the quantization scheme within that architecture. More specifically, smooth optimization problems are formulated with variables lying on the Grassmannian manifold. These formulations generate quantization codebooks with excellent distance properties and special features that facilitate their practical implementation.

2.1 Introduction

The structure of effective schemes for communicating wirelessly between nodes with multiple antennas is dependent on the extent to which the transmitter can adapt its transmission to the current state of the communication environment. As an example, in MIMO systems, CSI can be used by the transmitter to expand the achievable rate region through prudent management (and exploitation) of interference. In particular, consider the case of downlink transmission over a richly-scattered environment from a base station with M antennas to K mobile stations, each with a single antenna, in the presence of additive white Gaussian noise at the receivers. In such a system, if the base station has perfect knowledge of the channels to the receivers, then the sum of the ergodic achievable rates of coherent communication to the mobile users grows as $\min\{M, K\} \log(\mathsf{SNR})$ as the SNR increases (Caire and Shamai, 2003). When $M \geq K$ this rate of growth can be achieved by employing a simple linear zero-forcing transmission strategy (Caire and Shamai, 2003; Lee and Jindal, 2007). However, in the absence of any information about the state of the channel, the sum rate grows only as $\log(\mathsf{SNR})$ as the SNR increases (Caire and Shamai, 2003; Jafar and Goldsmith, 2005). In other words, the availability of perfect CSI at the transmitter enables a $\min\{M, K\}$-fold increase in the rate of growth of the achievable sum rate over that of a system with a single antenna at the transmitter, whereas in the absence of CSI at the transmitter the rate of growth is the same as that of a system with a single transmit antenna. (See also, Jindal, 2006).

Seeing as coherent receivers obtain an estimate of the channel, typically through training, prior to processing their received signals, a natural approach to providing the transmitter with information regarding the channel state would be for the receivers to feed back these estimates (or some function thereof) to the transmitter. However, account needs to be taken of the communication resources that are allocated to the feedback. When the feedback is digital, the question of how the channel state information should be quantized arises naturally (Love *et al.*, 2008; Jindal, 2006; Caire *et al.*, 2010). In essence, this is a source coding problem (Gersho and Gray, 1991), in which the channel (or some function thereof) is regarded as a source that needs to be represented using a small number of bits while introducing only a small "distortion". In this application, the notion of distortion is associated with the degradation of a chosen measure of the performance of the communication system with respect to the perfect CSI case. Based on insight from the source coding literature, a number of feedback strategies have been developed (Love *et al.*, 2008), many of which are based on the principles of memoryless vector quantization (Gersho and Gray, 1991). In those schemes, the receiver compares its channel estimate (or some function thereof) to "codewords" in a codebook and the index of the codeword that is closest to the channel estimate, in some appropriate sense, is fed back to the transmitter. The transmitter has a copy of the codebook and uses the index it receives to reconstruct the codeword that the receiver selected.

Although generic techniques for the design of quantization codebooks (e.g., Gersho and Gray, 1991) could be applied to the limited feedback problem, these techniques tend to be rather cumbersome and the resulting codebooks can be rather awkward to implement. As a result, in many scenarios, the quantization codebook is partitioned into a codebook that represents the directional information and a codebook that represents the gain (Love *et al.*, 2008). In a number of important scenarios, the relevant communication performance metric depends on the subspaces spanned by the directions, rather than the directions themselves. Since subspaces of a given dimension in a given ambient space can be represented by points on the Grassmannian manifold of the corresponding dimensions, in these scenarios the natural setting for the design of the codebook that represents the directional information is the Grassmannian manifold (Love *et al.*, 2003; Love and Heath, Jr., 2005a,b; Roh and Rao, 2006; Shenouda and Davidson, 2008; Dai *et al.*, 2008).

Despite the insight provided by posing the codebook design problem on the manifold, codebook design remains a rather intricate task, even when the scenario is such that the codewords should be uniformly distributed on the manifold with respect to a certain distance metric; a problem that is typically referred to as Grassmannian Subspace Packing (Conway *et al.*, 1996). One reason for this intricacy is that, with the exception of a few special cases (e.g., Pitaval *et al.*, 2011a), the problem has proven quite resistant to analysis of the structure of the optimal codebook. Furthermore, numerical optimization procedures are complicated by the fact that the manifold itself is inherently non-convex, as are the objectives that are typically used to capture the desired performance properties of the codebook. Some numerical design methods have been proposed, including adaptations of Lloyd's algorithm to the manifold (Roh and Rao, 2006; Xia and Giannakis, 2006), an alternating projection method (Dhillon *et al.*, 2008), and an "expansion–compression" algorithm (Schober *et al.*, 2009). In addition, methods for the design of Grassmannian constellations for non-coherent point-to-point MIMO communication (e.g., Agrawal *et al.*, 2001; Hochwald *et al.*, 2000; Kammoun *et al.*, 2007; Gohary and Davidson, 2009) can be employed.

Numerical methods have been quite successful in designing codebooks in which each codeword represents a subspace of dimension one; i.e., a single direction. Such codebooks are often referred to as being "rank one" codebooks, and the subspace packing problem is often referred to as the Grasssmannian Line Packing problem. Some of these numerical methods have been also used to design "rank two" codebooks of codewords representing two-dimensional subspaces for systems with small number of transmitter antennas. However, for larger problems, there are fewer results available, especially in the case in which distances other than the chordal distance are chosen as the metric on the manifold.

In addition to pure quantization performance, practical considerations suggest that codebooks ought to possess additional properties that facilitate their implementation and storage (Clerckx *et al.*, 2008). For example, (i) in the single-user case one might seek to constrain the elements of the codewords to have constant modulus, so as to avoid power imbalances at the transmitter; (ii) for reasons of storage and computational cost, it may be desirable to have the elements of the codewords come from a defined alphabet such as the phase-shift keying (PSK) alphabet (Ryan *et al.*, 2009); and (iii) in applications in which there is the option of multiple signalling modes (e.g., Love and Heath Jr., 2005), it is desirable for the codebook to be nested, in the sense that structure is imposed on the higher rank codewords so that they can also be used to generate codewords for codebooks of lower rank. In that case, the transmitter and the receiver need to store only a single codebook, and calculations results for a lower rank codeword can be re-used for the calculation of a higher rank one. A nested codebook also facilitates the transmitter's use of "rank overriding", in which the transmitter overrides the receiver's decision on how many data streams to transmit and uses lower-rank transmission (Clerckx *et al.*, 2008).

The desire to obtain codebooks that possess additional properties significantly complicates what is already quite a difficult problem. In the case of codebooks from a defined alphabet, the problem becomes a combinational one. For certain alphabets, codebooks of certain dimensions and sizes can be generated (Mondal *et al.*, 2007; Inoue and Heath Jr., 2009) using the concept of mutually unbiased bases (MUBs) from the theory of quantum mechanics (Tselniker *et al.*, 2009). By construction, codebooks derived from MUBs exhibit large distances between each pair of codewords, but the inherent restrictions on the technique limit the scenarios in which it can be applied.

Given the rather specialized nature of the existing approaches to structured Grassmannian codebook design, the goal of this chapter is to suggest a flexible design approach based on smooth optimization on the Grassmannian manifold (Edelman *et al.*, 1998; Manton, 2002). In particular, to optimize the quantization performance of otherwise unconstrained codebooks the insight from the technique of Gohary and Davidson (2009) is used and a sequence of smooth approximations to the objective is constructed. Since each approximation has continuous first and second order derivatives, a codebook that (locally) optimizes the approximation can be obtained by using adaptations of the conventional smooth optimization techniques, such as gradient descent, conjugate gradient and Newton's method, to the manifold (Edelman *et al.*, 1998; Manton, 2002). The optimization with respect to the next approximation in the sequence is initialized with the codebook obtained by optimizing the current approximation. In order to tackle problems with constraints on the codebook, the use of a combination of smooth penalty functions, which enable the algorithms for unconstrained optimization on the manifold to be retained is suggested, along with rounding techniques and low-dimensional searches.

Since both the original design problem and the smooth unconstrained approximations are not convex, the proposed approach does not necessarily yield optimal codebooks. However, in a number of the examples for the unconstrained case the codebooks that have been obtained achieve a known upper bound on the codebook quantization performance and hence are optimal. In the case of the Fubini-Study distance, for which a simple performance bound is not available, the unconstrained codebooks that have been obtained exhibit larger minimum distances than the existing packings (Love, 2004). Furthermore, the proposed approach has been able to generate codebooks with excellent properties for systems with a larger number of transmit antennas, larger subspace dimensions and larger size than those currently available. In the case of the chordal distance, the proposed approach has been able to design unconstrained codebooks that have essentially the same minimum distances as the codebooks of Dhillon *et al.* (2008). In the case of constrained codebooks, the proposed approach has been able to design some codebooks that achieve an upper bound on the performance and hence are optimal, and other codebooks with distance properties that are close to those of the best unconstrained codebooks that we have been able to generate.

In order to assess the impact of the obtained codebook designs on the performance of practical limited feedback systems, the performance of existing and proposed codebooks is evaluated under three transmission schemes for the downlink of a wireless system (from base station to mobile station) with limited feedback, namely Zero-Forcing Beam Forming (ZFBF) (Jindal, 2006; Spencer *et al.*, 2004b), the Per-User

Unitary beamforming and Rate Control (PU^{2}RC) system (Kim *et al.*, 2006), and zero-forcing block diagonalization (ZFBD) (Choi and Murch, 2004; Spencer *et al.*, 2004b; Ravindran and Jindal, 2008). In all three schemes, the transmitter employs linear precoding of the symbols intended for each user. In the first two schemes, the base station transmits a single data stream to each mobile station and hence rank-one codebooks designed using Grassmannian line packing are employed. The difference between those schemes lies in the way that precoding matrix is constructed from the quantized feedback. In the third scheme, the base station transmits multiple streams to each (multiple-antenna) mobile station, and higher rank subspace codebooks are employed. The presented results show that the achievable rate can be increased and the outage probability can be decreased by implementing well-designed codebooks. These results also show that the proposed approach yields structured codebooks with performance close to that of the unstructured codebooks.

2.2 System Model

Although the proposed approach to the design of structured Grassmannian quantization codebooks is quite generic and can be applied in many applications, the approach will be presented in the context of a multi-user multiple-input multiple-output (MU-MIMO) downlink communication system in which a base station with M_t antennas communicates over narrow-band channels to K receivers, the kth of which is equipped with $M_{r_k} < M_t$ antennas. The case of a point-to-point MIMO link arises as the special case in which $K = 1$. As in Section 1.2, at each channel use, the transmitter communicates simultaneously with all K receivers, sending $Q_k \leq \min\{M_t, M_{r_k}\}$ symbols to the kth receiver. The transmitter synthesizes the signal $\mathbf{x} = \sqrt{\frac{E_s}{Q}} \sum_i \mathbf{T}_i \mathbf{s}_i \in \mathbb{C}^{M_t \times 1}$

for transmission, where E_s is the total transmit power, Q is the total number of symbols, $\mathbf{s}_i \in \mathbb{C}^{Q_i \times 1}$ contains the symbols intended for the ith receiver, normalized so that $E\{\mathbf{s}_i\mathbf{s}_i^H\} = \mathbf{I}_{Q_i}$, and $\mathbf{T}_i \in \mathbb{C}^{M_t \times Q_i}$ is a linear precoding matrix. The average transmitted power per symbol for the ith receiver is $\text{trace}(\mathbf{T}_i\mathbf{T}_i^H)/Q_i$. The signal observed at the kth receiver is

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k = \sqrt{\frac{E_s}{Q}}\mathbf{H}_k\mathbf{T}_k\mathbf{s}_k + \sqrt{\frac{E_s}{Q}}\mathbf{H}_k\sum_{i \neq k}\mathbf{T}_i\mathbf{s}_i + \mathbf{n}_k, \tag{2.1}$$

where $\mathbf{H}_k \in \mathbb{C}^{M_{r_k} \times M_t}$ is the matrix of complex channel gains from the antenna inputs at the transmitter to the antenna outputs at the kth receiver. The first term on the right hand side of (2.1) denotes the desired signal term, the second denotes the interference from signals transmitted to other receivers and the third term represents the additive noise at the kth receiver. A more compact form of the model in (2.1) can be obtained by concatenating the $Q = \sum_i Q_i$ symbols sent to the receivers into a single vector $\mathbf{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \ldots, \mathbf{s}_K^T\right]^T$, and concatenating the precoding matrices $\mathbf{T}_i$ to form $\mathbf{T} = \left[\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_K\right] \in \mathbb{C}^{M_t \times Q}$. In that case, the expression in (2.1) can be rewritten as

$$\mathbf{y}_k = \sqrt{\frac{E_s}{Q}}\mathbf{H}_k\mathbf{T}\mathbf{s} + \mathbf{n}_k. \tag{2.2}$$

The scenarios that will be considered are those in which the channels can be modelled as remaining (approximately) constant for long enough that it is viable for the receiver(s) to feed back information regarding the channel state to the transmitter over a dedicated channel of limited capacity, and for the transmitter to adapt its linear precoding matrix $\mathbf{T}$ (and possibly the coding and modulation schemes that produce $\mathbf{s}$)

to the information that it receives. Though the channel may exhibit temporal correlation between fading blocks, in this chapter, the focus will be on designing codebooks for memoryless quantization schemes, while Chapter 3 studies quantization schemes for channels with memory. Therefore, some examples of memoryless quantization schemes are briefly reviewed before tackling the codebook design problem.

2.2.1 Single-user MIMO (SU-MIMO)

In the case of a point-to-point MIMO link, for which $K = 1$, a simplified limited feedback strategy that has proven effective in practice is for the receiver to inform the transmitter of the directions in which transmission should take place and for the transmitter to allocate power uniformly over these directions. In that case, many communication objectives depend on the subspace spanned by the directions and not on the directions themselves, and hence quantization on the Grassmannian manifold arises naturally. If the transmitter and the receiver are provided with a codebook $\mathcal{F} = \{\mathbf{F}_i\}_{i=1}^{L}$ containing $L = 2^B$ tall matrices of size $M_t \times Q$ with orthonormal columns, each of which is the Grassmannian representative of the subspace that it spans, then the receiver selects the matrix (codeword) $\mathbf{F}_{i^\star}$ that maximizes a chosen performance metric and sends the index $i^\star$ to the transmitter so it can identify this subspace (Love and Heath, Jr., 2005b). For example, if the performance metric is the Gaussian mutual information, then the receiver selects a codeword $\mathbf{F}_i \in \mathcal{F}$ according to

$$\mathbf{F}_{i^\star} = \arg\max_{\mathbf{F}_i \in \mathcal{F}} \log\det\Big(\mathbf{I}_{M_r} + \frac{E_s}{Q}\mathbf{F}_i^H\mathbf{H}^H\mathbf{H}\mathbf{F}_i\Big). \tag{2.3}$$

The index of $\mathbf{F}_{i^\star}$ is sent back to the transmitter which will use $\mathbf{F}_{i^\star}$ as the current preprocessing matrix, i.e, $\mathbf{T} = \mathbf{F}_{i^\star}$. Therefore, the achievable rate is the optimal value of the problem in (2.3). That rate, or the index of an appropriate coding and modulation scheme, can be also fed back to the transmitter.

2.2.2 Single Receive Antenna Downlink (MU-MISO)

In a multi-user downlink scenario in which each user has a single receive antenna ($M_{r_k} = 1$ and hence $Q = K$), the feedback of the channel state information is often partitioned into a channel direction vector and a scalar that represents the "quality" of the channel. Similar to the single user case, the transmitter and the receivers are provided with a codebook $\mathcal{F} = \{\mathbf{f}_i\}_{i=1}^{L}$ of unit norm vectors and the kth receiver selects the codeword (vector) that solves (Jindal, 2007)

$$\mathbf{f}_{i^\star,k} = \arg\max_{\mathbf{f}_i \in \mathcal{F}} |\mathbf{h}_k \mathbf{f}_i^*|, \tag{2.4}$$

and transmits $i^\star$ back to the transmitter. The transmitter uses these indices to reconstruct $\mathbf{f}_{i^\star,k}$ for all k, and then uses these vectors to calculate the (normalized) preprocessing matrix $\mathbf{T}$ in (2.1). A subsequent "dedicated training" step (e.g., Caire *et al.*, 2010) enables the kth receiver to determine $\mathbf{h}_k\mathbf{T}$, and the achievable rate for the kth user, which treats interference as noise, is given by

$$R_k^{BF} = \log\Big(1 + \frac{\frac{E_s}{K}|\mathbf{h}_k\mathbf{t}_k|^2}{1 + \sum_{j\neq k}\frac{E_s}{K}|\mathbf{h}_k\mathbf{t}_j|^2}\Big) \tag{2.5}$$

In the case that the transmitter performs Zero-Forcing Beamforming (ZFBF) with equal power loading, a matrix $\tilde{\mathbf{T}}$ is constructed by aggregating the quantized channel directions sent by the scheduled users according to:

$$\tilde{\mathbf{T}} = \Big[\mathbf{f}_{i^\star,1}, \mathbf{f}_{i^\star,2}, \ldots, \mathbf{f}_{i^\star,K}\Big]^\dagger, \tag{2.6}$$

where $\mathbf{A}^\dagger$ is the pseudo inverse of $\mathbf{A}$. The precoding matrix $\mathbf{T}^{ZFBF}$ is then formed by normalizing each column in $\tilde{\mathbf{T}}$, and if $\mathbf{t}_j$ denotes the jth column of $\mathbf{T}$, then $\mathbf{f}_{i^\star,k}^H \mathbf{t}_j^{ZFBF} = 0$ for all $k \neq j$. It can be observed that if two users select the same $\mathbf{f}_{i^\star} \in \mathcal{F}$, then the matrix $\mathbf{T}$ will have rank less than K. In practice, this scenario is avoided by providing each user a uniquely rotated version of the codebook $\mathcal{F}$. For comparison, in the case of perfect channel state information at the transmitter ZFBF results in $\mathbf{t}_j^{ZFBF}$ being orthogonal to $\mathbf{h}_k$ for $j \neq k$ and hence in that case the achievable rate simplifies to $R_k^{ZFBF-CSIT} = \log\big(1 + \frac{E_s}{K}|\mathbf{h}_k \mathbf{t}_k^{ZFBF}|^2\big)$.

2.2.3 Multiple Receive Antenna Downlink (MU-MIMO)

In this scenario, each receiver has more than one antenna and hence can receive more than one data stream. Similar to the single antenna downlink case, the feedback strategy involves sending information about the channel direction information and the quality of the channel. Assuming that a Grassmannian codebook $\mathcal{F} = \{\mathbf{F}_i\}_{i=1}^L$ is known to the transmitter and the receiver, the kth receiver selects the index of a codeword according to (Ravindran and Jindal, 2008)

$$\mathbf{F}_{i^\star,k} = \arg\max_{\mathbf{F}_i \in \mathcal{F}} \|\mathbf{H}_k \mathbf{F}_i\|_F^2, \tag{2.7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and sends $i^\star$ to the transmitter. The transmitter reconstructs $\mathbf{F}_{i^\star,k}$, for all k, from the received indices and then constructs the precoding matrix $\mathbf{T} = \big[\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_K\big]$, where $\mathbf{T}_k \in \mathbb{C}^{M_t \times Q_k}$ and $Q_k \leq \min\{M_t, M_{r_k}\}$ is the number of streams transmitted to the kth receiver. Following the dedicated training step the kth receiver can determine $\mathbf{H}_k\mathbf{T}$, and the achievable rate for the kth user, which treats interference as noise, is given by

$$R_k = \log\Big(\frac{\det\big(\mathbf{I}_{M_{r_k}} + \sum_j \frac{E_s}{Q_j}\mathbf{T}_j^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{T}_j\big)}{\det\big(\mathbf{I}_{M_{r_k}} + \sum_{j\neq k} \frac{E_s}{Q_j}\mathbf{T}_j^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{T}_j\big)}\Big) \tag{2.8}$$

In the case that the transmitter performs Block Diagonalization (BD) with equal power loading, each $\mathbf{T}_k$ is chosen such that $\mathbf{F}_{i^\star,j}\mathbf{T}_k^{BD} = 0$, for all $j \neq k$, and so that its columns have unit norm. This can be formed by finding an orthonormal basis for the null space of the matrix constructed by stacking all the $\mathbf{F}_{i^\star,j}$ matrices, $j \neq k$, together. For comparison, in the case of perfect channel state information at the transmitter, $\mathbf{H}_j\mathbf{T}_k^{BD} = 0$, for all $j \neq k$, and hence the achievable rate simplifies to $R_k^{BD-CSIT} = \log\det\big(\mathbf{I}_{M_{r_k}} + \frac{E_s}{Q_k}\mathbf{T}_k^{BD^H}\mathbf{H}_k^H\mathbf{H}_k\mathbf{T}_k^{BD}\big)$.

2.3 Grassmannian Packings

Despite the prominent role that Grassmannian subspace packings play in point-to-point and downlink communication schemes with limited feedback, the design of good packings is widely regarded as a rather difficult problem (e.g., Conway *et al.*, 1996; Love and Heath, Jr., 2005b; Dhillon *et al.*, 2008). The goal of this chapter is to propose effective algorithms for finding good packings and for finding good packings that are constrained to posses a structure that facilitates implementation.

To formalize the discussion, note that for a given dimension $M \leq M_t$, the complex Grassmannian manifold $\mathbb{G}_{M_t,M}$ is the set of all subspaces of dimension M in $\mathbb{C}^{M_t}$. In order to perform quantization on the manifold, a subspace $\mathcal{S}$ of dimension M in $\mathbb{C}^{M_t}$ can be expressed as the linear span of an orthonormal basis; i.e., $\mathcal{S} = \{\mathbf{F}\mathbf{x} | \mathbf{x} \in \mathbb{C}^M\}$, where each column of the matrix $\mathbf{F} \in \mathbb{C}^{M_t \times M}$ is an element of the orthonormal basis. Since the columns of $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{Q}$, where $\mathbf{Q}$ is an $M \times M$ unitary matrix, also form an orthonormal basis for $\mathcal{S}$, all such $\tilde{\mathbf{F}}$ can be deemed to be equivalent in terms of the description of $\mathcal{S}$. If for each subspace $\mathcal{S}$ one such $\tilde{\mathbf{F}}$ is selected as the representative of the equivalence class, then each of these representatives specifies a point on $\mathbb{G}_{M_t,M}$. For the ease of exposition, with a mild abuse of terminology, the set of representatives will be also referred to as a Grassmannian manifold, with the context making the distinction (see also Edelman *et al.*, 1998). The Grassmannian packing problem is the problem of finding a codebook $\mathcal{F} = \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N\}$ of representatives of N subspaces of dimension M in $\mathbb{C}^{M_t}$. The codewords $\mathbf{F}_i \in \mathbb{C}^{M_t \times M}$ have orthonormal columns, and hence such a codebook is often said to be a rank-M codebook. For the case of an unconstrained uniform codebook, the codebook design problem can be posed in terms of maximizing the minimum distance between codeword pairs. That is, find a codebook $\mathcal{F} = \{\mathbf{F}_i\}_{i=1}^N$ that solves:

$$\max_{\{\mathbf{F}_i\},\, \mathbf{F}_i \in \mathbb{G}_{M_t,M}} \; \min_{i \neq j} \; d(\mathbf{F}_i, \mathbf{F}_j), \tag{2.9}$$

where $d(\mathbf{F}_i, \mathbf{F}_j)$ is a measure of the distance between the subspaces spanned by its arguments.

There are many valid distance metrics on the Grassmannian manifold that have been used in the design of packings for a variety of limited feedback applications (e.g., Love *et al.*, 2003; Love and Heath, Jr., 2005b; Dai *et al.*, 2008; Roh and Rao, 2006; Shenouda and Davidson, 2008; Love and Heath, Jr., 2005a; Gohary and Davidson, 2009). Two of the more prominent metrics are the chordal distance,

$$d_{\text{ch}}(\mathbf{F}_i, \mathbf{F}_j) = \frac{1}{\sqrt{2}} \|\mathbf{F}_i \mathbf{F}_i^H - \mathbf{F}_j \mathbf{F}_j^H\|_F = \left(M - \|\mathbf{F}_j^H \mathbf{F}_i\|_F^2\right)^{1/2} \tag{2.10}$$

and the Fubini-Study distance,

$$d_{\text{FS}}(\mathbf{F}_i, \mathbf{F}_j) = \arccos\left|\det\left(\mathbf{F}_j^H \mathbf{F}_i\right)\right|. \tag{2.11}$$

Consistent with the definition of the manifold, these metrics are invariant to the right multiplication of any codeword matrix by a unitary matrix. The chordal distance has been used to quantify the effect of quantization on the achievable rate in limited feedback systems and to derive the bounds on the rate-distortion trade-off (Dai *et al.*, 2008). It has also been used to asses the impact of quantization in point-to-point links that employ unitary precoding of orthogonal space-time block codes (Love and Heath, Jr., 2005a). The Fubini-Study distance has been used to bound the degradation of the achievable rate over an i.i.d. Rayleigh channel caused by quantization (Love and Heath, Jr., 2005b).

In the case in which the subspaces are of dimension $M = 1$, the subspace packing problem in (2.9) collapses to the problem of packing Grassmannian lines and, from the perspective of that problem all the distances become equivalent. For that case, the distance

$$d_{\text{line}}(\mathbf{f}_i, \mathbf{f}_j) = \sqrt{\left(1 - |\mathbf{f}_j^H \mathbf{f}_i|^2\right)}, \tag{2.12}$$

will be used, where $\mathbf{f}_i$ denotes the ith element of a rank-1 codebook.

There are three aspects of the codebook design problem in (2.9) that make it difficult to solve. First, the constraint that the codewords $\mathbf{F}_i$ lie on the manifold is non-convex. Second, the min operator is not differentiable, and third, many distance metrics for the manifold are both non-convex and non-differentiable. These aspects suggest that the basic problem in (2.9) may be difficult to solve, even before seeking to modify it to obtain codebooks with additional features. Given these difficulties, it can be useful to have an upper bound on the optimal value of the problem in (2.9) so that the quality of potentially suboptimal solutions can be assessed.

If $\delta(M_t, M, N)$ denotes the optimal value of the problem in (2.9) for a particular distance metric, then for the case of the chordal distance in (2.10) the Rankin bounds can be used, i.e., the simplex and the orthoplex bounds, to evaluate the quality of a codebook (Conway *et al.*, 1996; Barg and Nogin, 2002; Henkel, 2005; Pitaval *et al.*, 2011b). The simplex bound is defined as

$$\delta_{\mathrm{ch}}(M_t, M, N) \leq \sqrt{\frac{M(M_t - M)N}{M_t(N-1)}}. \tag{2.13}$$

This bound is attainable only if $N \leq M_t^2$. For $N > M_t^2$, the orthoplex bound

$$\delta_{\mathrm{ch}}(M_t, M, N) \leq \sqrt{\frac{M(M_t - M)}{M_t}} \tag{2.14}$$

is a tighter bound, and this bound is attainable only if $N \leq 2(M_t^2 - 1)$. These bounds are also valid in the case of line packing; with mild abuse of notation, $\delta_{\mathrm{line}}(M_t, N) \leq \sqrt{\frac{(M_t-1)N}{M_t(N-1)}}$ for $N \leq M_t^2$ and $\delta_{\mathrm{line}}(M_t, N) \leq \sqrt{\frac{(M_t-1)}{M_t}}$ for $N > M_t^2$. This chapter will also consider codebooks that are constrained so that each element of each codeword has the same modulus; i.e., for $i = 1, 2, \ldots, N$, $|[\mathbf{F}_i]_{\ell m}|^2 = \frac{1}{M_t}$ for $\ell = 1, 2, \ldots, M_t, m =$

$1, 2, \ldots, M$. In this case, the ranges of the bounds can be refined (Pitaval *et al.*, 2011b). In particular for such constrained codebooks, the simplex bound is attainable only if $N \leq M_t^2 - M_t + 1$, while the orthoplex bound is attainable only if $N \leq 2(M_t^2 - M_t)$.

2.4 Subspace Packing Design via Sequential Smooth Optimization

The proposed strategy for constructing an effective technique for finding good solutions to the codebook design problem in (2.9) is to use algorithms developed for optimization of smooth functions on the Grassmannian manifold (Edelman *et al.*, 1998; Manton, 2002). These algorithms are adaptations to the manifold of conventional algorithms for unconstrained optimization of smooth functions, such as gradient descent, conjugate gradient and Newton's method, and so in order to use them a smooth approximation of the objective needs to be constructed. Since the Grassmannian manifolds of interest lie in a complex-valued ambient space, the notion of smooth approximation needs to be clarified. The notion that will be used herein is that the approximation has continuous derivatives with respect to the real and imaginary parts of the variable (cf. Kreutz-Delgado, 2009). In the following discussions, the Fubini-Study distance and the chordal distance will be considered as they often arise in limited feedback systems. That said, the principles that underlie the suggested procedures can be applied to any valid subspace distance or smoothed version thereof.

As a first step, the proposed approach defines $\tilde{d}_{\mathrm{FS}}(\mathbf{F}_i, \mathbf{F}_j) = \left|\det\left(\mathbf{F}_j^H \mathbf{F}_i\right)\right|$ and

$\tilde{d}_{\text{ch}}(\mathbf{F}_i, \mathbf{F}_j) = \|\mathbf{F}_j^H \mathbf{F}_i\|_F^2$. With these metrics, the problem in (2.9) can be rewritten as:

$$\min_{\{\mathbf{F}_i\}, \mathbf{F}_i \in \mathbb{G}_{M_t,M}} \max_{i \neq j} \tilde{d}(\mathbf{F}_i, \mathbf{F}_j). \tag{2.15}$$

In the case of the chordal distance, $\tilde{d}_{\text{ch}}(\mathbf{F}_i, \mathbf{F}_j)$ is smooth and one can move directly to dealing with the non-smoothness of the max operation in (2.15). In the Fubini-Study case, the magnitude operator in $\tilde{d}_{\text{FS}}(\mathbf{F}_i, \mathbf{F}_j)$ is not smooth, and hence a smooth approximation has to be developed. By employing some of the existing smooth approximations for the magnitude operator (e.g., Bertsekas 1999, Sec 5.4.5; Nesterov 2005), and by using $\mathbf{X}$ to denote $\mathbf{F}_j^H \mathbf{F}_i$, the following approximations of $|\det(\mathbf{X})|$ are obtained:

$$\sqrt{\mu^2 + \det(\mathbf{X}^H \mathbf{X})} - \mu, \tag{2.16}$$

$$\mu \log\Big(2 \cosh\Big(\sqrt{\det\left(\mathbf{X}^H \mathbf{X}\right)}/\mu\Big)\Big), \tag{2.17}$$

$$\frac{2}{\pi}\sqrt{\det(\mathbf{X}^H \mathbf{X})}\, \tan^{-1}\Big(\mu\sqrt{\det(\mathbf{X}^H \mathbf{X})}\Big), \tag{2.18}$$

where μ is an appropriate constant, and for a non-negative real number x the derivative of $\sqrt{x}$ at $x = 0$ is defined as the limit of that derivative as x approaches zero from above. Other approximations include the Huber penalty function (e.g., Boyd and Vandenberghe, 2004) and

$$\big(1 + \det\left(\mathbf{X}^H \mathbf{X}\right)\big) \log\big(1 + \det(\mathbf{X}^H \mathbf{X})\big). \tag{2.19}$$

The choice among these approximations does not affect the principles of the proposed approach, and hence, for simplicity, let $\hat{\tilde{d}}_{\mathrm{FS}}(\cdot,\cdot)$ denotes a chosen smooth approximation of $\tilde{d}_{\mathrm{FS}}(\cdot,\cdot)$. For notational convince, set $\hat{\tilde{d}}_{\mathrm{ch}}(\cdot,\cdot) = \tilde{d}_{\mathrm{ch}}(\cdot,\cdot)$. Having said that, the different shapes of the approximations do have an impact in terms of the convergence behavior of the proposed design approach. For instance, the approximation in (2.19) is quite flat around the origin and quite steep away from the origin. As a result, in numerical experiments that are not reported here, the approximation in (2.19) tended to generate good solutions in a few iterations, but the refinement of those good solutions tended to be slow. The approximation in (2.17) has a reasonably consistent slope and in the conducted experiments, it tended to refine good codebooks more quickly than the approximation in (2.19). However in the absence of a good initialization that approximation can become numerically unwieldy. Since the goal is to develop techniques that yield good codebooks with a modest effort, the approximation in (2.19) will be used in the designs, but techniques that switch between approximations can be implemented in a straightforward manner.

There are numbers of ways in which the expression $\max_{i\neq j} \hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j)$ in (2.15) can be smoothly approximated. One is to use the approximation $\max\{a, b\} \approx \log\left(e^a + e^b\right)$. (The reverse approximation is similar to the "max-log" approximation that is often used in soft decoding algorithms.) This approximation was successfully used by Gohary and Davidson (2009) for the construction of Grassmannian constellations for non-coherent MIMO communication and is also applicable here. An alternative is to use the approximation $\max_{i\neq j} \hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j) \approx \mu \log\big(\frac{1}{N(N-1)} \sum_i \sum_{j\neq i} \cosh\big(\hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j)/\mu\big)\big)$, for some $\mu > 0$, (Nesterov, 2005). As in the case of the distance metric, the proposed approach can incorporate a number of different smooth approximations, and approximations of $\max_{i \neq j} \hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j)$ of the form

$$J_1\big(\{\mathbf{F}_i\}\big) = \Big(\sum_{i \neq j} \hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j)^{\beta}\Big)^{1/\beta}, \tag{2.20}$$

will be employed, where $\beta \geq 1$. That is, the ∞–norm of the vector of pair-wise metrics is approximated by its β–norm. (For some distances, the construction and optimization of $J_1(\cdot)$ can be simplified by restricting β to be even.)

Even though the problem

$$\min_{\{\mathbf{F}_i\},\, \mathbf{F}_i \in \mathbb{G}_{M_t, M}} J_1\big(\{\mathbf{F}_i\}\big) \tag{2.21}$$

is smooth, it remains non-convex, due, in part, to the nature of the manifold. Nevertheless, a straightforward approach for using (2.21) to obtain good solutions to (2.15), and to the original problem in (2.9), would be to select a value for β and an initial codebook, and seek a locally optimal solution by applying an adaptation of a conventional smooth optimization technique, such as gradient descent, conjugate gradient or Newton's method, to the manifold (Edelman *et al.*, 1998; Manton, 2002); see also Gohary and Davidson (2011). The suggested approach enhances that approach by solving a sequence of problems with increasing values of β. Once a good codebook for the current problem has been found, that codebook is used to initialize the next problem. This sequential procedure takes advantage of the better conditioning of $J_1(\cdot)$ for smaller values of β, and the improved approximation for larger values of β.

In particular, the basic procedure is as follows:

Basic Sequential Optimization Procedure

1. Choose the initial value of β, denoted β_0, and the increment δ_β in β at each iteration.

2. Set $\beta = \beta_0$ and randomly select an initial codebook of matrices with orthonormal columns.

3. Starting from the codebook obtained in the previous iteration, obtain a good solution to (2.21) using an algorithm for smooth unconstrained optimization on the manifold (Edelman *et al.*, 1998; Manton, 2002).

4. Evaluate the quality of the codebook against a known bound (if any), and/or evaluate the progress of the algorithm in terms of the increase in the minimum distance.

5. Terminate if desired, else $\beta \leftarrow \beta + \delta_\beta$ and return to step 3.

In the implementations, β is chosen to be $\beta_0 = 2$, $\delta_\beta = 1$ or 2, and in step 3 the gradient-based algorithm of Manton (2002) is employed. The numerical experiments have provided good empirical evidence for monotonic convergence of the procedure, but the question of whether convergence is guaranteed remains open.

There are several ways in which the cost function $J_1(\cdot)$ can be modified to improve the basic procedure. In the case that an upper bound on the minimum distance is known, such as the Rankin bound on the chordal distance, $J_1(\cdot)$ can be replaced by

$$J_2\big(\{\mathbf{F}_i\}\big) = \Big(\sum_{i \neq j} \big(\hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j) - \alpha\big)^{\beta}\Big)^{1/\beta}, \tag{2.22}$$

where the constant α is set to the known bound. For the case of the Fubini-Study distance, and other metrics for which non-trivial bounds are not known, the value of the constant α can be adapted to the outcome of the previous iteration of the sequential procedure. For example, the value of α can be set to the average of the pair-wise distances obtained at the previous iteration. The numerical experiments have suggested that employing $J_2(\cdot)$ with α set to a known bound or adapted to the previous iteration, tends to improve the convergence properties and the quality of the generated packings. This idea is similar to idea presented by Conway *et al.* (1996) for constructing packings in 4-dimensional space.

An important property of the Grassmannian manifold that simplifies the design process for codebooks of different ranks is the isomorphism between the Grassmannian manifolds $\mathbb{G}_{M_t,M}$ and $\mathbb{G}_{M_t,M_t-M}$. Accordingly, an optimal codebook of rank M can be used to synthesize an optimal codebook of rank $M_t - M$ as follows: Given a codebook $\mathcal{F} = \{\mathbf{F}_i\}_{i=1}^N, \mathbf{F}_i \in \mathbb{G}_{M_t,M}$ that is optimal for a given distance metric

1. For $i = 1, 2, \ldots N$, find a matrix $\mathbf{K}_i$ whose columns form an orthonormal basis for the null space of the codeword $\mathbf{F}_i$.

2. Each $\mathbf{K}_i$ is an element of $\mathbb{G}_{M_t,M_t-M}$, and the codebook $\bar{\mathcal{F}} = \{\mathbf{K}_i\}_{i=1}^N$ is optimal in the same distance metric sense.

One conclusion that can be inferred from this property is that all the optimal codebooks on $\mathbb{G}_{3,M}$ are equivalent under any distance metric. That is, an optimal codebook on $\mathbb{G}_{3,1}$, which is the solution of a line packing problem, can be used to generate an optimal codebook on $\mathbb{G}_{3,2}$. Hence optimal chordal distance and Fubini Study distance based designs on $\mathbb{G}_{3,2}$ will result in codebooks with the same minimum distance. More generally, this conclusion can be applied to codebook pairs on $\mathbb{G}_{M_t,M}$ and $\mathbb{G}_{M_t,M_t-M}$.

Table 2.1: Minimum Fubini-Study distances of unconstrained codebooks.

$M_t \times M$	N	Designed Codebook	Codebook of Love (2004)
4×2	4	1.5708	1.2451
4×2	8	1.3444	1.0414
4×2	16	1.2309	0.8654
4×2	64	0.9613	0.6059
6×3	8	1.5594	N/A
6×3	16	1.5261	1.1936
6×3	32	1.4187	1.0724
6×3	64	1.3514	0.9722
8×2	8	1.5707	N/A
8×2	16	1.5414	N/A
8×2	32	1.4738	1.3153
8×2	64	1.4084	N/A

In order to assess the basic design procedure, in Table 2.1 the minimum Fubini-Study distances of some of the codebooks that have been designed are compared to those of the corresponding codebooks of Love (2004). That comparison shows that the designed codebooks have significantly larger minimum distances than those of Love (2004). Indeed, in the case of the 4×2 codebook of size 4, the basic procedure achieves the upper bound on the Fubini-Study distance of $\pi/2$; cf. (2.11). The results reported in the first two rows of Table 2.1 can also be compared to the results for codebooks designed using the method of Dhillon *et al.* (2008); see Table 6 in that paper. In those cases, the basic procedure produces codebooks with the same minimum distances as those of Dhillon *et al.* (2008). However, as evidenced by Table A.1 in Appendix A, the basic procedure has enabled the design of codebooks for systems with larger dimensions. The basic procedure has been also used to design codebooks for the chordal distance, and in that case the minimum distances of the designed codebooks are essentially the same as those of Dhillon *et al.* (2008).

2.5 Line Packings with Constant Modulus

In scenarios in which at most one data stream is sent to each user, the generic codebook design problem in (2.9) reduces to the line packing problem of finding a codebook of vectors $\mathbf{f}_i \in \mathbb{G}_{M_t,1}$, that are maximally separated with respect to the distance metric $d_{\text{line}}(\mathbf{f}_i, \mathbf{f}_j) = (1 - |\mathbf{f}_j^H \mathbf{f}_i|^2)^{1/2}$. As in the previous section, the principle of sequential smooth approximation of the problem in (2.15) can be invoked in a number of different ways. For reasons analogous to those that led to the selection of the approximation in (2.19), one way that has been found to be particularly effective is to choose $\hat{\tilde{d}}_{\text{line}}(\mathbf{f}_i, \mathbf{f}_j) = \bigl(1 + |\mathbf{f}_j^H \mathbf{f}_i|^2\bigr) \log \bigl(1 + |\mathbf{f}_j^H \mathbf{f}_i|^2\bigr)$, and to apply the basic sequential design procedure to the problem of minimizing $J_2(\{\mathbf{f}_i\})$ in (2.22), where α is set to the value of the corresponding Rankin bound. That procedure produces codebooks that exhibit essentially the same distance properties as the best of the existing codebooks (e.g., Dhillon *et al.*, 2008). In this section the basic sequential procedure is used to obtain codebooks with similar distance properties and the additional property that the elements of the codewords have (essentially) constant modulus. In feedback schemes for point-to-point links, this property ensures that the power transmitted from each antenna is (almost) the same and hence mitigates the effects of power imbalances at the transmitter.

For the line packing case, the constant modulus constraint is that the ℓth element of each vector $\mathbf{f}_i$ has modulus $1/\sqrt{M_t}$; i.e., $\bigl|[\mathbf{f}_k]_\ell\bigr| = 1/\sqrt{M_t}$. However, the basic design procedure is based on unconstrained optimization on the manifold. In order to use that procedure the smooth penalty term is defined as

$$P_{\text{cm}}\bigl(\{\mathbf{f}_i\}\bigr) = \Bigl(\sum_{k,\ell} \bigl(\bigl|[\mathbf{f}_k]_\ell\bigr|^2 - \tfrac{1}{M_t}\bigr)^\beta\Bigr)^{1/\beta}, \tag{2.23}$$

where $\beta \geq 1$, and apply the basic sequential optimization procedure with the cost function

$$J_3\big(\{\mathbf{f}_i\}\big) = w_1 J_2\big(\{\mathbf{f}_i\}\big) + w_2 P_{\mathrm{cm}}\big(\{\mathbf{f}_i\}\big) \tag{2.24}$$

for appropriately chosen weights w_1 and w_2. Since there are only two weights, a variety of line search strategies (e.g., **?**) can be used to select an appropriate value for the ratio of the weights. Perhaps the simplest strategy would be to augment the basic sequential procedure with an additional loop in which w_2 is initially set to zero and then increased sequentially until a codebook whose elements are sufficiently close to being constant modulus is obtained. As in the basic procedure, the optimization at the current iteration would be initialized by the codebook obtained from the previous iteration.

Table 2.2 compares the minimum distances of the unconstrained and constant-modulus Grassmannian line packings that have been obtained with those of the unconstrained line packings presented of Dhillon *et al.* (2008). For the unconstrained case, the proposed technique generates line packings with approximately the same minimum distances as those reported by Dhillon *et al.* (2008). In the constrained case, the constructed packings have elements with almost constant modulus, and have minimum distances that are very close to the unconstrained case.

2.6 Line Packings with PSK Alphabet

In this section, the focus is on the design of line packings with codeword elements that are not only of constant modulus, but are also from a finite alphabet, such as a scaled phase-shift keying (PSK) alphabet, $\mathcal{A} = 1/\sqrt{M_t}\left\{e^{j(\phi_0 + 2\pi(\ell-1)/L)}\right\}_{\ell=1}^{L}$. This property

Table 2.2: Minimum distances of unconstrained and constant modulus rank-1 codebooks, along with the Rankin bound. For the constant modulus packings $\max_{k,\ell} \left| |[\mathbf{f}_k]_\ell - 1/M_t|^2 \right| \leq 0.01/M_t$.

M_t	N	Designed Codebook	Unconstrained Codebook of Dhillon *et al.* (2008)	Rankin Bound	Const. Mod. Designed Codebook
4	5	0.9682	0.9682	0.9682	0.9680
4	6	0.9448	0.9448	0.9486	0.9441
4	7	0.9350	0.9353	0.9354	0.9188
4	8	0.9255	0.9257	0.9258	0.9114
4	9	0.9155	0.9150	0.9186	0.8969
4	10	0.9115	0.9114	0.9129	0.8977
4	16	0.8943	0.8943	0.8943	0.8656
4	20	0.8659	0.8458	0.8660	0.8240
5	6	0.9798	0.9798	0.9798	0.9797
5	7	0.9638	0.9637	0.9661	0.9522
5	8	0.9553	0.9553	0.9562	0.9444
5	9	0.9466	0.9468	0.9487	0.9350
5	10	0.9427	0.9427	0.9428	0.9354
5	16	0.9190	0.9183	0.9238	0.9145

significantly reduces the storage requirements of the codebook and may reduce the computational effort required to select the codeword at the receiver. In particular, in the case of the scaled 4–PSK alphabet, $1/\sqrt{M_t}\,\{1, -1, j, -j\}$, the scaling can be absorbed and the complex multiplications that are inherent in the selection process are reduced to sign changes and swaps of the real and imaginary parts.

The restriction to a defined alphabet $\mathcal{A}$ offers the possibility to design codebooks by exhaustively evaluating each admissible codebook. However, taking into account the rotational invariance of the codewords into account, there are $\binom{|\mathcal{A}|^{M_t-1}}{N}$ admissible codebooks and even for modestly sized codebooks the computational cost of this approach exceeds the computational resources that could reasonably be applied to the problem. This section describes two simple approaches to obtaining good codebooks with elements from a (scaled) PSK alphabet. In the first approach, the method of the previous section is used to generate partial codebooks with (scaled) PSK elements and good properties. These partial codebooks are then completed by exhaustive search, but that exhaustive search is typically over a much smaller number of codewords than the size of the codebook. In the second approach, an incremental construction is employed in which the size of the codebook is doubled at each step. The new codewords are Hadamard products of the existing vectors and a single vector that is obtained using a smooth optimization and rounding procedure similar to that used in the first approach. This incremental construction was motivated, in part, by some of the principles used in the generation of mutually unbiased bases (MUB) with elements from the (scaled) 4–PSK alphabet (Mondal *et al.*, 2007; Inoue and Heath Jr., 2009; Tselniker *et al.*, 2009).

2.6.1 Relax-round-expurgate-replace Approach

The first of the proposed approaches begins with the application of a variant of the method in Section 2.5 to design an initial codebook of codewords with constant modulus elements. Those elements are then rounded to the given alphabet. The "bad" codewords in this rounded codebook are expurgated and are replaced through an exhaustive search procedure. The selection of the number of codewords to be expurgated, and hence the size of the subsequent search is based on the notion of a satisfactory codebook.

Given that Grassmannian line packings are being considered, one way in which the quality of a finite alphabet codebook could be assessed would be to compare the achieved minimum distance to the Rankin bound. That approach can be refined by observing that the distance spectrum of a finite alphabet codebook is discrete and the set of distance values depends on the alphabet being used. A refined upper bound on the minimum distance is the largest of these discrete values that is no larger than the Rankin bound. As an example, the set of possible distances between pairs of codewords of a codebook of size 8 that resides on $\mathbb{G}_{4,1}$ and whose elements are selected from the (scaled) 4–PSK alphabet can be shown to be $\{1, 0.9354, 0.8660, 0.7071, 0.6123, 0\}$. The Rankin bound for a codebook of size 8 on $\mathbb{G}_{4,1}$ is 0.92582. Comparing the Rankin bound with the set of admissible distances, it can be deduced that the minimum distance of the (scaled) 4–PSK codebook is bounded above by 0.8660. A satisfactory codebook would achieve a large fraction of this "quantized" Rankin bound.

A more detailed description of the proposed approach is as follows: The first codeword of the codebook is chosen to be a randomly generated vector with elements from the scaled PSK alphabet. Then the finite alphabet constraint is relaxed on the remaining $N-1$ codewords and the procedure in Section 2.5 is used to design those $(N-1)$ codewords so that a good codebook of size N with (essentially) constant modulus elements is obtained. The elements of the $(N-1)$ designed codewords in that codebook are then rounded to the nearest point in the alphabet. This rounding process could be done in a more sophisticated way, but element-wise rounding of each codeword is simple to implement and has sufficed in the numerical examples. The quantized codebook is then analyzed to determine whether its minimum distance achieves a sufficiently large fraction of the quantized Rankin bound to be deemed satisfactory. If not, the codeword that appears in the largest number of pairwise distances that are deemed unsatisfactory is removed, and the resulting codebook is analyzed again. This process is repeated until a satisfactory codebook is found. If $\bar{N}$ denotes the number of codewords that are expurgated in this way, the satisfactory partial codebook, which is of size $N - \bar{N}$, is completed by adding $\bar{N}$ codewords via an exhaustive search. Typically, $\bar{N} \ll N$ and this reduced-dimension exhaustive search is often viable. In cases in which $\bar{N}$ is deemed to be too large to attempt an exhaustive search, the basic relaxation and rounding aspects of the proposed approach can be recapitulated, but with the existing $N - \bar{N}$ codewords being fixed, rather than the single codeword that was fixed in the first stage.

The above procedure is used to generate 16 lines in the 4 dimensional complex space using the (scaled) 4–PSK alphabet. The generated packings have a minimum distance of 0.8660, while the corresponding codebook reported by Clerckx *et al.* (2008) has a minimum distance of 0.7071. The codebooks generated incrementally using the notion of mutually unbiased bases (Mondal *et al.*, 2007; Inoue and Heath Jr., 2009) also achieve a minimum distance of 0.8660. As noted above, the quantized Rankin bound in this case is 0.8660. Hence the codebooks that have been obtained, and those by Mondal *et al.* (2007) and Inoue and Heath Jr. (2009), are essentially optimal.

2.6.2 Incremental Construction

The good performance of the incremental construction based on mutually unbiased bases (MUBs) (Mondal *et al.*, 2007; Inoue and Heath Jr., 2009; Tselniker *et al.*, 2009) in the previous example suggests that the MUB construction might also be of interest in other scenarios. While that is the case to some degree, in its raw form the incremental MUB construction is only applicable to cases in which the number of transmitter antennas, M_t, is a power of two. In this section, the earlier work is exploited to develop an incremental construction that is based on some of the ideas that underlie the MUB construction, but is more flexible in that it can be applied to systems with an arbitrary number of antennas. Furthermore, at each step the proposed approach doubles the size of the codebook, whereas the MUB construction only adds M_t codewords. Since the computational cost of each step is typically lower in the proposed approach, this facilitates the construction of sizeable codebooks with elements from (scaled) PSK alphabets.

In order to place the proposed approach in context, the existing MUB-based construction is first briefly described (Mondal *et al.*, 2007), which can be applied to systems for which M_t is a power of two: To design a codebook of size $N = (\check{n}+1)M_t$ in $\check{n}$ steps,

1. Let $\mathbf{A}^{(0)}$ to be the normalized Hadamard matrix of size $M_t \times M_t$, and let $\mathbf{a}_i^{(0)}$ denote its ith column. Set $n = 0$.

2. While $n < \check{n}$, increment n and find a length-M_t vector $\mathbf{u}^{(n)}$ with elements $\mathcal{A} = 1/\sqrt{M_t}\,\{\pm 1, \pm j\}$ such that all the elements of the vectors $\mathbf{A}^{(n-1)^H}\mathbf{u}^{(n)}$, $\mathbf{A}^{(n-2)^H}\mathbf{u}^{(n)}$, ..., $\mathbf{A}^{(0)^H}\mathbf{u}^{(n)}$ lie in $\mathcal{A}$.

3. Construct $\mathbf{A}^{(n)} = \sqrt{M_t}\left[\mathbf{u}^{(n)} \odot \mathbf{a}_1^{(0)}, \mathbf{u}^{(n)} \odot \mathbf{a}_2^{(0)}, \ldots, \mathbf{u}^{(n)} \odot \mathbf{a}_{M_t}^{(0)}\right]$, where $\odot$ denotes the element-wize (Hadamard) product, and return to 2.

4. The desired codebook is the columns of the matrix $\left[\mathbf{A}^{(0)}, \mathbf{A}^{(1)}, , \ldots, \mathbf{A}^{(\check{n})}\right]$.

In conventional implementations, the vector $\mathbf{u}^{(n)}$ in step 2) is found using an exhaustive search.

In the proposed approach, a codebook of size $N = 2^{\check{q}}N_0$ is designed in $\check{q}$ steps, where N_0 is the number of codewords in the initial codebook. To do so, the above steps are modified in a number of ways:

- The proposed approach is initialized by a matrix $\mathbf{B}^{(0)}$ of size $M_t \times N_0$ with elements from the scaled PSK alphabet whose columns yield a satisfactory codebook; cf. Section 2.6.1. In the implementations, when M_t is a power of two $\mathbf{B}^{(0)}$ is chosen to be the normalized Hadamard matrix of the appropriate size, in which case, $N_0 = M_t$. In other cases $N_0 = 1$ and $\mathbf{B}^{(0)} = (1/\sqrt{M_t})\,\mathbf{1}$, where $\mathbf{1}$ denotes the length-M_t vector with all elements equal to 1.

- Similar to step 2) in the existing method, at each iteration, starting from $q = 1$, the proposed approach seeks a vector $\mathbf{v}^{(q)}$ with elements in $\mathcal{A}$ such that the vector ${\mathbf{B}^{(q-1)}}^H\mathbf{v}^{(q)}$ has all its elements in $\mathcal{A}$. This can be done using a variant of the procedure in Section 2.6.1 in which the finite alphabet constraint is first relaxed and the basic sequential optimization procedure in Section 2.4 is applied to
$$J_4(\mathbf{v}) = \omega_1 P_{\mathrm{cm}}\big({\mathbf{B}^{(q-1)}}^H\mathbf{v}\big) + \omega_2 P_{\mathrm{cm}}(\mathbf{v}), \tag{2.25}$$
where $P_{\mathrm{cm}}(\cdot)$ was defined in (2.23). Once a good solution to that problem has been obtained, the elements of the resulting $\mathbf{v}$ are rounded to the nearest member of the alphabet to form a candidate for $\mathbf{v}^{(q)}$. That candidate is then used to generate a candidate for the incremented codebook, as described below.

- Distinct from the existing method, the candidate vector $\mathbf{v}^{(q)}$ is used to generate the vectors $\mathbf{w}_i = \sqrt{M_t}\mathbf{v}^{(q)} \odot \mathbf{b}_i^{(q-1)}$, $i = 1, 2, \ldots, 2^{(q-1)}N_0$, where $\mathbf{b}_i^{(q-1)}$ is the ith column of $\mathbf{B}^{(q-1)}$. A candidate codebook consisting of the union of these $2^{(q-1)}N_0$ new vectors and the existing codebook is evaluated. If it is deemed to be satisfactory (cf. Section 2.6.1), the candidate codebook becomes the incremented codebook and is represented by the columns of the matrix $\mathbf{B}^{(q)} = \left[\mathbf{B}^{(q-1)}, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2^{(q-1)}N_0}\right]$. Otherwise, the proposed approach seeks a new candidate for $\mathbf{v}^{(q)}$. One can be found by picking a vector with elements from $\mathcal{A}$ that lies in the neighborhood of the existing candidate, or by returning to the previous step and optimizing the expression in (2.25) from a different starting point. (Obviously, the search for new candidates for $\mathbf{v}^{(q)}$ must be controlled so that all possibilities are eventually explored.)

In applications in which the space required to store the codewords is limited, incremental constructions such as the MUB-based construction and that proposed above have the advantage that when $\mathbf{A}^{(0)}$, respectively $\mathbf{B}^{(0)}$, can be generated deterministically, one need only store the sequence of vectors $\mathbf{u}^{(n)}$, $n = 1, 2, \ldots, \check{n}$, respectively $\mathbf{v}^{(q)}$, $q = 1, 2, \ldots, \check{q}$. For a codebook of size $N = (\check{n}+1)M_t$, the MUB approach needs to store $\check{n}$ vectors, whereas for a codebook of size $N = 2^{\check{q}}N_0$, the proposed approach need only store $\check{q}$ vectors. Furthermore, since the size of the codebook that has been obtained doubles at each iteration, and since the relax-round approach typically finds an appropriate $\mathbf{v}^{(q)}$ quite quickly, for large codebooks the approach had a significant computational advantage in the numerical experiments. As such the proposed approach has been able to design significantly larger codebooks than those that have been previously obtained using the MUB approach.

Table 2.3 presents the minimum distances of some (scaled) 4-PSK line packings designed with the proposed incremental approach. Many of these packings are optimal, in the sense that they achieve the “quantized” Rankin bound. In the case of $M_t = 4$ and $N = 32$ the designed codebook does not achieve the quantized Rankin

Table 2.3: Minimum distances of rank-1 codebooks with (scaled) 4–PSK alphabet, along with the Rankin bound for constant modulus codebooks and an indication of whether the designed codebook is known to be optimal.

M_t	N	Designed PSK Codebook	Rankin Bound	Optimal?
4	8	0.8660	0.9258	Yes
4	16	0.8660	0.8660	Yes
4	32	0.7071	0.8660	Yes
5	8	0.8944	0.9562	Yes
5	16	0.8944	0.9238	Yes
5	32	0.8000	0.8944	?
6	8	0.9428	0.976	Yes
6	16	0.9428	0.9428	Yes
6	32	0.8819	0.9129	Yes
7	8	0.9897	0.9897	Yes
7	16	0.9035	0.9562	?
7	32	0.9035	0.9406	Yes
8	16	0.9354	0.9661	Yes
8	32	0.9354	0.9504	Yes
8	64	0.9354	0.9354	Yes
16	32	0.9682	0.9837	?
16	64	0.9682	0.9759	Yes
16	128	0.9682	0.9720	Yes
16	256	0.9682	0.9682	Yes

bound, but it can be shown to be optimal in the minimum distance sense via an exhaustive search. In the other three cases in Table 2.3 in which the designed codebook does not achieve the "quantized" Rankin bound, the optimality, or otherwise, of the obtained codebooks remains to be determined.

Interestingly, when M_t is a power of two, the codewords in the packings generated by the proposed technique can be arranged into sets of mutually unbiased bases. In particular, the union of the packing with $M_t = 16$ and $N = 256$ and an identity matrix generates a complete set of mutually unbiased bases for dimension 16; a result that may be of independent interest.

2.7 Subspace Packings with Nested Structure

In a number of scenarios, substantial performance gains can be made by enabling "multi-mode" precoding (Love and Heath Jr., 2005), in which the number of data streams to be transmitted to a given receiver is adapted to the state of the channel, in addition to the directions in which those streams are sent. To facilitate the implementation of a multi-mode system, it is desirable that the codebook have a nested structure, in which a codebook of higher rank contains codebooks of lower rank. For example, a nested rank-2 codebook, $\mathcal{F}_2 = \{\mathbf{F}_i\}_{i=1}^N = \{[\mathbf{f}_i, \check{\mathbf{f}}_i]\}_{i=1}^N$ contains the rank-1 codebook $\mathcal{F}_1 = \{\mathbf{f}_i\}_{i=1}^N$. The design of a nested codebook intrinsically involves making trade-offs between the quality of the codebook of each rank. To capture that trade-off without adding constraints, the design of a nested codebook is formulated using a weighted sum of metrics. For the case of rank-2 nested codebooks, that design problem can be written as

$$\max_{\{\mathbf{F}_i\},\, \mathbf{F}_i \in \mathbb{G}_{M_t,2}} \min_{i \neq j} \; \omega_2 d_2\big(\mathbf{F}_i, \mathbf{F}_j\big) + \omega_1 d_1(\mathbf{f}_i, \mathbf{f}_j), \tag{2.26}$$

where $\mathbf{f}_i$ denotes the first column of $\mathbf{F}_i$, $d_m(\cdot,\cdot)$ is any valid subspace distance on $\mathbb{G}_{M_t,m}$, and ω_m is the weight assigned to the distance metric for the codebook of rank m. (In practice, the distance metrics for the rank-1 and rank-2 codebooks would likely be chosen to be of the same type, but the approach can also handle metrics of different types.) The extension to nested codebooks of higher rank is conceptually straightforward but somewhat awkward from a notational perspective. By defining smooth functions $\hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j)$ and $\hat{\tilde{d}}_{\text{line}}(\mathbf{f}_i, \mathbf{f}_j)$ as in Section 2.4, good solutions to (2.26) can be obtained by applying the basic sequential optimization procedure.

In particular, if the dimensions of the argument of $J_2(\cdot)$ in (2.22) implicitly define the appropriate metric and if $\mathbf{f}_i$ denotes the first column of $\mathbf{F}_i$, then the basic sequential optimization procedure can be applied to

$$J_5\big(\{\mathbf{F}_i\}\big) = \omega_2 J_2(\{\mathbf{F}_i\}) + \omega_1 J_2(\{\mathbf{f}_i\}), \tag{2.27}$$

where $J_2(\{\mathbf{F}_i\}) = \big(\sum_{i\neq j} \big(\hat{\tilde{d}}(\mathbf{F}_i, \mathbf{F}_j) - \alpha_2\big)^{\beta}\big)^{1/\beta}$, $J_2(\{\mathbf{f}_i\}) = \big(\sum_{i\neq j} \big(\hat{\tilde{d}}_{\text{line}}(\mathbf{f}_i, \mathbf{f}_j) - \alpha_1\big)^{\beta}\big)^{1/\beta}$, with α_2 and α_1 set or adapted as described in Section 2.4 and the Grassmannian optimization steps are over $\mathbb{G}_{M_t,2}$. For simplicity, the same notation for the weights in (2.26) and (2.27) have been used, but they may be adjusted if desired.

Table 2.4 presents some results from the application of this approach to the design of nested codebooks of rank two in which the Fubini Study distance is chosen as the subspace metric. The weights ω_1 and ω_2 were chosen so as to achieve a balance between the properties of the rank-1 and rank-2 codebooks. The third and fourth columns of the table contain the minimum distance achieved by the rank-1 component of the nested codebook and the (unconstrained) Rankin bound for rank-1 codebooks, respectively. The fifth column contains the minimum Fubini-Study distance of the designed nested codebook, and the sixth column contains the best minimum Fubini-Study distance that has been obtained for unstructured rank-2 codebooks. (Some of those distances appear in Table 2.1.) Table 2.5 presents corresponding results for nested codebooks with the chordal distance as the subspace metric. In that table, the comparator for the quality of the rank-2 codebook is best of the corresponding minimum distances reported by Dhillon *et al.* (2008), where corresponding results are available.

Table 2.4: Minimum distances achieved by designed nested codebooks with Fubini-Study distance as the rank-2 distance. The Rankin bound of the rank-1 case and the minimum Fubini-Study distance of the designed unstructured rank-2 codebooks are provided for comparison.

$M_t \times M$	N	Rank 1		Rank 2, Fubini-Study dist.	
		Designed Nested	Rankin Bound	Designed Nested	Designed Unstructured
4×2	8	0.9118	0.9258	1.3069	1.3444
4×2	16	0.8594	0.8944	1.1659	1.2309
4×2	32	0.7687	0.8660	0.9689	1.1159
6×2	8	0.9740	0.9759	1.5245	1.5584
6×2	16	0.9358	0.9428	1.3855	1.4812
6×2	32	0.8980	0.9275	1.3258	1.3636
8×2	8	1	1	1.5282	1.5642
8×2	16	0.9629	0.9661	1.4829	1.5414
8×2	32	0.9388	0.9504	1.4203	1.4738

Table 2.5: Minimum distances achieved by designed nested codebooks with chordal distance as the rank-2 distance. The Rankin bound of the rank-1 case and the minimum chordal distance of the unstructured rank-2 codebooks reported by Dhillon *et al.* (2008) are provided for comparison.

$M_t \times M$	N	Rank 1		Rank 2, chordal dist.	
		Designed Nested	Rankin Bound	Designed Nested	Unstructured, by Dhillon *et al.* (2008)
4×2	8	0.9242	0.9258	1.0519	1.0690
4×2	16	0.8660	0.8944	1	1.0323
4×2	32	0.7512	0.8660	0.8925	N/A
6×2	8	0.9744	0.9759	1.2311	1.2344
6×2	16	0.9372	0.9428	1.1563	1.1925
6×2	32	0.8895	0.9275	1.1073	N/A
8×2	8	1	1	1.3093	N/A
8×2	16	0.9645	0.9661	1.2352	N/A
8×2	32	0.9372	0.9504	1.2024	N/A

2.8 Subspace Packings with Nested Structure and PSK Alphabet

In this section an effective design technique is developed for codebooks that not only have the nested structure, but also have codeword elements from a (scaled) PSK alphabet. One approach to developing such a technique would be to tackle the problem directly, by applying discretization techniques to the method in the previous section, perhaps employing an appropriate constant-modulus penalty along the way. However, such techniques can become quite unwieldy as the rank of the codebook and the number of antennas grow. As an alternative, in this section the line packing technique developed in Section 2.6.1 is used to develop a layered method in which the codebook is designed one rank at a time, in a greedy fashion. The greedy nature of this method is reasonable, because in many wireless systems lower rank codebooks will be employed more often.

To simplify the exposition of the proposed technique, the focus will be on the case of nested codebooks of rank 2. A rank-1 codebook $\{\mathbf{f}_i\}_{i=1}^{N}$ is first designed with codeword elements from the (scaled) PSK alphabet, possibly designed using one of the techniques in Section 2.6. Then vectors $\{\check{\mathbf{f}}_i\}_{i=1}^{N}$ with elements from the PSK alphabet are designed so that $\mathcal{F} = \big\{[\mathbf{f}_i, \check{\mathbf{f}}_i]\big\}_{i=1}^{N}$ forms a good nested rank-2 codebook with elements from the alphabet.

The proposed technique for designing the vectors $\{\check{\mathbf{f}}_i\}_{i=1}^{N}$ is based on the principles of the relax-round-expurgate-replace approach in Section 2.6.1. To employ those principles, one needs to capture the desired properties of $\{\check{\mathbf{f}}_i\}_{i=1}^{N}$ in the absence of the PSK alphabet constraint. To that end, the first observation is that the matrices $[\mathbf{f}_i, \check{\mathbf{f}}_i]$

should have orthonormal columns. Second, the rank-2 codebook should have a large minimum distance in the chosen subspace distance metric. Third, to facilitate the rounding procedure, the relaxed codewords should have elements of almost constant modulus. The following smooth objective captures these properties

$$J_6\big(\{\check{\mathbf{f}}_i\}\big) = \omega_1\Big(\sum_i \big|\mathbf{f}_i^H\check{\mathbf{f}}_i\big|^\beta\Big)^{1/\beta} + \omega_2 J_2\big(\big\{[\mathbf{f}_i, \check{\mathbf{f}}_i]\big\}\big) + \omega_3 P_{\mathrm{cm}}\big(\{\check{\mathbf{f}}_i\}\big), \tag{2.28}$$

where β is constrained to be even and the subspace metric is implicit in $J_2(\cdot)$. Then the basic sequential optimization procedure is applied using this objective. Next, the principles of the approach in Section 2.6.1 are applied, and the elements of the obtained vectors are rounded to the nearest point in the alphabet. Any rounded $\check{\mathbf{f}}_i$ that is not orthogonal to $\mathbf{f}_i$ is immediately expurgated, and the remaining set of rank-2 codewords, $\big\{[\mathbf{f}_i, \check{\mathbf{f}}_i]\big\}$ is examined. Any $\check{\mathbf{f}}_i$ that induces a distance, or distances, in the rank-2 codebook that is deemed unsatisfactory is also expurgated. The expurgated vectors can be replaced using an exhaustive search, or through a recapitulation of the relax-round-expurgate-replace method. However, in both cases the size of the new problem is the number of codewords to be replaced, which is typically much smaller than N.

Some of the distance results for codebooks designed using this technique have been summarized in Table 2.6. In this case, the distance metric in the optimization of the rank-2 codebook is the chordal distance. This table shows that in spite of the nesting and alphabet constraints, the proposed approach can find codebooks that come reasonably close to the Rankin bound at both ranks. The obtained minimum distances are also quite close to those of the otherwise unconstrained nested codebooks in Table 2.5.

Table 2.6: Minimum distances achieved by designed nested codebooks with (scaled) 4–PSK alphabet and the chordal distance as the rank-2 distance. The Rankin bounds for constant modulus codebooks for each rank are provided for comparison.

$M_t \times M$	N	Rank 1		Rank 2, chordal dist.	
		Designed Nested PSK	Rankin Bound	Designed Nested PSK	Rankin Bound
4×2	8	0.8660	0.9258	1	1.0690
4×2	16	0.8660	0.8660	1	1
4×2	32	0.7071	0.8660	0.8660	1.0000
6×2	8	0.9428	0.9759	1.0541	1.2344
6×2	16	0.9428	0.9428	1	1.1926
6×2	32	0.8819	0.9129	1	1.1547
8×2	8	1	1	1.1456	1.3093
8×2	16	0.9354	0.9661	1.0897	1.2649
8×2	32	0.9354	0.9504	1.0607	1.2443

2.9 Performance Evaluation

This section evaluates the performance of communication systems based on the codebooks obtained using the suggested approaches in the case of the MIMO downlink with zero-forcing beamforming (ZFBF) (Jindal, 2006; Spencer *et al.*, 2004b), per-user unitary precoding and rate control (PU^{2}RC) (Kim *et al.*, 2006), and zero-forcing block diagonalization (ZFBD) (Ravindran and Jindal, 2008) signalling under a Rayleigh fading channel model. In that model, the channel gains are modelled as i.i.d. circular complex Gaussian random variables with zero mean and unit variance, and each receiver is assumed to know its channel realization. The kth receiver has a codebook $\mathcal{F}_k$ that is a random rotation of the transmitter's codebook $\mathcal{F}$, (Ding *et al.*, 2007). (The transmitter knows the rotation.) For rank-1 codebooks used in the beamforming cases, each receiver determines the codeword $\mathbf{f}_{i^\star} \in \mathcal{F}_k$ that solves $\max_{\mathbf{f}_i \in \mathcal{F}_k} |\mathbf{h}_k \mathbf{f}_i^*|$, (Jindal, 2006), and for the higher-rank codebooks used in the ZFBD case, the receiver determines the codeword $\mathbf{F}_{i^\star} = \arg\max_{\mathbf{F}_i \in \mathcal{F}_k} \|\mathbf{H}\mathbf{F}_i\|_F$, (Ravindran and Jindal, 2008).

The index $i^\star$ is fed back to the transmitter, which collects all the quantized versions of the channel estimates from the scheduled users in order to compute the beamforming/precoding matrix according to the system architecture.

First, the ZFBF case is considered for a system with $M_t = 4$ transmitter antennas, $K = 4$ receivers and rank-1 codebooks of size $N = 16$, and 64. In Figure 2.1 the cumulative distributions (cdf) of the sum of the rates that can be achieved using ZFBF and Gaussian signalling at an average SNR of 15 dB have been plotted, where if σ^2 denotes the variance of the noise at each receiver, then the SNR is trace$(\mathbf{T}^H\mathbf{T})/(K\sigma^2)$. The considered codebooks are designed using the basic procedure, and Love's codebooks (Love, 2004). The average performance of a set of codebooks generated randomly using the uniform distribution on the manifold is also provided. For the codebooks of size $N = 16$, the basic procedure generated a codebook with a minimum distance of 0.8943, whereas the codebook of Love (2004) has a minimum distance of 0.8670. For the codebooks of size $N = 64$, the corresponding minimum distances are 0.7175 and 0.6035, respectively. Figure 2.1 demonstrates that the improved minimum distances of the designed codebooks generate improved sum rate statistics. For a given target rate, the designed codebooks yield a lower outage probability, and for a given target outage probability, the designed codebooks yield a larger rate. In Figure 2.2 the corresponding results for PU^{2}RC signalling are plotted. In that case, the tangible performance advantages of the designed codebooks extend over a broader range of rates.

Figure 2.3 examines the performance of rank-1 codebooks of size $N = 16$ and 32 with elements from the (scaled) 4-PSK defined alphabet in the PU^{2}RC scheme at an

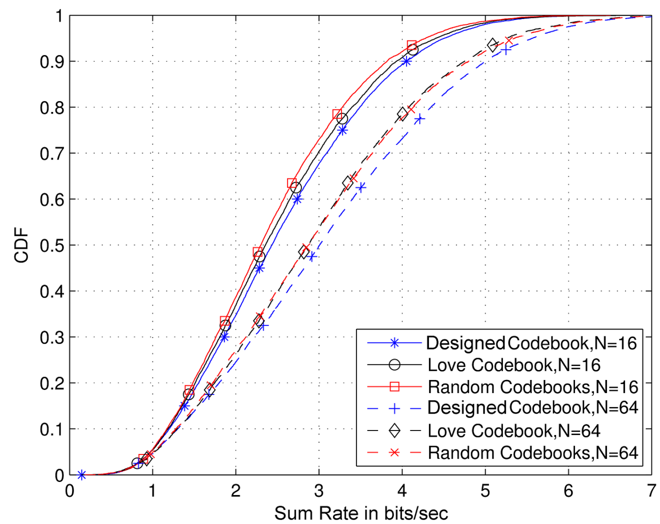

Figure 2.1: CDF of the sum rate for a ZFBF system with various unconstrained codebooks at SNR=15 dB.

SNR of 15 dB. For the codebooks of size $N = 16$, the incremental construction generated a 4-PSK codebook with a minimum distance of 0.8660 (cf. Table 2.3), whereas the unconstrained codebook has a minimum distance of 0.8943. For the codebooks of size $N = 32$, the corresponding distances are 0.7071 and 0.8115. As suggested by their good distance properties, the 4-PSK codebooks provide performance that is close to that of the unconstrained codebooks and better than the average performance of randomly generated codebooks (with unconstrained coefficients).

Finally, this section examines the performance of the proposed approach to the design of nested codebooks with PSK elements in the context of a ZFBD system with $M_t = 4$ transmit antennas, 2 users, and rank-2 codebooks of size $N = 16$. The CDF of the sum rate at an SNR of 15 dB for a nested codebook with 4-PSK elements designed using the approach in Section 2.8 with the chordal distance is plotted in Figure 2.4. To provide benchmarks against which the performance can be assessed,

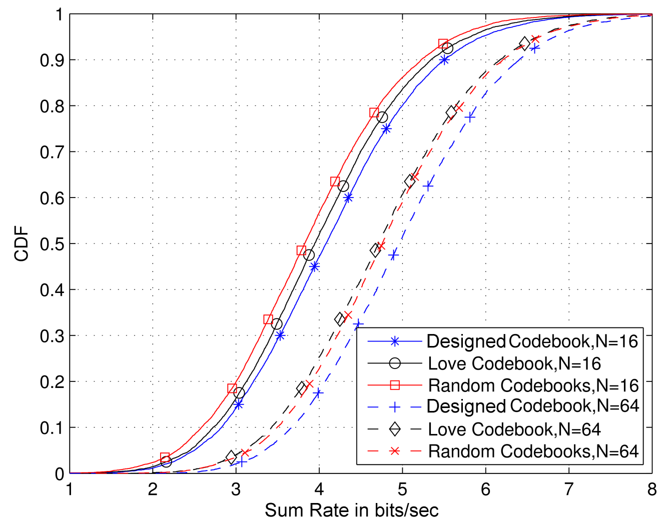

Figure 2.2: CDF of the sum rate for a PU^2RC system with various unconstrained codebooks at SNR=15 dB.

the CDFs for the average of a set of randomly generated unconstrained codebooks, for Love's codebook (Love, 2004), and for unconstrained codebooks designed using the technique in Section 2.4 with the Fubini-Study and the chordal distances have been also plotted. For reference, the minimum distances of the designed codebooks are provided in Table 2.7. The key feature of Figure 2.4 is that in this setting, the memory and computational savings associated with the nested 4-PSK codebook that has been designed, and the convenience of the nested structure, are obtained without incurring a substantial reduction in performance. Figure 2.4 also shows that both the unconstrained codebooks designed using the proposed technique provide improved performance over the codebook of Love (2004). In a similar setting with $M_t = 8$, 2 users and rank-4 codebooks (not shown here) the unconstrained codebook that was obtained using the proposed method with the chordal distance provided marginally better performance than the codebook that was obtained using the Fubini-Study

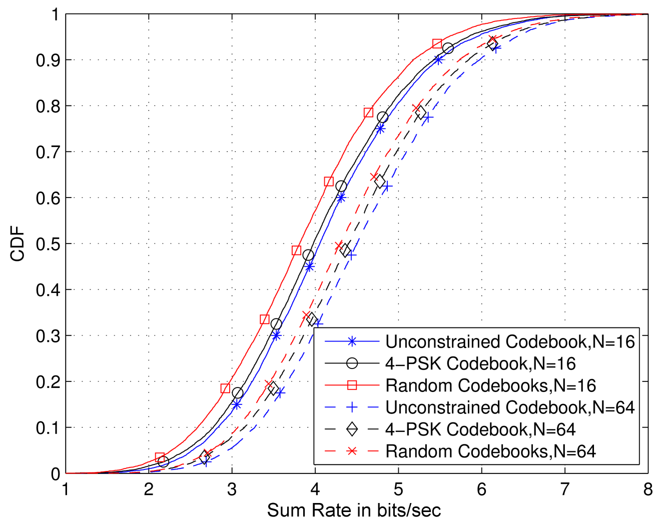

Figure 2.3: CDF of the sum rate for a PU^{2}RC system with unconstrained, 4-PSK and random codebooks at SNR=15 dB.

Table 2.7: Distance properties of the rank-2 codebooks used in Figure 2.4.

Type	Design Metric	Min. Ch. dist.	Min. FS dist.
Nested 4-PSK	Chordal dist.	1	1.0472
Unconstr. (Love, 2004)	—	0.8188	0.8654
Unconstr.	Chordal dist.	1.0309	1.0863
Unconstr.	FS dist.	0.9428	1.2309

distance.

2.10 Conclusion

A flexible approach to the design of structured Grassmannian codebooks (packings) for communication systems that employ limited feedback has been developed. The

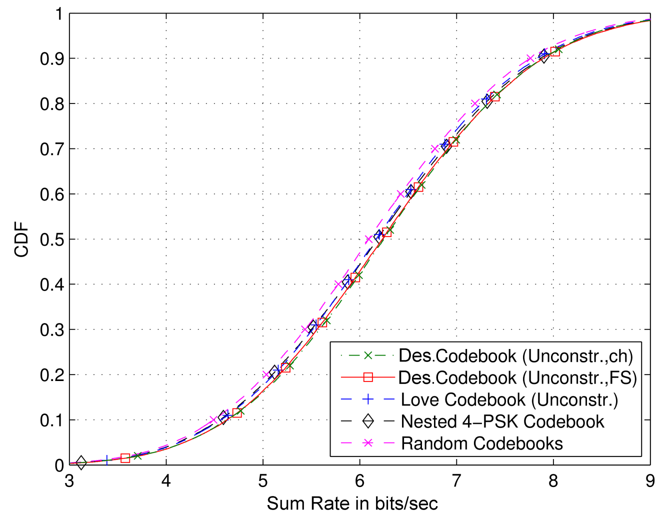

Figure 2.4: CDF of the sum rate for a rank-2 ZFBD system with unconstrained, and nested 4-PSK codebooks at SNR=15 dB.

proposed approach is based on (otherwise) unconstrained optimization over the Grassmannian manifold of a sequence of smooth objective functions. In addition to generating unstructured subspace codebooks with better minimum Fubini-Study distances than some existing codebooks, the proposed approach was also used to generate structured codebooks with other desirable properties. In particular, one variant of the approach was used to generate rank-2 codebooks with a nested structure and elements from the (scaled) 4-PSK alphabet. An outcome of another variant was the complete set of mutually unbiased bases of dimension 16; a result that may be of independent interest. In the unstructured case, the designed codebooks were shown to provide tangible performance gains when applied to a simple multiple antenna downlink communication system with limited feedback. In was also shown that the structured codebooks obtained using the proposed approaches do not incur a substantial degradation in performance.

The design of unstructured Grassmannian codebooks is widely acknowledged as a difficult problem, and the kinds of structure that have been imposed on the structured codebooks do not make that problem easier. As a result, the focus has been on the development of sound heuristics that guide us towards good codebooks. Inherently, the development of heuristics involves choices, and the choices that have been made herein could certainly be debated. However, the numerical results have shown that with the choices that have been proposed the proposed techniques have been able to obtain codebooks of considerable size that possess the desirable structural properties and yet have a minimum distance that comes close to the bound on the minimum distance of unstructured codebooks. Furthermore, the basic principles of the approach can accommodate, in a straightforward way, a myriad of other choices for the smooth objectives used to capture the desirable properties of the codebook.

In closing, attention should be drawn to a point that was implicit in step 3 of the basic sequential optimization procedure in Section 2.4, namely that the proposed approach can be extended in a straightforward way to certain other manifolds, including the Stiefel manifold. The extension to that manifold may be of interest in some related applications of limited feedback in wireless communications (e.g., Pitaval and Tirkkonen, 2012).

Chapter 3

Incremental Grassmannian Feedback Schemes for Multi-User MIMO Systems

In Chapter 2, optimization techniques were proposed for the design of codebooks that can be utilized in channel quantization schemes that are based on memoryless quantization techniques. While memoryless quantization techniques have been used efficiently in quantization of temporally-uncorrelated channels, they neglect the temporal correlation that often exists in the wireless channels. When such correlation exists, incremental/differential schemes that exploit the correlation lead to more accurate representation of the channel using the same feedback budget. Accordingly, in this chapter an incremental feedback quantization scheme is proposed for temporally correlated channels. The proposed scheme is based on the concept of the geodesic between two points on the Grassmannian manifold and has a number of interesting features. It requires only one codebook for initialization and updates.

This codebook can be designed using the same techniques proposed in Chapter 2 for memoryless quantization. Furthermore, the proposed scheme has an inherent ability to autonomously recover from feedback errors. Two implementations are proposed for the incremental scheme that differ in the way the feedback budget is partitioned and the presumed channel model.

3.1 Introduction

For the uncorrelated i.i.d. block fading Rayleigh channel, memoryless uniform vector quantization on the Grassmannian manifold is an effective strategy that can offer substantial performance gain at the price of only a few feedback bits (Love *et al.*, 2003; Love and Heath, Jr., 2005b). As discussed in Chapter 2, a quantization codebook is designed offline and known to all terminals (Clerckx *et al.*, 2008; Dhillon *et al.*, 2008; Schober *et al.*, 2009; Mondal *et al.*, 2007; Inoue and Heath Jr., 2009). For each fading block, the receiver sends to the transmitter the index of one element of the codebook, which represents the quantized version of the subspace. Since it is memoryless, this scheme neglects any correlation that may exist between fading blocks in practice, and hence it may require a larger amount of feedback than necessary. However, neglecting temporal correlation provides inherent robustness against mismatches in any model for the temporal correlation.

Several methods have been proposed for developing limited feedback schemes that take advantage of temporal correlation between fading blocks (e.g., Yang and Williams, 2007; Heath *et al.*, 2009; Choi *et al.*, 2012; Kim *et al.*, 2011a, 2008; El Ayach and Heath, 2012; Liu and Jafarkhani, 2006; Inoue and Heath, 2011; Kim *et al.*, 2010, 2012). A number of these schemes are based on a differential quantization strategy in which the scheme seeks to incrementally update the current estimate of the subspace. Typically, these schemes involve two codebooks, the first of which is a conventional codebook designed for a memoryless quantization scheme and is used to initialize the scheme. The second codebook (or family of codebooks) is used to quantize local changes around the current subspace. A variety of approaches have been proposed for designing the second codebook. One approach is to construct a simple parameterized codebook for the neighborhood of a point on the manifold (Heath *et al.*, 2009; Choi *et al.*, 2012; Kim *et al.*, 2008, 2011a). Since the geometry of the Grassmannian manifold is analogous to that of a sphere, these local codebooks are often called spherical or polar cap codebooks. In a number of schemes of this kind, the codebook consists of a set of points that are carefully distributed on a ring. In some schemes, the radius of the ring is scaled according to the channel statistics, while in others, the radius is successively shrunk. In some schemes, the spherical cap codebook is rotated each time it is used. An alternative to the spherical/polar cap approaches, is to employ a codebook of tangents to the Grassmannian manifold. Once the transmitter receives the index of the tangent, it can construct the updated version of the subspace by taking a step of a specified size on the manifold in a direction specified by the tangent (El Ayach and Heath, 2012; Liu and Jafarkhani, 2006). The geometric constructions in these differential methods can be extended to develop predictive quantization schemes (Inoue and Heath, 2011; Kim *et al.*, 2012). Finally, guidelines for selecting the feedback update period and the number of feedback bits in differential quantization schemes for beamforming based systems were derived by Kim *et al.* (2010)

In this chapter, an incremental feedback approach to the Grassmannian quantization problem is proposed that exploits the temporal correlation between fading blocks and yet requires only a single codebook that is designed using conventional techniques for memoryless Grassmannian quantization (such as those in Clerckx *et al.*, 2008; Dhillon *et al.*, 2008; Schober *et al.*, 2009; Mondal *et al.*, 2007; Inoue and Heath Jr., 2009) and the techniques presented in Chapter 2. Similar to the principles of some existing methods (Yang and Williams, 2007; Inoue and Heath, 2011; El Ayach and Heath, 2012), the current estimate of the subspace is updated by taking a step along a specified geodesic on the manifold. The proposed approach differs in that the direction is captured using a conventional Grassmannian codebook and the step size along the geodesic can be quantized in a variety of ways. In order to implement the principles of the proposed approach, two feedback schemes are developed that differ in the way in which the step size is communicated. The first scheme is a specialization of the incremental approach that is tailored to a first-order Gauss-Markov model with a known correlation coefficient. In this *model-based incremental scheme*, the step size is recursively updated in a manner that is dependent on the correlation coefficient. As a result, no bits need to be assigned to feed back the step size and hence a conventional Grassmannian codebook with higher resolution can be used. However, the performance of the model-based scheme is dependent on the accuracy with which the correlation coefficient is estimated and tracked.

In the second scheme, conventional Grassmannian codebooks of lower resolution are used and the remainder of the bit budget is used to adapt the step size to the channel realization (rather than to the channel statistics). Even with only one bit allocated to the step size, this scheme provides substantial robustness to mismatches in the presumed statistical model and hence this scheme will be referred to as the *robust incremental scheme*. An interesting feature of both of the proposed incremental schemes is that they exhibit an intrinsic property that allows them to recover from errors caused by the feedback channel. In the case of the robust scheme, this recovery property can be considered to be a "self reseting" feature, in which the transmitter and the receiver can perfectly re-synchronize to each other after a feedback error.

The performance of the proposed schemes will be compared to that of other existing schemes in the cases of single user (SU) and single-cell multi-user (MU) MIMO systems. The authors of these existing schemes have kindly shared their codebooks for the updates, and their willingness to do so is gratefully acknowledged. It is worth mentioning that the proposed schemes can be also implemented in more complicated scenarios, such as the multi-cell downlink, presented in Chapters 4-5 but the underlying principles for channel quantization in that setting are similar to those for the case of MU-MIMO systems. In the case of SU-MIMO systems, having CSI at the transmitter side can improve the achievable rate, but it is not necessary to achieve the multiplexing gain of the system. The case of MU-MIMO is quite different, in that the multiplexing gain provided by the channel cannot be achieved in the absence of CSI at the transmitter (Jindal, 2006). In the simulation section, the performance of the proposed incremental schemes in several channel scenarios and system models is evaluated. For example, the achievable rate in the case of SU-MIMO is compared with perfect channel estimation and with channel model mismatches. The results show the improved performance of the proposed schemes in comparison with some existing ones. In the case of MU-MIMO, the Zero-Forcing Beamforming (ZFBF) and Block-diagonalization (BD) systems are adopted and the results show the performance gains in the sum rate achieved by the proposed schemes, especially with estimation errors in the channel model. Finally, the performance of the proposed scheme is evaluated in the case of feedback error and the results show that the proposed schemes have an inherent ability to recover from feedback errors and provide performance gains, in contrast to other existing schemes.

3.2 System Model

In order to focus on the principles of the proposed schemes, the results will be presented in case of SU and MU-MIMO systems, as in Chapter 2. As in Sections 1.2 and 2.2, a MU-MIMO downlink system is considered with a base station with M_t transmit antennas communicating to K users, the kth of which has M_{rk} receive antennas. The received signal for the kth user is given by:

$$\mathbf{y}_k = \sqrt{\frac{E_s}{Q}}\mathbf{H}_k\mathbf{T}\mathbf{s} + \mathbf{z}_k \tag{3.1}$$

where $\mathbf{H}_k \in \mathbb{C}^{M_{rk}\times M_t}$ is the channel matrix from the transmitter to the kth user, $Q \leq \min\left(M_t, \sum_k M_{rk}\right)$ is the number of transmitted data streams, and $\mathbf{T} \in \mathbb{C}^{M_t\times Q}$ is the transmitter preprocessing matrix which is normalized so that $\mathrm{trace}(\mathbf{T}^H\mathbf{T}) = Q$. The transmitted symbols form the vector $\mathbf{s}$, which is multiplied by the preprocessing matrix $\mathbf{T}$ producing a signal vector $\mathbf{x} = \sqrt{\frac{E_s}{Q}}\mathbf{T}\mathbf{s}$, where E_s is the total transmit energy assuming that $\mathrm{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_Q$. Finally $\mathbf{z}_k$ is the additive noise at user k which is assumed to be zero-mean complex Gaussian with normalized covariance, $\mathrm{E}\{\mathbf{z}_k\mathbf{z}_k^H\} = \mathbf{I}_{M_{r_k}}$.

The scenario under consideration is the one in which the channel changes in a block fading manner with temporal correlation between blocks and each receiver is assumed to have perfect knowledge of its channel matrix. The receiver seeks to provide information about this channel to the base station. Furthermore, the receivers employ partitioned quantization schemes in which one of the partitions involves the quantization of a subspace.

3.3 Topology of the Grassmannian Manifold

Memoryless quantization schemes are effective schemes for block fading channel models in which the blocks are independent. In scenarios in which the blocks are correlated, memoryless quantization schemes can still be employed, but schemes that seek to exploit that correlation have the potential to reduce the quantization error (and implicitly its impact on the system performance) or reduce the amount of feedback that is required to achieve a given level of performance. The approach proposed in this chapter for exploiting correlation between blocks is based on the topology of the Grassmannian manifold (Edelman *et al.*, 1998; Manton, 2002; Conway *et al.*, 1996), which will be briefly reviewed.

The Grassmannian manifold $\mathbb{G}_{M_t,M}$ is a representation of subspaces of dimension M in $\mathbb{C}^{M_t}$. Such a subspace can be described as the linear span of an orthonormal basis, and that basis can be captured by the columns of a matrix $\mathbf{X} \in \mathbb{C}^{M_t \times M}$ that satisfies $\mathbf{X}^H\mathbf{X} = \mathbf{I}_M$. As there is a continuum of matrices that can represent a given subspace, these matrices can be deemed to form an equivalence class in terms of representing subspaces. Each "point" on the Grassmannian manifold $\mathbb{G}_{M_t,M}$ is a single matrix $\mathbf{X}$ that represents this equivalence class for the given subspace.

The intuition behind the proposed incremental schemes depends on the concept of the geodesic (the shortest path) between two points on the manifold. To write an equation for the geodesic from the point $\mathbf{F}_i$ on the manifold to the point $\mathbf{F}_j$, the notion of the tangent to the manifold at the point $\mathbf{F}_i$ in the direction of the point $\mathbf{F}_j$ will be used, which can be written as

$$\Delta(\mathbf{F}_i, \mathbf{F}_j) = (\mathbf{I} - \mathbf{F}_i\mathbf{F}_i^H)\mathbf{F}_j(\mathbf{F}_i^H\mathbf{F}_j)^{-1}. \tag{3.2}$$

If $\Delta(\mathbf{F}_i, \mathbf{F}_j) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ denotes the compact singular value decomposition of $\Delta(\mathbf{F}_i, \mathbf{F}_j)$ and $\mathbf{\Phi} = \tan^{-1}(\mathbf{\Sigma})$, with $\tan^{-1}(\cdot)$ being defined elementwise, then the points on the geodesic from $\mathbf{F}_i$ to $\mathbf{F}_j$ can be written as

$$\mathbf{F}(t) = \mathbf{F}_i\mathbf{V}\mathrm{cos}(\mathbf{\Phi}t)\mathbf{V}^H + \mathbf{U}\mathrm{sin}(\mathbf{\Phi}t)\mathbf{V}^H \tag{3.3}$$

where $t \in [0, 1]$ tracks the progress along the geodesic from $\mathbf{F}_i$ to $\mathbf{F}_j$. In the case of beamforming, where the points on the manifold are elements of $\mathbb{C}^{M_t}$, equation (3.3) can be simplified to $\mathbf{f}(t) = -\mathbf{f}_i\mathrm{cos}(\phi t) + \frac{\Delta(\mathbf{f}_i,\mathbf{f}_j)}{\|\Delta(\mathbf{f}_i,\mathbf{f}_j)\|}\mathrm{sin}(\phi t)$, where $\phi = \tan^{-1}\big(\|\Delta(\mathbf{f}_i, \mathbf{f}_j)\|\big)$.

3.4 Principles of the Proposed Incremental Feedback Schemes

The proposed approach to incremental Grassmannian feedback is based on the observation from (3.3) that the current representative of the subspace can be updated by taking a step of a specified size along the geodesic in the direction of a point selected from a conventional Grassmannian codebook. To describe that approach, it

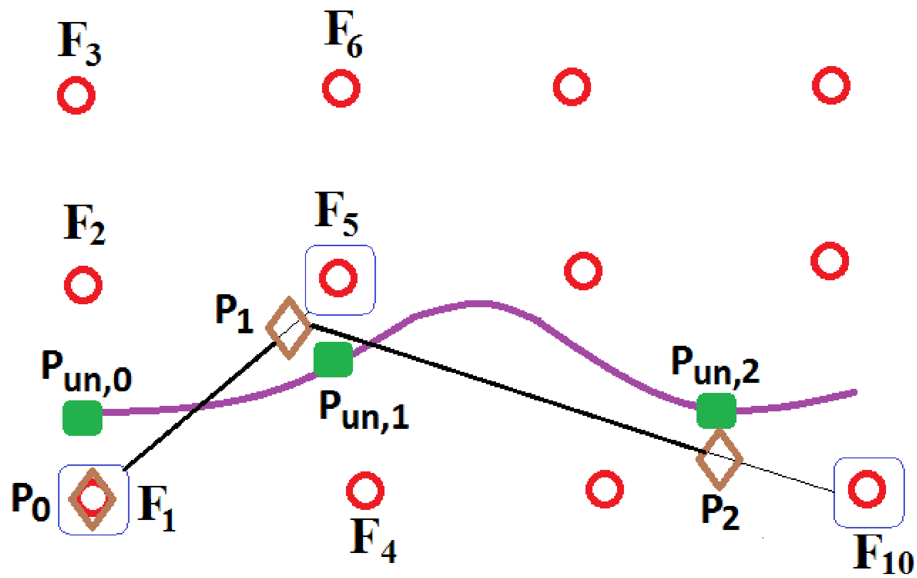

Figure 3.1: A pictorial representation of the proposed technique. The points marked $\boldsymbol{P}_{\mathrm{un},n}$ denote the sequence of points on the manifold that would be chosen if the feedback were unlimited. The circles denote the points $\mathbf{F}_i$ in the memoryless Grassmannian codebook (which do not have to be uniformly spaced), and the points $\boldsymbol{P}_n$ denote the points generated by the proposed technique. Note that the Grassmannian manifold is compact, and hence the edge effects that appear in this pictorial representation do not arise in practice.

will be assumed that B bits that are available for feedback, and that those bits are partitioned into two sets of size B_{cb} and B_{step}, respectively. The set of B_{cb} bits is used to index the elements of a Grassmannian codebook $\mathcal{F}_{\mathrm{cb}}$ of size $L_{\mathrm{cb}} = 2^{B_{\mathrm{cb}}}$ from a memoryless scheme, and the set of B_{step} bits is used to index a quantization of the interval $[0, 1]$, $\mathcal{T} = \{\bar{t}_1, \ldots, \bar{t}_{L_{\mathrm{step}}}\}$, where $L_{\mathrm{step}} = 2^{B_{\mathrm{step}}}$. The basic operation of the proposed scheme is illustrated in Fig. 3.1. In that figure, the points $\{\mathbf{P}_{un,n}\}$ denote the sequence of points on the manifold that would be chosen in case of unlimited feedback. That is, the true channel direction information (CDI) of the channel.[1] The

[1]Consistent with the established terminology, the term channel direction information (CDI) will be used, even though it is, strictly speaking, channel subspace information.

point $\mathbf{P}_n$ denotes the quantized version of the CDI for block n under the proposed limited feedback model and the points $\{\mathbf{F}_i\}$ denote the elements of the conventional Grassmannian codebook $\mathcal{F}_{\text{cb}}$. The initial quantized CDI $\mathbf{P}_0$ is determined using memoryless quantization scheme using the codebook $\mathcal{F}_{\text{cb}}$. The quantized CDI for the next block (the $(n+1)$th block) is obtained by informing the transmitter to take a step of size $t_n \in \mathcal{T}$ along the geodesic in the direction of a specified element of $\mathcal{F}_{\text{cb}}$. Accordingly, in Fig. 3.1, $\mathbf{P}_1$ is obtained by taking a step from $\mathbf{P}_0$ in the direction of $\mathbf{F}_5$, and $\mathbf{P}_2$ is obtained by taking a step from $\mathbf{P}_1$ in the direction of $\mathbf{F}_{10}$.

To formalize the procedure, let us consider the SU-MIMO case, in which the receiver selects a unitary precoder $\mathbf{P}$ to maximize the Gaussian mutual information,

$$\text{GMI}(\mathbf{P}) = \log\det\Big(\mathbf{I}_{M_r} + \frac{E_s}{Q}\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P}\Big). \tag{3.4}$$

The proposed incremental feedback approach proceeds as follows:

1. *Initialization:* As in memoryless quantization schemes, the receiver selects the initial precoder $\mathbf{P}_0$ as $\mathbf{P}_0 = \arg\max_{\mathbf{F}_i\in\mathcal{F}_{\text{cb}}} \text{GMI}(\mathbf{F}_i)$, and feeds back the corresponding index.

2. *Incremental updates:* Given the precoder for the nth fading block, $\mathbf{P}_n$, the receiver uses (3.2) to construct the tangent $\mathbf{\Delta}(\mathbf{P}_n, \mathbf{F}_j)$ for each $\mathbf{F}_j \in \mathcal{F}_{\text{cb}}$. For each tangent, the receiver considers steps along the corresponding geodesic of size $\bar{t}_i \in \mathcal{T}$, $i = 1, \ldots, L_{\text{step}}$. Using (3.3), this process generates $L_{\text{cb}}L_{\text{step}} = 2^B$ candidate precoders, denoted by $\mathbf{G}_i \in \mathcal{G}_n$. The receiver then determines

$$\mathbf{P}_{n+1} = \arg\max_{\mathbf{G}_i\in\mathcal{G}_n} \text{GMI}(\mathbf{G}_i), \tag{3.5}$$

and feeds back the indices that enable the transmitter to construct $\boldsymbol{P}_{n+1}$. Those indices are the index of the direction $\mathbf{F}_i \in \mathcal{F}_{\mathrm{cb}}$ and the step size $\bar{t}_i \in \mathcal{T}$ that generated $\mathbf{P}_{n+1}$ in (3.5).

Although the basic principle of the incremental approach is based on a rather simple concept, it has a number of interesting properties.

1. The codebook $\mathcal{F}_{\mathrm{cb}}$ that is used for the directions of the incremental update can be chosen to be a subset of the codebook $\mathcal{F}$ that is used in the memoryless quantization process in the initialization step. As such, there is no additional storage requirement. This enables a system designer to take advantage of existing codebooks for memoryless quantization (e.g., Love *et al.*, 2008; Love and Heath, Jr., 2005b; Clerckx *et al.*, 2008), including codebooks with a subset structure (e.g., Gohary and Davidson, 2009) and the codebooks designed in Chapter 2. Doing so alleviates some of the difficulties associated with constructing codebooks for increments (e.g., Heath *et al.*, 2009).

2. Since the proposed technique is based on updates along geodesics, the precoder in (3.5) lies on the manifold. As a result, the additional projection step employed in the method of Kim *et al.* (2011a) is not required.

3. Given its ability to interpolate between points in the codebook for memoryless feedback (cf., Fig. 3.1), the performance of the proposed technique is less sensitive to the quality of the codebook than conventional memoryless feedback schemes. Indeed, if the channel is constant, this interpolation yields a feedback scheme with higher resolution than the underlying memoryless scheme.

In the following two sections two schemes that tailor the basic principles described above to different scenarios will be developed:

1. The first scheme is a variant of the generic scheme that is specialized for a first-order Gauss-Markov model for the temporal correlation of the channel. The temporal correlation coefficient of the channel is assumed to be known to the transmitter and the receiver and is assumed to be constant. In this *model-based scheme*, the step size at channel use n can be adapted to the temporal correlation of the channel, and hence all the feedback bits can be assigned to the direction along the geodesic; i.e., $B_{\text{step}} = 0$. This scheme provides substantial performance gains over memoryless schemes when the temporal correlation of the channel is significant. However, like some other existing schemes, the performance degrades if the temporal correlation of the channel is only coarsely estimated, or if the temporal correlation changes significantly.

2. If the temporal correlation of the channel is not accurately known, or if the nature of the environment and the relative motion of the transmitter and the receivers mean that the temporal correlation coefficient may change, a fraction of the feedback budget should be reserved for the size of the geodesic step. This enables the step size to be adapted to the channel realization rather than the channel statistics. By doing so, the robust scheme exhibits robustness to the uncertainty in the temporal channel statistics estimation or abrupt changes in the channel conditions. This is in contrast to the first scheme and the methods of (Heath *et al.*, 2009), (Choi *et al.*, 2012), and (Kim *et al.*, 2011a), where the corresponding notions of step size are adapted to the channel statistics. Further, the quantization of the step size can be chosen so that this scheme exhibits an

interesting *self-reseting* feature; see Section 3.6.

3.5 Model-Based Incremental Feedback Scheme

In this section, a model-based incremental feedback scheme on the Grassmannian manifold in which the step size update for each block is determined according to the temporal correlation of the channel is presented. In this scheme, all the feedback bits B are used to represent the direction of the geodesic, i.e., $B_{\text{cb}} = B$ and $B_{\text{step}} = 0$, and the step size is computed at each channel use based on a statistical analysis provided below. The core assumption here is that the evolution of the channel matrix for a generic user k, $\mathbf{H}_{k,n}$, can be modelled by an independent first-order Gauss-Markov process. For notational convenience, the user index k will be dropped and model the channel at the nth block as

$$\mathbf{H}_n = \beta \mathbf{H}_{n-1} + \sqrt{1-\beta^2}\boldsymbol{\Theta}_n \tag{3.6}$$

where $\boldsymbol{\Theta}_n$ has independent entries distributed according to $\mathcal{CN}(0, 1)$. The temporal channel correlation coefficient β is modeled by Jakes model according to $\beta = J_0(2\pi f_{\text{d}})$, where $J_0(\cdot)$ is the zeroth order Bessel function and f_{d} is the normalized Doppler frequency. Let the singular value decomposition of $\mathbf{H}_n = \mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{V}_n^H$ and that of $\boldsymbol{\Theta}_n = \mathbf{Q}_n\boldsymbol{\Lambda}_n\mathbf{S}_n^H$, where the elements of the diagonal matrices $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\Lambda}_n$ are arranged in descending order. Generally, one is interested in tracking the dominant subspace of a particular dimension. In the SU-MIMO case that dimension is Q, for the MU-MISO case it is 1 and for the MU-MIMO case it is Q_k. In order to keep the notation generic, let M denote the dimension. Accordingly, $\bar{\boldsymbol{\Sigma}}_n$ denotes the upper-left $M \times M$

block of $\mathbf{\Sigma}_n$, and $\bar{\mathbf{V}}_n$ denotes the first M columns of $\mathbf{V}_n$.

To develop a methodology for choosing the step size to be used at the nth incremental step, first denote the previous quantized version of the CDI by $\mathbf{P}_{n-1}$ and second observe that the proposed model-based scheme has to move from $\mathbf{P}_{n-1}$ to $\bar{\mathbf{V}}_n$, the current (true) CDI. The model-based scheme does not necessarily move precisely in the direction of $\bar{\mathbf{V}}_n$, as the geodesic is specified by a point in the Grassmannian codebook, but, nevertheless, a reasonable guide for the choice of the step size can be obtained by looking at the average value of the distance between $\mathbf{P}_{n-1}$ and $\bar{\mathbf{V}}_n$. Unfortunately, the analysis of the average geodesic distance between $\mathbf{P}_{n-1}$ and $\bar{\mathbf{V}}_n$ is quite involved. Therefore, one may seek insight into that distance by examining the average chordal distance between $\mathbf{P}_{n-1}$ and $\bar{\mathbf{V}}_n$,

$$d_{\text{ch}}\big(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_n\big) = \sqrt{M - \big\|\bar{\mathbf{V}}_n^H \mathbf{P}_{n-1}\big\|_F^2}, \tag{3.7}$$

and then making the observation that for small geodesic steps the step size can be approximated by a linear function of the chordal distance. That is, given a point on the manifold $\mathbf{F}_i$ and another point $\mathbf{F}(t)$ of the form in (3.3) the approximation

$$d_{\text{ch}}\big(\mathbf{F}_i, \mathbf{F}(t)\big) \approx \gamma t, \tag{3.8}$$

will be adopted, where the choice of γ depends on the intended range of the approximation. As explained in Appendix B, it can be shown that for the model in (3.6) with M chosen to be M_r, the expected value of the chordal distance between $\mathbf{P}_{n-1}$

and $\bar{\mathbf{V}}_n$ is

$$\mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_n)\} \;=\; \beta^2\,\mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\} \,+\, (1\,-\,\beta^2)\,\mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_{n-1}, \mathbf{S}_n)\}, \quad (3.9)$$

where the expectation is taken over the channel model in (3.6), an i.i.d. Gaussian model for $\mathbf{H}_0$, and random codebooks generated according to the isotropic distribution on the manifold.

Using the above analysis, an appropriate choice for the step size for the nth feedback interval would be to choose t_n so that $(\gamma t_n)^2 = \mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_n)\}$. Now an analytical expression for such a t_n has to be developed, or an approximation thereof.

Since $\mathbf{S}_n$ is isotropically distributed on $\mathbb{G}_{M_t,M}$, and is independent of $\mathbf{P}_{n-1}$, it can be shown that (Kim *et al.*, 2011a)

$$\mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_{n-1}, \mathbf{S}_n)\} = \frac{M(M_t - M)}{M_t}. \quad (3.10)$$

Furthermore, for the memoryless quantization at step $n = 0$, it has been shown that (Dai *et al.*, 2008)

$$\mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_0, \bar{\mathbf{V}}_0)\} \approx \frac{1}{2}\Big(\frac{M(M_t - M)}{M(M_t - M) + 1} + \frac{\Gamma(\frac{1}{M(M_t-M)})}{M(M_t - M) + 1}\Big) \times \big(\mathcal{C}_{M_t,M} 2^B\big)^{\frac{-1}{M(M_t-M)}}, \quad (3.11)$$

where

$$\mathcal{C}_{M_t,M} = \begin{cases} \frac{1}{(M(M_t-M))!}\Pi_{i=1}^{M}\frac{(M_t-i)!}{(M-i)!} & 0 < M \leq \frac{M_t}{2} \\ \frac{1}{(M(M_t-M))!}\Pi_{i=1}^{M_t-M}\frac{(M_t-i)!}{(M_t-M-i)!} & \frac{M_t}{2} < M \leq M_t \end{cases} \quad (3.12)$$

and $\Gamma(\cdot)$ denotes the gamma function. Using these expressions, $(\gamma t_1)^2 = \mathrm{E}\{d^2_{\text{ch}}(\mathbf{P}_0, \bar{\mathbf{V}}_1)\}$

if

$$t_1 = \frac{1}{\gamma}\sqrt{\gamma^2\beta^2\xi_0^2 + \frac{(1-\beta^2)M(M_t - M)}{M_t}}, \tag{3.13}$$

where ξ_0^2 denotes $\mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_0, \bar{\mathbf{V}}_0)\}$ in (3.11).

Given the incremental nature of the proposed scheme, $\mathbf{P}_{n-1}$ is correlated with $\mathbf{P}_{n-2}$ and this significantly complicates the evaluation of $\mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\}$ for $n > 1$. However, as shown in Appendix C, by (i) modelling the region that can be spanned by a step of size t_{n-1} from $\mathbf{P}_{n-2}$ by a spherical cap; (ii) partitioning that cap according to a spherical cap approximation of the Voronoi cell of each element of the underlying Grassmannian codebook; and (iii) analyzing the relative volumes of those caps, a recursive approximation for $\mathrm{E}\big\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\big\}$ can be derived. Using that approximation the following recursive approximation for t_n can be obtained for $n \geq 2$ which ensures that $(\gamma t_n)^2 \approx \mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_n)\}$

$$t_n = \frac{1}{\gamma}\sqrt{\mu\gamma^2\beta^2 t_{n-1}^2 2^{\frac{-B}{M(M_t-M)}} + \frac{(1-\beta^2)M(M_t - M)}{M_t}}, \tag{3.14}$$

where μ is a correction factor due to the Voronoi region approximation and is dependent on the number of feedback bits assigned to the codebook.

The evolution of the step size for several values of β is plotted in Fig. 3.2. From that figure it can be seen that during the first few channel uses, the step size value is large in order to mitigate the quantization errors and the channel evolution. As the number of channel uses increases, the quantization error decreases and the step size converges to a steady state value (i.e., a fixed point) dependent on β. Using (3.14),

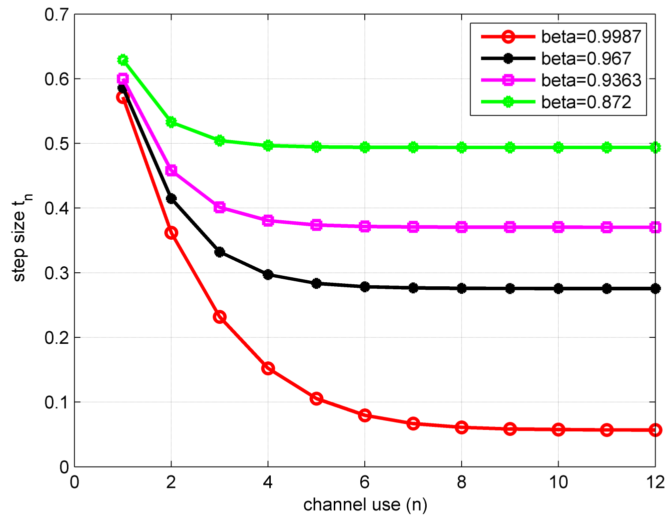

Figure 3.2: The evolution of the step size t_n with respect to the channel use (n) for various values of β.

that fixed point, which is denoted by t_∞, can be found to be

$$t_\infty(\beta) = \sqrt{\frac{(1-\beta^2)M(M_t - M)}{M_t\gamma^2\left(1-\mu\beta^2 2^{\frac{-B}{M(M_t-M)}}\right)}}. \tag{3.15}$$

An important feature of the proposed incremental feedback scheme is its intrinsic ability to recover, autonomously, from feedback errors. This feature is illustrated in Fig. 3.3. In that figure, the points $\mathbf{P}_{un,n}$ denote the sequence of CDI's that would be chosen in case of unlimited feedback, while the point $\mathbf{P}_n$ denotes the quantized version of the CDI for block n under the limited feedback model and the points $\{\mathbf{F}_i\}$ denote the elements of the conventional Grassmannian codebook $\mathcal{F}_{\text{cb}}$. The initial quantized CDI $\mathbf{P}_0$ is determined using memoryless quantization scheme using the codebook $\mathcal{F}_{\text{cb}}$. Now, let's assume that there were no feedback errors during the initialization phase nor in the first update ($n = 0$ and $n = 1$). As a result, at channel use $n = 1$, both the

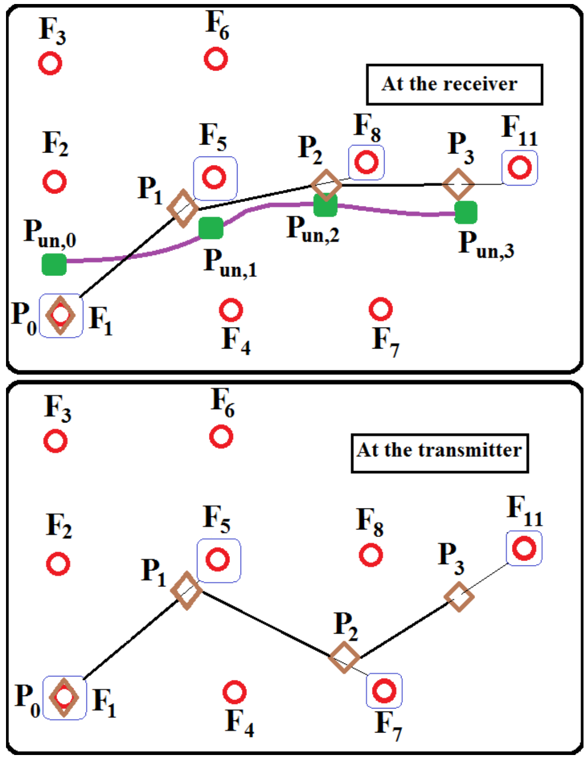

Figure 3.3: A pictorial representation of how the incremental schemes can mitigate error propagation in case of feedback errors. At the second channel use, P_2 at the transmitter and the receiver are different. With successive channel uses, new points P_n at the transmitter and the receiver become closer to each other.

transmitter and receiver have the same $\mathbf{P}_1$. Now consider the case that at the second channel use, the receiver decides to take a step in the direction of $\mathbf{F}_8$ and that errors in feeding back the indices lead the transmitter to interpret the direction of the geodesic as being that associated with $\mathbf{F}_7$. As can be seen by comparing the upper and the lower parts of Fig. 3.3 the transmitter and the receiver assume totally different values for the quantized CDI. Now, in the third channel use, the receiver chooses to take a step of size t_3 in the direction of $\mathbf{F}_{11}$. If there is no feedback error, the transmitter will move from the previous point $\mathbf{P}_2$ (which, due to the feedback error, is in the wrong place) with the a step size t_3 in the direction of $\mathbf{F}_{11}$. Even though the points $\mathbf{P}_2$ at the transmitter and the receiver are not the same, the new points $\mathbf{P}_3$ become closer to each other, and they will get closer still with successive channel uses. The main reason behind this property is that the incremental scheme moves from the current point to a *fixed* point determined by an element of the codebook $\mathcal{F}_{\text{cb}}$.

The simulation section will show that the above model-based incremental technique can achieve performance gains in highly correlated channel and provides better results, under a variety of channel conditions, than other similar schemes that employ some sort of geodesic interpolation.

3.6 A More Robust Incremental Feedback Scheme

In the previous section, a model-based incremental scheme, in which the step size is a function of the temporal correlation of the channel β was developed. Although that scheme enables us to retain the full resolution of the memoryless codebook for the directions of the updates, it is inherently sensitive to the accuracy of the channel model. This section describes how the quantization scheme for the step size in the generic form of the approach in Section 3.4 can be chosen so as to develop a more robust incremental scheme.

Once the bit budget B has been partitioned into components for the directions, B_{cb}, and the step size, B_{step}, of the geodesic updates, there are a variety of ways in which the scalar quantization scheme $\mathcal{T} = \{\bar{t}_i\}$ on $[0, 1]$ for the step size t can be chosen. A natural, if somewhat generic, approach would be to consider the Lloyd algorithm (e.g., Gray and Neuhoff, 1998; and references therein). If the temporal correlation of the channel model, β, is known and does not change in time, then obtaining a good quantization of $[0, 1]$ is quite straightforward. However, in practice the temporal correlation is likely to vary significantly in time, and a more sophisticated approach would be to postulate a distribution for β, and to apply Lloyd's algorithm to the resulting family of models.

Having said that, as will be now explained, in the proposed application a case can be made for choosing $\bar{t}_{L_{\text{step}}} = 1$. (Recall that $L_{\text{step}} = 2^{B_{\text{step}}}$.) First, it is observed that when the channel is uncorrelated in time, the representatives $\bar{\mathbf{V}}_n$ (c.f. (3.6)) are isotropically distributed on the manifold. Since good codebooks $\mathcal{F}_{\text{cb}}$ are approximately isotropically distributed, if the channel is likely to change rapidly, at least at some points in time, including a step size of 1 in $\mathcal{T}$ is a natural choice, as it will allow the algorithm to "track" the rapidly varying channel. This choice also imbues the approach with a desirable "self-resetting" property, in the sense that whenever the receiver chooses a step size of 1 (and there are no errors in the feedback at that step), both the transmitter and receiver will move to the same point on the manifold (i.e., the same codeword in the codebook), no matter where they were at the previous step. This can be illustrated using Fig. 3.3, in which there is a feedback error at the second channel use that results in the transmitter and receiver having different values for $\mathbf{P}_2$. If the step size chosen at step $n = 3$ is $t_3 = 1$ (and no feedback errors occur in this step), then both the transmitter and receiver take a step of size 1 in the direction of $\mathbf{F}_{11}$. Hence, even though they have different values for $\mathbf{P}_2$, both the receiver and the transmitter have $\mathbf{P}_3 = \mathbf{F}_{11}$. As such, in just one step the incremental system has recovered from the feedback error in step $n = 2$. Choosing a step-size quantization that includes $\bar{t}_{L_{\text{step}}} = 1$ also helps to develop insight into the choice of the codebook size. When the receiver chooses a step size of 1, the system behaves like a memoryless system with $B - B_{\text{step}}$ bits; i.e., a memoryless system with $2^{B-B_{\text{step}}}$ codewords.

There are a number of guidelines for choosing the size of memoryless codebooks in temporally-uncorrelated, spatially i.i.d. Gaussian channels (e.g., Jindal, 2006, 2007; Ravindran and Jindal, 2008; Dai *et al.*, 2008), and since the proposed robust incremental scheme has $2^{B_{\text{step}}} - 1$ choices for the step size other than 1, it will perform at least as well as a memoryless system with a codebook of $2^{B-B_{\text{step}}}$ codewords.

As the above discussion suggests, the choice of B_{step} involves trade-offs between the performance in highly-correlated channels and the performance in rapidly-varying channels. Given the possibility of significant changes in the channel correlation over time, a reasonable approach would be to choose $B_{\text{step}} = 1$, and to leave the remaining $B - 1$ bits to index the directions of the update on the manifold. As will be demonstrated in the simulations section, this choice results in a scheme that provides robust performance in the presence of significant changes in the channel correlation.

If B_{step} is chosen such that $B_{\text{step}} = 1$, and if one of the points in the step-size quantization should be 1, then $\mathcal{T} = \{\bar{t}_1, 1\}$ and the remaining design decision is to choose $\bar{t}_1$. While a (constrained) Lloyd approach could be used, a simple alternative is to consider first-order Gauss-Markov models of the form in (3.6) and to postulate a distribution for the temporal correlation parameter β based on the scenarios that are expected to be encountered. A reasonable choice for the remaining quantization point for the step size is then

$$\bar{t}_1 = \mathrm{E}_\beta\{t_\infty(\beta)\}, \tag{3.16}$$

where $t_\infty(\beta)$ is the fixed point of the step-size iteration for the model-based scheme; cf. (3.15). In the case of a finite number of choices, say N_β, for β, each with equal probability, this expression simplifies to $\bar{t}_1 = \frac{1}{N_\beta}\sum_\beta t_\infty(\beta)$.

3.7 Simulations

This section examines the performance of the proposed model-based and robust incremental schemes in a variety of different communication configurations. In each configuration, the channels are modelled using the first-order Gauss-Markov model in (3.6), with different, and possibly time-varying, values for the correlation coefficient, β. To provide context for the values of β, the relation $\beta = J_0(2\pi f_\mathrm{d})$ will be used from Jakes' model, and the parameters from the IEEE 802.16m standard (IEEE, 2008), in which the carrier frequency is 2.5 GHz and the feedback interval is 5 ms.

This section compares the performance of the proposed schemes against the conventional memoryless quantization scheme, and the differential rotation feedback scheme (Kim *et al.*, 2011a). In the case of systems based on beamforming, it will also compares against the polar-cap differential feedback scheme (Choi *et al.*, 2012). For the model-based and robust incremental schemes, the underlying Grassmannian codebook is designed using the technique in Chapter 2.

For the other differential schemes, I would like to thank the authors of these schemes for providing their codebooks for the updates. For the initialization of all the differential schemes the same Grassmannian codebook will be adopted, and, therefore, all the schemes start from the same point on the manifold. The base station is equipped with $M_t = 4$ antennas, while the number of receive antennas will be determined by the communication scheme. The SNR is fixed at 10 dB and the feedback budget is $B = 4$ bits per channel use. For this number of feedback bits, setting the Voronoi correction factor in the model-based scheme so that $0.9 < \mu < 0.95$ yields the best performance.

The case of a single-user multiple-input single-output (SU-MISO) system is first

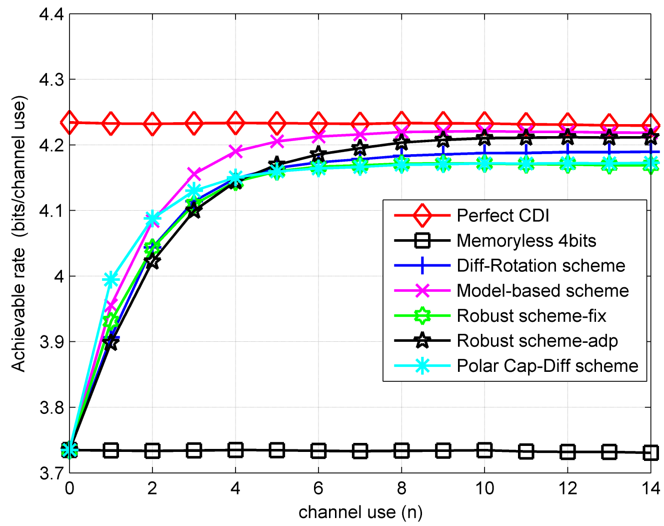

Figure 3.4: Performance comparison for a SU-MISO system under a channel model in which $\beta = 0.997$.

considered in which the transmitter uses the vector indexed by the feedback scheme as the beamforming vector; c.f. (2.3). Fig. 3.4 illustrates the average achievable rate over 1000 realizations of the channel model in (3.6) for different feedback schemes as a function of the feedback interval. In the considered scenario, the channel is temporally correlated with $\beta = 0.997$, which corresponds to a velocity of 1km/h. For the implementation of the robust incremental scheme, the results are presented for two different choices for the quantization of the step size, $\mathcal{T}$. In the first case (denoted 'Robust scheme-adp'), the step size quantization is $\mathcal{T}_1 = \{t_n(0.997), 1\}$, where $t_n(0.997)$ is the step size chosen for the model-based method with $\beta = 0.997$; cf. (3.14). In the second case (denoted 'Robust scheme-fix'), the step size quantization is constant in time and is chosen to be $\mathcal{T}_2 = \{\bar{t}_1, 1\}$, where $\bar{t}_1$ is the average of the fixed points of the model based scheme (cf. (3.16)) over equally likely correlation coefficients $\beta = \{0.999, 0.997, 0.936, 0.872, 0.5\}$. From Fig. 3.4, one can see that the model-based

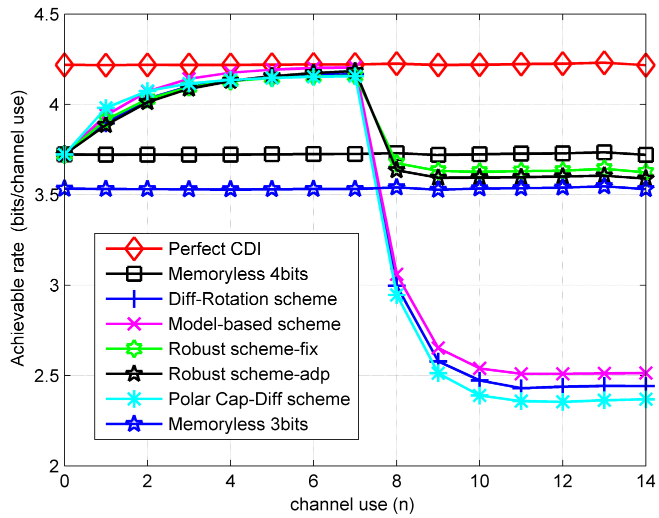

Figure 3.5: Performance comparison for a SU-MISO system under a channel model in which β drops from 0.997 to 0.5 at channel use 7.

incremental scheme provides better performance than the other schemes. The performance of the robust incremental scheme with fixed step size quantization $\mathcal{T}_2$ is close to that of the differential rotation scheme and approximately the same as that of the polar-cap differential scheme, while the robust incremental scheme with time-varying step size quantization, $\mathcal{T}_1$, provides improved performance compared to the differential rotation and polar-cap differential schemes. In the scenario of Fig. 3.4, all the differential schemes perform significantly better than the 4-bit memoryless scheme, but this scenario represents an ideal case in which the channel follows a first-order Gauss-Markov model with significant temporal correlation, the receiver has perfect knowledge of the temporal correlation, and that correlation remains unchanged during the data transmission period.

In Fig. 3.5, a scenario is considered in which the channel is, initially, temporally correlated with $\beta = 0.997$, and then, at channel use 7, the channel correlation drops to

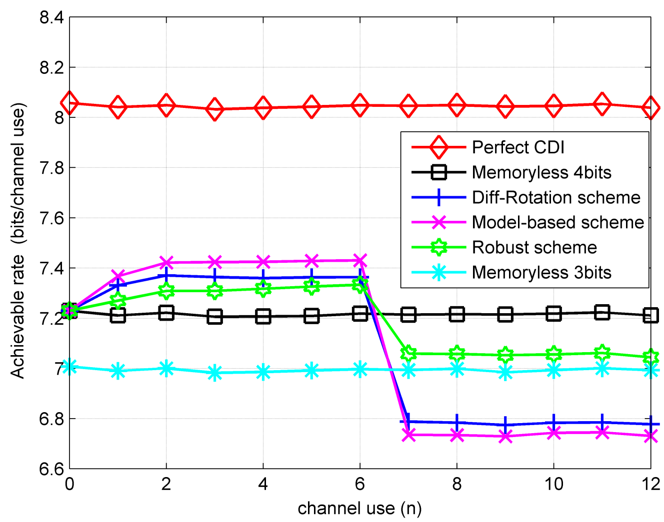

Figure 3.6: Performance comparison for a SU-MIMO system under a channel model in which β drops from 0.872 to 0 at channel use 7

.

$\beta = 0.5$. This corresponds to a velocity change from 1 km/h to roughly 20 km/h. Prior to the change in the channel correlation, all the schemes provide better performance than the 4-bit memoryless scheme, as the results in Fig. 3.4 predict. However, their performance is quite different after that change. As expected from the discussion in Section 3.6, the performance of the proposed robust scheme remains better than that of the memoryless scheme using 3 bits when the channel correlation changes. (The 3-bit memoryless scheme is equivalent to always choosing a step size of 1 in the proposed robust scheme.) In contrast, the performance of the other differential feedback schemes drops quite a long way below that of the 3-bit memoryless scheme. An additional observation from Fig. 3.5 is that the robust incremental scheme with the fixed step size quantization $\mathcal{T}_2$ has better performance than the robust scheme with the adaptive step-size quantization $\mathcal{T}_1$ when the channel correlation changes, and that it does not require accurate estimation of the channel temporal coefficient

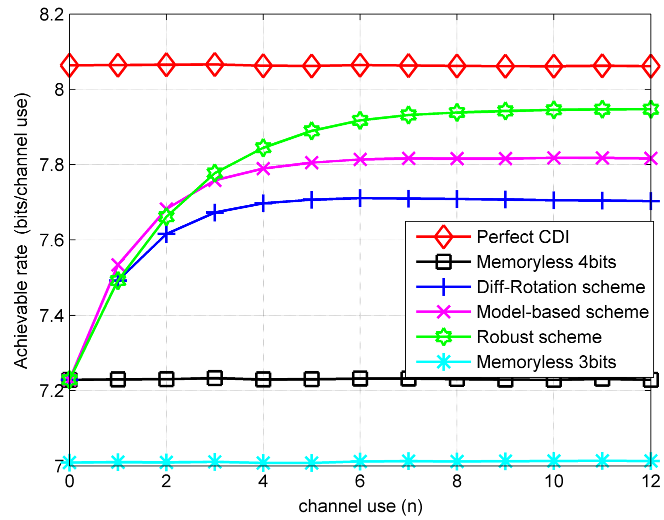

Figure 3.7: Performance comparison for a SU-MIMO system under a channel model in which β is constant, but underestimated (actual $\beta = 0.997$, but estimated to be $\beta = 0.872$).

β. Accordingly, in the following simulations, only results for the robust scheme with the fixed step-size quantization, $\mathcal{T}_2$ will be presented.

In Fig 3.6 a SU-MIMO system is considered in which the receiver has two antennas and the (four-antenna) transmitter sends two symbols per channel use. The transmitter performs linear precoding with equal power loading using the matrix specified by the feedback scheme; c.f. (2.3). The channel follows the Gauss-Markov model in (3.6), with the channel correlation changing from $\beta = 0.872$ to $\beta = 0$ at channel use 7. That is, the channel suddenly becomes uncorrelated. (For the IEEE 802.16m parameters, a correlation of $\beta = 0.872$ corresponds to a speed of 10 km/h.) Even though the channel becomes completely uncorrelated, the proposed robust incremental scheme still provides better performance than the 3-bit memoryless scheme. It also provides significantly better performance than the other competing schemes when the channel becomes uncorrelated. (The fact that the gain of all the schemes in the preliminary

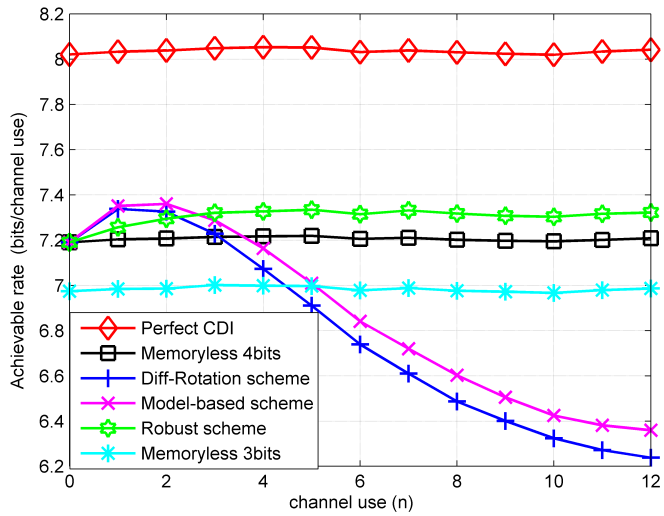

Figure 3.8: Performance comparison for a SU-MIMO system under a channel model in which β is constant, but overestimated (actual $\beta = 0.872$, but estimated to be $\beta = 0.997$).

stage in Fig 3.6 is less than the gain in Fig 3.5 is due, in large part, to the lower level of correlation, i.e., the smaller value of β.)

In Figs 3.7 and 3.8, the performance of the considered feedback schemes is examined in the above four-input two-output SU-MIMO system in cases in which the channel correlation β is constant, but is under- or over-estimated, respectively. In Fig. 3.7 the actual channel correlation is $\beta = 0.997$, whereas it is estimated to be 0.872. In Fig. 3.8, the actual correlation is $\beta = 0.872$, but it is estimated as being 0.997. Figs 3.7 and 3.8 demonstrate the sensitivities of the existing techniques and the sensitivity of the proposed model-based technique to misestimation of the channel correlation coefficient. The sensitivity arises from the resulting misadaptation of the step size in these techniques. The sensitivity is greater in the case of over-estimation of the correlation coefficient, because that results in small step sizes that are unable

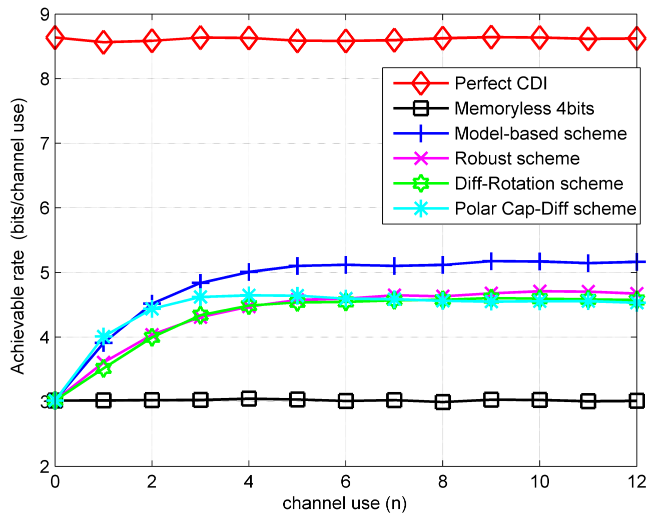

Figure 3.9: Performance comparison for a MU-MISO system that employs ZFBF under a channel model with $\beta = 0.936$.

to track the variation of the channel. Figs 3.7 and 3.8 also demonstrate the manner in which the proposed robust scheme overcomes the sensitivity to misestimation of the correlation coefficient. In both cases, the proposed robust scheme provides significantly better performance than the other competing schemes. In the case of over-estimation in Fig. 3.8, it provides better performance than the 4-bit memoryless scheme, whereas the performance of the other competing schemes falls significantly below that of the 3-bit memoryless scheme.

The performance of the considered feedback schemes in MU-MISO and MU-MIMO systems is now examined. Fig. 3.9 presents results for a MU-MISO system with 4 users that employs zero-forcing beamforming (ZFBF) with uniform power loading; c.f. (2.6). The correlation coefficient of the channel is $\beta = 0.936$, which corresponds to a speed of 7 km/h, and is fixed and precisely known. From Fig. 3.9, it can be seen that the model-based incremental scheme has the best tracking properties among

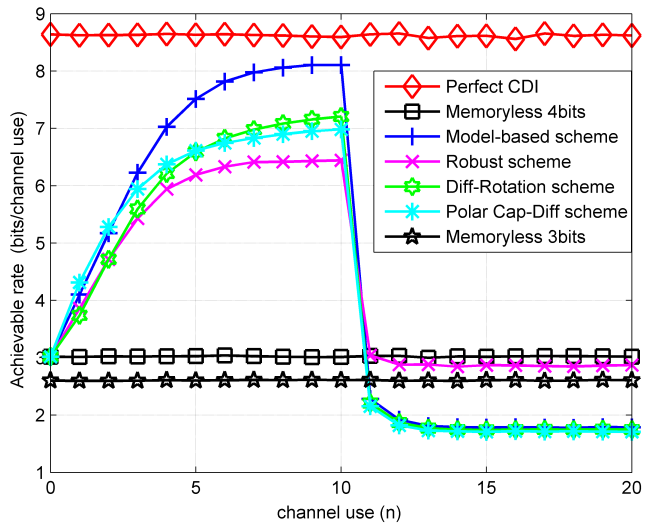

Figure 3.10: Performance comparison of MU-MISO systems with ZFBF and varying β (β drops from 0.997 to 0.5 at channel use 11).

the competing schemes. In many ways, the relative performance of the considered schemes in this scenario is similar to that for the SU-MISO scenario in Fig. 3.4.

In order to illustrate the performance of the various schemes in a multiuser scenario in which the channel correlation changes, in Fig. 3.10 the performance of the 4-user MU-MISO system is considered. This system employs ZFBF in an environment in which the channels are initially temporally correlated with $\beta = 0.997$ and then that correlation drops to 0.5 at channel use 11. As can be seen from the figure, initially all the competing methods provide better performance than the 4-bit memoryless scheme, but when the channel correlation drops, only the proposed robust scheme is able to maintain better performance than the 3-bit memoryless scheme.

In Fig. 3.11, a 2-user MU-MIMO system is considered in which each receiver has two antennas and the transmitter sends two symbols to each receiver at each channel use, using block-diagonalization (BD) with equal power loading; c.f. (2.8). The

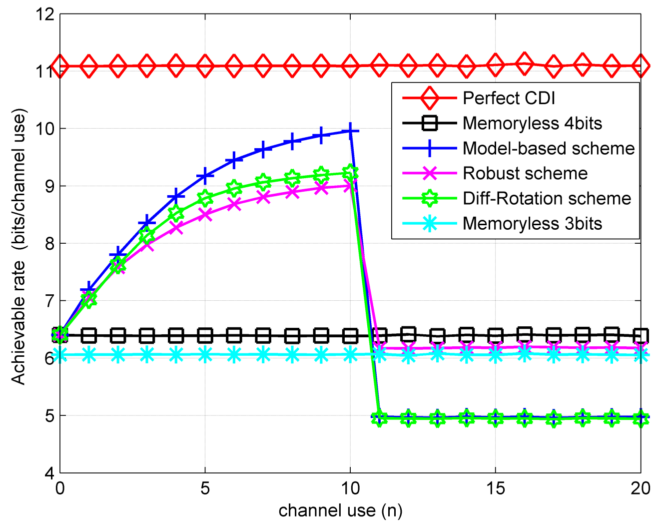

Figure 3.11: Performance comparison of MU-MIMO systems with BD and varying β (β drops from 0.997 to 0.872 at channel use 11).

channels are initially temporally correlated with $\beta = 0.997$ and then that correlation drops to 0.872 at channel use 11. In this scenario, the proposed robust scheme maintains performance better than the 4-bit memoryless scheme, even after the drop in correlation, whereas the performance of the other competing schemes drops below that of the 3-bit memoryless scheme.

One of the inherent weaknesses in incremental quantization schemes is their sensitivity to errors, and, in particular, the propagation of the effects of those errors. In Fig. 3.12, the performance of the considered schemes in the presence of a feedback error is examined, and in Fig. 3.13 the performance in the presence of both a feedback error and errors in the estimation of the correlation coefficient, β is examined as well. For simplicity, consider the case of an SU-MISO system. In the scenario considered in Fig. 3.12, the correlation coefficient is $\beta = 0.936$, and a single feedback error occurs at the 6th channel use. The error is modelled as a switch to another codeword, with each

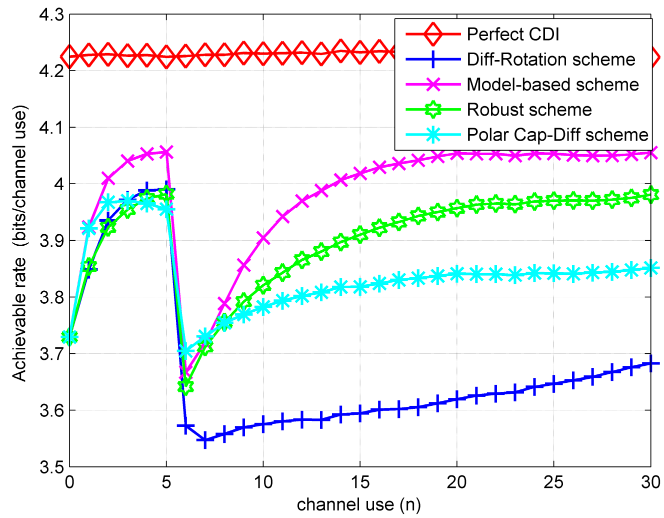

Figure 3.12: Performance comparison for a SU-MISO system under a channel model with $\beta = 0.936$ in the presence of a feedback error at channel use 6.

other codeword having equal probability, $1/(2^{B_{\text{cb}}} - 1)$. The figure shows how both the proposed model-based scheme and the proposed robust scheme are able to recover from the feedback error and return to the performance level that was achieved prior to the error, whereas for the polar cap and differential rotation schemes, the error has a lasting impact. To explore the combined impact of a feedback error and error in the estimation of the correlation coefficient, the previous experiment has been repeated for the case in which the actual correlation coefficient is $\beta = 0.936$, but it is estimated as being 0.967. This corresponds to an error of about 3%. The results for this case are provided in Fig. 3.13. In this scenario, the proposed robust scheme recovers more quickly than the proposed model-based scheme. Perhaps more importantly, both are able to recover from the feedback error, whereas for the existing competing schemes the presence of the estimation error in the correlation coefficient appears to hinder the ability to recover from the feedback error.

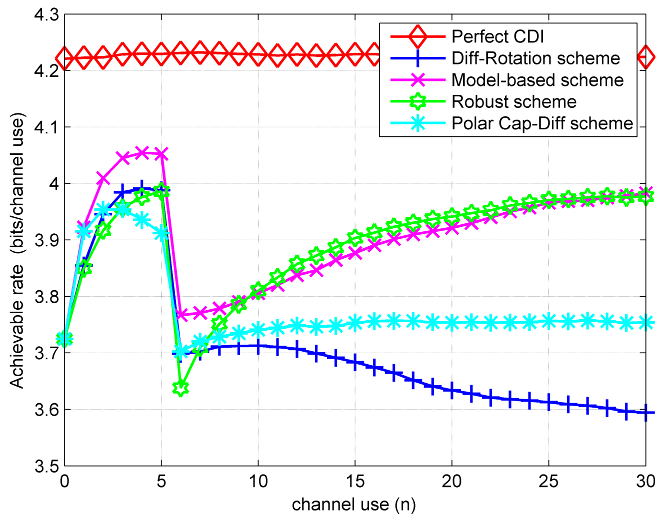

Figure 3.13: Performance comparison for a SU-MISO system under a channel model in which β is constant, but over-estimated, in the presence of a feedback error at channel use 6. (Actual $\beta = 0.936$, but estimated to be 0.967)

3.8 Conclusion

In this chapter, two implementations of an incremental feedback method for temporally correlated channels have been proposed. The first scheme is adapted to a first-order Gauss-Markov channel model and uses all the available resources for direction feedback, while the second scheme divides the resources between direction information and step size feedback. From the simulations, one can conclude that when the channel correlation coefficient β is perfectly estimated at the receiver side, the model-based incremental scheme provides the best performance when compared to other existing schemes. However, if the receiver does not have enough time or resources to estimate the long-term channel statistics with sufficient accuracy, or if those statistics change often, it may be desirable to use the robust incremental scheme. A distinguishing feature of both of the proposed schemes is their intrinsic ability to recover from feedback errors and return to their steady state performance levels.

Chapter 4

Interference Alignment for 2 and 3-cell MIMO Dowlink Networks

In Chapters 2 and 3 the problems of designing the quantization codebook for temporally-uncorrelated channels and the quantization scheme for temporally-correlated channels for SU-MIMO and MU-MIMO systems were studied. In those systems, the architecture for CSI feedback has been well established, and the linear precoding schemes for perfect CSI can be adapted to systems employing limited feedback for CSI in a straightforward way. In this chapter the goal is to tackle the problem of designing linear precoding schemes for MIMO downlink networks that are amenable to being implemented in limited feedback systems. In these systems, the appropriate architecture has yet to be established, and the contribution of the remainder of the thesis is to develop linear precoding schemes that induce feedback architectures that have desirable properties, such as achieving improved DoF compared to conventional interference avoidance schemes and requiring only local feedback. Once this architecture has been established, the underlying concepts for the codebook design procedures and the quantization schemes presented in Chapters 2 and 3 can be applied. Therefore, in this chapter and the following ones, the focus will be on designing the linear precoding schemes only. The implementation of these schemes has low computational complexity and it has been shown that, in many settings, they are sufficient to achieve the DoF of the network.

In this chapter, the basic concepts for linear precoding schemes for multi-cell downlink networks will be introduced including interference alignment (IA) precoding schemes, and the definition of the degrees of freedom (DoF) of a network. The application of the underlying concepts in the case of 2-cell MIMO downlink network is used in designing a linear IA scheme that can be implemented in a limited feedback model.

4.1 Introduction

One of the major bottlenecks in the process of improving the performance of wireless networks is interference. In the past, networks have been designed so that interference is avoided, but the loss of spectral efficiency that those schemes incur, means that interference will be actively managed in future networks. To provide more context for that statement, some of the interference management approaches that have been applied to multi-cell networks can be summarized as follows (Cadambe and Jafar, 2008):

- Decode: When the interfering signal power is much stronger than the desired signal power, the interfering signal can be decoded then subtracted from the desired signal, so long as the receiver has access to the codebook of the interferer and can synchronize appropriately. The receiver then decodes the desired signal after the removal of the interference. This method is less common in practice due to the complexity of sequential detection and the potential for error propagation.

- Treat as Noise: Treating interference as noise, combined with single user encoding and decoding often suffices when the interference is weak, and this has been used in practice in frequency-reuse cellular networks.

- Orthogonalize: When the interfering signal power is comparable to the desired signal power, interference can be avoided by orthogonalizing the access channel in time or frequency or space.

One of the more intriguing of the recently proposed approaches for manging the interference is Interference Alignment (IA) (Maddah-Ali *et al.*, 2008; Cadambe and Jafar, 2008; Jafar, 2011). The main principle that underlies IA is that the signalling scheme is designed so that interference arriving at a particular user from many sources is aligned in a way that it occupies only a portion of the received signal space. If the desired signal has a component that lies outside the signal space spanned by the interference, that component can be extracted using simple projection techniques. By being able to null out interfering sources that have comparable power to the desired signal power, without canceling the desired signal itself, the IA precoding scheme can be classified as an "Orthogonalize" approach for interference management. The development of IA schemes has helped characterize the fundamental limits on the spectral efficiency behavior of certain wireless network at high SNRs, or what is known as the "Degrees of Freedom" (DoF) of a network. The DoF can be interpreted as the number of resolvable interference-free signal dimensions and it is the slope of the achievable rate curve with respect to the logarithm of the SNR, at high SNRs (e.g., Cadambe and Jafar, 2008, 2009).

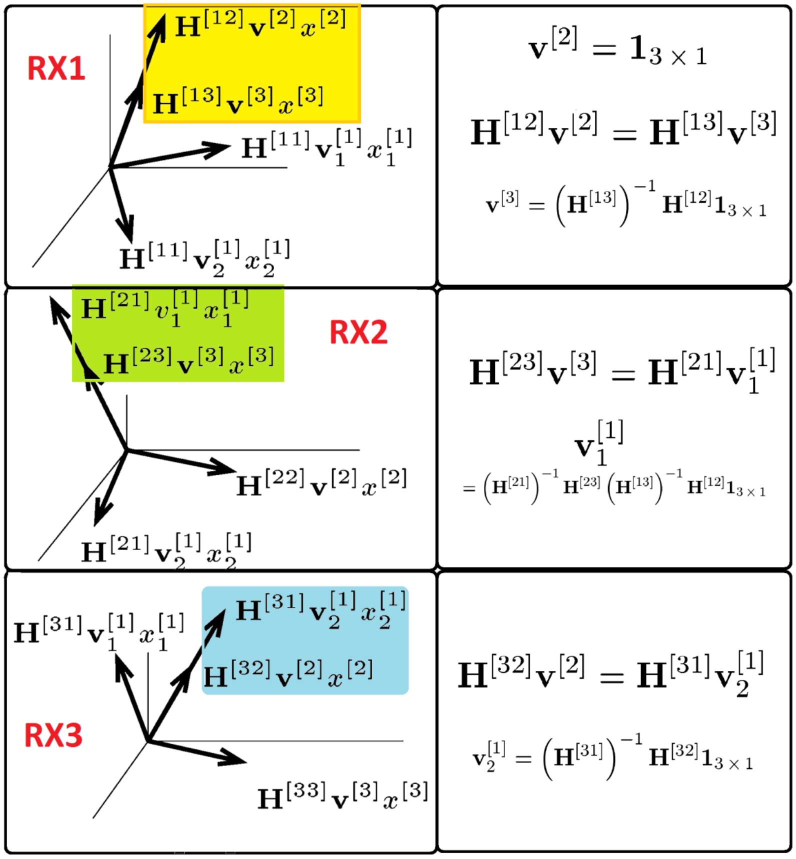

Figure 4.1: 3-user IC (adapted from Cadambe and Jafar, 2008)

The application of the concept of IA to wireless networks has shown that it is possible to achieve multiplexing gains larger than the gains that can be realized using the conventional techniques; i.e., higher slopes of the achievable rate curve at high SNRs. A simple example that illustrates the IA precoding scheme is presented in the next subsection.

4.1.1 3-user Interference Channel

Consider the 3-user interference channel (IC) shown in Figure 4.1, where each transmitter has a message for its receiver. Cadambe and Jafar (2008) showed that for

$K = 3$ it is possible to achieve $3n + 1$ degrees of freedom using $2n + 1$ dimensions. These dimensions can be time or frequency extensions or spatial dimensions. In other words, as n gets large, the DoF of the network approaches $3/2$, which means that the DoF per user is $\frac{1}{2}$. More generally, this result applies for any K. To gain some insight into these statements, the following analysis shows how for $n = 1$ one can transmit 4 symbols using 3 dimensions. Assume that each terminal has 3 antennas and the linear precoding scheme is designed in such a way so that receiver 1 can decode two messages $x_1^{[1]}$ and $x_1^{[2]}$ while each of receivers 2 and 3 decodes a single message $x^{[2]}$ and $x^{[3]}$, respectively, without interference. If $\mathbf{H}^{[ij]} \in \mathbb{C}^{3\times 3}$ is the channel matrix (of full rank) between transmitter j and receiver i and $\mathbf{v}_\ell^{[j]}$ is the transmit beamformer used at transmitter j for data stream ℓ, then the received signal at receiver 1 can be written as

$$\mathbf{y}^1 = \mathbf{H}^{[11]}\big(\mathbf{v}_1^{[1]}x_1^{[1]} + \mathbf{v}_2^{[1]}x_2^{[1]}\big) + \mathbf{H}^{[12]}\mathbf{v}^{[2]}x^{[2]} + \mathbf{H}^{[13]}\mathbf{v}^{[3]}x^{[3]} + \mathbf{z}^{[1]}, \tag{4.1}$$

where $\mathbf{z}^{[1]}$ is the additive white Gaussian noise. Now, if $\mathbf{v}^{[2]} = [1\ 1\ 1]^T$ and $\mathbf{v}^{[3]}$ is designed such that

$$\mathbf{H}^{[13]}\mathbf{v}^{[3]} = \mathbf{H}^{[12]}\mathbf{v}^{[2]}, \tag{4.2}$$

i.e., $\mathbf{v}^{[3]} = {\mathbf{H}^{[13]}}^{-1}\mathbf{H}^{[12]}\mathbf{v}^{[2]}$, then the interference arriving at receiver 1 due to the transmitted signals of transmitters 1 and 2 spans one dimensional space instead of 2, i.e., the interference aligned into reduced dimensional signal space at receiver 1. Therefore, there exists a two dimensional signal space at receiver 1 that is interference-free and the messages $x_1^{[1]}$ and $x_1^{[2]}$ can be extracted using linear projection techniques, such as zero-forcing. Similarly, if $\mathbf{v}_1^{[1]}$ is designed such that $\mathbf{H}^{[21]}\mathbf{v}_1^{[1]} = \mathbf{H}^{[23]}\mathbf{v}^{[3]}$, i.e., $\mathbf{v}_1^{[1]} = {\mathbf{H}^{[21]}}^{-1}\mathbf{H}^{[23]}\mathbf{v}^{[3]}$, then receiver 2 can project the interference arriving from transmitters 1 and 3 in order to decode $x^{[2]}$. Finally, if $\mathbf{v}_2^{[1]}$ is designed such that $\mathbf{H}^{[31]}\mathbf{v}_2^{[1]} = \mathbf{H}^{[32]}\mathbf{v}^{[2]}$, i.e., $\mathbf{v}_2^{[1]} = {\mathbf{H}^{[31]}}^{-1}\mathbf{H}^{[32]}\mathbf{v}^{[2]}$, then receiver 3 can project the interference arriving from transmitters 1 and 2 in order to decode $x^{[3]}$.

The application of the concept of IA to certain wireless network settings has shown that it is possible to achieve multiplexing gains larger than the gains that can be realized using the conventional interference avoidance techniques such as time division multiple access (TDMA) or frequency division multiple access (FDMA). In particular, for the 2-user multiple-input multiple-output (MIMO) X channel, with M antennas at each terminal, IA can achieve the degrees of freedom (DoF) of the channel, namely $4M/3$ (Jafar and Shamai, 2008). Furthermore, for the K-user fading interference channel, a scheme that achieves $K/2$-DoF has been developed (Cadambe and Jafar, 2008), although that scheme does require a number of linearly independent channel realizations that increases exponentially with K.

In the development of IA schemes, the assumption of global and perfect instantaneous CSI is typically imposed, as seen in the previous example. Some of the principles are being extended to certain other cases (e.g., Maddah-Ali and Tse, 2010; Maleki *et al.*, 2012; Jafar, 2012a). As the main target is to implement IA schemes in practice, the challenge is to design IA schemes that are amenable to systems with limitations on the rate of feedback of CSI, and to determine how the rate of feedback should scale with SNR in order to maintain desirable performance. In the next section, the system model for the G-cell MIMO IBC is presented and the required conditions for achieving interference-free reception and hence the DoF of the system.

In the following sections, IA schemes are developed that are well suited to a limited feedback implementation in the K-user 2 and 3-cell MIMO interfering broadcast channel (IBC).

4.2 System Model for the G-cell Network

The system model for the G-cell MIMO IBC consists of G transmitters or base stations (BSs), each of which has M_t transmit antennas and communicates to K users. Each user has M_r receive antennas and the kth user in the ith cell, user (i, k), receives $d^{i,k}$ data streams. The received signal at user (i, k) is

$$\tilde{\mathbf{y}}^{i,k} = \sum_{j=1}^{G} \tilde{\mathbf{H}}_j^{i,k} \tilde{\mathbf{x}}_j + \tilde{\mathbf{n}}^{i,k}, \quad (4.3)$$

where $\tilde{\mathbf{y}}^{i,k} \in \mathbb{C}^{M_r}$, $\tilde{\mathbf{H}}_j^{i,k} \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix between user (i, k) and BS j, $\tilde{\mathbf{x}}_j$ is the transmitted signal from BS j and is subject to the average power constraint $\mathrm{E}\big[\|\tilde{\mathbf{x}}_j\|^2\big] \leq P$, and $\tilde{\mathbf{n}}^{i,j}$ is the zero-mean additive white circular Gaussian noise with unit variance. The channel matrices $\tilde{\mathbf{H}}_j^{i,k}$ are assumed to be arbitrary time-invariant matrices with full rank. Since some of the linear precoding schemes presented in this thesis are based on T_c channel uses, let $\mathbf{y}^{i,k} = \big[\tilde{\mathbf{y}}^{i,k}[1], \tilde{\mathbf{y}}^{i,k}[2], \ldots, \tilde{\mathbf{y}}^{i,k}[T_\mathrm{c}]\big]^T$, where $\tilde{\mathbf{y}}^{i,k}[t]$ is the received signal at user (i, k) during the t-th channel use, and define $\mathbf{x}_j$ and $\mathbf{n}^{i,k}$ analogously. With these definitions in place, the received signal at user (i, k) over T_c channel uses can be expressed as

$$\mathbf{y}^{i,k} = \sum_{j=1}^{G} \mathbf{H}_j^{i,k} \mathbf{x}_j + \mathbf{n}^{i,k}, \quad (4.4)$$

where $\mathbf{H}_j^{i,k}$ is a block diagonal matrix, with diagonal blocks $\tilde{\mathbf{H}}_j^{i,k}$, which can be expressed as:

$$\mathbf{H}_j^{i,k} = \mathbf{I}_{T_\text{c}} \otimes \tilde{\mathbf{H}}_j^{i,k}, \tag{4.5}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{I}_{T_\text{c}}$ is the identity matrix of size T_c.

This thesis considers linear interference alignment schemes in which the signal transmitted from BS j takes the form

$$\mathbf{x}_j = \mathbf{T}_j\mathbf{s}_j = \mathbf{\Phi}_j\mathbf{V}_j\mathbf{s}_j = \mathbf{\Phi}_j\sum_{k=1}^{K}\mathbf{V}_j^k\mathbf{s}_j^k, \tag{4.6}$$

where $\mathbf{T}_j$ is a two-tier transmit precoder used at BS j in which $\mathbf{\Phi}_j$ is a projection matrix and is designed to eliminate the inter-cell interference (ICI), $\mathbf{V}_j^k$ is the matrix of beamformers for user (j,k) and is designed to eliminate the intra-cell interference inside cell j, and $\mathbf{s}_j^k \in \mathbb{C}^{d^{j,k}}$ is the vector of data symbols for user (j,k). The details of the construction of the matrices $\mathbf{\Phi}_j$ and $\mathbf{V}_j^k$ will be discussed in the following sections.

One typical assumption here is that the CSI is perfectly estimated at the receiver and that each receiver multiplies the received signal with a matrix of unit norm beamformers, $\mathbf{W}^{i,k} \in \mathbb{C}^{M_r\times d^{i,k}}$, so that scalar decoding can be employed. Accordingly, the "equalized" signal at user (i,k) is

$$\begin{aligned}\hat{\mathbf{y}}^{i,k} &= \mathbf{W}^{i,k\dagger}\Big(\sum_{j=1}^{G}\mathbf{H}_j^{i,k}\mathbf{x}_j + \mathbf{n}^{i,k}\Big)\\ &= \mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i\mathbf{V}_i^k\mathbf{s}_i^k + \mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i\sum_{\ell\neq k}^{K}\mathbf{V}_i^\ell\mathbf{s}_i^\ell + \mathbf{W}^{i,k\dagger}\sum_{j\neq i}^{G}\mathbf{H}_j^{i,k}\mathbf{x}_j + \tilde{\mathbf{n}}^{i,k},\end{aligned} \tag{4.7}$$

where the first term represents the desired signal, the second term represents the interference term due to transmissions to users in the same cell, the third term represents the interference term due to transmissions to users in other cells, and $\tilde{\mathbf{n}}^{i,k} = \mathbf{W}^{i,k\dagger}\mathbf{n}^{i,k}$ is the effective noise after combining.

In order to achieve interference-free transmission without removing the desired signal term, $\{\mathbf{W}^{i,k}\}_{i=1,k=1}^{G,K}$,$\{\mathbf{\Phi}_i\}_{i=1}^{G}$ and $\{\mathbf{V}_i\}_{i=1}^{G}$ must be designed so that:

$$\operatorname{rank}\left(\mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i\mathbf{V}_i^k\right) = d^{i,k}, \tag{4.8a}$$

$$\left(\mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i\mathbf{V}_i^\ell\right) = \mathbf{0} \quad \forall \ell \neq k, \tag{4.8b}$$

$$\left(\mathbf{W}^{i,k\dagger}\mathbf{H}_j^{i,k}\mathbf{\Phi}_j\sum_k \mathbf{V}_j^k\right) = \mathbf{0} \quad \forall j \neq i, \forall k. \tag{4.8c}$$

The DoF of a system that satisfies these simultaneous equations is $\sum_{i=1}^{G}\sum_{k=1}^{K} d^{i,k}/T_{\mathrm{c}}$ and the average DoF per cell is $\frac{\sum_{i=1}^{G}\sum_{k=1}^{K} d^{i,k}}{GT_c}$. The conventional TDMA approach to interference avoidance corresponds to a solution to (4.8) in which $T_{\mathrm{c}} = GK\min(M,N)$, each $\mathbf{F}_i^k$ has one non-zero block row (of height M) and $d^{i,k} = \min(M,N)$. As a result, the (generic) DoF of that scheme is $\min(M,N)$, which is the same DoF per cell for an isolated point-to-point MIMO link. One goal of IA is to obtain solutions to (4.8) that result in higher DoF. Unfortunately, the polynomial nature of these simultaneous equations tends to make them difficult to solve in the general case. Indeed, in the absence of special structure in the channel matrices $\{\mathbf{H}_j^{i,k}\}$ it is NP-hard to determine whether the set of equation in (4.8) has a solution, let alone finding a solution (Razaviyayn *et al.*, 2012; Bresler *et al.*, 2014). Fortunately, in certain settings involving a small number of antennas or structured channel matrices, the problem becomes simpler (Razaviyayn *et al.*, 2012) and certain classes of solutions have been obtained (e.g., Suh *et al.*, 2011; Suh and Tse, 2008; Shin *et al.*, 2010; Medra and Davidson, 2013). The linear precoding scheme for the isolated 2-cell network proposed in the following sections is another example. Further, in the next chapter general finite networks of arbitrary structure will be considered, and linear precoding schemes that satisfy (4.8) while only requiring local feedback will be developed. For unbounded networks with typical architectures, the proposed schemes leverage the designed linear precoding schemes for finite networks in order to develop linear precoding schemes that can completely eliminate the intra-cell interference and the dominant sources of inter-cell interference using only local feedback.

To obtain some insight into solutions for (4.8), it can be observed that for the G-cell MIMO IBC with constant channel that seeks to provide one interference-free dimension per user and hence provides GK DoF, the feasibility analysis in Liu and Yang (2013) and Zhuang *et al.* (2011) shows that any linear precoding scheme requires

$$M_t + M_r \geq GK + 1. \tag{4.9}$$

If the total DoF of the G-cell network normalized by the dimensions at any user is defined as

$$\mathrm{DoF}_n = \frac{GK}{\min(M_t, M_r)}, \tag{4.10}$$

then, when $M_t = M_r$, the DoF_n is upper bounded by

$$\mathrm{DoF}_n < 2. \tag{4.11}$$

One question of practical importance is how close can a precoding scheme come to that bound under constraints on the available CSI. For the 2-cell case, several schemes that achieve the DoF bound (4.11) using modest amounts of feedback, including the case of only local feedback, have been proposed (e.g., Shin *et al.*, 2010; Suh and Tse, 2008). For the G-cell network with a special class of $(G-1)$-level decomposable [1] channel matrices (e.g., line of sight (LOS) channels), Suh and Tse (2008) provided a precoding scheme for the uplink case that requires $(\sqrt[G-1]{K}+1)^{G-1}$ dimensions and can achieve

$$\frac{GK}{(\sqrt[G-1]{K}+1)^{G-1}} \tag{4.12}$$

DoF. As the number of users per cell, K, increases, this scheme can achieve G DoF. However, this scheme requires that the physical channel possess a specific structure and that structure might not be realized in practice. In Section 4.3, a linear precoding scheme will be proposed for the isolated 2-cell MIMO IBC for system with constant channel that does not require time extensions and achieves the optimal spatial-DoF for such a network; i.e., it achieves the bound in (4.9) with equality.

4.3 Linear Precoding Scheme for 2-cell MIMO IBC

In this section, a linear precoding scheme is proposed that achieves the optimal spatial-DoF for the isolated 2-cell MIMO IBC with constant channels without time extension, i.e., $T_c = 1$. In the model of the considered system, each BS has $M_t = K+1$ transmit antennas, where K is the number of users per cell. Each user has $M_r = K$ receive antennas. Further each BS sends one data stream to each of the K users in its cell; i.e. $d^{i,k} = 1$. In order to simplify the presentation, the case of $K = 2$ users per cell will be considered first and the extension to the case of $K > 2$ will then be established. For the limited feedback implementation, the scaling laws for the number of feedback bits required to maintain the (optimal) DoF of the system will be

[1] The m level decomposable matrix is the matrix which can decomposed into the Kronecker product of m matrices.

provided. An observation from this analysis is that having an extra antenna at each user greatly reduces the number of bits that need to be fed back to the base station. As an alternative to providing that additional hardware, this reduction can also be realized by allowing cooperation between users.

4.3.1 The Proposed Scheme, for $K = 2$

This subsection presents the proposed technique for achieving the DoF for 2-cell MIMO IBC in the ideal case where global and perfect CSI is assumed, while the next subsection shows how the scheme is suitable for limited feedback implementation. In order to simplify the discussion, the two user case, $K = 2$, illustrated in Figure 4.2 will be considered. (The extension is discussed in Subection 4.3.3.) In the 2-user case, the 4 DoF can be achieved if each BS has $M_t = 3$ transmit antennas and each user has $M_r = 2$ receive antennas (Kim *et al.*, 2011b). The principle that underlies the proposed methods is the sequential determination of components of the solution $\{\mathbf{w}^{i,k}, \mathbf{\Phi}_i, \mathbf{V}_i\}$ to (4.8), where the sequence in which the terms are designed is chosen in such a way that when a limited feedback implementation is developed, the amount of feedback to the base stations remains manageable.

The proposed scheme with $K = 2$ requires $M_t = K + 1 = 3$ and $M_r = K = 2$. In the operation of the scheme, the first step is for the users in one cell to design the projection matrix $\mathbf{\Phi}$ to be used by the base station in the other cell. To do so, user 1 in cell 1, i.e. user $(1, 1)$, estimates $\mathbf{H}_2^{1,1} \in \mathbb{C}^{2\times 3}$, the channel matrix between BS 2 and itself and then determines a column that lies in the null space of $\mathbf{H}_2^{1,1}$. User $(1, 2)$ can

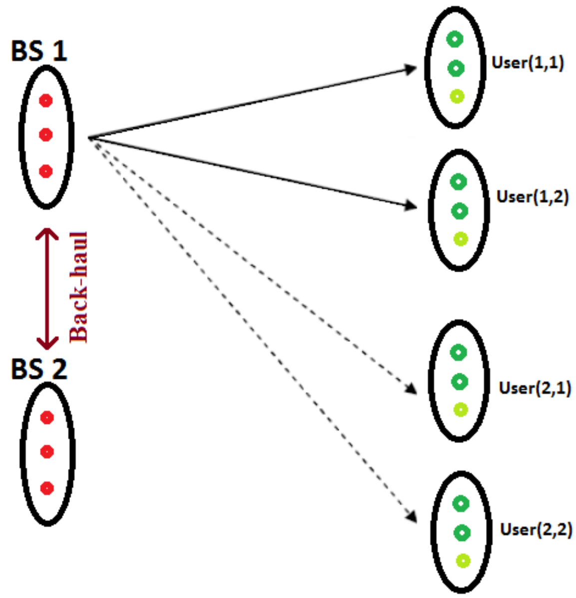

Figure 4.2: 2-cell 2-user MIMO IBC model

do the same. That is, the ith user in cell 1 designs $\boldsymbol{\phi}_2^i$ satisfying

$$\mathbf{H}_2^{1,i}\boldsymbol{\phi}_2^i = \mathbf{0}, \quad i = 1, 2. \tag{4.13}$$

The dimensions of $\mathbf{H}_2^{1,i}$ ensure that such a $\boldsymbol{\phi}_2^i$ exists. Now, users 1 and 2 in cell 1 feed back $\boldsymbol{\phi}_2^1$ and $\boldsymbol{\phi}_2^2$ to BS 1, which sends a copy to BS 2 through a dedicated back-haul for CSI exchange. The BSs allow their users to estimate $\mathbf{H}_1^{i,k}\boldsymbol{\Phi}_1$ and $\mathbf{H}_2^{i,k}\boldsymbol{\Phi}_2$ through a "dedicated" training step akin to that of Caire *et al.* (2010) for the single cell case.

Since both $\mathbf{H}_2^{1,1}\boldsymbol{\Phi}_2$ and $\mathbf{H}_2^{1,2}\boldsymbol{\Phi}_2$ have rank (at most) one, the second step is for users 1 and 2 in cell 1 to design their receive beamformers, $\mathbf{w}^{1,k}$, such that

$$\mathbf{w}^{1,k\dagger}\mathbf{H}_2^{1,k}\boldsymbol{\Phi}_2 = \mathbf{0}^T, \quad k = 1, 2. \tag{4.14}$$

Users 1 and 2 in cell 2 perform the analogous operation. Having determined $\mathbf{H}_1^{i,k}\boldsymbol{\Phi}_1$

and $\mathbf{H}_2^{i,k}\mathbf{\Phi}_2$ so that the inter-cell interference can be removed by receive beamforming, the final step is to design the transmit beamformers $\mathbf{V}$ to eliminate the intra-cell interference. In this step, each user in cell 1 feeds back its effective channel $\mathbf{h}_{\text{eff}} = \mathbf{w}^{1,k^\dagger}\mathbf{H}_1^{1,k}\mathbf{\Phi}_1$ and BS 1 defines $\mathbf{v}_1^1$ and $\mathbf{v}_1^2$ to lie in the null space of the effective channels of users 2 and 1 in cell 1 respectively, such that:

$$\mathbf{w}^{1,1^\dagger}\mathbf{H}_1^{1,1}\mathbf{\Phi}_1\mathbf{v}_1^2 = 0 \quad \text{and} \quad \mathbf{w}^{1,2^\dagger}\mathbf{H}_1^{1,2}\mathbf{\Phi}_1\mathbf{v}_1^1 = 0. \tag{4.15}$$

The transmit beamformers for users 1 and 2 in cell 2 are defined in a similar way. Being able to send 4 symbols without interference to their intended receivers, the proposed scheme is able to achieve 4 DoF.

One important point to mention here is that if the number of receive antennas is increased to $M_r = 3$ such that $M_t = M_r = K + 1$, then the 4 DoF can be achieved without cooperation between users in different cells using an approach akin to the subspace interference alignment (SIA) (Suh and Tse, 2008). To make this point clear, let us focus on the users in cell 1, user 1 and 2 design their receive beamformers $\mathbf{w}^{1,1}$ and $\mathbf{w}^{1,2}$ such that:

$$\mathbf{w}^{1,1^\dagger}\mathbf{H}_2^{1,1} \in \mathcal{N}_1 \tag{4.16a}$$

$$\mathbf{w}^{1,2^\dagger}\mathbf{H}_2^{1,2} \in \mathcal{N}_1 \tag{4.16b}$$

where $\mathcal{N}_1$ is a predetermined one-dimensional subspace that is chosen independently of the channels realizations, and hence can be made known to each user upon entry into the cell. Now, if $\mathbf{\Phi}_2$ is defined to lie in the orthogonal complement of $\mathcal{N}_1$, then the inter-cell interference will be completely eliminated. Finally the transmit beamformer

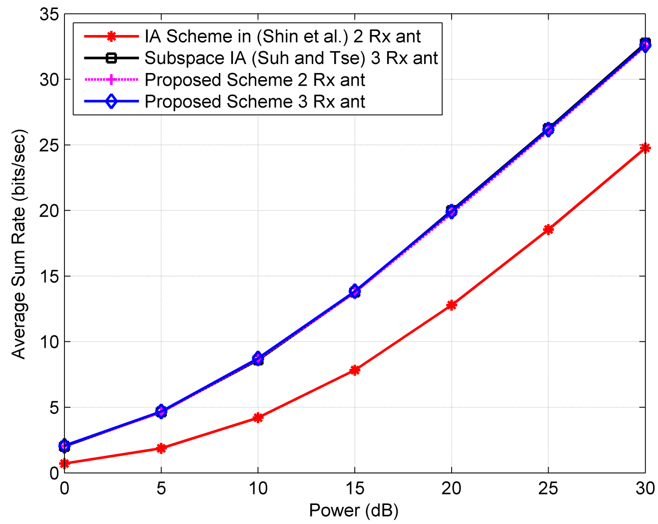

Figure 4.3: Sum rate comparison with perfect CSI for $K = 2$.

of each user is designed in the null space of the effective channel of the other users in the same cell. (The effective channel for user (i, k) is $\mathbf{w}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i$.)

Figure 4.3 compares the performance of the proposed schemes with the schemes of Suh and Tse (2008) and Shin *et al.* (2010) in the case of unlimited feedback. The figure shows that the proposed techniques achieve the maximum DoF of the channel, namely, 4, and the proposed scheme with $M_r = 2$ receive antennas achieves an improved power offset over the scheme of Shin *et al.* (2010), which also requires $M_r = 2$ receive antennas.

Before presenting the limited feedback implementation of the proposed scheme, it worth mentioning that the underlying principles for the proposed precoding scheme can be linked to the principles of the design of the precoding scheme for the 2-user X-channel developed by Maddah-Ali *et al.* (2008). In particular, both schemes exploit the structure of the channel matrices and take advantage of the null space of the communication links in order to cancel the interference and improve the system

DoF. However, the scheme proposed in this chapter differs the proposed scheme by Maddah-Ali *et al.* (2008) in three prominent ways. First, the proposed scheme is designed for a 2-cell K-user MIMO IBC, with possibly $K > 2$, while the scheme developed by Maddah-Ali *et al.* (2008) is only designed for the 2-user X-channel. Second, the design of the projection matrix Φ_j is based on the interfering channel matrices of the K users, while the corresponding aspect of thr design of Maddah-Ali *et al.* (2008) exploits the null space one direct link and one cross link. Finally, the proposed scheme develops a two-tier transmit precoder in order to cancel the intra-cell and inter-cell interference, whereas the precoder of Maddah-Ali *et al.* (2008) does not have this structure.

4.3.2 Limited Feedback Model

The limited feedback implementation of the proposed scheme is outlined in this subsection. Since the information that is to be exchanged is in the form of one-dimensional subspaces, Grassmannian quantization schemes, such as those discussed in Chapters 2 and 3, will form the core of the feedback scheme. In these vector quantization schemes, the receivers and the base station have copies of a codebook, $\mathcal{F}$, of 2^B representatives of one-dimensional subspaces. Given a representative, $\mathbf{p}$ of the subspace that it wishes to quantize, the receiver selects the index of an element of that codebook using a distortion metric, $d(\cdot,\cdot)$,

$$\hat{\mathbf{p}} = \arg\min\nolimits_{\mathbf{f}_i \in \mathcal{F}} d(\mathbf{f}_i, \mathbf{p}), \tag{4.17}$$

and then transmits that index to the base station using B bits. The base station can reconstruct $\hat{\mathbf{p}}$ using its own codebook.

To apply this scheme to the proposed method for $M_r = 2$, users in each cell have to define the projection matrix $\mathbf{\Phi} \in \mathbb{C}^{3\times 2}$ that should be used at the other cell. Accordingly, the following feedback procedures are suggested to design the transmit/receive beamformers:

1. Users 1 and 2 in cell 1 define $\boldsymbol{\phi}_2^1$ and $\boldsymbol{\phi}_2^2 \in \mathbb{C}^{3\times 1}$ that lie in the null space of $\mathbf{H}_2^{1,1}$ and $\mathbf{H}_2^{1,2}$ respectively. The receivers broadcast the indexes of the quantized versions of $\boldsymbol{\phi}_2^1$ and $\boldsymbol{\phi}_2^2$ by selecting the codewords according to:

$$\hat{\boldsymbol{\phi}}_2^i = \arg\min_{\mathbf{f}_\ell \in \mathcal{F}_1} \sqrt{1 - |\mathbf{f}_\ell^\dagger \boldsymbol{\phi}_2^i|^2}, \tag{4.18}$$

   where $\mathcal{F}_1$ is the codebook. Each user assign B_1 feedback bits to do so. Similarly, users in cell 2 define $\hat{\boldsymbol{\phi}}_1^i, i \in \{1, 2\}$ and broadcast the indexes of the corresponding codewords. Note that the BSs can retransmit the indexes of the codewords selected by users in the other cell if their users cannot access the feedback channel of the users in the other cell.

2. Knowing $\hat{\mathbf{\Phi}}_1$ and $\hat{\mathbf{\Phi}}_2$, users in each cell design their receive beamformers in order to remove ICI. For example, user 1 in cell 1 selects $\mathbf{w}^{1,1}$ to be in the null space of $\mathbf{H}_2^{1,1}\hat{\boldsymbol{\phi}}_2^2$.

3. Each user feeds back the index of a codeword representing the quantized version of its effective channel $\mathbf{h}_{\text{eff}}^{i,k} = \mathbf{w}^{i,k\dagger}\mathbf{H}_i^{i,k}\hat{\mathbf{\Phi}}_i$. Since $\mathbf{h}_{\text{eff}}^{i,k} \in \mathbb{C}^2$, a second codebook $\mathcal{F}_2$ is needed for the current quantization process where B_2 bits are assigned for feedback. Accordingly, each user feeds back a total number of bits $B = B_1 + B_2$.

Now, the number of feedback bits B_1 and B_2 should scale with the operating SNR ρ to maintain the DoF (Jindal, 2006; Lee and Ko, 2012). In order to derive a lower bound on the number of feedback bits B_1 and B_2, the rate loss, ΔR, due to quantization, will be expressed as a function of B_1 and B_2. Using this expression and taking expectation over different channel realizations and random codebooks, the rate loss can be upper bounded such that it is kept constant, $E\{\Delta R\} \leq \log_2 b$; (see Jindal, 2006; Lee and Ko, 2012; for related ideas in other contexts). Accordingly, the scaling laws of B_1 and B_2 are expressed as:

$$B_1 \geq (M_t - 1)\Big(\frac{\log_{10}\rho}{3} + \log_2 \frac{3\bar{\lambda}_{max}}{2(b-1)(M_t-1)}\Big) \tag{4.19a}$$

$$B_2 \geq \mu B_1 - (M_t - 2)\log_2 \mu \tag{4.19b}$$

where $\mu = \frac{M_t-2}{M_t-1}$, and $\bar{\lambda}_{max} = \frac{M_t M_r (M_t+M_r)^{(2/3)}}{M_t M_r + 1}$.

When $M_r = 3$, the previous procedures can be greatly simplified as no cooperation is required between terminals to achieve the DoF. For $M_r = 3$, the transmit/receive beamformer design can be summarized as follows:

1. Users in cell i design their receive beamformers $\mathbf{w}^{i,k}$ such that $\mathbf{w}^{i,k\dagger}\mathbf{H}^{i,k}_{j\neq i} \in \mathcal{N}_i$ where $\mathcal{N}_i$ is orthogonal to the projection matrix used by the BS in the other cell $\mathcal{N}_{j\neq i}$. Note that the projection matrices $\mathcal{N}_i$ are predefined and are known to each terminal.

2. Each user feeds back the B bits of the index of a codeword representing the quantized version of its effective channel $\mathbf{h}^{i,k}_{\text{eff}} = \mathbf{w}^{i,k\dagger}\mathbf{H}^{i,k}_i\mathbf{\Phi}_i$. In this case, the effective channel is quantized by the aid of a codebook $\mathcal{F}$ whose codewords are vectors of dimension 2.

The scaling law of the number of feedback bits B in step 2 can be expressed as:

$$B \geq (M_t - 2)\left[\frac{\log_{10}\rho}{3} + \log_2 \frac{\bar{\lambda}_{max}}{2(M_t-2)(b-1)}\right] \tag{4.20}$$

Figure 4.4 compares the number of total number of feedback bits for $M_r = 2$ and $M_r = 3$ for the case of $E\{\Delta R\} \leq \log_2 4$, where it is clear that there is a large reduction in the number of the feedback bits by adding one additional receive antenna at each user.

Even though the reduction in the feedback requirements is achieved by increasing the spatial dimensions at receivers, this reduction can be also achieved without any additional antennas by allowing users' cooperation. Shin *et al.* (2010) showed that the DoF for the 2-cell MIMO network can be achieved by assuming that users can cooperate using a Wi-Fi link. Users can also cooperate using cognitive radio techniques. Under these assumptions, the users can design the projection matrices $\mathbf{\Phi}_i$ and perform the quantization process with sufficiently high accuracy. The remaining step of quantizing the effective channel for each user can be performed by using one codebook and requires only B bits of feedback that scale with the same rule as in (4.20).

4.3.3 Generalization for $K > 2$

The extension of the proposed scheme for $K > 2$ users per cell is straightforward. For the case where $M_t = K + 1$ transmit antennas and $M_r = K$ receive antennas, the design process can be summarized as:

1. Users in cell i define the column space of the projection matrix $\mathbf{\Phi}_{j\neq i} \in \mathbb{C}^{M\times K}$

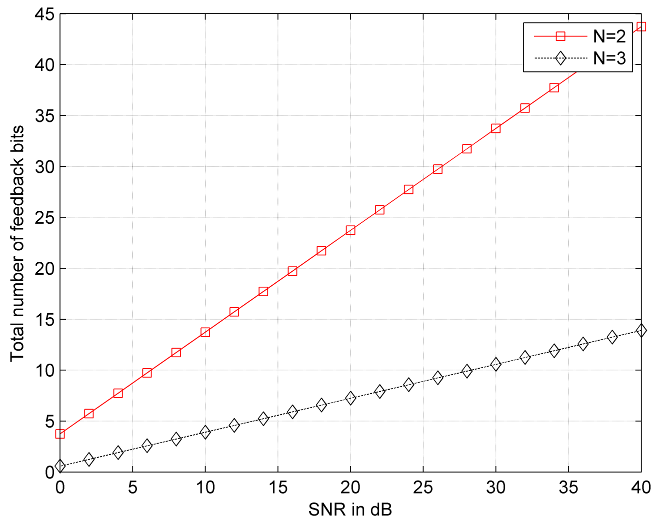

Figure 4.4: Number of feedback bits required to maintain a constant rate loss for the proposed schemes.

that will be used in the other cell. As explained before, each user defines $\boldsymbol{\phi}_{j \neq i}^{k}$ that lie in the null space of the channel matrix between this user and the BS in the other cell.

2. Based on a quantization codebook $\mathcal{F}_1$, each user broadcasts the index of the codeword representing the quantized version of $\boldsymbol{\phi}_{j \neq i}^{k}$. Each user assigns B_1 feedback bits for this quantization process.

3. Given $\hat{\boldsymbol{\Phi}}_1$ and $\hat{\boldsymbol{\Phi}}_2$, the users design their receive beamformers to remove the ICI.

4. Finally, each user feeds back the index of the codeword from a codebook $\mathcal{F}_2$ that represents the quantized version of its effective channel. This process requires B_2 bits. The total number of feedback bits per data stream is $B = B_1 + B_2$.

In the case of $M_r = K+1$, the design process of the transmit/receive beamformers is exactly the same process described in the previous section with one exception, which

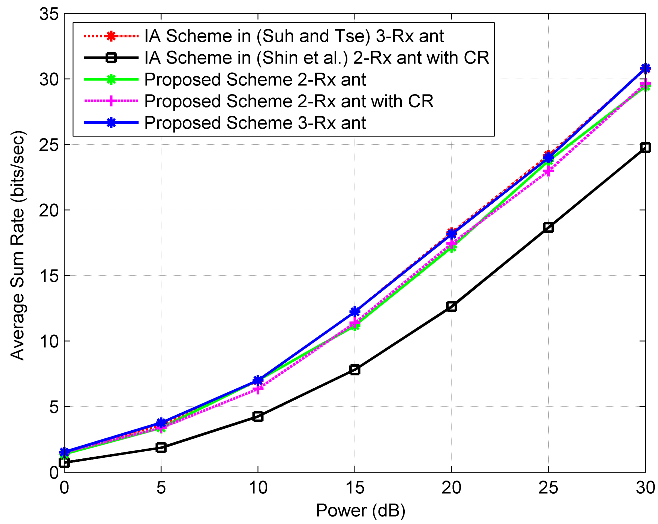

Figure 4.5: Sum rate comparison of various schemes under the limited feedback model

is the codebook $\mathcal{F}$ consists of unit norm vectors $\mathbf{f}_\ell$ of size K.

To conclude this subsection it is pointed out that the proposed scheme and the precoding scheme developed by Shin *et al.* (2010) achieve the optimal spatial DoF of $2K$. This observation illustrates the fact that, in general, there is more than one precoding scheme that can achieve the DoF of the channel. However, among these schemes that achieve the DoF of the channel, some may provide higher achievable rates than others, some may be able to achieve the DoF with a smaller feedback budget and some may offer other desirable characteristics.

4.3.4 Simulation Results

This subsection compares the performance of limited-feedback implementations of the proposed schemes with analogous implementations of the existing schemes (Suh and Tse, 2008; Shin *et al.*, 2010). In Figure 4.5, the average sum rate of these schemes is compared when each cell communicates with $K = 2$ users. Each BS has $M_t = 3$ antennas, and the users are equipped with $M_r = 2$ or $M_r = 3$ antennas. For the method of Shin *et al.* (2010) an assumption that the receivers cooperate through a cognitive radio (CR) is imposed. The performance of the proposed scheme with $M_r = 2$ will also be evaluated under that assumption. The number of feedback bits scales according to (4.19) and (4.20). From Figure 4.4, it can be deduced that adding an additional antenna at each receiver enables a substantial reduction in the number of feedback bits, especially at high SNR. Moreover, Figure 4.5 shows that the proposed scheme with $M_r = 2$ gives comparable performance with the scheme presented of Suh and Tse (2008), which requires 3 antennas at receiver, and provides a large power offset when compared to the scheme of Shin *et al.* (2010). That said, the proposed scheme with $M_r = 2$, and that of Shin *et al.* (2010), require more feedback than the scheme of Suh and Tse (2008) and the proposed scheme with $M_r = 3$, or cooperation between the receivers.

4.3.5 Discussion and Extensions

Although the scheme presented in this section achieves the optimal spatial-DoF of the isolated 2-cell MIMO network for systems that do not allow precoding over multiple channel uses, when $M_r = K$, the elimination of the inter-cell interference is dependent on the coordination between the BSs (to exchange the $\mathbf{\Phi}_j$'s.) This requires the presence of a dedicated back-haul between the BSs for CSI exchange. The proposed scheme also requires several rounds of dedicated training for the users. On the other hand, when the number of receive antennas is increased from $M_r = K$ to $M_r = K + 1$, the feedback requirements have been greatly reduced. In particular, with $M_t = M_r = K+1$, the proposed scheme for the isolated 2-cell MIMO network in this section and also the subspace interference alignment scheme (SIA) of Suh and Tse (2008) have many interesting features:

1. Each user need only feed back its effective channel with the scheduling BS. It does not need to feed back channels to the other BSs. The inter-cell interference can be totally eliminated at the user side.

2. No cooperation between BSs is required. Therefore, there is no need for a dedicated back-haul for CSI exchange.

3. The (normalized) DoF is
$$\mathrm{DoF}_n = \frac{2K}{K+1}, \tag{4.21}$$
which approaches 2 as K increases; cf. (4.11).

The extension of these techniques to the 3-cell case is quite straightforward. One way that it can be done is that each BS constructs a projection matrix $\mathbf{\Phi} \in \mathbb{C}^{\bar{M}_t \times K}$, where $\bar{M}_t = 2K+1$ in this case, and the users feed back their effective channels. Though this extension possesses the first two features of the 2-cell case, it results in a DoF loss. The (normalized) DoF in this case is

$$\mathrm{DoF}_n = \frac{3K}{2K+1}, \tag{4.22}$$

which approaches 1.5 as K increases. Fortunately, this scheme can be modified in such a way to achieve the 2 DoF in the 3-cell case, but it requires more feedback from users. Let us consider a modified version of the SIA (Mod-SIA), which works as follows:

1. BS 3 chooses $\mathbf{\Phi} \in \mathbb{C}^{(2K+1)\times 2K}$.

2. Users in cell 1 design a receive beamforming vector to cancel the interference resulting from BS 3 and feed back their effective channel to their BS (i.e., BS 1) and the effective channel between them and BS 2.

3. Given the previous information, BS 1 can design the projection matrix $\mathbf{\Phi}_2$ to lie in the null space of the concatenated matrix of the effective channels between users in cell 1 and BS 2. Using a dedicated back-haul, BS 1 sends $\mathbf{\Phi}_2$ to BS 2, which will be used at cell 2.

4. Similar procedures apply for users in cell 2.

5. Users in cell 3 can design their receive beamformers to cancel the inter-cell interference from BS 1 and 2 as it spans $2K$ dimensions out of the $2K+1$ dimensional signal space.

6. As in the 2-cell case, the BSs use the effective channels feedback between them and their scheduled users to design the transmit beamformers to cancel the intra-cell interference.

In the next chapter, structured linear precoding schemes that can be implemented for a G-cell MIMO IBC and require only local feedback for operation are proposed. As one example, a precoding scheme is proposed with $M_t = M_r$ for the isolated 3-cell MIMO IBC that can achieve the normalized 2-DoF using only local feedback within the same cell. This removes the need for a back-haul for the CSI exchange between base stations in different cells, in addition to the reduced feedback requirements.

Chapter 5

Spatial Reuse Precoding for Scalable Downlink Networks

In Chapter 4, a linear precoding scheme for the isolated 2-cell MIMO downlink network was developed that can achieve the optimal spatial-DoF of the network. One important conclusion from that development is that schemes that require only local feedback are quite attractive for implementation in limited feedback systems. Other conclusions were that the extension of the proposed isolated 2-cell scheme to the 3-cell case requires more feedback, otherwise the DoF will drop, and that it is not quite clear whether or not the proposed scheme can be extended to the general G-cell case.

In this chapter, linear precoding schemes are developed for MIMO downlink networks with slowly-varying channels that are scalable in the sense that they can be implemented with moderate complexity in networks with increasing numbers of cells and users. Further, a class of precoders that can be implemented in unbounded networks with typical architectures and in heterogeneous networks is developed.

5.1 Introduction

As has been stated several times in this thesis, one of the fundamental limitations on the communication efficiency of wireless networks is interference. Several approaches have been proposed to reduce the interference experienced by the users in a cellular network. The simplest are based on sending the signals orthogonally in time (TDMA) or frequency (FDMA). Although such interference avoidance approaches are simple, it may be possible to make better use of the available spectrum if interference is actively managed rather than being avoided (Gesbert *et al.*, 2007, 2010). An intriguing approach to doing so is that of interference alignment (IA) (Maddah-Ali *et al.*, 2008; Cadambe and Jafar, 2008; Jafar, 2011), which have been reviewed in Chapter 4. In Chapter 4, it was stated that the basic principle that underlies (linear) IA is to design the signalling in such a way that the interference components arriving at a particular receiver lie in a reduced dimensional subspace of the space spanned by the received signal. The potential spectral efficiency gains suggested by these schemes, and schemes developed for small networks as the isolated 2-cell MIMO IBC scheme in Chapter 4, have stimulated substantial research interest in the understanding of the performance of such schemes and their limitations, and in the development of pragmatic schemes that facilitate implementation in practice, (e.g., Gesbert *et al.*, 2010; Yetis *et al.*, 2010; Liu and Yang, 2013; Jindal, 2006; Cadambe *et al.*, 2010; Maddah-Ali and Tse, 2012; Maleki *et al.*, 2012; Suh *et al.*, 2011; Medra and Davidson, 2013; Lozano *et al.*, 2013). The focus of the work presented herein is on developing pragmatic schemes based on the principles of IA that are scalable as the size of the network grows. The proposed schemes mitigate the dominant sources of interference and can be implemented without requiring significantly more feedback or coordination between cells than current systems that manage interference on a cell-by-cell basis.

For small isolated networks (i.e., small values of G), the application of IA leads to substantial performance gains compared to the conventional interference avoidance techniques (Suh *et al.*, 2011; Gomadam *et al.*, 2011; Suh and Tse, 2008; Shin *et al.*, 2010; Medra and Davidson, 2013). For example, the subspace interference alignment (SIA) scheme of Suh *et al.* (2011) for the isolated 2-cell network permits the cancellation of the inter-cell interference, achieving interference-free transmission as the number of users increase. The application of a variant of the SIA for the isolated 3-cell network leads to 300% gain in the cell-edge user rate when compared with existing schemes (Suh *et al.*, 2011).

The linear precoding scheme in Section 4.3 and other related schemes (Suh *et al.*, 2011; Gomadam *et al.*, 2011; Suh and Tse, 2008; Shin *et al.*, 2010; Medra and Davidson, 2013) seek to design the precoding schemes in such a way that the fundamental DoF limits on the chosen networks are achieved. However, the extension from small scale isolated networks to a large scale network tends to result in schemes that are rather complicated to implement. One of the fundamental issues is the requirement of a central processing unit for channel state information (CSI) collection and precoder design. For a small scale network, the amount of resources that must be allocated for CSI exchange is sometimes out-weighed by the performance improvement that is obtained, but as the number of cooperating base stations increases, the mechanisms for CSI exchange process become rather involved and may consume a significant fraction of the available resources. As an example, the SIA scheme (Suh *et al.*, 2011) is capable of achieving the DoF of the isolated 2-cell downlink network by using only local "in-cell" feedback, but the extension of that scheme to the G-cell network appears to be difficult. Moreover, the problem of exchanging side information becomes more complicated in the case of heterogeneous networks that consists of different tiers of cells, e.g., macro and pico cells.

Although the extension to an arbitrary network remains intricate, the extension to many practical cellular networks presents an opportunity, namely that of partial connectivity in the network (Guillaud and Gesbert, 2011; Liu and Yang, 2014). In networks with a small number of cells, the receivers are often presumed to be close enough to all the transmitters for the interference to be deemed to be significant. In that scenario, examining the DoF and the high-SNR performance of the network generates considerable insight. However, as the size of the network grows, at moderate SNRs the power of interference from distant transmitters in the network may fall significantly below the noise level, and practical precoding schemes (and CSI feedback schemes) for such SNRs ought to take advantage of this partial connectivity; see also (Jafar, 2012b). A related perspective, albeit in a somewhat different context, arises in the development of the notion of a "DoF regime" in wireless networks (Lozano *et al.*, 2013).

With those perspectives in mind, the goal of this chapter is to develop linear precoding schemes for large downlink networks that provide improved performance over conventional interference avoidance schemes. The designed schemes can achieve this improved performance without requiring a significant increase in the fraction of the communication resources that must be allocated to the exchange of side information. In particular, it would be desirable if this could be achieved using only local feedback within each cell. If that is possible, there is no need for a dedicated back-haul for CSI exchange, nor for a central processing unit for precoder design. Further, approaches that can exploit the inherent path loss in the wireless networks that leads to partial connectivity of the network need to be developed. Finally, it would be desirable if the proposed precoding schemes were able to integrate some smaller cells, e.g. pico-cells, with low overhead and without affecting the performance of the original network.

The main contribution of this chapter is to develop precoding schemes that realize these goals in the case of quasi-static channels. The proposed precoders are designed to enable a concept that will be called Spatial Reuse Precoding (SRP). SRP describes a network precoding scheme that can be designed so that the signals from the dominant interfering sources at each receiver align in a reduced dimensional subspace. This enables the receiver to eliminate the interference using a simple projection operation. This interference alignment is designed to be achieved regardless of the exact values of the channel matrices between the interfering sources and the user experiencing that interference. The key observation that enables this property is that while a channel matrix can change the direction of a signal vector, it cannot change the subspace that it spans. Hence, if the subspaces spanned by the interference are aligned, interference alignment can be achieved irrespective of the channel. The notion of subspace alignment was also discussed by Suh *et al.* (2011). This subspace alignment property enables the possibility of network extension without any cooperation between existing cells and new cells. In other words, precoding schemes that are designed to exhibit the SRP property for a particular isolated 2 or 3-cell arrangement, can be used, with simple modification, in larger networks in which the BSs are deployed in certain sorts of site arrangements. In those cases, the SRP scheme is designed to eliminate the dominant sources of interference so that rates higher than those of conventional interference avoidance schemes can be achieved, and to do so using a finite number of channel uses with only local feedback.

In order to present the underlying concepts of the proposed schemes, this chapter begins by proposing precoding schemes for the isolated 2-cell and 3-cell networks. Here, the chapter focuses on the basic features of the proposed structured precoding schemes that facilitate the SRP property. In particular, a Kronecker structured IA precoding scheme (Kron-IA) is developed that, in those isolated networks, can achieve greater DoFs than the conventional interference avoidance techniques. The remainder of this chapter then explains how these schemes can be used to construct spatial reuse precoding schemes for unbounded cellular networks with linear or hexagonal architecture. Further, an example is provided of how such SRP schemes can lead to substantial performance gains in the sum achievable rate of a heterogeneous network that consists of linear arrangement of macro and micro-BSs in addition to hot-spots.

5.2 Principles of Proposed Approach for Arbitrary Networks of a Specified Size

This chapter relies on the system model for the G-cell MIMO IBC presented in Section 4.2. Further, the linear precoding schemes developed in this chapter are designed for systems that employ blocks of T_{c} channel uses over which the channel remains constant. As a result, the channel matrices $\mathbf{H}_j^{i,k}$ have the Kronecker structure in (4.5). Here, a class of two-tier precoders is considered,

$$\mathbf{T}_j^k = \mathbf{\Phi}_j \mathbf{V}_j^k, \tag{5.1}$$

where $\mathbf{\Phi}_j$ is a projection matrix designed so that the receivers can eliminate the inter-cell interference and $\mathbf{V}_j^k$ is the matrix of beamformers designed to eliminate the intra-cell interference inside cell j. This class of precoders has already proven effective in a number of settings (e.g., Chen and Lau, 2014; Medra and Davidson, 2013). For example, for the 2-cell MIMO interference broadcast channel, it was shown in Chapter 4 how the sets $\{\mathbf{\Phi}_j\}_j$, $\{\mathbf{W}^{i,k}\}_{i,k}$ and $\{\mathbf{V}_j^k\}_{j,k}$ could be chosen to achieve interference alignment for systems that restrict precoding over a single channel use (i.e., $T_\mathrm{c} = 1$), and to achieve the optimal (spatial) DoF in this setting (cf. Liu and Yang, 2013). However, like many related schemes, that scheme involves the exchange of certain forms of channel state information (CSI) between the BSs, requires several rounds of "dedicated" training (in the sense of Caire *et al.* (2010)), and appears to be difficult to extend to the general setting of $G > 2$ cells.

In the proposed approach, the design of the system components is performed sequentially. First, each transmitter constructs its projection matrix $\mathbf{\Phi}_j$ so that inter-cell interference can be eliminated by the users. A feature of the proposed approach is that construction of $\mathbf{\Phi}_j$ is performed without the need for the CSI. Next, each user designs its linear receive beamformer $\mathbf{W}^{i,k}$ to eliminate the inter-cell interference. (Variations on the choice of $\mathbf{W}^{i,k}$ are discussed in Section 5.6.6.) With $\mathbf{\Phi}_j$ and $\mathbf{W}^{i,k}$ designed in this way, the operation of each cell resembles that of an isolated single-cell downlink with effective channels

$$\mathbf{H}_{\mathrm{eff}}^{i,k} = \mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i. \tag{5.2}$$

Therefore, each receiver feeds back (a quantized version of) $\mathbf{H}_{\mathrm{eff}}^{i,k}$ to its serving BS and that BS designs the matrices $\mathbf{V}_i^k$ as if it were serving a single-cell downlink;

e.g., block diagonalization (Ravindran and Jindal, 2008), zero-forcing beamforming (Spencer *et al.*, 2004b), a quality of service design (Rashid-Farrokhi *et al.*, 1998; Bengtsson and Ottersten, 2001), or one of many other choices. Finally, there is a dedicated training phase within each cell so that each receiver (i,k) can determine $\mathbf{H}_{\text{eff}}^{i,k}\mathbf{V}_i^k$ and hence can perform coherent detection (e.g., Caire *et al.*, 2010). It is clear from (5.2) that the choice of $\{\boldsymbol{\Phi}_j\}_j$, and in particular its column rank, will affect the number of users that each BS can serve in each cell. This will play an important role in the proposed designs.

In the initial designs in Sections 5.3–5.5 for isolated fully-connected networks with a specified number of cells, the goal of the proposed schemes is to achieve more DoF than classical interference avoidance schemes while requiring only local feedback and also enabling spatial reuse precoding. To achieve these goals, the projection matrices $\boldsymbol{\Phi}_j$ in (5.1) will be designed independently of the channels, but will be designed so that with knowledge of the subspace spanned by $\{\mathbf{H}_j^{i,k}\boldsymbol{\Phi}_j\}_{j\neq i}$, user (i,k) can design $\mathbf{W}^{i,k}$ in such a way that the inter-cell interference is removed, i.e., ${\mathbf{W}^{i,k}}^{\dagger}\mathbf{H}_j^{i,k}\boldsymbol{\Phi}_j = \mathbf{0}$, $\forall j$. Users in cell i can determine the subspace spanned by $\{\mathbf{H}_j^{i,k}\boldsymbol{\Phi}_j\}_{j\neq i}$ from the signal received when BS i is turned off and all other BSs are transmitting (e.g., Krim and Viberg, 1996). That approach has the advantage that it does not require explicit training to determine each $\mathbf{H}_j^{i,k}\boldsymbol{\Phi}_j$ for $j \neq i$. Further, it lends itself naturally to a scheme in which each user (i,k) only attempts to project out the "dominant" interfering signals and hence need only estimate the dominant subspace of the signal received when BS i is not transmitting; see also Section 5.6.6.

To begin the development of the proposed spatial reuse precoding schemes, and in particular the choices for the set of projection matrices $\{\boldsymbol{\Phi}_j\}_j$, isolated 2-cell and 3-cell networks will be considered and linear interference alignment schemes will be developed for these networks that do not require BS cooperation. These schemes involve only a modest number of channel extensions, but are not necessarily DoF optimal for the networks that they are designed for. However, they provide better performance than conventional interference avoidance techniques, they have a number of features that facilitate their implementation, and they enable spatial reuse precoding.

5.3 Kronecker Structured Linear Precoding for 2-cell MISO-IBC

5.3.1 Basic Kronecker Structured Scheme

Let us consider an isolated 2-cell MISO downlink network with M_t transmit antennas and K receivers per cell, each with $M_r = 1$ receive antenna, that each receive a single data stream; i.e., $d^{i,k} = 1$. For a system that does not allow time extensions, it is optimal, from the DoF perspective, to turn off one BS, and have the other BS communicate to $K = M_t$ users. The resulting system has M_t DoF and achieves them without the need for CSI exchange between the BSs.

Now, let us consider the case in which a block of T_c channel uses is permitted, rather than just a single channel use (and the channels remain constant over that block). In that case, the equivalent channel matrices have the Kronecker structure in (4.5). Motivated by that structure, a scheme in which each BS transmits (up to) βM_t symbols in those T_c channel uses, with $\beta < T_c$, is proposed. The linear projection matrix $\mathbf{\Phi}_j$ at BS j will be constructed as the Kronecker product of two matrices.

That is,

$$\mathbf{\Phi}_j = \mathbf{\Gamma}_j^1 \otimes \mathbf{\Gamma}_j^2, \tag{5.3}$$

where the matrices $\mathbf{\Gamma}_j^1 \in \mathbb{C}^{T_c \times \beta}$ and $\mathbf{\Gamma}_j^2 \in \mathbb{C}^{M_t \times M_t}$ are randomly and independently generated from a continuous distribution. By construction, the generic rank of $\mathbf{\Phi}_j$ is βM_t.

To examine the performance of this precoder, the interference that BS 2 imposes on a user $(1,k)$ in cell 1 is first considered. Since $\mathbf{V}_2$ will be a matrix of full column rank, the subspace spanned by the interference from BS 2 is the column span of the "interference matrix"

$$\mathbf{Z}^{1,k} = \mathbf{H}_2^{1,k}\mathbf{\Phi}_2 \in \mathbb{C}^{T_c \times \beta M_t}. \tag{5.4}$$

The generic rank of $\mathbf{H}_2^{1,k}$ is T_c and that of $\mathbf{\Phi}_2$ is βM_t, but $\mathbf{H}_2^{1,k}$ has the Kronecker structure in (4.5) and $\mathbf{\Phi}_2$ has the Kronecker structure in (5.3). As a result, $\mathbf{Z}^{1,1}$ can be written as

$$\mathbf{Z}^{1,k} = \mathbf{\Gamma}_2^1 \otimes (\tilde{\mathbf{h}}_2^{1,k}\mathbf{\Gamma}_2^2), \tag{5.5}$$

where $\tilde{\mathbf{h}}_2^{1,k} \in \mathbb{C}^{1 \times M_t}$. Therefore,

$$\begin{aligned}\operatorname{rank}(\mathbf{Z}^{1,k}) &= \operatorname{rank}(\mathbf{\Gamma}_2^1)\operatorname{rank}(\tilde{\mathbf{h}}_2^{1,k}\mathbf{\Gamma}_2^2) \\ &\leq \beta \min(\operatorname{rank}(\tilde{\mathbf{h}}_2^{1,k}), \operatorname{rank}(\mathbf{\Gamma}_2^2)) = \beta. \end{aligned} \tag{5.6}$$

That is, the structure of the components of $\mathbf{Z}^{1,k}$ results in $\mathbf{Z}^{1,k}$ being rank deficient, with rank at most β. That means that the dimension of the subspace spanned by the interference at user $(1,k)$ is at most β. Since β was chosen to be less than T_c, there is a signal subspace in which receivers in cell 1 can receive signals from their serving

BS 1, free from interference from BS 2. This result has the interesting interpretation that allowing precoding over T_c uses of a channel that remains constant enables for time-interference alignment. That is, by choosing $\mathbf{\Phi}_j$ in the form of (5.3), the received interference dimension collapses as the result of the alignment of the interfering signal subspaces in the time domain.

The fact that the interference matrix $\mathbf{Z}^{1,k}$ is rank deficient for every user $(1,k)$ in cell 1 is promising, in the sense that with the knowledge of the space spanned by $\mathbf{H}_2^{1,k}\mathbf{\Phi}_2$ each of these users can design a unit norm receive vector $\mathbf{w}^{1,k}$ to cancel the intra-cell interference, while leaving dimensions available for communication with assigned BS. Closer study of (5.5) reveals that at each user the interfering subspace is the column span of $\mathbf{\Gamma}_2^1$ and hence is independent of the realization of the channel between BS 2 and user $(1,k)$, $\tilde{\mathbf{h}}_2^{1,k}$. That has the advantage that each user user $(1,k)$ can eliminate the inter-cell interference using only the knowledge of $\mathbf{\Gamma}_2^1$; knowledge of the realization of the interfering channel $\tilde{\mathbf{h}}_2^{1,k}$ is not required. However, the generic dimension of the null space of $\mathbf{\Gamma}_2^1$ is only $T_c - \beta$ and hence BS 1 can serve at most $(T_c - \beta)M_t$ users in an interference-free manner. The general form of that result is stated in the following theorem.

Theorem 1. *For an isolated 2-cell MISO-IBC operating in a quasi-static environment and implementing the precoding scheme in* (5.1) *with* $\mathbf{\Phi}_j$ *chosen according to* (5.3)*, the number of users per cell* K *that can be served in an interference-free manner is upper bounded by* $K \leq (T_c - \beta)M_t$*. As a result, the DoF of the network is upper bounded by* M_t*.*

Proof. The design of the precoding scheme in (5.3) enables each user to cancel the inter-cell interference using a linear receive beamformer $\mathbf{W}^{i,k}$. The receivers feed back their "in-cell" effective channels for their assigned BS, where $\mathbf{H}_{\text{eff}}^{i,k} = \mathbf{W}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i$; cf. (5.2). The number of users K that BS i can serve is the rank of the matrix $\mathbf{H}_{\text{eff}}^{i} = [\mathbf{H}_{\text{eff}}^{i,1^T}, \mathbf{H}_{\text{eff}}^{i,2^T}, \ldots, \mathbf{H}_{\text{eff}}^{i,K^T}]^T$. The structure of the channel matrix $\mathbf{H}_i^{i,k}$ in (4.5) implies that the channel matrix of the (M_t+1)th user can be written as a linear combination of the effective channel matrices of the previous M_t users. Further, since $\mathbf{W}^{i,k}$ is designed to eliminate the inter-cell interference, it lies in the null space of $\mathbf{Z}^{i,k}$, i.e., $\text{rank}(\mathbf{W}^i) = \mathcal{N}(\mathbf{Z}^{i,k}) = \mathcal{N}(\mathbf{\Gamma}_j^1), i \neq j$, where $\mathcal{N}(\mathbf{A})$ is the dimension of the null space of $\mathbf{A}$. Since $\mathbf{\Gamma}_j^1 \in \mathbb{C}^{T_c \times \beta}$, then its null space has a dimension of $T_c - \beta$. Therefore, $K \leq (T_c - \beta)M_t$. $\square$

The result in Theorem 1 is somewhat disappointing in the sense that using the basic Kron-IA precoding scheme in (5.3) in a generic isolated 2-cell network does not provide any DoF improvement when compared to systems that employ only a single channel use nor systems utilizing conventional interference avoidance schemes, such as TDMA. Fortunately, this basic scheme can be modified so that it can achieve more DoF, as will be shown in Subsection 5.3.3, and it enables SRP, as will be shown in Section 5.6. Before doing so, it is important to point out that there are scenarios in which the basic scheme in (5.3) does offer improved performance.

5.3.2 Two Case studies

One case where the physical channel between any user and the interfering BS is constant while the physical channel between any user and its scheduling BS is time-varying. In that case, the basic scheme in (5.3) can achieve the optimal $\frac{2\beta M_t}{\beta+1}$DoF, which approaches $2M_t$ as β, and hence T_c, increases (Park and Lee, 2009). Interestingly, not only does it achieve the optimal-DoF of the network, but it does so without

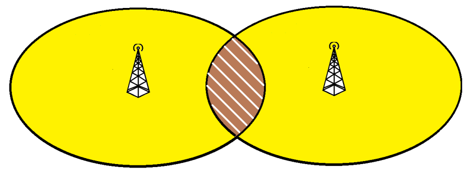

Figure 5.1: Partially connected 2-cell MISO IBC. Users in the shaded brown area suffer from significant inter-cell interference, while other users do not.

any cooperation between cells and using only local feedback. Although this channel model may appear to be quite abstract, it is a useful limiting case for scenarios in which the users in each cell are distant from the interfering BS and close enough to their BS. Accordingly, any small movements by the users will significantly affect the direct links with the scheduling BS and will have negligible effects on the links to the interfering BS.

A second case, which may be of more interest, is the partially-connected 2-cell MISO network illustrated in Figure 5.1. In this model, some users in each cell suffer from significant inter-cell interference (those in the brown shaded area in Figure 5.1), while the interference incurred by the other users is modeled as being negligible (with respect to the noise). In order to discuss a tangible case, let us assume that 50% of the users in each cell experience significant inter-cell interference from the other cell, while the remaining users can be modelled as having interference-free reception (regardless of the transmit power). Those users that do not suffer from interference can design their receive beamformers $\mathbf{w}^{i,k}$ in order to improve the desired signal power. Due to the partial connections between users and BSs, if the basic scheme in (5.3) is implemented with $T_c = \frac{3\beta}{2}$ for any (even) value of β, each BS is capable of communicating to $K = \beta M_t$ users per cell and the DoF of this network is $\frac{2K}{T_c} = \frac{4M}{3}$. In this partially connected network, the proposed basic scheme achieves $4M_t/3$ DoF, which is larger than the M_t DoF that can be achieved using time or frequency division between the cells, and it does so without any additional feedback requirements.

One might ask what happens if the percentage of users that suffer from interference decreased to 25%? In that case, setting $T_{\rm c} = \frac{5\beta}{4}$ for any value of β (that is a multiple of 4) will produce an increase in the DoF to $8M_t/5$. In other words, in the proposed basic scheme, the number of channel uses $T_{\rm c}$ can be adapted to the number of users that suffer from significant interference, and as this number decreases, $T_{\rm c}$ also decreases. However, one has to keep in mind that in this setting partial connectivity is assumed, where a fraction of users suffer from interference while the remaining fraction do not. In case of a fully connected network, it would be desirable to implement the modified version of the basic scheme that is described in the next subsection.

5.3.3 Modified Kronecker Structured Scheme

The weakness of the basic Kronecker structured scheme in Section 5.3.1 in the case of fully connected network arises from the fact that although that scheme enables the receivers to eliminate the inter-cell interference, without the need for coordination between BSs, it does not leave many dimensions available for intra-cell communications. In particular, in each cell i, the matrix formed by stacking the row vectors $\mathbf{h}_{\text{eff}}^{i,k} = \mathbf{w}^{i,k\dagger}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i$ in (5.2) is not full row rank. This precludes BS i from being able to communicate to all its users using linear precoding without (intra-cell) interference. To modify the basic Kronecker structured scheme the proposed scheme seeks a set of matrices $\{\mathbf{\Phi}_j\}_j$ that have more (linearly independent) columns than the matrices generated by (5.3), and yet still enable the receivers to remove the inter-cell interference without the need for BS coordination or inter-cell feedback. One way to do this is to augment the projection matrices in (5.3) with randomly generated matrices (from a continuous distribution) $\mathbf{\Theta}_j \in \mathbb{C}^{T_c M_t \times \alpha}$, where α is a parameter of the design. That is, $\mathbf{\Phi}_j \in \mathbb{C}^{T_c M_t \times \beta M_t}$ in (5.3) is replaced by

$$\mathbf{\Phi}_j = \begin{bmatrix} \mathbf{\Gamma}_j^1 \otimes \mathbf{\Gamma}_j^2 & \mathbf{\Theta}_j \end{bmatrix} \tag{5.7}$$

with $\mathbf{\Gamma}_j^1$, $\mathbf{\Gamma}_j^2$ as defined after (5.3). The generic rank of $\mathbf{\Phi}_j$ in (5.7) is $\beta M_t + \alpha$. The following theorem shows that for a particular class of triples $(T_\mathrm{c}, \beta, \alpha)$, parametrized by β, this modified Kronecker scheme can achieve an improved DoF (without the need for BS cooperation or inter-cell feedback).

Theorem 2. *For a fully-connected isolated* 2*-cell MISO-IBC operating in a quasi-static environment, a scheme that implements the precoders in* (5.1) *with* $\mathbf{\Phi}_j$ *chosen according to* (5.7)*,* T_c *chosen to be* $\lceil \frac{2\beta M_t - \beta - 1}{M_t - 1} \rceil$ *and* α *chosen to be* $T_c - \beta - 1$*, can achieve* $2(1 + \frac{(M_t-1)^2}{2M_t - 1})$ *DoF as* β*, and hence* T_c*, increases.*

Proof. Using the fact that the number of scheduled users per cell cannot exceed the rank of $\mathbf{\Phi}_j$, $\mathrm{rank}(\mathbf{\Phi}_j) = \beta M_t + \alpha$. Now, the interference matrix $\mathbf{Z}^{1,k}$ can be written as

$$\mathbf{Z}^{1,k} = (\mathbf{I}_{T_\mathrm{c}} \otimes \tilde{\mathbf{h}}_2^{1,k}) \mathbf{\Phi}_j, = \begin{bmatrix} \mathbf{Z}_1^{1,k} & \mathbf{Z}_2^{1,k} \end{bmatrix} \tag{5.8}$$

where $\mathbf{Z}_1^{1,k} = \mathbf{\Gamma}_1^2 \otimes \hat{\mathbf{h}}_2^{1,k}$ and $\mathbf{Z}_2^{1,k} = (\mathbf{I}_{T_\mathrm{c}} \otimes \tilde{\mathbf{h}}_2^{1,k})\mathbf{\Theta}_2$. Since $\mathrm{rank}(\mathbf{Z}^{1,k}) \leq \mathrm{rank}(\mathbf{Z}_1^{1,k}) + \mathrm{rank}(\mathbf{Z}_2^{1,k}) \leq \beta + \alpha$, in order to have a rank deficient $\mathbf{Z}^{1,k}$ that would enable a linear receiver to project out the interference, the number of channel uses T_c is lower bounded by $T_\mathrm{c} \geq \beta + \alpha + 1$. Further, to have a full rank matrix of effective channels for the design of the transmit beamformers, $\mathbf{v}^{i,k}$, the number of users cannot exceed

$K \leq (T_\mathrm{c} - \beta)M_t$. By setting $\alpha = T_\mathrm{c} - \beta - 1$ and using the fact that $K \leq \beta M_t + \alpha$, then $(T_\mathrm{c} - \beta)M_t = \beta M_t + \alpha = \beta M_t + T_\mathrm{c} - \beta - 1$ which implies that

$$T_\mathrm{c} = \left\lceil \frac{2\beta M_t - \beta - 1}{M_t - 1} \right\rceil. \tag{5.9}$$

Therefore, the DoF of the proposed scheme is $2K/T_\mathrm{c}$, i.e.,

$$\mathrm{DoF} = 2\Big(1 + \frac{(\beta M_t - \beta - 1)}{T_\mathrm{c}}\Big). \tag{5.10}$$

As T_c increases, the DoF approaches $2(1 + \frac{(M_t-1)^2}{2M_t-1})$. $\square$

To provide some numerical context for that result, let us consider the case of $M_t = 3$ transmit antennas and $M_r = 1$ receive antenna. If attention is restricted to schemes that involve a single channel use (i.e., $T_\mathrm{c} = 1$), no linear precoding scheme can achieve more than 3 DoF, no matter which signaling scheme we use. However, using the modified Kronecker scheme proposed above, it can achieve up to 3.6 DoF, even though the channel remains constant over the signaling interval. Moreover, that result is achieved without BS cooperation and using only local feedback.

5.4 Kronecker Structured Linear Precoding For G-cell MIMO-IBC with $M_t > M_r$ Antennas

This section briefly outlines the natural extension of the modified Kronecker structured scheme in Section 5.3.3 to the general case of an arbitrary G-cell MIMO IBC with M_r antennas at each user and $M_t > M_r$ antennas at each BS. In this setting, projection matrices of the form in (5.7) can be employed directly. The following theorem shows that for a particular choice of the triple $(T_{\mathrm{c}}, \beta, \alpha)$ that is dependent on M_t, M_r, and G, an improved DoF can be achieved, again without BS cooperation and using only local feedback.

Theorem 3. *For a fully connected G-cell MIMO-IBC with $M_t \geq 1$ receive antennas and $M_t > M_r$ transmit antennas operating in a quasi-static environment, a scheme that implements the precoders in* (5.1) *with $\mathbf{\Phi}_j$ chosen according to* (5.7)*, T_c chosen to be $\lceil \frac{\beta(G-1)(GM_t-N)-1}{(G-1)M_t-N} \rceil$ and α chosen to be $\lfloor \frac{T_c M_r-(G-1)\beta M_r-1}{G-1} \rfloor$ can achieve $G\big(\frac{M_r}{G-1} + \frac{((G-1)M_t-M_r)(M_t-M_r)}{(G-1)(GM_t-M_r)}\big)$ DoF as β, and hence T_c, increases.*

Proof. See Appendix D. □

For the isolated 2-cell MIMO-IBC with $M_t = 3$ transmit antennas and $M_r = 2$ receive antennas, Theorem 3 reveals that this system can achieve up to 4.5 DoF as T_{c} increases, which is substantially larger than the 3 DoF achieved using conventional interference avoidance and is larger than the 4 DoF of the system discussed in Chapter 4. (The system in Chapter 4 also requires cooperation between BSs.) Furthermore, a large fraction of this DoF can be achieved using only few channel uses. For example, with $\beta = 2$, then $\alpha = 9$, $T_{\mathrm{c}} = 7$. In that case, the analysis in Appendix D shows that the DoF = 4.28. When $\beta = 3$, then $\alpha = 15$ and $T_{\mathrm{c}} = 11$ and the DoF is 4.36.

5.5 Kronecker Structured Linear Precoding for MIMO IBC with $M_t = M_r$ Antennas

In this section, some specialized choices for the projection matrices $\mathbf{\Phi}_j$ are developed for the case of $M_t = M_r$ antennas and isolated 2 and 3-cell cases.

5.5.1 $G = 2$ and $M_t = M_r$ Antennas

In this setting, the subspace interference alignment (SIA) precoding scheme in (Suh *et al.*, 2011) can be easily extended to include the case of block based signaling with blocks of T_c channel uses. That extension corresponds to using the precoding scheme in (5.7) with $\beta = 0$ and $T_\mathrm{c} = (\alpha + 1)/M_r$. Using the analysis of Suh *et al.* (2011), it can be shown that the DoF of that scheme is $\frac{2(T_\mathrm{c} M_r - 1)}{T_\mathrm{c}}$ which approaches $2M_r$ as T_c increases and is achieved without cooperation between BSs.

5.5.2 $G = 3$ and $M_t = M_r$ Antennas

For the isolated 3-cell MIMO IBC with $M_t = M_r$ antennas, a linear precoding scheme is considered in which the projection matrices $\mathbf{\Phi}_j$ takes a similar structure to that in (5.3), but with slightly different dimensions:

$$\mathbf{\Phi}_j = \mathbf{\Gamma}_j^1 \otimes \mathbf{\Gamma}_j^2, \tag{5.11}$$

where $\mathbf{\Gamma}_j^1 \in \mathbb{C}^{T_\mathrm{c} \times \beta}$ and $\mathbf{\Gamma}_j^2 \in \mathbb{C}^{M_t \times (M_t - 1)}$ are independently generated random matrices from continuous distributions and $\beta < T_\mathrm{c}$. By construction, the generic rank of $\mathbf{\Phi}_j$ (5.11) is $\beta(M_t - 1)$. Using the analysis in Appendix E, it can be shown that with $T_\mathrm{c} = \lceil \frac{3\beta}{2} \rceil$ this scheme can achieve a DoF of $2(M_t - 1)$ without BS cooperation and using only local feedback for any finite extension of channel uses T_c. In Section 5.6, it will be shown that it gives rise a particularly straightforward implementation of spatial reuse precoding in the case of a hexagonal cell layout.

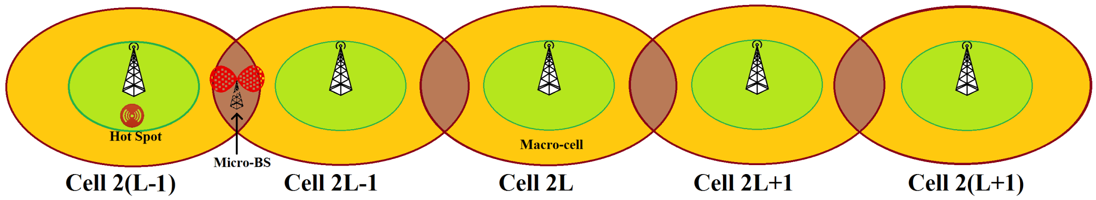

Figure 5.2: Cell-arrangement in a linear model. Users at the cell edge (brown area) suffer from significant inter-cell interference, while other users do not. Micro-BSs and hot spots can be deployed in this model, and an example is given in cells $2(L-1)$ and $2L-1$.

5.6 Spatial Reuse Precoding for Structured Networks of Unspecified Size

In this section, the proposed schemes leverage insights into the basic properties of the Kronecker structured precoding schemes for isolated 2-cell and 3-cell networks to extend them to unbounded networks in which the structure and operation of the network result in the neighboring connections being dominant. The proposed schemes scale to networks of arbitrary size, do not require BS cooperation, and employ only local feedback. In the conceptual development in this section abstract partially-connected models for the networks will be used. In those models weak connections are modelled as being absent. The design process is guided by the goal of achieving an improved DoF on those partially-connected network models over the DoF of conventional interference avoidance schemes (which also scale to networks of arbitrary size, without BS cooperation and using only local feedback).

5.6.1 A Motivating Example

Consider a simple example based on the linear downlink network illustrated in Figure 5.2, in the absence of the micro-BS and the hotspot. If the power transmitted by each BS is controlled so that users that are close to their BSs (i.e., in the green area) do not suffer from significant interference, then the abstract partially-connected network model is rather sparse, with only cell-edge users suffering significant interference. One way to improve the performance of these users is to structure the transmissions in such a way that each user can cancel one source of interference at its side. As one example, the precoders designed for the isolated 2-cell case in Section 5.3 provide this structure.

Now, if the BSs increase their transmit power in an attempt to provide better levels of service to their assigned users, the size of the green areas in Figure 5.2 will shrink and more users will suffer from significant interference. Let us consider the case in which each user receives non-negligible interference from both of the neighboring cells and the interference from all other cells is negligible. The abstract partially-connected network model for this scenario is considerably more dense than in the previous scenario. In particular, it appears that in order to apply signaling techniques from the isolated 2-cell model to this case, every third cell should be turned off. Unfortunately, doing so results in the same DoF (and in some cases even worse) as that achieved using conventional interference avoidance techniques.

To examine this scenario more closely, let us consider the case in which each BS has M_t antennas, each receiver has a single antenna, and signaling is performed in blocks of T_{c} channel uses. Now, let us examine cell $2L$ in Figure 5.2, in which the users experience interference from BSs $2L-1$ and $2L+1$. If each BS uses a linear precoding scheme with two-tier precoders of the form of (5.1), then the subspace spanned by the interference at user $(2L, k)$ is the column span of

$$\mathbf{Z}^{2L,k} = \begin{bmatrix}\mathbf{H}^{2L,k}_{2L-1}\mathbf{\Phi}_{2L-1} & \mathbf{H}^{2L,k}_{2L+1}\mathbf{\Phi}_{2L+1}\end{bmatrix}. \tag{5.12}$$

However, if BSs $2L-1$ and $2L+1$ use the same precoder, $\mathbf{\Phi}_{\mathrm{odd}}$, and if $\mathbf{\Phi}_{\mathrm{odd}} = \mathbf{\Gamma}^1_{\mathrm{o}} \otimes \mathbf{\Gamma}^2_{\mathrm{o}}$ is designed according to (5.3), then $\mathbf{Z}^{2L,k}$ takes the form

$$\mathbf{Z}^{2L,k} = \begin{bmatrix}\mathbf{\Gamma}^1_{\mathrm{o}} \otimes \hat{\mathbf{h}}^{2L,k}_{2L-1} & \mathbf{\Gamma}^1_{\mathrm{o}} \otimes \hat{\mathbf{h}}^{2L,k}_{2L+1}\end{bmatrix}, \tag{5.13}$$

where $\hat{\mathbf{h}}^{2L,k}_{2L\pm1} = \tilde{\mathbf{h}}^{2L,k}_{2L\pm1}\mathbf{\Gamma}^2_{\mathrm{o}}$. Interestingly, since BSs $2L-1$ and $2L+1$ use the same projection matrix $\mathbf{\Phi}_{\mathrm{odd}}$, the subspace spanned by $\mathbf{Z}^{2L,k}$ depends only on $\mathbf{\Gamma}^1_{\mathrm{o}}$; i.e., the interfering signals arriving from BSs $2L-1$ and $2L+1$ arrive in the same subspace, regardless of the channel realization. Generically, that subspace is of dimension β. Since β is chosen to be less than T_{c}, this enables each receiver to remove the inter-cell interference, without cooperation between the BSs, and leaves interference-free dimensions available for BS $2L$ to transmit to its desired users. Choosing the same precoder $\mathbf{\Phi}_{\mathrm{even}}$ for all even-indexed BSs provides the users in the odd-indexed cells with analogous ability to remove the inter-cell interference.

5.6.2 Definitions

The above example conjures a notion of "Spatial Reuse Precoding" (SRP), a precoding scheme designed in such a way that the signals transmitted by all the interfering sources that use the same precoder align together into a reduced dimensional subspace of the signals received at the unintended receivers. This enables the alignment of the inter-cell interference without requiring the knowledge of the inter-cell channels at the interfering BSs. In a complementary way, one can define the notion of "Spatial Reuse Factor" (SRF) as the rate at which the same precoder is used in the network. In the previous example, the SRF is 2.

The following sections will show below that the structured precoding schemes presented in the previous sections can be modified in such a way to allow SRP in unbounded downlink cellular networks that are either linear or approximately hexagonal in structure.

5.6.3 Unbounded Linear Network

This subsection considers an unbounded MIMO downlink network with cells that are deployed in a linear manner analogous to the deployment in Figure 5.2. The proposed structured precoding scheme is akin to that in (5.7), but with a slight modification. Let us define the projection matrix $\mathbf{\Phi}_j$ as

$$\mathbf{\Phi}_j = \begin{cases} \left[\mathbf{\Gamma}_{\mathrm{o}}^1 \otimes \mathbf{\Gamma}_{\mathrm{o}}^2 \quad \mathbf{\Theta}_j\right] & \text{if } j \text{ is odd} \\ \left[\mathbf{\Gamma}_{\mathrm{e}}^1 \otimes \mathbf{\Gamma}_{\mathrm{e}}^2 \quad \mathbf{\Theta}_j\right] & \text{if } j \text{ is even} \end{cases} \tag{5.14}$$

where $\mathbf{\Gamma}^1 \in \mathbb{C}^{T_{\mathrm{c}} \times \beta}$, $\mathbf{\Gamma}^2 \in \mathbb{C}^{M_t \times M_t}$ and $\mathbf{\Theta}_j \in \mathbb{C}^{T_{\mathrm{c}} M_t \times \alpha}$ are independent randomly generated matrices from continuous distributions. This scheme assigns the same structured matrix $\mathbf{\Gamma}_{\mathrm{o}}^1 \otimes \mathbf{\Gamma}_{\mathrm{o}}^2$ to every odd-indexed BS and $\mathbf{\Gamma}_{\mathrm{e}}^1 \otimes \mathbf{\Gamma}_{\mathrm{e}}^2$ to every even-indexed BS. If each user experiences interference from the two adjacent BSs, this kind of implementation permits IA regardless of the exact channel matrix between the user and each intefering BS, under the condition that the channel remains constant for T_{c} channel uses. Indeed, for the kth user in cell $2L$, the interference matrix

$\mathbf{Z}^{2L,k} = \left[\mathbf{H}_{2L-1}^{2L,k}\mathbf{\Phi}_{2L-1} \quad \mathbf{H}_{2L+1}^{2L,k}\mathbf{\Phi}_{2L+1}\right]$ can be written as

$$\mathbf{Z}^{2L,k} = \left[\mathbf{\Gamma}_{\mathrm{o}}^{1} \otimes (\tilde{\mathbf{H}}_{2L-1}^{2L,k}\mathbf{\Gamma}_{\mathrm{o}}^{2}) \quad \mathbf{H}_{2L-1}^{2L,k}\mathbf{\Theta}_{2L-1} \quad \mathbf{\Gamma}_{\mathrm{o}}^{1} \otimes (\tilde{\mathbf{H}}_{2L+1}^{2L,k}\mathbf{\Gamma}_{\mathrm{o}}^{2}) \quad \mathbf{H}_{2L+1}^{2L,k}\mathbf{\Theta}_{2L+1}\right]. \quad (5.15)$$

Since $\mathrm{Span}\left[\mathbf{\Gamma}_{\mathrm{o}}^{1} \otimes (\tilde{\mathbf{H}}_{1}^{2,k}\mathbf{\Gamma}_{\mathrm{o}}^{2})\right] = \mathrm{Span}\left[\mathbf{\Gamma}_{\mathrm{o}}^{1} \otimes (\tilde{\mathbf{H}}_{3}^{2,k}\mathbf{\Gamma}_{\mathrm{o}}^{2})\right]$, if the number of channel uses T_{c} satisfies

$$T_{\mathrm{c}} \geq (\beta M_r + 2\alpha + 1)/M_r \quad (5.16)$$

then $\mathbf{Z}^{2L,k}$ is rank-deficient and each user is capable of removing the inter-cell interference at its side. Theorem 4 determines the DoF of the proposed precoding scheme for the unbounded linear model with interference only from neighboring cells.

Theorem 4. *For a partially-connected unbounded linear MIMO-IBC operating in a quasi-static environment, in which each user suffers from interfering signals from the neighbor cells only, a scheme that implements the precoders in* (5.1) *with* $\mathbf{\Phi}_j$ *chosen according to* (5.14)*,* T_c *chosen to be* $\lceil\frac{\beta(4M_t-M_r)-1}{2M_t-M_r}\rceil$ *and* α *chosen to be* $\lfloor\frac{T_cM_r-\beta M_r-1}{2}\rfloor$ *can achieve* $\left(\frac{M_r}{2} + \frac{(2M_t-M_r)^2}{2(4M_t-M_r)}\right)$ *DoF per cell as* β*, and hence* T_c*, increases.*

Proof. See Appendix F. □

To gain some insight into the achievable DoF (under the abstract partially-connected model) of this scheme and to compare it to conventional methods, let us assume that each BS has $M_t = 3$ transmit antennas. The following observations hold

1. Using the conventional TDMA technique, eliminating the interference at any user requires that half the BSs are transmitting and the other half are turned off. Therefore, the average DoF of each cell is $M_t/2 = 1.5$ symbols per channel use. This 1.5 DoF/cell can be achieved without any cooperation between cells,

other than synchronization of the activation of the BSs.

2. Now, let us consider the scheme developed by Suh *et al.* (2011) and the scheme developed in Chapter 4 for the isolated 2-cell MIMO-IBC in which each receiver has $N = 3$ and $N = 2$ receive antennas respectively. In order to eliminate the inter-cell interference, every third BS must be off. For example, when BS 1 and 2 cooperate together in designing the precoding scheme, BS 3 is turned off. Since these scheme can achieve 4 DoF for the isolated 2-cell system, the average DoF/cell in this case is $4/3 = 1.33$. Surprisingly TDMA achieves a larger DoF than pairing neighboring cells as isolated 2-cell networks and turning every third BS off. This is despite the fact that the scheme of Suh *et al.* (2011) and the scheme in Chapter 4 achieve more DoF than TDMA in the case of an isolated 2-cell network. Furthermore, the scheme in Chapter 4 for the isolated 2-cell model requires coordination between BSs in order to eliminate the inter-cell interference.

3. Theorem 4 shows that by using the precoding scheme in (5.14), as T_c increases, the DoF/cell approaches 1.64 when there is $M_r = 1$ receive antenna at each user, approaches 1.8 when $M_r = 2$, and approaches 2 when $M_r = 3$. Furthermore, a large fraction of these DoF can be achieved in only a few channel uses; when $M_r = 3$, a DoF of 1.9 in 9 channel uses can be achieved. Moreover, the proposed precoding scheme does not require BS cooperation and uses only local feedback.

To illustrate the impact of the improved DoF of the proposed scheme, in Figure 5.3 the achievable sum rates of various schemes have been plotted for nodes with $M_t = M_r = 3$ antennas under an abstract partially-connected model for the linear cell arrangement with interference only from the neighboring cells. For the proposed

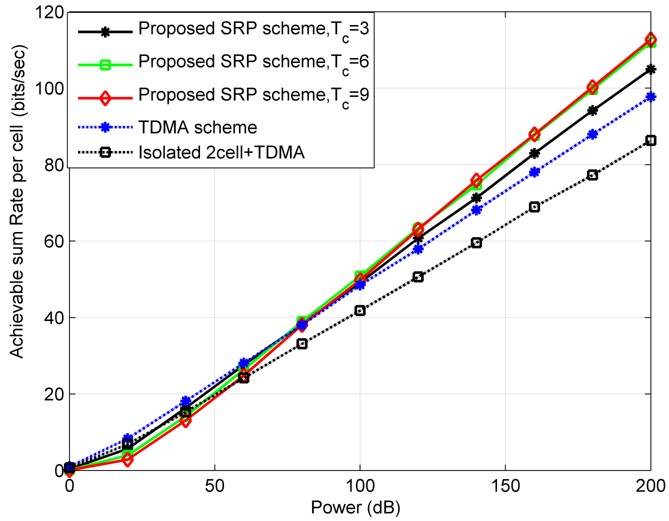

Figure 5.3: Illustration of the performance of the proposed schemes in an abstract partially-connected model for the linear cell arrangement in which each user suffers from interference from the neighboring cells only

schemes, three different values for β have been chosen, namely $\beta = 1$, which results in $\alpha = 2$, $T_\mathrm{c} = 3$ and $K = 5$ active users per cell; $\beta = 2$, which results in $(\alpha, T_\mathrm{c}, K) = (5, 6, 11)$, and $\beta = 3$ which results in $(\alpha, T_\mathrm{c}, K) = (8, 9, 17)$. These schemes are compared against the TDMA and "isolated 2-cell+TDMA" (with $M_r = 3$, (Suh *et al.*, 2011)) schemes in items 1) and 2) above, respectively. Figure 5.3 illustrates how the improved DoF of the proposed schemes manifests itself in steeper slopes for the achievable rate curves at high SNRs.

5.6.4 Hot-Spots or Micro-cells in a Linear Arrangement Setting

The network that was considered in the previous section was a single-tier homogeneous linear network. This subsection describes how the proposed precoding scheme can effectively accommodate certain heterogeneous elements. In particular, we will discuss the case of micro-BSs inserted at the edge of the macro cells and the case of the addition of a "hot spot"; see Figure 5.2.

In Figure 5.2, a hot-spot is installed close to the BS in cell $2(L-1)$ and a micro-cell is deployed at the cell edge between cell $2(L-1)$ and $2L-1$. The micro-BS is assumed to be equipped with 2-sector antennas that ensure that the radiation pattern of each sector is directed towards one cell. Now, if the hot-spot and the sector of the micro-BS that is directed towards cell $2(L-1)$ use a precoder $\mathbf{\Phi}_{s1} = \mathbf{\Phi}_{hs} = \mathbf{\Gamma}_{\text{o}}^1 \otimes \mathbf{\Gamma}_{\text{o}}^2$, the users in cell $2(L-1)$ will not suffer from any additional interference due to the implementation of the hot spot or the micro-BS. This is due to the fact that the interfering signals will arrive in the same subspace as the interference arriving from the two macro-BSs adjacent to cell $2(L-1)$ due to the interference alignment properties of the proposed scheme. The interference that the users associated with the hot-spot or the micro-BS receive from BS $2(L-1)$ lies in a proper subspace of their received signals and hence can be projected out. These users will, however, suffer from interference from the odd indexed BS. That said, those BSs are often quite distant from the users associated with the hot-spot or the micro-BS and hence, as will be illustrated in Section 5.7, there will often be a significant range of SNRs over which these users can achieve performance gains before the performance saturates as a result of the interference.

5.6.5 Unbounded Hexagonal Network

This subsection examines the hexagonal arrangement of cells illustrated in Figure 5.4. In contrast to the linear case, each user can be viewed as suffering from 2 or 3 dominant sources of interference. A scenario with $M_r = M_t$ receive antennas is considered and

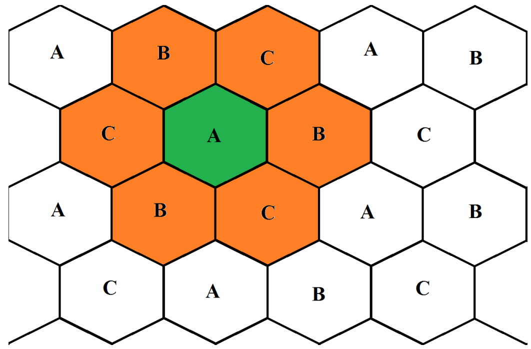

Figure 5.4: A spatial reuse precoding scheme for a hexagonal cell arrangement Cells are divided into three groups **A**, **B** and **C**. Each group implements the same structured precoder.

an SRP scheme is developed based on the Kronecker structured precoding scheme in (5.11), which was developed for the isolated 3-cell network. In particular, three inter-cell precoders, denoted $\mathbf{\Phi}_{\mathrm{A}}$, $\mathbf{\Phi}_{\mathrm{B}}$, and $\mathbf{\Phi}_{\mathrm{C}}$, are designed according to (5.11) and assign them to the BSs according to the pattern in Figure 5.4, which has an SRF of 3. The claim is that this precoding scheme enables users in a given cell to project out the interference from the 2 or 3 dominant sources of interference arriving from the neighboring cells, while preserving dimensions for communication with the assigned BS.

To verify that claim, consider the interference matrix for users in the shaded cell marked **A** in Figure 5.4 under an abstract partially-connected model in which only interference from the 3 neighboring cells. Further, consider a user whose dominant interferers are two BSs using **B** and one BS using **C**. If $\mathbf{H}_{\mathbf{B},i}^{\mathbf{A},k}$ denotes the channel from the ith BS that uses $\mathbf{\Phi}_{\mathrm{B}}$ to this user, and if $\mathbf{H}_{\mathbf{C},i}^{\mathbf{A},k}$ is defined analogously, the interference matrix can be written as

$$\mathbf{Z}^{\mathbf{A},k} = \begin{bmatrix} \mathbf{H}_{\mathbf{B},1}^{\mathbf{A},k}\mathbf{\Phi}_{\mathrm{B}} & \mathbf{H}_{\mathbf{B},2}^{\mathbf{A},k}\mathbf{\Phi}_{\mathrm{B}} & \mathbf{H}_{\mathbf{C},1}^{\mathbf{A},k}\mathbf{\Phi}_{\mathrm{C}} \end{bmatrix}, \tag{5.17}$$

where $\mathbf{H}_{\mathbf{B},m}^{\mathbf{A},k}\mathbf{\Phi}_{\mathrm{B}} = \mathbf{\Gamma}_{\mathrm{B}}^{1} \otimes \tilde{\mathbf{H}}_{\mathbf{B},m}^{\mathbf{A},k}\mathbf{\Gamma}_{\mathrm{B}}^{2}$ and $\mathbf{H}_{\mathbf{C},m}^{\mathbf{A},k}\mathbf{\Phi}_{\mathrm{C}} = \mathbf{\Gamma}_{\mathrm{C}}^{1} \otimes \tilde{\mathbf{H}}_{\mathbf{C},m}^{\mathbf{A},k}\mathbf{\Gamma}_{\mathrm{C}}^{2}$. If the receiver designs a vector $\mathbf{q}_1$ to lie in the null space of $\mathbf{\Gamma}_{\mathrm{B}}^{1}$, and $\mathbf{q}_2$ to lie in the null space of $\tilde{\mathbf{H}}_{\mathbf{C},1}^{\mathbf{A},k}\mathbf{\Gamma}_{\mathrm{C}}^{2}$, then the receive beamformer $\mathbf{w}^{\mathbf{A},k^{\dagger}} = \mathbf{q}_1^{\dagger} \otimes \mathbf{q}_2^{\dagger}$ eliminates the inter-cell interference, i.e., $\mathbf{w}^{\mathbf{A},k^{\dagger}}\mathbf{Z}^{\mathbf{A},k} = \mathbf{0}$. This is achieved without BS cooperation, and the users perform only local feedback so that the in-cell beamforming matrices $\mathbf{V}_j$ can be designed; see Section 5.2. (In the more densely connected networks, the other BSs will cause interference to users in this cell.) As was the case for the linear arrangement, the precoding scheme described in this section can also effectively accommodate certain heterogeneous network arrangements.

5.6.6 Variations on the Theme

The simple SRP schemes for unbounded networks that have been described in the subsections above have been based on direct extensions of the interference alignment schemes for small networks that were presented in Section 5.3 and 5.5.2. As outlined in Section 5.2, one of the basic principles of the schemes for these small networks was for each transmitter to select its projection matrix $\mathbf{\Phi}_j$ in such a way that each receiver can cancel the inter-cell interference using a receive beamforming matrix, $\mathbf{W}^{i,k}$. By doing so, the receivers convert the network into isolated single cells with effective channel gains given by (5.2), and each BS designs its transmit beamformers $\mathbf{V}_j$ using conventional single-cell techniques. As illustrated in Figure 5.3, in the presence of significant interference from all interfering sources, these straightforward techniques provide tangible gains over conventional interference avoidance techniques, while maintaining essentially the same feedback and coordination requirements.

The choice to select the receive beamforming matrix so that all the interfering sources are projected out is based on insight from the analysis of the DoF of small networks, and is an appropriate strategy when all interfering sources generate significant interference. In typical cellular architectures the dynamic range of the powers of the interfering signals at a given receiver can be large, and at a typical operating SNR the requirement that the receivers eliminate all the inter-cell interference may not be necessary, or even desirable. As will be demonstrated in the simulations, the proposed approach can accommodate other receive beamforming strategies that seek a balance between inter-cell interference cancellation and the gain of the desired signal. A natural choice is the receive beamforming matrix that maximizes the SINR (e.g., Gomadam *et al.*, 2011),

$$\mathbf{W}^{i,k} \propto (\mathbf{Q}^{i,k})^{-1}\mathbf{H}_i^{i,k}\mathbf{\Phi}_i\mathbf{V}_i^k, \tag{5.18}$$

where $\mathbf{Q}^{i,k}$ is the interference plus noise covariance matrix at user (i,k). In the proposed system, the in-cell beamforming matrices $\mathbf{V}_j^k$ are designed after the effective channels are fed back to the assigned BSs and hence are not available when $\mathbf{W}^{i,k}$ is designed. Although an iterative design scheme along the lines of Gomadam *et al.* (2011) can be envisioned, the simulations suggest that a substantial performance gains can be obtained by determining $\mathbf{W}^{i,k}$ in (5.18) as if $\mathbf{V}_j$ was an identity matrix.

5.7 Simulation Results

This section evaluates the performance of the proposed schemes in the case of the homogeneous network with hexagonal arrangement of cells shown in Figure 5.4 and in the case of the heterogeneous network with linear arrangement of cells shown in

Table 5.1: Break points in piece-wise linear path loss model

Distance	$\alpha_{\mathbf{loss}}$
$\leq$ 200m	2
500m	3
2km	4
10km	5

Figure 5.2. The effect of the distance between any transmitter and any receiver is captured by a piece-wise linear path loss model (McCune and Feher, 1997; Oda *et al.*, 2000), where the path loss exponent α_{loss} varies with distance according to linear interpolation between points in Table 5.1.

5.7.1 Hexagonal Cell Arrangement Model

The first experiment examines the performance of the hexagonal arrangement of cells shown in Figure 5.4, with a cell radius of 500m. A model of 19 cells that wraps around itself is considered. The BSs and terminals each have four antennas, i.e., $M_t = M_r = 4$, and each receiver is sent a single data stream; i.e., $d^{i,k} = 1, \ \forall i, k$. In all the schemes that will be considered the BSs allocate the same power to each data stream; i.e., with the data symbols normalized so that they have unit average energy, the single column "matrices" in (4.6) have norm $\|\mathbf{T}_j^k\|_2 = \sqrt{\gamma_j/K}$. To ensure that the average power transmitted by each BS, $\mathrm{E}\{\mathbf{x}_j^H\mathbf{x}_j\}/T_{\mathrm{c}}$, is the same for all the considered schemes, the parameter γ_j is chosen to be PT_{c}. The achievable rate of users will be evaluated at different positions within their cells, as a function of P, where, as in (4.3), the receiver noise variance is normalized to 1.

The current experiment will compare the performance of the proposed scheme (prop), against schemes based on designs for isolated single cell (Gesbert *et al.*, 2010), 2-cell (Suh *et al.*, 2011) and 3-cell (Medra and Davidson, 2014) networks. In all of the considered networks, there is no cooperation between BSs and only local feedback is employed. In each cell, $K = 9$ users are served. In all of the considered schemes, intra-cell interference is cancelled by having the receivers feed back their effective channels (cf. (5.2)) and then choosing $\mathbf{V}_j$ to be the zero-forcing beamforming matrix (Spencer *et al.*, 2004b). (A subsequent "dedicated" training step enables receiver (i, k) to estimate $\mathbf{H}_j^{i,k}\mathbf{\Gamma}_j^k$ Caire *et al.* (cf. 2010). In the proposed SRP scheme, the precoding matrix $\mathbf{\Phi}_j$ is chosen according to (5.11), with $\beta = 3$ and the SRP pattern in Figure 5.4. This results in a block length of $T_\mathrm{c} = 4$. For the reasons outlined in Section 5.6.6, each receiver will employ the Max-SINR receive beamforming in (5.18).

As its name suggests, the scheme based on insight from the isolated single-cell case (1-cell) ignores (inter-cell) interference. The precoder $\mathbf{\Phi}$ is a random matrix of size $MT_{\mathrm{c,1\text{-}cell}} \times K$ and hence, a block length $T_{\mathrm{c,1\text{-}cell}} = 3$ is chosen to enable $K = 9$ users to be served in each cell. Here, there is no spatial reuse; each BS (randomly) chooses its precoder $\mathbf{\Phi}$ individually. Since the 1-cell design ignores the inter-cell interference, the receive beamformer $\mathbf{w}^{i,k}$ is chosen to be the matched filter, i.e., $\mathbf{w}^{i,k}$ is aligned with the left singular vector of $\mathbf{H}_i^{i,k}\mathbf{\Phi}_i$ that corresponds to the largest singular value.

In the scheme based on the isolated 2-cell case (2-cell), the precoding matrix $\mathbf{\Phi}$ at each BS is chosen using the subspace interference alignment technique (Suh *et al.*, 2011); which was described in Section 5.5.1. For $K = 9$ users per cell, this results in block length $T_{\mathrm{c,2\text{-}cell}} = 3$. In an isolated 2-cell network, this choice of $\mathbf{\Phi}$ enables each receiver to project out the interference that it receives from the other BS. To extend that notion to the case of an unbounded network, each receiver chooses its receive beamformer $\mathbf{w}^{i,k}$ to project out the dominant interference source, i.e.,

$\mathbf{w}^{i,k} \in \mathcal{N}(\mathbf{H}_{j^*}^{i,k}\mathbf{\Phi}_{j^*})$, where j^* is the index of the dominant interfering BS, and $\mathcal{N}(\cdot)$ denotes the null-space. In practice this null-space can be obtained using principal component analysis of the signals received when BS i is turned off (cf. Krim and Viberg, 1996).

The scheme based on the isolated 3-cell case (3-cell) is developed in a similar way. The precoders are chosen using the subspace interference alignment scheme of Medra and Davidson (2014) with $T_{\mathrm{c,3\text{-}cell}} = 5$. In an isolated 3-cell network, this choice of $\mathbf{\Phi}$ enables each receiver to project out the interference from both the interfering BSs. Analogous to the 2-cell case, when bringing this design into an unbounded network, each receiver projects out the two dominant sources of interference.

In Figure 5.5, the achievable rates of three users in the network are compared under the four schemes described above. The first user (Ucenter) is located close to its serving base station, the second user (Umid) is located half way to the cell edge, and the third user (Uedge) is located at the cell edge (at the center of a face of a hexagon). Figure 5.5 shows that at high SNRs the proposed SRP scheme has a significant impact on the rates that can be achieved by the cell edge user. In particular, it provides 110% and 65% increases over the achievable rates of the schemes based on the 2-cell and 3-cell schemes, respectively, and more than a 10 fold increase over the 1-cell scheme (which does not manage interference). For the users at the cell center interference has a smaller impact, and the interference cancellation properties of the proposed SRP, which requires a block length of 4, are outweighed by the increased symbol rate enabled by the schemes based on the 1-cell and 2-cell designs. Those schemes are able to serve the 9 users per cell in a block of 3 channel uses. The isolated 2-cell design has a slight advantage over the 1-cell design in that it enables some interference mitigation

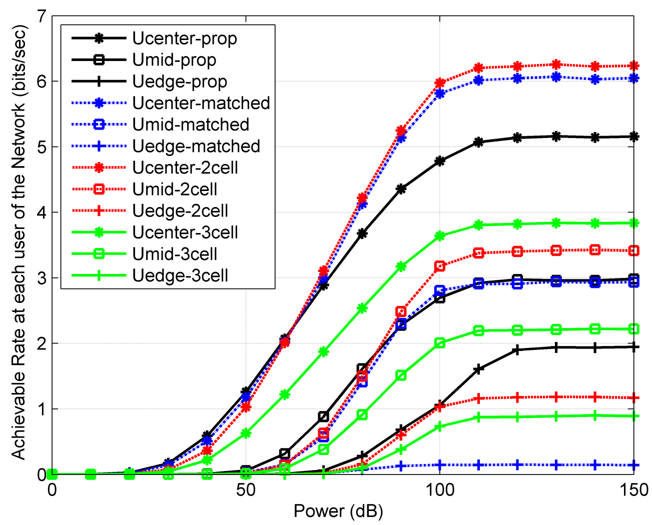

Figure 5.5: Achievable rates of various users in a hexagonal cellular network under four different signalling schemes.

and that scheme provides a 17% increase in the achievable rate of users close to the center over that provided by the proposed SRP scheme. That said, even for central users the proposed SRP scheme provides 32% increase over the scheme based on the 3-cell design.

5.7.2 A Heterogeneous Network Model

This subsection considers the heterogeneous network based on the linear arrangement of cells depicted in Figure 5.2. The network consists of macro-cells of radius 1.5km, plus micro-cells that are deployed mid-way between each pair of macro-cell BSs to provide service to the cell-edge users of the macro-cells. The micro-cell is modelled as having a radius of 400m and the micro BS is equipped with 2-sector antennas, where each radiation pattern is directed towards one cell. The transmitting power of each sector is 15 dB lower than the transmitting power of the macro-BS. Finally, a hot-spot is installed near BS $2(L-1)$ and its transmitting power is 30 dB lower than the transmitting power of the macro-BS. The hot-spot is assumed to serve users within a radius of 50m. The BSs and the hot-spot implement the precoding scheme in (5.14), as explained in Section 5.6.4. $\beta = 2$ is chosen, and, using Theorem 4, $\alpha = 5$ and $T_c = 6$. The number of users in the macro-cell is 11 users per cell, while the number of users per sector in the micro-cell is 6. Another modelling assumption is that the hot-spot serves up to 6 users.

In order to demonstrate the potential performance gains in terms of the achievable rates of users in the network, the current experiment simulates the performance of three users associated with each transmitting source, i.e., the macro-BS, the micro-BS and the hot-spot. These three users are located in the center, mid-way and boundary locations in the cell or covering area of each transmitting source, analogous to what was done in the hexagonal arrangement setting. The result of this experiment is illustrated in Figure 5.6

As can be seen from Figure 5.6, the users in the macro-cell are capable of removing the interference from the two adjacent BSs, the hot-spot and the micro-cell. They achieve improved performance as the transmitting power increases, until the interference from the next macro-BSs with the same precoder becomes significant, which is at powers beyond the scope of the graph. Although the users associated with the micro-cell or the hot-spot cannot remove all the interfering signals and their achievable rates saturate with increasing transmitting power, there is a wide range of transmitting powers before saturation takes place. The rate gains achieved over this range of powers are achieved without any sort of coordination between the existing macro-cell network and the micro-cell or the hot-spot and using only local feedback.

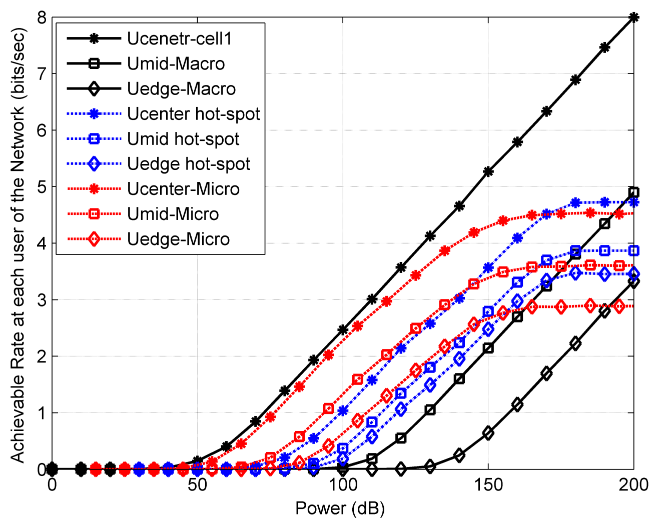

Figure 5.6: Achievable rates of various users in the heterogeneous linear model in Figure 5.2 under the proposed signalling scheme. The heterogeneous network consists of macro-cells, micro-cells and a hot-spot.

A key observation from Figure 5.6 is that the performance of the original macro-cells network was not altered as a result of the introduction of the micro-cells or the hot-spot, and these additional users can achieve substantial performance gains for a wide range of operating SNRs. Moreover, the presence of a micro-cell can provide improved performance for cell-edge users of the corresponding macro-cell. As seen in Figure 5.6, the achievable rate of any user associated with the micro-cell is higher than that of the cell-edge user of the macro-cell for a large range of transmitting powers before saturation.

5.8 Conclusion

In this chapter an approach to interference management in large structured networks has been developed. The approach is based on the observation that a linear time-invariant channel cannot expand the subspace spanned by a transmitted signal. This observation inspired the development of a notion of spatial reuse precoding in which the arrangement of the signal subspaces occupied by transmitters in a structured network is managed so that the dominant interfering subspaces align at each receiver. It has been shown that the alignment of the inter-cell interference provided by such a "spatial reuse precoding" scheme can be achieved without the need for inter-cell channel state information at the base stations, and it requires only a modest number of channel uses. By employing a two-level precoder structure in which the "inter-cell" precoder has a Kronecker structure, it has been shown how the receivers could eliminate the dominant components of the interference. By doing so they reduce the problem of designing the "intra-cell" precoder to the well-studied problem of downlink precoding for a single isolated cell. In addition to demonstrating the substantially increased rates that the proposed scheme can provide to cell edge users in single-tier homogeneous networks, it has also been illustrated how the principles that underlie spatial reuse precoding can be extended to certain classes of multi-tier heterogeneous networks.

Chapter 6

Summary and Future Work

6.1 Summary

This thesis developed approaches to problems that arise in the design and implementation of linear precoding schemes for linear downlink MIMO networks. As described in Chapter 1, these systems consist of a (quantized) feedback mechanism by which the transmitter or transmitters are informed of some of the properties of the channels to the receivers, and the linear precoding scheme. In point-to-point links and in single-cell downlink networks, the architecture of the feedback network is well established and the focus was on the design of the quantization scheme, and, in particular, on the design of the codebook for memoryless quantization and the design of an incremental quantization scheme that leverages codebooks design for the memoryless case. In multi-cell downlink networks, the design of the linear precoding scheme can reshape the architecture of the feedback network. The second half of the thesis focused on the development of precoding schemes for multi-cell downlink networks that provide improved performance over conventional schemes and yet requires only "local" feedback networks of the form considered in the first half of the thesis.

Chapter 2 discussed the problem of constructing quantization codebooks that can be used in memoryless quantization systems. In particular, smooth optimization objectives were developed that can be minimized using optimization techniques on the Grassmannian manifold. Chapter 2 started the presentation by designing unconstrained codebooks according to the chordal distance or the Fubini-Study distance. In several settings, it was shown that the designed codebooks are optimal when compared to a known bound on the minimum distance between any codewords in the codebook. Using the insights from the developed optimization objectives for the unconstrained codebooks, other objective functions were proposed in order to design more complicated codebooks that possess attractive features for practical implementation as constant modulus codebooks and finite alphabet codebooks. Further, an incremental method was proposed, inspired by the generation of mutually unbiased bases, to construct larger codebooks with elements selected from defined alphabet. The performance gains that can be achieved using the designed codebooks when implemented in limited feedback systems were shown to be significant in the simulations under different system models.

While the codebooks in Chapter 2 were designed to be used in memoryless quantization schemes, in Chapter 3 an incremental feedback scheme was proposed for temporally-correlated channels that utilizes these codebooks. The basic concept of the incremental scheme depends on moving from one point on the manifold to another point along the geodesic. Despite of the rather simple concept, the proposed incremental scheme exhibits many attractive features, one of which is the requirement of only one codebook for initialization and updates and this codebook can be designed by the methods described in Chapter 2. Furthermore, this scheme can recover autonomously from errors in the feedback path. In order to implement the incremental feedback scheme, two different approaches were proposed that differ in the way that the feedback budget is partitioned and the underlying channel model. In particular, the robust incremental scheme provided improved performance compared to some existing schemes in a variety of channel settings.

In Chapter 4, the thesis moved to the study of linear precoding schemes for multi-cell downlink networks. That chapter reviewed an intriguing approach to actively manage interference instead of avoiding it, namely Interference Alignment (IA). Then it proposed an IA linear precoding scheme that achieves the optimal spatial DoF for the isolated 2-cell MIMO downlink network. Although that scheme is suitable for feedback implementation, it requires several rounds of dedicated training. Further, it was concluded that increasing the spatial dimensions at the users resulted in a substantial reduction in the feedback requirements by enabling the use of a linear precoding scheme that requires only local feedback for operation. This suggested that it is preferable to design linear precoding schemes that can achieve improved DoF using only local feedback.

The extension from the isolated 2-cell downlink network to the general case of a G-cell MIMO interference broadcast channel (IBC) was presented in Chapter 5. In particular, Kronecker structured linear precoding schemes were designed that can achieve improved DoF when compared to the conventional interference avoidance schemes such as TDMA, while requiring only local feedback. Based on the insight from the developed schemes, the notion of spatial reuse precoding SRP was introduced. Using two different arrangements of cells, it was shown that precoding schemes

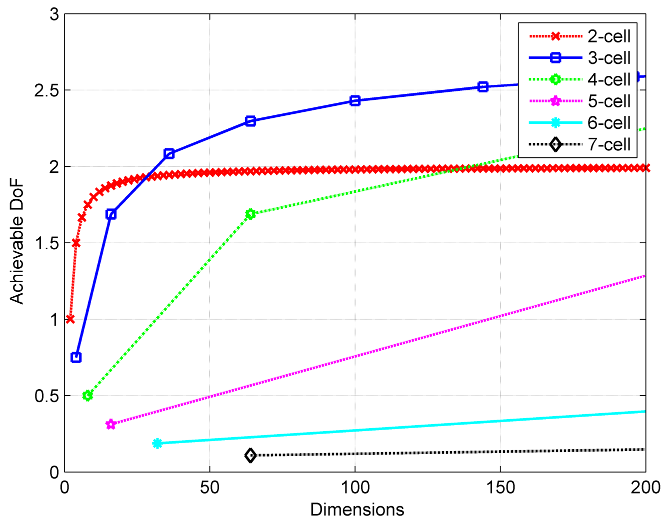

Figure 6.1: Achievable DoF for different values of G.

that exhibit the SRP can provide improved achievable rates compared to some existing schemes. Moreover, a case study was presented in which the precoders with SRP property can be implemented effectively in a heterogeneous network. The main outcome of this study is that users associated with different transmitting sources can achieve good performance in terms of achievable rate for a wide range of operating SNRs.

6.2 Future Work

This thesis has focused on designing linear precoding schemes that require only local feedback for operation. Although the structured precoding schemes in Chapter 5 may be sub-optimal in terms of DoF, they still achieve improved performance when compared to conventional interference avoidance schemes (which also require only local feedback). Further, those schemes highlighted the potential gains that can be achieved by precoding over multiple channel uses over which the channel is constant. In particular, the notion of spatial reuse precoding was introduced in which the same precoder can be used by multiple BSs in certain site arrangement of cells. A research direction that appears to be worthy of investigation is the implementation of a Kronecker-structured precoding schemes for different channel models and different assumptions. As one example, let us consider an OFDM based system model such as the one presented by Suh and Tse (2008) and a hexagonal arrangement of cells. For further simplification, let us assume that the physical channel between any transmitter and receiver consists only of one path, i.e., LOS channel, and the bandwidth is large enough to achieve the DoF of the network. Suh and Tse (2008) developed a precoding scheme for the uplink network that can achieve

$$\mathrm{DoF} = \frac{G\beta^{(G-1)}}{(\beta+1)^{(G-1)}} = \frac{GK}{(\sqrt[G-1]{K}+1)^{(G-1)}} \tag{6.1}$$

which approaches G as the number of users per cell K increases. Figure 6.1 illustrates the achievable DoF of the network for different values of G. Using the uplink-downlink network duality (Yu, 2006), let us assume that there exists a precoding scheme for the downlink network that can achieve the DoF in (6.1). Though it is clear that the precoding scheme can achieve the G-DoF of the network as the number of dimensions increases, the rate of convergence of the DoF to the optimal value significantly decreases by increasing G. For example, in the isolated 2-cell case, the precoding scheme can achieve a large fraction of the optimal 2-DoF using fewer dimensions when compared to the isolated 3-cell case. Now, for a hexagonal network with $G = 7$ cells, a question that may arise is that whether it is desirable to design the linear precoder assuming that $G = 7$ or not. In fact, if such a precoding scheme exists, it can designed for the isolated 7-cell, which is capable of providing interference free reception for every user in the entire network. However, as seen in Figure 6.1, this requires a very large number of dimensions, which may not be available. This is in addition to the required bandwidth scaling in order to achieve the target DoF.

An alternative approach that uses the insights from Chapter 5 is to design a Kronecker-structured linear precoder based on the assumption of $G = 3$ cells for an OFDM based system, then allow multiple BSs to reuse the same precoder, with a spatial reuse factor of 3, as had been done in the hexagonal arrangement of cells in Chapter 5. Several other factors should be also taken into account in that design, including

- Number of users per cell: A point that is worthy of investigation is the effect of a multi-user scheduler on the performance in terms of DoF and achievable sum rate.

- Network connectivity: In Chapters 5, the losses due to signal propagation were modelled using a piece-wise linear path loss model. Indeed, there are other models that need to be examined and may result in different conclusions on the amount of connectivity in the network. As one example, one may need to consider a model for an indoor path losses and examine how many dominant interfering sources are realized by a user.

- Effect of receive beamforming: In the case of point to point network, a user designs its receive beamformer in order to maximize the received signal power. In contrast, in a fully connected network at high SNRs, the user designs the receive beamformer in such a way to cancel the interference. A more practical case can be made when a user suffers from dominant sources of interference, but not all of them. This is the case of partially connected network, where it may be advantageous to design the receive beamformer to cancel only the dominant interfering sources while aiming at improving the received signal power.

- The site arrangement of cells: The topology of the cellular network need not to be hexagonal or linear, and, accordingly, some insights regarding the effect of the BS arrangement on the system performance have to be discussed.

- Heterogeneous networks: As demonstrated in Chapter 5, designing precoding schemes that allow the implementation of multi-tier networks without requiring significant resource exchange can provide better quality of service and coverage for cell-edge users. This is a topic that will likely require significant investigation.

- Power allocation: In the discussions and simulations in Chapters 5, one assumption is that uniform power loading is performed across all data streams. However, in practice, the power allocated for a user close to the base stations is likely to be smaller than the power allocated for a cell-edge user. Accordingly, applying different power assignment techniques can result in different performance gains for users located at different places in the cell and in the sum achievable rate of the network. Further, the interference pattern experienced by each user will be quite different compared to the uniform power loading case. This may allow for further improvement of the performance of cell-edge users.

- Fractional SRP: In an analogous way to fractional frequency reuse (Novlan *et al.*, 2011; Rahman and Yanikomeroglu, 2010), SRP can be designed in such a way that different structured precoders with different power loading can be assigned to users in different locations in the cell. This enables the base stations in the network in assigning higher power levels to cell-edge users, and, at the same time, the structure of the linear precoders enables these users to eliminate the dominant sources of interference.

Appendix A

Further Results for the Minimum Fubini-Study Distance of Unconstrained Codebooks

Table A.1 provides the minimum Fubini-Study distances of unconstrained codebooks that have been designed for dimensions for which there is no corresponding codebook in the catalogue of Love (2004).

Table A.1: Minimum Fubini-Study distances of unconstrained codebooks (continued from Table 2.1).

$M_t \times M$	N	Designed Codebook
4×3	8	1.1820
4×3	16	1.1070
4×3	32	0.8805
4×3	64	0.7413
6×2	8	1.5691
6×2	16	1.4812
6×2	32	1.3636
6×2	64	1.2986
8×3	8	1.5652
8×3	16	1.5420
8×3	32	1.5080
8×3	64	1.4628
8×4	8	1.5679
8×4	16	1.5410
8×4	32	1.5109
8×4	64	1.4703
10×2	8	1.5707
10×2	16	1.5468
10×2	32	1.5277
10×2	64	1.4708
10×3	8	1.5706
10×3	16	1.5624
10×3	32	1.5230
10×3	64	1.5002
12×2	8	1.5708
12×2	16	1.5600
12×2	32	1.5247
12×2	64	1.4945
16×2	8	1.5708
16×2	16	1.5684
16×2	32	1.5254
16×2	64	1.4979
16×3	8	1.5708
16×3	16	1.5704
16×3	32	1.5580
16×3	64	1.5394
16×4	8	1.5708
16×4	16	1.5708
16×4	32	1.5652
16×4	64	1.5546

Appendix B

Derivation of (3.9)

The derivation begins with the channel model in (3.6), $\mathbf{H}_n = \beta\mathbf{H}_{n-1} + \sqrt{1-\beta^2}\boldsymbol{\Theta}_n$, and the singular value decompositions $\mathbf{H}_n = \mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{V}_n^H$ and $\boldsymbol{\Theta}_n = \mathbf{Q}_n\boldsymbol{\Lambda}_n\mathbf{S}_n^H$. Using the analysis in the Appendix of Kim *et al.* (2011a), it can be shown that when M is chosen to be equal to M_r,

$$\begin{aligned}
&\mathrm{E}\{||\mathbf{H}_n\mathbf{P}_{n-1}||_F^2\} \\
&= \mathrm{E}\{\beta^2||\mathbf{H}_{n-1}\mathbf{P}_{n-1}||_F^2 + (1-\beta^2)||\boldsymbol{\Theta}_n\mathbf{P}_{n-1}||_F^2\} && \text{(B.1a)} \\
&= \mathrm{E}\{\beta^2||\boldsymbol{\Sigma}_{n-1}\mathbf{V}_{n-1}^H\mathbf{P}_{n-1}||_F^2 + (1-\beta^2)||\boldsymbol{\Lambda}_n\mathbf{S}_n^H\mathbf{P}_{n-1}||_F^2\}. && \text{(B.1b)}
\end{aligned}$$

Furthermore, using the unitary invariance of the Frobenius norm

$$\mathrm{E}\{||\mathbf{H}_n\mathbf{V}_n||_F^2\} = \mathrm{E}\{||\mathbf{U}_n\boldsymbol{\Sigma}_n\mathbf{V}_n^H\mathbf{V}_n||_F^2\} = \mathrm{E}\{\mathrm{tr}(\boldsymbol{\Sigma}_n^2)\} \qquad \text{(B.2)}$$

Using the fact that (Kim *et al.*, 2011a)

$$\mathrm{E}\{\mathrm{tr}(\boldsymbol{\Sigma}_n^2)\} = \mathrm{E}\{\mathrm{tr}(\boldsymbol{\Sigma}_{n-1}^2)\} = \mathrm{E}\{\mathrm{tr}(\boldsymbol{\Lambda}_n^2)\}, \tag{B.3}$$

it can then be deduced that

$$\begin{aligned}
&\mathrm{E}\{||\mathbf{H}_n\mathbf{V}_n||_F^2\} - \mathrm{E}\{||\mathbf{H}_n\mathbf{P}_{n-1}||_F^2\} \\
&\quad = \mathrm{E}\{\mathrm{tr}(\boldsymbol{\Sigma}_n^2)\} - \mathrm{E}\{\beta^2||\boldsymbol{\Sigma}_{n-1}\mathbf{V}_{n-1}^H\mathbf{P}_{n-1}||_F^2 + (1-\beta^2)||\boldsymbol{\Lambda}_n\mathbf{S}_n^H\mathbf{P}_{n-1}||_F^2\} && \text{(B.4a)} \\
&\quad = \mathrm{E}\{\mathrm{tr}\left(\beta^2\boldsymbol{\Sigma}_{n-1}^2(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}) + (1-\beta^2)\boldsymbol{\Lambda}_n^2(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n)\right)\}. && \text{(B.4b)}
\end{aligned}$$

Using the properties of the trace operator and the fact that $||\mathbf{X}||_F^2 = \mathrm{tr}(\mathbf{X}^H\mathbf{X})$, it can be shown that

$$\begin{aligned}
&\mathrm{E}\{||\mathbf{H}_n\mathbf{V}_n||_F^2\} - \mathrm{E}\{||\mathbf{H}_n\mathbf{P}_{n-1}||_F^2\} \\
&\quad = \mathrm{E}\{\mathrm{tr}(\boldsymbol{\Sigma}_n^2)\} - \mathrm{E}\{||\boldsymbol{\Sigma}_n\mathbf{V}_n^H\mathbf{P}_{n-1}||_F^2\} && \text{(B.5a)} \\
&\quad = \mathrm{E}\{\mathrm{tr}\left(\boldsymbol{\Sigma}_n^2(\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n)\right)\}. && \text{(B.5b)}
\end{aligned}$$

Using the fact that the expectation and the trace operations are linear, by equating (B.4b) and (B.5b), then

$$\begin{aligned}
&\mathrm{tr}\Big(\mathrm{E}\{\boldsymbol{\Sigma}_n^2(\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n)\}\Big) \\
&\quad = \mathrm{tr}\left(\mathrm{E}\{\beta^2\boldsymbol{\Sigma}_{n-1}^2(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}) + (1-\beta^2)\boldsymbol{\Lambda}_n^2(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n)\}\right). && \text{(B.6)}
\end{aligned}$$

Furthermore, using the fact that the channel gains and directions are independent and by using (B.3), the relation in (B.6) can be rewritten as

$$\begin{aligned}
&\mathrm{tr}\Big(\mathrm{E}\{\boldsymbol{\Sigma}_n^2\}\,\mathrm{E}\{\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n\}\Big)\\
&\quad= \mathrm{tr}\,\Big(\mathrm{E}\{\boldsymbol{\Sigma}_{n-1}^2\}\,\mathrm{E}\{\beta^2(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1})\}\\
&\qquad\quad + (1-\beta^2)\,\mathrm{E}\{\boldsymbol{\Lambda}_n^2\}\,\mathrm{E}\{\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n\}\Big) \qquad \text{(B.7a)}\\
&\quad= \mathrm{tr}\,\Big(\mathrm{E}\{\boldsymbol{\Sigma}_n^2\}\,\mathrm{E}\{\beta^2\big(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}\big) + (1-\beta^2)\big(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n\big)\}\Big). \qquad \text{(B.7b)}
\end{aligned}$$

By defining

$$\mathbf{A} = \mathrm{E}\{\boldsymbol{\Sigma}_n^2\} \tag{B.8}$$

and

$$\begin{aligned}
\mathbf{B} = \mathrm{E}\{&\beta^2\big(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}\big) + (1-\beta^2)\big(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n\big)\\
&- \big(\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n\big)\},
\end{aligned} \tag{B.9}$$

and by moving the right hand side of (B.7) to the left hand side, the relation in (B.7) can be rewritten as $\mathrm{tr}(\mathbf{AB}) = 0$. Since $\mathbf{A}$ and $\mathbf{B}$ are symmetric, the quantity $\mathrm{tr}(\mathbf{AB})$ can be bounded by (Fang *et al.*, 1994)

$$\lambda_{\min}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}) \leq \mathrm{tr}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}) \tag{B.10}$$

where $\lambda_{\max}(\mathbf{A})$ is the greatest eigenvalue of the matrix $\mathbf{A}$ and $\lambda_{\min}(\mathbf{A})$ is the least eigenvalue. Under channel models in which $\mathrm{E}\{\boldsymbol{\Sigma}_n^2\}$ has full rank, such as the case of the model in (3.6), $\lambda_{\min}(\mathbf{A}) > 0$, and therefore the condition $\mathrm{tr}(\mathbf{AB}) = 0$ implies that

$\mathrm{tr}(\mathbf{B}) = 0$. Therefore,

$$\begin{aligned}
\mathrm{tr}(\mathbf{B}) = \mathrm{tr}\Big(\mathrm{E}\Big\{&\beta^2\big(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}\big) + (1-\beta^2)\big(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n\big) \\
&- \big(\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n\big)\Big\}\Big) \qquad \text{(B.11a)} \\
= \mathrm{E}\Big\{\mathrm{tr}\Big(&\beta^2\big(\mathbf{I}_M - \mathbf{V}_{n-1}^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_{n-1}\big) + (1-\beta^2)\big(\mathbf{I}_M - \mathbf{S}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{S}_n\big) \\
&- \big(\mathbf{I}_M - \mathbf{V}_n^H\mathbf{P}_{n-1}\mathbf{P}_{n-1}^H\mathbf{V}_n\big)\Big)\Big\} \qquad \text{(B.11b)} \\
= \beta^2\,\mathrm{E}\big\{&d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{V}_{n-1})\big\} + (1-\beta^2)\,\mathrm{E}\big\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{S}_n)\big\} - \mathrm{E}\big\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{V}_n)\big\} = 0 \\
&\qquad \text{(B.11c)}
\end{aligned}$$

This concludes the derivation of (3.9).

For the case in which M is chosen to be less than M_r, the corresponding steps yield an inequality in the form of $\mathrm{tr}(\bar{\mathbf{A}}\bar{\mathbf{B}}) \geq 0$, where $\bar{\mathbf{A}} = \mathrm{E}\{\bar{\mathbf{\Sigma}}_n^2\}$ and $\bar{\mathbf{B}}$ takes the form in (B.11) with $\mathbf{V}_n$ replaced by $\bar{\mathbf{V}}_n$. That implies that $\mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{V}_n)\} \leq \beta^2\,\mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{V}_{n-1})\} + (1-\beta^2)\,\mathrm{E}\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1},\mathbf{S}_n)\}$. Accordingly, the step size in the model-based scheme is designed based on the upper bound on the expected chordal distance between $(\mathbf{P}_{n-1},\mathbf{V}_n)$.

Appendix C

Approximation of $\mathrm{E}\left\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\right\}$

In order to find an expression for $\mathrm{E}\left\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\right\}$, a Voronoi region approximation akin to that of Choi *et al.* (2012) is employed. To do so, first define the spherical cap centered at $\mathbf{U}$ with radius r to be (Dai *et al.*, 2008):

$$\mathcal{S}_{\mathbf{U}}(r) = \{\mathbf{V} : d_{\mathrm{ch}}(\mathbf{U}, \mathbf{V}) < r; \mathbf{V} \in \mathbb{G}_{M_t,M}\}. \tag{C.12}$$

The volume of that spherical cap is:

$$\mathrm{Vol}\big(\mathcal{S}_{\mathbf{U}}(r)\big) = \mathcal{C}_{M_t,M} r^{2M(M_t-M)}, \tag{C.13}$$

where $\mathcal{C}_{M_t,M}$ was defined in (3.12).

The analysis begins with the point $\mathbf{P}_{n-2}$ on the manifold, and the set of all points on the manifold that can be obtained by taking a step of size t_{n-1} along a geodesic from that point. Using the approximate relationship between the geodesic and chordal distances in (3.8), one can represent that set of points by the spherical cap $\mathcal{S}_{\mathbf{P}_{n-2}}(\gamma t_{n-1})$. When updating $\mathbf{P}_{n-2}$ to $\mathbf{P}_{n-1}$ a step of size t_{n-1} is taken in the direction of one of the 2^B elements of the codebook $\mathcal{F}$. If those elements are modelled as being isotropically distributed, then implicitly $\mathcal{S}_{\mathbf{P}_{n-2}}(\gamma t_{n-1})$ can be partitioned into 2^B spherical caps of equal radii. If the (high-resolution) Voronoi approximation is employed such that those 2^B spherical caps cover $\mathcal{S}_{\mathbf{P}_{n-2}}(\gamma t_{n-1})$ without overlap, then each has a volume that is $1/2^B$ of the volume of $\mathcal{S}_{\mathbf{P}_{n-2}}(\gamma t_{n-1})$. Using (C.13), this means that the radius, r_n of each of the 2^B spherical caps is such that

$$2^B \mathcal{C}_{M_t,M} r_n^{2M(M_t-M)} = \mathcal{C}_{M_t,M} (\gamma t_{n-1})^{2M(M_t-M)}. \tag{C.14}$$

That is,

$$r_n^2 = (\gamma t_{n-1})^2 2^{\frac{-B}{M(M_t-M)}}. \tag{C.15}$$

Now, when actually taking the step from $\mathbf{P}_{n-2}$ the proposed scheme moves in the direction of $\mathbf{P}_{n-1}$, which is the centre of the one of the 2^B spherical caps that contains $\bar{\mathbf{V}}_{n-1}$. That is, $\bar{\mathbf{V}}_{n-1}$ lies in $\mathcal{S}_{\mathbf{P}_{n-1}}(r_n)$. Therefore, up to the accuracy of the above approximations, $\mathrm{E}\big\{d_{\mathrm{ch}}^2(\mathbf{P}_{n-1}, \bar{\mathbf{V}}_{n-1})\big\} = r_n^2 = (\gamma t_{n-1})^2 2^{\frac{-B}{M(M_t-M)}}$. In order to account for the errors incurred in those approximations, and in particular the errors incurred in the Voronoi approximation, a correction factor $\mu \in [0, 1]$ is applied to that result and the approximation is refined to

$$\mathrm{E}\big\{d_{\mathrm{ch}}^2(\mathbf{P}_n, \bar{\mathbf{V}}_n)\big\} \approx \mu (\gamma t_n)^2 2^{\frac{-B}{M(M_t-M)}}. \tag{C.16}$$

Although it appears to be difficult to obtain an analytic expression for μ, it depends on the number of codewords in the codebook and the dimension of the manifold and can be determined, off-line, using straightforward numerical techniques. In the simulations, it was found that for a codebook of 2^4 codewords, setting $0.9 < \mu < 0.95$ gives good performance. As the number of feedback bits increases, the accuracy of the Voronoi region approximation also increases, and hence μ approaches 1.

Appendix D

Achievable DoF for $N \geq 1$ Case, $M > N$

The number of scheduled users per cell cannot exceed the rank of the matrix $\mathbf{\Phi}_j$, which is $\beta M_t + \alpha$. The interference matrix for the kth user in cell 1, $\mathbf{Z}^{1,k}$ can be written as

$$\mathbf{Z}^{1,k} = \begin{bmatrix} \mathbf{Z}_2^{1,k} & \mathbf{Z}_3^{1,k} & \dots & \mathbf{Z}_G^{1,k} \end{bmatrix} \tag{D.17}$$

where $\mathbf{Z}_j^{1,k}$ is the interference matrix between user k in cell 1 and BS j. Further $\mathbf{Z}_2^{1,k}$ can expressed as

$$\mathbf{Z}_2^{1,k} = \left[(\mathbf{I}_{T_c} \otimes \tilde{\mathbf{H}}_2^{1,k})(\mathbf{\Gamma}_1^2 \otimes \mathbf{\Gamma}_2^2) \quad (\mathbf{I}_{T_c} \otimes \tilde{\mathbf{H}}_2^{1,k})\mathbf{\Theta}_2 \right] = \left[\mathbf{Z}_{21}^{1,k} \quad \mathbf{Z}_{22}^{1,k} \right] \tag{D.18}$$

where $\mathbf{Z}_{21}^{1,k} = \mathbf{\Gamma}_1^2 \otimes \hat{\mathbf{H}}_2^{1,k}$, $\hat{\mathbf{H}}_2^{1,k} = \tilde{\mathbf{H}}_2^{1,k}\mathbf{\Gamma}_2^2$ and $\mathbf{Z}_{22}^{1,k} = (\mathbf{I}_{T_c} \otimes \tilde{\mathbf{H}}_2^{1,k})\mathbf{\Theta}_2$. Since $\operatorname{rank}(\mathbf{Z}_2^{1,k}) \leq \operatorname{rank}(\mathbf{Z}_{21}^{1,k}) + \operatorname{rank}(\mathbf{Z}_{22}^{1,k}) \leq \operatorname{rank}(\mathbf{\Gamma}_1^2)\operatorname{rank}(\hat{\mathbf{H}}_2^{1,k}) + \alpha \leq \beta M_r + \alpha$, and the rank of $\mathbf{Z}_g^{1,k}$

can be bounded in an analogous way, the rank of $\mathbf{Z}^{1,k}$ is upper bounded by

$$\text{rank}(\mathbf{Z}^{1,k}) \leq (G-1)(\beta M_r + \alpha). \tag{D.19}$$

In order to allow each user to cancel the inter-cell interference, $\mathbf{Z}^{1,k}$ should be rank-deficient. Therefore, the number of channel uses T_c is lower bounded by

$$T_\text{c} \geq \big((G-1)(\beta M_r + \alpha) + 1\big)/M_r \tag{D.20}$$

If α is set to be $\alpha = \lfloor \frac{T_\text{c} M_r - (G-1)\beta M_r - 1}{G-1} \rfloor$ and account is taken of the fact that the number of users cannot exceed $\big(T_\text{c} - (G-1)\beta\big) M_t$ in order to have full rank effective channels (which is required to enable the design of the transmit beamformers $\mathbf{v}^{i,k}$), then

$$\big(T_\text{c} - (G-1)\beta\big) M_t = \beta M_t + \alpha \tag{D.21}$$

Therefore, the smallest T_c that can be chosen is

$$T_\text{c} = \left\lceil \frac{\beta(G-1)(GM_t - M_r) - 1}{(G-1)M_t - M_r} \right\rceil. \tag{D.22}$$

The DoF of the proposed scheme can then be calculated as $\frac{GK}{T_\text{c}}$. As T_c increases, the DoF approaches $G\big(\frac{M_r}{G-1} + \frac{((G-1)M_t - M_r)(M_t - M_r)}{(G-1)(GM_t - M_r)}\big)$.

Appendix E

Achievable DoF of the 3-cell MIMO IBC Scheme

The interference matrix at the kth user in cell 1 is

$$\mathbf{Z}^{1,k} = \left[\mathbf{H}_2^{1,k}\boldsymbol{\Phi}_2 \quad \mathbf{H}_3^{1,k}\boldsymbol{\Phi}_3\right] \tag{E.23}$$

and it can be shown that $\mathbf{Z}^{1,k}$ is rank deficient as follows. The transpose of $\mathbf{Z}^{1,k}$ can be written as

$$\mathbf{Z}^{1,k^T} = \begin{bmatrix} \boldsymbol{\Gamma}_2^{1^T} \otimes (\tilde{\mathbf{H}}_2^{1,k}\boldsymbol{\Gamma}_2^2)^T \\ \boldsymbol{\Gamma}_3^{1^T} \otimes (\tilde{\mathbf{H}}_3^{1,k}\boldsymbol{\Gamma}_3^2)^T \end{bmatrix}. \tag{E.24}$$

Let $\boldsymbol{\Gamma}_2^{1^T} = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^\dagger$ and $(\tilde{\mathbf{H}}_2^{1,1}\boldsymbol{\Gamma}_2^2)^T = \mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^\dagger$ be the singular value decompositions (SVDs) of $\boldsymbol{\Gamma}_2^{1^T}$ and $(\tilde{\mathbf{H}}_2^{1,1}\boldsymbol{\Gamma}_2^2)^T$, respectively. The matrices $\mathbf{U}_1$, $\mathbf{U}_2$, $\mathbf{V}_1$ and $\mathbf{V}_2$ are unitary matrices of dimensions $\beta \times \beta$, $(M_t - 1) \times (M_t - 1)$, $T_c \times T_c$ and $M_t \times M_t$, respectively. The matrix $\mathbf{S}_1$ is $\beta \times T_c$ "diagonal" matrix, of the form of $\mathbf{S} = [\mathrm{diag}(s_1, s_2, \ldots s_\beta) \ \mathbf{0}_{\beta\times(T_c-\beta)}]$, where $\mathrm{diag}(s_1, s_2, \ldots s_\beta)$ is a diagonal matrix with $(s_1, s_2, \ldots s_\beta)$ on its diagonal and $\mathbf{0}_{\beta\times(T_c-\beta)}$ is a $\beta \times (T_c - \beta)$ zero matrix. Similarly, the matrix is $(M_t - 1) \times M_t$ "diagonal" matrix, of the form of $\mathbf{S}_2 = [\mathrm{diag}(s_1, s_2, \ldots s_{(M_t-1)})\ \mathbf{0}_{(M_t-1)\times 1}]$.

Define ${\mathbf{\Gamma}_3^1}^T\mathbf{V}_1 = \mathbf{Q}_3\mathbf{R}_3$ and $(\tilde{\mathbf{H}}_3^{1,k}\mathbf{\Gamma}_3^2)^T\mathbf{V}_2 = \mathbf{Q}_4\mathbf{R}_4$ to be the QR decompositions of ${\mathbf{\Gamma}_3^1}^T\mathbf{V}_1$ and $(\tilde{\mathbf{H}}_3^{1,k}\mathbf{\Gamma}_3^2)^T\mathbf{V}_2$, respectively. The matrices $\mathbf{Q}_3$ and $\mathbf{Q}_4$ are unitary matrices of dimensions $\beta \times \beta$ and $(M_t - 1) \times (M_t - 1)$ respectively. The matrices $\mathbf{R}_3$ and $\mathbf{R}_4$ are upper triangular matrices of dimensions $\beta \times T_c$ and $(M_t - 1) \times M_t$, respectively. The matrix ${\mathbf{Z}^{1,k}}^T$ in (E.24) can be rewritten as

$$
\begin{aligned}
{\mathbf{Z}^{1,k}}^T &= \begin{bmatrix} \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^\dagger \otimes \mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^\dagger \\ \mathbf{Q}_3\mathbf{R}_3\mathbf{V}_1^\dagger \otimes \mathbf{Q}_4\mathbf{R}_4\mathbf{V}_2^\dagger \end{bmatrix} && \text{(E.25)} \\
&= \begin{bmatrix} (\mathbf{U}_1 \otimes \mathbf{U}_2)(\mathbf{S}_1 \otimes \mathbf{S}_2)(\mathbf{V}_1^\dagger \otimes \mathbf{V}_2^\dagger) \\ (\mathbf{Q}_3 \otimes \mathbf{Q}_4)(\mathbf{R}_3 \otimes \mathbf{R}_4)(\mathbf{V}_1^\dagger \otimes \mathbf{V}_2^\dagger) \end{bmatrix} && \text{(E.26)} \\
&= \begin{bmatrix} \mathbf{U}_1 \otimes \mathbf{U}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_3 \otimes \mathbf{Q}_4 \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 \otimes \mathbf{S}_2 \\ \mathbf{R}_3 \otimes \mathbf{R}_4 \end{bmatrix} \mathbf{V}_1^\dagger \otimes \mathbf{V}_2^\dagger && \text{(E.27)} \\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger && \text{(E.28)}
\end{aligned}
$$

where

$$
\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \otimes \mathbf{U}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_3 \otimes \mathbf{Q}_4 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{S}_1 \otimes \mathbf{S}_2 \\ \mathbf{R}_3 \otimes \mathbf{R}_4 \end{bmatrix}
$$

and $\mathbf{V}^\dagger = \mathbf{V}_1^\dagger \otimes \mathbf{V}_2^\dagger$. The matrix $\mathbf{U}$ has full rank since it is a block diagonal matrix whose diagonal entry is the Kronecker product of two unitary matrices. A similar argument can be made for the matrix $\mathbf{V}$. If $\mathbf{R}_3 = \mathbf{W}_3 + \mathbf{Y}_3$, where $\mathbf{W}_3 = [\mathbf{0}_{\beta\times\beta}\ \ \mathbf{L}_3]$ and $\mathbf{L}_3$ is a matrix consisting of the $(T_c - \beta)$ right most columns of the matrix $\mathbf{R}_3$, then $\mathbf{Y}_3 = [\mathbf{R}_{\beta\times\beta} \quad \mathbf{0}_{T_c\times(T_c-\beta)}]$. Further, if $\mathbf{R}_4 = \mathbf{W}_4 + \mathbf{Y}_4$, where $\mathbf{W}_4 = [\mathbf{0}_{(M_t-1)\times(M_t-1)} \quad \mathbf{L}_4]$ and $\mathbf{L}_4$ is the right most column of the matrix $\mathbf{R}_4$, then $\mathbf{Y}_4 = [\mathbf{R}_{(M_t-1)\times(M_t-1)} \quad \mathbf{0}_{(M_t-1)\times 1}]$. Since augmenting zeros to any matrix does not change its rank, zeros can be added to $\mathbf{R}_3$ and by definition to $\mathbf{W}_3$, $\mathbf{L}_3$, and $\mathbf{Y}_3$ to match the dimensions of $\mathbf{R}_4$ or vise versa. Now

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{S}_1 \otimes \mathbf{S}_2 \\ \mathbf{R}_3 \otimes \mathbf{R}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \otimes \mathbf{S}_2 \\ \mathbf{Y}_3 \otimes \mathbf{Y}_4 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_2 \end{bmatrix} \tag{E.29}$$

$$= \boldsymbol{\Sigma}_1 + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_2 \end{bmatrix} \tag{E.30}$$

where $\boldsymbol{\Sigma}_2 = \mathbf{W}_3 \otimes \mathbf{Y}_4 + \mathbf{Y}_3 \otimes \mathbf{W}_4 + \mathbf{W}_3 \otimes \mathbf{W}_4$. Due to the specific structure of the matrices $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{Y}_3$ and $\mathbf{Y}_4$, the matrix $\boldsymbol{\Sigma}_1$ has $(T_c M_t - \beta(M_t - 1))$ zero columns and its rank $\text{rank}(\boldsymbol{\Sigma}_1) \leq \beta(M_t - 1)$. Since $\mathbf{W}_3 \otimes \mathbf{R}_4 = \mathbf{W}_3 \otimes \mathbf{Y}_4 + \mathbf{W}_3 \otimes \mathbf{W}_4$, but $\text{rank}(\mathbf{W}_3 \otimes \mathbf{R}_4) = \text{rank}(\mathbf{W}_3 \otimes \mathbf{Y}_4) = (T_c - \beta)(M_t - 1)$, this implies that

$$\text{Span}(\mathbf{W}_3 \otimes \mathbf{W}_4) \subset \text{Span}(\mathbf{W}_3 \otimes \mathbf{Y}_4). \tag{E.31}$$

By using similar argument, $\text{Span}(\mathbf{W}_3 \otimes \mathbf{W}_4) \subset \text{Span}(\mathbf{Y}_3 \otimes \mathbf{W}_4)$ and the rank of the matrix $\boldsymbol{\Sigma}_2$ cannot exceed $(T_c - \beta)(M_t - 1) + \beta - (T_c - \beta)$. Since

$$\text{rank}(\boldsymbol{\Sigma}) \leq \text{rank}(\boldsymbol{\Sigma}_1) + \text{rank}(\boldsymbol{\Sigma}_2), \tag{E.32}$$

then

$$\operatorname{rank}(\mathbf{\Sigma}) \leq \beta(M_t-1)+(T_{\mathrm{c}}-\beta)(M_t-1)+\beta-(T_{\mathrm{c}}-\beta) \tag{E.33}$$

$$= T_{\mathrm{c}}M_t - 2(T_{\mathrm{c}}-\beta). \tag{E.34}$$

Finally, the rank of the interference matrix $\mathbf{Z}^{1,k^T}$

$$\operatorname{rank}(\mathbf{Z}^{1,k^T}) \leq \min\big(\operatorname{rank}(\mathbf{U}), \operatorname{rank}(\mathbf{\Sigma}), \operatorname{rank}(\mathbf{V})\big) \tag{E.35}$$

$$\leq \operatorname{rank}(\mathbf{\Sigma}) \tag{E.36}$$

$$\leq T_{\mathrm{c}}M_t - 2(T_{\mathrm{c}}-\beta). \tag{E.37}$$

For any $T_{\mathrm{c}} > \beta$, $\mathbf{Z}^{1,k}$ is rank deficient. For example, if T_{c} is chosen such that $T_{\mathrm{c}} = \beta+1$, the rank of $\mathbf{Z}^{i,k}$ is $T_{\mathrm{c}}M_t - 2$ and hence, it has a null space of dimension 2. Since $\mathbf{Z}^{i,k}$ is rank deficient, user (i,k) designs each receive beamformer $\mathbf{w}^{i,k}$ to lie in the null space of $\mathbf{Z}^{1,k}$ such that $\mathbf{w}^{i,k^\dagger}\mathbf{Z}^{i,k} = 0$.

Following the design of the equalizers, each user feeds back its effective channel to its BS in order for the BS to design the transmit beamformers $\mathbf{v}^{i,k}$ that will eliminate the intra-cell interference. In order to characterize the DoF of the 3-cell MIMO-IBC with $M_t = M_r$ antennas, consider the fact that the number of scheduled users per cell cannot exceed the rank of the matrix $\mathbf{\Phi}_j$, which is $\beta(M_t-1)$. Further, the interference matrix at any user $\mathbf{Z}^{i,k}$ is rank deficient and has a null space of dimension $2(T_{\mathrm{c}}-\beta)$. Hence to have a full rank effective channels from the users in order to design the transmit beamformers $\mathbf{v}^{i,k}$, the number of scheduled users cannot exceed $2(T_{\rm c}-\beta)(M_t-1)$. Therefore, one can simply deduce that

$$K=\beta(M_t-1)=2(T_{\rm c}-\beta)(M_t-1), \tag{E.38}$$

and $T_{\rm c}=\frac{3\beta}{2}$. The DoF of the proposed scheme is

$$\mathrm{DoF}=\frac{3K}{T_{\rm c}}=\frac{3(\beta(M_t-1))}{T_{\rm c}}=2(M_t-1) \tag{E.39}$$

This scheme can achieve $2(M_t-1)$ DoF using any finite number of channel uses $T_{\rm c}$.

Appendix F

Achievable DoF for Linear G-cell Model

To determine the DoF of the proposed precoding scheme for the linear G-cell MIMO IBC as in Figure 5.2, where each user experiences interference from the two adjacent BSs and assuming that the BSs implement the precoding scheme in (5.14), it can shown that for each user to be able to cancel the inter-cell interference, the number of channel uses T_c is lower bounded by $T_\mathrm{c} \geq (\beta M_r + 2\alpha + 1)/M_r$. Now, if α is chosento be $\alpha = \lfloor \frac{T_\mathrm{c} M_r - \beta M_r - 1}{2} \rfloor$ and by using the fact that the number of users cannot exceed $\big(T_\mathrm{c} - \beta\big) M_t$ in order that the matrix of effective channels has full rank, then $\big(T_\mathrm{c} - \beta\big) M_t = \beta M_t + \alpha$, which implies that

$$T_\mathrm{c} = \left\lceil \frac{\beta(4M_t - M_r) - 1}{2M_t - M_r} \right\rceil. \tag{F.40}$$

The DoF/cell of the proposed scheme can be calculated by

$$\text{DoF/cell} = \frac{K}{T_{\mathrm{c}}} = \frac{(\beta M_t + \alpha)}{T_{\mathrm{c}}}. \tag{F.41a}$$

As T_{c} increases, the DoF/cell approaches $\left(\frac{M_r}{2} + \frac{(2M_t - M_r)^2}{2(4M_t - M_r)}\right)$.

Bibliography

Agrawal, D., Richardson, T. J., and Urbanke, R. L. (2001). Multiple-antenna signal constellations for fading channels. *IEEE Trans. Inf. Theory*, **47**(6), 2618–2626.

Barg, A. and Nogin, D. Y. (2002). Bounds on packings of spheres in the Grassmann manifold. *IEEE Trans. Inf. Theory*, **48**(9), 2450–2454.

Bengtsson, M. and Ottersten, B. (2001). Optimal and suboptimal transmit beamforming. In L. C. Godara, editor, *Handbook of Antennas in Wireless Communications*, chapter 18. CRC Press.

Bertsekas, D. P. (1999). *Nonlinear Programming*. Athena Scientific.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bresler, G., Cartwright, D., and Tse, D. (2014). Feasibility of interference alignment for the MIMO interference channel. *IEEE Trans. Inf. Theory*, **60**(9), 5573–5586.

Cadambe, V. R. and Jafar, S. A. (2008). Interference alignment and degrees of freedom of the K-user interference channel. *IEEE Trans. Inf. Theory*, **54**(8), 3425–3441.

Cadambe, V. R. and Jafar, S. A. (2009). Interference alignment and the degrees of freedom of wireless X networks. *IEEE Trans. Inf. Theory*, **55**(9), 3893–3908.

Cadambe, V. R., Jafar, S. A., and Wang, C. (2010). Interference alignment with asymmetric complex signaling: Settling the Host-Madsen–Nosratinia conjecture. *IEEE Trans. Inf. Theory*, **56**(9), 4552–4565.

Caire, G. and Shamai, S. (2003). On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans. Inf. Theory*, **49**(7), 1691–1706.

Caire, G., Jindal, N., Kobayashi, M., and Ravindran, N. (2010). Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Trans. Inf. Theory*, **56**(6), 2845–2866.

Chen, J. and Lau, V. (2014). Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks. *IEEE J. Sel. Areas Commun.*, **32**(6), 1230–1238.

Choi, J., Clerckx, B., Lee, N., and Kim, G. (2012). A new design of polar-cap differential codebook for temporally/spatially correlated MISO channels. *IEEE Trans. Wireless Commun.*, **11**(2), 703–711.

Choi, L.-U. and Murch, R. D. (2004). A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach. *IEEE Trans. Wireless Commun.*, **3**(1), 20–24.

Clerckx, B., Zhou, Y., and Kim, S. (2008). Practical codebook design for limited feedback spatial multiplexing. In *Proc. IEEE Int. Conf. Commun.*, pages 3982–3987.

Conway, J. H., Hardin, R. H., and Sloane, N. J. A. (1996). Packing lines, planes, etc.: Packings in Grassmannian spaces. *Exper. Math.*, **5**, 139–159.

Dai, W., Liu, Y., and Rider, B. (2008). Quantization bounds on Grassmann manifolds and applications to MIMO communications. *IEEE Trans. Inf. Theory*, **54**(3), 1108–1123.

de Kerret, P. and Gesbert, D. (2013). CSI sharing strategies for transmitter cooperation in wireless networks. *IEEE Wireless Commun. Mag.*, **20**(1), 43–49.

Dhillon, I. S., Heath, R. W., Strohmer, T., and Tropp, J. A. (2008). Constructing packings in Grassmannian manifolds via alternating projection. *Exper. Math.*, **17**(1), 9–35.

Ding, P., Love, D. J., and Zoltowski, M. D. (2007). Multiple antenna broadcast channels with shape feedback and limited feedback. *IEEE Trans. Signal Process.*, **55**(7-1), 3417–3428.

Edelman, A., Arias, T., and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, **20**(2), 303–353.

El Ayach, O. and Heath, R. (2012). Grassmannian differential limited feedback for interference alignment. *IEEE Trans. Signal Process.*, **60**(12), 6481–6494.

Fang, Y., Loparo, K., and Feng, X. (1994). Inequalities for the trace of matrix product. *IEEE Trans. Autom. Control*, **39**(12), 2489–2490.

Gersho, A. and Gray, R. M. (1991). *Vector quantization and signal compression*. Kluwer.

Gesbert, D., Kountouris, M., Heath, R., Chae, C.-B., and Salzer, T. (2007). Shifting the MIMO paradigm. *IEEE Signal Process. Mag.*, **24**(5), 36–46.

Gesbert, D., Hanly, S., Huang, H., Shamai, S., Simeone, O., and Yu, W. (2010). Multi-cell MIMO cooperative networks: A new look at interference. *IEEE J. Sel. Areas Commun.*, **28**(9), 1380–1408.

Gohary, R. H. and Davidson, T. N. (2009). Noncoherent MIMO communication: Grassmannian constellations and efficient detection. *IEEE Trans. Inf. Theory*, **55**(3), 1176–1205.

Gohary, R. H. and Davidson, T. N. (2011). An explicit expression for the Newton direction on the complex Grassmann manifold. *IEEE Trans. Signal Process.*, **59**(3), 1303–1309.

Gomadam, K., Cadambe, V., and Jafar, S. (2011). A distributed numerical approach to interference alignment and applications to wireless interference networks. *IEEE Trans. Inf. Theory*, **57**(6), 3309–3322.

Gray, R. and Neuhoff, D. (1998). Quantization. *IEEE Trans. Inf. Theory*, **44**(6), 2325–2383.

Guillaud, M. and Gesbert, D. (2011). Interference alignment in partially connected interfering multiple-access and broadcast channels. In *IEEE Glob. Telecommun. Conf.*, pages 1–5.

Hanly, S. (1996). Capacity and power control in spread spectrum macrodiversity radio networks. *IEEE Trans. Commun.*, **44**(2), 247–256.

Heath, R. W., Wu, T., and Soong, A. C. K. (2009). Progressive refinement of beamforming vectors for high-resolution limited feedback. *EURASIP J. Adv. Signal Process*, **2009**, 6:1–6:13.

Henkel, O. (2005). Sphere-packing bounds in the Grassmann and Stiefel manifolds. *IEEE Trans. Inf. Theory*, **51**(10), 3445–3456.

Hochwald, B. M., Marzetta, T. L., Richardson, T. J., Sweldens, W., and Urbanke, R. (2000). Systematic design of unitary space-time constellations. *IEEE Trans. Inf. Theory*, **46**(6), 1962–1973.

IEEE (2008). IEEE P802.16m-2008 draft standard for local and metropolitan area network. *IEEE Standard 802.16m*.

Inoue, T. and Heath, R. (2011). Grassmannian predictive coding for limited feedback multiuser MIMO systems. In *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pages 3076–3079.

Inoue, T. and Heath Jr., R. W. (2009). Kerdock codes for limited feedback precoded MIMO systems. *IEEE Trans. Signal Process.*, **57**(9), 3711–3716.

Jafar, S. A. (2011). Interference alignment: A new look at signal dimensions in a communication network. *F&T Commun. Info. Theory*, **7**(1), 1–136.

Jafar, S. A. (2012a). Blind interference alignment. *IEEE J. Sel. Topics Signal Process.*, **6**(3), 216–227.

Jafar, S. A. (2012b). Elements of cellular blind interference alignment-aligned frequency reuse, wireless index coding and interference diversity. Available online at `http://arxiv.org/abs/1203.2384`.

Jafar, S. A. and Goldsmith, A. J. (2005). Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom. *IEEE Trans. Inf. Theory*, **51**(3), 848–857.

Jafar, S. A. and Shamai, S. (2008). Degrees of freedom region of the MIMO X channel. *IEEE Trans. Inf. Theory*, **54**(1), 151–170.

Jindal, N. (2006). MIMO broadcast channels with finite-rate feedback. *IEEE Trans. Inf. Theory*, **52**(11), 5045–5060.

Jindal, N. (2007). Antenna combining for the MIMO downlink channel. *IEEE Trans. Wireless Commun.*, **7**(10), 3834–3844.

Kaltenberger, F., Jiang, H., Guillaud, M., and Knopp, R. (2010). Relative channel reciprocity calibration in MIMO/TDD systems. In *Future Network and Mobile Summit*, pages 1–10.

Kammoun, I., Cipriano, A. M., and Belfiore, J.-C. (2007). Non-coherent codes over the Grassmannian. *IEEE Trans. Wireless Commun.*, **6**(10), 3657–3667.

Kim, J. S., Kim, H., Park, C. S., and Lee, K. B. (2006). On the performance of multiuser MIMO systems in WCDMA/HSDPA: Beamforming, feedback and user diversity. *IEICE Trans. Commun.*, **E89–B**(8), 2161–2169.

Kim, K., Kim, T., Love, D., and Kim, I. H. (2012). Differential feedback in codebook-based multiuser MIMO systems in slowly varying channels. *IEEE Trans. Commun.*, **60**(2), 578–588.

Kim, T., Love, D., Clerckx, B., and Kim, S. (2008). Differential rotation feedback

MIMO system for temporally correlated channels. In *IEEE Global Telecommun. Conf.*, pages 1–5.

Kim, T., Love, D., and Clerckx, B. (2010). Leveraging temporal correlation for limited feedback multiple antennas systems. In *IEEE Int. Conf. Acoustics Speech and Signal Process.*, pages 3422–3425.

Kim, T., Love, D., and Clerckx, B. (2011a). MIMO systems with limited rate differential feedback in slowly varying channels. *IEEE Trans. Commun.*, **59**(4), 1175–1189.

Kim, T., Love, D. J., Clerckx, B., and Hwang, D. (2011b). Spatial degrees of freedom of the multicell MIMO multiple access channel. In *Proc. IEEE Global Telecommun. Conf.*, pages 1–5.

Kreutz-Delgado, K. (2009). The complex gradient operator and the $\mathbb{CR}$-calculus. Technical report, Dept. Elec. & Comp. Engineering, University of California. Available at: `http://arxiv.org/abs/0906.4835`.

Krim, H. and Viberg, M. (1996). Two decades of array signal processing research: The parametric approach. *IEEE Trans. Signal Process.*, **13**(4), 67–94.

Lee, H.-H. and Ko, Y.-C. (2012). Interference alignment with random vector quantization for MIMO interference channels. In *Proc. IEEE Veh. Tech. Conf.*, pages 1–5.

Lee, J. and Jindal, N. (2007). High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding. *IEEE Trans. Inf. Theory*, **53**(12), 4787–4792.

Liu, L. and Jafarkhani, H. (2006). Novel transmit beamforming schemes for time-selective fading multiantenna systems. *IEEE Trans. Signal Process.*, **54**(12), 4767–4781.

Liu, T. and Yang, C. (2013). On the spatial degrees of freedom of asymmetric MIMO interference broadcast channel. Available online at `http://arxiv.org/pdf/1310.7311.pdf`.

Liu, T. and Yang, C. (2014). On the degrees of freedom of partially-connected symmetrically-configured MIMO interference broadcast channels. In *IEEE Int. Conf. Acoustics, Speech, Signal Process.*, pages 1145–1149.

Love, D. J. (2004). Grassmannian subspace packing webpage. [Online]: `https://engineering.purdue.edu/~djlove/grass.html`.

Love, D. J. and Heath, Jr., R. W. (2005a). Limited feedback unitary precoding for orthogonal space-time block codes. *IEEE Trans. Signal Process.*, **53**(1), 64–73.

Love, D. J. and Heath, Jr., R. W. (2005b). Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Trans. Inf. Theory*, **51**(8), 2967–2976.

Love, D. J. and Heath Jr., R. W. (2005). Multimode precoding for MIMO wireless systems. *IEEE Trans. Signal Process.*, **53**(10), 3674–3687.

Love, D. J., Heath, Jr., R. W., and Strohmer, T. (2003). Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Trans. Inf. Theory*, **49**(10), 2735–2747.

Love, D. J., Heath, Jr., R. W., Lau, V. K. N., Gesbert, D., Rao, B. D., and Andrews,

M. (2008). An overview of limited feedback in wireless communication systems. *IEEE J. Sel. Areas Commun.*, **26**(8), 1341–1365.

Lozano, A., Heath, R., and Andrews, J. (2013). Fundamental limits of cooperation. *IEEE Trans. Inf. Theory*, **59**(9), 5213–5226.

Maddah-Ali, M. A. and Tse, D. (2010). Completely stale transmitter channel state information is still very useful. *In Proc. 48th Annual Allerton Conf. Commun. , Control, Computing*, pages 1188–1195.

Maddah-Ali, M. A. and Tse, D. (2012). Completely stale transmitter channel state information is still very useful. *IEEE Trans. Inf. Theory*, **58**(7), 4418–4431.

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

Maleki, H., Jafar, S. A., and Shamai, S. (2012). Retrospective interference alignment over interference networks. *IEEE J. Sel. Topics Signal Process.*, **6**(3), 228–240.

Manton, J. H. (2002). Optimization algorithms exploiting unitary constraints. *IEEE Trans. Signal Process.*, **50**(3), 635–650.

McCune, E. and Feher, K. (1997). Closed-form propagation model combining one or more propagation constant segments. In *IEEE 47th Veh. Tech. Conf.*, volume 2, pages 1108–1112.

Medra, A. and Davidson, T. N. (2013). On the achievability of the degrees of freedom for the three-cell MIMO interfering broadcast channel with minimum spatial dimensions. In *13th Canadian Wkshp. Info. Theory*, pages 87–91.

Medra, A. and Davidson, T. N. (2014). Kronecker structured linear precoding with local feedback for the 3-cell downlink. In *Proc. IEEE Int. Wkshp Signal Process. Adv. Wireless Commun.*

Mondal, B., Thomas, T. A., and Harrison, M. (2007). Rank-independent codebook design from a quaternary alphabet. In *Conf. Rec. 41st Asilomar Conf. Signals, Systems, Computers*, pages 297–301.

Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Math. Program*, **103**, 127–152.

Novlan, T., Ganti, R., Ghosh, A., and Andrews, J. (2011). Analytical evaluation of fractional frequency reuse for OFDMA cellular networks. *IEEE Trans. Wireless Commun.*, **10**(12), 4294–4305.

Oda, Y., Tsunekawa, K., and Hatay, M. (2000). Advanced los path-loss model in microcellular mobile communications. *IEEE Trans. Veh. Technol.*, **49**(6), 2121–2125.

Park, S.-H. and Lee, I. (2009). Analysis of degrees of freedom of interfering MISO broadcast channels. In *IEEE Global Telecommun. Conf.*, pages 1–6.

Pitaval, R. and Tirkkonen, O. (2012). Incorporating Stiefel geometry in codebook design and selection for improved base station cooperation. In *Proc. 75th IEEE Veh. Tech. Conf.*, pages 1–5.

Pitaval, R. A., Määttänen, H. L., Schober, K., Tirkkonen, O., and Wichman, R. (2011a). Beamforming codebooks for two transmit antenna systems based on optimum Grassmannian packings. *IEEE Trans. Inf. Theory*, **57**(10), 6591–6602.

Pitaval, R.-A., Tirkkonen, O., and Blostein, S. (2011b). Low complexity MIMO precoding codebooks from Orthoplex packings. In *Proc. IEEE Int. Conf. Commun.*, pages 1–5.

Rahman, M. and Yanikomeroglu, H. (2010). Enhancing cell-edge performance: A downlink dynamic interference avoidance scheme with inter-cell coordination. *IEEE Trans. Wireless Commun.*, **9**(4), 1414–1425.

Rashid-Farrokhi, F., Tassiulas, L., and Liu, K. J. R. (1998). Joint optimal power control and beamforming in wireless networks using antenna arrays. *IEEE Trans. Commun.*, **46**(10), 1313–1324.

Ravindran, N. and Jindal, N. (2008). Limited feedback-based block diagonalization for the MIMO broadcast channel. *IEEE J. Sel. Areas Commun.*, **26**(8), 1473–1482.

Razaviyayn, M., Sanjabi, M., and Luo, Z.-Q. (2012). Linear transceiver design for interference alignment: Complexity and computation. *IEEE Trans. Inf. Theory*, **58**(5), 2896–2910.

Roh, J. C. and Rao, B. D. (2006). Design and analysis of MIMO spatial multiplexing systems with quantized feedback. *IEEE Trans. Signal Process.*, **54**(8), 2874–2886.

Ryan, D. J., Clarkson, I. V. L., Collings, I. B., Guo, D., and Honig, M. L. (2009). QAM and PSK codebooks for limited feedback MIMO beamforming. *IEEE Trans. Commun.*, **57**, 1184–1196.

Schober, K., Janis, P., and Wichman, R. (2009). Geodesical codebook design for precoded MIMO systems. *IEEE Commun. Lett.*, **13**(10), 773–775.

Shenouda, M. and Davidson, T. (2007). Convex conic formulations of robust downlink precoder designs with quality of service constraints. *IEEE J. Sel. Topics Signal Process.*, **1**(4), 714–724.

Shenouda, M. and Davidson, T. (2008). On the design of linear transceivers for multiuser systems with channel uncertainty. *IEEE J. Sel. Areas Commun.*, **26**(6), 1015–1024.

Shin, W., Lee, N., Lim, J. B., Shin, C., and Jang, K. (2010). Interference alignment through user cooperation for two-cell MIMO interfering broadcast channels. In *Proc. IEEE GLOBECOM Workshops*, pages 120–125.

Spencer, Q., Peel, C., Swindlehurst, A., and Haardt, M. (2004a). An introduction to the multi-user MIMO downlink. *IEEE Commun. Mag.*, **42**(10), 60–67.

Spencer, Q., Swindlehurst, A., and Haardt, M. (2004b). Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels. *IEEE Trans. Signal Process.*, **52**(2), 461–471.

Suh, C. and Tse, D. (2008). Interference alignment for cellular networks. In *Proc. 46th Ann. Allerton Conf. Commun., Control, Computing*, pages 1037–1044.

Suh, C., Ho, M., and Tse, D. (2011). Downlink interference alignment. *IEEE Trans. Commun.*, **59**(9), 2616–2626.

Telatar, I. E. (1999). Capacity of multi-antenna Gaussian channels. *European Trans. Telecommun.*, **10**, 585–595.

Tse, D. and Viswanath, P. (2005). *Fundamentals of Wireless Communication*. Cambridge University Press.

Tselniker, I., Nazarathy, M., and Orenstein, M. (2009). Mutually unbiased bases in 4, 8, and 16 dimensions generated by means of controlled-phase gates with application to entangled-photon QKD protocols. *IEEE J. Sel. Topics Quantum Electron.*, **15**(6), 1713–1723.

Xia, P. and Giannakis, G. B. (2006). Design and analysis of transmit-beamforming based on limited-rate feedback. *IEEE Trans. Signal Process.*, **54**(5), 1853–1863.

Yang, J. and Williams, D. (2007). Transmission subspace tracking for MIMO systems with low-rate feedback. *IEEE Trans. on Commun.*, **55**(8), 1629–1639.

Yetis, C., Gou, T., Jafar, S., and Kayran, A. (2010). On feasibility of interference alignment in MIMO interference networks. *IEEE Trans. Signal Process.*, **58**(9), 4771–4782.

Yu, W. (2006). Uplink-downlink duality via minimax duality. *IEEE Trans. Inf. Theory*, **52**(2), 361–374.

Yu, W. and Cioffi, J. (2006). Constant-power waterfilling: Performance bound and low-complexity implementation. *IEEE Trans. Commun.*, **54**(1), 23–28.

Zhuang, B., Berry, R. A., and Honig, M. L. (2011). Interference alignment in MIMO cellular networks. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pages 3356–3359.